Inter-Comparison of Multiple Gridded Precipitation Datasets over Different Climates at Global Scale

Wenyan Qi¹*, Shuhong Wang¹ and Jianlong Chen²

¹ School of Civil Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China; shhwang@lzu.edu.cn
² College of Earth and Environmental Science, Lanzhou University, Lanzhou 730070, China; chjianlong2023@lzu.edu.cn
* Correspondence: qwydwyyx@163.com

Abstract: Comprehensive evaluations of global precipitation datasets are imperative for gaining insights into their performance and potential applications. However, the existing evaluations of global precipitation datasets are often constrained by limitations regarding the datasets, specific regions, and hydrological models used for hydrologic predictions. The accuracy and hydrological utility of eight precipitation datasets (including two gauged-based, five reanalysis and one merged precipitation datasets) were evaluated on a daily timescale from 1982 to 2015 in this study by using 2404 rain gauges, 2508 catchments, and four lumped hydrological models under varying climatic conditions worldwide. Specifically, the characteristics of different datasets were first analyzed. The accuracy of precipitation datasets at the site and regional scale was then evaluated with daily observations from 2404 gauges and two high-resolution gridded gauge-interpolated regional datasets. The effectiveness of precipitation datasets in runoff simulation was then assessed by using 2058 catchments around the world in combination with four conceptual hydrological models. The results show that: (1) all precipitation datasets demonstrate proficiency in capturing the interannual variability of the annual mean precipitation, but with magnitudes deviating by up to 200 mm/year among the datasets; (2) the precipitation datasets directly incorporating daily gauge observations outperform the uncorrected precipitation datasets. The Climate Precipitation Center dataset (CPC), Global Precipitation Climatology Center dataset (GPCC) and multi-source weighted-ensemble precipitation V2 (MSWEP V2) can be considered the best option for most climate regions regarding the accuracy of precipitation datasets; (3) the performance of hydrological models driven by different datasets is climate dependent and is notably worse in arid regions (with median Kling–Gupta efficiency (KGE) ranging from 0.39 to 0.65) than in other regions. The MSWEP V2 posted a stable performance with the highest KGE and Nash–Sutcliffe Efficiency (NSE) values in most climate regions using various hydrological models.

Keywords: global gridded precipitation; hydrological models; climatic conditions; evaluation

1. Introduction

As a vital driver of the hydrological cycle, precipitation is essential in multiple fields, such as agriculture, environmental science, hydrological modeling and water resource management [1–3]. Accurate precipitation can be obtained from gauge observations. However, the uneven distribution of rain gauges across various regions poses a significant challenge in acquiring precise precipitation data. Therefore, global-scale gridded precipitation datasets with high spatio–temporal resolution are required [4]. To address these requirements, a number of gridded precipitation datasets on a global scale have been developed over the past several decades [5–7]. These datasets exhibit substantial variability with respect to data sources, spatial coverage and resolution, temporal span and resolution, among other factors [8–10]. According to data sources, these precipitation datasets fall into four categories: (1) gauge-based, (2) satellite-based, (3) reanalysis and (4) merged [8,11–13].
In brief, gauge-based precipitation datasets have been created using historical observations from ground stations, such as the Climate Research Unit dataset (CRU), the Precipitation Reconstruction over Land (PRECL), the Climate Precipitation Center dataset (CPC) and the Global Precipitation Climatology Center dataset (GPCC) [12,14,15]. Satellite-based precipitation datasets have been produced by using polar-orbiting passive microwave sensors and geosynchronous infrared sensors, including the Climate Hazards Group Infrared Precipitation with Station Data (CHIRPS), the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIAN), the Tropical Rainfall Measuring Mission (TRMM), and so on [16–18]. Reanalysis systems are innovative tools for estimating global precipitation by reintegrating irregular observations and models, e.g., the European Centre for Medium-Range Weather Forecast Reanalysis 5 (ERA5), the National Centers for Environmental Prediction–National Center (NCEP–NCAR and NCEP–DOE), the Japanese 55-year ReAnalysis (JRA55) and WATCH forcing data methodology applied to the ERA-Interim dataset (WFDEI) [19,20]. In order to obtain optimal precipitation analyses, various attempts have been made to merge different sources of information while leveraging the strengths of each method to overcome associated issues, for instance, the Global Precipitation Climatology Project (GPCP), the CPC Merged Analysis of Precipitation (CMAP), and the multi-source weighted-ensemble precipitation V2 (MSWEP V2) [12,20,21].

Precipitation datasets are subject to various uncertainties, including quality control schemes, errors from data sources (gauge-based, satellite estimates, reanalysis), estimation procedures and scaling effects [22–24]. Therefore, it is of great significance to perform an exhaustive assessment of extant precipitation datasets to furnish a robust foundation for scientific research and practical application. Many studies have evaluated these datasets in the past. Some assessed accuracy by comparing with gauge observations assuming that they represent the real rainfall [25–29]. For example, Chen et al. [30] assessed three satellite-based precipitation datasets in China using rain gauges to analyze and assess their spatio–temporal accuracy features. They concluded that the spatial expression of precipitation datasets varies significantly across multiple basins, indicating noteworthy differences in performance. Iqbal et al. [31] evaluated the reliability of four daily precipitation datasets in monitoring changes in precipitation extremes. They concluded that most datasets indicated notable enhancements in extreme precipitation, but with substantial discrepancies in the estimation results among the different precipitation datasets. Lu et al. [32] assessed the accuracy of 12 gridded precipitation datasets in depicting extreme precipitation events. The analysis showed that these precipitation datasets effectively identified the significant increasing trends of extreme precipitation, but most datasets tend to overestimate the increasing trend, especially the satellite retrieval and reanalysis datasets.

Some have used hydrological models to analyze the effectiveness of precipitation datasets through the comparison of simulated and observed soil moisture [33,34] or river discharge [35–37]. For example, Pan et al. [33] tested the performance of precipitation datasets in predicting hydrologic states, including evapotranspiration, soil moisture and runoff. They found that inaccuracies in the precipitation datasets significantly affected the total column soil moisture. In addition, Alexopoulos et al. [38] assessed the performance of two precipitation datasets using the Génie Rural à 4 Paramètres Journalier model (GR4J) in 16 meso-scale catchments located in Slovenia. They found that precipitation datasets can serve as possible inputs for runoff simulations in regions with limited data. However, the two precipitation datasets tend to underestimate the rainfall extreme values. Sabbaghi et al. [39] compared the efficacy of three satellite precipitation datasets as inputs to the Hydrologic Engineering Center’s Hydrologic Modeling System. They concluded that these three satellite products provided unreliable and poor streamflow simulations. The advantages and potential of these precipitation datasets and their limitations have been widely considered. Marked differences were found among these datasets, revealing the significance of evaluating different precipitation datasets and their influence on runoff simulations to ascertain the most suitable choice for operational applications as well as research [12,40,41].
Numerous studies have considered the implications of different precipitation datasets on derived predictions of hydrological models. However, some considered only a single precipitation dataset or a single hydrological model, despite the uncertainties occurring in runoff simulation. In addition, many studies have been carried out only at catchment, regional and continental scales [39,42–46], rarely at global scale, and therefore it needs to be clarified as to what extent these findings can be applied. Only a few studies have a global focus. Fekete et al. [47] examined the impact of precipitation uncertainty on streamflow estimation at a global scale. However, this study only performed monthly simulations and did not corroborate the fidelity of their simulation with observations. Voisin et al. [48] assessed three precipitation datasets by utilizing observed streamflow derived from only nine very large catchments with an area of more than 290,000 km$^2$. Beck et al. [15] evaluated multiple global daily rainfall datasets in terms of discharge for more than 9000 basins over the globe by only using one hydrological model (HBV). Gebrechorkos et al. [49] evaluated six global precipitation datasets in streamflow modeling based on 1825 gauging stations, employing the WBMsed hydrological model. The current assessments of global precipitation datasets present a fragmented landscape, characterized by limitations in the number of precipitation datasets evaluated, the diversity of hydrological models employed, and the geographic specificity of regions investigated.

To address these concerns, we assess the accuracy and the impact of eight different precipitation datasets on water resource estimation by utilizing thousands of catchments in diverse climatic conditions around the world in combination with four conceptual hydrological models. Specifically, the objectives of this study are: (1) to compare precipitation characteristics of different precipitation datasets at temporal and spatial scales; (2) to assess the accuracy of precipitation datasets at both the point and regional scales through validation with in situ observations and high-resolution gridded gauge-interpolated datasets and verify whether they are affected by climate conditions; (3) to assess the influence of different precipitation inputs on the calibration and validation of hydrological models across 2058 catchments worldwide and how these effects differ with climate conditions and hydrological models. The thorough evaluation helped us to identify the accuracy and hydrological utility of different precipitation datasets under different climate conditions. This information holds significance in providing a valuable reference for readers in different research fields who require long-term precipitation data that better suits their needs.

2. Materials and Methods
2.1. Data Availability
2.1.1. Global Precipitation Datasets

Considering the spatial and temporal coverages, eight gridded precipitation datasets with a temporal span greater than 30 years (from 1982 to 2015) covering global land surface were evaluated in this study (Table 1). These include two gauge-based precipitation datasets (CPC and GPCC), five reanalysis datasets (ERA5, NCEP–NCAR, NCEP–DOE, JRA55 an WFDEI), and one satellite-gauge-reanalysis dataset (MSWEP V2). The CPC is operated by the National Oceanic and Atmospheric Administration. It integrates data from over 30,000 stations and covers the period from 1979 to the present [9]. The GPCC is available from 1981 to 2016. The data sources used in developing this dataset house more than 50,000 weather stations around the world [6,14]. The ERA5 is a reanalysis product developed by ECMWF using the 4D-Var data assimilation techniques and is available from 1950 to present [50]. The NCEP–NCAR [51] is developed using a fixed assimilation system and is available for the period from 1948 to the present. The NCEP–DOE is an improved version of the NCEP–NCAR, which fixed errors and updated parametrization of physical processes [52]. The MSWEP V2 [20] is based on the weighted averaging of several satellites, gauges, and reanalysis products and is available from 1979 to the present. The JRA55 is produced by the Japan Meteorological Agency (JMA) covering the period 1958 until the present [19]. In addition, the WFDEI is derived from the reanalysis of the European Centre for Medium-Range Weather Forecasts and is covers the period from 1979 to 2016 [7].
Table 1 presents the specifics of these eight gridded precipitation datasets. The evaluated precipitation datasets were classified into two categories, namely uncorrected precipitation datasets (UPDs) and corrected precipitation datasets (CPDs). Among the eight datasets, ERA5, JRA55, NCEP–NCAR and NCEP–DOE belong to UPDs, and CPC, GPCC, MSWEP and WFDEI belong to CPDs. To ensure a fair comparison in this study, GPCC and MSWEP were subjected to resampling at a resolution of 0.5° employing the bilinear averaging technique, given the different spatial resolutions. Concurrently, JRA55, NCEP–NCAR and NCEP–DOE were adjusted to a 0.5° resolution through the application of the inverse distance weighting interpolation method. These methods have been extensively utilized in prior studies and have shown a high level of effectiveness [11,15,53,54].

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Temporal Resolution</th>
<th>Spatial Resolution</th>
<th>Data Source</th>
<th>Category</th>
<th>Period</th>
<th>Reference</th>
<th>Download</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPC</td>
<td>Daily</td>
<td>0.5°</td>
<td>Gauged-based</td>
<td>CPDs</td>
<td>1979–present</td>
<td>Chen et al. [9]</td>
<td><a href="https://psl.noaa.gov/data/gridded/data.cpc.globalprecip.html">https://psl.noaa.gov/data/gridded/data.cpc.globalprecip.html</a>, accessed on 1 April 2024</td>
</tr>
<tr>
<td>ERA5</td>
<td>Daily</td>
<td>0.5°</td>
<td>Reanalysis</td>
<td>UPDs</td>
<td>1950–present</td>
<td>Hersbach et al. [50]</td>
<td><a href="https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-complete?tab=form">https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-complete?tab=form</a>, accessed on 1 April 2024</td>
</tr>
<tr>
<td>NCEP–NCAR</td>
<td>Daily</td>
<td>1.875°</td>
<td>Reanalysis</td>
<td>UPDs</td>
<td>1948–present</td>
<td>Kalnay et al. [51]</td>
<td><a href="https://psl.noaa.gov/data/gridded/data.ncep.reanalysis.html">https://psl.noaa.gov/data/gridded/data.ncep.reanalysis.html</a>, accessed on 1 April 2022</td>
</tr>
<tr>
<td>NCEP–DOE</td>
<td>Daily</td>
<td>1.875°</td>
<td>Reanalysis</td>
<td>UPDs</td>
<td>1979–present</td>
<td>Kanamitsu et al. [52]</td>
<td><a href="https://www.cpc.ncep.noaa.gov/products/wesley/reanalysis2/">https://www.cpc.ncep.noaa.gov/products/wesley/reanalysis2/</a>, accessed on 1 April 2022</td>
</tr>
<tr>
<td>MSWEP V2</td>
<td>Daily</td>
<td>0.1°</td>
<td>Satellite-gauge-reanalysis</td>
<td>CPPs</td>
<td>1979–present</td>
<td>Beck et al. [8]</td>
<td><a href="https://gloh2o.org/mswep/">https://gloh2o.org/mswep/</a>, accessed on 1 April 2024</td>
</tr>
<tr>
<td>WFDEI</td>
<td>Daily</td>
<td>0.5°</td>
<td>Reanalysis</td>
<td>CPPs</td>
<td>1979–2016</td>
<td>Weedon et al. [7]</td>
<td><a href="https://rda.ucar.edu/datasets/ds314.2/">https://rda.ucar.edu/datasets/ds314.2/</a>, accessed on 1 April 2024</td>
</tr>
</tbody>
</table>

2.1.2. Rain Gauge Observations

The daily rain gauge observations used in this paper were collected from the Global Summary of the Day database (https://data.noaa.gov, accessed on 1 April 2024) and the China Meteorological Administration website (http://data.cma.cn/, accessed on 1 April 2024). Two criteria were used in selecting suitable rain gauges: (1) gauges with available data from 1982 to 2015; (2) missing and invalid observations <1% of the period length. A total of 2404 gauges were selected and their distributions are shown in Figure 1a. In accordance with the Köppen–Geiger climate classification system, 2404 measurement
gauges are categorized into five primary climate categories: Equatorial (A), Arid (B), Warm Temperate (C), Snow (D), and Arctic (E) [55].

Figure 1. Location of the basins used in this study.

2.1.3. Other Meteorological Data

The reliability of these precipitation datasets underwent evaluation through a comparative analysis with the high-resolution gridded gauge-interpolated datasets in China (CN05) [56] and Europe (E-OBS) [57]. CN05 provides daily precipitation estimates spanning the years 1961 to the present, specifically targeting the quasi-China region. This dataset encompasses a broad latitudinal range from 54° N to 18° S, with a fine spatial resolution of 0.5 degrees (https://ccrc.iap.ac.cn/resource/detail?id=228, accessed on 1 April 2024). It is produced by integrating daily precipitation measurements from 2472 Chinese national weather gauges with mainland Digital Elevation Model data using the Thin Plate Spline algorithm [58]. E-OBS is a 0.5° gridded dataset providing daily precipitation for Europe from 1950 to 2018. E-OBS is generated by interpolating data from more than 11,000 meteorological stations after undergoing accurate pre-processing (https://www.ecad.eu/download/ensembles/download.php, accessed on 1 April 2024).

The 0.25° Global Land Evaporation Amsterdam Model potential evaporation dataset from 1980 to 2015 was also employed in the simulation of runoff (https://www.gleam.eu/, accessed on 1 April 2024). It employs the Priestley and Taylor equation for the accurate estimation of potential evapotranspiration, incorporating observations of surface net radiation along with near-surface air temperature measurements [34]. Those datasets possessing spatial resolutions finer than 0.5° were resampled to a 0.5° resolution utilizing the bilinear averaging method [53,54].
2.1.4. Observed Streamflow

The daily streamflow data were systematically collected from three primary sources: the Global Runoff Data Centre (GRDC, http://www.bafg.de/GRDC/, accessed on 1 April 2024) [59], the Canadian Model Parameter Experiment database [60] and several watersheds across China. To discard erroneous observation, only gauges with an area of 2500 km$^2$ to 50,000 km$^2$ and for at least 10 years (not necessarily consecutive) during 1982–2015 were retained. In addition, the Hydro1K was employed to define the basin’s boundary and flow paths utilizing the ArcGIS software 10.8 [54,61]. We have compared the catchment areas delineated by hydro1k with the areas provided by GRDC, retaining only catchments with an error of less than 5%. A total of 2058 catchments were chosen, as delineated in Figure 1b, which depicts the geographical distribution of these gauges. Consistent with rain gauges, these catchments have been classified into five main climate types.

2.2. Methodology

This section delineates the workflow for evaluating precipitation datasets (Figure 2), followed by details of the hydrological model and statistical analysis methods utilized in this study. All the methods considered in this study were implemented using MATLAB R2018b.

(1) The four UPDs were first evaluated on a daily timescale by comparing them with data from 2404 rain gauges worldwide.

(2) The performance of the four UPDs and four CPDs (see Table 1) was then evaluated on a daily timescale using two high-resolution gridded gauge-interpolated datasets in China (i.e., CN05) and Europe (i.e., E-OBS).

(3) The hydrological utility of eight precipitation datasets was subsequently assessed through hydrological modeling for 2058 catchments worldwide on a daily timescale.

Figure 2. The workflow for evaluating precipitation datasets used in this study.

2.2.1. Hydrological Models and Calibration

Numerous hydrological models have been devised and utilized for several purposes. Research conducted over the previous decade has shown that lumped and distributed
hydrological models have similar accuracy for streamflow simulation. Given the scope of this study, four conceptual hydrological models with computational efficiency were implemented to assess the efficacy of eight precipitation datasets for hydrological modeling. These models have garnered widespread application in hydrological research and have demonstrated efficacy globally. The first is the GR4J \cite{62,63}. It is a conceptual model that contains four parameters (see Table 2). It simulates the rainfall–runoff process using two storage reservoirs and a unit hydrograph-based routing scheme. In addition, the CemaNeige snow model \cite{64,65} was added to simulate snow processes. The second is the simple lumped conceptual daily rainfall–runoff model (SIMHYD) \cite{66,67}. SIMHYD features two linear reservoirs to calculate interflow and base flow, and a nonlinear reservoir for routing process. As with GR4J, the CemaNeige snow model was added to simulate snow processes. Therefore, it has 11 parameters in total. The third is the Xinanjiang model (XAJ) \cite{68–70}. The calculation process of XAJ comprises four modules: evaporation, runoff yield, runoff sources partition, and runoff concentration. As with GR4J, the CemaNeige snow model was added to simulate snow processes. Therefore, it has 17 parameters in total. The final is the hydrological model of École de Technologie Supérieure model (HMETS) \cite{71,72}. HMETS is a conceptual model with 21 calibration parameters. This model simulates snow accumulation, snowmelt and refreezing, evapotranspiration, infiltration, and flow routing.

The observed streamflow and simultaneous forcing data was delineated into two periods: the calibration period covering the initial 70% of the dataset, and the validation period including the subsequent 30%. The calibration process was executed employing the Shuffled Complex Evolution University of Arizona (SCE-UA) \cite{73,74} algorithm to refine hydrological model parameters. The optimization process was driven by the objective function of Kling–Gupta Efficiency (KGE) \cite{54,75}. In addition, all eight precipitations were employed as input to calibrate the four conceptual models against observed daily streamflow.

Table 2. Overview of the four hydrological models.

<table>
<thead>
<tr>
<th>Model</th>
<th>Snow Module</th>
<th>Simulated Processes (Number of Parameters)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>GR4J</td>
<td>CemaNeige</td>
<td>Flow routing (1) Snow modeling (2) Vertical budget (3)</td>
<td>Perrin et al. [62,76]</td>
</tr>
<tr>
<td>SIMHYD</td>
<td>CemaNeige</td>
<td>Flow routing (1) Snow modeling (2) Vertical budget (8)</td>
<td>Chiew et al. [66,67]</td>
</tr>
<tr>
<td>XAJ</td>
<td>CemaNeige</td>
<td>Flow routing (7) Snow modeling (2) Vertical budget (8)</td>
<td>Zhao et al. [68,69]</td>
</tr>
<tr>
<td>HMETS</td>
<td>HMETS</td>
<td>Evapotranspiration (1) Flow routing (4) Snow modeling (10) Vertical budget (6)</td>
<td>Martel et al. [71] and Chen et al. [72]</td>
</tr>
</tbody>
</table>

2.2.2. Statistical Analysis Methods

A series of statistical metrics was employed to assess the effectiveness of precipitation datasets in accurately capturing precipitation events and estimating runoff. Table 3 presents the nine metrics used for precipitation evaluation and the two metrics applied for hydrological assessment. Further elucidation of these metrics is provided in the subsequent sections.

The Maximum Percentage Difference (MPD) metric was employed to assess the discrepancies in estimating multi-annual average values among various precipitation datasets. The correlation coefficient (CC), the relative bias (BIAS) and the root mean squared error (RMSE) were adopted in this study to reflect the correlations and errors of precipitation
datasets. CC quantifies the level of consistency between the two series, and the ability of each product to describe the dynamic changes of precipitation. BIAS indicates the systemic error, denoting the relative divergence in the long-term average values between two comparative series. RMSE reflects the magnitude of discrepancies between two comparative series. The Probability of Detection (POD), False Alarm Ratio (FAR) and Critical Success Index (CSI) were utilized in this study to reflect the detection capability and accuracy. POD reflects the proportion of correctly identified precipitation events. Conversely, FAR delineates the proportion of precipitation events that were erroneously forecast. CSI combines the comprehensive features of POD and FAR, which considers accurate prediction, false alarms, and missed events. In this study, any precipitation event with a value greater than 1 mm/d is considered as a precipitation event [77,78].

Table 3. List of statistical indexes used in this study.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Unit</th>
<th>Formula</th>
<th>Perfect Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum Percentage Difference (MPD)</td>
<td>%</td>
<td>( \frac{\text{max}P - \text{min}P}{\text{mm}} )</td>
<td>0</td>
</tr>
<tr>
<td>Correlation Coefficient (CC)</td>
<td>NA</td>
<td>( \frac{\sum (G - G')(P - P')}{\sqrt{\sum (G - G')^2 \sum (P - P')^2}} )</td>
<td>1</td>
</tr>
<tr>
<td>Relative Bias (BIAS)</td>
<td>%</td>
<td>( \frac{P - G'}{\mu} )</td>
<td>0</td>
</tr>
<tr>
<td>Root Mean Square Error (RMSE)</td>
<td>mm</td>
<td>( \sqrt{\frac{\sum (G - P')^2}{n}} )</td>
<td>0</td>
</tr>
<tr>
<td>Critical Success Index (CSI)</td>
<td>NA</td>
<td>( \frac{a}{a + b + c} )</td>
<td>1</td>
</tr>
<tr>
<td>False Alarm Ratio (FAR)</td>
<td>NA</td>
<td>( \frac{b}{a + b + c} )</td>
<td>0</td>
</tr>
<tr>
<td>Probability of Detection (POD)</td>
<td>NA</td>
<td>( \frac{a}{a + b + c} )</td>
<td>1</td>
</tr>
<tr>
<td>PBias-Rx1</td>
<td>%</td>
<td>( \frac{P_{g,Rx1} - P_{s,Rx1}}{P_{s,Rx1}} )</td>
<td>0</td>
</tr>
<tr>
<td>PBias-Rx5</td>
<td>%</td>
<td>( \frac{P_{g,Rx5} - P_{s,Rx5}}{P_{s,Rx5}} )</td>
<td>0</td>
</tr>
<tr>
<td>Kling-Gupta Efficiency (KGE)</td>
<td>NA</td>
<td>( \text{KGE} = 1 - \sqrt{(R - 1)^2 + \left(\frac{\overline{O_{obs}} - 1}{\overline{O_{obs}}}\right)^2 + \left(\frac{\overline{C_{obs}} - 1}{\overline{C_{obs}}}\right)^2} )</td>
<td>1</td>
</tr>
<tr>
<td>Nash-Sutcliffe Efficiency (NSE)</td>
<td>NA</td>
<td>( \text{NSE} = 1 - \frac{\sum (Q_{s} - Q_{obs})^2}{\sum (Q_{s} - \overline{Q_{obs}})^2} )</td>
<td>1</td>
</tr>
</tbody>
</table>
3. Results

3.1. Discrepancies in Annual Precipitation

The spatial distributions of global long-term average annual precipitation of eight precipitation datasets are similar (Figure 3). The precipitation near the equator is the highest. The precipitation is significant on the east coast of the mainland due to the summer monsoons coming from the ocean. In addition, the west coast of the mainland around the Tropic of Capricorn and Cancer experiences a reduced amount of precipitation due to the subtropical high-pressure belt [81,82]. For example, West Asia and North Africa on the west coast of the mainland are dry and rainless in the prevailing northeast trade wind and the subsidence of air. However, the climatic conditions along the east coast of the mainland, exemplified by regions in East Asia, are humid and rainy due to the prevailing southerly winds from the ocean in summer. In addition, mid-latitudes and inland areas receive less precipitation as they are far from the sea, such as the northwestern region of China.

Figure 3. Long-term average annual total precipitation (time period 1982–2015), in mm/year.

Figure 4a shows the maximum differences in long-term average annual precipitation among eight precipitation datasets, which outlines the solid spatial variability. Regions with significant differences are the Intertropical Convergence Zone (ITCZ), northwestern
North America and southwestern South America. In regions like the ITCZ, minor alterations in the positioning of climatic features can result in substantial local differences in precipitation, due to large spatial gradients [83]. In addition, precipitation variations in coastal and mountainous regions exhibit more significant variability than in other areas. This phenomenon may be partly attributable to the intricate interplay between steep topographical elevations and the positioning of storm tracks [84]. Also of note is that, in general, the heavy rainfall in these areas may also result in more noticeable differences.

Regarding MPD (Figure 4b), regions with more than 100% MPD are primarily situated in North Africa, Greenland, and western Asia. Most of these regions belong to arid and Arctic regions and have few rain gauges due to complex terrain and low population. Regions with less than 40% of MPD are primarily concentrated in Europe, North America and South America. Most of these regions belong to equatorial, warm temperate and snow regions and have high-density rain gauges. In total, the MPDs between precipitation datasets are marginally higher in arid and Arctic regions in comparison to other regions [12,85]. This indicates that precipitation estimates for regions with sparse measurements remain challenging and largely uncertain, and we need to be prudent in selecting the precipitation datasets for these regions.

Figure 3. Long-term average annual total precipitation (time period 1982–2015), in mm/year.

Figure 4. (a) The spatial distribution of the maximum differences of long-term average annual precipitation between various precipitation datasets (mm/year); (b) The spatial distribution of the MPD of long-term average annual precipitation between various precipitation datasets (%).

The annual temporal evolution of global land-averaged precipitation is depicted in Figure 5. It can be seen that the interannual variability in the annual means is reasonably consistent among precipitation datasets, but with magnitudes deviating by up to
200 mm/year among the datasets. CPC estimates the lowest precipitation, which aligns with previous studies [12], followed by GPCC, WFDEI, MSWEP V2, NCEP–NCAR, NCEP–DOE, ERA5 and JRA55. In total, the annual global precipitation of the UPDs exceeds that of the CPPs, which supports the previous findings that reanalysis datasets tend to overestimate precipitation amounts [12,86]. This may be partly due to reanalysis datasets overestimating precipitation rates at higher elevations and precipitation amounts of small to medium-sized precipitation events [12,87]. The highest annual precipitation occurred in 2010 for all datasets, except for ERA5 (which occurred in 1986). The lowest precipitation occurred in 1987 for GPCC, MSWEP V2 and WFDEI, 1992 for CPC and NCEP–DOE, and 2002 for ERA5, NCEP–NCAR and JRA55. WFDEI and MSWEP V2 match well with the average of the eight precipitation datasets over the time series [88].

Figure 5. Temporal evolution of global land-averaged precipitation on annual timescales (time period 1982–2015).

3.2. Evaluation of the Precipitation Datasets’ Performance Based on Ground Precipitation Observations

3.2.1. The Performance of the Four UPDs Using Gauge Observations

The spatial distribution of the CC, CSI, FAR and POD between the gauge observations and four uncorrected precipitation datasets are elucidated in Figure 6. Four UPDs show a similar spatial pattern for different metrics. However, different metrics exhibit pronounced spatial variability worldwide. Among them, CC, CSI and POD exhibit pronounced spatial variability. CC, CSI and POD in East Asia, Europe and northern North America are good, indicating high precipitation detection capabilities in these regions. Towards northern South America, southwest Asia and North Africa, CC, CSI and POD decreased gradually. Regarding FAR, the performance in topographically complex regions is the worst (e.g., the Hindu Kush, Rocky and Andes Mountains), demonstrating the significant challenges associated with estimating precipitation in these areas.

The spatial configurations of the RMSE, BIAS, PBias-Rx1 and PBias-Rx5 comparing the gauge observations with four UPDs are depicted in Figure 7. The distribution characteristics of RMSE exhibit consistency with the distribution properties of the long-term average annual total precipitation, as illustrated in Figure 3. The RMSE is highest near the equator and significant on the mainland’s east coast. Towards both sides of the tropic of Capricorn and the west coast of the mainland (e.g., northern North America, southwest Asia and North Africa), RMSE decreased gradually. Regarding BIAS, four UPDs exhibited negative biases in arid regions, and positive biases in other regions. The BIAS in warm temperate climates was consistently lower in comparison to other climatic regions. In the context of PBias-Rx1 and PBias-Rx5, it was observed that four UPDs consistently demonstrated negative biases in most regions. This indicated that these precipitations tend to underestimate the severity of extreme precipitation events. Notably, this underestimation decreases as the temporal scale increases (i.e., from 1 to 5 days).
Figure 6. The CC (the first column—(a,e,i,m)), CSI (the second column—(b,f,j,n)), FAR (the third column—(c,g,k,o)) and POD (the last column—(d,h,l,p)) between the UPDs and observed precipitation data (during 1982–2015).
Figure 7. The RMSE (the first column—(a,e,i,m)) and Bias (the second column—(b,f,j,n)), PBias-Rx1 (the third column—(c,g,k,o)) and PBias-Rx5 (the last column—(d,h,l,p)) between the UPDs and observed precipitation data (during 1982–2015).
Figure 8 presents the boxplot of the eight metrics between the gauge observations and four uncorrected precipitation datasets regarding five Köppen–Geiger climate types. ERA5 outperforms other datasets in terms of CC, CSI, FAR, PBias-Rx1 and PBias-Rx5 for all climate types (with the highest median values in most regions). However, regarding BIAS, RMSE and POD, ERA5 no longer has an edge. Particularly with respect to RMSE, ERA5 underperforms against other precipitation datasets across most climate types. This indicates that ERA5 lacks the sensitivity required to identify small-sized rainfall events. ERA5 has the highest accuracy of Rx1 and Rx5 in most regions, except for the Arctic region (with median PBias-Rx1 ranging from −10.9% for the global land to −38.8% for the equatorial region). The PBias-Rx1 of the other three precipitation datasets are below −38.8%, which indicates the underestimation of extreme precipitation events. The underestimation of PBias-Rx5 is slight compared to PBias-Rx1. In particular, JRA55 performs better in warm temperate, snow, and Arctic regions regarding RMSE, while it performs well in all climate types regarding POD. However, JRA55 performs worse than other precipitation datasets for all climate types in terms of BIAS. In the evaluation of PBias-Rx1 and PBias-Rx5 metrics, JRA55 is positioned third in the ranking, significantly below the performance of ERA5, yet closely approaching the second-placed NCEP–DOE. This indicates that JRA55 effectively detects rainfall events but with significant quantitative errors. NCEP–NCAR and NCEP–DOE are slightly stronger than others in terms of BIAS. NCEP–NCAR performs better in equatorial, arid and warm temperate regions, while NCEP–DOE excels in snow and Arctic regions. Nevertheless, in other dimensions, their performance is comparatively inferior.

**Figure 8.** Boxplots of CC (a), BIAS (b), RMSE (c), CSI (d), FAR (e), POD (f), PBias-Rx1 (g) and PBias-Rx5 (h) between gauge-observed precipitation and four UPDs in different climate region. A, B, C, D and E represent the equatorial (A), arid (B), warm temperate (C), snow (D), and Arctic (E) region, respectively.
Generally, precipitation datasets exhibit lower accuracy in arid climates compared to other climate types, as measured by most of the metrics (except for BIAS and RMSE). The short-lived and localized rainfall behavior and the sub-cloud evaporation of falling rain bring some challenges in estimating precipitation for this region. Compared to other climate regions, the BIAS of four precipitation datasets is more considerable in the Arctic region and the RMSE of four precipitation datasets is more significant in the equatorial region. The high RMSE in the equatorial region is probably due to the high precipitation intensity, extended duration rainfall, and the large amount of precipitation. The high BIAS in the Arctic region is probably due to the gauge observations not being corrected for wind-induced under-catch at high latitudes [15].

3.2.2. Performance Evaluation Using Two Gridded Gauge-Interpolated Datasets

Figure 9 illustrates the daily performance of eight global precipitation datasets compared to the gridded gauge-interpolated datasets in China (CN05) and Europe (EOBS). The performance of each precipitation dataset across these two regions exhibits fundamental consistency when analyzed through various statistical measures (i.e., CC, BIAS, RMSE, CSI, FAR, POD, PBias-Rx5 and PBias-Rx1). GPCC presents a better CC, RMSE and CSI than the other seven precipitation datasets in both China and Europe, followed by CPC. CPC presents a better BIAS and FAR than the other seven precipitation datasets in both China and Europe, followed by GPCC and MSWEP V2. These indicate smaller systematic errors of gauged-based precipitation datasets in these two regions. The high BIAS and RMSE of the UPDs may be partly due to the significant overestimation of precipitation amounts, as shown in Figure 5. However, in terms of POD, JRA55 presents a better POD than the other seven precipitation datasets in both China and Europe. CPC has the most miniature POD. The ranking of the accuracy for extreme precipitation metrics (median PBias-Rx1) is MSWEP V2 (−4.8%), ERA5 (7.6%), WFDEI (−7.9%), CPC (−10.2%), GPCC (−32.1%), NCEP–DOE (−44.8%), JRA55 (−47.8%) and NCEP–NCAR (−49.6%) in China (CN05), and WFDEI (13.7%), CPC (16.1%), GPCC (−17.4%), ERA5 (23.6%), NCEP–DOE (−25.6%), MSWEP V2 (26.5%), JRA55 (−34.4%), and NCEP–NCAR (−38.7%) in Europe (E-OBS). Most precipitation datasets tend to underestimate Rx1day and Rx5day in China, except for ERA5. In Europe, ERA5, CPC, MSWEP V2 and WFDEI tend to overestimate Rx1day, Rx5day and GPCC, JRA55, NCEP-NSAR and NCEP–DOE tend to underestimate Rx1day, Rx5day. It could be concluded that CPC, WFDEI and GPCC show great accuracy as an extreme precipitation index. Overall, the corrected precipitation datasets (CPDs, GPCC, CPC, MSWEP V2 and WFDEI) perform slightly better than the UPDs (ERA5, JRA55, NCEP-NSAR and NCEP–DOE).

The Taylor diagram of different precipitation datasets in 0.5° grids of China and Europe using CN05 and EOBS as references has been plotted (Figure 10). In general, CPC is closer to the gridded gauge-interpolated datasets in China for most climate types (except for the arid region), followed by GPCC. In Europe, GPCC is closer to the gridded gauge-interpolated datasets for most climate types (except for the snow region). Generally, precipitation datasets perform relatively better in warm temperate and equatorial regions than in arid and snow regions. Compared to arid and snow regions, precipitation datasets in other regions have a smaller deviation from the gridded gauge-interpolated dataset. The difference in standard deviation between different precipitation datasets is more significant in the arid region, especially in China.
The Taylor diagram of different precipitation datasets in 0.5° grids of China and Europe, respectively.

Europe (EOBS) in terms of five Köppen–Geiger climate types (1982–2015). A, B, C, D and E represent the equatorial (A), arid (B), warm temperate (C), snow (D), and Arctic (E) region, respectively.

Figure 9. Boxplots of CC (a), BIAS (b), RMSE (c), CSI (d), FAR (e), POD (f), PBias-Rx1 (g) and PBias-Rx5 (h) between two high-resolution gridded gauge-interpolated datasets (CN05 and EOBS) and eight precipitation datasets.

Figure 10. Taylor diagram for comparison of different precipitation datasets in China (CN05) and Europe (EOBS) in terms of five Köppen–Geiger climate types (1982–2015). A, B, C, D and E represent the equatorial (A), arid (B), warm temperate (C), snow (D), and Arctic (E) region, respectively.
3.3. Evaluation of Precipitation Datasets’ Performance Based on Hydrological Modeling

The hydrological utilities of eight precipitation datasets have been assessed using four hydrological models across 2058 global catchments. Figure 11 shows the global maps with KGE values obtained from GR4J (see the Figures A1–A3 for Spatial patterns of SIMHYD, XAJ and HMETS) driven by eight precipitation datasets in both calibration (Figure 11a–h) and validation periods (Figure 11i–p). The precipitation datasets exhibit similar behavior across continental areas. That is, the KGE values for most catchments in southern China, eastern Canada, the United States, and along the Atlantic Coast of Europe exceed 0.8. The excellent model performances of those regions are likely to be due to the dense network of precipitation gauges and the high quality of precipitation data. However, catchments in areas such as the Andes (South America), northwest China, and the American tropics exhibit low KGE values due to the intricate topography and the absence of precipitation gauges.

Figure 12 presents the best performing precipitation datasets (i.e., with the highest calibration and validation KGE) for each catchment (see Figure A4 for Spatial patterns of NSE). The ranking of the best performed precipitation dataset in GR4J (i.e., the proportion of catchments the dataset exhibited optimal performance) is MSWEP V2 (33.1%), CPC (16.3%), GPCC (15.4%), JRA55 (12.1%), ERA5 (9.3%), WFDEI (7.8%), NCEP–NCAR (5.2%) and NCEP–DOE (0.8%) in the calibration period, and MSWEP V2 (27.2%), CPC (13.1%), GPCC (12.1%), JRA55 (9.4%), ERA5 (22.6%), WFDEI (6.9%), NCEP–NCAR (5.2%) and NCEP–DOE (3.5%) in the validation period. Comparable outcomes are achieved in other models. The MSWEP V2, in combination with various satellites, gauges, and reanalysis products, outperforms other precipitation datasets.

To further explore the impact of climatic conditions on the efficacy of precipitation datasets in runoff simulation affected by climatic conditions and whether the best performing precipitation datasets vary depending on the climate type, Table 4 shows the median KGE values of four hydrological models driven by eight precipitation datasets for different climate types. The graphic demonstration of the comparison results is shown in Figure 13 (see Figure A5 for NSE). The analysis reveals that, in comparison to the other four climatic classifications, the performance of these models is markedly inferior within arid climate regions (with median KGE ranging from 0.39 to 0.65), as evidenced across all precipitation datasets. The intricate hydrological behavior, characterized by extreme floods and droughts, along with the variability in meteorological and hydrological conditions, presents significant challenges to hydrological modeling in these regions [89–91]. The hydrological performance in warm temperate regions is best (with the median KGE values fluctuating between 0.47 and 0.84), which may be partly because of the presence of dense monitoring networks. In addition, MSWEP V2 demonstrates superior performance in comparison to others across most climate regions, with the exception of the Arctic region. In the Arctic region, the efficacy of MSWEP V2 exhibits no significant deviation from those obtained from optimal precipitation (with the median KGE values no more than 0.03). Overall, the MSWEP V2 model maintains a relatively stable performance across diverse climatic conditions. The four UPDs, especially the NCEP–DOE and NCEP–NCAR datasets, perform much worse than the other four CPDs in the equatorial, warm temperate and arid regions. However, in the snow and Arctic regions, the performance order changed. The gauged-based CPC demonstrates the poorest performance, as evidenced by the lowest KGE value. The NCEP–DOE and NCEP–NCAR show a substantial rise in these two regions, which is in line with previous evaluations [92].
Figure 11. Spatial distribution of KGE values obtained from GR4J driven by eight precipitation datasets in calibration (a–h) and validation periods (i–p).
Figure 12 presents the best performing precipitation datasets (i.e., with the highest calibration and validation KGE) for each catchment (see Figure A4 for Spatial patterns of NSE). The ranking of the best performed precipitation dataset in GR4J (i.e., the proportion of catchments the dataset exhibited optimal performance) is MSWEP V2 (33.1%), CPC (16.3%), GPCC (15.4%), JRA55 (12.1%), ERA5 (9.3%), WFDEI (7.8%), NCEP–NCAR (5.2%) and NCEP–DOE (0.8%) in the calibration period, and MSWEP V2 (27.2%), CPC (13.1%), GPCC (12.1%), JRA55 (9.4%), ERA5 (22.6%), WFDEI (6.9%), NCEP–NCAR (5.2%) and NCEP–DOE (3.5%) in the validation period. Comparable outcomes are achieved in other models. The MSWEP V2, in combination with various satellites, gauges, and reanalysis products, outperforms other precipitation datasets.

Table 4. Median calibration KGE scores for the four hydrological models driven by eight precipitation datasets. Values in bold represent the highest score in each group.

<table>
<thead>
<tr>
<th>Climate Type</th>
<th>All</th>
<th>A: Equatorial</th>
<th>B: Arid</th>
<th>C: Warm Temperate</th>
<th>D: Snow</th>
<th>E: Arctic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n = 2058)</td>
<td>(n = 248)</td>
<td>(n = 217)</td>
<td>(n = 665)</td>
<td>(n = 885)</td>
<td>(n = 43)</td>
</tr>
<tr>
<td>CPC</td>
<td>0.73</td>
<td>0.66</td>
<td>0.57</td>
<td>0.81</td>
<td>0.70</td>
<td>0.71</td>
</tr>
<tr>
<td>ERA5</td>
<td>0.72</td>
<td>0.65</td>
<td>0.49</td>
<td>0.74</td>
<td>0.77</td>
<td>0.81</td>
</tr>
<tr>
<td>GPCC</td>
<td>0.75</td>
<td>0.67</td>
<td>0.53</td>
<td>0.77</td>
<td>0.78</td>
<td>0.80</td>
</tr>
<tr>
<td>JRA55</td>
<td>0.67</td>
<td>0.62</td>
<td>0.46</td>
<td>0.65</td>
<td>0.74</td>
<td>0.84</td>
</tr>
<tr>
<td>NCEP–DOE</td>
<td>0.54</td>
<td>0.37</td>
<td>0.38</td>
<td>0.50</td>
<td>0.65</td>
<td>0.79</td>
</tr>
<tr>
<td>NCEP–NCAR</td>
<td>0.58</td>
<td>0.48</td>
<td>0.46</td>
<td>0.50</td>
<td>0.70</td>
<td>0.83</td>
</tr>
<tr>
<td>MSWEP V2</td>
<td>0.79</td>
<td>0.70</td>
<td>0.59</td>
<td>0.83</td>
<td>0.80</td>
<td>0.82</td>
</tr>
<tr>
<td>WFDEI</td>
<td>0.73</td>
<td>0.66</td>
<td>0.54</td>
<td>0.76</td>
<td>0.76</td>
<td>0.72</td>
</tr>
</tbody>
</table>

Figure 12. Spatial distribution of best performing precipitation datasets. For each catchment, color indicates which precipitation dataset produced the best KGE.
Table 4. Cont.

<table>
<thead>
<tr>
<th>Climate Type</th>
<th>All ((n = 2058))</th>
<th>A: Equatorial ((n = 248))</th>
<th>B: Arid ((n = 217))</th>
<th>C: Warm Temperate ((n = 665))</th>
<th>D: Snow ((n = 885))</th>
<th>E: Arctic ((n = 43))</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SIMHYD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPC</td>
<td>0.74</td>
<td>0.82</td>
<td>0.60</td>
<td>0.80</td>
<td>0.69</td>
<td>0.64</td>
</tr>
<tr>
<td>ERA5</td>
<td>0.78</td>
<td>0.78</td>
<td>0.61</td>
<td>0.77</td>
<td>0.81</td>
<td>0.85</td>
</tr>
<tr>
<td>GPCC</td>
<td>0.77</td>
<td>0.79</td>
<td>0.59</td>
<td>0.78</td>
<td>0.79</td>
<td>0.75</td>
</tr>
<tr>
<td>JRA55</td>
<td>0.71</td>
<td>0.71</td>
<td>0.52</td>
<td>0.68</td>
<td>0.77</td>
<td>0.85</td>
</tr>
<tr>
<td>NCEP–DOE</td>
<td>0.59</td>
<td>0.57</td>
<td>0.45</td>
<td>0.51</td>
<td>0.70</td>
<td>0.81</td>
</tr>
<tr>
<td>NCEP–NCAR</td>
<td>0.62</td>
<td>0.58</td>
<td>0.48</td>
<td>0.50</td>
<td>0.73</td>
<td>0.82</td>
</tr>
<tr>
<td>MSWEP V2</td>
<td><strong>0.83</strong></td>
<td><strong>0.84</strong></td>
<td><strong>0.65</strong></td>
<td><strong>0.84</strong></td>
<td><strong>0.83</strong></td>
<td><strong>0.85</strong></td>
</tr>
<tr>
<td>WFDEI</td>
<td>0.77</td>
<td>0.78</td>
<td>0.62</td>
<td>0.77</td>
<td>0.80</td>
<td>0.75</td>
</tr>
<tr>
<td><strong>XAJ</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPC</td>
<td>0.74</td>
<td>0.76</td>
<td>0.55</td>
<td>0.80</td>
<td>0.73</td>
<td>0.71</td>
</tr>
<tr>
<td>ERA5</td>
<td>0.74</td>
<td>0.73</td>
<td>0.49</td>
<td>0.74</td>
<td>0.78</td>
<td>0.85</td>
</tr>
<tr>
<td>GPCC</td>
<td>0.77</td>
<td>0.74</td>
<td>0.60</td>
<td>0.77</td>
<td>0.80</td>
<td>0.83</td>
</tr>
<tr>
<td>JRA55</td>
<td>0.68</td>
<td>0.67</td>
<td>0.42</td>
<td>0.66</td>
<td>0.73</td>
<td>0.84</td>
</tr>
<tr>
<td>NCEP–DOE</td>
<td>0.57</td>
<td>0.56</td>
<td>0.44</td>
<td>0.50</td>
<td>0.68</td>
<td>0.82</td>
</tr>
<tr>
<td>NCEP–NCAR</td>
<td>0.59</td>
<td>0.52</td>
<td>0.46</td>
<td>0.49</td>
<td>0.71</td>
<td>0.83</td>
</tr>
<tr>
<td>MSWEP V2</td>
<td><strong>0.82</strong></td>
<td><strong>0.78</strong></td>
<td><strong>0.62</strong></td>
<td><strong>0.84</strong></td>
<td><strong>0.84</strong></td>
<td><strong>0.88</strong></td>
</tr>
<tr>
<td>WFDEI</td>
<td>0.75</td>
<td>0.74</td>
<td>0.52</td>
<td>0.76</td>
<td>0.79</td>
<td>0.77</td>
</tr>
<tr>
<td><strong>HMETS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPC</td>
<td>0.75</td>
<td>0.81</td>
<td>0.48</td>
<td>0.80</td>
<td>0.71</td>
<td>0.75</td>
</tr>
<tr>
<td>ERA5</td>
<td>0.75</td>
<td>0.78</td>
<td>0.45</td>
<td>0.74</td>
<td>0.78</td>
<td><strong>0.83</strong></td>
</tr>
<tr>
<td>GPCC</td>
<td>0.75</td>
<td>0.79</td>
<td>0.52</td>
<td>0.75</td>
<td>0.77</td>
<td>0.80</td>
</tr>
<tr>
<td>JRA55</td>
<td>0.65</td>
<td>0.72</td>
<td>0.39</td>
<td>0.63</td>
<td>0.71</td>
<td>0.78</td>
</tr>
<tr>
<td>NCEP–DOE</td>
<td>0.56</td>
<td>0.59</td>
<td>0.39</td>
<td>0.48</td>
<td>0.66</td>
<td>0.77</td>
</tr>
<tr>
<td>NCEP–NCAR</td>
<td>0.58</td>
<td>0.52</td>
<td>0.41</td>
<td>0.47</td>
<td>0.68</td>
<td>0.76</td>
</tr>
<tr>
<td>MSWEP V2</td>
<td><strong>0.80</strong></td>
<td><strong>0.83</strong></td>
<td><strong>0.56</strong></td>
<td><strong>0.82</strong></td>
<td><strong>0.80</strong></td>
<td><strong>0.80</strong></td>
</tr>
<tr>
<td>WFDEI</td>
<td>0.75</td>
<td>0.78</td>
<td>0.49</td>
<td>0.76</td>
<td>0.78</td>
<td>0.79</td>
</tr>
</tbody>
</table>

Figure 13. Boxplots of KGE values of different precipitation datasets for both calibration period (Cal) and validation period (Val) in different climate regions ((a) GR4J, (b) SIMHYD, (c) XAJ, (d) HMETS). A, B, C, D and E represent the equatorial (A), arid (B), warm temperate (C), snow (D), and Arctic (E) region, respectively.
4. Discussion

Based on the findings of the present investigation, it is clear that there is no one precipitation dataset that excels in every aspect. This emphasizes the uncertainties and errors within existing precipitation datasets and the necessity for meticulous evaluation to ascertain their suitability for particular utilities and climate conditions. The uncertainties are usually induced by errors from data sources, estimation procedures, quality control schemes, and scaling effects. For satellite-based precipitation, the errors may stem from the retrieval algorithms, the sensor used and the sampling frequency. The observational constraints, assimilation techniques and parameterization strategies may lead to uncertainties in reanalysis datasets. Uncertainties in gauge-based precipitation can be due to the network density, the interpolation method and the inconsistent time series of gauges [8,12,50]. As found in the current study, precipitation datasets show poor performance in regions with sparse measurements.

The scaling effects also lead to uncertainty of precipitation, e.g., the spatial and temporal resolution. Behrangi et al. [93] quantified the errors attributed to temporal and spatial sampling of precipitation events and concluded that the error of precipitation datasets decreases with increasing spatial and time resolution. Peng et al. [78] found that, while precipitation products demonstrate a robust correlation with rain gauge measurements at the annual and seasonal scales, their accuracy decreases when evaluated on a monthly scale. In addition, the spatial discontinuities between the catchment scale and grid scale may lead to uncertainties when using gridded precipitation as input for runoff simulation. For example, the MSWEP V2 precipitation dataset, which excels in runoff simulation, does not demonstrate clear advantages in precipitation estimation accuracy, especially for extreme precipitation. High spatial and temporal resolution precipitation can offer more detailed precipitation patterns, thereby improving the reliability of precipitation estimations and reducing the uncertainties in hydrological utility [78]. The validation conducted here shows the imperative need for advancements in precipitation datasets. Enhanced spatial and temporal resolution combined with expanded temporal coverage is crucial for achieving these improvements.

The fidelity of various precipitation datasets in capturing extreme precipitation events exhibits regional variability. In this study, four uncorrected precipitations underestimate Rx1 and Rx5 in most climate regions. The results of Rx5 showed overall improvement compared to Rx1. The bias of Rx5 in most regions is within 50%. The degree of bias has improved, and the range of severe overestimates or underestimates has narrowed. Similar results were obtained when compared to CN05. However, ERA5, CPC, MSWEP V2, and WFDEI overestimated extreme precipitation with positive PBias-Rx1 and PBias-Rx5 compared with E-OBS. Our results provided obvious evidence that precipitation datasets show poor accuracy in extreme precipitation events, even though estimation accuracy improves as the time scale increases (i.e., from 1 to 5 days). The poor accuracy of the maximum 1-day precipitation amount makes it difficult to meet the needs of risk management. Extreme precipitation events can result in devastating consequences, such as storms, floods, mudslides, and other related disasters, presenting a significant risk to human life and property. Climate change significantly impacts the increase in extreme precipitation events. In the future, a significant proportion of meteorological stations may encounter more intense extreme precipitation events than have been recorded historically [79]. The non-stationarity of precipitation may result in previously gauged stations becoming data-sparse regions, increasing our dependence on precipitation products [94]. This serves as a reminder that it is essential to improve the ability to faithfully represent extreme events, rather than simply increasing the quantity of precipitation products.

5. Conclusions

This research undertook a thorough evaluation of eight global grid precipitation datasets (i.e., CPC, GPCC, ERA5, NCEP–NCAR, NCEP–DOE, MSWEP V2, JRA55 and WFDEI) from multiple aspects. The daily observations from 2404 gauges and two high-
resolution gridded gauge-interpolated datasets (i.e., CNO5 and E-OBS) were employed to assess the precipitation datasets at the site and regional scales. Subsequently, four conceptual hydrological models were employed to evaluate the efficacy of different precipitation datasets on hydrological modeling over 2058 catchments from five climate regions. The following conclusions can be made:

All precipitation datasets are capable of capturing the interannual variability observed in the mean annual precipitation values. The deviations in estimated annual precipitation were up to 200 mm/year among these precipitation datasets. The degree of variability in precipitation estimates also differed across regions. Regions with significant differences are the Intertropical convergence zone, northwestern North America and southwestern South America. Regions with more than 100% MPD are mainly distributed in North Africa, Greenland and western Asia, primarily arid and Arctic regions with sparse measurements compared to other regions.

The accuracy difference of these four uncorrected precipitation datasets at the gauge site scale varies according to statistical metrics across different climate regions. Generally, ERA5 surpasses others for all climate regions with the highest median CC, CSI and FAR and least bias of Rx1 and Rx5. Regionally, the gauged-based dataset GPCC is more correlated with CNO5 and E-OBS, closely followed by CPC and blended dataset MSWEP V2. Overall, the CPDs perform slightly better than the UPDs. The high BIAS and RMSE of the UPDs may be partly due to the significant overestimation of precipitation amounts. Notable disparities are observed across various climatic regions, with precipitation datasets exhibiting significantly inferior performance in arid regions than in the other four climate types for most metrics.

The differences among different datasets in hydrological simulations are significant. The datasets that incorporate daily gauge observations (MSWEP V2, CPC, GPCC and WFDEI) outperform others. The efficacy of hydrological modeling with different precipitation datasets is climate dependent, with lower median KGE values (ranging from 0.39 to 0.65) in arid regions compared to the other four climate types. The four UPDs perform much worse in the equatorial, arid and warm temperate regions. However, these UPDs demonstrate a marked improvement in performance in the snow and Arctic regions. Among the eight precipitation datasets, MSWEP V2 posted a stable performance with the highest KGE and NSE values in most climate regions.

Overall, the thorough evaluation investigates the accuracy and hydrological utility of different precipitation datasets under different climate conditions and highlights the strengths and limitations of individual datasets. This information holds significance in providing a valuable reference for readers where care is required when undertaking analyses. The study still has some limitations. For example, only precipitation datasets covering the global land surface and with a time span greater than 30 years are considered in this study. The latest precipitations with short temporal coverage are not included in the evaluation. Furthermore, only lumped hydrological models were used in this study, and distributed hydrological models were not considered. Therefore, further studies with various latest high-resolution precipitation datasets and distributed hydrological models should be conducted to generalize the conclusions of this study.

Author Contributions: Conceptualization, W.Q.; Data curation, W.Q. and J.C.; Formal analysis, W.Q., S.W. and J.C.; Funding acquisition, W.Q.; Investigation, W.Q., S.W. and J.C.; Methodology, W.Q. and S.W.; Supervision, W.Q.; Validation, W.Q. and J.C.; Visualization, W.Q. and J.C.; Writing—original draft, W.Q., S.W. and J.C.; Writing—review and editing, W.Q., S.W. and J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Science and Technology Program of Gansu Province, grant number 22JR11RA143 and the Young Scholars Science Foundation of Lanzhou Jiaotong University, grant number 1200061160.

Data Availability Statement: The datasets used in this study are available by writing to the authors.
Acknowledgments: The authors would like to acknowledge all of the organizations for providing the precipitation datasets and other data. The authors also thank the editors and reviewers for their valuable comments and thoughtful suggestions.

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

Abbreviations

Precipitation
CPC Climate Precipitation Center dataset
GPCCC Global Precipitation Climatology Center dataset
ERA5 European Centre for Medium-Range Weather Forecast Reanalysis 5
NCEP-NCAR National Centers for Environmental Prediction-National Center for Atmosphere Research
NCEP-DOE National Centers for Environmental Prediction-Department of Energy
WFDEI WATCH forcing data methodology is applied to ERA-Interim dataset
MSWEP V2 Multi-source weighted-ensemble precipitation V2
UPDs Uncorrected precipitation datasets
CPDs Corrected precipitation datasets

Köppen–Geiger climate classification
A Equatorial
B Arid
C Warm temperate
D Snow
E Arctic

Other Data
CN05 A high-quality gridded daily observation dataset over China
E-OBS European high-resolution gridded dataset
GRDC Global Runoff Data Centre

Hydrological model
GR4J Génie Rural à 4 Paramètres Journalier model
SIMHYD Simple lumped conceptual daily rainfall-runoff model
XAJ Xinanjiang model
HMETS Hydrological model of École de Technologie supérieure model

Criteria
KGE Kling–Gupta efficiency
NSE Nash–Sutcliffe Efficiency
MPD Maximum Percentage Difference
CC Correlation coefficient
BIAS Relative bias
RMSE Root mean squared error
POD Probability of Detection
FAR False Alarm Ratio
CSI Critical Success Index
Rx1 annual maximum 1-day precipitation amount
Rx5 annual maximum 5-day precipitation amount

Others
ITCZ Intertropical Convergence Zone
Figure A1. Spatial distribution of KGE values obtained from SIMHYD driven by eight precipitation datasets in calibration (a–h) and validation periods (i–p).
Figure A2. Spatial distribution of KGE values obtained from XAJ driven by eight precipitation datasets in calibration (a–h) and validation periods (i–p).
Figure A3. Spatial distribution of KGE values obtained from HMETS driven by eight precipitation datasets in calibration (a–h) and validation periods (i–p).
Figure A4. Spatial distribution of best performing precipitation datasets. For each catchment, color indicates which precipitation dataset produced the best NSE.
Figure A5. Boxplots of NSE values of different precipitation datasets for both calibration period (Cal) and validation period (Val) in different climate region ((a) GR4J, (b) SIMHYD, (c) XAJ, (d) HMETS). A, B, C, D and E represent the equatorial (A), arid (B), warm temperate (C), snow (D), and Arctic (E) region, respectively.

References


23. Tuo, Y.; Duan, Z.; Disse, M.; Chiogna, G. Evaluation of precipitation input for SWAT modeling in Alpine catchment: A case study in the Adige river basin (Italy). *Sci. Total Environ.* 2016, 573, 66–82. [CrossRef]


39. Sabbaghi, M.; Shahnazari, A.; Soleimaniyan, E. Evaluation of high-resolution precipitation products (CMORPH-CRT, PERSIANN, and TRMM-3B42RT) and their performances as inputs to the hydrological model. Model. Earth Syst. Environ. 2024, 10, 1–17. [CrossRef]

40. Maggioni, V.; Meyers, P.C.; Robinson, M.D. A review of merged high-resolution satellite precipitation product accuracy during the Tropical Rainfall Measurement Mission (TRMM) era. J. Hydrometeorol. 2016, 17, 1101–1117. [CrossRef]


44. Jiang, Q.; Li, W.; Fan, Z.; He, X.; Sun, W.; Chen, S.; Wen, J.; Gao, J.; Wang, J. Evaluation of the ERA5 reanalysis precipitation dataset over Chinese Mainland. J. Hydrol. 2021, 595, 125660. [CrossRef]


46. Araghi, A.; Adamowski, J.F. Assessment of 30 gridded precipitation datasets over different climates on a country scale. Earth Sci. Inform. 2024, 17, 1301–1313. [CrossRef]


Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.