

Article

Deep Learning-Based Multi-Source Precipitation Fusion and Its Utility for Hydrological Simulation

Zihao Huang¹, Changbo Jiang^{1,2,*}, Yuannan Long^{2,3}, Shixiong Yan^{1,2}, Yue Qi³, Munan Xu¹ and Tao Xiang¹¹ School of Civil and Environmental Engineering, Hunan University of Technology, Zhuzhou 412007, China² Key Laboratory of Dongting Lake Aquatic Eco-Environmental Control and Restoration of Hunan Province, Changsha 410114, China³ School of Hydraulic and Ocean Engineering, Changsha University of Science & Technology, Changsha 410114, China

* Correspondence: jiangchb@csust.edu.cn

Abstract

High-resolution satellite precipitation products are key inputs for basin-scale rainfall estimation, but they still exhibit substantial biases in complex terrain and during heavy rainfall. Recent multi-source fusion studies have shown that simply stacking multiple same-type microwave satellite products yields only limited additional gains for high-quality precipitation estimates and may even introduce local degradation, suggesting that targeted correction of a single, widely validated high-quality microwave product (such as IMERG) is a more rational strategy. Focusing on the mountainous, gauge-sparse Lüshui River basin with pronounced relief and frequent heavy rainfall, we use GPM IMERG V07 as the primary microwave product and incorporate CHIRPS, ERA5 evaporation, and a digital elevation model as auxiliary inputs to build a daily attention-enhanced CNN–LSTM (A-CNN–LSTM) bias-correction framework. Under a unified IMERG-based setting, we compare three network architectures—LSTM, CNN–LSTM, and A-CNN–LSTM—and test three input configurations (single-source IMERG, single-source CHIRPS, and combined IMERG + CHIRPS) to jointly evaluate impacts on corrected precipitation and SWAT runoff simulations. The IMERG-driven A-CNN–LSTM markedly reduces daily root-mean-square error and improves the intensity and timing of 10–50 mm·d⁻¹ rainfall events; the single-source IMERG configuration also outperforms CHIRPS-including multi-source setups in terms of correlation, RMSE, and performance across rainfall-intensity classes. When the corrected IMERG product is used to force SWAT, daily Nash-Sutcliffe Efficiency increases from about 0.71/0.70 to 0.85/0.79 in the calibration/validation periods, and RMSE decreases from 87.92 to 60.98 m³ s⁻¹, while flood peaks and timing closely match simulations driven by gauge-interpolated precipitation. Overall, the results demonstrate that, in gauge-sparse mountainous basins, correcting a single high-quality, widely validated microwave product with a small set of heterogeneous covariates is more effective for improving precipitation inputs and their hydrological utility than simply aggregating multiple same-type satellite products.



Academic Editor: Stephan Havemann

Received: 2 December 2025

Revised: 5 January 2026

Accepted: 6 January 2026

Published: 8 January 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).

Keywords: satellite precipitation; GPM IMERG; deep-learning-based precipitation bias correction; multi-source precipitation data fusion; runoff simulation using SWAT

1. Introduction

High spatiotemporal-resolution precipitation information underpins flood early warning, water resources regulation, and distributed hydrological modeling, and is particularly

critical for small and medium-sized mountainous catchments with sparse rain gauge networks [1]. In such basins, gauges are typically concentrated along urban areas and transportation corridors, leaving upstream mountains and storm centers poorly monitored; interpolated areal precipitation fields thus tend to smooth extremes and spatial gradients, causing discrepancies in runoff volume, flood peaks, and peak timing and increasing uncertainty in flood protection and operational regulation [2]. Where substantial densification of the gauge network is impractical, satellite and reanalysis precipitation products with continuous spatial coverage have become an important means of improving basin-scale precipitation inputs through regional evaluation and bias correction [3].

Satellite-based precipitation products such as TRMM, GSMaP, and CHIRPS, together with reanalysis datasets including ERA5 and ERA5-Land, have been widely used for basin-scale precipitation monitoring and hydrological modeling, and provide a fundamental data basis for small and medium-sized basins [4–7]. Previous studies show that these datasets can effectively compensate for sparse gauge coverage at medium to large scales, yet individual products still exhibit systematic biases, distorted frequency distributions, and deficiencies in representing extremes [8,9]. To reduce these errors, a range of statistical correction and fusion methods—linear and nonlinear regression, quantile mapping, probability distribution matching, Bayesian merging, and multi-source weighted blending—have been applied to satellite and reanalysis precipitation [10,11]. However, most of these approaches rely on prescribed distributional forms or linear assumptions and are sensitive to regional conditions, limiting their ability to capture nonlinear error structures and extreme precipitation in complex terrain and leaving considerable room for improvement in typical small mountainous basins.

The rapid development of deep learning has opened new avenues for multi-source satellite precipitation fusion and bias correction. Convolutional neural networks (CNNs) can extract multi-scale spatial features from gridded fields, whereas long short-term memory (LSTM) networks are well suited to representing variability across multiple temporal scales. Hybrid spatiotemporal architectures such as CNN–LSTM and ConvLSTM, which combine these strengths, generally outperform traditional statistical approaches and conventional machine-learning methods in terms of correlation, error metrics, and spatial pattern consistency, and can partly mitigate the impact of satellite-product errors on hydrological simulations [12–15]. Building on this, deep learning models that incorporate temporal and spatial attention or transformer-based architecture further enhance the response to salient information and have been used for downscaling, multi-source fusion, and numerical model bias correction [16,17]. These studies suggest that attention mechanisms help emphasize key periods of heavy precipitation and critical regions, improve the depiction of extremes and complex-terrain precipitation, and enhance the interpretability of input-feature importance [18,19]. Nevertheless, most work still focuses on improving statistical metrics of the precipitation field, with relatively few systematic evaluations across rainfall intensities—especially for moderate-to-heavy and storm events that control flood generation—or in typical small mountainous basins.

Fusion-based correction studies further indicate that merging multiple satellite precipitation products can enhance spatial consistency and improve certain hydrological indicators, but that the magnitude of improvement depends strongly on the quality and error structure of the inputs. When products share similar observation mechanisms and error characteristics, simply stacking several microwave-based datasets often yields only modest benefits and may even produce mixed local improvements and degradations in mountainous regions [20]. Accordingly, rather than further pursuing marginal gains from “quantity accumulation” among multiple homogeneous microwave products, it is more meaningful to focus on targeted regional correction and application-oriented evaluation

centered on a high-quality, representative product that has been extensively validated across different regions. Among current high-resolution satellite precipitation datasets, the Integrated Multi-satellite Retrievals for GPM (IMERG) is one of the most widely used. The latest seventh-generation algorithm (IMERG V07) integrates multi-constellation passive microwave observations, geostationary infrared brightness temperatures, and rain-gauge data, and provides quasi-global precipitation estimates at $0.1^\circ/30$ min resolution. Extensive global and regional evaluations have shown that IMERG can reasonably reproduce spatial patterns and seasonal evolution, and that for moderate to heavy rainfall, it generally outperforms many traditional satellite products; after modest regional bias correction, daily and monthly errors are typically acceptable for hydrological modeling and risk assessment across a range of climate regimes and basin scales [21–23]. Taken together, these findings indicate that IMERG V07 provides relatively high overall accuracy and a robust basis for hydrological applications and can be regarded as a representative “core” high-resolution microwave precipitation product; however, in typical small mountainous basins, deep learning-based correction frameworks that build on a single high-quality product, incorporate heterogeneous auxiliary information, and systematically evaluate performance across rainfall-intensity classes and runoff responses remain relatively scarce [24–27].

In this context, this study focuses on the Lushui River Basin, a small mountainous catchment in the upper Xiangjiang River. We employ IMERG V07 as the primary satellite precipitation product, introduce CHIRPS—based mainly on infrared brightness-temperature retrievals—as a heterogeneous auxiliary source, and incorporate ERA5 evaporation and digital elevation model (DEM) data as additional predictors. On the basis of multi-source precipitation correction, we further couple the corrected fields with the SWAT model to jointly evaluate precipitation statistics and hydrological response [28]. The specific objectives are to: (1) develop a daily A-CNN-LSTM correction framework that combines convolutional encoding, long short-term memory networks, and attention mechanisms, and within a unified architecture comparing single-source IMERG, single-source CHIRPS, and combined IMERG + CHIRPS inputs, together with ERA5 and DEM-based auxiliary features, to clarify how input combinations and product quality influence correction performance; (2) assess, using station-based spatial cross-validation, stratified rainfall statistics, and representative storm-event metrics, the generalization ability of the model across space and rainfall intensities, with emphasis on flood-critical situations; and (3) examine, from a “precipitation correction–runoff response” perspective, the applicability and practical effectiveness of an IMERG-dominated deep learning framework that integrates heterogeneous satellite and reanalysis information for flood-risk assessment and water resources management in small and medium-sized mountainous basins.

2. Study Area and Data

2.1. Overview of the Study Area

The Lushui River Basin is located on the eastern side of the middle–upper reaches of the Xiangjiang River in southern China. It flows through Liuyang and Liling in Hunan Province and Pingxiang in Jiangxi Province, and joins the Xiangjiang River near Xiangtan, draining a catchment area of approximately 2871 km². The basin generally drains from east to west. The upper reaches are dominated by low mountains and dissected hills, whereas the middle and lower reaches gradually transition into gently undulating hills and alluvial plains and valleys. Pronounced topographic relief and highly heterogeneous land-surface conditions create complex runoff generation and routing processes.

The region is characterized by a subtropical monsoon climate, with a mean annual precipitation of about 1500–1700 mm that is highly concentrated within the year. From April to June, precipitation is mainly associated with frontal systems and mesoscale convective

events, whereas from July to September it is often influenced by instability along the subtropical high and the remnants of tropical cyclones, leading to short-duration intense rainfall and multi-peak flood events. Historical flood events indicate that the Lushui River Basin has a short concentration time and steeply rising and falling flood peaks, making it highly sensitive to the accuracy with which the spatiotemporal distribution of precipitation is represented [29]. However, the existing rain gauge and hydrological station network is sparse, with most stations concentrated along towns and transportation corridors, leaving the upstream mountainous areas and some headwater subcatchments poorly monitored. As a result, conventional gauge-interpolated precipitation fields have difficulty accurately capturing storm centers and orographic rainfall gradients [30]. These characteristics make the Lushui River Basin a typical small to medium-sized catchment for testing high-resolution satellite precipitation products, multi-source fusion and deep-learning-based correction methods, and their hydrological performance. The approximate location of the basin and the distribution of hydro-meteorological stations are shown in Figure 1.

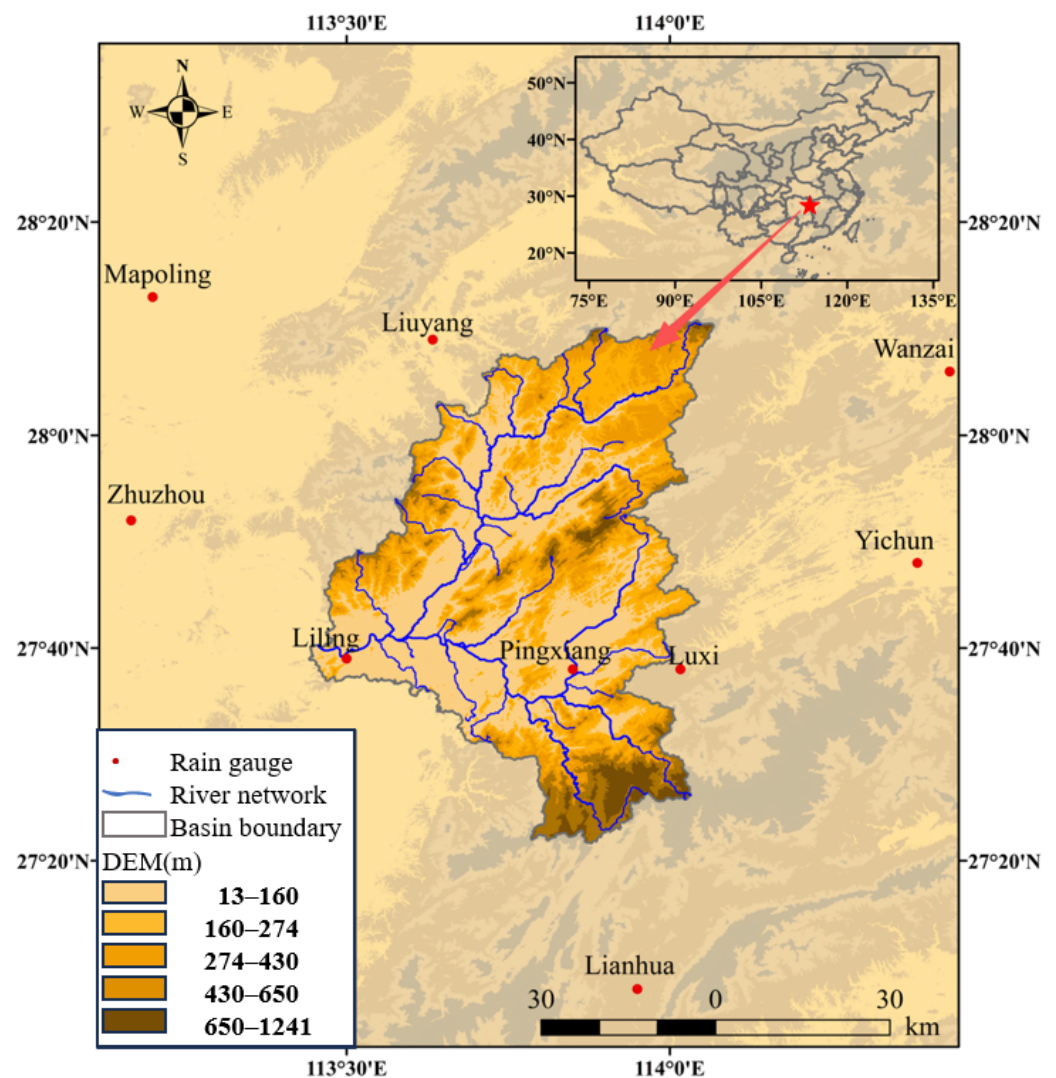


Figure 1. Location of the Lushui River Basin and distribution of hydrometeorological stations. Elevation classes follow a left-inclusive and right-exclusive convention; e.g., 13–160 m denotes $13 \leq z < 160$ m.

2.2. Datasets

This study employs multi-source datasets for the period 2011–2022, including gauge observations, satellite precipitation products, reanalysis datasets, and geospatial data, for multi-source precipitation correction and hydrological modeling. The analysis period (2011–2022) was selected as the longest time span over which daily rain-gauge observations and the outlet discharge record are both continuous and simultaneously overlap with all remote-sensing/reanalysis inputs used in this study. This period also provides sufficient samples to support model training under spatial cross-validation and to perform SWAT calibration/validation using temporally consistent forcings.

2.2.1. Gauge-Based Precipitation and Runoff Observations

Gauge-based precipitation data consist of daily records from eight rain gauges located within and around the Lushui River Basin. These data are obtained from the National Meteorological Information Center of the China Meteorological Administration and the China Meteorological Data Service Center (CMDSC, <https://data.cma.cn/>) and are supplemented and cross-checked using operational datasets from the relevant provincial hydrometeorological agencies [31]. These data serve as the benchmark for evaluating the accuracy of satellite precipitation products and for training and validating the deep learning model. To ensure the reliability of the reference data, unified quality control was applied to the daily precipitation series at all stations. First, completeness checks were performed to quantify the availability of valid records for each station. Second, physical plausibility checks were conducted, and negative values or evidently unreasonable records were flagged as invalid. Third, temporal consistency and change-point checks were implemented to identify anomalous values that deviated markedly from the continuity of adjacent periods. In addition, spatial consistency was verified by comparing precipitation processes with those at nearby stations during the same period; abnormal records lacking support from regional precipitation events were removed. After quality control, only station datasets with stable records and meeting modeling requirements were used for subsequent accuracy assessment and model training.

The runoff observations consist of daily measured streamflow at the watershed outlet control station (Daxitan Station) from 2011 to 2022, obtained from the operational database of the basin hydrological authority. These data were used for SWAT model calibration and validation, as well as for comparative analyses of runoff responses under different precipitation input schemes. The streamflow series was also screened for missing values and anomalies, with clearly unreasonable records excluded. Only valid records were used in metric calculations and comparative analyses to ensure comparability across scenarios and consistency in interpretation.

2.2.2. Satellite Precipitation Products

GPM IMERG V07: The Integrated Multi-satellite Retrievals for GPM (IMERG) Version 07 Final Run daily product with a spatial resolution of $0.1^\circ \times 0.1^\circ$ is used in this study. IMERG V07 combines multi-constellation microwave observations, infrared imagery, and ground-based rain-gauge information to generate high-resolution global precipitation estimates and represents one of the latest-generation satellite precipitation products. In this work, it is treated as the primary satellite precipitation data source and as the target of the deep learning-based correction. The data were obtained from the NASA GPM data portal (<https://gpm.nasa.gov/data>, accessed on 5 April 2025).

CHIRPS: The Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) daily product, with an original spatial resolution of $0.05^\circ \times 0.05^\circ$, is used in this study. CHIRPS combines infrared brightness temperature fields with ground-based station

observations and features a relatively long temporal record and high spatial resolution. In this work, CHIRPS is bilinearly interpolated and resampled to $0.1^\circ \times 0.1^\circ$ to match the IMERG grid and is used for comparative evaluation and as an additional input in multi-source fusion experiments [32]. The data were obtained from the CHIRPS archive hosted by the Climate Hazards Center (<https://www.chc.ucsb.edu/data/chirps>, accessed on 5 April 2025).

Both satellite precipitation products are clipped to the extent of the Lushui River Basin and an external buffer zone and are used for evaluating the accuracy of the raw products and for constructing input features for the deep learning models.

2.2.3. Reanalysis and Geospatial Data

ERA5 daily statistics (single levels): This study uses the “ERA5 post-processed daily statistics on single levels from 1940 to present” dataset distributed by the Copernicus Climate Data Store (CDS) [33]. The dataset is derived from ERA5 hourly single-level fields and computes daily statistics on retrieval according to user-specified options (e.g., daily accumulation for accumulated variables). The daily aggregation window can be defined via a UTC offset, enabling the generation of day-scale data. In our retrieval, the UTC offset was set to UTC + 8, so that each daily window corresponds to Beijing time from 00:00 to 24:00. For the ERA5 evaporation variable, we used the daily sum as the daily value and selected the evapotranspiration-related accumulated evaporation term to characterize the surface-to-atmosphere moisture flux, which implicitly includes a simplified representation of vegetation transpiration. Under the ECMWF IFS sign convention for fluxes (downward positive), negative values indicate evaporation (upward moisture flux), whereas positive values indicate condensation. The data were obtained from the Copernicus Climate Data Store (CDS).

Digital elevation model (DEM): A 30 m-resolution digital elevation model is used, obtained from the “Geospatial Data Cloud” platform operated by the Computer Network Information Center of the Chinese Academy of Sciences. Elevation, slope, and other terrain attributes are derived from the DEM and aggregated to the 0.1° grid scale to characterize topographic relief and potential orographic rainfall effects, and to provide basic inputs for the delineation of the watershed and sub-basins in the SWAT model. The data were obtained from the Geospatial Data Cloud (<https://www.gscloud.cn>).

2.2.4. Other Data

Land-use data: The 2020 China Land-Use/Cover Change (LUCC) remote-sensing monitoring dataset at 1 km resolution is used, obtained from the Resource and Environment Science and Data Center (<https://www.resdc.cn>). The corresponding annual land-cover dataset is documented in related studies [34]. According to the requirements of the SWAT model, the original land-use classes are reclassified into ten categories, including cropland, forest, grassland, water bodies, and built-up land, for the delineation of hydrological response units (HRUs) and the characterization of land-surface conditions.

Soil data: The Harmonized World Soil Database (HWSD) v1.1 soil dataset for China, with a spatial resolution of approximately 1 km, is used, obtained from the National Cryosphere Desert Data Center (<https://www.ncdc.ac.cn>) [35]. Based on the HWSD soil attribute table and informed by typical parameterization schemes adopted in previous SWAT applications over Chinese basins, a SWAT soil-parameter database tailored to the Lushui River Basin is constructed to constrain the simulation of runoff generation and routing processes.

2.3. Data Preprocessing

To ensure consistency and comparability among the multi-source datasets, all data were subjected to a unified preprocessing procedure prior to model development. First, in the spatial domain, all gridded datasets, including IMERG, CHIRPS, ERA5 evaporation, and the DEM, were resampled to a resolution of $0.1^\circ \times 0.1^\circ$. The study area was then restricted to a regular grid spanning $26.0\text{--}30.0^\circ$ N and $112.0\text{--}115.0^\circ$ E, resulting in a 30×30 grid of cells. The original 0.05° CHIRPS product was resampled to 0.1° using bilinear interpolation, whereas ERA5 evaporation and DEM data were aggregated to 0.1° by grid averaging and related methods, thereby enabling the overlay of all datasets within a common grid framework.

In the temporal dimension, IMERG, CHIRPS, and ERA5 are provided in UTC, whereas gauge precipitation and discharge observations follow Beijing Time (UTC + 8). We first converted the timestamps of satellite and reanalysis datasets to Beijing Time and adopted 00:00–24:00 (Beijing Time) as the daily aggregation window. All datasets were then aligned day by day, and only dates with valid records across all required data sources were retained for accuracy assessment and model training.

Missing data and outlier handling: For gauge precipitation and discharge time series, missing values and outliers were detected and clearly unreasonable records were removed. Specifically, a station was retained only if at least 95% of daily records were available during 2011–2022. Negative precipitation values were removed, and daily totals exceeding $500 \text{ mm}\cdot\text{d}^{-1}$ were flagged as suspicious. In addition, a gauge value was flagged when it exceeded three times the median of the three nearest stations within 50 km while their median was $<5 \text{ mm}\cdot\text{d}^{-1}$; flagged records were excluded from training and accuracy evaluation without imputation. For discharge, missing/negative values were removed, and isolated spikes exceeding five times the median of a ± 3 -day window were flagged and excluded from metric calculation. To avoid introducing artificial errors, imputed values were not used as targets for supervised learning; when a gauge record on a given day was missing or flagged as abnormal, the corresponding sample was excluded from model training and precipitation accuracy evaluation. For discharge, days with missing or abnormal values were excluded from metric calculation and comparative analyses; the selection of calibration and validation periods also prioritized the continuity of valid records to ensure reliable SWAT parameter optimization and evaluation. For satellite precipitation, a physical plausibility check was applied and negative values as well as physically unreasonable outliers were removed. It should be noted that the ERA5 evaporation variable follows the ECMWF flux sign convention, under which evaporation is typically represented by negative values. Therefore, negative values were not directly removed. Instead, we converted it to a non-negative evapotranspiration intensity variable using $ET = -E$. When condensation occurs (represented by positive evaporation values), the sign conversion yields negative values, which were set to 0 so that this auxiliary predictor represents evaporation only and remains non-negative. Note that the ERA5 variable used in this study is the model-diagnosed surface accumulated evaporation (E ; daily sum), which is sign-corrected to a non-negative evaporation/ET—intensity proxy $ET_{ERA5} = \max(-E, 0)$; it is used only as an auxiliary predictor rather than being interpreted as potential evapotranspiration (PET). Importantly, ERA5 is kept identical across all precipitation-input scenarios and is not used as a SWAT PET forcing (the hydrological evaluation is driven only by the corresponding precipitation products), so it does not affect the relative comparison among scenarios or the independence of the SWAT-based validation.

When constructing the deep learning training samples, the locations of the rain gauges were used as spatial anchors. For each station, multi-source information within a 15×15 grid window in the unified grid was extracted, including IMERG, CHIRPS, ERA5

evaporation, and terrain attributes derived from the DEM, to characterize the spatial structure of precipitation and the underlying surface conditions in the station neighborhood. The daily precipitation observations at each station were taken as prediction targets and paired with the contemporaneous sub-grid features to form supervised learning samples [36,37]. To reduce the influence of differing units and value ranges on the training process, the multi-source input features were normalized or standardized, whereas the observed precipitation values were retained in their physical units to facilitate error analysis and hydrological interpretation. Details of the neighborhood window size selection and the sensitivity analysis are provided in Section 3 and summarized in Table A1.

Dataset partitioning followed a station-based spatial cross-validation scheme to evaluate the model's generalization ability at stations that were not used for training. Specifically, among the eight rain gauges within and around the basin, one station was selected in turn as the test site, and the remaining seven stations were used as the training set, yielding an eight-fold spatial cross-validation. In each fold, the model parameters were fitted using only samples from the training stations and then evaluated at the independent test station. The same station-based partitioning strategy was consistently applied to satellite-precipitation evaluation, multi-source deep learning correction, and SWAT-based runoff simulations, thereby ensuring comparability among different experimental configurations and consistency in the interpretation of results [38].

3. Methodology

3.1. Conceptual Framework for Multi-Source Precipitation Correction

When applied in mountainous catchments, satellite precipitation products are often affected by complex topography and pronounced spatiotemporal rainfall variability, resulting in biases in precipitation magnitude, spatial patterns, and the representation of heavy rainfall events. To improve the suitability of satellite precipitation for the study region, we used daily rain-gauge observations as the reference and incorporated spatial neighborhood information around stations together with auxiliary environmental covariates to develop a deep-learning-based bias-correction framework for IMERG daily precipitation. In addition, we designed two experimental scenarios—single-source-input correction and multi-source-input correction—to quantify the contribution of multi-source information to correction performance.

To enable an unambiguous interpretation of the experimental design and result comparisons, we first describe the construction, naming conventions, and intended use of each precipitation scheme (Table 1). The sources and basic descriptions of the underlying datasets (e.g., IMERG, CHIRPS, ERA5, and DEM) are provided in Section 2 and are not repeated here; this section focuses on how these datasets are integrated within our framework and the corresponding correction workflow.

Table 1. Summary of precipitation schemes and experimental scenarios.

Scenario	Nature	Primary Inputs	Framework/ Processing	Auxiliary Predictors	Purpose	Notes
OBS	Baseline	Gauge observations	SWAT built-in gauge interpolation (station-weighting)	—	Hydrological reference	See Section 3.4
IMERG	Raw	IMERG	Original product	—	Baseline comparison	Pre-correction benchmark
CHIRPS	Raw	CHIRPS	Original product	—	Baseline comparison	Comparative product

Table 1. Cont.

Scenario	Nature	Primary Inputs	Framework/ Processing	Auxiliary Predictors	Purpose	Notes
IMERG-LSTM	Corrected	IMERG (single-source)	LSTM	DEM-derived variables, ERA5 evaporation	Model selection	—
IMERG-CNN-LSTM	Corrected	IMERG (single-source)	CNN-LSTM	Same as IMERG-LSTM	Model selection	—
IMERG-A	Corrected	IMERG (single-source)	A-CNN-LSTM	Same as IMERG-LSTM	Main results	Recommended scheme
IMERG-CL	Corrected	IMERG (single-source)	A-CNN-LSTM	Same as IMERG-LSTM	Input-configuration comparison	Single-source benchmark
CHIRPS-CL	Corrected	CHIRPS (single-source)	A-CNN-LSTM	Same as IMERG-LSTM	Input-configuration comparison	Single-source benchmark
ICS-CL	Fused	IMERG + CHIRPS (multi-source)	A-CNN-LSTM	Same as IMERG-LSTM	Input-configuration comparison	Feature-level multi-source input

Note: (1) Raw: uncorrected precipitation products. (2) Corrected: deep-learning-corrected output using a single precipitation input. (3) Fused: deep-learning-corrected output with multiple precipitation inputs (in-model fusion), not simple averaging. (4) “Auxiliary predictors”: non-precipitation predictors (e.g., DEM- and ERA5-derived); see Sections 2 and 3.3.

3.1.1. Experimental Scenarios and Naming of Precipitation Schemes

We categorize the precipitation schemes into three groups: raw products, corrected products, and fused products. Here, *raw* denotes the original gridded precipitation products that are not processed by our framework (e.g., IMERG and CHIRPS). *Corrected* refers to the bias-corrected outputs trained within a unified deep-learning correction framework using a single precipitation product as the primary input. *Fused* denotes the corrected results produced by a single model that takes multiple precipitation products (e.g., IMERG and CHIRPS) as concurrent inputs and performs joint multi-source representation learning at the feature level. Notably, *fused* in this study refers to in-model multi-source input fusion rather than a simple superposition/averaging of multiple products, and it is distinct from standalone statistical fusion approaches. For consistency in subsequent comparisons, we use the scenario names and classification system in Table 1 to refer to all precipitation schemes throughout the Results and Hydrological Simulation sections.

3.1.2. Bias-Correction Workflow and Experimental Design

This study uses daily rain-gauge observations as reference data and exploits the spatial distribution of satellite information in the vicinity of each station to correct daily IMERG precipitation. Data preprocessing has already harmonized all datasets to a spatial resolution of 0.1° and a consistent daily time axis. For each rain gauge, a 15×15 sub-grid window (approximately $147 \text{ km} \times 147 \text{ km}$) centered on the grid cell containing the station is extracted to form a multi-channel input comprising IMERG precipitation and auxiliary variables such as evapotranspiration and terrain information, thereby characterizing the precipitation structure and underlying surface conditions in the station neighborhood. The corresponding daily gauge observations at each station serve as supervisory signals and are paired with the same-day sub-grid features to construct “multi-source sub-grid features–station precipitation” sample pairs.

In this study, the “ $N \times N$ gridded window” refers to a spatial neighborhood sub-grid extracted from the unified 0.1° grid, centered at the grid cell containing the rain gauge. It is used to provide the point-scale correction model with surrounding precipitation spatial structure and land-surface/background information. It should be emphasized that gauge observations are point-scale references, and we do not assume that a gauge value represents the areal-mean precipitation over the $N \times N$ window. Instead, the model learns

a statistical mapping from multi-source neighborhood gridded predictors to the point-scale precipitation observed at the gauge. The neighborhood window therefore serves as an interpretable spatial context that helps characterize local precipitation variability under complex terrain conditions.

To enhance the transparency of the input-scale setting, we conducted a sensitivity comparison across candidate window sizes of 7×7 , 9×9 , 11×11 , 13×13 , and 15×15 (using the final A-CNN-LSTM framework as a representative case, while keeping the training strategy and the station-based eight-fold spatial cross-validation split identical). The results show that increasing the window size from 7×7 to 15×15 improves the mean CC across the eight stations by about 0.016 ($0.7915 \rightarrow 0.8071$) and reduces the mean RMSE by about $0.25 \text{ mm} \cdot \text{d}^{-1}$ ($6.949 \rightarrow 6.702$). In contrast, the differences between 13×13 and 15×15 are marginal (CC increases by only ~ 0.001 and RMSE decreases by only $\sim 0.015 \text{ mm} \cdot \text{d}^{-1}$, indicating that performance gains saturate within the 13×13 – 15×15 range (Table A1). Therefore, we adopt 15×15 as the default neighborhood window for subsequent experiments and product generation. Given that the study domain is a 30×30 grid, further increasing the window size to 17×17 would prevent extracting complete neighborhoods for samples near the domain boundary and would require additional boundary handling (e.g., padding or cropping), which could introduce unnecessary uncertainty; thus, larger windows were not included in the comparison.

To assess the spatial transferability of the proposed approach, a station-based eight-fold spatial cross-validation scheme is adopted: in each fold, one rain gauge is designated as an independent test station, and the samples from the remaining seven stations are used for training; the model parameters are fitted exclusively on the training stations and the correction performance is then evaluated at the previously unseen test station. This design reflects the potential applicability of the model in sparsely gauged or ungauged areas, rather than merely its fitting ability at the training stations.

Once a model with stable cross-validation performance is obtained, the trained network is applied grid by grid: for each grid cell, the corresponding 15×15 sub-grid inputs centered on that cell are extracted and fed into the network to correct the original satellite precipitation, thereby producing a spatially continuous corrected precipitation product. The corrected fields are subsequently used to drive the SWAT model, allowing the correction performance to be evaluated jointly from the dual perspective of meteorological-field accuracy and hydrological response.

3.2. A-CNN-LSTM Model Architecture

3.2.1. Model Formulation and Training Objective

Let $X_t \in R^{N \times N \times C}$ denote the multi-channel neighborhood window extracted for a given station on day t , where $N \times N$ is the spatial window size and C is the number of predictor channels (e.g., precipitation products and auxiliary variables). For each station, the study period (2011–2022) contains $T = 4383$ daily records, forming an input sequence $\{X_t\}_{t=1}^T$. The corresponding supervision signal is the point-scale daily gauge precipitation $\{y_t\}_{t=1}^T$. The proposed model directly predicts the corrected daily precipitation series $\{\hat{y}_t\}_{t=1}^T$.

Spatial feature extraction (CNN). For each day t , a convolutional encoder is applied to the neighborhood window to extract a spatial feature map:

$$F_t = f_{CNN}(X_t, \theta_c), \quad F_t \in R^{H \times W \times d} \quad (1)$$

where $f_{CNN}(\cdot)$ represents a stack of convolutional operations with nonlinear activations, θ_c denotes the CNN parameters, and H , W , and d are the spatial dimensions and the number of feature channels, respectively.

Spatial attention. To emphasize spatial locations within the neighborhood that are more informative for point-scale precipitation correction, a spatial attention module is applied to F_t . Specifically, F_t is unfolded into $P = H \times W$ location-wise feature vectors $\{f_{t,p}\}_{p=1}^P$, where $f_{t,p} \in R^d$. An attention score for each location is computed as

$$u_{t,p} = v_s^T \tanh(W_s f_{t,p} + b_s), p = 1, \dots, P \tag{2}$$

where W_s, b_s, v_s and are trainable parameters and $(\cdot)^T$ denotes the transpose. The scores are normalized by a softmax function to obtain spatial attention weights:

$$\alpha_{t,p} = \frac{\exp(u_{t,p})}{\sum_{q=1}^P \exp(u_{t,q})} \tag{3}$$

The attention-weighted spatial representation is then derived by

$$s_t = \sum_{p=1}^P \alpha_{t,p} f_{t,p}, s_t \in R^d \tag{4}$$

This mechanism enables the model to adaptively focus on informative sub-regions in the neighborhood and thus better capture localized precipitation variability in complex terrain.

Temporal modeling (LSTM). The sequence $\{s_t\}_{t=1}^T$ is modeled by an LSTM to capture temporal dependencies:

$$\begin{aligned} i_t &= \sigma(W_i s_t + U_i h_{t-1} + b_i) \\ f_t &= \sigma(W_f s_t + U_f h_{t-1} + b_f) \\ \check{c}_t &= \tanh(W_c s_t + U_c h_{t-1} + b_c) \\ c_t &= f_t \odot c_{t-1} + i_t \odot \check{c}_t \\ o_t &= \sigma(W_o s_t + U_o h_{t-1} + b_o) \\ h_t &= o_t \odot \tanh(c_t) \end{aligned} \tag{5}$$

where $\sigma(\cdot)$ denotes the sigmoid function, \odot is element-wise multiplication, and c_t and h_t are the hidden and cell states.

Temporal attention. To adaptively emphasize informative time steps, we apply an additive temporal attention mechanism to the LSTM hidden states. For a target day t , an alignment score for each time step k is computed as

$$e_{t,k} = v_s^T \tanh(W_a h_k + U_a h_t + b_a), k = 1, \dots, T \tag{6}$$

and normalized to obtain attention weights:

$$\beta_{t,k} = \frac{\exp(e_{t,k})}{\sum_{j=1}^T \exp(e_{t,j})} \tag{7}$$

The attention-pooled context vector is then

$$z_t = \sum_{k=1}^T \beta_{t,k} h_k \tag{8}$$

This temporal attention allows the model to assign higher weights to time steps that contribute more to the precipitation correction for day t .

Output layer. The corrected daily precipitation is produced by a regression head:

$$\hat{y}_t = w^T z_t + b \tag{9}$$

where w and b are trainable parameters.

Training objective. Let Θ denote all trainable parameters in the model. Let $y_{m,t}$ be the observed daily precipitation at station m on day t , and $\hat{y}_{m,t}$ be the corresponding model prediction. We introduce a validity indicator $I_{m,t} \in \{0, 1\}$ to mask missing or flagged observations ($I_{m,t} = 0$), so that only valid samples contribute to optimization. The model is trained by minimizing a weighted combination of mean squared error (MSE) and mean absolute error (MAE):

$$\gamma(\Theta) = \alpha \frac{1}{N_v} \sum_{m=1}^M \sum_{t=1}^T I_{m,t} (\hat{y}_{m,t} - y_{m,t})^2 + (1 - \alpha) \frac{1}{N_v} \sum_{m=1}^M \sum_{t=1}^T I_{m,t} |\hat{y}_{m,t} - y_{m,t}| \quad (10)$$

where M is the number of training stations in a given cross-validation fold. T is the number of days, $N_v = \sum_{m=1}^M \sum_{t=1}^T I_{m,t}$ is the number of valid samples, and $\alpha \in [0, 1]$ controls the trade-off between overall fitting accuracy and robustness to large deviations.

3.2.2. Network Architecture and Parameter Settings

This study develops an A-CNN-LSTM model to correct satellite precipitation at the station scale, using IMERG as the primary input while incorporating CHIRPS precipitation, ERA5 reanalysis evapotranspiration, and DEM-derived information as auxiliary features. Specifically, for each rain-gauge station and each day, a 15×15 grid window centered on the station is extracted from the preprocessed gridded datasets. IMERG, CHIRPS, ERA5 evaporation, and DEM-based variables are stacked along the channel dimension to form an input tensor of “time series $\times 15 \times 15$ spatial window \times multiple channels”, while the corresponding daily station precipitation serves as the output for supervised training. The overall architecture consists of three stages: spatial feature extraction, temporal attention weighting, and temporal sequence modeling with regression output, as illustrated in Figure 2 [39,40].

In the spatial feature extraction stage, a “per-time-step convolution” strategy is adopted, whereby, for each day, the 15×15 multi-channel grid is passed through two successive convolution–pooling blocks. The first convolutional layer uses 3×3 kernels with 32 channels and a ReLU activation function to extract local spatial features of precipitation, topography, and related fields around the station. It is followed by a 2×2 max-pooling layer that downsamples the feature maps, reducing spatial dimensionality and attenuating noise. The second convolutional layer also employs 3×3 kernels but increases the number of channels to 64 to extract higher-level spatial features and is again followed by a 2×2 max-pooling layer to further reduce the size of the feature maps. After the two convolution–pooling blocks, a global average pooling layer is applied at each time step to compress the two-dimensional feature maps into a one-dimensional vector. This vector is then passed through a fully connected layer with 32 units and ReLU activation, followed by a Dropout layer with a rate of 0.2 (20% random dropout) to mitigate the risk of overfitting. Through this sequence of operations, each daily 15×15 multi-channel input is ultimately compressed into a 32-dimensional spatial feature vector, and these vectors are concatenated in chronological order to form the feature sequence required for subsequent temporal modeling.

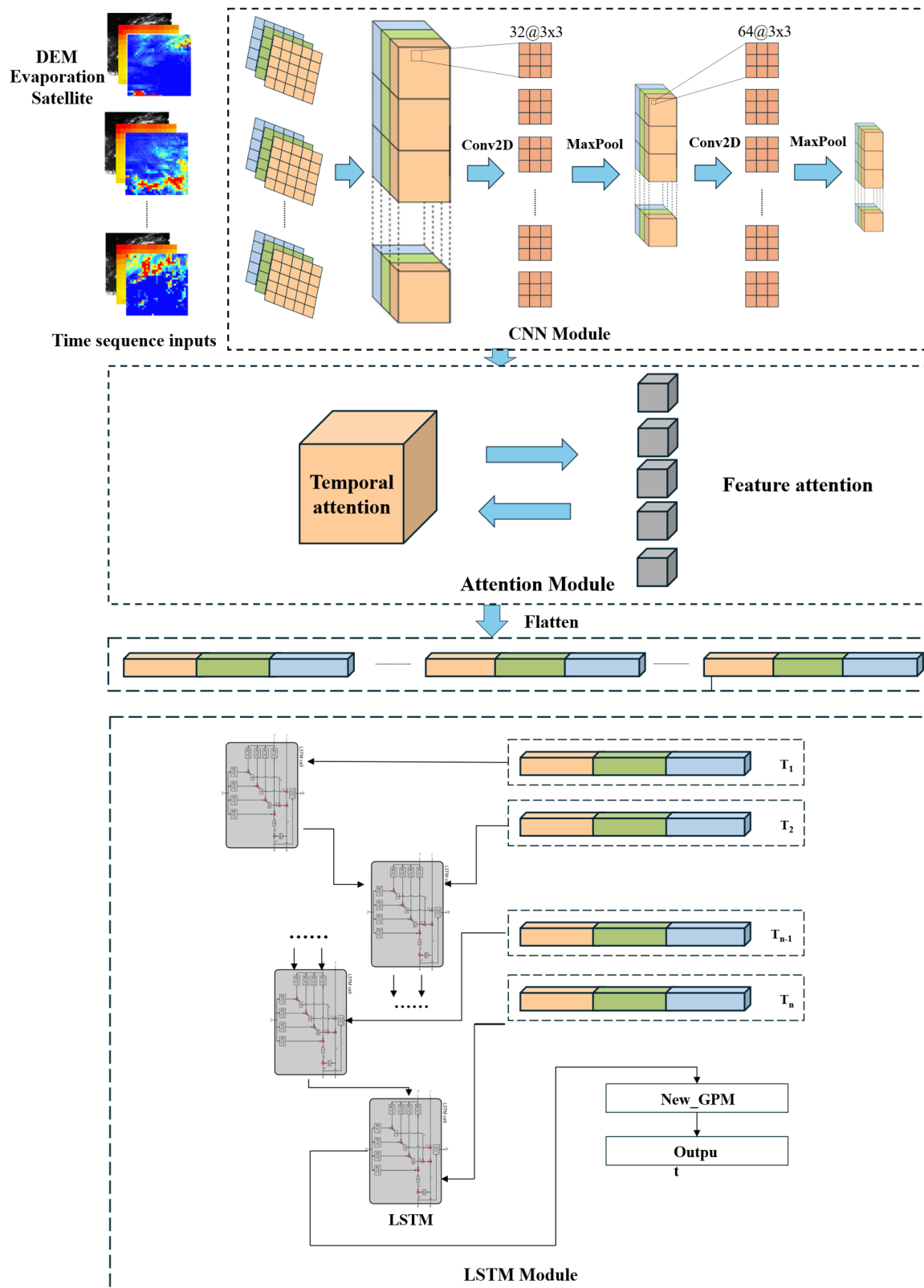


Figure 2. Multi-source Precipitation Data Fusion Framework.

Once the spatial feature sequences are obtained, a temporal attention mechanism is introduced after the convolutional block to enable the model to automatically distinguish the relative importance of different time periods for precipitation correction at each station. Specifically, a shared fully connected layer is first used to compute a scalar “score” for

each 32-dimensional feature vector at every time step, and a Softmax function is then applied to normalize these scores across the sequence into a set of attention weights. These weights reflect the importance assigned by the model to each time step: larger weights indicate a greater influence on the corresponding features in subsequent computations, whereas smaller weights down-weight their contribution. Through this adaptive weighting, the model can focus more on precipitation events or synoptic conditions that are most representative for the station and reduce the impact of weak-precipitation periods and noisy samples on the results [40,41]. The model was implemented in Python (v3.8, Ubuntu 20.04) using TensorFlow (v2.5.0, Keras API), with GPU acceleration enabled via CUDA (v11.2).

In the temporal modeling stage, a two-layer LSTM architecture is employed to capture the temporal evolution of daily precipitation and satellite biases over the period 2011–2022. The first LSTM layer is configured with 128 hidden units and returns the full sequence of outputs to preserve complete temporal information, while L2 regularization is applied to its weights to control model complexity. The second LSTM layer reduces the number of hidden units to 64 and outputs only the final hidden state, which serves as a compact representation of the entire period. This aggregated feature vector is then passed through a fully connected layer with 8 units and ReLU activation for further compression and fusion, followed by a linear output layer that yields the corrected daily precipitation. For single-station modeling, the output layer has a dimensionality of one, whereas for joint modeling of multiple stations, its dimensionality can be extended to the number of stations. The dataset partitioning, cross-validation scheme, and optimizer settings used for model training and validation have already been described in the data preprocessing and modeling workflow sections and are therefore not repeated here [42,43].

3.3. Model Training and Validation

The input features are normalized, whereas the observed station precipitation used as the regression target is retained in its physical units. Model parameters are optimized by minimizing the masked weighted MSE–MAE objective defined in Section 3.2.1. An Adam-type adaptive optimizer is employed, together with learning-rate scheduling, Dropout, and gradient clipping, to mitigate overfitting [44].

In the eight-fold spatial cross-validation, the model is trained in each fold using samples from only seven stations, and its performance is evaluated at the held-out test station, thereby avoiding information leakage and enabling an assessment of spatial generalization. During training, an early-stopping criterion based on validation performance is used to select the optimal model. If necessary, a small ensemble of models initialized with different random weights can be constructed to enhance the stability of the results, without altering the overall methodological framework.

Finally, by aggregating the results from all folds, the correction performance of the model is analyzed across different stations and rainfall regimes, and an appropriate model configuration is selected for application to the entire basin.

3.4. SWAT Model Setup and Runoff Simulation Scenarios

SWAT (Soil and Water Assessment Tool), developed in the 1990s by the United States Department of Agriculture–Agricultural Research Service (USDA-ARS), is a widely used semi-distributed hydrological model operating at the basin scale. It is primarily designed to simulate surface water, sediment, and pollutant transport, as well as the impacts of agricultural management practices on water resources [45]. SWAT delineates sub-basins and the channel network from a digital elevation model (DEM) and, in combination with soil, land-use, and meteorological data, defines hydrological response units (HRUs). Runoff generation, evapotranspiration, soil erosion, and related processes are simulated at the

HRU scale and then aggregated at the sub-basin and reach scales, enabling the model to output water quantity, water quality, and sediment fluxes. This structure makes SWAT well suited for analyzing the effects of different precipitation inputs on runoff while explicitly accounting for heterogeneity in underlying surface conditions.

In this study, a 30 m-resolution DEM is used within SWAT to delineate the Lushui River Basin boundary and extract the river network. The 2020 LUCC land-use dataset and the HWSO soil dataset are employed for HRU definition, with threshold values for land-use, soil, and slope classes set to 10%, 10%, and 5%, resulting in a total of 518 HRUs [46]. This HRU-threshold setting (10/10/5 for land use/soil/slope) is used to control the number of HRUs and maintain computational efficiency. Meanwhile, it preserves the key heterogeneity associated with land use, soil, and slope, which is a common practice in SWAT/ArcSWAT applications [47,48]. Notably, Jiang et al. (2021) [49] reported that daily streamflow evaluation statistics tend to be more sensitive to HRU-threshold choices than monthly results, highlighting the need for a conservative and consistent threshold when the modeling objective focuses on daily runoff. During model runs, the original or bias-corrected satellite precipitation products under different scenarios are used as precipitation inputs, while all other model settings are kept identical so that subsequent comparisons primarily reflect differences among the precipitation schemes.

Previous studies have shown that, if the model is calibrated independently for different precipitation inputs, parameter adjustments tend to partially compensate for precipitation errors and make the simulated streamflow under different forcings converge, which is unfavorable for objectively evaluating the merits of the precipitation inputs themselves [50]. Accordingly, this study employs the SUFI-2 algorithm in SWAT-CUP to perform a unified calibration and validation of the SWAT model [51]. The selection of calibration parameters is mainly guided by previous analyses of SWAT runoff sensitivity, with priority given to parameters that exert strong controls on runoff generation and flow routing [52]. Once the parameter set is determined, a single precipitation-forcing configuration is used for parameter optimization, and the Nash–Sutcliffe efficiency (NSE) and the coefficient of determination (R^2) are adopted as the primary performance metrics [53]. The simulation period spans 2011–2022, with 2011–2012 used as the warm-up period, 2013–2018 as the calibration period, and 2019–2022 as the validation period. The parameter sensitivity analysis and calibration results are presented in Section 4.5.

After establishing a unified parameter set, we designed multiple precipitation-forcing scenarios to quantify how different precipitation inputs affect runoff simulation. The scenarios include: (i) an interpolated gauge-based precipitation scenario as the hydrological reference; (ii) raw IMERG precipitation; (iii) raw CHIRPS precipitation (when included in the comparison); (iv) IMERG precipitation corrected by the proposed A-CNN-LSTM model (IMERG-A); and (v) additional deep-learning-based scenarios derived under the same framework, including single-source corrected products and multi-source-input corrected products (feature-level fusion within one network). For the gauge-based reference scenario, SWAT converts point gauge observations into sub-basin precipitation using its built-in station-weighting interpolation scheme, in which weights are assigned based on the relative distance between each gauge and the target sub-basin. The resulting sub-basin precipitation series are then used to drive SWAT, and the simulated runoff is evaluated against observations to assess hydrological consistency under each precipitation scenario. For clarity and consistency, the scenario names and the classification scheme in Table 1 are used throughout the subsequent results and hydrological evaluation sections.

3.5. Evaluation Metrics for Precipitation Accuracy and Hydrological Simulation

3.5.1. Evaluation of Precipitation Product Accuracy

To evaluate the performance of the original satellite products and the model, this study adopts the correlation coefficient (CC), root mean square error (RMSE), mean absolute error (MAE), and relative bias (RB) as accuracy metrics for precipitation products and uses them to compare the performance of different datasets, as defined below:

$$CC = \frac{\sum_{i=1}^n (P_{obs,i} - P_{obs,avg})(P_{RS,i} - P_{RS,avg})}{\sqrt{\sum_{i=1}^n (P_{obs,i} - P_{obs,avg})^2} \sqrt{\sum_{i=1}^n (P_{RS,i} - P_{RS,avg})^2}} \quad (11)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_{obs,i} - P_{RS,i})^2}{n}} \quad (12)$$

$$MAE = \frac{\sum |P_{obs,i} - P_{RS,i}|}{n} \quad (13)$$

$$Rbias = \frac{\sum_{i=1}^n (P_{RS,i} - P_{obs,i})}{\sum_{i=1}^n P_{obs,i}} \times 100\% \quad (14)$$

Here, $P_{obs,i}$ represents the observed value at the station, while $P_{RS,i}$ denotes the value from the original precipitation product or the fused precipitation prediction. $P_{obs,avg}$ is the average of the observed station values, and $P_{RS,avg}$ is the average of the precipitation product values or the fused precipitation predictions.

Probability of detection (POD), false alarm ratio (FAR), and critical success index (CSI) are widely used event-based metrics for evaluating the detection skill and reliability of precipitation products. POD measures the ability to capture observed precipitation events (range: 0–1), with values closer to 1 indicating fewer missed events. FAR quantifies the proportion of false alarms among predicted precipitation events, where lower values indicate better control of spurious detections. CSI provides an overall measure of event-detection performance by jointly considering hits, misses, and false alarms, representing the fraction of correctly detected events among all occasions where an event was observed and/or predicted. In addition, the frequency bias index (FBI) is reported to diagnose the frequency bias of extreme-event occurrence at a given threshold T (here $T = 50 \text{ mm} \cdot \text{d}^{-1}$). FBI is defined as the ratio of the number of predicted exceedance events to the number of observed exceedance events; $FBI > 1$ indicates overestimation of event frequency, whereas $FBI < 1$ indicates underestimation. Given the limited sample size of $\geq 50 \text{ mm} \cdot \text{d}^{-1}$ events, FBI is used as a supplementary descriptor and interpreted cautiously. The formulas for these metrics are as follows:

$$POD = \frac{H}{H+M} \quad (15)$$

$$FAR = \frac{F}{H+F} \quad (16)$$

$$CSI = \frac{H}{H+M+F} \quad (17)$$

$$FBI(T) = \frac{H(T)+F(T)}{H(T)+M(T)} = \frac{N_{pred \geq T}}{N_{obs \geq T}} \quad (18)$$

Here, H represents the number of hits, i.e., the instances where precipitation actually occurred and was successfully detected by the product; M denotes the number of misses, i.e., the instances where precipitation occurred but was not detected by the product; F indicates the number of false alarms, i.e., the instances where no precipitation occurred but was erroneously detected by the product. Their values range from 0 to 1. The corresponding event-detection metrics (POD, FAR, and CSI) are ratios derived from H , M , and F , and thus range from 0 to 1. In addition, the frequency bias index (FBI) is

used to quantify the bias in extreme-event occurrence frequency at a threshold T . Values greater than 1 indicate overestimation of extreme-event frequency, whereas values less than 1 indicate underestimation.

3.5.2. Evaluation of Hydrological Runoff Simulation

This study utilized the coefficient of determination (R^2) and the Nash-Sutcliffe efficiency (NSE) coefficient to assess the accuracy of runoff flow simulations. The formulas for these metrics are as follows:

$$R^2 = \left(\frac{\sum_{i=1}^n (Q_{obs,i} - \overline{Q_{obs,i}})(Q_{RS,i} - \overline{Q_{RS,i}})}{\sqrt{\sum_{i=1}^n (Q_{obs,i} - \overline{Q_{obs,i}})^2} \sqrt{\sum_{i=1}^n (Q_{RS,i} - \overline{Q_{RS,i}})^2}} \right)^2 \tag{19}$$

$$NSE = 1 - \frac{\sum_{i=1}^n (Q_{obs,i} - Q_{RS,i})^2}{\sum_{i=1}^n (Q_{obs,i} - \overline{Q_{obs,i}})^2} \tag{20}$$

Here, $Q_{obs,i}$ represents the observed runoff value, $Q_{RS,i}$ denotes the simulated runoff value, $\overline{Q_{obs,i}}$ is the average observed runoff, and $\overline{Q_{RS,i}}$ is the average simulated runoff.

4. Results

4.1. Comparison of Deep Learning-Based Correction Models

Building on the explicit designation of IMERG as the primary input product, we first compare the precipitation-correction performance of different deep learning architectures under a unified data setting, so as to avoid any logical coupling between product choice and model structure. In this subsection, only IMERG and its surrounding 15×15 neighborhood sub-grid information are used to construct three models—LSTM, CNN-LSTM, and A-CNN-LSTM—which are evaluated using station-based eight-fold spatial cross-validation, as illustrated in Figure 3.

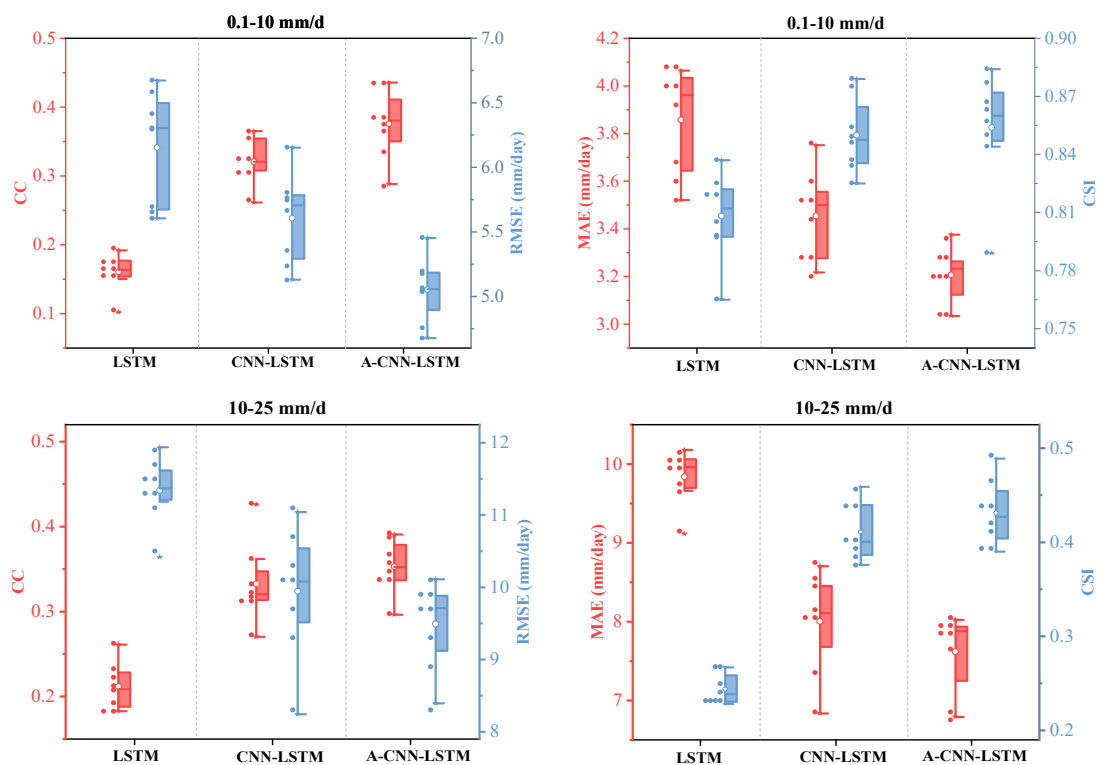


Figure 3. Cont.

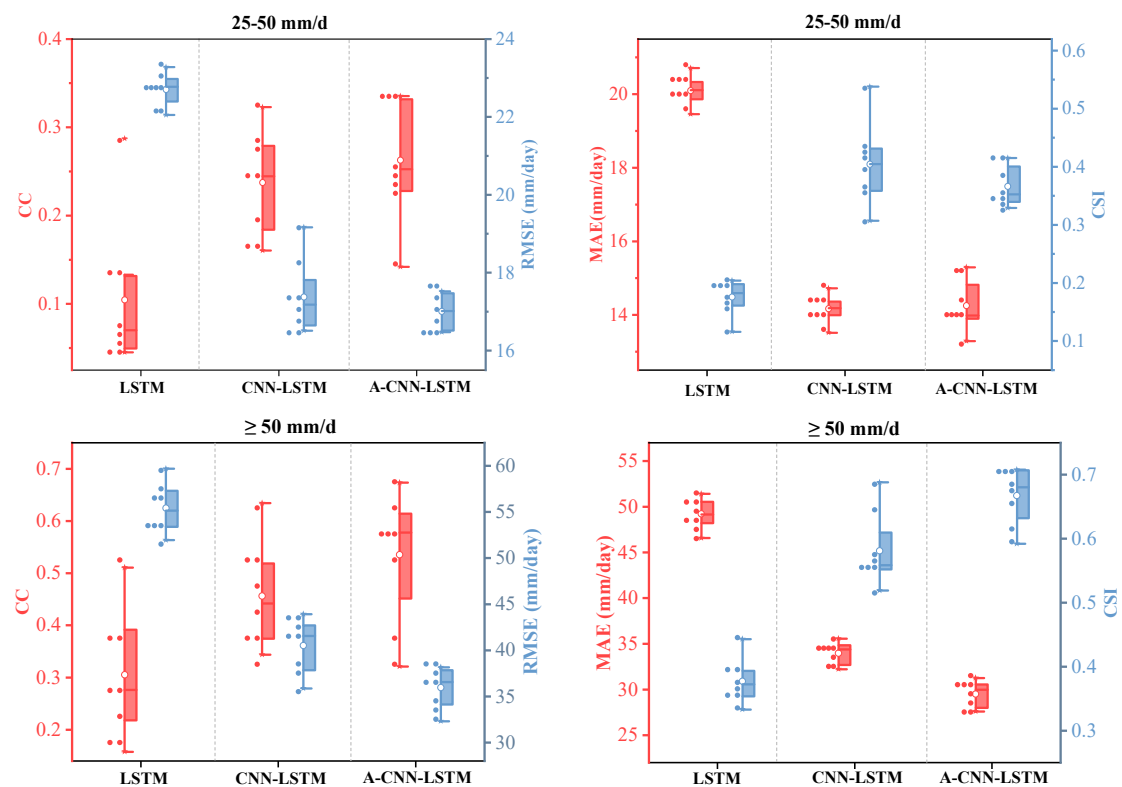


Figure 3. Accuracy assessment of the three models under different precipitation intensities. (left) CC and RMSE. (right) MAE and CSI. Rainfall-intensity classes are 0.1–<10, 10–<25, 25–<50, and ≥ 50 mm·d⁻¹.

Overall, the three models all reduce the systematic biases of IMERG to varying degrees and improve its agreement with gauge observations. The LSTM, which relies solely on temporal information, can reduce bias and RMSE at some stations, but it does not adequately capture spatial heterogeneity over complex terrain or the response to localized heavy rainfall, leading to pronounced performance variability across test stations in the cross-validation. Building on this, the CNN-LSTM introduces convolutional layers to extract neighborhood spatial structure, thereby improving the representation of precipitation spatial patterns and rainband locations; as a result, correlation coefficients increase, RMSE and MAE decrease, and spatial generalization is superior to that of the LSTM.

The A-CNN-LSTM model further augments the CNN-LSTM convolution–temporal framework with temporal and feature attention, assigning larger weights to key periods and influential predictors. It achieves the best and most stable performance in most folds and at most stations: correlation coefficients are further increased, error metrics are reduced overall, and the dispersion of performance among test stations is diminished. These results indicate that, when IMERG is used as the input, A-CNN-LSTM can more effectively exploit neighborhood sub-grid information while suppressing redundant noise and is therefore the most suitable unified correction framework for subsequent comparisons of product configurations and multi-source fusion schemes.

4.2. Correction Performance Under Different Rainfall Intensities

To evaluate the ability of the different models to represent hydrologically sensitive precipitation ranges, and under the common setting that IMERG is used as the sole input product, the raw IMERG field and the three corrected products are stratified by daily rainfall classes and subjected to categorical statistical analysis (Figure 3).

For the light-rain category (e.g., $0.1\text{--}10\text{ mm}\cdot\text{d}^{-1}$), the LSTM, CNN-LSTM, and A-CNN-LSTM models all reduce the systematic biases of the raw IMERG product to some extent, with no pronounced differences among the three. The A-CNN-LSTM does not exhibit any marked suppression of light-rain frequency or excessive smoothing, indicating that the overall performance gain is not achieved at the expense of light-rain events.

For the moderate- and heavy-rain categories (e.g., $10\text{--}25\text{ mm}\cdot\text{d}^{-1}$ and $25\text{--}50\text{ mm}\cdot\text{d}^{-1}$), the superiority of A-CNN-LSTM becomes more evident. Compared with raw IMERG, the intensity biases on moderate- to heavy-rain days are substantially reduced, and the corrected products show improved event-level agreement with gauge observations. Relative to the LSTM and CNN-LSTM, the A-CNN-LSTM attains higher correlation coefficients, lower RMSE and MAE, and overall better critical success index (CSI) values in these classes, and its performance is more stable across the folds of the spatial cross-validation. These findings indicate that the attention mechanisms, combined with neighborhood sub-grid information, effectively enhance the representation of precipitation events that exert a critical influence on flood processes [54,55].

For the extreme-rainfall category (e.g., $\geq 50\text{ mm}\cdot\text{d}^{-1}$), the limited sample size leads to considerable uncertainty in the results of all models; nevertheless, the A-CNN-LSTM generally maintains higher event-level agreement than the other two models, consistent with the overall improvements and the subsequent runoff simulations. To provide a concise quantitative summary of the intensity-stratified results in Figure 3, Table A2 reports the mean \pm standard deviation of CSI across the eight folds. Overall, the A-CNN-LSTM achieves the highest CSI in the light-rain ($0.1\text{--}<10\text{ mm}\cdot\text{d}^{-1}$) and extreme-rainfall ($\geq 50\text{ mm}\cdot\text{d}^{-1}$) classes, while the CNN-LSTM attains comparable performance and slightly higher CSI in the $25\text{--}<50\text{ mm}\cdot\text{d}^{-1}$ class. For the $\geq 50\text{ mm}\cdot\text{d}^{-1}$ class, the larger dispersion among folds reflects limited event samples, and the differences should therefore be interpreted as indicative rather than definitive. This uncertainty is further contextualized by Table A3, which lists the sample sizes of each rainfall-intensity class at the eight gauges; notably, the $\geq 50\text{ mm}\cdot\text{d}^{-1}$ class contains only 44–65 days per station (about 1.0–1.5% of the 4383-day record), and is therefore more sensitive to a small number of events. To further quantify the frequency bias of extreme precipitation ($P \geq 50\text{ mm}\cdot\text{d}^{-1}$), the FBI values are summarized in Table A4.

4.3. Impact of Introducing an ERA5 Evaporation Variable

To quantify the marginal contribution of auxiliary environmental variables to precipitation bias correction and assess the model's sensitivity to auxiliary feature selection, we conducted an ablation study using ERA5 evaporation, i.e., the sign-corrected ET feature. Experiments were performed under the selected best-performing framework. We kept the network architecture, input window, training protocol, and the station-based 8-fold spatial cross-validation split identical and varied only the input features: with ET (including ERA5 ET) versus w/o ET (excluding ERA5 ET). Thus, differences between the two runs can be interpreted as the net effect of the ET auxiliary predictor.

Table 2 summarizes the performance at eight independent test stations under the two settings, including the correlation coefficient (CC), root mean square error (RMSE), mean absolute error (MAE), and event-detection metrics (POD, FAR, and CSI). Overall, removing ET leads to only marginal performance changes, indicating that the correction skill primarily stems from the precipitation inputs and model architecture, while ET serves as a supplementary constraint in the current framework. Specifically, relative to with ET, w/o ET slightly reduces the mean CC from 0.807 to 0.804 ($\Delta \approx -0.003$; range: -0.010 to $+0.002$), increases RMSE from 6.70 to $6.73\text{ mm}\cdot\text{d}^{-1}$ ($\Delta \approx +0.024\text{ mm}\cdot\text{d}^{-1}$; range: 0.000 to $+0.060\text{ mm}\cdot\text{d}^{-1}$), and increases MAE from 3.00 to $3.02\text{ mm}\cdot\text{d}^{-1}$ ($\Delta \approx +0.018\text{ mm}\cdot\text{d}^{-1}$; range: $+0.003$ to

+0.060 mm·d⁻¹). For event detection, POD and CSI decrease slightly (mean $\Delta \approx -0.006$ and -0.002 , respectively), whereas FAR shows a small increase (mean $\Delta \approx +0.005$). These results suggest that the ERA5 evaporation (ET) feature provides limited gains for some stations and metrics, but it does not dominate overall performance and does not alter the main conclusions regarding correction effectiveness and input-scheme comparisons.

Table 2. Evaluation of the Contribution of the ERA5 Evaporation (ET) Predictor.

Metric	with ET (Mean \pm SD)	with ET (Median)	w/o ET (Mean \pm SD)	w/o ET (Median)	Δ Mean (w/o-with)
CC	0.807 \pm 0.013	0.805	0.804 \pm 0.014	0.799	-0.003
RMSE (mm·d ⁻¹)	6.702 \pm 0.359	6.739	6.726 \pm 0.363	6.779	+0.024
MAE (mm·d ⁻¹)	3.004 \pm 0.181	3.000	3.023 \pm 0.186	3.024	+0.019
POD	0.943 \pm 0.023	0.944	0.937 \pm 0.023	0.940	-0.006
FAR	0.249 \pm 0.018	0.253	0.253 \pm 0.016	0.257	+0.006
CSI	0.627 \pm 0.011	0.630	0.624 \pm 0.011	0.627	-0.003

Note: Statistics are computed from the test-station results of the 8-fold spatial cross-validation; Δ denotes the change after removing ET (w/o – with).

To visualize the station-wise changes, Figure 4 presents paired comparisons of CC and RMSE between the with ET and w/o ET settings. For clarity in the annotations, the eight rain-gauge stations are abbreviated as: MPL (Mapoling), LY (Liuyang), ZZ (Zhuzhou), LL (Liling), WZ (Wanzai), YC (Yichun), PX (Pingxiang), and LH (Lianhua). Overall, the curves under the two settings nearly overlap for most stations, further indicating low sensitivity to the ET predictor and that its contribution is marginal.

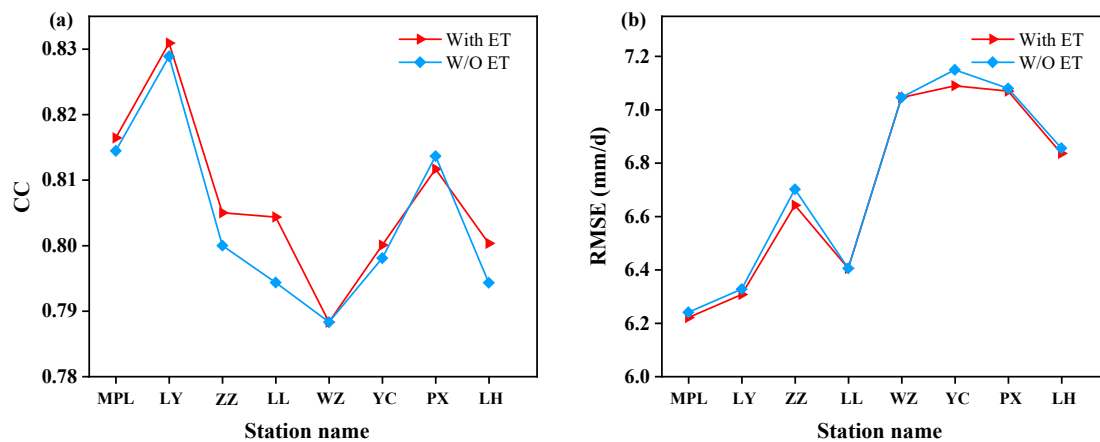


Figure 4. Station-wise paired comparisons between the with-ET and w/o-ET settings: (a) CC; (b) RMSE (mm·d⁻¹).

4.4. Comparison Between Single-Source and Multi-Source Input Schemes

Before comparing different input combinations, we first carried out a brief evaluation of the original daily precipitation accuracy of IMERG and CHIRPS. Using daily observations from rain gauges within and around the Lushui River Basin as the reference, we computed the correlation coefficient (CC), root mean square error (RMSE), mean absolute error (MAE), probability of detection (POD), false alarm ratio (FAR), and critical success index (CSI). The corresponding results for IMERG and CHIRPS are shown as the two bar sets in Figure 5. It is evident that IMERG exhibits higher correlations and smaller RMSE and MAE at most stations, and its ability to detect moderate-to-heavy rainfall events is clearly superior to that of CHIRPS. In contrast, CHIRPS daily precipitation generally shows positive bias and event mismatches, together with considerable inter-station variability. These findings are broadly consistent with the understanding outlined in the Introduction, namely that

IMERG is suitable to serve as the primary satellite precipitation product, whereas CHIRPS is better used as a comparative and auxiliary information source representing infrared–gauge blended products [56,57].

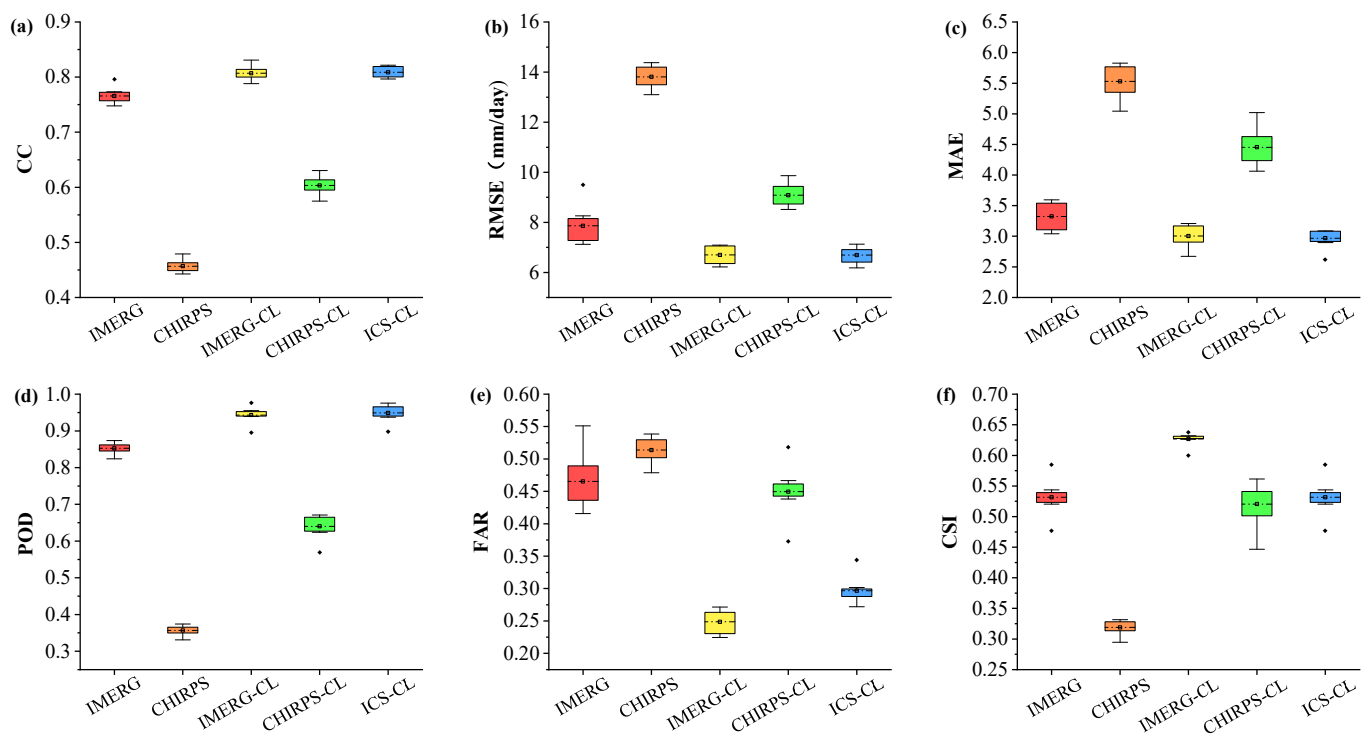


Figure 5. Comparison of performance metrics for raw and corrected precipitation products (daily scale): (a) CC; (b) RMSE; (c) MAE; (d) POD; (e) FAR; (f) CSI.

After identifying A-CNN-LSTM as the precipitation-correction framework, this subsection further examines how different combinations of input products affect the correction performance. Under a consistent network architecture, loss function, and training strategy, three input scenarios are defined: (i) a single-source scheme using only IMERG as input (denoted IMERG-CL); (ii) a single-source scheme using only CHIRPS as input (CHIRPS-CL), intended for comparison with raw CHIRPS to assess the ability of the same deep learning framework to improve the accuracy of a comparatively lower-accuracy infrared–gauge blended product; and (iii) a multi-source scheme that simultaneously ingests IMERG and CHIRPS (ICS-CL). In all three schemes, ERA5 evaporation, DEM-based variables, and other auxiliary factors are included, while the primary precipitation inputs differ. Station-based eight-fold spatial cross-validation is used to compare the performance of each scenario in terms of CC, RMSE, MAE, POD, FAR, and CSI, and the corresponding results are likewise summarized in Figure 5.

As shown in Figure 5, IMERG-CL exhibits relatively stable performance across stations. Compared with raw IMERG, the correlation coefficients increase overall, RMSE and MAE decrease markedly, POD remains at a high level, FAR is reduced, and CSI is improved, indicating that applying spatio-temporal deep learning correction to a single high-quality microwave product can effectively enhance daily precipitation accuracy. CHIRPS-CL also shows some improvement relative to raw CHIRPS; however, owing to the initial systematic biases and local errors in CHIRPS, its overall accuracy and stability remain clearly inferior to those of IMERG-CL, further suggesting that the intrinsic quality of the input product constrains the upper bound of deep learning-based correction [58,59]. It is worth noting that, at some stations and for certain metrics, the relative improvement of CHIRPS-CL over

its original product even exceeds that of IMERG-CL over raw IMERG. This indicates that the same deep learning framework can also offer substantial potential for enhancement when applied to satellite precipitation products with lower initial accuracy.

The multi-source input scheme ICS-CL exhibits modest improvements over IMERG-CL for some metrics; for example, the CC and CSI at certain stations increase slightly, while RMSE and MAE are further reduced. This suggests that, in specific regions and periods, part of the information contained in CHIRPS still provides a supplementary benefit to IMERG. At the same time, because CHIRPS in the study area suffers from positive bias and temporal instability, the performance of ICS-CL is more sensitive to station and year. In some cases, the improvement is limited, and at a few stations a slight degradation is even observed, indicating that its transferability and robustness are inferior to those of the single-source IMERG-CL [60].

Considering overall accuracy, stability, and data availability, this study ultimately adopts the A-CNN-LSTM correction product based on single-source IMERG (denoted IMERG-A) as the primary recommended scheme for subsequent runoff simulations and application analyses. In contrast, CHIRPS-CL and ICS-CL are treated as comparative scenarios for examining the sensitivity of the results to input quality and multi-source fusion, and for discussing how different products and input combinations influence the performance of the deep learning-based correction.

4.5. Parameter Sensitivity and Runoff Simulation Response

4.5.1. Parameter Sensitivity Analysis and Calibration Results

Based on daily streamflow observations, this study conducts parameter sensitivity analysis and calibration for the Lushui River Basin using the previously constructed SWAT model. The SUFI-2 algorithm implemented in SWAT-CUP is employed to perform sensitivity tests on a set of representative parameters that are closely related to runoff generation, groundwater processes, and channel routing, and the results are summarized in Table 3 [61].

Table 3. Parameter sensitivity analysis.

Parameters	Daily	
	T-Stat	p-Value
CN2	12.163	0.000
ALPHA_BF	10.312	0.000
GW_DELAY	−9.232	0.000
GW_REVAP	5.522	0.000
CANMX	−3.239	0.001
CH_N2	−3.236	0.001
CH_K2	−2.201	0.028
SOL_K	−1.733	0.084
GWQWN	1.617	0.106
ESCO	−1.316	0.189
SURLAG	−1.156	0.248
SOL_BD	0.634	0.527

As shown in Table 3, the model output at the daily scale is most sensitive to the following parameters: the SCS runoff curve number (CN2), baseflow recession constant (ALPHA_BF), groundwater delay time (GW_DELAY), groundwater “revap” coefficient (GW_REVAP), maximum canopy storage (CANMX), Manning’s roughness coefficient for the main channel (CH_N2), and saturated hydraulic conductivity of the main channel (CH_K2). Specifically, CN2 directly controls runoff generation under storm conditions; ALPHA_BF and GW_DELAY govern the response speed and lag of baseflow to antecedent

precipitation; GW_REVAP represents the strength of groundwater “revap” to the soil layer; CANMX affects canopy interception and the resulting effective precipitation; and CH_N2 and CH_K2 determine channel routing velocity and flood-peak characteristics. All of these parameters exhibit high statistical sensitivity, indicating that they are key controlling factors in representing the hydrological processes of the Lushui River Basin.

On the basis of the identified sensitive parameters, preliminary calibrations were carried out by driving SWAT with different precipitation datasets, yielding a corresponding optimal parameter set for each case; the results are listed in Table 4. It can be seen that, under different precipitation forcings—including raw IMERG, LSTM-corrected IMERG, CNN-LSTM-corrected IMERG, and A-CNN-LSTM-corrected IMERG—the optimal values of some sensitive parameters differ to varying degrees, suggesting that the quality of the precipitation input is reflected in the representation of runoff generation and routing processes through parameter adjustment. Overall, the optimal values of all parameters fall within reasonable physical ranges. Moreover, the calibrated values of key parameters such as CN2, ALPHA_BF, and GW_REVAP are consistent with the characteristics of a humid monsoon climate and hilly–mountainous topography—namely, pronounced surface runoff and relatively rapid baseflow response—indicating that the constructed SWAT model is physically consistent and credible.

Table 4. Optimal Values of Sensitive Parameters for Different Datasets.

Parameters	Daily			
	IMERG	LSTM	CNN-LSTM	A-CNN-LSTM
CN2	−0.269	−0.068	0.031	0.255
ALPHA_BF	0.980	0.954	0.857	0.059
GW_DELAY	0.175	1.352	0.609	373.428
GW_REVAP	0.159	0.184	0.052	0.154
CANMX	11.172	2.715	6.373	18.297
CH_N2	0.124	0.166	0.171	0.081
CH_K2	189.913	74.931	173.221	102.023
SOL_K	−0.498	−0.489	−0.462	−0.276
GWQWN	1232.480	1207.927	850.960	4307.956
ESCO	0.452	0.265	0.316	0.797
SURLAG	6.057	10.407	17.375	9.651
SOL_BD	1.805	2.270	1.928	1.178

Given that the primary objective of this study is to compare the impacts of different precipitation products and their deep learning-based correction schemes on runoff simulations, rather than to investigate the extent to which parameter adjustments can compensate for precipitation errors, calibrating a separate parameter set for each precipitation dataset would risk masking the true differences among precipitation inputs through parameter-compensation effects. In light of the preliminary calibration results, this study adopts the parameter set derived from the A-CNN-LSTM-corrected precipitation as a unified parameter set for the SWAT model. Subsequent runoff simulations and comparative analyses under the various precipitation-input scenarios are all carried out using this unified parameter set, so as to highlight, as far as possible, the influence of the precipitation schemes themselves on runoff simulation performance.

4.5.2. Runoff Simulations and Integrated Evaluation

To comprehensively evaluate the effectiveness of different precipitation schemes from a hydrological-response perspective, five precipitation-forcing scenarios are configured using the calibrated SWAT model: raw IMERG (IMERG), interpolated gauge observations (OBS), the LSTM-corrected IMERG product (IMERG-LSTM), the CNN-LSTM-corrected IMERG product (IMERG-CNN-LSTM), and the A-CNN-LSTM-corrected IMERG product (IMERG-A). The evaluation results for the calibration period (P1) and validation period (P2) are summarized in Table 5, and the corresponding daily hydrographs are compared in Figure 6.

Table 5. Evaluation of runoff simulation using different precipitation Datasets.

Product	R ²		NSE		RMSE (m ³ /s)	
	P1	P2	P1	P2	P1	P2
IMERG	0.72	0.68	0.71	0.70	71.87	87.92
OBS	0.83	0.80	0.83	0.79	61.02	71.72
IMERG-LSTM	0.75	0.70	0.75	0.72	66.05	74.48
IMERG-CNN-LSTM	0.80	0.75	0.78	0.75	65.70	65.24
IMERG-A-CNN-LSTM	0.85	0.80	0.85	0.79	61.74	60.98

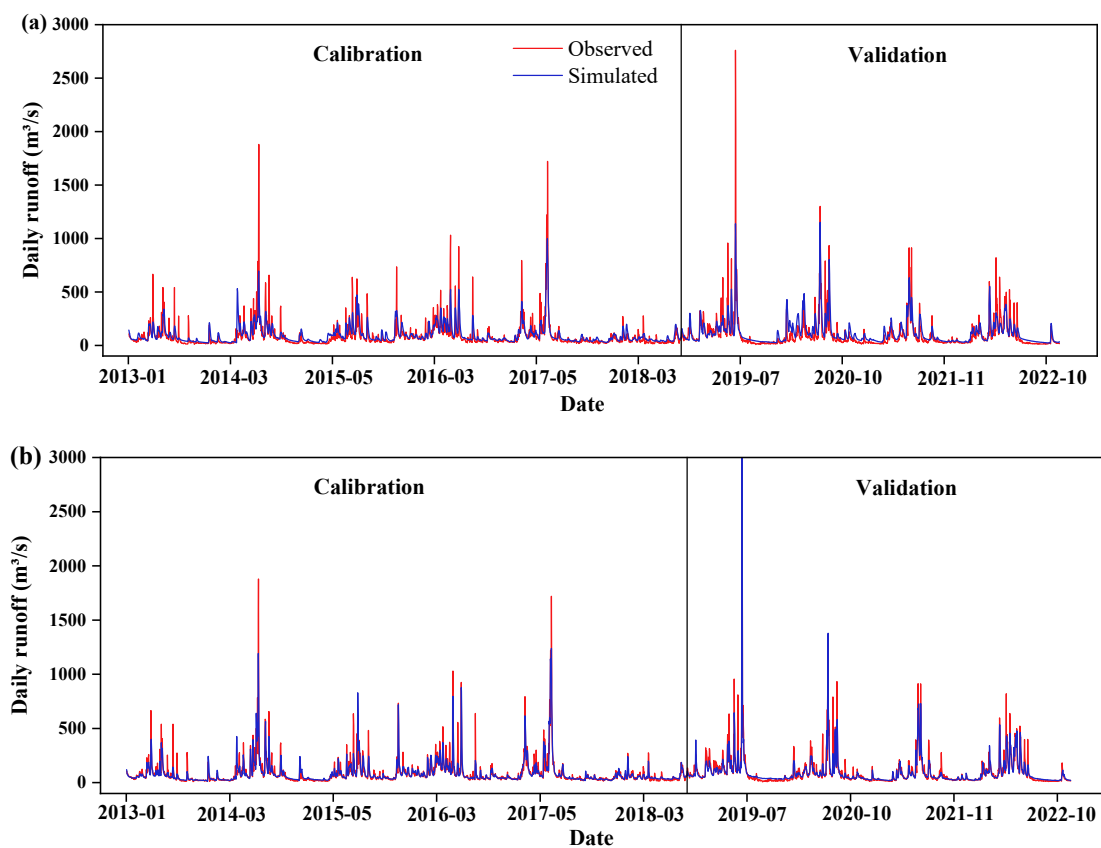


Figure 6. Cont.

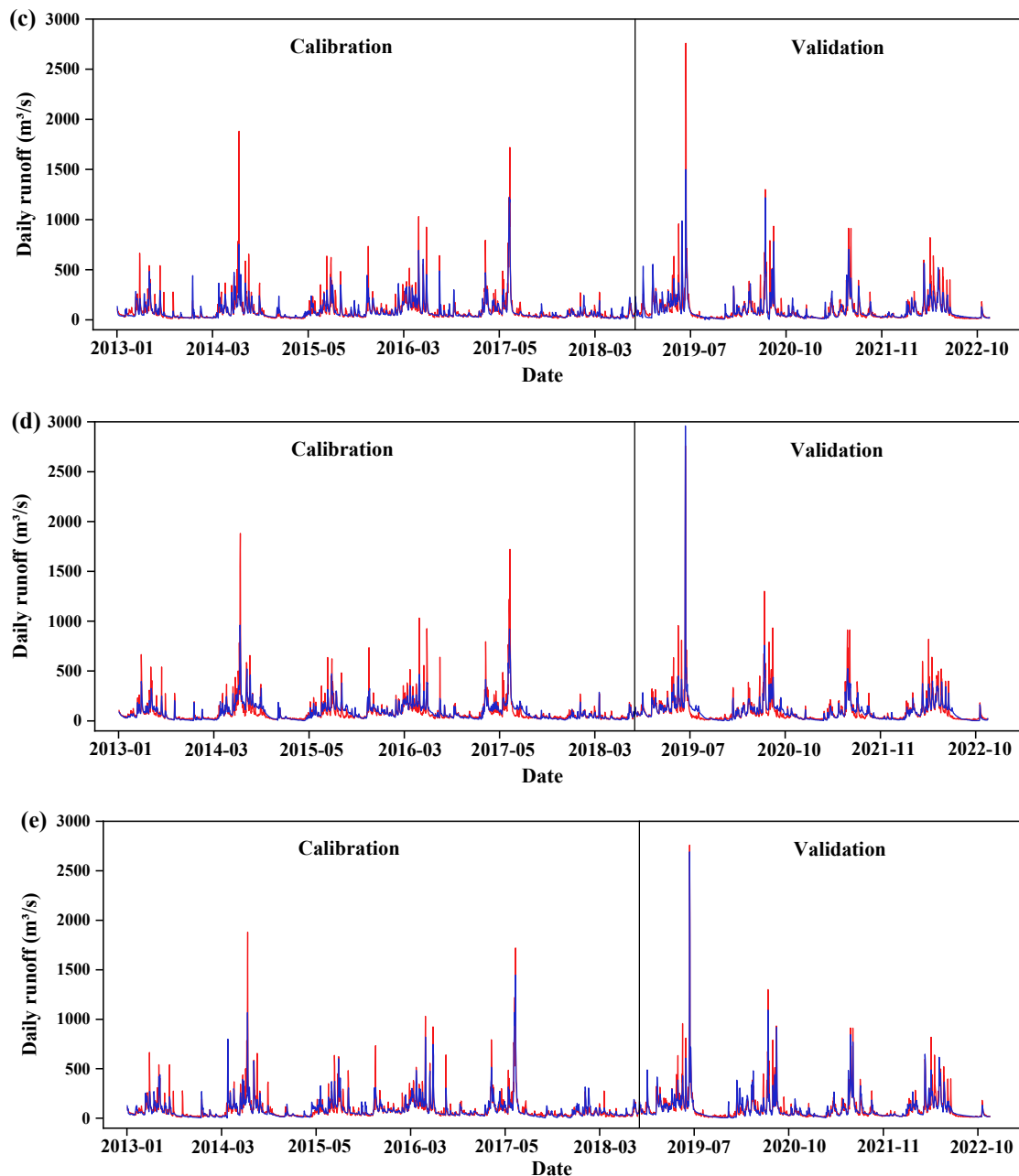


Figure 6. Runoff simulation results driven by different precipitation datasets: (a) IMERG; (b) OBS; (c) IMERG-LSTM; (d) IMERG-CNN-LSTM; (e) IMERG-A.

Under the OBS scenario, R^2 reaches 0.83 and 0.80 for P1 and P2, respectively, NSE is 0.83 and 0.79, and RMSE is markedly lower than for the other schemes ($61.02 \text{ m}^3 \text{ s}^{-1}$ in P1 and $71.72 \text{ m}^3 \text{ s}^{-1}$ in P2), which can be regarded as a “reference upper bound” given the existing gauge network. By contrast, the simulations driven by raw IMERG exhibit noticeably weaker performance: the NSE for P1/P2 is about 0.71/0.70, R^2 is 0.72/0.68, and RMSE is considerably higher (71.87 and $87.92 \text{ m}^3 \text{ s}^{-1}$), with some flood peaks being underestimated and temporally shifted, indicating that directly using uncorrected satellite precipitation for runoff simulation in small- to medium-sized basins still entails substantial uncertainty.

With the introduction of deep learning-based correction, runoff simulation performance improves progressively. For IMERG-LSTM, the NSE for P1/P2 increases to 0.75/0.72 and the RMSE decreases to $66.05/74.48 \text{ m}^3 \text{ s}^{-1}$, indicating that part of the systematic bias is removed, although the improvement in peak-flow reproduction remains

limited. IMERG-CNN-LSTM further enhances the simulations by exploiting neighborhood spatial information: NSE rises to 0.78 and 0.75, the RMSE in P2 drops markedly to about $65.24 \text{ m}^3 \text{ s}^{-1}$, and the simulated daily hydrographs more closely track the observations, demonstrating that convolutional extraction of spatial structure makes a substantial contribution to the hydrological response.

On this basis, the IMERG-A scenario performs best. Its R^2 values for the calibration and validation periods reach 0.85 and 0.80, respectively, and the corresponding NSE values are 0.85 and 0.79, all close to those of the OBS scenario; the RMSE is 61.74 and $60.98 \text{ m}^3 \text{ s}^{-1}$, with the validation-period error already lower than that of the OBS interpolation. As shown in Figure 6, the runoff series driven by IMERG-A closely matches the observed hydrograph in terms of transitions between low- and high-flow conditions, ordinary flood peaks, and several extreme events; both peak magnitude and time to peak exhibit markedly reduced biases, and the model performance remains highly consistent between the calibration and validation stages.

Taken together, Table 5 and Figure 6 yield two key insights. First, deep learning-based correction can substantially transform the unfavorable errors in raw IMERG into high-quality precipitation inputs suitable for basin-scale simulation. The purely temporal LSTM brings only limited gains, whereas the inclusion of spatial convolution leads to clear improvements, and the attention-augmented A-CNN-LSTM achieves the best overall robustness and accuracy. Second, the IMERG-based A-CNN-LSTM correction product (IMERG-A) already approaches, and in some respects even surpasses, the gauge-interpolated OBS scenario in runoff simulation. This indicates that, in small- to medium-sized basins with sparse gauge networks, targeted correction of a high-quality satellite product can deliver hydrological responses comparable to those from interpolated observations while preserving spatial continuity, thereby providing a practically feasible pathway for flood forecasting and water-resources assessment in ungauged or poorly gauged regions.

Second, comparison of the model architecture shows that the ways in which different networks exploit spatio-temporal information lead to marked differences in their representation of moderate-to-heavy rainfall. The LSTM, which relies solely on temporal sequences, can achieve a certain level of overall correlation and error performance, but its ability to reconstruct localized storm centers and spatial gradients is limited. By incorporating neighborhood convolutions, the CNN-LSTM can more effectively utilize surrounding grid information and substantially improve the spatial distribution of moderate-to-heavy rainfall. Building on this, the A-CNN-LSTM employs attention mechanisms to emphasize key periods and key regions, further enhancing the detection of intense precipitation events and the associated runoff response. Taken together, these comparisons indicate that, for precipitation correction in small mountainous basins, fully exploiting neighborhood spatial information and incorporating physically meaningful auxiliary predictors such as topography and evapotranspiration is more beneficial than relying solely on temporal information for improving the representation of hydrologically sensitive rainfall (e.g., moderate-to-heavy rain). At the same time, the temporal and spatial weights produced by the attention mechanisms provide an intuitive indication of the model's focus on different precipitation processes and regions, laying a basis for subsequent feature-importance analysis and for exploring the relationships between the learned weight patterns and terrain or land-surface characteristics, as well as facilitating comparison between deep learning-based corrections and traditional hydrological understanding.

Finally, from the perspective of the “precipitation correction–runoff response” chain, the NSE and RMSE of SWAT simulations driven by IMERG-A are already comparable to those obtained under the gauge-interpolated OBS scenario. This indicates that, when the characteristics of basin physiography and runoff generation–routing are properly ac-

counted for, targeted correction of satellite precipitation can support reasonably reliable runoff simulations in small mountainous catchments. However, the differences among precipitation scenarios in terms of flood-peak magnitude, timing, and spatial distribution also show that improvements in the precipitation field are not propagated linearly into runoff. For example, during some storm events, IMERG-A reproduces peak magnitude close to the gauge-interpolated case but still exhibits biases in time to peak or in the relative contributions of tributaries; CHIRPS-CL and ICS-CL substantially reduce daily scale errors at some stations but do not yield commensurate improvements in peak-flow simulation. These findings suggest that the sensitivity of hydrological models to precipitation errors is strongly controlled by factors such as basin area, slope, underlying surface conditions, and storm type, and that evaluation of precipitation correction should therefore consider both statistical metrics and the hydrological response of key events. Future work could extend the present framework to multiple basins, diverse climatic regions, and higher temporal resolutions, both to test its robustness under different hydrological settings and model structures and, in combination with uncertainty analysis and scenario-based comparisons, to provide decision-relevant guidance for flood forecasting and water-resources management that is better aligned with practical needs [62].

5. Conclusions

On the basis of the integrated satellite-precipitation evaluation, deep learning-based correction, and SWAT runoff simulations, the main conclusions can be summarized as follows:

(1) IMERG is suitable to serve as the primary satellite precipitation product for the study basin. The initial evaluation shows that, over the Lushui River Basin, IMERG outperforms CHIRPS at daily and monthly scales in terms of correlation, error metrics, and the detection of moderate-to-heavy rainfall events, and more faithfully reproduces the evolution and seasonal distribution of precipitation. By contrast, CHIRPS exhibits positive bias and local instability in this region and is therefore better suited as a comparative and auxiliary information source rather than as the sole primary input.

(2) Combining a high-quality single-source IMERG input with the A-CNN-LSTM model constitutes the most suitable correction strategy under the current conditions. With IMERG as a common input, A-CNN-LSTM yields markedly higher station-scale correction accuracy than LSTM and CNN-LSTM, particularly for moderate-to-heavy rainfall in the $10\text{--}50\text{ mm}\cdot\text{d}^{-1}$ range, where correlations are higher and errors smaller. Within the same A-CNN-LSTM framework, the IMERG-based single-source correction product systematically outperforms corrections based on CHIRPS and the multi-source IMERG + CHIRPS scheme, indicating that the intrinsic quality of the satellite product exerts a primary constraint on deep learning-based correction and that simple multi-source aggregation does not necessarily lead to robust improvements.

(3) Using IMERG-A as precipitation input substantially improves runoff simulations for the Lushui River Basin. Relative to the raw IMERG scenario, SWAT simulations driven by IMERG-A increase NSE from 0.71/0.70 to 0.85/0.79 for the calibration/validation periods and reduce RMSE from 71.87/87.92 to 61.74/60.98 $\text{m}^3\text{ s}^{-1}$, yielding performance comparable to the gauge-interpolated scenario and markedly improving the representation of low–high flow transitions and flood peaks. This demonstrates that, under sparse gauge conditions, targeted deep learning-based correction of high-quality satellite products such as IMERG, when coupled with a distributed hydrological model, provides a viable pathway to enhance precipitation inputs and runoff simulation reliability in small mountainous basins.

Author Contributions: Conceptualization, Z.H. and S.Y.; Methodology, Z.H. and S.Y.; Formal analysis, Z.H.; Writing—original draft preparation, Z.H.; Writing—review and editing, Z.H., C.J., Y.L., S.Y., M.X., Y.Q. and T.X.; Supervision and project administration, C.J. and Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are publicly available from the following sources: [GPM-IMERG] (https://disc.gsfc.nasa.gov/datasets/GPM_3IMERGDF_07/summary?keywords=GPM, accessed on 5 January 2026); [CHIRPS] (https://data.chc.ucsb.edu/products/CHIRPS-2.0/global_daily/netcdf/p05/, accessed on 5 January 2026); [ERA5] (<https://cds.climate.copernicus.eu/datasets/derived-era5-single-levels-daily-statistics?tab=overview>, accessed on 5 April 2026); [DEM] (<https://www.gscloud.cn>, accessed on 5 April 2026); [soil data] (<https://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/harmonized-world-soil-database-v12/en/>, accessed on 6 May 2026); and [land-use data] (<https://www.resdc.cn>, accessed on 6 May 2025).

Acknowledgments: The authors gratefully acknowledge the GPM mission team for providing the IMERG precipitation products and the local hydrological authorities for supplying rain gauge and streamflow observations used in this study. The authors also thank the colleagues and technical staff who assisted with data preprocessing, model calibration, and result visualization. The insightful and constructive comments from the anonymous reviewers are sincerely appreciated and have substantially improved this manuscript.

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

Appendix A. Tables

Table A1. Sensitivity of neighborhood window size (A-CNN-LSTM; eight-fold spatial cross-validation across 8 stations).

Window Size	CC (Mean \pm SD)	RMSE (Mean \pm SD, $\text{mm}\cdot\text{d}^{-1}$)
7 \times 7	0.791 \pm 0.012	6.949 \pm 0.392
9 \times 9	0.797 \pm 0.012	6.856 \pm 0.396
11 \times 11	0.797 \pm 0.012	6.850 \pm 0.386
13 \times 13	0.806 \pm 0.014	6.717 \pm 0.337
15 \times 15	0.807 \pm 0.013	6.702 \pm 0.359

Table A2. Critical Success Index (CSI) of the three correction models under different daily precipitation intensity ranges.

Model	0.1–<10	10–<25	25–<50	≥ 50
LSTM	0.408 \pm 0.022	0.244 \pm 0.016	0.175 \pm 0.030	0.377 \pm 0.034
CNN-LSTM	0.600 \pm 0.019	0.411 \pm 0.031	0.404 \pm 0.069	0.581 \pm 0.056
A-CNN-LSTM	0.654 \pm 0.029	0.431 \pm 0.035	0.366 \pm 0.035	0.667 \pm 0.045

Table A3. Sample sizes (number of days) of gauge precipitation in each rainfall-intensity class at the eight stations (2011–2022).

Station	0.1–<10	10–<25	25–<50	≥ 50	Total Wet Days (≥ 0.1)
Mapoling	1367	366	135	44	1912
Liuyang	1305	365	163	60	1893
Zhuzhou	1370	357	146	55	1928
Liling	1347	356	148	51	1902

Table A3. *Cont.*

Station	0.1–<10	10–<25	25–<50	≥50	Total Wet Days (≥0.1)
Wanzai	1345	417	153	65	1980
Yichun	1376	394	175	60	2005
Pingxiang	1412	402	160	64	2038
Lianhua	1409	377	178	49	2013

Note: Rainfall classes are defined as $0.1 \leq P < 10$, $10 \leq P < 25$, $25 \leq P < 50$, and $P \geq 50$ mm·d⁻¹. Boundary values (10, 25, and 50 mm·d⁻¹) are assigned to the higher-intensity class.

Table A4. Frequency bias index (FBI) for extreme precipitation events ($P \geq 50$ mm·d⁻¹) at eight stations (2011–2022).

Model	Mapoling	Liuyang	Zhuzhou	Liling	Wanzai	Yichun	Pingxiang	Lianhua	Pooled
LSTM	0.091	0.050	0.073	0.216	0.154	0.100	0.141	0.245	0.132
CNN-LSTM	0.523	0.383	0.400	0.549	0.446	0.433	0.344	0.653	0.458
A-CNN-LSTM	0.545	0.533	0.509	0.604	0.462	0.583	0.516	0.633	0.544

Note: “Pooled” denotes the FBI computed by pooling all station–day samples across the eight stations (i.e., a single FBI value based on aggregated counts of exceedance events).

References

- Zubieta, R.; Getirana, A.; Espinoza, J.C.; Lavado-Casimiro, W.; Aragon, L. Impacts of Satellite-Based Precipitation Datasets on Rainfall–Runoff Modeling of the Western Amazon Basin of Peru and Ecuador. *J. Hydrol.* **2015**, *528*, 599–612. [[CrossRef](#)]
- Kidd, C.; Becker, A.; Huffman, G.J.; Müller, C.L. So, How Much of the Earth’s Surface Is Covered by Rain Gauges? *Bull. Am. Meteorol. Soc.* **2017**, *98*, 69–78. [[CrossRef](#)] [[PubMed](#)]
- Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al. The ERA5 Global Reanalysis. *Q. J. R. Meteorol. Soc.* **2020**, *146*, 1999–2049. [[CrossRef](#)]
- 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. Am. Meteorol. Soc.* **2019**, *100*, 473–500. [[CrossRef](#)]
- Funk, C.; Peterson, P.; Landsfeld, M.; Pedreros, D.; Verdin, J.; Shukla, S.; Husak, G.; Rowland, J.; Harrison, L.; Hoell, A.; et al. A Global Satellite-Assisted Precipitation Climatology. *Earth Syst. Sci. Data* **2015**, *7*, 275–287. [[CrossRef](#)]
- Maggioni, V.; Massari, C. On the Performance of Satellite Precipitation Products in Riverine Flood Modeling: A Review. *J. Hydrol.* **2018**, *558*, 214–224. [[CrossRef](#)]
- Liu, Y.; Peters-Lidard, C.D.; Kumar, S.V.; Arsenault, K.R.; Mocko, D.M. Blending Satellite-Based Snow Depth Products with In Situ Observations for Streamflow Predictions in the Upper Colorado River Basin. *Water Resour. Res.* **2015**, *51*, 1182–1202. [[CrossRef](#)]
- Wu, H.; Yang, Q.; Liu, J.; Wang, G. A Spatiotemporal Deep Fusion Model for Merging Satellite and Gauge Precipitation in China. *J. Hydrol.* **2020**, *584*, 124664. [[CrossRef](#)]
- Akbari Asanjan, A.; Yang, T.; Hsu, K.; Sorooshian, S.; Lin, J.; Peng, Q. Short-Term Precipitation Forecast Based on the PERSIANN Cloud Classification System. *J. Geophys. Res. Atmos.* **2018**, *123*, 12543–12563. [[CrossRef](#)]
- Yumnam, K.; Guntu, R.K.; Rathinasamy, M.; Agarwal, A. Quantile-Based Bayesian Model Averaging Approach towards Merging of Precipitation Products. *J. Hydrol.* **2022**, *604*, 127206. [[CrossRef](#)]
- Wei, L.; Jiang, S.; Dong, J.; Ren, L.; Liu, Y.; Zhang, L.; Wang, M.; Duan, Z. Fusion of Gauge-Based, Reanalysis, and Satellite Precipitation Products Using Bayesian Model Averaging Approach: Determination of the Influence of Different Input Sources. *J. Hydrol.* **2023**, *618*, 129234. [[CrossRef](#)]
- Sadeghi, M.; Asanjan, A.A.; Faridzad, M.; Nguyen, P.; Hsu, K.; Sorooshian, S.; Braithwaite, D. PERSIANN-CNN: Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks–Convolutional Neural Networks. *J. Hydrometeorol.* **2019**, *20*, 2273–2289. [[CrossRef](#)]
- Shi, X.; Gao, Z.; Lausen, L.; Wang, H.; Yeung, D.-Y.; Wong, W.-K.; Woo, W.-C. Deep Learning for Precipitation Nowcasting: A Benchmark and a New Model. In *Advances in Neural Information Processing Systems 30*; Curran Associates, Inc.: Red Hook, NY, USA, 2017; pp. 5617–5627. [[CrossRef](#)]
- Tian, B.; Chen, H.; Yan, X.; Sheng, S.; Lin, K. A Downscaling–Merging Scheme for Monthly Precipitation Estimation with High Resolution Based on CBAM-ConvLSTM. *Remote Sens.* **2023**, *15*, 4601. [[CrossRef](#)]
- Yang, X.; Yang, S.; Tan, M.L.; Pan, H.; Zhang, H.; Wang, G.; He, R.; Wang, Z. Correcting the Bias of Daily Satellite Precipitation Estimates in Tropical Regions Using Deep Neural Network. *J. Hydrol.* **2022**, *608*, 127656. [[CrossRef](#)]
- Jing, Y.; Lin, L.; Li, X.; Li, T.; Shen, H. An Attention Mechanism Based Convolutional Network for Satellite Precipitation Downscaling over China. *J. Hydrol.* **2022**, *613*, 128388. [[CrossRef](#)]

17. Xiang, L.; Guan, J.; Xiang, J.; Zhang, L.; Zhang, F. Spatiotemporal Model Based on Transformer for Bias Correction and Temporal Downscaling of Forecasts. *Front. Environ. Sci.* **2022**, *10*, 1039764. [[CrossRef](#)]
18. Gao, Y.; Guan, J.; Zhang, F.; Wang, X.; Long, Z. Attention-Unet-Based Near-Real-Time Precipitation Estimation from Fengyun-4A Satellite Imageries. *Remote Sens.* **2022**, *14*, 2925. [[CrossRef](#)]
19. Liu, H.; Yang, Q.; Liu, Z.; Shao, J.; Wang, G. An Attention-Mechanism-Based Deep Fusion Model for Improving Quantitative Precipitation Estimation in a Sparsely-Gauged Basin. *J. Hydrol.* **2024**, *628*, 130568. [[CrossRef](#)]
20. Zhang, L.; Li, X.; Zheng, D.; Zhang, K.; Ma, Q.; Zhao, Y.; Ge, Y. Merging Multiple Satellite-Based Precipitation Products and Gauge Observations Using a Novel Double Machine Learning Approach. *J. Hydrol.* **2021**, *594*, 125969. [[CrossRef](#)]
21. Hu, Y.; Zhang, L. Added Value of Merging Techniques in Precipitation Estimates Relative to Gauge-Interpolation Algorithms of Varying Complexity. *J. Hydrol.* **2024**, *645*, 132214. [[CrossRef](#)]
22. Kao, Y.-C.; Tsou, H.-E.; Chen, C.-J. Development of Multi-Source Weighted-Ensemble Precipitation: Influence of Bias Correction Based on Recurrent Convolutional Neural Networks. *J. Hydrol.* **2024**, *629*, 130621. [[CrossRef](#)]
23. Huffman, G.J.; Bolvin, D.T.; Joyce, R.; Nelkin, E.J.; Tan, J.; Braithwaite, D.; Hsu, K.; Kelley, O.A.; Nguyen, P.; Sorooshian, S.; et al. *Integrated Multi-Satellite Retrievals for GPM (IMERG) Algorithm Theoretical Basis Document (ATBD)*; Version 07; NASA Goddard Space Flight Center: Greenbelt, MD, USA, 2023.
24. Zhang, J.; Xu, J.; Dai, X.; Ruan, H.; Liu, X.; Jing, W. Multi-Source Precipitation Data Merging for Heavy Rainfall Events Based on Cokriging and Machine Learning Methods. *Remote Sens.* **2022**, *14*, 1750. [[CrossRef](#)]
25. Xu, Z.; Wu, Z.; He, H.; Wu, X.; Ren, J. Evaluating the accuracy of MSWEP V2.1 and its performance for drought monitoring over Mainland China. *Atmos. Res.* **2019**, *226*, 17–31. [[CrossRef](#)]
26. Wei, G.; Lü, H.; Crow, W.T.; Zhu, Y.; Wang, J.; Su, J. Comprehensive evaluation of GPM-IMERG, CMORPH, and TMPA precipitation products with gauged rainfall over mainland China. *Adv. Meteorol.* **2018**, *2018*, 3024190. [[CrossRef](#)]
27. Baudouin, J.-P.; Herzog, M.; Petrie, C.A. Cross-validating precipitation datasets in the Indus River basin. *Hydrol. Earth Syst. Sci.* **2020**, *24*, 427–450. [[CrossRef](#)]
28. Lv, A.; Qi, S.; Wang, G. Multi-model driven by diverse precipitation datasets increases confidence in identifying dominant factors for runoff change in a subbasin of the Qaidam Basin of China. *Sci. Total Environ.* **2022**, *802*, 149831. [[CrossRef](#)]
29. Yan, S.; Long, Y.; He, H.; Wen, X.; Lv, Q.; Zheng, M. Flood response to urban expansion in the Lushui River Basin. *Nat. Hazards* **2023**, *115*, 779–805. [[CrossRef](#)]
30. Xu, H.; Xu, C.-Y.; Chen, H.; Zhang, Z.; Li, L. Assessing the influence of rain gauge density and distribution on hydrological model performance in a humid region of China. *J. Hydrol.* **2013**, *505*, 1–12. [[CrossRef](#)]
31. 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](#)]
32. Climate Hazards Center (CHC). *Climate Hazards Center InfraRed Precipitation with Stations*; Version 3 (CHIRPS v3); CHIRPS3 Data Repository: Santa Barbara, CA, USA, 2025. [[CrossRef](#)]
33. Copernicus Climate Change Service (C3S). *ERA5 Post-Processed Daily Statistics on Single Levels from 1940 to Present*; ECMWF, Copernicus Climate Data Store (CDS): Reading, UK, 2024. [[CrossRef](#)]
34. Yang, J.; Huang, X. The 30 m annual land cover dataset and its dynamics in China from 1990 to 2019. *Earth Syst. Sci. Data* **2021**, *13*, 3907–3925. [[CrossRef](#)]
35. Nachtergaele, F.; van Velthuizen, H.; Verelst, L.; Wiberg, D.; Henry, M.; Chiozza, F.; Yigini, Y.; Aksoy, E.; Batjes, N.; Boateng, E.; et al. *Harmonized World Soil Database Version 2.0*; Food and Agriculture Organization of the United Nations (FAO): Rome, Italy; International Institute for Applied Systems Analysis (IIASA): Laxenburg, Austria, 2023. [[CrossRef](#)]
36. Wang, F.; Tian, D.; Carroll, M. Customized Deep Learning for Precipitation Bias Correction and Downscaling in Mountain Regions. *Geosci. Model Dev.* **2023**, *16*, 535–556. [[CrossRef](#)]
37. Le, X.-H.; Lee, G.; Jung, K.; An, H.-u.; Lee, S.; Jung, Y. Application of Convolutional Neural Network for Spatiotemporal Bias Correction of Daily Satellite-Based Precipitation. *Remote Sens.* **2020**, *12*, 2731. [[CrossRef](#)]
38. El Garnaoui, M.; Boudhar, A.; Nifa, K.; El Jabiri, Y.; Karaoui, I.; El Aloui, A.; Midaoui, A.; Karroum, M.; Mosaid, H.; Chehbouni, A. Nested Cross-Validation for HBV Conceptual Rainfall–Runoff Model Spatial Stability Analysis in a Semi-Arid Context. *Remote Sens.* **2024**, *16*, 3756. [[CrossRef](#)]
39. Cheng, Y.-Y.; Chang, C.-T.; Chen, B.-F.; Kuo, H.-C.; Lee, C.-S. Extracting 3D Radar Features to Improve Quantitative Precipitation Estimation in Complex Terrain Based on Deep Learning Neural Networks. *Weather Forecast.* **2023**, *38*, 273–289. [[CrossRef](#)]
40. Gamboa-Villafruela, C.J.; Fernández-Alvarez, J.C.; Márquez-Mijares, M.; Pérez-Alarcón, A.; Batista-Leyva, A.J. Convolutional LSTM Architecture for Precipitation Nowcasting Using Satellite Data. *Environ. Sci. Proc.* **2021**, *8*, 33. [[CrossRef](#)]
41. Guo, Q.; He, Z.; Wang, Z. Monthly Climate Prediction Using Deep Convolutional Neural Network and Long Short-Term Memory. *Sci. Rep.* **2024**, *14*, 17748. [[CrossRef](#)]
42. Guo, L.; Pu, Y.; Zhao, W. CNN-BiLSTM Daily Precipitation Prediction Based on Attention Mechanism. *Atmosphere* **2025**, *16*, 333. [[CrossRef](#)]

43. Dao, V.; Jimenez Arellano, C.; Nguyen, P.; Almutlaq, F.; Hsu, K.; Sorooshian, S. Bias Correction of Satellite Precipitation Estimation Using Deep Neural Networks and Topographic Information Over the Western U.S. *J. Geophys. Res. Atmos.* **2025**, *130*, e2024JD042181. [CrossRef]
44. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958. [CrossRef]
45. Dubey, S.K.; Kim, J.; Her, Y.; Sharma, D.; Jeong, H. Hydroclimatic Impact Assessment Using the SWAT Model in India—State of the Art Review. *Sustainability* **2023**, *15*, 15779. [CrossRef]
46. Taia, S.; Erraioui, L.; Arjdal, Y.; Chao, J.; El Mansouri, B.; Scozzari, A. The Application of SWAT Model and Remotely Sensed Products to Characterize the Dynamic of Streamflow and Snow in a Mountainous Watershed in the High Atlas. *Sensors* **2023**, *23*, 1246. [CrossRef]
47. Her, Y.; Frankenberger, J.; Chaubey, I.; Srinivasan, R. Threshold Effects in HRU Definition of the Soil and Water Assessment Tool. *Trans. ASABE* **2015**, *58*, 367–378. [CrossRef]
48. Femeena, P.V.; Karki, R.; Cibir, R.; Sudheer, K.P. Reconceptualizing HRU Threshold Definition for SWAT to Improve Consistency and Performance. *J. Am. Water Resour. Assoc.* **2022**, *58*, 81–96. [CrossRef]
49. Jiang, L.; Zhu, J.; Chen, W.; Hu, Y.; Yao, J.; Yu, S.; Jia, G.; He, X.; Wang, A. Identification of Suitable Hydrologic Response Unit Thresholds for Soil and Water Assessment Tool Streamflow Modelling. *Chin. Geogr. Sci.* **2021**, *31*, 696–710. [CrossRef]
50. Wang, J.; Zhuo, L.; Han, D.; Liu, Y.; Rico-Ramirez, M.A. Hydrological model adaptability to rainfall inputs of varied quality. *Water Resour. Res.* **2023**, *59*, e2022WR032484. [CrossRef]
51. Anand, V.; Singh, S.K.; Oinam, B. Parameterization and calibration of the SWAT hydrological model using SUFI-2 and GLUE algorithms in Bay of Plenty, New Zealand. *J. Water Clim. Change* **2025**, *16*, 2444–2461. [CrossRef]
52. Xiang, X.; Ao, T.; Xiao, Q.; Li, X.; Zhou, L.; Chen, Y.; Bi, Y.; Guo, J. Parameter sensitivity analysis of SWAT modeling in the upper Heihe River Basin using four typical approaches. *Appl. Sci.* **2022**, *12*, 9862. [CrossRef]
53. Brighenti, T.M.; Gassman, P.W.; Gutowski, W.J., Jr.; Thompson, J.R. Assessing the Influence of a Bias Correction Method on Future Climate Scenarios Using SWAT as an Impact Model Indicator. *Water* **2023**, *15*, 750. [CrossRef]
54. Sneha, M.R.; Nair, A.; Somasundaram, K. Spatiotemporal Bias Correction of Satellite Precipitation Products Using Multimodel Techniques over Temporally Coherent Clusters in South Peninsular India. *Atmos. Res.* **2025**, *325*, 108244. [CrossRef]
55. Yao, N.; Ye, J.; Wang, S.; Yang, S.; Lu, Y.; Zhang, H.; Yang, X. Bias Correction of the Hourly Satellite Precipitation Product Using Machine Learning Methods Enhanced with High-Resolution WRF Meteorological Simulations. *Atmos. Res.* **2024**, *310*, 107637. [CrossRef]
56. Guo, H.; Chen, S.; Bao, A.; Hu, J.; Yang, B.; Stepanian, P.M. Comprehensive Evaluation of High-Resolution Satellite-Based Precipitation Products over China. *Atmosphere* **2016**, *7*, 6. [CrossRef]
57. Du, H.; Tan, M.L.; Zhang, F.; Chun, K.P.; Li, L.; Kabir, M.H. Evaluating the Effectiveness of CHIRPS Data for Hydroclimatic Studies: A Review. *Theor. Appl. Climatol.* **2024**, *155*, 1519–1539. [CrossRef]
58. Baig, F.; Ali, L.; Faiz, M.A.; Chen, H.; Sherif, M. From Bias to Accuracy: Transforming Satellite Precipitation Data in Arid Regions with Machine Learning and Topographical Insights. *J. Hydrol.* **2025**, *653*, 132801. [CrossRef]
59. Azimi, S.; Massari, C.; Roati, G.; Barbeta, S.; Rigon, R. A New Tool for Correcting the Spatial and Temporal Pattern of Global Precipitation Products across Mountainous Terrain: Precipitation and Hydrological Analysis. *J. Hydrol.* **2025**, *660*, 133530. [CrossRef]
60. Shi, Y.; Chen, C.; Chen, J.; Yang, S.; Liu, Y.; Li, M.; Guo, B. Evaluation of the RF-MEP Method for Merging Multiple Gridded Precipitation Products in Chongqing City, China. *Remote Sens.* **2023**, *15*, 4230. [CrossRef]
61. Abbaspour, K.C. *SWAT-CUP: SWAT Calibration and Uncertainty Programs—A User Manual*; Swiss Federal Institute of Aquatic Science and Technology (Eawag): Dübendorf, Switzerland, 2015. Available online: https://swat.tamu.edu/media/114860/usermanual_swatcup.pdf (accessed on 5 January 2026).
62. Chawanda, C.J.; van Griensven, A.; Nkwasa, A.; Teran Orsini, J.P.; Jeong, J.; Choi, S.-K.; Srinivasan, R.; Arnold, J.G. CoSWAT Model v1: A high-resolution global SWAT+ hydrological model. *Hydrol. Earth Syst. Sci.* **2025**, *29*, 6901–6916. [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.