https://doi.org/10.5194/essd-2026-497
Preprint. Discussion started: 22 September 2026

Earth System Science Data Discussions

A global terrestrial water budget discrete grid dataset constrained by GRDC observations (2000–2020)

Kai Li[1,2], Juanle Wang[1,2,5*], Congrong Li[1], Xianglin Ji[3], Lizhi Pan[1,4]

[1]State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

[2]College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China

[3]State Key Laboratory of Earth Surface Process and Resource Ecology, Beijing Normal University, Beijing 100875, China

[4]College of Geoscience and Surveying Engineering, China University of Mining and Technology (Beijing), Beijing 100083, China

[5]Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China

Correspondence to: Juanle Wang (wangjl@igsnrr.ac.cn)

Abstract. With the development of large-scale data-driven research paradigm, spatially continuous and physically consistent gridded water-cycle data have become an essential basis for global hydrological process assessments, model training, and numerical simulations. Previous studies on terrestrial water budget closure have encountered persistent difficulty in simultaneously accounting for the observational realism of basin-scale runoff and the spatial continuity of gridded data. To address this problem, we developed a technical pathway for transferring constraints from real basin-scale observations to global gridded water-cycle variables. This approach overcomes the limitations of traditional basin-scale methods, which cannot adequately describe the internal spatial heterogeneity, and those of gridded-scale methods, which typically lack constraints from real runoff observations. The ISEA3H10 equal-area hexagonal discrete global grid was used as a unified spatial framework for this new dataset. Through basin–grid spatial overlay and area weighting, GRDC basin outlet runoff observations were converted into supervisory information that can be used by a gridded model. A deep-learning model with physical loss constraints was designed to jointly optimize key water-cycle variables, including precipitation, evapotranspiration, runoff, and terrestrial water storage change. The model further introduced natural background variables, including topography, vegetation, land-surface temperature, snow cover, and soil moisture. So that the water-cycle estimation does not rely solely on individual hydrological flux information, but also uses the spatial heterogeneity of more environmental factors to support water budget closure. This study integrated multiple remote-sensing, reanalysis, and land-surface model products, and further incorporated 710,274 GRDC runoff records from 3,471 basins worldwide as observational constraints. Based on these data, this study developed a monthly gridded dataset of precipitation, evapotranspiration, runoff, and terrestrial water storage changes covering global land areas from 2000 to 2020, for a total of 252 months. The optimized gridded runoff had a correlation coefficient of approximately 0.9 with GRDC observations, and its root mean square error (RMSE) was substantially lower than that of the original gridded runoff proxy products. Independent validation at 524 FLUXNET sites worldwide showed that precipitation

https://doi.org/10.5194/essd-2026-497
Preprint. Discussion started: 22 September 2026

and evapotranspiration maintained an accuracy comparable to, or more stable than, that of mainstream input products. Among 672 original *P-E-R-ΔS* product combinations, the combination exhibiting the optimal performance in terms of closure had a global monthly water budget residual RMSE of 43.37 mm/month. The optimized product reduced the RMSE to 14.33 mm/month, with a mean residual of −0.44 mm/month. The dataset preserved the seasonal consistency of each water-cycle variable and the consistency of the land-cover type zonal differences. Typical regions validation indicated that it also captured the coordinated anomalous responses of precipitation, evapotranspiration, runoff, and terrestrial water storage during special flood and drought events.

1 Introduction

The terrestrial water cycle is a key process that links the atmosphere, land surface, and groundwater systems, and its changes directly affect global water resource security (Oki and Kanae, 2006). Therefore, accurately characterizing the relationships between different water-cycle variables is a fundamental scientific issue in hydrology and Earth system science. Based on the principle of mass conservation, the water cycle equation describes the balance between water input and output over a given spatial and temporal scale (Zheng et al., 2025), which is a basic criterion for evaluating the physical consistency and reliability of water-cycle data. However, existing water-cycle variable products are typically derived from different observation platforms, retrieval algorithms, and modeling systems. Their spatiotemporal scales, error sources, and physical assumptions differ substantially (Anjaneyulu et al., 2025), leading to widespread water budget non-closure in multi-source datasets at global and regional scales (Huang et al., 2025; Lehmann et al., 2022).

Closure residuals caused by inconsistencies among multi-source water-cycle variables directly affect the reliability of hydrological process diagnosis and water-resource assessment. Pascolini-Campbell et al. (2020) found that across 11 major basins in the United States, the land-surface model and remote-sensing evapotranspiration products were generally lower than evapotranspiration estimates derived from a GRACE-based water balance. Sheffield et al. (2009) showed that directly using existing precipitation, evapotranspiration, and terrestrial water storage change products to infer runoff substantially overestimated the observed discharge in the Mississippi River Basin, with a bias of ≤ 1.4 mm/day. Michailovsky et al. (2023) also found that estimating runoff from existing water-cycle variables across 931 global basins caused large residuals to be allocated to runoff, with a median Nash-Sutcliffe efficiency coefficient of only −0.02. In contrast, water-cycle variable datasets constrained by closure better satisfy mass conservation requirements, thereby providing a more consistent and reliable basis for global and regional hydrological process analysis, basin water resource assessment, hydrological model calibration, and extreme hydrological event identification.

Given the widespread non-closure and uncertainty in existing water-cycle products, global water-cycle analysis requires the joint use of multi-source water-cycle datasets under physically consistent constraints (Zhang et al., 2026). Previous water-cycle closure studies have mainly been conducted at the basin and grid scales. Basin-scale studies typically aggregate multi-source water budget components within watershed boundaries and use station runoff or river discharge observations such as

https://doi.org/10.5194/essd-2026-497
Preprint. Discussion started: 22 September 2026

GRDC (Global Runoff Data Centre) data to constrain the total water budget (Abolafia-Rosenzweig et al., 2021; Pan et al., 2012). Although outlet-discharge-based methods provide hydrologically meaningful constraints on the overall basin water balance, their focus on basin averages or totals limits their ability to represent the internal spatial heterogeneity of water-cycle variables within a basin (Pellet et al., 2021).

In contrast, gridded data preserve the continuous spatial structures of multi-source remote sensing, reanalysis, and land-surface model products. Existing grid-scale studies commonly use optimal interpolation, constrained Kalman filtering, neural networks, and related methods to distribute water balance residuals (Heberger et al., 2025; Zhang et al., 2018). These studies involve multi source precipitation, evapotranspiration, runoff, and water storage products. In grid products such as ERA5-Land, FLDAS, and GLDAS, runoff generally represents surface runoff, subsurface runoff, or related quantities generated within a cell; whereas, basin outlet discharge is an integrated response formed after upstream runoff generation, hillslope concentration, river network routing, and regulation. Therefore, direct substitution introduces additional uncertainty into water budget closures (Ghiggi et al., 2019; Harrigan et al., 2020; Muñoz-Sabater et al., 2021).

Previous studies have found that using runoff from publicly available gridded products as a substitute for basin outlet discharge in closure constraints can introduce a water imbalance error of approximately 8.31% (Luo et al., 2026). Therefore, how to introduce measured runoff observations into water-balance constraints while preserving the spatial continuity of gridded products has become a critical issue. In addition, current closure studies still rely mainly on information contained within the water-cycle variables themselves. Such optimization strategy remains within an internal adjustment framework among the original components, making it difficult to acquire additional information from external environmental variables to achieve a more accurate closure optimization.

To address these limitations, we constructed a global-gridded terrestrial water budget closure dataset constrained by GRDC observations. Using the ISEA3H10 equal-area hexagonal discrete global grid as a unified spatial framework, this study integrated multi-source water-cycle variables and external environmental information such as topography. The GRDC basin runoff observations were used to provide supervised constraints on gridded runoff estimates, thereby alleviating the excessive dependence of traditional gridded closure studies on land surface-model runoff proxies. The resulting dataset reduced the residual water balance and improved the physical consistency and spatial representation.

2 Methods

To address these problems of existing global terrestrial water balance products, this study combined GRDC runoff observations, multi-source water-cycle variables, and multi-source background variables within a global discrete grid framework. A physically constrained deep-learning algorithm was used to construct a global terrestrial water budget closure dataset. The overall technical workflow included four major stages: multi-source water cycle data input, data preprocessing and harmonization, water budget closure model training based on physical constraints, and data reliability assessment (Fig. 1).

https://doi.org/10.5194/essd-2026-497
Preprint. Discussion started: 22 September 2026

This study first integrated multi-source water-cycle variables, including four core variables: precipitation, evapotranspiration, runoff proxies, and terrestrial water storage change. Auxiliary variables, including the normalized difference vegetation index (NDVI), land-surface temperature (LST), snow water equivalent, soil moisture, digital elevation model (DEM), and land cover, were also introduced to characterize vegetation, thermal conditions, terrain, and land-surface conditions. This design allows the closure model to use external environmental information to give more details. During pre-processing, different data sources were first converted to a unified monthly water depth unit (mm/month). These gridded variables were then mapped to the ISEA3H10 discrete global hexagonal grid using area-weighted resampling. GRDC basin boundaries were used to establish basin–grid correspondence, allowing measured runoff observations to serve as gridded-scale constraints for water-budget closure optimization. The training and validation samples were divided by time for model training and evaluation of extrapolation capability. For model construction, a HexCNN-based water budget closure model was adopted. Model training used a multi-objective loss function such that the model approached GRDC runoff observations and satisfied closure while preserving the physical consistency and original characteristics of each water-cycle component as much as possible. The optimal scheme was selected through evaluation, and a global monthly closed water cycle dataset for 2000–2020 was generated. Based on this dataset, global water balance residual diagnosis and validation using typical extreme hydrological events were conducted.

https://doi.org/10.5194/essd-2026-497
Preprint. Discussion started: 22 September 2026

Open Access Earth System Science Data Discussions

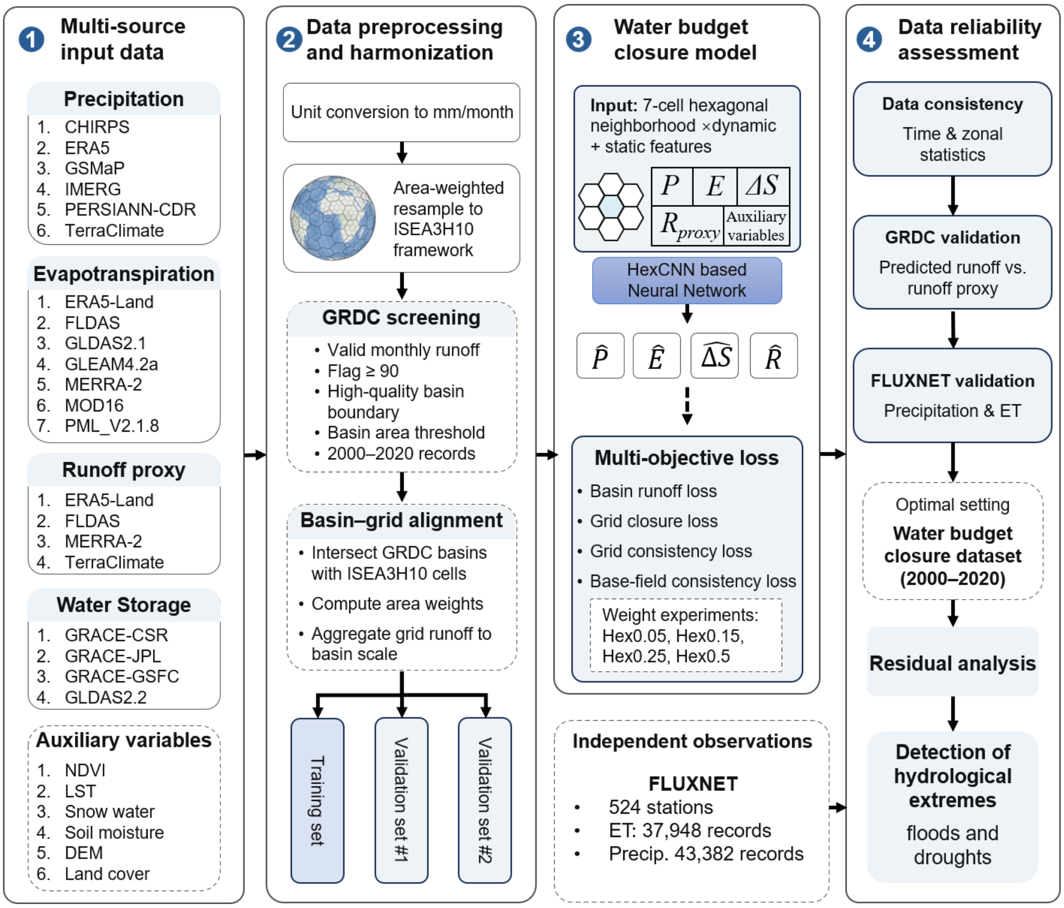


Figure 1: Research workflow.

2.1 Data collection

The data used in this study comprised three main categories: water-cycle closure constraint data, background variable data, and site observations. Water-cycle closure constraint data included multiple precipitation, evapotranspiration, runoff proxy, and water storage change products (Table 1). The precipitation datasets used included CHIRPS, ERA5, GSMaP, IMERG, PERSIANN-CDR, and TerraClimate. Evapotranspiration data are selected from ERA5-Land, FLDAS, GLDAS2.1, GLEAM4.2a, MERRA-2, MOD16, and PML_V2.2a. These products cover reanalysis data, land-surface model simulations, and remote-sensing retrievals. Runoff proxy data were mainly obtained from ERA5-Land, FLDAS, MERRA-2, and TerraClimate, including surface runoff, subsurface runoff, and total runoff variables used to characterize basin runoff generation and land surface water output. Water storage change data included CSR, JPL, and GSFC GRACE/GRACE-FO mascon products and the GLDAS2.2 terrestrial water storage product. Most datasets are publicly available from the Google Earth Engine (GEE), except for GLEAM evapotranspiration (Miralles et al., 2025a, 2025b), GSFC terrestrial water storage

https://doi.org/10.5194/essd-2026-497
Preprint. Discussion started: 22 September 2026

anomaly data, and CSR terrestrial water storage anomaly data (Loomis et al., 2019; Save, 2020; Save et al., 2016). To facilitate unified processing, these three datasets were uploaded also, and the access link are provided in Table 1.

In total, the water-cycle closure constraint data comprised precipitation products from six sources, evapotranspiration products from seven sources, runoff proxy products from four sources, and terrestrial water storage change products from four sources. Direct combinations of these existing water-cycle variables generally contained large water-balance residuals. After unit conversion and monthly-scale transformation, 6 × 7 × 4 × 4 = 672 *P-E-R-ΔS* combinations were generated. For ERA5 and FLDAS, total runoff was calculated as the sum of surface runoff and subsurface runoff. The statistics showed that the root mean square error (RMSE) values of the residuals were mainly concentrated between 50 and 70 mm/month, with an overall median of 56.87 mm/month (Fig. 2). The combination with the smallest RMSE consisted of IMERG precipitation, FLDAS evapotranspiration and runoff, and GLDAS water storage change. Its RMSE was 43.37 mm/month, and the mean land residual was 11.09 mm/month. Even the optimal-performing original data source combination in terms of closure still had a high residual monthly water balance. Original products showed widespread closure inconsistencies, and simply selecting different combinations of original data sources remained insufficient for obtaining low-closure residuals.

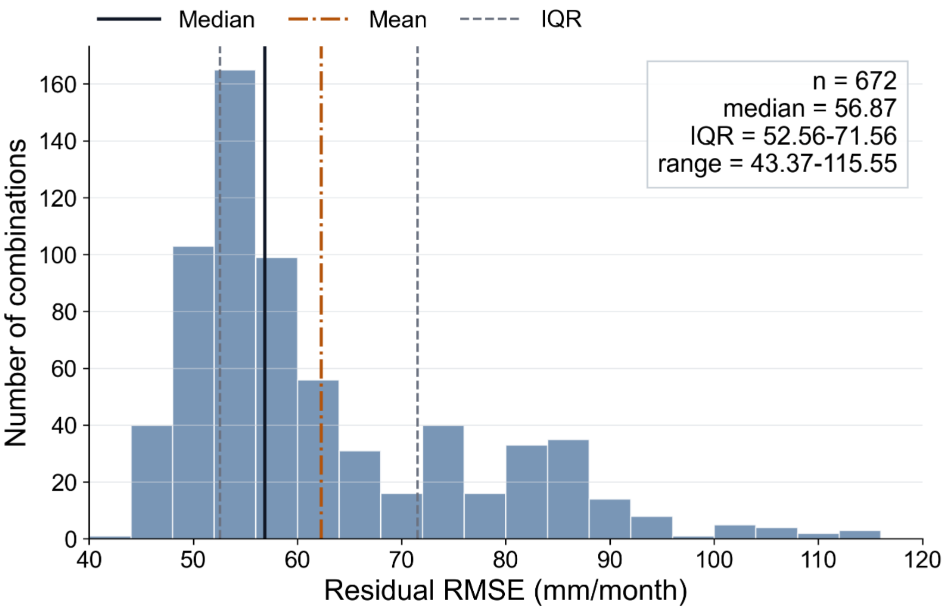


Figure 2: Water balance residual RMSE of the 672 multi-source input combinations.

For water-cycle variables, this study introduced multi-source background variables, including dynamic and static environmental variables (Table 2). Dynamic environmental variables refer to monthly changes in environmental variables, including snow water equivalent and volumetric soil water content in layers 1–4 from ERA5-Land. These variables characterize the spatiotemporal variations in snow and soil moisture conditions. The 1-km NDVI product from MOD13A2 was used to reflect vegetation cover and growth conditions. The 1-km LST product from MOD11A2 characterizes land-surface thermal conditions and their seasonal variations. The static environmental variables included the Copernicus DEM GLO-30

https://doi.org/10.5194/essd-2026-497
Preprint. Discussion started: 22 September 2026

and GlobCover. The DEM describes terrain relief and elevation differences, whereas GlobCover land-cover data distinguish different surface types and their spatial distributions. Introducing background variables helps overcome the limitation of optimizing data only from statistical relationships among water-cycle variables and provides the model with external information reflecting land-surface conditions and environmental regulation.

The site observation data included the GRDC-observed runoff data (http://grdc.bafg.de/) and FLUXNET flux observations (Pastorello et al., 2020). The GRDC provides river discharge records from global river stations. Introducing GRDC constraints can reduce the errors caused by runoff proxy flow data (Luo et al., 2026). In our study, GRDC data are used for supervised model training and validation. FLUXNET data are used only as independent site observations for model-result validation and are not involved in model training. To ensure dimensional consistency, FLUXNET precipitation and latent heat flux variables are converted proportionally to monthly total precipitation and monthly evapotranspiration, respectively, both in mm/month (Chen et al., 2023).

https://doi.org/10.5194/essd-2026-497
Preprint. Discussion started: 22 September 2026

Open Access Earth System Science Data Discussions

Table 1: Data sources for water-cycle closure constraints.

Water-cycle variable	Data product	Temporal coverage	Spatial coverage	Variable name	GEE address
Precipitation	CHIRPS	1981–present, daily	Land between 60°S and 60°N, 0.05°	precipitation (mm/day)	UCSB-CHC/CHIRPS/V3/DAILY_SAT
	ERA5	1979–present, monthly	Global, 0.25°	total_precipitation (m/month)	ECMWF/ERA5/MONTHLY
	GSMaP	1998–present, hourly	60°S–60°N, 0.1°	hourlyPrecipRateGC (mm/hour)	JAXA/GPM_L3/GSMaP/v8/operational
	IMERG	1998–present, monthly	Global, 0.1°	precipitation (mm/hour)	NASA/GPM_L3/IMERG_MONTHLY_V07
	PERSIANN-CDR	1983–present, daily	60°S–60°S–60°N, 0.25°	precipitation (mm/day)	NOAA/PERSIANN-CDR
	TerraClimate	1958–present, monthly	Global land, 4 km	pr (mm/month)	IDAHO_EPSCOR/TERRACLIMATE
Evapotranspiration	ERA5_Land	1950–present, monthly	Global land, 0.1°	total_evaporation_sum (m/month, scale factor = −1)	ECMWF/ERA5_LAND/MONTHLY_AGGR
	FLDAS	1982–present, monthly	Global land excluding Antarctica, 0.1°	Evap_tavg (mm/second)	NASA/FLDAS/NOAH01/C/GL/M/V001

https://doi.org/10.5194/essd-2026-497
Preprint. Discussion started: 22 September 2026

Water-cycle variable	Data product	Temporal coverage	Spatial coverage	Variable name	GEE address
	GLDAS2.1	2000–present, 3-hourly	Global land excluding Antarctica, 0.25°	Evap_tavg (mm/second)	NASA/GLDAS/V021/NOAH/G025/T3H
	GLEAM4.2a	1980–present, monthly	Global land, 0.1°	E (mm/month)	projects/ee-carylee/assets/GLEAM_v4_2a
	MERRA-2	1980–present, hourly	Global land excluding Greenland and Antarctica, 0.5° × 0.625°	EVLAND (mm/second)	NASA/GSFC/MERRA/lnd/2
	MOD16	2000–present, 8-day	Most global land areas, 500 m	ET (mm/8days, scale factor = 0.1)	MODIS/061/MOD16A2GF
	PML_V2.2a	2000–2023 monthly	Global land, 500 m	ET (mm/day, scale factor = 0.01)	projects/pml_evapotranspiration/PML/OUTPUT/PML_V22a
Runoff proxy	ERA5_Land	1950–present, monthly	Global land, 0.1°	surface_runoff_sum (m/month) sub_surface_runoff_sum (m/month)	ECMWF/ERA5_LAND/MONTHLY_AGGR
	FLDAS	1982–present, monthly	Global land excluding Antarctica, 0.1°	Qs_tavg (mm/second) Qsb_tavg (mm/second)	NASA/FLDAS/NOAH01/C/GL/M/V001

https://doi.org/10.5194/essd-2026-497
Preprint. Discussion started: 22 September 2026

Water-cycle variable	Data product	Temporal coverage	Spatial coverage	Variable name	GEE address
	MERRA-2	1980–present, hourly	Global land excluding Greenland and Antarctica, 0.5° × 0.625°	RUNOFF (mm/second)	NASA/GSFC/MERRA/lnd/2
	TerraClimate	1958–present, monthly	Global land, 4 km	ro (mm/month)	IDAHO_EPSCOR/TERRACLIMATE
Water storage	CSR GRACE/GRACE-FO	2002–present, monthly	Global, 0.25°	b1 (cm/month)	projects/ee-linken173230/assets/CSR_GRACE
	JPL GRACE/GRACE-FO	2002–present, monthly	Global, 0.5°	lwe_thickness (cm/month)	NASA/GRACE/MASS_GRIDS_V04/MASCON_CRI
	GSFC GRACE/GRACE-FO	2002–present, monthly	Global, 0.5°	b1 (cm/month)	projects/ee-linken173230/assets/GSFC_GRACE
	GLDAS2.2	2003–present, daily	Global land excluding Antarctica, 0.25°	TWS_tavg (mm/day)	NASA/GLDAS/V022/CLSM/G025/DA1D

Table 2: Data sources of background variables.

https://doi.org/10.5194/essd-2026-497
Preprint. Discussion started: 22 September 2026

Background variable type	Data product	Temporal coverage	Spatial coverage	Variable name	GEE address
Dynamic variables	ERA5_Land	1950–present, monthly	Global land, 0.1°	snow_depth_water_equivalent (m/month) volumetric_soil_water_layer_1 (%) volumetric_soil_water_layer_2 (%) volumetric_soil_water_layer_3 (%) volumetric_soil_water_layer_4 (%)	ECMWF/ERA5_LAND/MONTHLY_AGGR
	MOD13A2	2000–present, 16-day	Global land excluding Antarctica, 1 km	NDVI	MODIS/061/MOD13A2
	MOD11A2	2000–present, 8-day	Global land, 1 km	LST_Day_1km	MODIS/061/MOD11A2
Static variables	Copernicus DEM GLO-30	2010–2015	Global land, 30 m	DEM	COPERNICUS/DEM/GLO30
	GlobCover	2009	Global excluding Antarctica, 300 m	landcover	ESA/GLOBCOVER_L4_200901_200912_V2_3

https://doi.org/10.5194/essd-2026-497
Preprint. Discussion started: 22 September 2026

Open Access Earth System Science Data Discussions

2.2 Data screening and alignment

Because these multi-source water-cycle closure constraint datasets differ in temporal coverage, spatial extent, and spatiotemporal resolution, preprocessing was required to harmonize them for subsequent analysis. First, all the water-cycle variables were converted to a unified unit of mm/month. Each water-cycle product was initially averaged monthly. According to the original unit of each water-cycle data, the values were multiplied by the corresponding number of days, hours, or seconds, and converted to millimetres. The units of kg/m^2 were treated as equivalent to 1 mm of water depth. The negative sign of ERA5-Land evapotranspiration and the 0.1 scale factor of MOD16 evapotranspiration were corrected before subsequent calculations. The CSR, JPL, and GSFC GRACE/GRACE-FO mascon products represented terrestrial water storage anomalies, whereas GLDAS2.2 represented instantaneous or daily terrestrial water storage states. To obtain terrestrial water storage change data, we calculated the difference between adjacent monthly terrestrial water storage values (Feng et al., 2025; Li et al., 2022).

Second, spatial grid mapping was performed to establish a unified spatial framework. This study adopted the Icosahedral Snyder Equal Area (ISEA) aperture 3 hexagonal (3H) discrete global grid system (DGGS) as the unified data grid framework. Compared with equal longitude–latitude grids, each grid cell in this system is hexagonal, except for 12 pentagons. The cells have equal areas, and the distances to the six neighboring cells are equal. This grid effectively avoids systematic biases in the retrieved variables caused by area distortions across latitudes (Li and Wang, 2026a). Within this framework, the aforementioned water-cycle and background variables from different sources were resampled. Quantitative data products were resampled using area-weighted zonal statistics, whereas categorical products such as land cover were resampled by taking the modal category within each hexagonal cell.

In addition, necessary screening was performed on the site observation data. Monthly runoff records from 11,266 GRDC basins were initially obtained. Among them, 2,286 basins were removed owing to missing data. The remaining 8,980 basins contained 5,063,115 monthly runoff records, with each record representing runoff observations for a basin in a given month. To improve the data accuracy, quality control screening was applied to these runoff records. The final reliable dataset contained 710,274 monthly runoff records covering 3,471 basins. The screening criteria were as follows.

(1) Screening of monthly runoff record validity. The calculated field in the GRDC data was used as the monthly flow input. This field represents the monthly mean flow calculated from the daily flow data (m^3/s). Records with missing calculated values or values < 0 were removed.

(2) Screening monthly flow quality flags. The Flag field in the original GRDC file is the percentage of valid daily flow records used to calculate the monthly mean flow. This study required Flag > = 90; monthly runoff records below this threshold were removed.

(3) Screening of basin boundary quality. Based on the GRDC basin boundary quality information, only records from basins with the boundary quality field marked as High were retained.

https://doi.org/10.5194/essd-2026-497
Preprint. Discussion started: 22 September 2026

(4) Basin area screening. To reduce the area errors caused by overlaying small basins with ISEA3H10 hexagonal cells, only basins with an effective overlay area not smaller than the average area of one ISEA3H10 cell were retained. The current average grid area threshold is approximately 863.8 km^2 (Kevin, 2018).

(5) Screening of model input availability. Only the monthly runoff records from January, 2000 to December, 2020 were retained. Each basin must contain at least one valid ISEA3H10 hexagonal cell in the basin-grid mapping table, and this cell and its six neighboring cells must have corresponding dynamic, static, and neighborhood indices in the model input cache.

To align the basin-scale GRDC runoff observations with the model outputs at the ISEA3H10 hexagonal grid scale, we used an area-weighted mapping method (Michailovsky et al., 2023). First, the spatial overlay area between each GRDC basin and all ISEA3H10 cells within its coverage area was calculated. For a given basin, if the overlay area between the basin and the ith grid cell was Ai, and the total overlay area between the basin and all relevant grid cells was $\sum_i A_i$, then the area weight of the i-th grid cell within the basin was defined as follows:

$$w_i = \frac{A_i}{\sum_i A_i}, \tag{1}$$

where w_i represents the contribution weight of the i-th grid cell to the basin-scale runoff aggregation. Based on this weight, the gridded model output runoff depth was area-weighted and aggregated at the basin scale as follows:

$$\hat{R}_{basin} = \sum_i w_i \hat{R}_i, \tag{2}$$

where $\grave{R}_i$ is the predicted runoff depth of the i-th grid cell and $\grave{R}_{basin}$ is the area-weighted basin-scale predicted runoff depth. Through this area-weighted mapping, the gridded simulation results and the GRDC basin-scale observed runoff were spatially consistent, allowing the basin-scale supervised loss to be calculated.

All the valid sites in the FLUXNET site list were used as independent validation samples. A total of 667 FLUXNET sites were obtained. After screening, 524 sites remained, including 37,948 monthly evapotranspiration records and 43,382 monthly precipitation records. The screening criteria were as follows.

(1) Temporal screening. The study period was from January, 2000 to December, 2020. Observations outside this period were not included in the subsequent matching and statistical analyses.

(2) Spatial screening. A sample was included only when valid FLUXNET data and the corresponding gridded product data were simultaneously available for the same site and month.

(3) Plausibility screening. Missing values were removed. A maximum precipitation threshold of 1000 mm/month was set. Data above this threshold were excluded from the validation, and 154 records were removed.

2.3 Data architecture and training

In this study, a deep-learning model was used to impose water budget closure constraints (Fig. 3). Because ISEA3H10 was used as the training framework, each central grid cell and its six adjacent cells were combined into a seven-cell neighborhood as the basic input unit, so that spatial information could be learned from the neighboring grid cells. For each grid cell, the input features included time-varying dynamic variables, which are the dynamic variables listed in Tables 1 and 2, and static variables,

https://doi.org/10.5194/essd-2026-497
Preprint. Discussion started: 22 September 2026

which are the land-cover and DEM variables listed in Table 2. The dynamic inputs contained 30 data sources, including six precipitation products, seven evapotranspiration products, four water-storage-change products, six runoff proxies, and seven background products consisting of land-surface temperature, NDVI, snow, and four soil-moisture layers (Table 1 & 2). Each data source included one numerical variable channel and one variable mask channel, resulting in 60 dynamic channels. The mask value of 1 was the data source had valid observations for the corresponding grid cell and month, and a mask value of 0 was the value was missing or invalid. With the DEM and land cover included as two static features, the total number of input dimensions for a single grid cell were 62. Because each central grid cell was combined with its six first-order neighboring cells, the basic model input consisted of a seven-cell neighborhood, giving an input tensor of $X \in \mathbb{R}^{7\times62}$, and the input features were divided into five branches: precipitation ($X_1 \in \mathbb{R}^{7\times12}$), evapotranspiration ($X_2 \in \mathbb{R}^{7\times14}$), water storage change ($X_3 \in \mathbb{R}^{7\times8}$), runoff proxies ($X_4 \in \mathbb{R}^{7\times12}$), and background variables ($X_5 \in \mathbb{R}^{7\times16}$).

Each branch first passed through an independent HexEncoder to extract local spatial features from the central grid cell and its six neighbors. Within the HexEncoder, the input features were first mapped to a 64-dimensional hidden space through a linear layer. A hexagonal neighborhood convolution module (HexCNN) was then used for local information transmission (Li and Wang, 2026a). The central-grid, neighborhood mean, and neighborhood maximum features, and the difference between the central-grid and neighborhood mean features, were concatenated, and a readout layer was used to obtain a 64-dimensional branch representation. The resulting branch features for precipitation, evapotranspiration, water storage change, runoff proxies, and background variables were denoted as h_P, hE, h_S, h_R, and h_{BG}, respectively.

To account for the regulatory effect of background environmental conditions on different hydrological processes, the model used the background branch features to gate and modulate the other four hydrological branches, yielding h'_P, $h'E$, h'_S, and h'_R. The four modulated hydrological branch features and background feature h_{BG} were then concatenated to form a 320-dimensional feature vector. This vector was input into the Feature Fusion module, where a linear transformation compressed the features into a 64-dimensional fused feature, h_{fused}. Four linear prediction heads were used to output the correction terms for precipitation, evapotranspiration, water storage change, and runoff, denoted by δP, δE, δS, and δR, respectively. The model adopted residual learning, and it did not predict water balance components from zero but instead learned correction terms based on background fields obtained through multi-source data fusion. The final estimates $\hat{P}$, $\hat{E}$, $\widehat{\Delta S}$, and $\hat{R}$ were then recovered.

https://doi.org/10.5194/essd-2026-497
Preprint. Discussion started: 22 September 2026

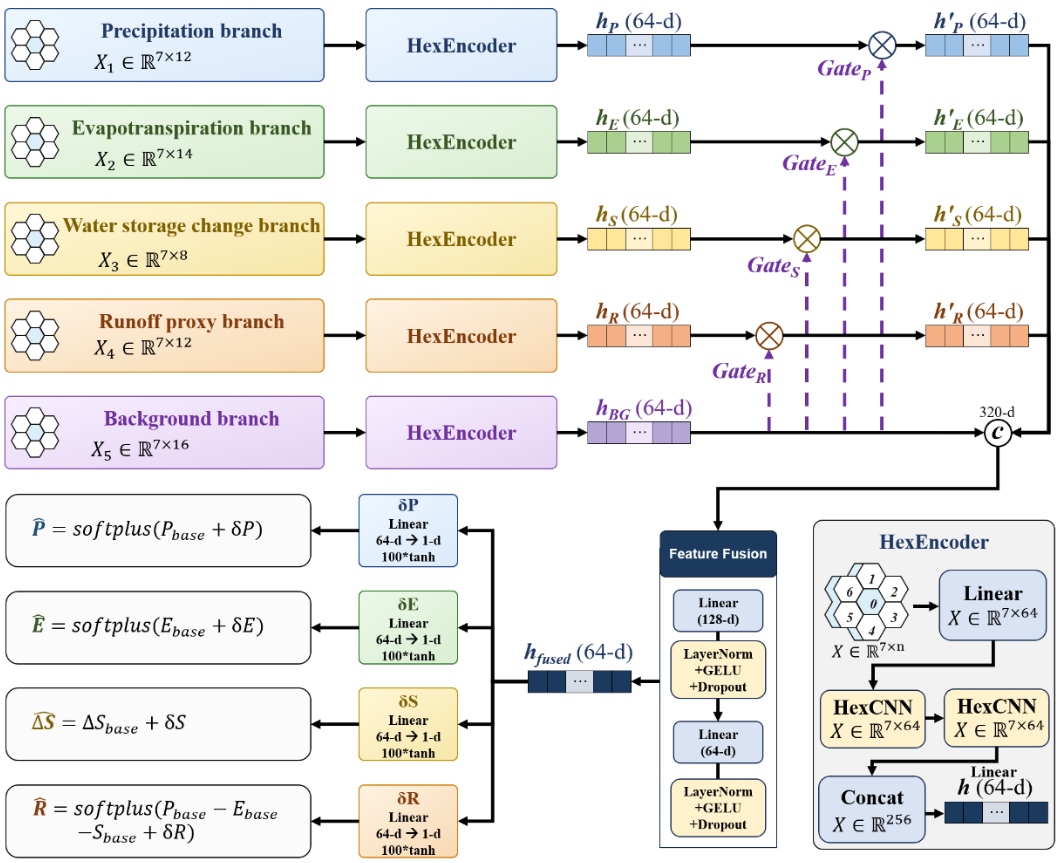


Figure 3: Model architecture.

The model outputs $\hat{P}$, $\hat{E}$, $\widehat{\Delta S}$, and $\hat{R}$, which were then combined with the corresponding input water-cycle variables, background fields, and basin-scale runoff observations to calculate the loss for backpropagation and model optimization. Four loss function types were defined in this study: basin runoff loss (L_{basin}), grid-scale closure loss ($L_{closure}$), grid-scale observation consistency loss (L_{grid}), and median consistency loss (L_{base}). The total loss was obtained by a weighted allocation among these four loss functions (Eq. (3)). The basin runoff loss compared the area-weighted runoff predictions with the GRDC observations (Eq. (4)). The grid-scale closure loss penalized the absolute water-budget residual at locations where all three median background fields were valid (Eq. (5)). The grid-source consistency loss constrained the corrected variables to remain close to the contributing gridded products (Eq. (6)). The median background-field preservation loss constrained the corrected precipitation, evapotranspiration, and water-storage-change estimates to remain close to their corresponding multi-source median background fields (Eq. (7)).

$$L = \lambda_1 L_{basin} + \lambda_2 L_{closure} + \lambda_3 L_{grid} + \lambda_4 L_{base} \tag{3}$$

$$L_{basin} = MAE\left(\hat{R}, R_{GRDC}\right) = MAE\left(\frac{\sum_i a_i \hat{R}_i}{\sum_i a_i}, R\right) \tag{4}$$

$$L_{closure} = MAE(\hat{P} - \hat{E} - \widehat{\Delta S} - \hat{R}, 0) \tag{5}$$

https://doi.org/10.5194/essd-2026-497
Preprint. Discussion started: 22 September 2026

$$L_{\text{grid}} = \frac{1}{3}\left(MAE(\hat{P}, P_{\text{obs}}) + MAE(\hat{E}, E_{\text{obs}}) + MAE(\widehat{\Delta S}, \Delta S_{\text{obs}})\right) \tag{6}$$

$$L_{\text{base}} = \frac{1}{3}\left(MAE(\hat{P}, P_{\text{base}}) + MAE(\hat{E}, E_{\text{base}}) + MAE(\widehat{\Delta S}, \Delta S_{\text{base}})\right) \tag{7}$$

$$MAE(\hat{P}, P_{\text{obs}}) = \frac{1}{K_P}\sum_{k=1}^{K_P}\frac{\sum_{i=1}^{N} m_{i,k}^{P}|\hat{P}_i - P_{\text{obs},i,k}|}{\sum_{i=1}^{N} m_{i,k}^{P}} \tag{8}$$

$$MAE(\hat{E}, E_{\text{obs}}) = \frac{1}{K_E}\sum_{k=1}^{K_E}\frac{\sum_{i=1}^{N} m_{i,k}^{E}|\hat{E}_i - E_{\text{obs},i,k}|}{\sum_{i=1}^{N} m_{i,k}^{E}} \tag{9}$$

$$MAE(\widehat{\Delta S}, \Delta S_{\text{obs}}) = \frac{1}{K_{\Delta S}}\sum_{k=1}^{K_{\Delta S}}\frac{\sum_{i=1}^{N} m_{i,k}^{\Delta S}|\widehat{\Delta S}_i - \Delta S_{\text{obs},i,k}|}{\sum_{i=1}^{N} m_{i,k}^{\Delta S}} \tag{10}$$

Where, P_{obs}, E_{obs}, ΔS_{obs} denote the multi-source input products for precipitation, evapotranspiration, and water-storage change, respectively. K_P, K_E, $K_{\Delta S}$ denote the numbers of valid data sources for the corresponding variables, and $m_{i,k}^{P}$ denotes the data-availability mask. P_{obs}, E_{obs}, ΔS_{obs} represent the median values of the corresponding multi-source input products for the same grid cell and month and were used as the background fields for residual learning. Because the grid-scale runoff proxies and observed runoff at basin outlets are not fully equivalent in either spatial scale or physical meaning, no grid-source consistency loss or median background-field preservation loss was directly imposed on the runoff component. Instead, runoff was primarily optimized through basin-scale supervision using GRDC observations and the grid-scale water-budget closure constraint.

Four weighting schemes were designed to identify the optimal loss function configuration: Scheme 1: λ_1=1， λ_2=0.5， λ_3=0.1， λ_4=0.05 (Hex0.05); Scheme 2: λ_1=1， λ_2=0.5， λ_3=0.3， λ_4=0.15 (Hex0.15); Scheme 3: λ_1=1， λ_2=0.5， λ_3=0.5， λ_4=0.25 (Hex0.25); and Scheme 4: λ_1=1， λ_2=0.5， λ_3=1， λ_4=0.5 (Hex0.5). Basin runoff loss was assigned the highest weight of 1 because supervised samples were available. Grid closure loss was the objective of this study, however strict closure can cause water-cycle variables to deviate from their physically plausible values; therefore, its weight was set to 0.5. From Schemes 1–4, the weights of the grid consistency and median consistency gradually increased, allowing examination of the model performance under different settings.

For each weighting scheme, the model was trained for 200 epochs using the Adam optimizer with a fixed learning rate of $5*10^{-4}$. During each training iteration, the model generated $\hat{P}$, $\hat{E}$, $\widehat{\Delta S}$, and $\hat{R}$, and the corresponding loss function were calculated according to Eq. (3). The weighted total loss was then backpropagated to update the model parameters. The training batches were randomly shuffled at the beginning of each epoch. After each epoch, the mean training loss was recorded and a model checkpoint was saved. The same training settings were applied to all four weighting schemes to ensure a consistent comparison of their performance.

2.4 Validation

This study used a validation set and independent site observations for validation and further examined the event-mining capability of the data product using typical hydrological events. Data from January, 2000 to December, 2015 were used as the

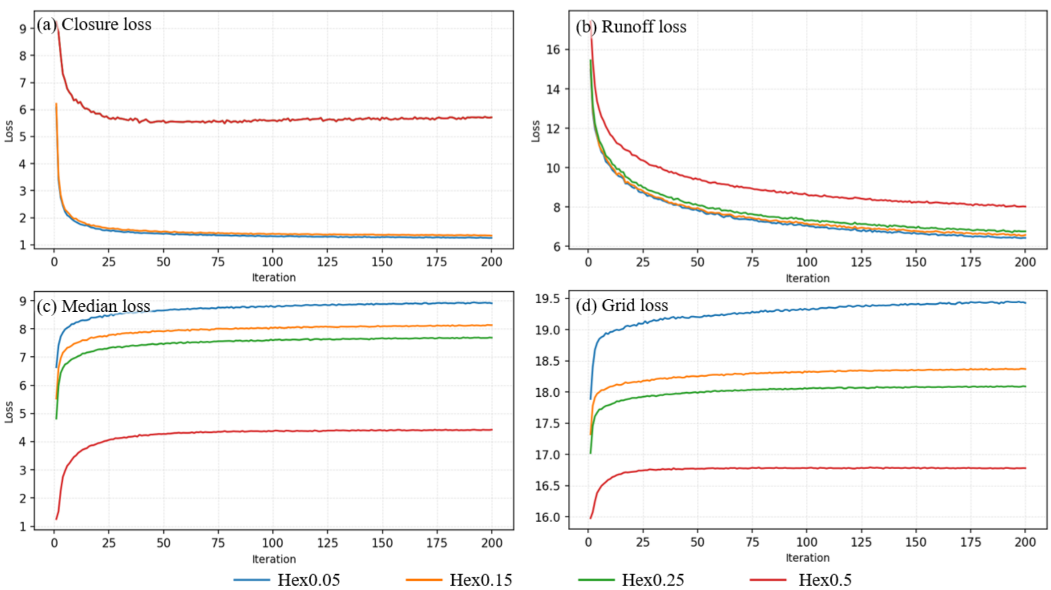
https://doi.org/10.5194/essd-2026-497
Preprint. Discussion started: 22 September 2026

training set, and data from January, 2016 to December, 2020 were used as the validation set. The training set contained 554,747 monthly runoff records covering 3,452 basins, and the validation set contained 155,527 monthly runoff records covering 3,116 basins. Independent validation sites were selected from the FLUXNET data, and the precipitation and evapotranspiration fluxes were statistically compared with the predicted and multi-source data products.

3 Results and discussion

3.1 Model training results

Model training results revealed a trade-off among water-budget closure, agreement with GRDC runoff observations, and preservation of the original multi-source data characteristics. This study involved four parallel experiments with different weight configurations, and the iterative changes in the closure loss, basin runoff loss, median consistency loss, and grid consistency loss were recorded (Fig. 4). When the weights of the median and grid consistency losses were reduced, the closure and basin runoff losses decreased more rapidly, whereas the median and grid consistency losses increased accordingly. This indicated that to further approach the GRDC basin runoff observations and satisfy the water budget closure constraints, the model must modify the original water-cycle variables to a certain extent. The iteration curves showed that Hex0.05 had the most rapid decrease in closure loss and basin runoff loss but performed less well in preserving median consistency and grid consistency. At the 200th iteration, for example, the median consistency loss and grid consistency loss of Hex0.05 were 8.90 and 19.43, respectively, both of which were relatively high among the four experiments. Although this scheme can more strongly optimize the closure and runoff objectives, it also produces the largest correction magnitude for the original water-cycle variables. Therefore, additional consistency checks are required.



https://doi.org/10.5194/essd-2026-497
Preprint. Discussion started: 22 September 2026

Figure 4: Iteration curves of model training for iterations 1–200: (a) closure loss, (b) runoff loss, (c) median loss, and (d) grid loss. The blue, orange, green, and red solid lines denote the four experimental settings, respectively.

The extrapolation capability of the model was evaluated by using an independent validation set (Table 3). Although the model performance differed slightly among the experimental settings, none of the schemes showed obvious overfitting, and the performance of the validation set was superior to that of the training set in some experiments. The model has a strong generalization ability and can be stably applied to data from different periods. Overall, Hex0.05, and Hex0.15 exhibited the optimal performances and were the most stable, achieving the highest correlation coefficients in the training and validation sets from different periods (i.e., all > 0.89). For the RMSE, Hex0.05 exhibited a slightly higher performance than Hex0.15. However, the difference between them was extremely small. Hex0.25 and Hex0.5 had higher RMSE values than the first two schemes, and their correlation coefficients were also slightly lower.

Table 3: Fitting performance between runoff estimates and GRDC observed runoff under different datasets.

	RMSE (mm/month)			Pearson r		
Scheme	Train set (2000-2015)	Val set (2016-2018)	Val set (2019-2020)	Train set (2000-2015)	Val set (2016-2018)	Val set (2019-2020)
Hex0.05	24.54	22.23	23.31	0.893	0.895	0.899
Hex0.15	24.65	22.26	23.27	0.891	0.894	0.899
Hex0.25	25.18	22.97	23.85	0.888	0.888	0.897
Hex0.5	26.95	25.24	25.63	0.870	0.863	0.880

3.2 Data consistency comparison

To evaluate whether the closure optimization preserved the basic hydroclimatic characteristics of the original datasets, the optimized products were compared with the multi-source inputs from both monthly and land-cover zonal perspectives.

3.2.1 Monthly data consistency comparison

The global monthly mean values from 2000–2020 showed that all water balance components had consistent seasonal structures. However, marked differences existed among the data sources (Fig. 5). Precipitation and evapotranspiration generally peaked in July for most of the original products, whereas runoff peaked earlier, in May–June. The change in the terrestrial water storage was positive in winter and negative in summer. The differences among the data sources were most pronounced for evapotranspiration, ranging from 37.97 mm for GLEAM to 96.00 mm for PML. In comparison, the annual mean precipitation ranged from 57.82–81.07 mm, and runoff ranged from 16.32–26.68 mm.

Compared with the observational and reanalysis products, the optimized products were more consistent with each other. The annual mean evapotranspiration ranged from 43.06–44.50 mm, similar to that of ERA5 (44.14 mm). The annual mean precipitation ranged from 61.01–63.30 mm, similar to that of GSMaP and TerraClimate. Runoff showed even stronger consistency, with estimates ranging from 21.92–22.73 mm, which were within the range of the original products and closest

https://doi.org/10.5194/essd-2026-497
Preprint. Discussion started: 22 September 2026

to that of FLDAS (21.91 mm). The terrestrial water storage changes differed among the experimental settings. Hex0.05 showed an obvious negative bias, with smaller monthly mean values than the minimum monthly mean of the multi-source input data for every month (Fig. 5d). This implied that under lower grid and median consistency weights, the model tended to achieve runoff simulation and physical closure of the water cycle by strongly modifying the water storage change. As the consistency constraint weights increased, the negative biases of Hex0.15, Hex0.25, and Hex0.5 were substantially reduced, and their consistency with the multi-source input data was enhanced accordingly.

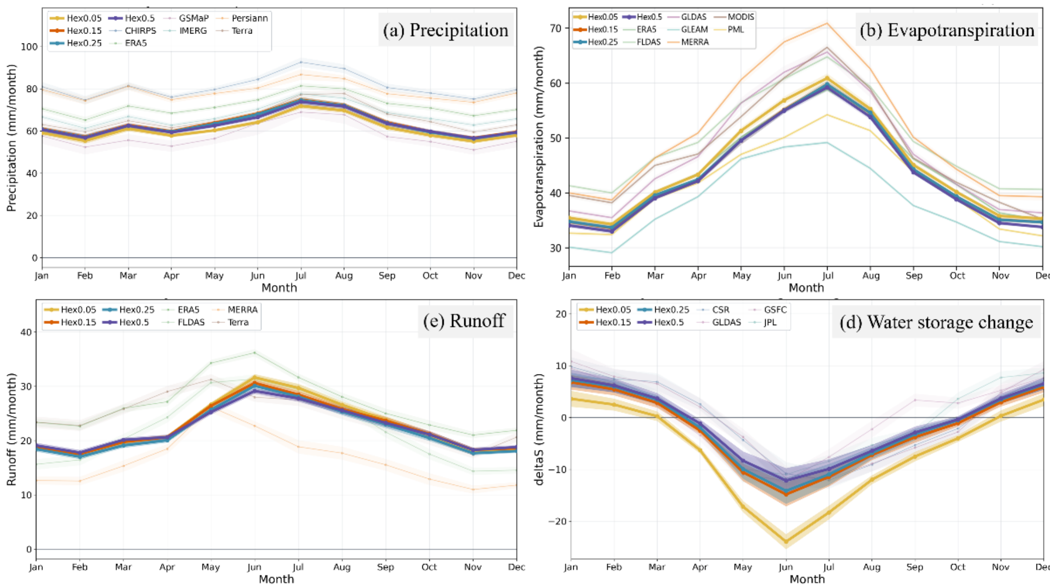


Figure 5: Seasonal variations in monthly mean values of multi-source inputs and optimized water-cycle variables: (a) precipitation, (b) evapotranspiration, (c) runoff, and (d) water storage change.

3.2.2 Zonal data consistency by land cover

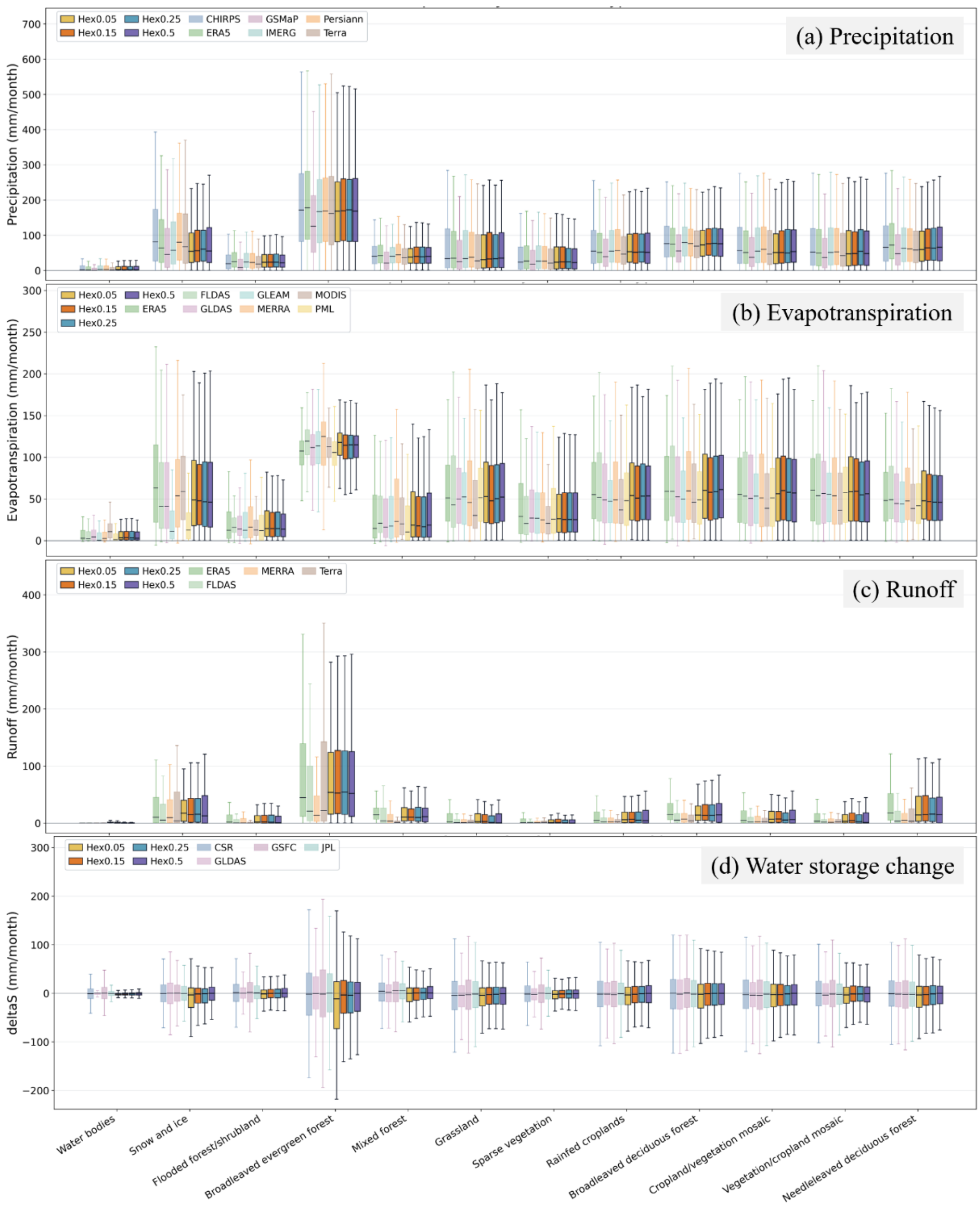
From the perspective of land-cover zones, the optimized data products reasonably reproduced the spatial differentiation of the four water-balance components (Fig. 6). For precipitation, the optimized products generally preserved the land-cover ranking shown by the reference products. Water bodies and flooded forests/shrubland were at the low-value end, and evergreen broadleaf forests were at the high-value end. The estimates generally fell within the range of the multi-source input data. For evapotranspiration, the optimized products also captured land-cover differences, with evergreen broadleaf forests having the highest evapotranspiration and water bodies or flooded forests/shrublands having relatively low values. Across all schemes, the values for the 12 main land-cover classes fell within the ranges of the original products, indicating high zonal consistency. Runoff showed strong consistency across land-cover zones. The optimized products preserved the basic pattern of low runoff in water bodies and sparse vegetation and relatively high runoff in forest types, with evergreen broadleaf forests still located at the high-value end.

https://doi.org/10.5194/essd-2026-497
Preprint. Discussion started: 22 September 2026

In contrast, changes in terrestrial water storage remained the component with the greatest differences. The reference products showed storage changes close to zero for most land-cover types. Whereas, Hex0.05 provided negative values for all land-cover types, and all 12 classes fell outside the range of the multi-source input products (Fig. 6d). Increasing the consistency weight in Hex0.15 reduced the negative bias from 7.09 mm in Hex0.05 to 2.83 mm. In addition, Hex0.15 reduced the within-class variability of the original multi-source products. The average within-class standard deviation across the 12 land-cover classes was 38.39 mm/month, which was lower than the values of CSR, GLDAS, GSFC, and JPL (47.17–53.49 mm/month) as well as Hex0.05 (42.57 mm/month). Hex0.15 provided more stable terrestrial water storage change estimates within each surface type and alleviated the high dispersion found in the original products and the weak-consistency-constraint scheme.

The land-cover zonal results indicated that the optimized data products generally preserved the land-surface type differences in the multi-source input data for precipitation, evapotranspiration, and runoff. For terrestrial water storage change, Hex0.15 weakened the systematic negative bias of Hex0.05 while reducing the within-class uncertainty of the original multi-source products.

https://doi.org/10.5194/essd-2026-497
Preprint. Discussion started: 22 September 2026



370 **Figure 6: Distributions of water-cycle variables from multi-source inputs and optimized products under different land-cover types: (a) precipitation, (b) evapotranspiration, (c) runoff, and (d) water storage change.**

3.3 Effectiveness analysis of GRDC simulation

The GRDC basin-scale validation showed that the optimized runoff products substantially outperformed the original runoff proxy products in representing monthly basin runoff variations. Using the basin-scale GRDC observations, the trained runoff data and proxies were mapped to the corresponding basins and compared (Fig. 7). The optimized runoff datasets from all four schemes exhibited a substantially higher performance than the original runoff proxy products in the GRDC validation. Hex0.05 exhibited the optimal overall performance, with an RMSE of 24.15 mm/month and a correlation coefficient of 0.894. Hex0.15 ranked second, with an RMSE of 24.23 mm/month and a correlation coefficient of 0.893, and the difference between the two was small. The RMSE values of Hex0.25 and Hex0.5 were 24.79 and 26.62 mm/month, respectively, and their correlation coefficients were 0.889 and 0.871, respectively. As the consistency constraints strengthened, the agreement with the GRDC decreased observed runoff decreased slightly. However, the differences among the four experiments were small. Among the original runoff proxies, ERA5 had an RMSE of 38.23 mm/month and a correlation coefficient of 0.743. FLDAS had an RMSE of 37.53 mm/month and a correlation coefficient of 0.741. Single-source MERRA and TerraClimate runoff proxies had RMSE values of 46.43 and 47.41 mm/month, respectively. Thus, the optimized runoff data reduced the RMSE by approximately 31–49% and increased the correlation coefficient to nearly 0.9. These results demonstrate that all four optimized datasets better captured the runoff variations more accurately than the original runoff proxy products

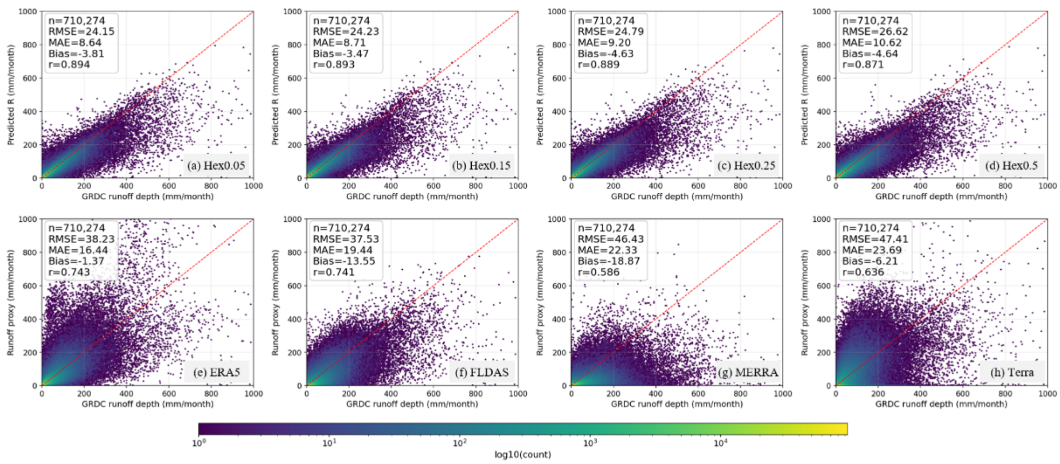


Figure 7: Comparison of multi-source input runoff and optimized runoff with GRDC runoff at the basin scale: (a) Hex0.05, (b) Hex0.15, (c) Hex0.25, (d)Hex0.5, (e) ERA5, (f) FLDAS, (g) MERRA, and (h) Terra.

3.4 Independent validation analysis

To further validate the reliability of the optimized data, this study used FLUXNET site data for independent validation of precipitation and evapotranspiration (Figs. 8 and 9). These two data types were not included in the model training. The validation of evapotranspiration showed that the optimized results outperformed some original products, and the differences between the four schemes and the original evapotranspiration products were small (Fig. 8). For the correlation coefficient, the results of the four schemes were superior than 4-6 out of 7 sets of original products. Hex0.05 and Hex0.5 had the highest value among the four parallel experiments (0.77), second only to PML (0.81), and comparable to that of ERA5 (0.77). The other two schemes had a correlation coefficient of 0.76, with only small differences. For RMSE, Hex0.5 had the lowest (26.6 mm/month), followed by Hex0.15 and Hex0.25 (27.0 mm/month). Among the original evapotranspiration products, PML had the lowest

https://doi.org/10.5194/essd-2026-497
Preprint. Discussion started: 22 September 2026

RMSE (22.7 mm/month). Overall, the four optimized products showed performance comparable to ERA5 and generally outperformed FLDAS, GLDAS, and MERRA. This suggested that the evapotranspiration data in the proposed product can maintain evapotranspiration accuracy while satisfying closure constraints.

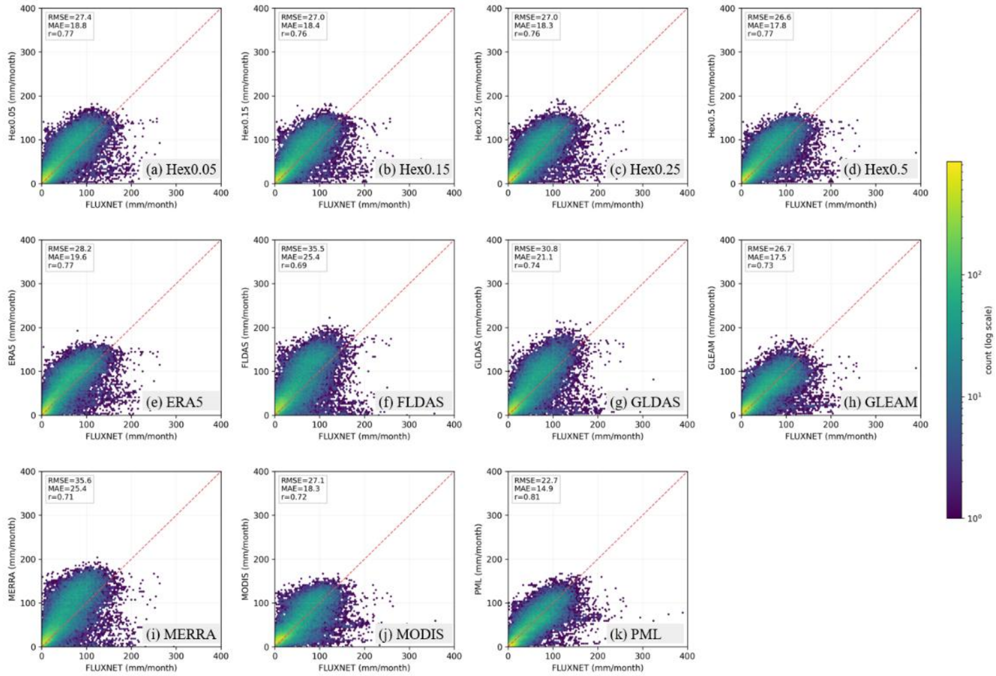


Figure 8: Comparison of multi-source input and optimized evapotranspiration data with FLUXNET evapotranspiration: (a) Hex0.05, (b) Hex0.15, (c) Hex0.25, (d)Hex0.5, (e) ERA5, (f) FLDAS, (g) GLDAS, (h) GLEAM, (i) MERRA, (j) MODIS, and (k) PML.

The precipitation validation showed no significant differences among the four products. However, all four outperformed the original precipitation input data (Fig. 9). Among the original precipitation products, ERA5 and IMERG performed relatively well, with RMSE values of 133.40 and 134.08 mm/month, mean absolute error values of 65.84 and 64.66 mm/month, and correlation coefficients of 0.596 and 0.590, respectively. The RMSE values for CHIRPS, GSMaP, PERSIANN, and TerraClimate were 134.72, 140.83, 139.01, and 137.39 mm/month, respectively. The four schemes had the same correlation coefficient (0.61), which is slightly higher than that of the original multi-source input data. The RMSE behavior of precipitation was similar to that of evapotranspiration. As the consistency constraint increased, the RMSE decreased, however the differences remained small. The RMSE of Hex0.5 improved by only 0.77% relative to that of Hex0.05, indicating that the improvement among the four schemes was limited.

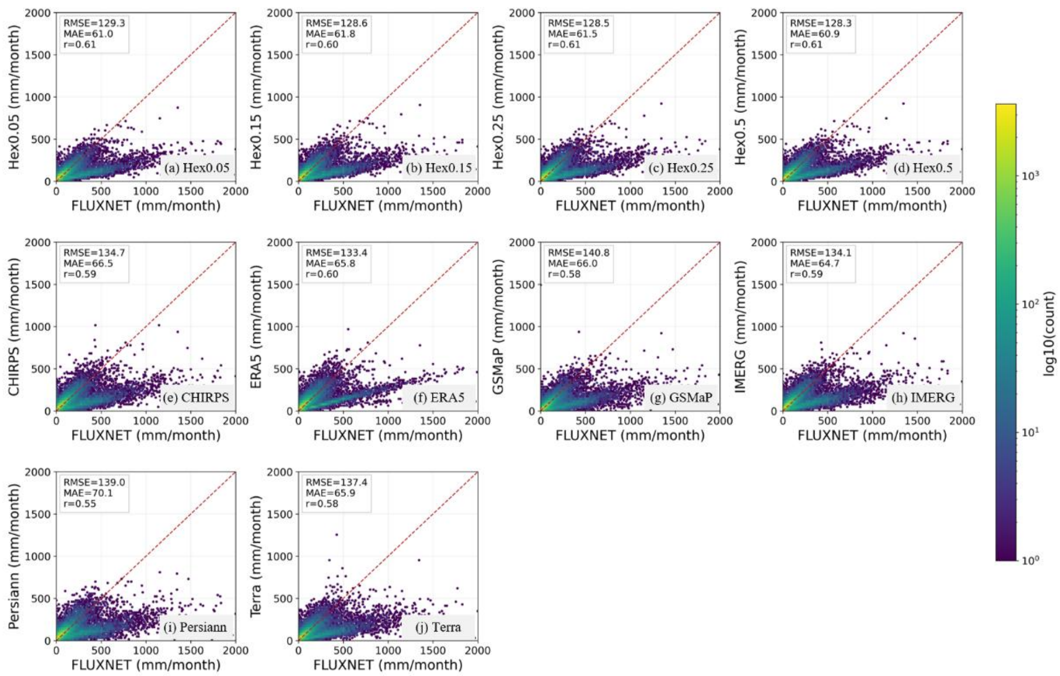


Figure 9: Comparison of multi-source input and optimized precipitation data with FLUXNET precipitation: (a) Hex0.05, (b) Hex0.15, (c) Hex0.25, (d)Hex0.5, (e) CHIRPS, (f) ERA5, (g) GSMaP, (h) IMERG, (i) Persiann, and (j) Terra.

3.5 Closure analysis

Based on the preceding results, the four settings produced broadly similar water-cycle variables, however Hex0.05 had a severe negative bias in the characterization of terrestrial water storage change. The Hex0.15 data product was not inferior to Hex0.05 in terms of data preservation, closure constraints, and basin runoff representation, and it performed comparably to the other settings in the FLUXNET independent validation. Therefore, it provided the optimal overall representation of the water cycle. Under this scheme, monthly global water balance residuals from 2000–2020 were calculated, with the residual defined as $P-E-R-\Delta S$ (Fig. 10). The global residual was generally close to zero, with a full-sample mean residual of −0.44 mm/month. The median absolute value of the multi-year mean gridded residual was 0.73 mm/month.

In terms of the spatial distribution, the annual mean residual generally fluctuated around zero. The residual magnitudes were small and close to zero across most mid- and high-latitude regions and arid and semi-arid regions. More pronounced positive and negative residuals were mainly distributed in regions with strong hydrological processes, complex terrain, or high input data uncertainty, including the Caspian Sea region at the Eurasian continental boundary, the Amazon River basin in South America, North America, high-latitude snow- and ice-covered regions, and some coastal mountainous areas. The zonal statistics along the longitude and latitude further showed that the residual fluctuated around zero and did not exhibit a persistent accumulated bias with latitude or longitude. The residuals mainly appeared as local anomalies rather than as systematic biases at the global scale. The spatial median of the residual standard deviation was 11.72 mm/month, the median RMSE was 11.77 mm/month, and the overall RMSE was 14.33 mm/month, which was 66.97% lower than the optimal multi-source input

https://doi.org/10.5194/essd-2026-497
Preprint. Discussion started: 22 September 2026

combination in Fig. 1, corresponding to a reduction of 29.04 mm/month. Regions with high standard deviations were primarily concentrated in humid tropical, monsoon, alpine snow- and ice-covered regions, and areas with strong terrain relief. These regions had strong seasonality in terms of precipitation, evapotranspiration, runoff, and terrestrial water storage changes, and their hydrological processes were highly nonlinear, resulting in large monthly residual fluctuations. In contrast, arid regions and some stable mid- and high-latitude regions had lower standard deviations, indicating that water-cycle fluxes varied less strongly and that closure was more stable in these areas.

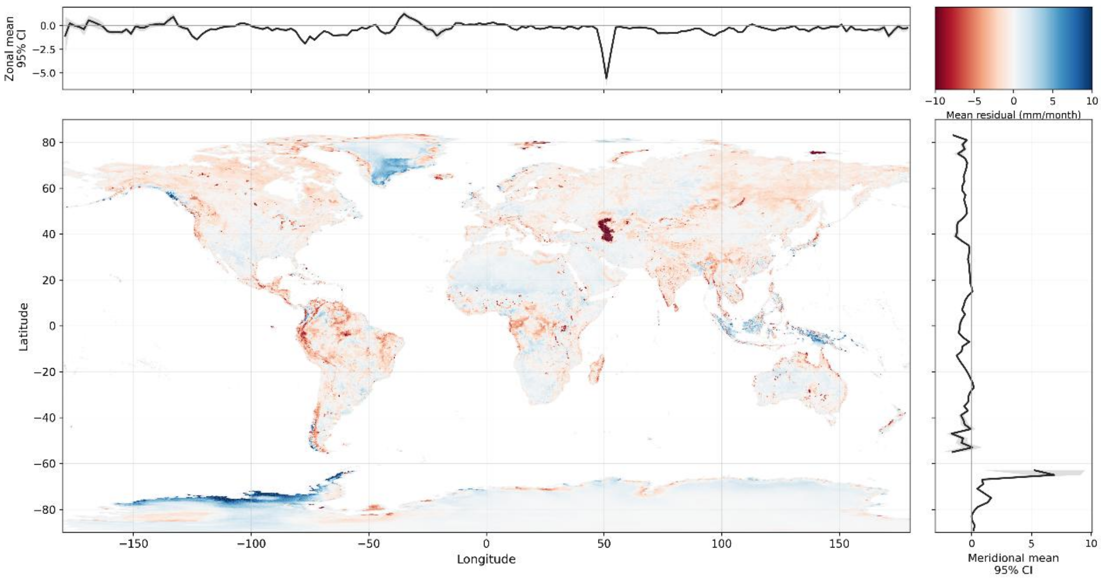


Figure 10: Spatial distribution of monthly-scale mean water-cycle closure residuals.

3.6 Analysis of typical event detection

The diagnosis of typical hydrological events further demonstrated the effectiveness of the data product. In this study, the 2020 Yangtze River Basin flood, 2010 Indus River Basin flood, 2012 United States drought, and 2019 Australian drought were selected as the representative hydrological events (Table 4). Anomalies in precipitation, evapotranspiration, runoff, and terrestrial water storage changes relative to the 2000−2020 monthly climatology data were used as statistical indicators.

Table 4. Typical hydrological events.

No.	Region	Event type	Time
(a)	East Asia	Flood	2020-07
(b)	South Asia	Flood	2010-08
(c)	North America	Drought	2012-07
(d)	Oceania	Drought	2019-12

For flood events, both the 2020 Yangtze River Basin and 2010 Indus River Basin floods showed clear increases in precipitation input and terrestrial water storage. In July, 2020, precipitation, runoff, and terrestrial water storage change anomalies in the

Yangtze River Basin were +88.24, +43.56, and +42.12 mm/month, respectively, and the proportions of grids with positive anomalies reached 0.81, 0.67, and 0.80, respectively (Fig. 11a). This flood was among the most severe flooding events in the region for nearly 60 years. Under the influence of the East Asian monsoon and persistent plum rain processes, summer precipitation in the Yangtze River Basin has increased substantially, accompanied by rising river and lake water levels and increased runoff (Wei et al., 2020; Xiong et al., 2021). During the Indus River Basin flood in August, 2010, precipitation and terrestrial water storage change anomalies were +56.40 and +33.59 mm/month, respectively, and the proportions of grids with positive anomalies reached 0.92 and 0.89, respectively (Fig. 11b). The mean runoff anomaly was +6.81 mm/month, however the proportions of positive and negative runoff anomaly grids were similar. This flood was driven jointly by anomalous moisture transport, orographic lifting, and widespread precipitation systems (Houze et al., 2011). Flash floods in mountain tributaries, mainstem flooding, and floodplain storage and detention jointly affect the spatial distribution of flooding and downstream flood peaks (Sayama et al., 2012).

In contrast to the synchronous strengthening of water input and terrestrial storage during flood events, drought events are mainly characterized by precipitation deficits and the resulting decline in actual evapotranspiration, runoff recession, and terrestrial storage depletion. Both the 2012 United States and 2019 Australian droughts showed coordinated negative anomalies across multiple variables. The former was characterized by rapid development, whereas the latter exhibited a more pronounced long-term cumulative effect. In July, 2012, the anomalies of precipitation, evapotranspiration, runoff, and terrestrial water storage change in the United States were −11.25, −4.14, −3.08, and −7.69 mm/month, respectively, and the proportions of grids with negative anomalies were 0.58, 0.53, 0.68, and 0.68, respectively. Drought is triggered by precipitation deficits caused by weakened moisture transport and suppressed convection (Hoerling et al., 2014). Precipitation deficits can propagate further into runoff and storage anomalies (Van Loon et al., 2012). During the Australian drought in December, 2019, the anomalies of precipitation, evapotranspiration, runoff, and terrestrial water storage change were −46.59, −26.19, −4.33, and −14.57 mm/month, respectively, and the proportions of grids with negative anomalies were 0.92, 0.91, 0.79, and 0.81, respectively. Australia experienced its warmest and driest year on record in 2019, and a strong positive Indian Ocean Dipole, together with antecedent accumulated precipitation deficits, intensified widespread drought (Bureau of Meteorology, 2020). In southeastern Australia, where the impacts were severe, persistent precipitation deficits and anomalously high temperatures further constrained actual evapotranspiration and cause sustained decreases in soil moisture, terrestrial water storage, groundwater, and runoff (Devanand et al., 2024).

The above diagnosis of typical events indicated that the Hex0.15 data product can capture coordinated responses among multiple variables during extreme hydrological events and can effectively characterize the basic process features of different types of hydrological extremes. During flood events, increased precipitation is typically accompanied by increased runoff and terrestrial water storage. During drought events, precipitation deficits manifest as constrained actual evapotranspiration, runoff recession, and terrestrial storage depletion.

https://doi.org/10.5194/essd-2026-497
Preprint. Discussion started: 22 September 2026

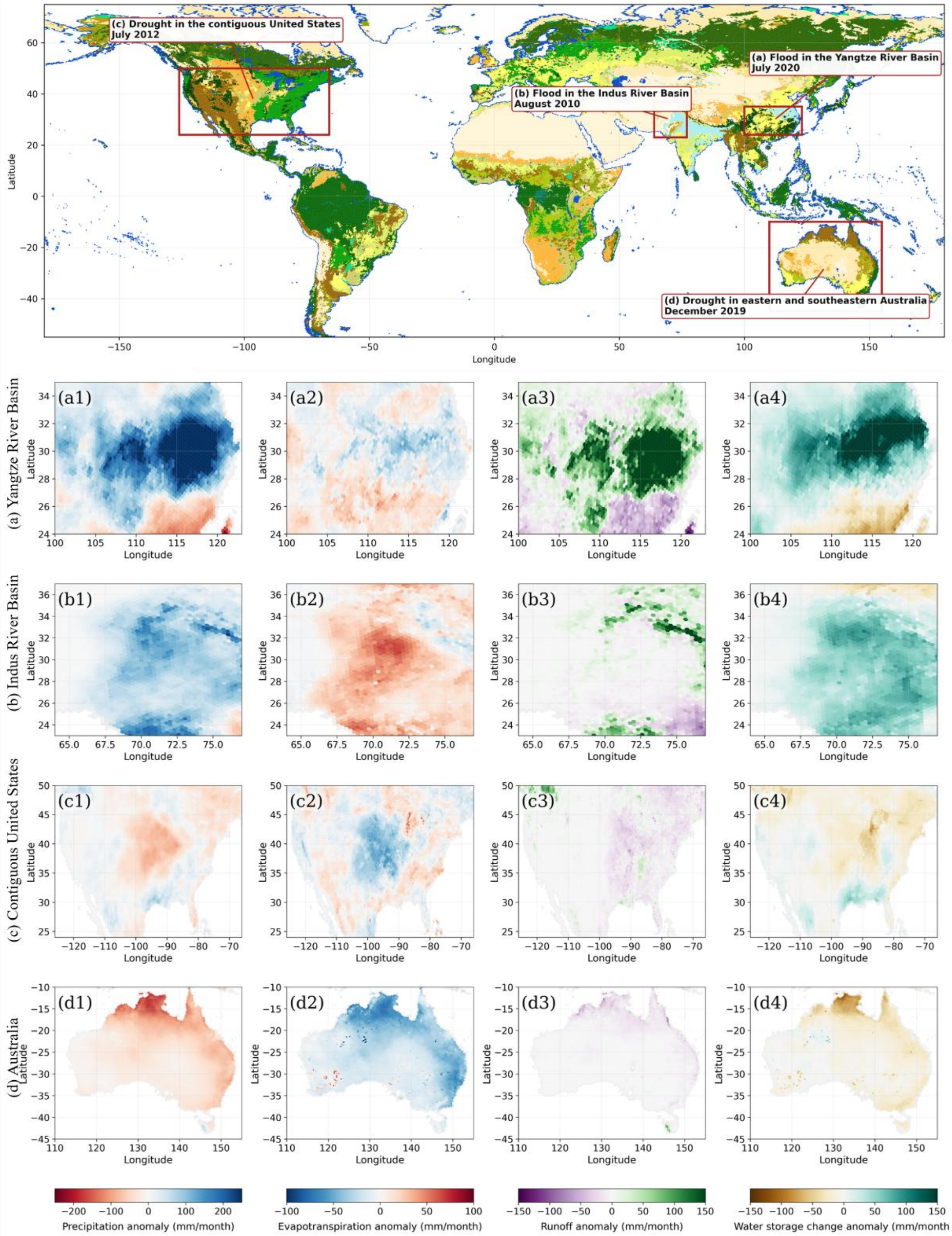


Figure 11: Distribution of the optimized data product during typical hydrological events.

https://doi.org/10.5194/essd-2026-497
Preprint. Discussion started: 22 September 2026

4. Code and Data availability

All the datasets used in our study are publicly available online and have been introduced in Sect. 2.1 Data collection. The water budget closure dataset is also publicly available in various formats (Shapefile and Zarr) (Li and Wang, 2026b) and can be freely downloaded on the Science Data Bank (https://doi.org/10.57760/sciencedb.40260). The source code associated with this study is also publicly available through the same link, which provides the main workflow from data preprocessing and feature construction to model training and prediction.

5. Conclusions

To address the difficulty in simultaneously preserving the observational realism of basin-scale measurements and the spatial continuity of gridded data, this study developed a technical pathway for transferring observational constraints from basin-scale runoff measurements to global gridded water-cycle variables. This approach bridges the gap between traditional basin-scale methods, which are unable to resolve internal spatial heterogeneity, and gridded-scale methods, which lack direct constraints on observed runoff. Using the ISEA3H10 equal-area hexagonal discrete global grid as a unified spatial framework, this method converts GRDC basin outlet runoff observations into supervised information usable by gridded models through spatial overlay and area-weighted aggregation between basins and grid cells. Under physical loss constraints, a deep-learning model was used to jointly optimize key water-cycle variables, including precipitation, evapotranspiration, runoff, and terrestrial water storage changes. Furthermore, the model further incorporates natural background variables, such as topography, vegetation, LST, snow cover, and soil moisture, allowing the estimation of water-cycle variables to use the spatial heterogeneity of external environmental factors to support water budget closure.

Comprehensive validation results showed that the data product achieved strong closure consistency while preserving major hydrological signals and spatiotemporal variability. Compared with the minimum residual RMSE of 43.37 mm/month among multi-source combinations of water-cycle variables, the residual RMSE of the optimized product was reduced to 14.33 mm/month, with a global mean residual of only −0.44 mm/month. The optimized gridded runoff had a correlation coefficient of approximately 0.90 with the GRDC-observed runoff, outperforming the original runoff proxy products from ERA5, FLDAS, MERRA, and TerraClimate, with RMSE reductions of approximately 31–49% relative to those of the original runoff proxies. Independent validation against FLUXNET sites further indicated that the optimized precipitation and evapotranspiration maintained an accuracy comparable to, or more stable than, the mainstream input products. Analysis of typical hydrological events showed that the dataset can capture coordinated multivariate responses during different types of extreme events across different regions. Overall, this study provides a spatially continuous, scale-consistent, and water budget-balanced global gridded terrestrial water cycle dataset that offers physically consistent foundational data support for global water-cycle assessment, water resource change analysis, and extreme hydrological event research.

Author contributions

Conceptualization: KL, JW; Data curation: KL, JW; Formal analysis: KL, XJ, LP; Funding acquisition: JW, CL; Investigation: JW, CL; Methodology: KL, JW; Supervision: JW; Validation: KL, XJ, LP; Visualization: KL, XJ, LP; Writing – original draft: KL, JW; Writing – review & editing: All.

Competing interests

The authors declare that they have no conflict of interest.

Financial support

This work was supported in part by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB0740202, XDB0740100), and the Key Project of Innovation LREIS (LREIS202504).

References

Abolafia-Rosenzweig, R., Pan, M., Zeng, J.L., Livneh, B., 2021. Remotely sensed ensembles of the terrestrial water budget over major global river basins: An assessment of three closure techniques. Remote Sensing of Environment 252, 112191. https://doi.org/10.1016/j.rse.2020.112191

Anjaneyulu, R., Kashyap, P., Xiong, J., Abhishek, 2025. WATcycle: A software for terrestrial water cycle and budget analysis with case studies on the Amazon and Mississippi Basins. Environmental Modelling & Software 194, 106656. https://doi.org/10.1016/j.envsoft.2025.106656

Bureau of Meteorology, 2020. Annual Australian Climate Statement 2019 [WWW Document]. URL https://www.bom.gov.au/climate/current/annual/aus/2019/ (accessed 6.15.26).

Chen, W., Liu, S., Zhao, S., Zhu, Y., Feng, S., Wang, Z., Wu, Y., Xiao, J., Yuan, W., Yan, W., Ju, H., Wang, Q., 2023. Temporal dynamics of ecosystem, inherent, and underlying water use efficiencies of forests, grasslands, and croplands and their responses to climate change. Carbon Balance Manage 18, 13. https://doi.org/10.1186/s13021-023-00232-2

Devanand, A., Falster, G.M., Gillett, Z.E., Hobeichi, S., Holgate, C.M., Jin, C., Mu, M., Parker, T., Rifai, S.W., Rome, K.S., Stojanovic, M., Vogel, E., Abram, N.J., Abramowitz, G., Coats, S., Evans, J.P., Gallant, A.J.E., Pitman, A.J., Power, S.B., Rauniyar, S.P., Taschetto, A.S., Ukkola, A.M., 2024. Australia's Tinderbox Drought: An extreme natural event likely worsened by human-caused climate change. Science ADvAnceS.

Feng, J., Li, B., Song, J., Tang, B., Nyein, M.M., Tani, B.P., 2025. Spatiotemporal Variations of Terrestrial Water Storage and Driving Factors in the Water Towers of Northwest China Based on GRACE and Multi-Source Data Sets. Water Resources Research 61, e2024WR039490. https://doi.org/10.1029/2024WR039490

Ghiggi, G., Humphrey, V., Seneviratne, S.I., Gudmundsson, L., 2019. GRUN: an observation-based global gridded runoff dataset from 1902 to 2014. Earth System Science Data 11, 1655–1674. https://doi.org/10.5194/essd-11-1655-2019

Harrigan, S., Zsoter, E., Alfieri, L., Prudhomme, C., Salamon, P., Wetterhall, F., Barnard, C., Cloke, H., Pappenberger, F., 2020. GloFAS-ERA5 operational global river discharge reanalysis 1979–present. Earth System Science Data 12, 2043–2060. https://doi.org/10.5194/essd-12-2043-2020

Heberger, M., Aires, F., Pellet, V., 2025. Improving satellite remote sensing estimates of the global terrestrial water cycle via neural network modeling. Journal of Hydrology 662, 133825. https://doi.org/10.1016/j.jhydrol.2025.133825

Hoerling, M., Eischeid, J., Kumar, A., Leung, R., Mariotti, A., Mo, K., Schubert, S., Seager, R., 2014. Causes and Predictability of the 2012 Great Plains Drought. Bulletin of the American Meteorological Society 95, 269–282. https://doi.org/10.1175/BAMS-D-13-00055.1

Houze, R.A., Rasmussen, K.L., Medina, S., Brodzik, S.R., Romatschke, U., 2011. Anomalous Atmospheric Events Leading to the Summer 2010 Floods in Pakistan. Bulletin of the American Meteorological Society 92, 291–298. https://doi.org/10.1175/2010BAMS3173.1

Huang, H., Liu, J., Chen, A., Ruiz-Vásquez, M., Orth, R., 2025. State-of-the-art hydrological datasets exhibit low water balance consistency globally. Earth System Science Data Discussions 1–23. https://doi.org/10.5194/essd-2025-376

Kevin, S., 2018. DGGRID version 6.4 User Documentation for Discrete Global Grid Generation Software.

Lehmann, F., Vishwakarma, B.D., Bamber, J., 2022. How well are we able to close the water budget at the global scale? Hydrology and Earth System Sciences 26, 35–54. https://doi.org/10.5194/hess-26-35-2022

Li, K., Wang, J., 2026a. Hexagonal discrete global grid system enhances model latitudinal spatial generalization and deep learning capabilities. Big Earth Data 0, 1–23. https://doi.org/10.1080/20964471.2026.2666953

Li, K., Wang, J., 2026b. A GRDC observation-constrained terrestrial water budget dataset under a discrete global grid framework (2000-2020). https://doi.org/10.57760/sciencedb.40260

Li, Xueying, Long, D., Scanlon, B.R., Mann, M.E., Li, Xingdong, Tian, F., Sun, Z., Wang, G., 2022. Climate change threatens terrestrial water storage over the Tibetan Plateau. Nat. Clim. Chang. 12, 801–807. https://doi.org/10.1038/s41558-022-01443-0

Loomis, B.D., Luthcke, S.B., Sabaka, T.J., 2019. Regularization and error characterization of GRACE mascons. J Geod 93, 1381–1398. https://doi.org/10.1007/s00190-019-01252-y

Luo, Z., Zhang, S., Ding, X., Wang, L., Shao, Q., Huang, H., Chen, X., Li, H., 2026. A framework for assessing the impact trends of neglecting water surface evaporation and substituting streamflow on water budget closure. Journal of Hydrology 664, 134391. https://doi.org/10.1016/j.jhydrol.2025.134391

Michailovsky, C.I., Coerver, B., Mul, M., Jewitt, G., 2023. Investigating sources of variability in closing the terrestrial water balance with remote sensing. Hydrology and Earth System Sciences 27, 4335–4354. https://doi.org/10.5194/hess-27-4335-2023

Miralles, D.G., Bonte, O., Koppa, A., Baez-Villanueva, O.M., Tronquo, E., Zhong, F., Beck, H.E., Hulsman, P., Dorigo, W., Verhoest, N.E.C., Haghdoost, S., 2025a. GLEAM4: global land evaporation and soil moisture dataset at 0.1° resolution from 1980 to near present. Sci Data 12, 416. https://doi.org/10.1038/s41597-025-04610-y

Miralles, D.G., Bonte, O., Koppa, A., Baez-Villanueva, O.M., Tronquo, E., Zhong, F., Beck, H.E., Hulsman, P., Dorigo, W., Verhoest, N.E.C., Haghdoost, S., 2025b. GLEAM4. https://doi.org/10.5281/zenodo.14724263

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D.G., Piles, M., Rodríguez-Fernández, N.J., Zsoter, E., Buontempo, C., Thépaut, J.-N., 2021. ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. Earth System Science Data 13, 4349–4383. https://doi.org/10.5194/essd-13-4349-2021

Oki, T., Kanae, S., 2006. Global Hydrological Cycles and World Water Resources. Science 313, 1068–1072. https://doi.org/10.1126/science.1128845

Pan, M., Sahoo, A.K., Troy, T.J., Vinukollu, R.K., Sheffield, J., Wood, E.F., 2012. Multisource Estimation of Long-Term Terrestrial Water Budget for Major Global River Basins. Journal of Climate 25, 3191–3206. https://doi.org/10.1175/JCLI-D-11-00300.1

Pascolini-Campbell, M.A., Reager, J.T., Fisher, J.B., 2020. GRACE-based Mass Conservation as a Validation Target for Basin-Scale Evapotranspiration in the Contiguous United States. Water Resources Research 56, e2019WR026594. https://doi.org/10.1029/2019WR026594

Pastorello, G., Trotta, C., Canfora, E., et al., 2020. The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. Sci Data 7, 225. https://doi.org/10.1038/s41597-020-0534-3

Pellet, V., Aires, F., Yamazaki, D., Papa, F., 2021. Coherent Satellite Monitoring of the Water Cycle Over the Amazon. Part 1: Methodology and Initial Evaluation. Water Resources Research 57, e2020WR028647. https://doi.org/10.1029/2020WR028647

Save, H., 2020. CSR GRACE and GRACE-FO RL06 Mascon Solutions v02. https://doi.org/10.1002/2016JB013007

https://doi.org/10.5194/essd-2026-497
Preprint. Discussion started: 22 September 2026

Save, H., Bettadpur, S., Tapley, B.D., 2016. High-resolution CSR GRACE RL05 mascons. Journal of Geophysical Research: Solid Earth 121, 7547–7569. https://doi.org/10.1002/2016JB013007

Sayama, T., Ozawa, G., Kawakami, T., Nabesaka, S., Fukami, K., 2012. Rainfall–runoff–inundation analysis of the 2010 Pakistan flood in the Kabul River basin. Hydrological Sciences Journal 57, 298–312. https://doi.org/10.1080/02626667.2011.644245

Sheffield, J., Ferguson, C.R., Troy, T.J., Wood, E.F., McCabe, M.F., 2009. Closing the terrestrial water budget from satellite remote sensing. Geophysical Research Letters 36. https://doi.org/10.1029/2009GL037338

Van Loon, A.F., Van Huijgevoort, M.H.J., Van Lanen, H. a. J., 2012. Evaluation of drought propagation in an ensemble mean of large-scale hydrological models. Hydrology and Earth System Sciences 16, 4057–4078. https://doi.org/10.5194/hess-16-4057-2012

Wei, K., Ouyang, C., Duan, H., Li, Y., Chen, M., Ma, J., An, H., Zhou, S., 2020. Refections on the catastrophic 2020 Yangtze River basin flooding in southern China. innovation 1.

Xiong, J., Guo, S., Yin, J., Gu, L., Xiong, F., 2021. Using the Global Hydrodynamic Model and GRACE Follow-On Data to Access the 2020 Catastrophic Flood in Yangtze River Basin. Remote Sensing 13, 3023. https://doi.org/10.3390/rs13153023

Zhang, Y., Blöschl, G., Wei, H., Kong, D., Ma, N., Wagener, T., Tian, J., Xia, J., Li, C., Wang, L., Chiew, F.H.S., Leung, L.R., Liu, X., Zheng, H., Zhang, X., Liu, C., 2026. Overestimation of past and future increases in global river flow by Earth system models. Nat. Geosci. 19, 311–316. https://doi.org/10.1038/s41561-025-01897-9

Zhang, Y., Pan, M., Sheffield, J., Siemann, A.L., Fisher, C.K., Liang, M., Beck, H.E., Wanders, N., MacCracken, R.F., Houser, P.R., Zhou, T., Lettenmaier, D.P., Pinker, R.T., Bytheway, J., Kummerow, C.D., Wood, E.F., 2018. A Climate Data Record (CDR) for the global terrestrial water budget: 1984–2010. Hydrology and Earth System Sciences 22, 241–263. https://doi.org/10.5194/hess-22-241-2018

Zheng, X., Liu, D., Huang, S., Wang, H., Meng, X., 2025. Achieving water budget closure through physical hydrological process modelling: insights from a large-sample study. Hydrology and Earth System Sciences 29, 627–653. https://doi.org/10.5194/hess-29-627-2025