A novel method for correcting water budget components and reducing their uncertainties by optimally distributing the imbalance residual without full closure

Zengliang Luo[a,b], Hanjia Fu[a,b], Quanxi Shao[c]*, Wenwen Dong[a,b], Xi Chen[d], Xiangyi Ding[e]*, Lunche Wang[a,b], Xihui Gu[a], Ranjan Sarukkalige[f], Heqing Huang[a], Huan Li[g]

[a] Hubei Key Laboratory of Regional Ecology and Environmental Change, State Key Laboratory of Biogeology and Environmental Geology, China University of Geosciences, Wuhan 430074, China.

[b] Engineering Research Center of Natural Resource Information Management and Digital Twin Engineering Software, Ministry of Education, Wuhan 430074, China.

[c] CSIRO Data61, Australian Resources Research Centre, 26 Dick Perry Avenue, Kensington, WA 6155, Australia.

[d] State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, 100101, China.

[e] Department of Water Resources, China Institute of Water Resources and Hydropower Research, Beijing 100038, China.

[f] School of Civil and Mechanical Engineering, Curtin University, GPO Box U1987, Perth, WA 6845, Australia.

[g] HUN-REN Balaton Limnological Research Institute, Hungary.

Corresponding author to: Quanxi Shao (Quanxi.Shao@data61.csiro.au) and Xiangyi Ding (dingxy@iwhr.com)

Highlights:

- Existing correction methods may introduce large errors, and more seriously cause unrealistic negative values in P, ET and Q in up to 10% of cases.
- A novel IWE-Res method is proposed to improve the accuracy and consistency of corrected satellite-based water budget component data.
- In most river basins (except cold regions), the best correction is achieved by adjusting 40% to 90% of the total water imbalance error.

Abstracts: Achieving water budget closure improves the consistency of water budget component datasets, including precipitation (P), evapotranspiration (ET), streamflow (Q) and terrestrial water storage change (TWSC), thereby advancing our

understanding of basin-scale water cycle dynamics. Existing water budget closure

correction (BCC) methods typically aim to eliminate the entire water imbalance error

(ΔRes) by fully redistributing it across budget components. However, this often

overlooks the trade-off between achieving perfect closure and the errors introduced

into the corrected components through this redistribution. Moreover, inaccurate

estimation of redistribution weights can lead to contradictory outcomes, such as

negative values in P, ET, or Q. In this study, we quantify the uncertainties introduced

by four existing BCC methods (CKF, MCL, MSD, and PR) at the monthly scale

across 84 basins spanning diverse climate zones. We then propose a novel method,

IWE-Res, which identifies an optimal redistributing strategy by minimizing the

combined error from both the errors introduced to individual budget components and

the remaining ΔRes error. This method also reduces the occurrence of negative values

in the corrected datasets. Our results show: (1) Existing BCC methods can result in

negative values in 0–10% of the time series for each corrected budget component

(typically < 5%); (2) The proposed IWE-Res method improves the accuracy of

corrected components compared to existing methods, reducing RMSE by 29.5% for P,

24.7% for ET, 69.0% for Q, and 6.8% for TWSC; and (3) For most basins, excluding

those in cold regions, the optimal redistribution is achieved when 40%–90% of ΔRes

is redistributed. By offering a more balanced approach to water budget closure, this

study improves the accuracy and reliability of corrected budget component datasets.

Keywords: Water budget closure; Budget components; Water imbalance; Uncertainty

identification; Global hydrology

1 Introduction

The terrestrial water balance represents a fundamental physical framework that describes the distribution and movement of water across the Earth's land surface (Lehmann et al., 2022) and is governed by four interconnected components—precipitation (P), evapotranspiration (ET), streamflow (Q), and terrestrial water storage change (TWSC)—that together regulate the exchange of water among the atmosphere, land, and oceans (Abolafia-Rosenzweig et al., 2021; Sahoo et al., 2011; Chen et al., 2020; Wang et al., 2015). These components are dynamically linked and respond to climatic variability, land surface heterogeneity, and human interventions across a range of spatial and temporal scales. Achieving water budget closure (that is, ensuring internal consistency among these fluxes and storages), Equation (1) is essential for advancing our understanding of hydrological processes (Li et al., 2024; Mourad et al., 2024).

$$P - ET - Q - TWSC = 0 \qquad (1)$$

where P represents precipitation, ET represents evapotranspiration, Q represents streamflow, and TWSC represents terrestrial water storage change. It is worth noting that TWSC refers to the change in total terrestrial water storage, including but not limited to surface water, soil moisture, groundwater, water infiltrating into aquifers, and ice/snow (Mehrnegar et al., 2023; Pellet et al., 2020; Wang et al., 2022). Infiltrated water into aquifers is not permanently stored, but eventually returned to major water bodies sooner or later (Levison et al., 2016). The ability of aquifers to retain or transmit infiltrated water is strongly influenced by local geological

characteristics, particularly the spatial heterogeneity, presence of fractures, or high-permeability pathways (Levison et al., 2016; Schiavo, 2023).

Despite its importance, obtaining observational datasets that achieve water balance closure remains a major challenge. In practice, no single observational system can simultaneously measure all four water budget components at the required resolution and accuracy. Each budget component is typically derived from independent data sources or models with differing spatial and temporal characteristics, which complicates the direct closure of the terrestrial water budget.

P is typically derived from point-based rain gauge networks, which are generally reliable but often incomplete, requiring gap-filling (Esquivel-Arriaga et al., 2024; Nassaj et al., 2022; Bai et al., 2021; Lockhoff et al., 2014). The main source of uncertainty lies in the spatial distribution and representativeness of these gauges, particularly in relation to P type (Bai et al., 2019; Trenberth et al., 2014). Spatial uncertainty tends to be low for widespread frontal systems but can be substantial for localized convective storms (Palharini et al., 2020). Gauge placement is often dictated by accessibility and logistical convenience, which may lead to underestimation of the uncertainty in daily P inputs (Wang et al., 2017; Bai et al., 2019; Wu et al., 2018). Satellite-based P estimates have demonstrated good performance in capturing frontal rainfall, but not in other rainfall types (Masunaga et al., 2019; Petković et al., 2017; Palharini et al., 2020). ET is commonly estimated by empirical or physically based models (Jacobs et al., 1998; McMahon et al., 2016; Allen et al., 1998). Although these models are generally well calibrated, uncertainties persist due to the complex

influence of advection and localized meteorological variability, especially in small catchments. At larger spatial scales, energy balance approaches tend to provide sufficiently accurate estimates (Hua et al., 2020; Hao et al., 2018; Ruhoff et al., 2022). Q measurements typically exhibit low uncertainty when rating curves are well established and regularly maintained (Jian et al., 2015; Krabbenhoft et al., 2022). However, uncertainty can still arise from the delineation of watershed boundaries, particularly in regions where groundwater flow does not align with surface catchment divides (Huang et al., 2023; Bouaziz et al., 2018). This mismatch can result in misrepresentation of actual hydrological contributions. TWSC generally has a negligible impact on water balance calculations over multi-year periods, but can significantly affect short-term (e.g., daily) balances (He et al., 2023; Zhang et al., 2016). A key challenge is to define the effective depth over which TWSC should be quantified, as changes in soil moisture near the surface are more easily observed than those occurring at greater depths.

Hydrological models, which are grounded in the principle of mass conservation and explicitly implement the water balance equation, offer an alternative to direct observation for achieving water budget closure. However, in practice, model structure simplifications, parameter uncertainties, and errors in meteorological forcing data introduce substantial biases and propagate uncertainty across simulated components. These limitations make it equally difficult to achieve water budget closure using hydrological modeling alone.

In recent years, the rapid expansion of remote sensing and reanalysis datasets has

significantly improved global access to budget components, offering new

opportunities for data-driven analysis of hydrological processes. However, even these

advanced products often exhibit internal water budget inconsistencies. To address this

issue, a growing number of studies have adopted water budget closure correction

(BCC) methods to reduce water imbalance error (ΔRes), with the goal of forcing

ΔRes from a non-zero value (ΔRes $\neq$ 0) to theoretical closure (ΔRes = 0), where

ΔRes = P – ET – Q – TWSC (Zhou et al., 2024; Munier et al., 2018; Zhang et al.,

2016). Common methods include Proportional Redistribution (PR), the Constrained

Kalman Filter (CKF), Multiple Collocation (MCL), and the Minimized Series

Deviation (MSD) method (Pan et al., 2012; Luo et al., 2023). For example, Abhishek

et al. (2021) applied the PR, CKF, and MCL methods to quantify water budget closure

and uncertainties in budget components in the upper Chao Phraya River basin;

Abolafia-Rosenzweig et al. (2021) evaluated the effectiveness of PR, CKF, and MCL

methods in closing the water budget for 24 global basins; Dastjerdi et al. (2024)

developed a precipitation data merging method to improve precipitation estimates

based on existing BCC methods.

Existing BCC methods redistribute the entire ΔRes error among water budget

components to enforce strict water budget closure. This redistribution is typically

guided by the relative uncertainties of the individual components, based on the

assumption that the entire residual error originates from observational or modeling

errors in these datasets. However, this assumption overlooks the fact that ΔRes is not

solely the result of measurement or estimation errors in P, ET, Q, or TWSC. Rather, it

is a composite residual that also reflects contributions from systematic biases and the omission of unmeasured components. These include deep groundwater exchanges that may cross basin boundaries, snow and glacier storage changes (particularly in high-altitude or high-latitude regions), and anthropogenic influences such as irrigation withdrawals, reservoir operations, and inter-basin water transfers. Because existing BCC methods do not explicitly account for these additional sources of imbalance, forcing strict closure by allocating the entire ΔRes to the measured components can introduce unrealistic uncertainties. As a result, the application of existing BCC methods—despite their goal of improving internal consistency—often leads to limited improvements, or, in some cases, even a decline in the accuracy of the corrected hydrological datasets.

A clear manifestation of this limitation is the occurrence of negative values in corrected budget component datasets when applying existing BCC methods at the monthly scale, such as negative P, ET, and Q. These unrealistic negative values arise when an excessive share of the ΔRes is redistributed to specific components. For instance, if the BCC method overestimates the error in a specific component, it may assign an excessively large portion of ΔRes to that component. When the magnitude of the correction exceeds the component's original value, the result is a negative flux, which is hydrologically incorrect. Beyond introducing negative values, such imbalanced redistribution compromises the integrity of the remaining components. Overcorrecting one variable necessarily reduces the share of ΔRes available for others, potentially degrading their accuracy. Our previous work demonstrated that enforcing

water budget closure can, to some extent, reduce the accuracy of individual components and tends to introduce an ET regulation factor to mitigate accuracy loss in ET caused by existing BCC methods (Luo et al., 2023). A more hydrologically sound approach may involve partial closure, whereby only the portion of ΔRes attributable to quantified uncertainties is redistributed, while the residual linked to unmeasured processes is retained.

The key question we aim to answer in this study is the extent of uncertainty introduced into budget components by existing BCC methods for enforcing water budget closure and, more critically, whether this uncertainty exceeds the reduction in the ΔRes error. If the introduced uncertainty outweighs the error reduction, fully closing the water budget may not be necessary. As noted earlier, ΔRes represents a composite error, whereas existing BCC methods primarily address errors in budget components. Therefore, an optimal balance for redistributing the ΔRes error should be identified—one that minimizes the combined error from budget components and the remaining water imbalance. This optimal balance allows for redistributing only the portion of ΔRes attributable to errors in budget components, rather than the entire ΔRes, thereby preventing the occurrence of negative values in budget components due to improper error redistribution. However, research on identifying this optimal balance, which is crucial for improving existing BCC methods, remains lacking.

The primary goals of this study are to quantify the uncertainties introduced by existing BCC methods in closing the water budget from multiple perspectives and to propose a new IWE-Res method for identifying the optimal balance in ΔRes

redistribution. To enhance the robustness of error analysis and validate the proposed IWE-Res method, we applied four existing BCC methods with varying principles and complexities (PR, CKF, MCL, and MSD) across 84 global basins with diverse climatic characteristics. The specific objectives of this study are:

(1) To quantify the uncertainties introduced into budget components by enforcing water budget closure using existing BCC methods from multiple perspectives, including uncertainties relative to observations, the occurrence of negative values in budget components, and deviations from the original budget component datasets. This analysis provides a more comprehensive understanding of the trade-offs between achieving water budget closure and the associated errors;

(2) To analyze in detail the occurrence of negative corrected values in budget components caused by existing BCC methods, including the proportion of negative values within the time series of each budget component and their spatial distribution under varying climatic conditions;

(3) To compare the reduction in ΔRes with the corresponding increase in budget component errors resulting from enforced water budget closure;

(4) To propose a new method (IWE-Res) for identifying the optimal balance in ΔRes redistribution, minimizing the combined error from both introduced budget component errors and the remaining ΔRes error. The accuracy and reliability of the proposed IWE-Res method were validated through comparisons with existing BCC methods (PR, CKF, MCL, MSD).

2 Study area and data

210   To robustly quantify the uncertainties introduced by existing BCC methods in

211   closing the water budget and to assess the accuracy of the proposed IWE-Res method

212   across different climate zones, multiple river basins worldwide were selected as study

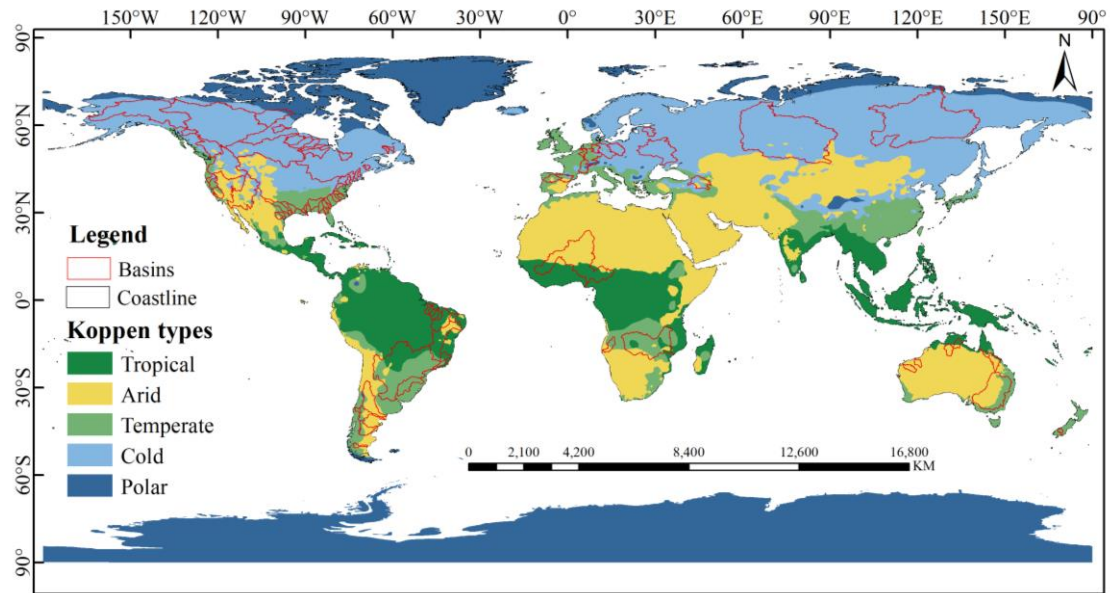
213   areas. In total, 84 basins (Fig. 1) were chosen based on the availability of streamflow

214   observations from the Global Runoff Data Centre (GRDC) for the period 2002–2020.

215   To ensure data reliability, the proportion of missing data was kept below 10%, with

216   missing values interpolated using a linear method. Notably, approximately 90% of the

217   basins used in this study had less than 5% missing data.

218

219   **Fig. 1** Overview of the Study Area. The climate classification used in this study is

220   based on the Köppen climate classification system.

221

222   The climate classifications presented in Fig. 1 were determined using the Köppen

223   climate classification system, a widely adopted framework that categorizes global

224   climates based on temperature and precipitation thresholds (Crosbie et al., 2012;

225   Hansford et al., 2020; Liu et al., 2022; Papacharalampous et al., 2023). This system

divides the world into five primary climate types—Tropical, Arid, Temperate, Cold, and Polar. Its key strength lies in its integration of climate data with vegetation distribution, making it highly relevant to ecological environments.

For each budget component, multiple datasets are typically available, with accuracy varying across different basins. No single dataset consistently performs best across all global basins. Therefore, multiple datasets were selected for each budget component to generate various data combinations (Equations 2–3). This approach ensures the inclusion of the most suitable dataset combinations while mitigating uncertainties associated with reliance on a single dataset.

Given the biases in the outputs of global P and ET models, observationally constrained datasets that have undergone bias correction or rigorous quality control are generally considered more accurate and reliable (Ehret et al., 2012). Accordingly, priority was given to datasets that incorporate extensive ground-based observations and provide bias-corrected or quality-controlled products. We selected four P datasets—GPCC, GPM IMERG, MSWEP, and PERSIANN-CDR; three ET datasets—GLDAS, GLEAM, and TerraClimate; and three TWSC datasets derived from GRACE satellite observations—GRACE CSR, GRACE GFZ, and GRACE JPL. All datasets were either bias-corrected according to the standards of their respective data providers or subjected to systematic quality control. Observed Q data were obtained from the GRDC platform. The above datasets were upscaled to the basin and monthly scales using spatial and temporal averaging. By combining these datasets, a total of 36 distinct data combinations were generated for each basin (Equation 3).

$$248 \qquad C_{jkl} = \begin{bmatrix} P_j & ET_k & TWSC_l & Q \end{bmatrix} \qquad (2)$$

where j, k, and l represent the indices of the datasets corresponding to each budget

component. Table 1 provides basic information on the datasets used in this study,

along with their corresponding indices. Equation 3 represents a matrix composed of

the elements defined in Equation 2.

$$253 \qquad C = \begin{bmatrix} C_{111} & C_{112} & C_{113} & C_{121} & C_{122} & C_{123} & C_{131} & C_{132} & C_{133} \\ C_{211} & C_{212} & C_{213} & C_{221} & C_{222} & C_{223} & C_{231} & C_{232} & C_{233} \\ C_{311} & C_{312} & C_{313} & C_{321} & C_{322} & C_{323} & C_{331} & C_{332} & C_{333} \\ C_{411} & C_{412} & C_{413} & C_{421} & C_{422} & C_{423} & C_{431} & C_{432} & C_{433} \end{bmatrix} \qquad (3)$$

The Global Precipitation Climatology Centre (GPCC) dataset, provided by the

German Weather Service (DWD), is derived from a dense global network of rain

gauge observations, and incorporates strict quality control procedures such as station

data validation, temporal consistency checks, and outlier removal (Becker et al., 2013;

Schneider et al., 2008). The dataset is available at 0.25° spatial resolution and daily to

monthly temporal scales. The Global Precipitation Measurement Integrated

Multi-Satellite Retrievals (GPM IMERG) Final Run product, developed by NASA

and JAXA, integrates multiple satellite-based precipitation estimates and applies

monthly bias correction using ground-based gauge data (Wang et al., 2017; Cui et al.,

2020; Huang et al., 2019). The Multi-Source Weighted-Ensemble Precipitation

(MSWEP) dataset combines satellite, gauge, and reanalysis data using an

ensemble-weighted approach, incorporating over 77,000 ground stations for

daily-scale bias correction (Beck et al., 2019; Beck et al., 2017). The

PERSIANN-CDR dataset, based on satellite remote sensing and artificial neural

network technology, spans 60°S to 60°N with 0.25° daily resolution, and is

bias-corrected using the GPCP monthly product, which includes extensive rain gauge

observations (Chen et al., 2020; Kaprom et al., 2025; Sadeghi et al., 2019).

For ET, the Global Land Data Assimilation System (GLDAS), developed by

NASA and NOAA, uses land surface modeling and data assimilation to produce

physically consistent estimates of land surface fluxes. The GLEAM dataset,

developed by the Miralles team at the University of Bristol, estimates actual ET using

satellite-derived net radiation and air temperature via the Priestley-Taylor model, and

applies a stress factor derived from vegetation optical depth (VOD) and soil moisture

to adjust potential evaporation. TerraClimate dataset provides global monthly actual

ET estimates based on the Penman Montieth approach (Abatzoglou et al., 2018).

Notably, bias correction in global ET products is generally less systematic than for P

products, mainly due to the limited availability and spatial coverage of in situ flux

tower observations. As a result, bias adjustments in ET datasets are typically indirect,

relying on corrections applied to the climate forcing variables rather than to ET itself.

The launch of the GRACE and GRACE Follow-On (GRACE-FO) satellite

missions has provided new opportunities for more accurate observations of large-scale

TWSC. GRACE operated from 2002 to 2017, followed by GRACE-FO starting in

2018 (Boergens et al., 2024). These missions infer terrestrial total TWSC by tracking

temporal variations in Earth's gravity field, which are primarily attributed to changes

in terrestrial water mass. The GRACE TWSC datasets used in this study are provided

by the University of Texas Center for Space Research (CSR), the German Research

Centre for Geosciences (GFZ), and NASA's Jet Propulsion Laboratory (JPL), all of

291 which include multiple bias correction procedures to improve data quality (Landerer

292 et al., 2012; Shamsudduha et al., 2017). These bias correction procedures include

293 filtration to suppress correlated noise and striping artifacts (Swenson et al., 2006),

294 replacement of poorly resolved spherical harmonic coefficients (e.g., degree-2 term

295 C20) with satellite laser ranging data (Loomis et al., 2020), and correction for glacial

296 isostatic adjustment (GIA) (Peltier et al., 2012; Mu et al., 2017). Numerous studies

297 have demonstrated the sensitivity and reliability of GRACE satellite data for

298 monitoring TWSC (Swenson and Wahr, 2006; Resende et al., 2019; Majid et al., 2019;

299 Reager et al., 2014).

300 The GRDC provides the most comprehensive open-access river discharge data

301 available worldwide, collected from national hydrological agencies. This dataset

302 includes river streamflow measurements from over 10,000 stations across 159

303 countries (Su et al., 2024). To minimize the impact of missing data on the reliability of

304 the results, hydrological stations were selected based on the criterion that missing

305 values accounted for less than 10% of the total dataset. Linear interpolation was then

306 applied to fill any remaining data gaps.

307 **Table 1** Datasets used for this study.

Variable	Data Source	Number	Resolution	Reference
	Global Precipitation Climatology Centre (GPCC)	1	0.25°/month	Schneider et al. (2008)
	Global Precipitation Measurement (GPM IMERG)	2	0.1°/month	Huffman et al. (2015)
P	Multi-Source Weighted-Ensemble Precipitation (MSWEP)	3	0.1°/month	Beck et al. (2019)
	Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks —	4	0.25°/month	Hsu et al. (1997)

	Climate Data Record (PERSIANN-CDR)				
ET	global land data assimilation system (GLDAS)	1	0.25°/month		Park and Choi (2015)
	Global Land Evaporation Amsterdam Model (GLEAM)	2	0.25°/month		Miralles et al. (2011)
	TerraClimate	3	1/24°/month		Abatzoglou et al. (2018)
TWSC	Gravity Recovery and Climate Experiment (GRACE CSR)	1	1.0°/month		Watkins et al. (2015)
	Gravity Recovery and Climate Experiment (GRACE GFZ)	2	1.0°/month		Watkins et al. (2015)
	Gravity Recovery and Climate Experiment (GRACE JPL)	3	1.0°/month		Watkins et al. (2015)
Q	Global Runoff Data Centre (GRDC)	-	-		Burek and Smilovic (2022)

308

3 Methods

3.1 Water imbalance error

The water balance equation describes the conservation of mass between water inflows, outflows, and changes in storage within a given region (Equation 1). However, in practice, this balance is rarely achieved due to various sources of error. These include systematic biases (such as missed portions of outflow resulted from unclosed basin boundaries and inaccuracies in catchment area delineation, particularly in small basins), measurement uncertainties, and the omission of unmeasured components. Consequently, each budget component (P, ET, Q, and TWSC) is subject to an associated error term (denoted as ε_P, ε_{ET}, ε_Q, ε_{TWSC}, respectively), leading to a non-closure of the water budget (i.e., Equation 1 becomes Equation 4) (Aires, 2014; Wong et al., 2021). The resulting imbalance is represented by the residual error term ΔRes (Equation 5), which quantifies the inconsistency among the observed or

estimated components of the water cycle.

Minimizing the ΔRes error is a key objective in practical hydrological applications, as it enhances the accuracy and reliability of budget component datasets. However, it is important to note that smaller ΔRes values may arise from error compensation among budget components rather than genuine improvements in data accuracy. Therefore, a high-precision water balance dataset is characterized not only by a near-zero ΔRes error but also by budget components that closely approximate their true values (Luo et al., 2023).

$$(P + \varepsilon_P) - (ET + \varepsilon_{ET}) - (Q + \varepsilon_Q) - (TWSC + \varepsilon_{TWSC}) = 0 \qquad (4)$$

$$\Delta Res = \varepsilon_{ET} + \varepsilon_Q + \varepsilon_{TWSC} - \varepsilon_P = P - ET - Q - TWSC \qquad (5)$$

where ε_P, ε_{ET}, ε_Q, ε_{TWSC} are the errors in budget components of P, ET, Q, and TWSC relative to their true values, respectively.

3.2 Existing water budget closure correction methods

To minimize the ΔRes error in Equation 5 (reducing ΔRes from ≠0 to 0), various statistical BCC methods have been developed. These methods differ in their principles for redistributing the ΔRes error, leading to varying levels of introduced uncertainty. To systematically assess the uncertainties associated with existing BCC methods in closing the water budget and to reduce uncertainty in method selection, we evaluated four representative methods: PR, CKF, MCL, and MSD (Luo et al., 2023; Abolafia-Rosenzweig et al., 2021; Dastjerdi et al., 2024). In the following application of these BCC methods, the TWSC data used in this study refer to the basin-scale total terrestrial water storage change observed by GRACE satellite data.

344   For each basin, these four methods were applied to 36 different data

345 combinations (Equation 3), yielding 144 uncertainty estimates. The optimal

346 combinations were identified using a 5% threshold. By averaging the errors

347 introduced into budget components across these selected optimal combinations, we

348 quantified the uncertainty associated with existing BCC methods. This approach

349 minimizes uncertainties arising from both BCC method selection and budget

350 component data selection, enabling a more objective evaluation of the errors

351 introduced by existing BCC methods. A brief overview of the four BCC methods is

352 provided below:

353   (1) PR method

354   The PR method assumes that the error in budget components is proportional to

355 their magnitudes (Abatzoglou et al., 2018). Based on the relative magnitudes of these

356 variables, the ΔRes error is redistributed across them to achieve water budget closure

357 (Equation 6).

$$F_i = X_i - \Delta Res(G_i)\left(\frac{|X_i|}{\sum_{j=1}^{n}|X_j|}\right) \tag{6}$$

359 where F_i and X_i represent the corrected and original data for budget components (P,

360 ET, Q and TWSC), respectively; n denotes the number of budget components

361 involved in the water budget closure calculation; ΔRes represents the water

362 imbalance error; G is a constant vector defined as $G = [1\ -1\ -1\ -1]$.

363   (2) CKF method

364   The CKF method is developed based on the Kalman filter method (Pan and wood,

365 2006). For a given set of estimated budget components $X = [P\ ET\ Q\ TWSC]^T$ and

366 their estimated errors $\Delta Res = GX \neq 0$ (where G is a constant vector, $G = [1 \ -1 \ -$

367 $1 \ -1]$), the goal is to find a new set of estimates $F = [P' \ ET' \ Q' \ TWSC']^T$

368 such that $GX' = 0$, achieving water budget closure (Pan et al., 2012). In simple

369 terms, the CKF method redistributes the ΔRes among the budget components based

370 on the error covariance of X, defined as $\Delta \varepsilon_{XX}$ (Equation 7), to obtain a closured

371 dataset.

$$\Delta \varepsilon_{XX} = \overline{(X - X_0)(X - X_0)^T} \tag{7}$$

373 where X_0 refers to the reference values of the estimated budget components,

374 and the bar over an expression denotes expectation. For P, ET and TWSC, the

375 reference values X_0 were calculated by averaging all considered datasets, following

376 previous studies (Zhang et al., 2018; Abolafia-Rosenzweig et al., 2021). For Q, we

377 adopted observed Q. Due to the difficulty in quantifying the uncertainty in observed Q,

378 previous studies have reported gauge-based uncertainty as a percent error for some of

379 the basins, ranging from 2.3%–28.8% (Clarke, 1999; Mueller, 2003; Shiklomanov et

380 al., 2006; Abolafia-Rosenzweig et al., 2021). We followed a similar approach to

381 estimate the uncertainty associated with Q in this study.

382 The error covariance matrix $\Delta \varepsilon_{XX}$ is of dimension 4×4 and represents the

383 covariances among errors in the four budget components:

$$\Delta \varepsilon_{XX} = \begin{bmatrix} \Delta \varepsilon_{P-P} & \Delta \varepsilon_{P-ET} & \Delta \varepsilon_{P-Q} & \Delta \varepsilon_{P-TWSC} \\ \Delta \varepsilon_{ET-P} & \Delta \varepsilon_{ET-ET} & \Delta \varepsilon_{ET-Q} & \Delta \varepsilon_{ET-TWSC} \\ \Delta \varepsilon_{Q-P} & \Delta \varepsilon_{Q-ET} & \Delta \varepsilon_{Q-Q} & \Delta \varepsilon_{Q-TWSC} \\ \Delta \varepsilon_{TWSC-P} & \Delta \varepsilon_{TWSC-ET} & \Delta \varepsilon_{TWSC-Q} & \Delta \varepsilon_{TWSC-TWSC} \end{bmatrix} \tag{8}$$

385 Following Pan et al. (2012), the off-diagonal elements representing

386 cross-variable error covariances were assumed to be zero, under the assumption that

387 errors among different budget components are uncorrelated. Accordingly, the matrix F

388 can be computed as shown in Equation 9.

$$F = X + K(0 - GX) \tag{9}$$

390 where $K = \Delta\varepsilon_{XX}C^T(C\Delta\varepsilon_{XX}C^T)^{-1}$ is the Kalman gain. Setting $GX = \Delta Res$, and

391 Equation 9 can be rewritten as Equation 10.

$$F = X - \Delta\varepsilon_{XX}G^T(G\Delta\varepsilon_{XX}G^T)^{-1}\Delta Res \tag{10}$$

393 where error covariance ε_{XX} is calculated entry by entry according to Equation 8.

394   (3) MCL method

395   The MCL method is an extension of the triple collocation (TC) method. It

396 calculates the weights for redistributing the ΔRes error among budget components by

397 estimating the errors relative to their true values (expressed as distances, without

398 requiring knowledge of the true values). The fundamental equations of the MCL

399 method are shown in Equations 11-12.

$$F_i = X_i - \Delta Res(G^i)(d^i_{xx_0-norm}) \tag{11}$$

$$d^i_{xx_0-norm} = \frac{d^i_{xx_0}}{\sum_{j=1}^{4}|d^j_{xx_0}|} \tag{12}$$

402   In these equations, F_i represents the corrected data for the i-th budget

403 component; X_i denotes the original data for the i-th budget component; ΔRes

404 represents the water imbalance error; $d^i_{xx_0-norm}$ represents the weight assigned to

405 the i-th budget component, and $d^i_{xx_0}$ represents the distance between the i-th

406 budget component and the true value, as calculated using the Monte Carlo (MC)

407 method. For example, in the case of five precipitation data products ($N = 5$), the

408 calculation of $d^i_{xx_0}$ ($d1t$, $d2t$, $d3t$, $d4t$, and $d5t$) is shown in Equations 13-14.

409
$$A_{(N)}y_{(N)} = b_{(N)} \tag{13}$$

410
$$A_{(5)}=\begin{bmatrix}1 & 1 & 0 & 0 & 0\\1 & 0 & 1 & 0 & 0\\1 & 0 & 0 & 1 & 0\\1 & 0 & 0 & 0 & 1\\0 & 1 & 1 & 0 & 0\\0 & 1 & 0 & 1 & 0\\0 & 1 & 0 & 0 & 1\\0 & 0 & 1 & 1 & 0\\0 & 0 & 1 & 0 & 1\\0 & 0 & 0 & 1 & 1\end{bmatrix},\ y_{(5)}=\begin{bmatrix}d_{1t}^2\\d_{2t}^2\\d_{3t}^2\\d_{4t}^2\\d_{5t}^2\end{bmatrix},\ b_{(5)}=\begin{bmatrix}d_{12}^2\\d_{13}^2\\d_{14}^2\\d_{15}^2\\d_{23}^2\\d_{24}^2\\d_{25}^2\\d_{34}^2\\d_{35}^2\\d_{45}^2\end{bmatrix} \tag{14}$$

411 (4) MSD method

412 The MSD method redistributes the ΔRes to each budget component based on

413 minimizing the time-series deviation error, aiming to reduce model uncertainties

414 caused by errors in estimating time-point deviations (Luo et al., 2023). Specifically,

415 the MSD method first calculates the minimum time-series deviation distance between

416 remote sensing data for budget components and multi-source integrated data products

417 (EO) (Equation 15).

418
$$D_{x,\to n} = -\frac{\left[\sum_{j=1}^n (y_{(EO,j)} - \overline{y_{(EO,\to n)}})(x_{(RS,j)} - \overline{x_{(RS,\to n)}})\right]^2}{\sum_{j=1}^n (x_{(RS,j)} - \overline{x_{(RS,\to n)}})^2} + \sum_{j=1}^n (y_{(EO,j)} - \overline{y_{(EO,\to n)}})^2 \tag{15}$$

419 where $D_{x,\to n}$ represents the minimum time-series deviation distance for budget

420 component x (e.g., P, ET, TWSC); $y_{(EO,j)}$ and $x_{(RS,j)}$ refer to the integrated value

421 and raw value of the budget component x, respectively; $\overline{y_{(EO,\to n)}}$ and $\overline{x_{(RS,\to n)}}$

422 denote the average deviation of budget component x from the first to the n-th time

423 point.

424 Next, the MSD method calculates the weights for each budget component based

425 on $D_{x,\to n}$ (Equation 16).

426
$$w_{x,j} = \frac{D_{x,\to j}}{\sum_{i=1}^{4} D_{i,\to j}} \tag{16}$$

427 where $w_{x,j}$ is the weight of budget component x at time point j.

428 Finally, the weight calculation results from Equation 16 are substituted into

429 Equation 17 to achieve water budget closure.

430
$$\begin{bmatrix} F_{P,j}^{BCC} \\ F_{ET,j}^{BCC} \\ F_{R,j}^{BCC} \\ F_{TWSC,j}^{BCC} \end{bmatrix} = \begin{bmatrix} F_{P,j}^{Raw} \\ F_{ET,j}^{Raw} \\ F_{R,j}^{Raw} \\ F_{TWSC,j}^{Raw} \end{bmatrix} - \Delta Res \begin{bmatrix} 1 \\ -1 \\ -1 \\ -1 \end{bmatrix} \begin{bmatrix} w_{P,j} \\ w_{ET,j} \\ w_{R,j} \\ w_{TWSC,j} \end{bmatrix} \tag{17}$$

431 where F^{BCC} represents the budget components (P, ET, Q, and TWSC) corrected for

432 water budget closure, while F^{Raw} denotes the raw, uncorrected values of the budget

433 components.

3.3 Uncertainties introduced by existing BCC methods for closing water budget

435 When the existing BCC methods described in Section 3.2 are applied to close the

436 water budget, they redistribute ΔRes based on the estimated errors of budget

437 components but neglect unmeasured components. This inevitably leads to an

438 unreasonable redistribution of the ΔRes error, introducing new uncertainties. The

439 magnitude of these introduced errors and whether they can be ignored remain

440 unresolved, primarily due to insufficient observational data for some budget

441 components, making it difficult to quantify the associated uncertainties.

442 Our analysis in this study reveals that when existing BCC methods are used for

443 water budget closure, certain budget components that typically have positive values,

444 such as P, ET, and Q, occasionally become negative in some months. Previous studies

445 have also mentioned this issue (Lehmann et al., 2022). This clearly indicates an

446 unreasonable redistribution of ΔRes errors, underscoring the urgent need for

447   methodological improvements. Despite this issue, research on negative values remains

448   limited. Key questions persist regarding the proportion of negative values in each

449   budget component under current BCC methods, which variables are most susceptible

450   to severe negative values, and how these errors vary throughout the year. Addressing

451   these questions is critical for refining existing BCC methods.

452   Notably, quantifying negative values does not require observational data. To

453   comprehensively assess the uncertainties introduced by forced water budget closure,

454   we consider three aspects: errors of individual budget components relative to observed

455   values (Section 4.2.1), negative values arising from budget closure (Section 4.2.2),

456   and ensemble errors (Section 4.2.3).

457   (1) Errors of individual budget components

458   Quantifying this type of error requires determining reference values for budget

459   components. However, for certain variables, such as ET, observational data are

460   insufficient across global watersheds, posing a major challenge in accurately

461   characterizing global ET patterns. As a result, approximate reference values must be

462   used to ensure the reliability of the results.

463   In this study, reference values for budget components were established based on

464   the following principles. For Q, long-term observational records from hydrological

465   stations were available for all selected basins, meeting the study's requirements. For

466   TWSC, we utilized three observational datasets from the GRACE satellite, which

467   currently provides the only large-scale measurements of basin water storage changes

468   under rigorous quality control. The reliability of GRACE data has been validated

through ground-based observations (Famiglietti et al., 2011; Landerer et al., 2020;

Rodell et al., 2009; Tapley et al., 2004; Yeh et al., 2006). Thus, GRACE TWSC data

can be considered approximately reliable. To further enhance its accuracy, we applied

data fusion techniques, as described in Equation 18, to merge the three GRACE

TWSC products into a single reference dataset (Munier & Aires, 2018; Zhang et al.,

2018).

The uncertainty introduced by existing BCC methods for precipitation was

evaluated from two perspectives. First, 13 basins with sufficient observational

precipitation data were selected, using observed precipitation as the reference. This

sample size was sufficient for assessing the uncertainties associated with existing

BCC methods. Second, 71 additional basins lacking sufficient observational

precipitation data were included, for which fused precipitation values, derived using

Equation 18, served as reference. This approach enabled cross-validation of the

reliability of the fused dataset by comparing results with those from basins with

observational data, allowing the study to be extended to a larger number of basins.

ET is the most challenging budget component to measure directly. The scarcity

of globally available ET observational data precludes the direct use of observed ET as

a reference. To address this limitation, previous studies have either focused on a few

basins with available observational data or compared multiple existing ET datasets.

ET products are generally considered reliable if their magnitudes and trends align

with those of other datasets (Chen et al., 2021; Pan et al., 2020; Xu et al., 2019). Some

studies have also employed the fusion of multiple data products as a reference for ET

491 validation (Jiménez et al., 2018; Mueller et al., 2011; Yao et al., 2014). Following this

492 approach, we assessed the uncertainty introduced by existing BCC methods for ET by

493 utilizing a fusion-based reference dataset.

$$\overline{M_x} = \sum_{i=1}^{n} M_{x,i} * \omega_i \ and \ \omega_i = \frac{1}{\sigma_i^2} / \sum_{i=1}^{1} \frac{1}{\sigma_i^2} \tag{18}$$

494

495 where $\overline{M_x}$ represents the fused value of the budget component, $M_{x,i}$ denotes the

496 i-th product of the budget component; ω_i denotes the weight of the i-th product, and

497 σ_i^2 refers to the covariance error of the i-th product, n is the total number of budget

498 components, and x refers to P, ET or TWSC.

499 After establishing reference values for budget components, we quantify errors in

500 the original data relative to these references, using the positive metric CC and inverse

501 metric RMSE as examples, denoted as CC_1 and $RMSE_1$, respectively. Similarly, errors

502 in the BCC-corrected data relative to the reference values are calculated, represented

503 as CC_2 and $RMSE_2$.

504 To assess the uncertainties introduced by water budget closure, changes in CC

505 and RMSE (CC′ and RMSE′) are computed using Equations 19 and 22. Positive

506 values of CC′ and RMSE′ indicate an improvement in data accuracy following BCC

507 correction, whereas negative values suggest a decline. In addition to CC and RMSE,

508 other statistical metrics used in this study include the positive indicator NSE and the

509 negative indicator MAE.

$$\text{CC}' = CC_2 - CC_1 \tag{19}$$

510

$$NSE' = NSE_2 - NSE_1 \tag{20}$$

511

$$MAE' = MAE_1 - MAE_2 \tag{21}$$

512

$$RMSE' = RMSE_1 - RMSE_2 \tag{22}$$

$$CC = \frac{\sum_{i=1}^{n}(Obs_i - \overline{Obs})(Sim_i - \overline{Sim})}{\sqrt{\sum_{i=1}^{n}(Obs_i - \overline{Obs})^2}\sqrt{\sum_{i=1}^{n}(Sim_i - \overline{Sim})^2}} \tag{23}$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}(Sim_i - Obs_i)^2}{\sum_{i=1}^{n}(Obs_i - \overline{Obs})^2} \tag{24}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Sim_i - Obs_i| \tag{25}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Obs_i - Sim_i)^2} \tag{26}$$

where Obs_i represents the reference value at time i, and Sim_i represents either the original data or the BCC-corrected data. $\overline{Obs}$ and $\overline{Sim}$ represent the mean values of Obs and Sim, respectively, and n is the sample size.

(2) Negative values

Negative values are defined as the issue that arises when the BCC method is used to close the water budget, and the redistributed ΔRes error exceeds the actual values of budget components (P, ET, Q, and TWSC), causing P, ET, and Q to become negative. For TWSC, a negative value occurs when the corrected TWSC has an opposite sign to its raw value. These negative values represent only a subset of the errors introduced during water budget closure but reflect an extreme case of unreasonable ΔRes error redistribution, serving as an indicator of the BCC method's effectiveness.

When a budget component exhibits a negative value, the redistribution of ΔRes errors to other components is significantly affected, reducing the overall accuracy of the corrected datasets. Thus, negative values are a critical factor influencing the performance of existing BCC methods and should be prioritized for improvement. To better understand this issue, we analyze the proportion of negative values for each

535   budget component, their seasonal distribution, and their sensitivity to climatic

536   conditions (i.e., their prevalence in arid versus humid basins). Insights from this

537   analysis were incorporated into the proposed IWE-Res method to address the

538   occurrence of negative values (Section 3.4).

539   (3) Ensembled error of four budget components

540   The aforementioned evaluations (1) and (2) assess errors for individual budget

541   components. To gain a more comprehensive understanding of the uncertainties

542   introduced by water budget closure, we also evaluate the combined error. First, the

543   absolute error (AE) of each budget component is calculated (using P as an example,

544   see Equation 29). Second, the relative absolute error (RAE) is determined for each

545   budget component (Equation 28). Finally, by aggregating the relative errors of

546   individual components, we define the ensembled relative error (Equation 27) to

547   quantify the overall error introduced by BCC methods.

548
$$F(Re) = \frac{1}{n}\sum_{i=1}^{n} \frac{|AE(P')| - |AE(P_{Raw})| + |AE(ET')| - |AE(ET_{Raw})| + |AE(Q')| - |AE(Q_{Raw})| + |AE(TWSC')| - |AE(TWSC_{Raw})|}{P_0 + ET_0 + Q_0 + |TWSC_0|}$$
$$= \frac{1}{n}\sum_{i=1}^{n} \frac{RAE(P) + RAE(ET) + RAE(Q) + RAE(TWSC)}{P_0 + ET_0 + Q_0 + |TWSC_0|}$$

549   (27)

550                    $$RAE(P) = |AE(P')| - |AE(P_{Raw})|$$ (28)

551                         $$AE(P) = |P - P_0|$$ (29)

552   where, $F(Re)$ represents the ensembled relative error, and RAE refers to the relative

553   value of absolute error, with i denoting the month. The subscript "Raw" corresponds

554   to the raw data of the budget components, the subscript 0 represents the observed data,

555   the superscript "'" denotes the BCC-corrected data for the budget components. The

556   degree of alteration induced by the BCC methods for each budget component are

557 defined based on the value of $F(Re)$, and four intervals are established in 5%

558 increments: no significant change [0-5%], minor change (5-10%], moderate change

559 (10-15%], and significant change (>15%).

3.4 Proposed IWE-Res method for closing water budget

561   In this section, we propose the IWE-Res method to identify the optimal balance

562 for redistributing ΔRes, minimizing the sum of the introduced error to budget

563 components and the remaining ΔRes error while reducing the negative values

564 introduced by closing the water budget. Unlike existing BCC methods that fully

565 redistribute the ΔRes term in a single step, the IWE-Res method adopts a gradual,

566 iterative redistribution strategy that allows for more consistent correction. Specifically,

567 the method incrementally allocates fractions of ΔRes to P, ET, Q and TWSC, based on

568 fixed percentage steps and guided by existing BCC weighting schemes. At each

569 iteration, the redistribution process seeks to minimize the combined error—defined as

570 the sum of the induced changes in the water budget components and the remaining

571 unexplained ΔRes. This dual-objective criterion ensures that the method balances

572 error reduction while maintaining hydrological plausibility. Importantly, the approach

573 includes a mechanism to avoid introducing implausible negative values. If, during any

574 iteration, the corrected value of a component becomes negative—violating

575 hydrological constraints such as non-negative precipitation or runoff—further

576 redistribution to that component is halted. Subsequent iterations reallocate the

577 remaining ΔRes among the unaffected components. From a hydrological perspective,

578 this strategy acknowledges that not all of the residuals can be attributed to known

components. Some portion of ΔRes may originate from unmeasured or poorly constrained processes. By partially closing the water budget in a controlled and iterative manner, the IWE-Res method reduces the risk of overcorrecting well-characterized components while better preserving the consistency of the entire budget. The specific steps of the proposed IWE-Res method are as follows:

First, the ΔRes error is calculated using Equation 5 and the original datasets of budget components.

Second, an iterative loop is constructed to compute the errors introduced into budget components during the gradual redistribution of the ΔRes error and to address negative values. To more accurately identify the optimal balance, a step size of 0.1% of ΔRes is used in each iteration in this study. We denote the ΔRes redistributed to budget components in each iteration as x, where $x \in [0, \Delta Res]$.

During each redistribution of ΔRes, two error terms are computed: (1) the remaining ΔRes error, defined as $\Delta Res^* = \Delta Res - x$, and (2) the error introduced to budget components due to the redistribution of the x error, denoted as IWE (Equation 31). When these errors are plotted in a coordinate system, two distinct curves emerge (Fig. 2), each representing a different error relationship. For ΔRes* (Equation 30), Figure 2 shows a fixed, monotonically decreasing linear trend, as 0.1% increments of ΔRes are uniformly redistributed to budget components using existing BCC methods. In contrast, the IWE curve exhibits a non-fixed shape, reflecting the cumulative error introduced to budget components during the redistribution of a portion of ΔRes (Equations 31–32). This variability in the IWE curve arises from the nonlinear

601 relationship between the introduced budget component errors and the reduction in

602 ΔRes error.

$$\Delta Res^* = ax + b = -x + \Delta Res \tag{30}$$

$$IWE = F(\varepsilon_P, \varepsilon_{ET}, \varepsilon_Q, \varepsilon_{TWSC}) = F(x, RAE) \tag{31}$$

$$RAE = \frac{1}{4}\sum_{i=1}^{4}(RAE(P) + RAE(ET) + RAE(Q) + RAE(TWSC)) \tag{32}$$

606 where x represents the portion of ΔRes redistributed to the budget components, with a

607 range from 0 to ΔRes. The terms ε_P, ε_{ET}, ε_Q, ε_{TWSC} represent the errors introduced to P,

608 ET, Q and TWSC, respectively, due to the redistribution of x to the budget

609 components. F(x, RAE) denotes the RAE error calculated by the redistribution of the x

610 error to budget components.

611 During the iterative correction process, if any of the water budget components (P,

612 ET, and Q) becomes negative, the redistribution of water imbalance error to that

613 component is immediately suspended. In subsequent iterations, redistribution is

614 recalculated to ensure that only components with physically meaningful positive

615 values receive the imbalance correction. For example, if ET becomes negative in a

616 given iteration, the imbalance is subsequently redistributed to P, Q, and TWSC only,

617 in accordance with Equation 33. For TWSC, if a sign reversal occurs during iteration

618 (i.e., from positive to negative or vice versa), the redistribution of the water imbalance

619 error to TWSC is suspended in the following iteration.

$$F_i = X_i - x(G_i)(\frac{|\varepsilon_i|}{\sum_{j=1}^{n}|\varepsilon_j|}) \tag{33}$$

621 where F_i denotes the corrected dataset, and X_i denotes the original dataset of budget

622 components. Since ET does not participate in the redistribution of the residual error x

623 based on the example above, the weighting vector is defined as G=[1, 0, −1, −1]. The

624 term ε represents the error in budget components estimated using existing BCC

625 methods, as described in Section 3.2.

626   Third, the IWE-Res curve is plotted (Fig. 2) to provide an intuitive comparison

627 between the introduced budget component errors and the remaining water imbalance

628 error. The error calculation results from Equations (30) and (31) are presented within

629 the same coordinate system.

630   The IWE-Res method is illustrated in Fig. 2 using four curves. The x-axis

631 represents the percentage of water imbalance error redistributed to budget components

632 using existing BCC methods, while the y-axis denotes the percentage of the remaining

633 water imbalance error (ΔRes*) after each iteration. The black dashed line represents

634 the redistributed x-error value among the budget components. The thin blue solid line

635 represents the ΔRes* error curve. Since the sum of redistributed x and remaining

636 ΔRes* equals the total ΔRes error, this curve forms a monotonically decreasing 45°

637 line. The thin green solid line represents the introduced budget component error (IWE)

638 after a given percentage of ΔRes is redistributed (x-axis), with its shape varying

639 depending on the redistribution process (Fig. 2 is illustrative). Initially, when no ΔRes

640 is redistributed (x = 0), the IWE error is zero. As more ΔRes is redistributed (with

641 increasing x values), IWE increases due to the growing uncertainty introduced. The

642 thin red solid line represents the total error, defined as the sum of ΔRes* and IWE

643 after applying BCC methods. This curve varies depending on the redistribution

644 process, and its minimum value identifies the optimal balance where combined ΔRes*

645 and IWE errors are minimized. The intersection of the ΔRes* and IWE curves

646 indicates only the point at which these errors are equal, not necessarily the optimal

647 balance.

648   To determine the optimal redistribution of the water imbalance error, we plot

649 the IWE-Res curve (the green solid line) for each basin, identifying the minimum of

650 the red total error curve. We then analyze its patterns across basins with different

651 characteristics to optimize water budget closure and improve the accuracy of budget

652 component datasets.

653   The IWE error in Fig. 2 also serves as a metric for evaluating the performance

654 of existing BCC methods. If a BCC method perfectly redistributed ΔRes without

655 introducing additional errors, the IWE curve would be a flat line at zero, and the red

656 total error line would coincide with the blue ΔRes* error line. This scenario indicates

657 that full redistribution of water imbalance error achieves the optimal balance,

658 providing indirect validation of the IWE-Res method's effectiveness.

659   Finally, the optimal balance is identified, enabling the generation of a

660 high-precision dataset that improves water budget closure. The optimal balance

661 corresponds to the minimum of the total error curve (IWE + ΔRes*), where the sum

662 of remaining water imbalance error and introduced budget component errors is

663 minimized. Ideally, both ΔRes* and IWE would reach their minimum values

664 simultaneously, meaning minimal error is introduced while fully redistributing ΔRes.

665 However, since this ideal state may not always be achievable, identifying the point

666 where combined error is minimized is essential. This principle defines the proposed

667   IWE-Res method (Fig. 2).

668

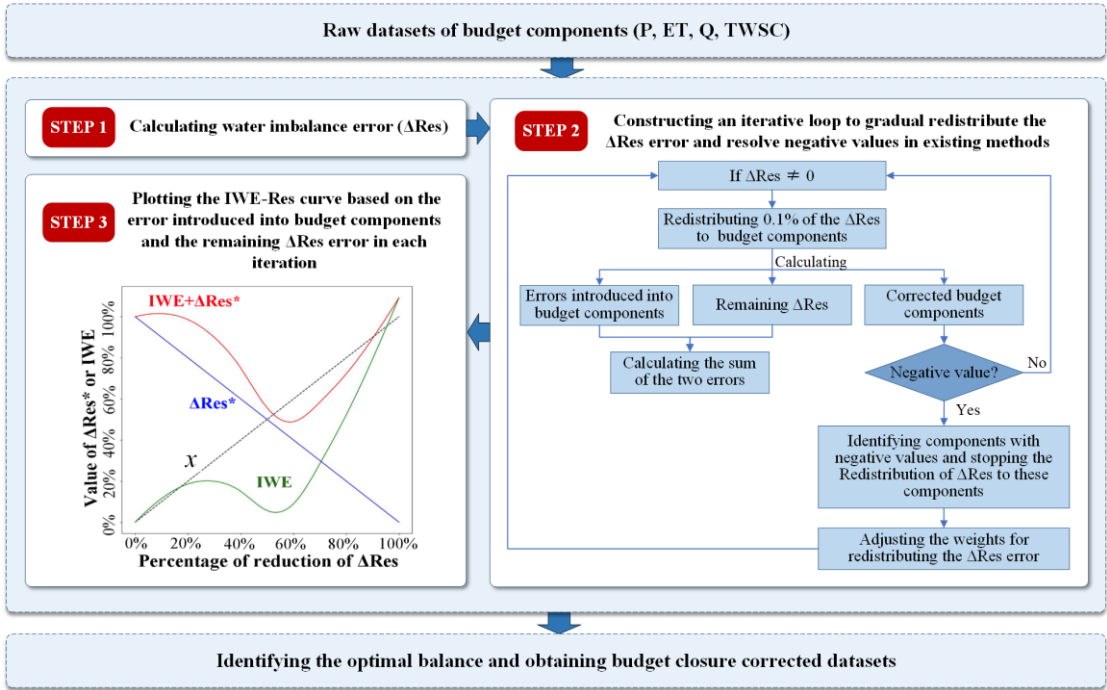



Fig. 2 Framework of the IWE-Res method to identify the optimal balance for redistributing the ΔRes error. The x-axis represents the proportion of ΔRes redistributed to budget components, while the y-axis reflects the proportion of the remaining ΔRes error. The black dashed line represents the redistributed x-error value among the budget components. The blue solid line represents the ΔRes* curve, while the green solid line shows the IWE error introduced into budget components after redistributing the corresponding percentage of ΔRes. The red solid line represents the total error curve.

4 Results

4.1 Water imbalance error

This section presents a comparative analysis of variations in water imbalance errors across different basins and data combinations, aiming to clarify how errors in

682   budget components contribute to these discrepancies. Figure 3 illustrates the spatial

683   distribution of monthly ΔRes errors across various data combinations. To prevent the

684   cancellation of positive and negative values, the absolute values of monthly ΔRes

685   errors were first computed for each basin and then averaged.

686   As shown in Figure 3, ΔRes values vary significantly across basins. Most basins

687   in Africa, South America, and Europe exhibit high ΔRes values, typically exceeding

688   20 mm. In North America, ΔRes values generally range from 15 to 45 mm. Due to

689   inconsistencies among budget component datasets, substantial differences in ΔRes

690   also emerge across different data combinations. In combinations where only P data

691   varied while other budget component datasets remained constant (combinations in Fig.

692   3 where the first digit varies while the second and third remain constant), pronounced

693   changes in water imbalance errors were observed in parts of southern Africa, northern

694   Asia, and North America. This suggests substantial estimation errors in P for these

695   regions.

696

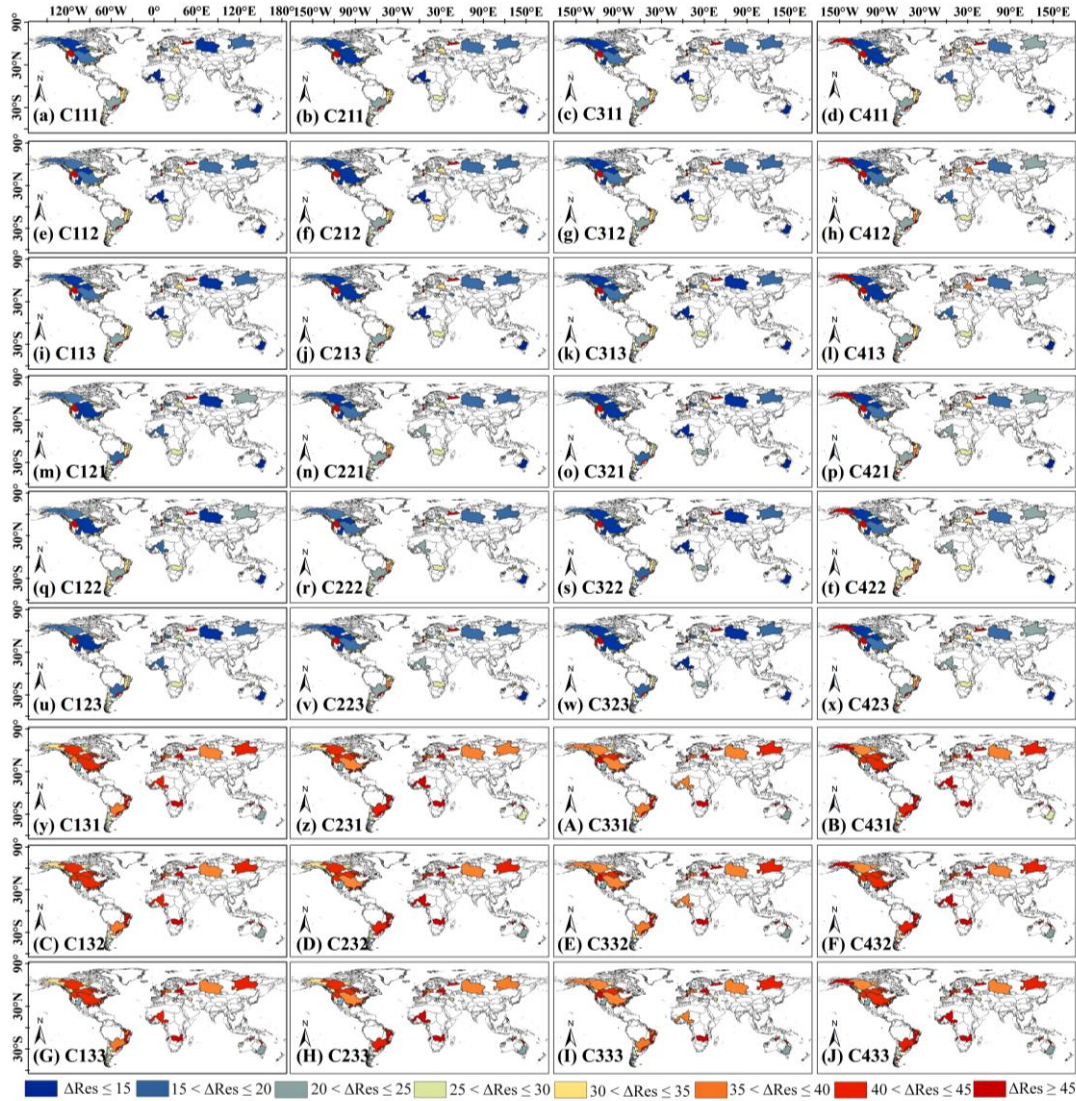



Fig. 3 Spatial distribution of the ΔRes error on a monthly scale for different combinations of budget components. The unit of ΔRes is mm. Each subplot represents a distinct combination, where the first digit corresponds to the P product, the second to the ET product, and the third to the TWSC product. The detailed definitions of these combinations are provided in Equation 3.

When different ET products were used (combinations where the second digit varies while the first and third remain constant), water imbalance errors changed

significantly in most basins. Specifically, in combinations using the TerraClimate ET dataset, water imbalance errors exceeded 35 mm in the majority of basins, indicating severe water imbalance. This underscores the considerable discrepancies among ET products and their substantial impact on accurately representing basin water balance. In contrast, when TWSC data from different GRACE products were used (combinations where the third digit varies while the first and second remain constant), variations in water imbalance errors across basins were relatively small.

Overall, ET and P are the primary variables influencing water imbalance in most basins, consistent with previous findings (Pan et al., 2012; Zhang et al., 2018). The uncertainty in budget component datasets remains a key challenge for water balance research (Dagan et al., 2019; Lv et al., 2017; Luo et al., 2023).

4.2 Uncertainties of budget components introduced by closing water budget

To gain a more comprehensive understanding of the uncertainties introduced into budget components when closing the water budget, this section analyzes the errors introduced by fully closing the water budget using existing BCC methods from three perspectives: the errors of individual budget components, the occurrence of negative values, and ensemble errors (Section 3.3).

4.2.1 Errors of individual budget components

Figure 4 presents the relative statistical metrics calculated using Equations 19–22 to evaluate the uncertainties introduced into budget components by existing BCC methods. Positive values indicate an improvement in the accuracy of corrected budget components, whereas negative values indicate a decline in accuracy.

727   Overall, existing BCC methods exhibit notable limitations in enhancing the

728   accuracy of budget components. In particular, for P, nearly all statistical metrics (CC',

729   NSE', MAE', RMSE') across various basins yield negative values. For instance, under

730   the CKF method, these values are approximately –0.05, –0.15, –3.82 mm, and –8.47

731   mm, respectively, indicating a significant reduction in the accuracy of the corrected P

732   dataset when BCC methods are applied to enforce water budget closure. Specifically,

733   the accuracy of the corrected P dataset decreases by approximately 6%, 34%, 11%,

734   and 55%, as reflected in the CC, NSE, MAE, and RMSE metrics, respectively.

735   Analysis of 13 selected basins with sufficient P observations further confirms this

736   decline, showing a reduction in the accuracy of budget-corrected P (Figure 5). A

737   possible explanation for this decrease is the inherently high accuracy of raw P datasets,

738   supported by advancements in remote sensing technologies, meteorological models,

739   and observational networks. However, when BCC methods are applied, water

740   imbalance errors from other budget components, such as ET, may be inappropriately

741   redistributed to the corrected P dataset in an effort to enforce overall water budget

742   closure. As a result, while the total water budget is balanced, the accuracy of the

743   corrected P data is compromised.

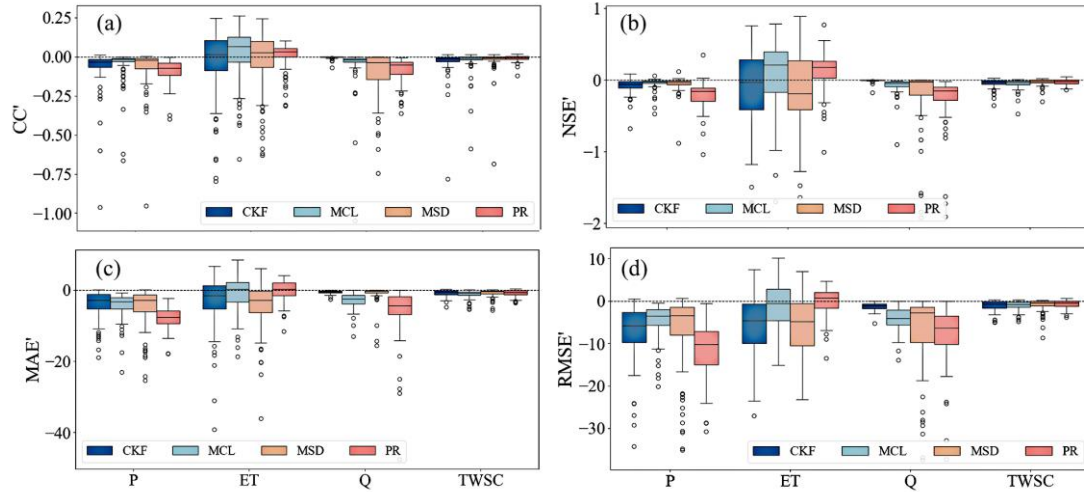
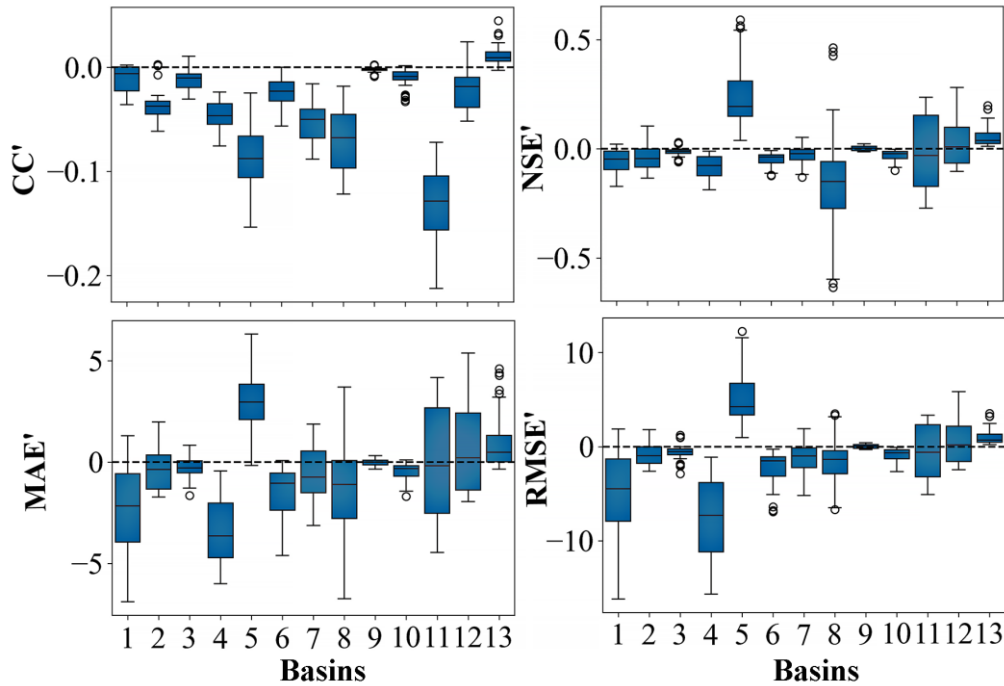



Fig. 4 Box plot quantifying the errors introduced into budget components by existing BCC methods when closing the water budget. (a) - (d) represent the results of the CC', NSE', MAE', RMSE' indicators, respectively. Positive values indicate an improvement in accuracy relative to the reference values after applying existing BCC methods, while negative values indicate a decline. Different colors represent different BCC methods.

The impact of enforcing water budget closure using existing BCC methods on ET was particularly significant (Fig. 4), with approximately 50% of basins exhibiting improved accuracy in corrected ET. For TWSC, most basins showed decreased accuracy. For Q, CC' and NSE' values ranged from 0 to -0.5, while MAE' and RMSE' were primarily concentrated between 0 mm and -20 mm. Consequently, the accuracy of corrected Q declined, with CC, NSE, MAE, and RMSE decreasing by approximately 0.1, 0.2, 3 mm, and 5 mm, respectively. These findings indicate that while redistributing the entire ΔRes enhances the consistency of budget components, it provides limited improvement in their accuracy and may even introduce further

761 errors. Identifying an optimal redistribution strategy for ΔRes errors could help

762 mitigate this issue.

763

764 **Fig. 5** Box plot illustrating precipitation errors introduced by correcting ΔRes using

765 existing BCC methods across 13 basins with sufficient observational precipitation

766 data. The x-axis represents the 13 basins in the following order: NIGER, OB,

767 MISSISSIPPI, SACRAMENTO, SAN JOAQUIN, SUSQUEHANNA, BRAZOS,

768 FRASER, NELSON, MURRAY, RIO EBRO, ELBE, and KURA.

769

770 4.2.2 Negative values

771   This section examines the occurrence of negative values in budget components

772 arising from the application of existing BCC methods to close the water budget. For

773 each budget component, the proportion of months with negative values relative to the

774 total time series was computed (Fig. 6). Overall, the fraction of negative values across

775 budget components ranges from 0% to 10%, with the majority falling below 5%. This

proportion is notable, as negative values indicate substantial inaccuracies in the

redistribution of water imbalance errors by existing BCC methods. When a budget

component exhibits a negative value, the accuracy of the remaining budget

components is also compromised. The relatively high occurrence of negative values

highlights the need for methodological improvements to enhance the performance of

existing BCC methods.

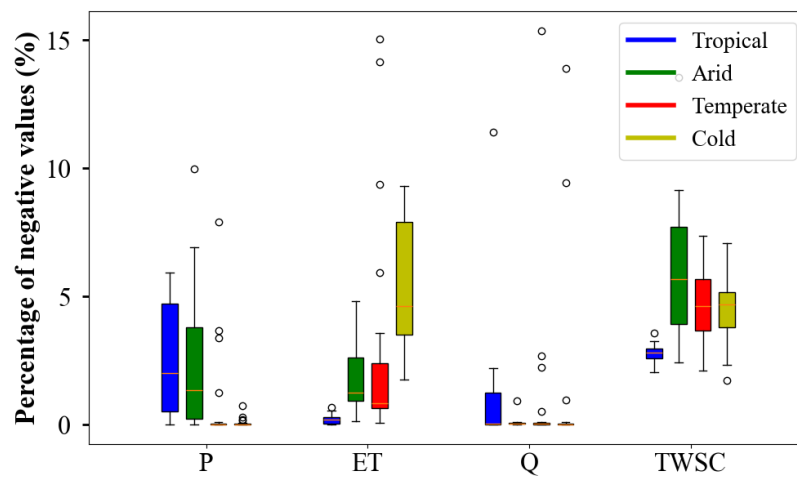



Fig. 6 Percentage of negative values for corrected datasets of budget components

induced by closing the water budget. Different colors indicate distinct climate

classifications.

Among the individual budget components, ET and TWSC exhibit the most

pronounced negative values, followed by P, while Q shows the least (Fig. 6). Notably,

the proportion of negative values in budget components varies significantly across

climate types. For P, negative values generally remain below 5% but can occasionally

reach 7% in arid regions. The likelihood of negative P values is higher in tropical and

arid climates (mostly below 5%) compared with temperate and cold regions (around

1%). For ET, the proportion of negative values is largely below 5%, but it is notably

higher in cold climates (reaching 9%), followed by arid and temperate regions

(approximately 1%–3%). Tropical climates exhibit the lowest proportion of negative

ET values, with most instances below 1%. Q consistently shows a low occurrence of

negative values across all climate types (generally below 3%), with a slightly higher

probability in tropical regions than in other zones. The proportion of negative TWSC

values ranges from 3% to 10%, being lowest in tropical climates (below 5%), while

other climate types exhibit values between 3% and 10%. Previous studies based on

the Budyko framework (ignoring TWSC) at the annual scale have shown that water

balance is primarily governed by P and potential ET (Sankarasubramanian & Vogel,

2002; Zhang et al., 2008; Koster & Suarez, 1999). However, these influences vary

across climatic regions. For example, in tropical and arid regions, P tends to be the

dominant controlling factor (Du et al., 2024; Wu et al., 2018; Liu et al., 2017; Guo et

al., 2022). In cold regions, the Budyko model exhibits relatively limited accuracy in

estimating ET at the annual scale (Lute et al., 2014; Gao et al., 2010; Potter et al.,

2005). These previous findings at the annual scale provide indirect support for our

results derived at the monthly scale.

Fig. 7 presents the seasonal cycle of negative values across different climate

zones, examining whether these values exhibit significant seasonal patterns. Negative

P values predominantly occur in winter and spring, with a higher proportion from

January to March in tropical climates compared to arid regions. ET tends to show

negative values more frequently in winter and spring, with a lower likelihood in

815 summer and autumn. Except in summer, cold climate zones are most susceptible to

816 negative ET values. Among the four budget components, Q has the lowest occurrence

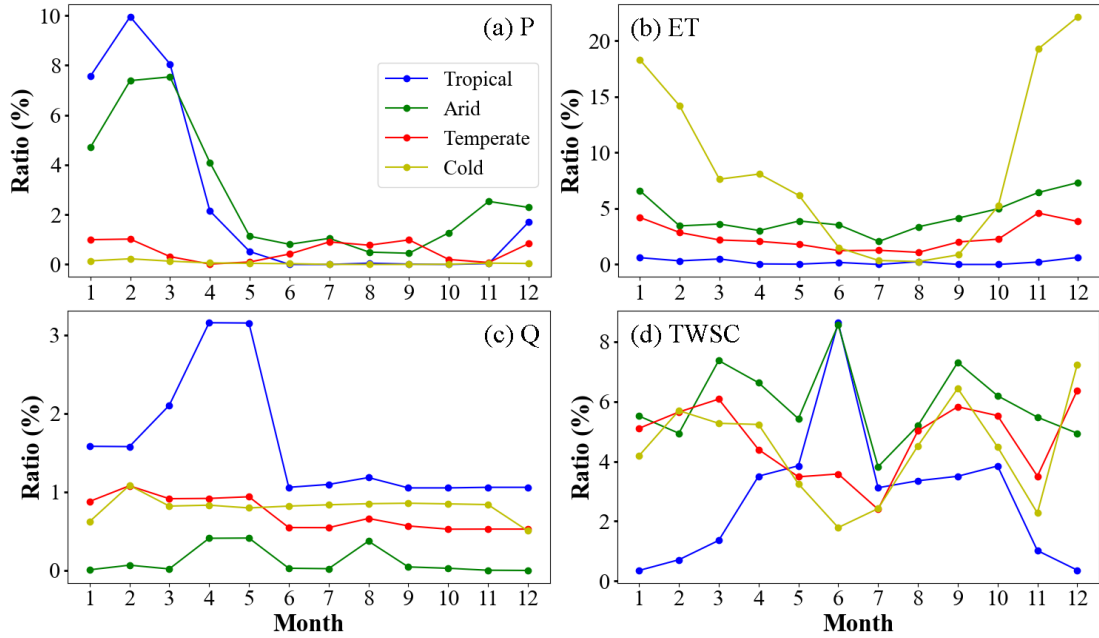
817 of negative values. Negative TWSC values exhibit no obvious seasonal pattern, with

818 arid regions exhibiting a higher likelihood of negative values throughout the year

819 compared to other climate types. These findings indicate that the occurrence of

820 negative values varies significantly across seasons and climate zones. Future research

821 should account for this seasonal variability to further refine existing BCC methods.

822

823 **Fig. 7** Seasonal cycle of the proportion of negative errors for budget components.

824 Different colors representing various climate types. The Southern Hemisphere data by

825 applying a 6-month shift to align its seasonal phases with those of the Northern

826 Hemisphere.

827

828 4.2.3 Ensemble errors

829 Fig. 8 presents the ensemble errors in budget components (i.e., F(Re) in Equation

830 27) introduced by existing BCC methods (CKF, MCL, MSD, and PR). All four

831 methods exhibit similar spatial distribution patterns. Notably, high ensemble errors

832 (F(Re) > 10%) are concentrated in the northwestern basins of North America,

833 particularly in Alaska, suggesting substantial variations in budget components in these

834 regions. Basins with minor ensemble errors (5% < F(Re) ≤ 10%) generally cover

835 larger areas, such as African and Northern Asia. Although these errors are relatively

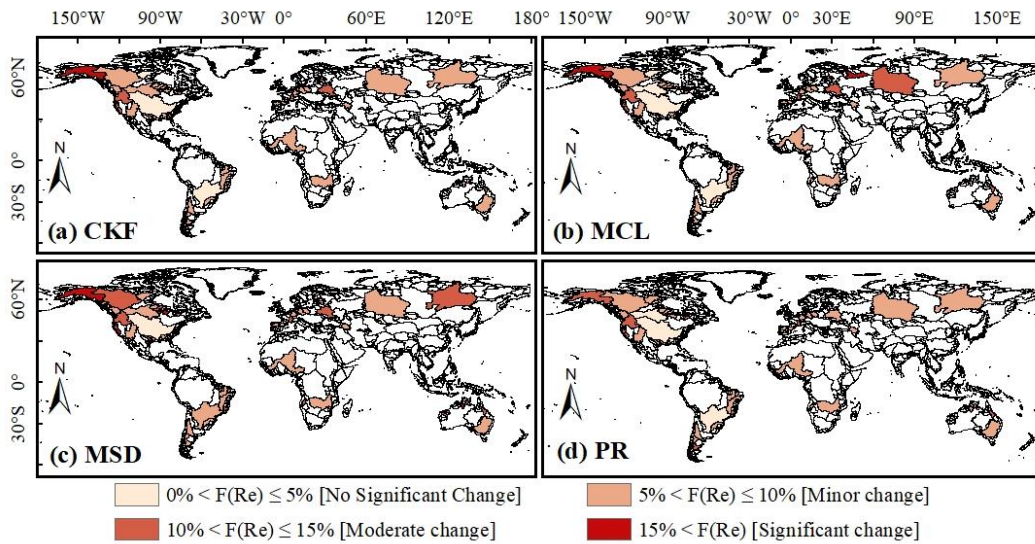
836 small, they remain non-negligible. Basins with lower ensemble errors (F(Re) ≤ 5%)

837 also cover some basins. Further analysis of ΔRes in basins with higher F(Re) values

838 reveals a strong correlation, as these basins also exhibit larger ΔRes. This finding

839 highlights the limitations of existing BCC methods in effectively redistributing ΔRes

840 errors.

841

842 **Fig. 8** Ensemble errors in budget components introduced by closing the water budget

843 using existing BCC methods.

844

845   To determine whether the error cost introduced by existing BCC methods in

846 closing the water budget outweighs the reduction in water imbalance error, we

847 analyzed the relationship between the reduction in ΔRes error and the introduced

848 budget component errors (Fig. 9). As shown in Fig. 9, with the exception of the PR

849 method, the basins where |RAE| exceeds |Res| are largely consistent across the other

850 three BCC methods. This discrepancy arises because the PR method redistributes

851 ΔRes based on the magnitude of budget components, whereas the CKF, MCL, and

852 MSD methods redistribute ΔRes according to the estimated errors in budget

853 components.

854

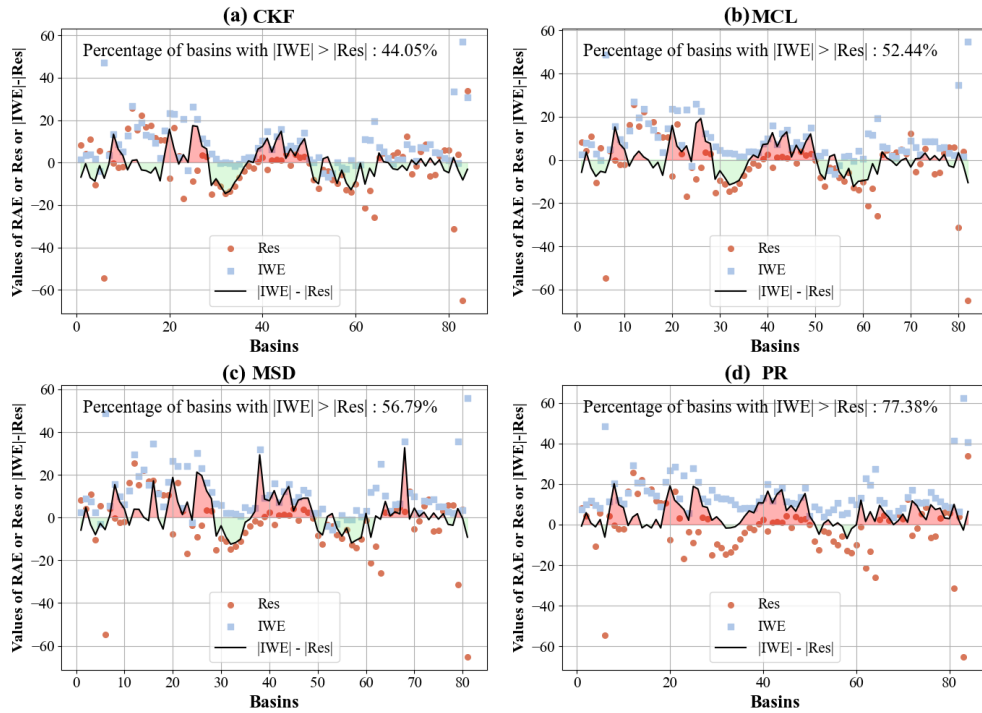



Fig. 9 Comparison of relative absolute error (RAE) and residual error (Res) for four

BCC methods (CKF, MCL, MSD, PR) across various basins. The black lines in the

red shaded area on the upper half of the y-axis indicate that the error introduced by the

BCC methods for budget components exceeds the reduction in ΔRes error (|RAE| >

|Res|), while the green shaded area on the lower half of the y-axis represents cases

860   where the error introduced is less than the reduction in ΔRes error ($|RAE| < |Res|$).

861

862   For the CKF, MCL, MSD, and PR methods, the proportions of basins where

863 |RAE| exceeds |Res| are 44.05%, 52.44%, 56.79%, and 77.38%, respectively. This

864   indicates that, for all four methods, the introduced |RAE| error in budget components

865   surpasses the reduction in water imbalance error in more than 40% of the basins.

866   These findings underscore the non-negligible uncertainties introduced by these

867   methods. Striking a balance between reducing water imbalance error and minimizing

868   the impact of budget component errors remains a critical challenge, motivating us to

869   propose the IWE-Res method to identify optimal balance.

870   *4.3 Verifying the accuracy of the proposed IWE-Res method*

871   Based on the error analysis of existing BCC methods in Section 4.2, this section

872   assesses the accuracy and reliability of the proposed IWE-Res method. The evaluation

873   is conducted through a comparative analysis with PR, CKF, MCL, and MSD, focusing

874   on three key aspects: the errors of individual budget components, the occurrence of

875   negative values, and ensemble errors.

876   Fig. 10 compares the accuracy of the proposed IWE-Res method with existing

877   PR, CKF, MCL, and MSD methods from the perspective of errors in individual

878   budget components. The red and blue lines represent the IWE-Res method and the

879   existing BCC methods, respectively, while the bars indicate the relative accuracy

880   improvement of the IWE-Res method compared to the BCC methods. As shown in

881   Fig. 10, the proposed IWE-Res method exhibits consistently higher accuracy than all

existing CKF, MCL, MSD, and PR methods for budget components P, ET, Q, and TWSC. This result highlights the superior capability of the IWE-Res method in optimizing errors in budget corrected datasets. According to the statistical metrics CC, NSE, MAE, and RMSE, the proposed IWE-Res method improves the corrected P data by 4.2%, 21.3%, 25.5% and 29.5%, respectively, compared to the existing BCC methods. For corrected ET, the improvements are 6.9%, 265.7%, 17.6% and 24.7%, respectively; for corrected Q, the improvements are 3.4%, 185.1%, 67.1%, and 69.0%; and for corrected TWSC, the improvements are 0.0%, 7.0%, 7.5%, and 6.8%.

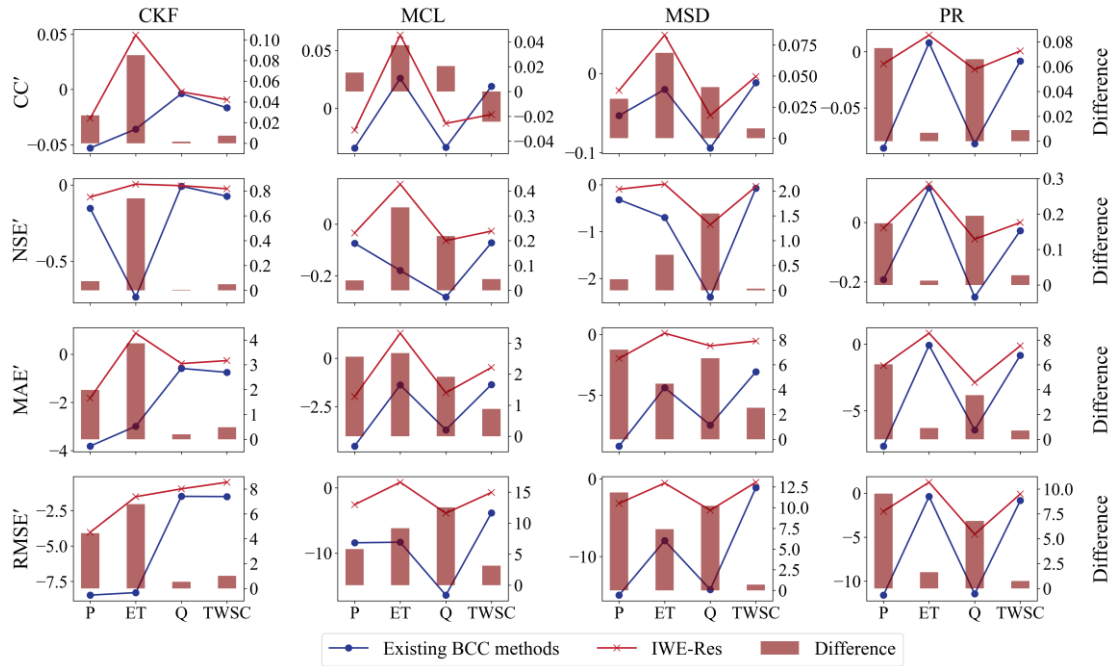



Fig. 10 Performance comparison of the proposed IWE-Res method with existing BCC methods in corrected individual budget components. The red and blue lines in the figure represent the average values across all basins considered in this study.

Table 2 presents the percentage of negative values observed in the corrected budget components for the proposed IWE-Res method and existing BCC methods.

897 One of the key contributions of the IWE-Res method is its ability to address the

898 critical limitation of negative value generation in existing BCC methods. As a result,

899 the percentage of negative values in the corrected P, ET, Q, and TWSC data using the

900 proposed IWE-Res method is zero. In contrast, the corrected P, ET, Q, and TWSC

901 data obtained from existing BCC methods contain negative values to varying degrees

902 (for a detailed analysis of negative values, see Section 4.2.2). These results

903 demonstrate that, in addition to improving the accuracy of budget components relative

904 to observations, the proposed IWE-Res method effectively eliminates the issue of

905 negative values inherent in existing BCC methods.

906 **Table 2** The percentage of months with negative values in the corrected datasets of

907 budget components P, ET, Q, and TWSC for the proposed IWE-Res method and

908 existing BCC methods. The percentages in the table represent the average values

909 across all basins considered in this study.

	P		ET		Q		TWSC	
	Existing	IWE-Res	Existing	IWE-Res	Existing	IWE-Res	Existing	IWE-Res
CKF	0.31%	0%	6.73%	0%	0.75%	0%	4.81%	0%
MCL	1.82%	0%	4.78%	0%	0.77%	0%	7.61%	0%
MSD	1.68%	0%	7.03%	0%	0.82%	0%	5.40%	0%
PR	0%	0%	0.57%	0%	0.72%	0%	0.47%	0%

910

911 We further evaluate the accuracy and reliability of the proposed IWE-Res

912 method using the ensemble error metric defined by Equation (27) (Fig. 11), where

913 lower values indicate better model performance. As shown in Fig. 11, the IWE-Res

914 method significantly reduces ensemble errors compared to existing BCC methods. For

915 instance, in the CKF method, the median ensemble error decreases from above 5% to

below 5%. This reduction is even more pronounced in the MCL, MSD, and PR methods. Additionally, the interquartile ranges under IWE-Res are notably narrower, suggesting improved control over stochastic variability. For example, in the PR method, the interquartile range shrinks from 5–8% (existing BCC methods) to 1–2% (IWE-Res), reflecting an approximate 67% reduction in variability. These findings highlight the robustness of the IWE-Res method in minimizing integrated errors, aligning with its previously demonstrated excellence in single-variable error optimization and negative value elimination.

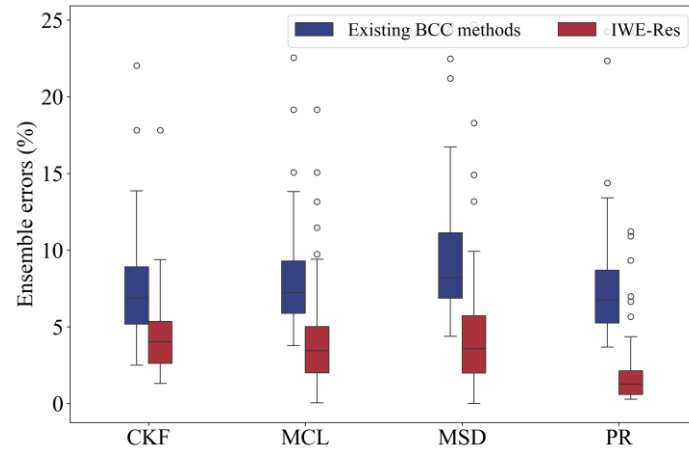



Fig. 11 Performance comparison of the proposed IWE-Res method and existing BCC methods based on the ensemble errors of budget components.

4.4 Identifying the optimal balance for redistributing water imbalance error

Based on the proposed IWE-Res method, this section aims to determine the optimal balance for redistributing water imbalance errors across different climate zones (Tropical, Arid, Temperate, and Cold climate zones) to achieve the best trade-off (Figs. 12-15). Specifically, it seeks to minimize both water imbalance errors and the uncertainties in budget components introduced by enforcing water budget

934 closure. The findings offer a valuable reference for generating high-precision datasets

935 of budget components with a closed water budget in diverse climate regions. When

936 developing the IWE-Res method, we incorporated multiple BCC methods, each based

937 on different principles. As a result, the identified optimal balance results vary across

938 methods. This section presents results for the CKF method only, while results for the

939 MSD, MCL, and PR methods are provided in the supplementary materials.

940 Overall, the optimal balance varied among basins located in different climate

941 zones (Figs. 12–15). In most basins within the Tropical, Arid, and Temperate zones,

942 the optimal balance was achieved when only a portion of the water imbalance error,

943 rather than the entire error, was redistributed to budget components. However, this

944 pattern was not observed in the Cold region.

945 For most basins in the Tropical climate zone (Fig. 12), the optimal balance was

946 reached when 40%–90% of ΔRes was reallocated to budget components, suggesting

947 that the corrected budget datasets achieve their highest accuracy within this range.

948 Notably, approximately 20% of basins attained their optimal balance when 80%–90%

949 of ΔRes was redistributed, while about 70% did so within the 40%–50% range.

950 Therefore, in Tropical basins, if sufficient observational data are unavailable to

951 precisely determine the optimal balance, redistributing 40%–50% of ΔRes to budget

952 components is recommended to obtain the most accurate dataset.

953

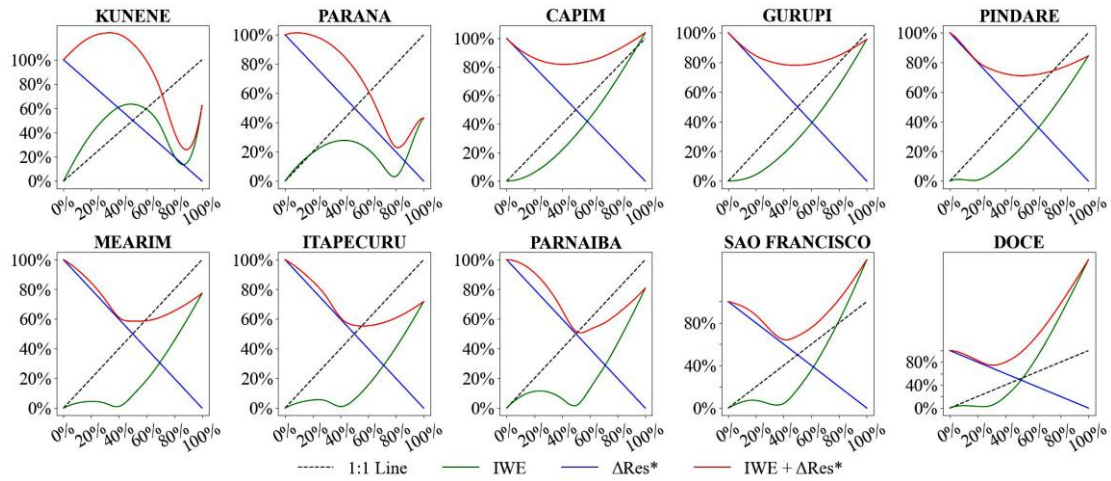



Fig. 12 IWE-Res curve in basins of Tropical climate zone for identifying the optimal

balance that enhances water budget closure and reduces uncertainty.

956

For basins in the Arid climate zone (Fig. 13), optimal balance are generally found

when 40%–90% of ΔRes is redistributed, indicating that the corrected budget

component datasets achieve their highest accuracy within this range. Specifically,

approximately 31% of basins reach their optimal balance at 40%–50% redistribution,

38% at 60%–80%, and over 20% at 90%. Thus, the distribution of optimal balance in

Arid basins does not follow a distinct pattern.

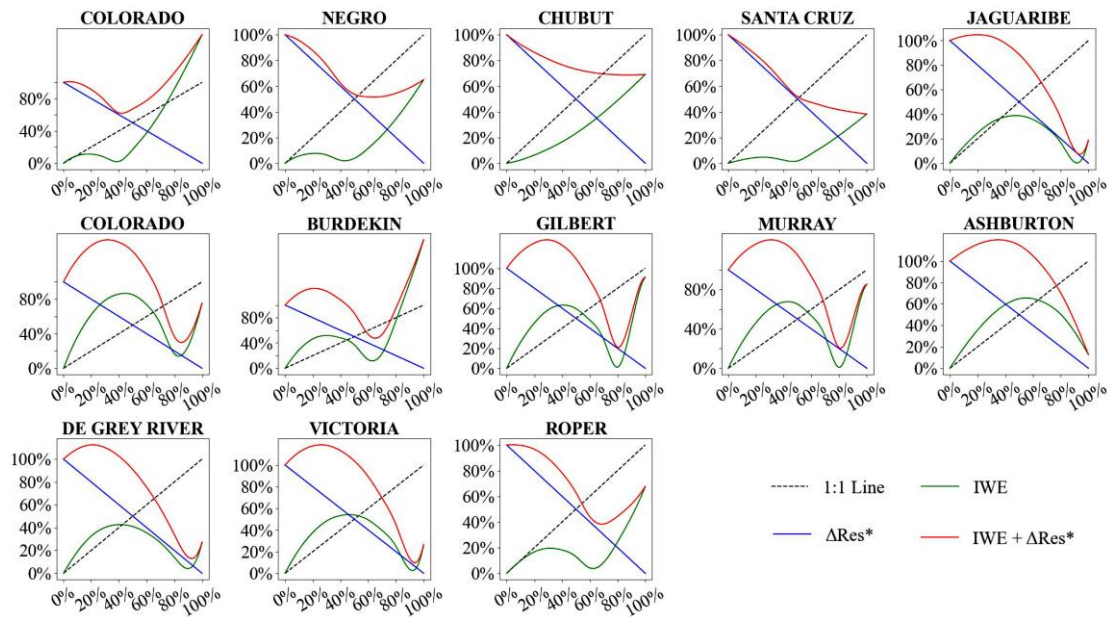



Fig. 13 IWE-Res curve in watersheds of Arid climate zone for identifying the optimal balance that enhances water budget closure and reduces uncertainty.

In the Temperate climate zone (Fig. 14), optimal balance are concentrated within the 40%–90% range. Approximately 53% of basins achieve their optimal balance when 40%–50% of ΔRes is redistributed, while 17% and 13% reach it at 70% and 90% of the ΔRes redistribution. A smaller proportion of basins achieve optimal balance at 60% and 80% of the ΔRes redistribution. Overall, redistributing 40%–50% of ΔRes minimizes the combined error from both the introduced budget component error and the remaining water imbalance error in most basins.

974

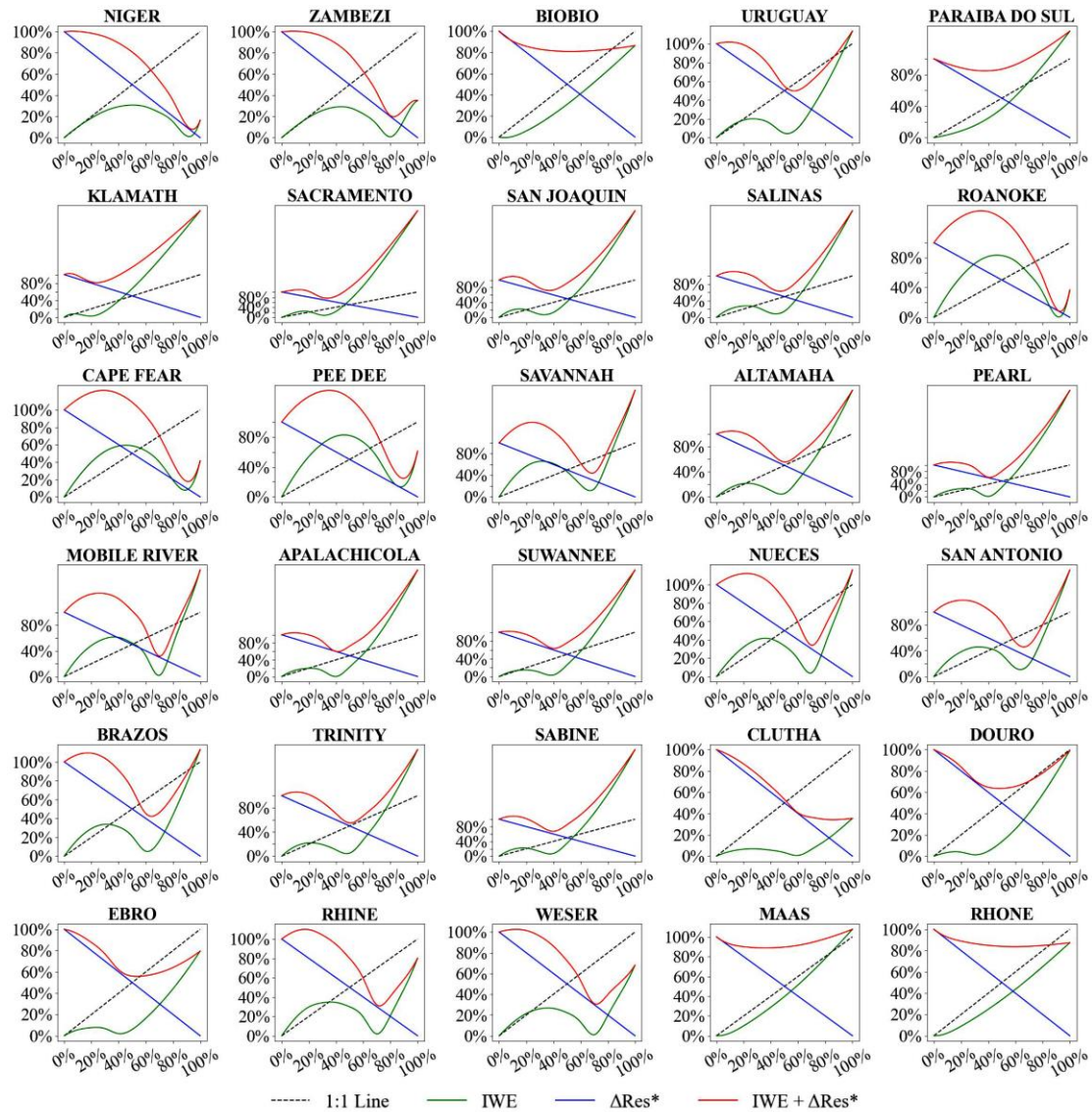



Fig. 14 IWE-Res curve in watersheds of Temperate climate zone for identifying the

optimal balance that enhances water budget closure and reduces uncertainty.

977

In Cold climate zone basins (Fig. 15), the optimal balance is typically reached

when the entire ΔRes is fully redistributed. This suggests that complete redistribution

of ΔRes does not compromise the accuracy of the budget components. This is

primarily due to the trend observed in the IWE curve, which initially

increases—indicating rising error—before decreasing, in contrast to the patterns seen

in most basins in Figs. 12–14. A comparison of ΔRes and the negative values introduced by full redistribution of ΔRes across climate zones reveals that, in Cold regions, negative values predominantly occur in ET. This is likely due to the inherently lower ET values in Cold regions, which increases the likelihood of negative values when ΔRes is redistributed. However, errors introduced in other budget components, such as P and Q, remain relatively low under full redistribution of ΔRes.

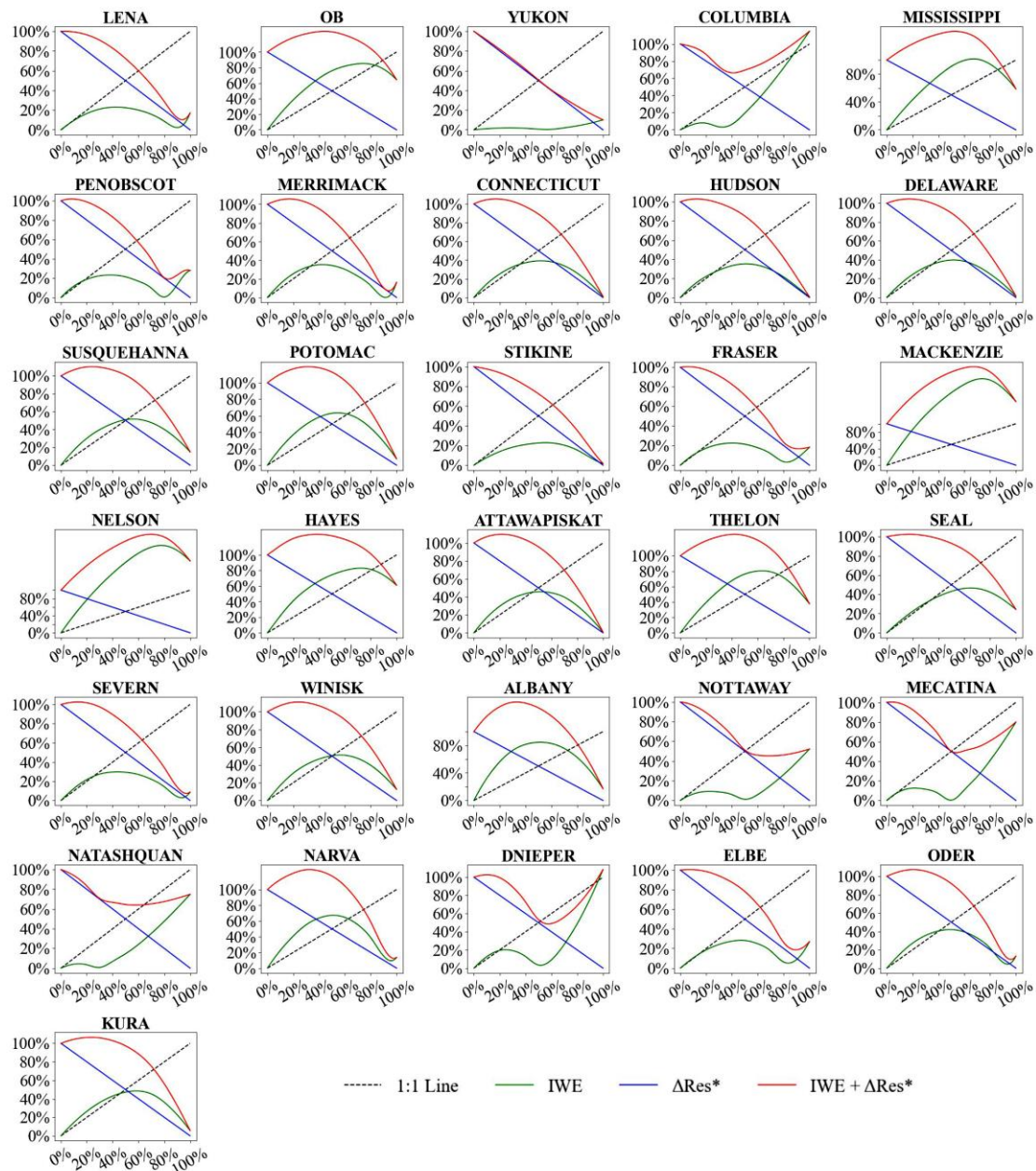



Fig. 15 IWE-Res curve in watersheds of Cold climate zone for identifying the optimal

balance that enhances water budget closure and reduces uncertainty.

5 Discussion

5.1 Uncertainty introduced by existing BCC methods

To quantify the uncertainty introduced by existing BCC methods in closing the water balance, we evaluated four BCC methods across 84 global basins. The assessment focused on errors in individual budget components, occurrences of negative values, and ensemble errors in budget components. Our findings indicate that while existing BCC methods improve the consistency of budget components, their ability to enhance the accuracy of these components is limited and, in some cases, may even reduce it. It is worth noting that the datasets generated by both the existing BCC methods and the IWE-Res method proposed in this study were not further bias-corrected against independent observations. For basin-specific applications requiring higher reliability, we recommend additional bias correction.

Several factors may contribute to this reduction in accuracy. First, most existing BCC methods estimate errors in budget components without incorporating independent observational data. These methods then redistribute water imbalance errors based on these internally estimated uncertainties (Section 3.2). However, the absence of observational constraints undermines the reliability of the estimated component errors, which in turn leads to a suboptimal and potentially biased allocation of the imbalance. As previously noted, inaccurate error estimates for a single variable can propagate through the redistribution process, biasing the residual redistribution to the remaining budget components and ultimately lowering the

accuracy of all water budget components (Abolafia-Rosenzweig et al., 2021). Incorporating high-quality observational data into the error estimation process is therefore essential to improve the robustness of BCC methods; Second, existing BCC methods are limited by their assumption that the entire water imbalance error can be fully attributed to errors in the measured budget components. These methods enforce water budget closure by completely redistributing the water imbalance error among the budget components, yet this residual may also stem from systematic biases and unmeasured processes—not just estimation errors of measured budget components. In this study, we propose an iterative optimization approach that seeks a balanced redistribution of the ΔRes, aiming to minimize both the errors introduced to individual budget components and the remaining ΔRes. This method significantly improves the accuracy of the corrected datasets. Future research may further enhance this framework by integrating it with physically based hydrological or land surface models, which could provide a promising pathway toward more physically consistent and realistic water budget estimates; Third, the observational datasets themselves often fail to strictly satisfy water budget closure due to measurement limitations and sampling errors. This introduces uncertainty when using these datasets to validate the accuracy of BCC-corrected estimates. For instance, even if the corrected datasets more closely approximate the true values of budget components, the lack of ground-truth observations presents a fundamental challenge for objectively evaluating the effectiveness of these corrections. Future work should prioritize the development of more objective and physically grounded evaluation metrics to assess the accuracy

of BCC-corrected datasets. Although this challenge lies beyond the scope of the present study, addressing it will be critical for advancing the reliability of water budget assessments.

5.2 Identification of the optimal balance

Each budget component inherently contains observational or model-based errors. Indiscriminately redistributing water imbalance errors across all budget components to achieve complete water budget closure can introduce additional uncertainties. By identifying the optimal balance for error redistribution across different climate zones, we observed significant variations in distribution patterns. In tropical and temperate regions, most basins achieved their optimal balance when 40%–90% of the water imbalance error was redistributed, with a concentration around the 40%–50% range. In arid regions, the distribution of optimal balance was more dispersed, lacking a clear concentration within any specific redistribution range but generally falling within the 40%–90% range. Cold climate regions exhibited distinct characteristics, with most basins achieving the smallest error when the water imbalance error was fully redistributed.

Overall, optimizing the redistribution ratio of water imbalance errors is critical for improving the accuracy of corrected budget components. However, the sensitivity of these components to error redistribution varies, and both over- and under-correction can propagate new imbalances across the remaining terms, ultimately misrepresenting the underlying hydrological processes. While existing BCC methods estimate redistribution weights based on the relative uncertainty of each

component, future research should examine the physical rationale behind these redistributions. The spatiotemporal variability of residual errors offers valuable insight into their dominant sources, which can serve as an independent reference to validate the influence weights computed by BCC methods. For instance, as shown in previous studies, the contribution of TWSC to residual errors diminishes at annual and especially decadal timescales, where P and ET uncertainties become more dominant. Spatial patterns of residuals also reflect the nature of regional precipitation regimes. In regions dominated by frontal systems, such as temperate zones, remotely sensed precipitation products tend to capture rainfall events more accurately, leading to smaller residuals. In contrast, in areas characterized by convective rainfall—such as the tropics and arid zones—larger residuals are observed, likely due to the higher uncertainty in capturing short-lived and spatially localized storm events.

Notably, the choice of spatial resolution has a significant impact on the results (Aziz et al., 2022; Bormann et al., 2006; Senan et al., 2022). Following many previous studies (Lehmann et al., 2022; Abolafia-Rosenzweig et al., 2023; Luo et al., 2023; Wang et al., 2014; Tan et al., 2022; Sahoo et al., 2011), the BCC method in this study is also applied at the basin scale rather than the grid scale for the following reasons: 1) Achieving water budget closure at the grid scale is complex and challenging due to the difficulty of quantifying all water flux and storage components flowing into and out of the grid, including P, ET, TWSC, lateral inflow and outflow, leakage losses, and human water withdrawals and returns. Several of these components, such as lateral flow and leakage, are poorly observed or highly uncertain, and their omission

introduces substantial error; 2) The datasets of different variables have varying spatial resolutions, and resampling them to a common resolution introduces uncertainties, which in turn affect the accuracy of water budget closure correction; 3) The coarse spatial resolution of GRACE-derived TWSC data limits their applicability for water budget closure calculation at the grid scale. At monthly resolution, TWSC is a critical component and cannot be neglected. Averaging GRACE data to the basin scale helps reduce random errors by offsetting positive and negative biases, thereby increasing the reliability of water budget closure correction; 4) Despite advances in remote sensing and in situ observation networks, grid-scale uncertainties remain substantial for some budget components, such as ET. Basin-scale analysis therefore reduces uncertainty and improves the reliability of water budget closure correction results.

6 Conclusions

Existing BCC methods introduce new uncertainties when closing the water budget due to challenges in accurately estimating errors in budget components and the integrated concept of water imbalance error. This study first evaluates the issues arising from existing BCC methods by comparing the errors introduced in budget components with the improvement in water budget closure precision. A new method, termed IWE-Res, is proposed to identify the optimal redistribution of ΔRes, aiming to minimize the sum of the remaining residual error and the introduced budget component error. To assess the reliability of the IWE-Res method, we compare it with four different BCC methods across 84 basins spanning various global climate zones. The main conclusions are as follows:

1101   (1) While applying existing BCC methods reduces water imbalance error, it

1102 simultaneously introduces new errors in budget components. For P, a decline in

1103 accuracy is observed in most basins. For Q, the corrected data exhibits lower

1104 performance than the raw data, with reductions in CC, NSE, MAE, and RMSE of

1105 approximately 0.1, 0.2, 3 mm, and 5 mm, respectively. At the basin scale, more than

1106 40% of basins experience budget component errors greater than the reduction in ΔRes

1107 after applying existing BCC methods.

1108   (2) The proportion of negative corrected values in each budget component is

1109 predominantly within 0%–5%. For ET, negative corrected values are mostly below

1110 5%, though they reach 9% in cold climate regions. For P, the proportion is primarily

1111 below 5%, with rare occurrences around 7%. Q generally exhibits a lower proportion

1112 of negative values, mostly below 3%. In TWSC, negative values are concentrated

1113 between 3% and 10%.

1114   (3) The proposed IWE-Res method improves the accuracy of corrected budget

1115 components compared to existing BCC methods. Based on RMSE, it improves the

1116 accuracy of corrected P by 29.5%, corrected ET by 24.7%, corrected Q by 69.0%, and

1117 corrected TWSC by 6.8%.

1118   (4) Except in cold regions, redistributing 40%–90% of ΔRes to budget

1119 components yields the optimal balance, minimizing the sum of the remaining ΔRes

1120 and the introduced budget component error. In tropical and temperate regions, the

1121 optimal balance is typically achieved when 40%–50% of ΔRes is redistributed.

1122 Similarly, in arid regions, redistributing 40%–90% of ΔRes effectively reduces errors,

though the optimal redistribution ratio varies across basins. In most cold-region basins,

the total error is minimized when the entire ΔRes is redistributed.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or

personal relationships that could have appeared to influence the work reported in this

paper.

Acknowledgments

This research was supported by the National Natural Science Foundation of

China (No. 42201038); The Belt and Road Special Foundation of the National Key

Laboratory of Water Disaster Prevention (No. 2024nkms02); a research project

funded by China Yangtze Power Co., Ltd. (ZCJSYS-JJ-202506);

IWHR-SKL-KF202301; and China Scholarship Council.

Reference

Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., & Hegewisch, K. C. (2018).
TerraClimate, a high-resolution global dataset of monthly climate and climatic
water balance from 1958–2015. Scientific Data, 5(1), 1-12.
https://doi.org/10.1038/sdata.2017.191

Abhishek, Kinouchi, T., Abolafia-Rosenzweig, R., & Ito, M. (2021). Water budget
closure in the Upper Chao Phraya River basin, Thailand using multisource
data. Remote Sensing, 14(1), 173. https://doi.org/10.3390/rs14010173

Abolafia-Rosenzweig, R., Pan, M., Zeng, J., & 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

Aires, F. (2014). Combining datasets of satellite-retrieved products. Part I: Methodology and water budget closure. Journal of Hydrometeorology, 15(4), 1677-1691. https://doi.org/10.1175/JHM-D-13-0148.1

Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. Fao, Rome, 300(9), D05109.

Aziz, K. M. A., & Rashwan, K. S. (2022). Comparison of different resolutions of six free online DEMs with GPS elevation data on a new 6th of October City, Egypt. Arabian Journal of Geosciences, 15(20), 1585. https://doi.org/10.1007/s12517-022-10845-5

Bai, X., Wu, X., & Wang, P. (2019). Blending long-term satellite-based precipitation data with gauge observations for drought monitoring: Considering effects of different gauge densities. Journal of Hydrology, 577, 124007. https://doi.org/10.1016/j.jhydrol.2019.124007

Bai, X., Wang, P., He, Y., Zhang, Z., & Wu, X. (2021). Assessing the accuracy and drought utility of long-term satellite-based precipitation estimation products using the triple collocation approach. Journal of Hydrology, 603, 127098. https://doi.org/10.1016/j.jhydrol.2021.127098

Beck, H. E., Pan, M., Roy, T., Weedon, G. P., Pappenberger, F., Van Dijk, A. I., . . . Wood, E. F. (2019). Daily evaluation of 26 precipitation datasets using

Stage-IV gauge-radar data for the CONUS. Hydrology and Earth System Sciences, 23(1), 207-224. https://doi.org/10.5194/hess-23-207-2019

Beck, H. E., Vergopolan, N., Pan, M., Levizzani, V., Van Dijk, A. I., Weedon, G. P., . . . Wood, E. F. (2017). Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. Hydrology and Earth System Sciences, 21(12), 6201-6217. https://doi.org/10.5194/hess-21-6201-2017

Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. G., Van Dijk, A. I., . . . Adler, R. F. (2019). MSWEP V2 global 3-hourly 0.1 precipitation: methodology and quantitative assessment. Bulletin of the American Meteorological society, 100(3), 473-500. https://doi.org/10.1175/BAMS-D-17-0138.1

Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., Schamm, K., Schneider, U., & Ziese, M. (2013). A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901–present. Earth System Science Data, 5(1), 71-99. https://doi.org/10.5194/essd-5-71-2013

Burek, P., & Smilovic, M. (2022). The use of GRDC gauging stations for calibrating large-scale hydrological models. Earth System Science Data Discussions, 2022, 1-18. https://doi.org/10.5194/essd-15-5617-2023

Boergens, E., Güntner, A., Sips, M., Schwatke, C., & Dobslaw, H. (2024). Interannual variations of terrestrial water storage in the East African Rift region. Hydrology and Earth System Sciences, 28(20), 4733-4754.

https://doi.org/10.5194/hess-28-4733-2024

Bormann, H. (2006). Impact of spatial data resolution on simulated catchment water balances and model performance of the multi-scale TOPLATS model. Hydrology and Earth System Sciences, 10(2), 165-179. https://doi.org/10.5194/hess-10-165-2006

Bouaziz, L., Weerts, A., Schellekens, J., Sprokkereef, E., Stam, J., Savenije, H., & Hrachowitz, M. (2018). Redressing the balance: quantifying net intercatchment groundwater flows. Hydrology and Earth System Sciences, 22(12), 6415-6434. https://doi.org/10.5194/hess-22-6415-2018

Chen, J., Tapley, B., Rodell, M., Seo, K. W., Wilson, C., Scanlon, B. R., & Pokhrel, Y. (2020). Basin‐scale river runoff estimation from GRACE gravity satellites, climate models, and in situ observations: A case study in the Amazon basin. Water resources research, 56(10), e2020WR028032. https://doi.org/10.1029/2020WR028032

Chen, S., Liu, B., Tan, X., & Wu, Y. (2020). Inter-comparison of spatiotemporal features of precipitation extremes within six daily precipitation products. Climate Dynamics, 54, 1057-1076. https://doi.org/10.1007/s00382-019-05045-z

Chen, X., Su, Z., Ma, Y., Trigo, I., & Gentine, P. (2021). Remote sensing of global daily evapotranspiration based on a surface energy balance method and reanalysis data. Journal of Geophysical Research: Atmospheres, 126(16), e2020JD032873. https://doi.org/10.1029/2020JD032873

Clarke, R. T. (1999). Uncertainty in the estimation of mean annual flood due to rating-curve indefinition. Journal of Hydrology, 222(1-4), 185-190. https://doi.org/10.1016/S0022-1694(99)00097-9

Crosbie, R., Pollock, D., Mpelasoka, F., Barron, O., Charles, S., & Donn, M. (2012). Changes in Köppen-Geiger climate types under a future climate for Australia: hydrological implications. Hydrology and Earth System Sciences, 16(9), 3341-3349. https://doi.org/10.5194/hess-16-3341-2012

Cui, W., Dong, X., Xi, B., Feng, Z. H. E., & Fan, J. (2020). Can the GPM IMERG final product accurately represent MCSs' precipitation characteristics over the central and eastern United States?. Journal of Hydrometeorology, 21(1), 39-57. https://doi.org/10.1175/JHM-D-19-0123.1

Dagan, G., Stier, P., & Watson‑Parris, D. (2019). Analysis of the atmospheric water budget for elucidating the spatial scale of precipitation changes under climate change. Geophysical Research Letters, 46(17-18), 10504-10511.

Dastjerdi, P. A., Ghomlaghi, A., & Nasseri, M. (2024). A new approach to ensemble precipitation Estimation: Coupling satellite hydrological products with backward water balance models in Large-Scale. Journal of Hydrology, 629, 130564. https://doi.org/10.1016/j.jhydrol.2023.130564

Du, H., Zeng, S., Liu, X., & Xia, J. (2024). An improved Budyko framework model incorporating water-carbon relationship for estimating evapotranspiration under climate and vegetation changes. Ecological Indicators, 169, 112887. https://doi.org/10.1016/j.ecolind.2024.112887

Ehret, U., Zehe, E., Wulfmeyer, V., & Liebert, J. (2012). Should we apply bias correction to global and regional climate model data? HESS, 16, 3391–3404. https://doi.org/10.5194/hess-16-3391-2012, 2012.

Esquivel-Arriaga, G., Huber-Sannwald, E., Reyes-Gómez, V. M., Bravo-Peña, L. C., Dávila-Ortiz, R., Martínez-Tagüeña, N., & Velázquez-Zapata, J. A. (2024). Performance Evaluation of Global Precipitation Datasets in Northern Mexico Drylands. Journal of Applied Meteorology and Climatology, 63(12), 1545-1558. https://doi.org/10.1175/JAMC-D-23-0227.1

Famiglietti, J. S., Lo, M., Ho, S. L., Bethune, J., Anderson, K., Syed, T. H., . . . Rodell, M. (2011). Satellites measure recent rates of groundwater depletion in California's Central Valley. Geophysical Research Letters, 38(3). https://doi.org/10.1029/2010GL046442

Gao, H., Tang, Q., Ferguson, C. R., Wood, E. F., & Lettenmaier, D. P. (2010). Estimating the water budget of major US river basins via remote sensing. International Journal of Remote Sensing, 31(14), 3955-3978. https://doi.org/10.1080/01431161.2010.483488

Guo, W., Hong, F., Yang, H., Huang, L., Ma, Y., Zhou, H., & Wang, H. (2022). Quantitative evaluation of runoff variation and its driving forces based on multi-scale separation framework. Journal of Hydrology: Regional Studies, 43, 101183. https://doi.org/10.1016/j.ejrh.2022.101183

Hansford, M. R., Plink-Björklund, P., & Jones, E. R. (2020). Global quantitative analyses of river discharge variability and hydrograph shape with respect to

climate types. Earth-science reviews, 200, 102977. https://doi.org/10.1016/j.earscirev.2019.102977

Hao, X., Zhang, S., Li, W., Duan, W., Fang, G., Zhang, Y., & Guo, B. (2018). The uncertainty of Penman‐Monteith method and the energy balance closure problem. Journal of Geophysical Research: Atmospheres, 123(14), 7433-7443. https://doi.org/10.1029/2018JD028371

He, Q., Fok, H. S., Ferreira, V., Tenzer, R., Ma, Z., & Zhou, H. (2023). Three-dimensional Budyko framework incorporating terrestrial water storage: Unraveling water-energy dynamics, vegetation, and ocean-atmosphere interactions. Science of the Total Environment, 904, 166380. https://doi.org/10.1016/j.scitotenv.2023.166380

Hsu, K.-l., Gao, X., Sorooshian, S., & Gupta, H. V. (1997). Precipitation estimation from remotely sensed information using artificial neural networks. Journal of Applied Meteorology and Climatology, 36(9), 1176-1190. https://doi.org/10.1175/1520-0450(1997)036<1176:PEFRSI>2.0.CO;2

Hua, D., Hao, X., Zhang, Y., & Qin, J. (2020). Uncertainty assessment of potential evapotranspiration in arid areas, as estimated by the Penman-Monteith method. Journal of Arid Land, 12, 166-180. https://doi.org/10.1007/s40333-020-0093-7

Huang, C., Hu, J., Chen, S., Zhang, A., Liang, Z., Tong, X., ... & Zhang, Z. (2019). How well can IMERG products capture typhoon extreme precipitation events over southern China?. Remote Sensing, 11(1), 70. https://doi.org/10.3390/rs11010070

1277 Huang, P., Wang, G., Guo, L., Mello, C. R., Li, K., Ma, J., & Sun, S. (2023). Most

1278   global gauging stations present biased estimations of total catchment discharge.

1279   Geophysical   Research   Letters,   50(15),   e2023GL104253.

1280   https://doi.org/10.1029/2023GL104253

1281 Huffman, G. J., Bolvin, D. T., Braithwaite, D., Hsu, K., Joyce, R., Xie, P., & Yoo,

1282   S.-H. (2015). NASA global precipitation measurement (GPM) integrated

1283   multi-satellite retrievals for GPM (IMERG). Algorithm theoretical basis

1284   document (ATBD) version, 4(26), 2020-2005.

1285 Jacobs, A. F. G., & De Bruin, H. A. R. (1998). Makkink's equation for

1286   evapotranspiration applied to unstressed maize. Hydrological processes, 12(7),

1287   1063-1066.https://doi.org/10.1002/(SICI)1099-1085(19980615)12:7<1063::AI

1288   D-HYP640>3.0.CO;2-2

1289 Jiménez, C., Martens, B., Miralles, D. M., Fisher, J. B., Beck, H. E., &

1290   Fernández-Prieto, D. (2018). Exploring the merging of the global land

1291   evaporation WACMOS-ET products based on local tower measurements.

1292   Hydrology   and   Earth   System   Sciences,   22(8),   4513-4533.

1293   https://doi.org/10.5194/HESS-22-4513-2018

1294 Jian, J., Ryu, D., Costelloe, J. F., & Su, C. H. (2015). Towards reliable hydrological

1295   model calibrations with river level measurements. In 21st International

1296   Congress on Modelling and Simulation, Modelling and Simulation Society of

1297   Australia and New Zealand.

1298 Kaprom, C., Williams, J. A., Mehrotra, R., Ophaphaibun, C., & Sriwongsitanon, N.

(2025). A comprehensive evaluation of the accuracy of satellite-based precipitation estimates over Thailand. Journal of Hydrology: Regional Studies, 59, 102380. https://doi.org/10.1016/j.ejrh.2025.102380

Koster, R. D., & Suarez, M. J. (1999). A simple framework for examining the interannual variability of land surface moisture fluxes. Journal of Climate, 12(7), 1911-1917. https://doi.org/10.1175/1520-0442(1999)012<1911:ASFFET>2.0.CO;2

Krabbenhoft, C. A., Allen, G. H., Lin, P., Godsey, S. E., Allen, D. C., Burrows, R. M., ... & Olden, J. D. (2022). Assessing placement bias of the global river gauge network. Nature Sustainability, 5(7), 586-592. https://doi.org/10.1038/s41893-022-00873-0

Landerer, F. W., & Swenson, S. C. (2012). Accuracy of scaled GRACE terrestrial water storage estimates. Water resources research, 48(4). https://doi.org/10.1029/2011WR011453

Landerer, F. W., Flechtner, F. M., Save, H., Webb, F. H., Bandikova, T., Bertiger, W. I., . . . Dobslaw, H. (2020). Extending the global mass change data record: GRACE Follow‐On instrument and science data performance. Geophysical Research Letters, 47(12), e2020GL088306. https://doi.org/10.1029/2020GL088306

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(1), 35-54. https://doi.org/10.5194/hess-26-35-2022

1321   Levison, J., Larocque, M., Ouellet, M. A., Ferland, O., & Poirier, C. (2016).

1322   Long-term trends in groundwater recharge and discharge in a fractured

1323   bedrock aquifer–past and future conditions. Canadian Water Resources

1324   Journal/Revue canadienne des ressources hydriques, 41(4), 500-514.

1325   https://doi.org/10.1080/07011784.2015.1037795

1326   Li, L., Dai, Y., Wei, Z., Wei, S., Zhang, Y., Wei, N., & Li, Q. (2024). Enforcing Water

1327   Balance in Multitask Deep Learning Models for Hydrological Forecasting.

1328   Journal   of   Hydrometeorology,   25(1),   89-103.

1329   https://doi.org/10.1175/JHM-D-23-0073.1

1330   Liu, J., Zhang, Q., Singh, V. P., & Shi, P. (2017). Contribution of multiple climatic

1331   variables and human activities to streamflow changes across China. Journal of

1332   Hydrology, 545, 145-162. https://doi.org/10.1016/j.jhydrol.2016.12.016

1333   Liu, X., Yang, K., Ferreira, V. G., & Bai, P. (2022). Hydrologic model calibration with

1334   remote sensing data products in global large basins. Water Resources Research,

1335   58(12), e2022WR032929. https://doi.org/10.1029/2022WR032929

1336   Lockhoff, M., Zolina, O., Simmer, C., & Schulz, J. (2014). Evaluation of

1337   satellite-retrieved extreme precipitation over Europe using gauge observations.

1338   Journal of climate, 27(2), 607-623. https://doi.org/10.1175/JCLI-D-13-00194.1

1339   Loomis, B. D., Rachlin, K. E., Wiese, D. N., Landerer, F. W., & Luthcke, S. B. (2020).

1340   Replacing GRACE/GRACE‐FO with satellite laser ranging: Impacts on

1341   Antarctic Ice Sheet mass change. Geophysical Research Letters, 47(3),

1342   e2019GL085488. https://doi.org/10.1029/2019GL085488

Lute, A. C., & Abatzoglou, J. T. (2014). Role of extreme snowfall events in interannual variability of snowfall accumulation in the western United States. Water Resources Research, 50(4), 2874-2888. https://doi.org/10.1002/2013WR014465

Luo, Z., Gao, Z., Wang, L., Wang, S., & Wang, L. (2023). A method for balancing the terrestrial water budget and improving the estimation of individual budget components. Agricultural and Forest Meteorology, 341, 109667. https://doi.org/10.1016/j.agrformet.2023.109667

Luo, Z., Li, H., Zhang, S., Wang, L., Wang, S., & Wang, L. (2023). A novel two‐step method for enforcing water budget closure and an intercomparison of budget closure correction methods based on satellite hydrological products. Water Resources Research, 59(3), e2022WR032176. https://doi.org/10.1029/2022WR032176

Luo, Z., Yu, H., Liu, H., & Chen, J. (2023). Assessing the Water Budget Closure Accuracy of Satellite/Reanalysis-Based Hydrological Data Products over Mainland China. Remote Sensing, 15(21), 5230. https://doi.org/10.3390/rs15215230

Lv, M., Ma, Z., Yuan, X., Lv, M., Li, M., & Zheng, Z. (2017). Water budget closure based on GRACE measurements and reconstructed evapotranspiration using GLDAS and water use data for two large densely-populated mid-latitude basins. Journal of Hydrology, 547, 585-599. https://doi.org/10.1016/j.jhydrol.2017.02.027

Majid, R., & Ardalan, E. S. (2019). Performance of the Gravity Recovery and Climate Experiment (GRACE) method in monitoring groundwater-level changes in local-scale study regions within Iran. Hydrogeology Journal, 27(7), 2497-2509. https://doi.org/10.1007/s10040-019-02007-x

Masunaga, H., Schröder, M., Furuzawa, F. A., Kummerow, C., Rustemeier, E., & Schneider, U. (2019). Inter-product biases in global precipitation extremes. Environmental Research Letters, 14(12), 125016. https://doi.org/10.1088/1748-9326/ab5da9

McMahon, T. A., Finlayson, B. L., & Peel, M. C. (2016). Historical developments of models for estimating evaporation using standard meteorological data. Wiley Interdisciplinary Reviews: Water, 3(6), 788-818. https://doi.org/10.1002/wat2.1172

Mehrnegar, N., Schumacher, M., Jagdhuber, T., & Forootan, E. (2023). Making the best use of GRACE, GRACE‐FO and SMAP data through a constrained Bayesian data‐model integration. Water Resources Research, 59(9), e2023WR034544. https://doi.org/10.1029/2023WR034544

Miralles, D. G., Holmes, T., De Jeu, R., Gash, J., Meesters, A., & Dolman, A. (2011). Global land-surface evaporation estimated from satellite-based observations. Hydrology and Earth System Sciences, 15(2), 453-469. https://doi.org/10.5194/hess-15-453-2011

Mourad, R., Schoups, G., Bastiaanssen, W., & Kumar, D.N. (2024). Expert-based prior uncertainty analysis of gridded water balance components: Application to

the irrigated Hindon River Basin, India. Journal of Hydrology: Regional

Studies, 55, 2214-5818. https://doi.org/10.1016/j.ejrh.2024.101935.

Mu, D., Yan, H., Feng, W., & Peng, P. (2017). GRACE leakage error correction with

regularization technique: Case studies in Greenland and Antarctica.

Geophysical Journal International, 208(3), 1775-1786.

https://doi.org/10.1093/gji/ggw494

Mueller, D. S. (2003, March). Field evaluation of boat-mounted acoustic Doppler

instruments used to measure streamflow. In Proceedings of the IEEE/OES

Seventh Working Conference on Current Measurement Technology, 2003. (pp.

30-34). IEEE. https://doi.org/10.1109/CCM.2003.1194278

Mueller, B., Seneviratne, S. I., Jimenez, C., Corti, T., Hirschi, M., Balsamo, G., . . .

Guo, Z. (2011). Evaluation of global observations‐based evapotranspiration

datasets and IPCC AR4 simulations. Geophysical Research Letters, 38(6).

https://doi.org/10.1029/2010GL046230

Munier, S., & Aires, F. (2018). A new global method of satellite dataset merging and

quality characterization constrained by the terrestrial water budget. Remote

Sensing of Environment, 205, 119-130.

https://doi.org/10.1016/j.rse.2017.11.008

Nassaj, B. N., Zohrabi, N., Shahbazi, A. N., & Fathian, H. (2022). Evaluating the

performance of eight global gridded precipitation datasets across Iran.

Dynamics of Atmospheres and Oceans, 98, 101297.

https://doi.org/10.1016/j.dynatmoce.2022.101297

Palharini, R. S. A., Vila, D. A., Rodrigues, D. T., Quispe, D. P., Palharini, R. C., de Siqueira, R. A., & de Sousa Afonso, J. M. (2020). Assessment of the extreme precipitation by satellite estimates over South America. Remote Sensing, 12(13), 2085. https://doi.org/10.3390/rs12132085

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(9), 3191-3206. https://doi.org/10.1175/JCLI-D-11-00300.1

Pan, S., Pan, N., Tian, H., Friedlingstein, P., Sitch, S., Shi, H., . . . Kato, E. (2020). Evaluation of global terrestrial evapotranspiration using state-of-the-art approaches in remote sensing, machine learning and land surface modeling. Hydrology and Earth System Sciences, 24(3), 1485-1509. https://doi.org/10.5194/hess-24-1485-2020

Papacharalampous, G., Tyralis, H., Markonis, Y., Máca, P., & Hanel, M. (2023). Features of the Earth's seasonal hydroclimate: characterizations and comparisons across the Köppen–Geiger climates and across continents. Progress in Earth and Planetary Science, 10(1), 46. https://doi.org/10.1186/s40645-023-00574-y

Park, J., & Choi, M. (2015). Estimation of evapotranspiration from ground-based meteorological data and global land data assimilation system (GLDAS). Stochastic Environmental Research and Risk Assessment, 29, 1963-1992. https://doi.org/10.1007/s00477-014-1004-2

1431 Pellet, V., Aires, F., Papa, F., Munier, S., & Decharme, B. (2020). Long-term total

1432 water storage change from a Satellite Water Cycle reconstruction over large

1433 southern Asian basins. Hydrology and Earth System Sciences, 24(6),

1434 3033-3055. https://doi.org/10.5194/hess-24-3033-2020

1435 Peltier, W. R., Drummond, R., & Roy, K. (2012). Comment on "Ocean mass from

1436 GRACE and glacial isostatic adjustment" by DP Chambers et al. Journal of

1437 Geophysical Research: Solid Earth, 117(B11).

1438 https://doi.org/10.1029/2011JB008967

1439 Petković, V., & Kummerow, C. D. (2017). Understanding the sources of satellite

1440 passive microwave rainfall retrieval systematic errors over land. Journal of

1441 Applied Meteorology and Climatology, 56(3), 597-614.

1442 https://doi.org/10.1175/JAMC-D-16-0174.1

1443 Potter, N. J., Zhang, L., Milly, P. C. D., McMahon, T. A., & Jakeman, A. J. (2005).

1444 Effects of rainfall seasonality and soil moisture capacity on mean annual water

1445 balance for Australian catchments. Water Resources Research, 41(6).

1446 https://doi.org/10.1029/2004WR003697

1447 Reager, J. T., Thomas, B. F., & Famiglietti, J. S. (2014). River basin flood potential

1448 inferred using GRACE gravity observations at several months lead time.

1449 Nature Geoscience, 7(8), 588-592. https://doi.org/10.1038/ngeo2203

1450 Resende, T. C., Longuevergne, L., Gurdak, J., Leblanc, M., Favreau, G., Ansems,

1451 N., ... & Aureli, A. (2019). Assessment of the impacts of climate variability on

1452 total water storage across Africa: implications for groundwater resources

management. Hydrogeology Journal, 27(1), 493-512. https://doi.org/10.1007/s10040-018-1864-5

Rodell, M., Velicogna, I., & Famiglietti, J. S. (2009). Satellite-based estimates of groundwater depletion in India. Nature, 460(7258), 999-1002. https://doi.org/10.1038/nature08238

Ruhoff, A., de Andrade, B. C., Laipelt, L., Fleischmann, A. S., Siqueira, V. A., Moreira, A. A., ... & Biggs, T. (2022). Global evapotranspiration datasets assessment using water balance in South America. Remote Sensing, 14(11), 2526. https://doi.org/10.3390/rs14112526

Sadeghi, M., Akbari Asanjan, A., Faridzad, M., Afzali Gorooh, V., Nguyen, P., Hsu, K., ... & Braithwaite, D. (2019). Evaluation of PERSIANN-CDR constructed using GPCP V2. 2 and V2. 3 and a comparison with TRMM 3B42 V7 and CPC unified gauge-based analysis in global scale. Remote Sensing, 11(23), 2755. https://doi.org/10.3390/rs11232755

Sahoo, A. K., Pan, M., Troy, T. J., Vinukollu, R. K., Sheffield, J., & Wood, E. F. (2011). Reconciling the global terrestrial water budget using satellite remote sensing. Remote Sensing of Environment, 115(8), 1850-1865. https://doi.org/10.1016/j.rse.2011.03.009

Sankarasubramanian, A., & Vogel, R. M. (2002). Annual hydroclimatology of the United States. Water Resources Research, 38(6), 19-1. https://doi.org/10.1029/2001WR000619

Schiavo, M. (2023). The role of different sources of uncertainty on the stochastic

quantification of subsurface discharges in heterogeneous aquifers. Journal of Hydrology, 617, 128930. https://doi.org/10.1016/j.jhydrol.2022.128930

Schneider, U., Fuchs, T., Meyer-Christoffer, A., & Rudolf, B. (2008). Global precipitation analysis products of the GPCC. Global Precipitation Climatology Centre (GPCC), DWD, Internet Publikation, 112, 3819-3837.

Senan, S., Thomas, J., Vema, V. K., Jainet, P. J., Nizar, S., Sivan, S., & Sudheer, K. P. (2022). A study of the influence of rainfall datasets' spatial resolution on stream simulation in Chaliyar River Basin, India. Journal of Water and Climate Change, 13(12), 4234-4254. https://doi.org/10.2166/wcc.2022.273

Shamsudduha, M., Taylor, R. G., Jones, D., Longuevergne, L., Owor, M., & Tindimugaya, C. (2017). Recent changes in terrestrial water storage in the Upper Nile Basin: an evaluation of commonly used gridded GRACE products. Hydrology and Earth system sciences, 21(9), 4533-4549. https://doi.org/10.5194/hess-21-4533-2017

Shiklomanov, A. I., Yakovleva, T. I., Lammers, R. B., Karasev, I. P., Vörösmarty, C. J., & Linder, E. (2006). Cold region river discharge uncertainty—Estimates from large Russian rivers. Journal of Hydrology, 326(1–4), 231–256. https://doi.org/10.1016/j.jhydrol.2005.10.037

Su, Y., & Zhang, S. (2024). Optimizing Parameters in the Common Land Model by Using Gravity Recovery and Climate Experiment Satellite Observations. Land, 13(4), 508. https://doi.org/10.3390/land13040508

Swenson, S., & Wahr, J. (2006). Post‑processing removal of correlated errors in

GRACE data. Geophysical research letters, 33(8). https://doi.org/10.1029/2005GL025285

Tan, X., Liu, B., Tan, X., & Chen, X. (2022). Long‑term water imbalances of watersheds resulting from biases in hydroclimatic data sets for water budget analyses. Water Resources Research, 58(3), e2021WR031209. https://doi.org/10.1029/2021WR031209

Tapley, B. D., Bettadpur, S., Ries, J. C., Thompson, P. F., & Watkins, M. M. (2004). GRACE measurements of mass variability in the Earth system. science, 305(5683), 503-505. https://dor.org/10.1126/science.1099192

Trenberth, K. E., Dai, A., Van Der Schrier, G., Jones, P. D., Barichivich, J., Briffa, K. R., & Sheffield, J. (2014). Global warming and changes in drought. Nature Climate Change, 4(1), 17-22. https://doi.org/10.1038/nclimate2067

Wang, L., Wang, J., Li, M., Wang, L., Li, X., & Zhu, L. (2022). Response of terrestrial water storage and its change to climate change in the endorheic Tibetan Plateau. Journal of Hydrology, 612, 128231. https://doi.org/10.1016/j.jhydrol.2022.128231

Wang, S., McKenney, D. W., Shang, J., & Li, J. (2014). A national‑scale assessment of long‑term water budget closures for Canada's watersheds. Journal of Geophysical Research: Atmospheres, 119(14), 8712-8725. https://doi.org/10.1002/2014JD021951

Wang, S., Huang, J., Yang, D., Pavlic, G., & Li, J. (2015). Long‑term water budget imbalances and error sources for cold region drainage basins. Hydrological

Processes, 29(9), 2125-2136. https://doi.org/10.1002/hyp.10343

Wang, Z., Zhong, R., Lai, C., & Chen, J. (2017). Evaluation of the GPM IMERG satellite-based precipitation products and the hydrological utility. Atmospheric Research, 196, 151-163. https://doi.org/10.1016/j.atmosres.2017.06.020

Watkins, M. M., Wiese, D. N., Yuan, D. N., Boening, C., & Landerer, F. W. (2015). Improved methods for observing Earth's time variable mass distribution with GRACE using spherical cap mascons. Journal of Geophysical Research: Solid Earth, 120(4), 2648-2671. https://doi.org/10.1002/2014JB011547

Wei, L., Jiang, S., Ren, L., Yuan, F., & Zhang, L. (2019). Performance of two long-term satellite-based and GPCC 8.0 precipitation products for drought monitoring over the Yellow River Basin in China. Sustainability, 11(18), 4969. https://doi.org/10.3390/su11184969

Wong, J. S., Zhang, X., Gharari, S., Shrestha, R. R., Wheater, H. S., & Famiglietti, J. S. (2021). Assessing water balance closure using multiple data assimilation– and remote sensing–based datasets for canada. Journal of Hydrometeorology, 22(6), 1569-1589. https://doi.org/10.1175/JHM-D-20-0131.1

Wu, C., Yeh, P. J. F., Hu, B. X., & Huang, G. (2018). Controlling factors of errors in the predicted annual and monthly evaporation from the Budyko framework. Advances in water resources, 121, 432-445. https://doi.org/10.1016/j.advwatres.2018.09.013

Wu, Z., Zhang, Y., Sun, Z., Lin, Q., & He, H. (2018). Improvement of a combination of TMPA (or IMERG) and ground-based precipitation and application to a

typical region of the East China Plain. Science of the Total Environment, 640, 1165-1175. https://doi.org/10.1016/j.scitotenv.2018.05.272

Xu, T., Guo, Z., Xia, Y., Ferreira, V. G., Liu, S., Wang, K., . . . Zhao, C. (2019). Evaluation of twelve evapotranspiration products from machine learning, remote sensing and land surface models over conterminous United States. Journal of Hydrology, 578, 124105. https://doi.org/10.1016/j.jhydrol.2019.124105

Yao, Y., Liang, S., Li, X., Hong, Y., Fisher, J. B., Zhang, N., . . . Zhang, X. (2014). Bayesian multimodel estimation of global terrestrial latent heat flux from eddy covariance, meteorological, and satellite observations. Journal of Geophysical Research: Atmospheres, 119(8), 4521-4545. https://doi.org/10.1002/2013JD020864

Yeh, P. J. F., Swenson, S. C., Famiglietti, J. S., & Rodell, M. (2006). Remote sensing of groundwater storage changes in Illinois using the Gravity Recovery and Climate Experiment (GRACE). Water Resources Research, 42(12). https://doi.org/10.1029/2006WR005374

Zhang, D., Liu, X., Zhang, Q., Liang, K., & Liu, C. (2016). Investigation of factors affecting intra-annual variability of evapotranspiration and streamflow under different climate conditions. Journal of Hydrology, 543, 759-769. https://doi.org/10.1016/j.jhydrol.2016.10.047

Zhang, L., Potter, N., Hickel, K., Zhang, Y., & Shao, Q. (2008). Water balance modeling over variable time scales based on the Budyko framework–Model

development and testing. Journal of Hydrology, 360(1-4), 117-131. https://doi.org/10.1016/j.jhydrol.2008.07.021

Zhang, Y., Pan, M., & Wood, E. F. (2016). On creating global gridded terrestrial water budget estimates from satellite remote sensing. Remote sensing and water resources, 59-78. https://doi.org/10.1007/978-3-319-32449-4_4

Zhang, Y., Pan, M., Sheffield, J., Siemann, A. L., Fisher, C. K., Liang, M., . . . Houser, P. R. (2018). A Climate Data Record (CDR) for the global terrestrial water budget: 1984–2010. Hydrology and Earth System Sciences, 22(1), 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(3), 627-653. https://doi.org/10.5194/hess-29-627-2025

Zhou, L., Cao, Y., Shi, C., Liang, H., & Fan, L. (2024). Quantifying the Atmospheric Water Balance Closure over Mainland China Using Ground-Based, Satellite, and Reanalysis Datasets. Atmosphere, 15(4), 497. https://doi.org/10.3390/atmos15040497