Hydrology and
Earth System
Sciences

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

Zengliang Luo[1,2], Hanjia Fu[1,2], Quanxi Shao[3], Wenwen Dong[1,2], Xi Chen[4], Xiangyi Ding[5], Lunche Wang[1,2], Xihui Gu[1], Ranjan Sarukkalige[6], Heqing Huang[1], and Huan Li[7]

[1]Hubei Key Laboratory of Regional Ecology and Environmental Change, State Key Laboratory of Biogeology and Environmental Geology, China University of Geosciences, Wuhan 430074, China
[2]Engineering Research Center of Natural Resource Information Management and Digital Twin Engineering Software, Ministry of Education, Wuhan 430074, China
[3]CSIRO Data61, Australian Resources Research Centre, 26 Dick Perry Avenue, Kensington, WA 6155, Australia
[4]State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China
[5]Department of Water Resources, China Institute of Water Resources and Hydropower Research, Beijing 100038, China
[6]School of Civil and Mechanical Engineering, Curtin University, GPO Box U1987, Perth, WA 6845, Australia
[7]HUN-REN Balaton Limnological Research Institute, Tihany, Hungary

Correspondence: Quanxi Shao (quanxi.shao@data61.csiro.au) and Xiangyi Ding (dingxy@iwhr.com)

Abstract. 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.

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.

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., 2020a; 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), Eq. (1) is essential for advancing our understanding of hydrological processes (Li et al., 2024; Mourad et al., 2024).

$$P - \mathrm{ET} - Q - \mathrm{TWSC} = 0 \tag{1}$$

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 and De Bruin, 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 $= P - \mathrm{ET} - Q -$ TWSC (Zhou et al., 2024; Munier and Aires, 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., 2023a). For ex-

ample, 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., 2023a). 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 vali-

https://doi.org/10.5194/hess-29-4607-2025

Hydrol. Earth Syst. Sci., 29, 4607–4635, 2025

dated through comparisons with existing BCC methods (PR, CKF, MCL, MSD).

2 Study area and data

To robustly quantify the uncertainties introduced by existing BCC methods in closing the water budget and to assess the accuracy of the proposed IWE-Res method across different climate zones, multiple river basins worldwide were selected as study areas. In total, 84 basins (Fig. 1) were chosen based on the availability of streamflow observations from the Global Runoff Data Centre (GRDC) for the period 2002–2020. To ensure data reliability, the proportion of missing data was kept below 10 %, with missing values interpolated using a linear method. Notably, approximately 90 % of the basins used in this study had less than 5 % missing data.

The climate classifications presented in Fig. 1 were determined using the Köppen climate classification system, a widely adopted framework that categorizes global climates based on temperature and precipitation thresholds (Crosbie et al., 2012; 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 (Eqs. 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 (Eq. 3).

$$C_{jkl} = \left[P_j \mathrm{ET}_k \mathrm{TWSC}_l Q \right] \tag{2}$$

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 Eq. (2).

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

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., 2019a, 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., 2020b; 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 vari-
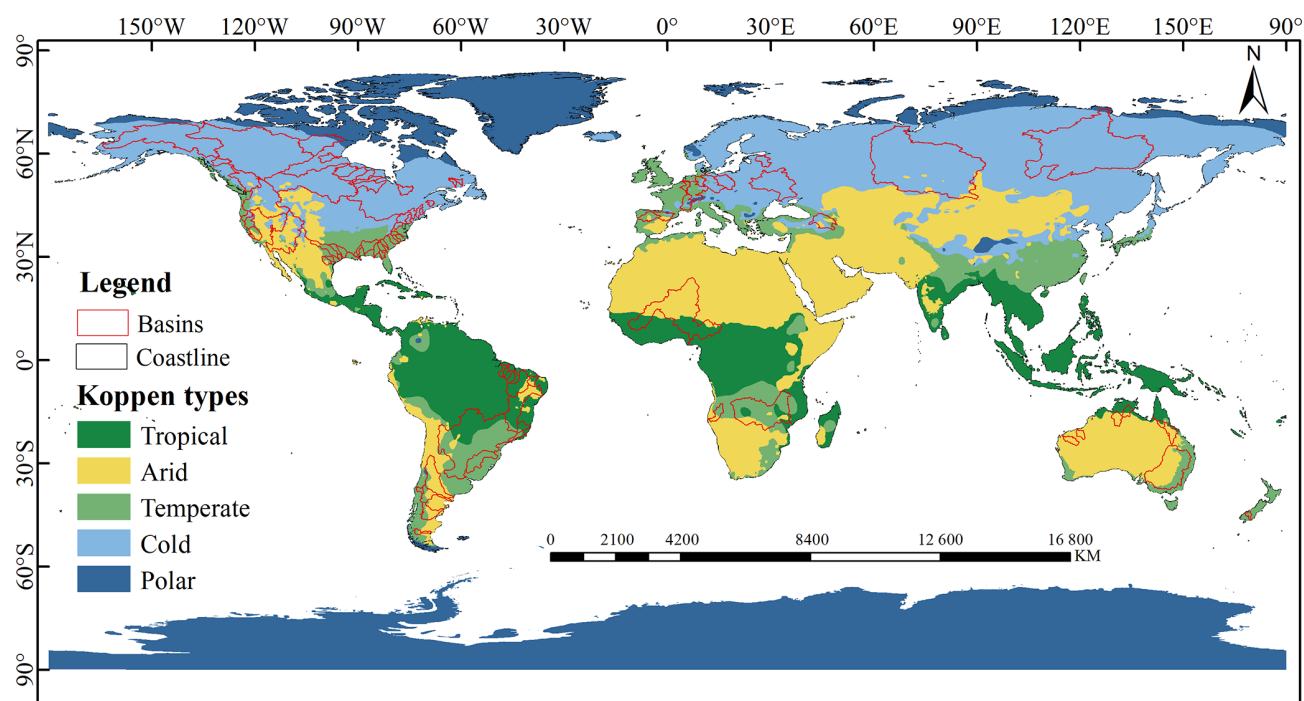
Figure 1. Overview of the Study Area. The climate classification used in this study is based on the Köppen climate classification system.

ations 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 which include multiple bias correction procedures to improve data quality (Landerer et al., 2012; Shamsudduha et al., 2017). These bias correction procedures include filtration to suppress correlated noise and striping artifacts (Swenson and Wahr, 2006), replacement of poorly resolved spherical harmonic coefficients (e.g., degree-2 term C20) with satellite laser ranging data (Loomis et al., 2020), and correction for glacial isostatic adjustment (GIA) (Peltier et al., 2012; Mu et al., 2017). Numerous studies have demonstrated the sensitivity and reliability of GRACE satellite data for monitoring TWSC (Swenson and Wahr, 2006; Resende et al., 2019; Majid et al., 2019; Reager et al., 2014).

The GRDC provides the most comprehensive open-access river discharge data available worldwide, collected from national hydrological agencies. This dataset includes river streamflow measurements from over 10 000 stations across 159 countries (Su and Zhang, 2024). To minimize the impact of missing data on the reliability of the results, hydrological stations were selected based on the criterion that missing values accounted for less than 10 % of the total dataset. Linear interpolation was then applied to fill any remaining data gaps.

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 (Eq. 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., Eq. 1 becomes Eq. 4) (Aires, 2014; Wong et al., 2021). The resulting imbalance is represented by the residual error term ΔRes (Eq. 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 ap-

Table 1. Datasets used for this study.

Variable	Data Source	Number	Resolution	Reference
P	Global Precipitation Climatology Centre (GPCC)	1	0.25° per month	Schneider et al. (2008)
	Global Precipitation Measurement (GPM IMERG)	2	0.1° per month	Huffman et al. (2015)
	Multi-Source Weighted-Ensemble Precipitation (MSWEP)	3	0.1° per month	Beck et al. (2019b)
	Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks – Climate Data Record (PERSIANN-CDR)	4	0.25° per month	Hsu et al. (1997)
ET	global land data assimilation system (GLDAS)	1	0.25° per month	Park and Choi (2015)
	Global Land Evaporation Amsterdam Model (GLEAM)	2	0.25° per month	Miralles et al. (2011)
	TerraClimate	3	1/24° per month	Abatzoglou et al. (2018)
TWSC	Gravity Recovery and Climate Experiment (GRACE CSR)	1	1.0° per month	Watkins et al. (2015)
	Gravity Recovery and Climate Experiment (GRACE GFZ)	2	1.0° per month	Watkins et al. (2015)
	Gravity Recovery and Climate Experiment (GRACE JPL)	3	1.0° per month	Watkins et al. (2015)
Q	Global Runoff Data Centre (GRDC)	–	–	Burek and Smilovic (2023)

proximate their true values (Luo et al., 2023a).

$$(P + \varepsilon_P) - (\text{ET} + \varepsilon_{\text{ET}}) - \left(Q + \varepsilon_Q \right)$$
$$- (\text{TWSC} + \varepsilon_{\text{TWSC}}) = 0 \tag{4}$$
$$\Delta\text{Res} = \varepsilon_{\text{ET}} + \varepsilon_Q + \varepsilon_{\text{TWSC}} - \varepsilon_P$$
$$= P - \text{ET} - Q - \text{TWSC} \tag{5}$$

where ε_P, ε_{ET}, ε_Q, $\varepsilon_{\text{TWSC}}$ 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 Eq. (5) (reducing ΔRes from $\neq 0$ 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., 2023b; 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.

For each basin, these four methods were applied to 36 different data combinations (Eq. 3), yielding 144 uncertainty estimates. The optimal combinations were identified using a 5 % threshold. By averaging the errors introduced into budget components across these selected optimal combinations, we quantified the uncertainty associated with existing BCC methods. This approach minimizes uncertainties arising from both BCC method selection and budget component data selection, enabling a more objective evaluation of the errors introduced by existing BCC methods. A brief overview of the four BCC methods is provided below:

1. *PR method.*

   The PR method assumes that the error in budget components is proportional to their magnitudes (Abatzoglou et al., 2018). Based on the relative magnitudes of these variables, the ΔRes error is redistributed across them to achieve water budget closure (Eq. 6).

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

where F_i and X_i represent the corrected and original data for budget components (P, ET, Q and TWSC), respectively; n denotes the number of budget components involved in the water budget closure calculation; $\Delta\mathrm{Res}$ represents the water imbalance error; G is a constant vector defined as $G = [1 \; -1 \; -1 \; -1]$.

2. *CKF method.*

The CKF method is developed based on the Kalman filter method (Pan and wood, 2006). For a given set of estimated budget components $X = [P \, \mathrm{ET} \, Q \, \mathrm{TWSC}]^T$ and their estimated errors $\Delta\mathrm{Res} = \boldsymbol{G}X \neq 0$ (where $\boldsymbol{G}$ is a constant vector, $\boldsymbol{G} = [1\,{-}1\,{-}1\,{-}1]$), the goal is to find a new set of estimates $F = [P' \, \mathrm{ET}' \, Q' \, \mathrm{TWSC}']^T$ such that $\boldsymbol{G}X' = 0$, achieving water budget closure (Pan et al., 2012). In simple terms, the CKF method redistributes the $\Delta\mathrm{Res}$ among the budget components based on the error covariance of X, defined as $\Delta\varepsilon_{XX}$ (Eq. 7), to obtain a closured dataset.

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

where X_0 refers to the reference values of the estimated budget components, and the bar over an expression denotes expectation. For P, ET and TWSC, the reference values X_0 were calculated by averaging all considered datasets, following previous studies (Zhang et al., 2018; Abolafia-Rosenzweig et al., 2021). For Q, we adopted observed Q. Due to the difficulty in quantifying the uncertainty in observed Q, previous studies have reported gauge-based uncertainty as a percent error for some of the basins, ranging from 2.3 %–28.8 % (Clarke, 1999; Mueller, 2003; Shiklomanov et al., 2006; Abolafia-Rosenzweig et al., 2021). We followed a similar approach to estimate the uncertainty associated with Q in this study.

The error covariance matrix $\Delta\varepsilon_{XX}$ is of dimension 4×4 and represents the covariances among errors in the four budget components:

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

Following Pan et al. (2012), the off-diagonal elements representing cross-variable error covariances were assumed to be zero, under the assumption that errors among different budget components are uncorrelated. Accordingly, the matrix $\mathbf{F}$ can be computed as shown in Eq. (9).

$$\mathbf{F} = X + K(0 - \boldsymbol{G}X) \tag{9}$$

where $K = \Delta\varepsilon_{XX}C^T(C\Delta\varepsilon_{XX}C^T)^{-1}$ is the Kalman gain. Setting $\boldsymbol{G}X = \Delta\mathrm{Res}$, and Eq. (9) can be rewritten as Eq. (10).

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

where error covariance ε_{XX} is calculated entry by entry according to Eq. (8).

3. *MCL method.*

The MCL method is an extension of the triple collocation (TC) method. It calculates the weights for redistributing the $\Delta\mathrm{Res}$ error among budget components by estimating the errors relative to their true values (expressed as distances, without requiring knowledge of the true values). The fundamental equations of the MCL method are shown in Eqs. (11)–(12).

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

$$d^i_{xx_0-\mathrm{norm}} = \frac{d^i_{xx_0}}{\sum\limits_{j=1}^{4}\left|d^j_{xx_0}\right|} \tag{12}$$

In these equations, F_i represents the corrected data for the ith budget component; X_i denotes the original data for the ith budget component; $\Delta\mathrm{Res}$ represents the water imbalance error; $d^i_{xx_0-\mathrm{norm}}$ represents the weight assigned to the ith budget component, and $d^i_{xx_0}$ represents the distance between the ith budget component and the true value, as calculated using the Monte Carlo (MC) method. For example, in the case of five precipitation data products ($N = 5$), the calculation of $d^i_{xx_0}$ ($d1t$, $d2t$, $d3t$, $d4t$, and $d5t$) is shown in Eqs. (13)–(14).

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

$$A_{(5)} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix},$$

$$y_{(5)} = \begin{bmatrix} d^2_{1t} \\ d^2_{2t} \\ d^2_{3t} \\ d^2_{4t} \\ d^2_{5t} \end{bmatrix}, \; b_{(5)} = \begin{bmatrix} d^2_{12} \\ d^2_{13} \\ d^2_{14} \\ d^2_{15} \\ d^2_{23} \\ d^2_{24} \\ d^2_{25} \\ d^2_{34} \\ d^2_{35} \\ d^2_{45} \end{bmatrix} \tag{14}$$

4. *MSD method.*

The MSD method redistributes the $\Delta\mathrm{Res}$ to each budget component based on minimizing the time-series deviation error, aiming to reduce model uncertainties caused

by errors in estimating time-point deviations (Luo et al., 2023b). Specifically, the MSD method first calculates the minimum time-series deviation distance between remote sensing data for budget components and multi-source integrated data products (EO) (Eq. 15).

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

where $D_{x,\to n}$ represents the minimum time-series deviation distance for budget component x (e.g., P, ET, TWSC); $y_{(\mathrm{EO},j)}$ and $x_{(\mathrm{RS},j)}$ refer to the integrated value and raw value of the budget component x, respectively; $\overline{y_{(\mathrm{EO},\to n)}}$ and $\overline{x_{(\mathrm{RS},\to n)}}$ denote the average deviation of budget component x from the first to the nth time point.

Next, the MSD method calculates the weights for each budget component based on $D_{x,\to n}$ (Eq. 16).

$$w_{x,j} = \frac{D_{x,\to j}}{\sum_{i=1}^4 D_{i,\to j}} \quad (16)$$

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

Finally, the weight calculation results from Eq. (16) are substituted into Eq. (17) to achieve water budget closure.

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

where F^{BCC} represents the budget components (P, ET, Q, and TWSC) corrected for water budget closure, while F^{Raw} denotes the raw, uncorrected values of the budget components.

3.3 Uncertainties introduced by existing BCC methods for closing water budget

When the existing BCC methods described in Sect. 3.2 are applied to close the water budget, they redistribute ΔRes based on the estimated errors of budget components but neglect unmeasured components. This inevitably leads to an unreasonable redistribution of the ΔRes error, introducing new uncertainties. The magnitude of these introduced errors and whether they can be ignored remain unresolved, primarily due to insufficient observational data for some budget components, making it difficult to quantify the associated uncertainties.

Our analysis in this study reveals that when existing BCC methods are used for water budget closure, certain budget components that typically have positive values, such as P, ET, and Q, occasionally become negative in some months. Previous studies have also mentioned this issue (Lehmann et al., 2022). This clearly indicates an unreasonable redistribution of ΔRes errors, underscoring the urgent need for methodological improvements. Despite this issue, research on negative values remains limited. Key questions persist regarding the proportion of negative values in each budget component under current BCC methods, which variables are most susceptible to severe negative values, and how these errors vary throughout the year. Addressing these questions is critical for refining existing BCC methods.

Notably, quantifying negative values does not require observational data. To comprehensively assess the uncertainties introduced by forced water budget closure, we consider three aspects: errors of individual budget components relative to observed values (Sect. 4.2.1), negative values arising from budget closure (Sect. 4.2.2), and ensemble errors (Sect. 4.2.3).

1. *Errors of individual budget components*.

   Quantifying this type of error requires determining reference values for budget components. However, for certain variables, such as ET, observational data are insufficient across global watersheds, posing a major challenge in accurately characterizing global ET patterns. As a result, approximate reference values must be used to ensure the reliability of the results.

   In this study, reference values for budget components were established based on the following principles. For Q, long-term observational records from hydrological stations were available for all selected basins, meeting the study's requirements. For TWSC, we utilized three observational datasets from the GRACE satellite, which currently provides the only large-scale measurements of basin water storage changes 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 Eq. (18), to merge the three GRACE TWSC products into a single reference dataset (Munier and 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 lack-

ing sufficient observational precipitation data were included, for which fused precipitation values, derived using Eq. (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 validation (Jiménez et al., 2018; Mueller et al., 2011; Yao et al., 2014). Following this approach, we assessed the uncertainty introduced by existing BCC methods for ET by utilizing a fusion-based reference dataset.

$$\overline{M_x} = \sum_{i=1}^{n} M_{x,i} \cdot \omega_i \text{ and } \omega_i = \frac{1}{\sigma_i^2} / \sum_{i=1}^{1} \frac{1}{\sigma_i^2} \quad (18)$$

where $\overline{M_x}$ represents the fused value of the budget component, $M_{x,i}$ denotes the ith product of the budget component; ω_i denotes the weight of the ith product, and σ_i^2 refers to the covariance error of the ith product, n is the total number of budget components, and x refers to P, ET or TWSC.

After establishing reference values for budget components, we quantify errors in the original data relative to these references, using the positive metric CC and inverse metric RMSE as examples, denoted as CC_1 and $RMSE_1$, respectively. Similarly, errors in the BCC-corrected data relative to the reference values are calculated, represented as CC_2 and $RMSE_2$.

To assess the uncertainties introduced by water budget closure, changes in CC and RMSE (CC' and $RMSE'$) are computed using Eqs. (19) and (22). Positive values of CC' and $RMSE'$ indicate an improvement in data accuracy following BCC correction, whereas negative values suggest a decline. In addition to CC and RMSE, other statistical metrics used in this study include the positive indicator NSE and the negative indicator MAE.

$$CC' = CC_2 - CC_1 \quad (19)$$
$$NSE' = NSE_2 - NSE_1 \quad (20)$$
$$MAE' = MAE_1 - MAE_2 \quad (21)$$
$$RMSE' = RMSE_1 - RMSE_2 \quad (22)$$
$$CC = \frac{\sum_{i=1}^{n}(Obs_i - \overline{Obs})(Sim_i - \overline{Sim})}{\sqrt{\sum_{i=1}^{n}(Obs_i - \overline{Obs})^2}\sqrt{\sum_{i=1}^{n}(Sim_i - \overline{Sim})^2}} \quad (23)$$

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

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

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

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 budget component, their seasonal distribution, and their sensitivity to climatic conditions (i.e., their prevalence in arid versus humid basins). Insights from this analysis were incorporated into the proposed IWE-Res method to address the occurrence of negative values (Sect. 3.4).

3. *Ensembled error of four budget components.*

The aforementioned evaluations (1) and (2) assess errors for individual budget components. To gain a more comprehensive understanding of the uncertainties introduced by water budget closure, we also evaluate the combined error. First, the absolute error (AE) of each budget component is calculated (using P as an example, see Eq. 29). Second, the relative absolute error (RAE) is determined for each budget component (Eq. 28). Finally, by aggregating the relative errors of individual components, we define the ensembled relative error (Eq. 27) to quantify the overall error introduced by

BCC methods.

$$F(\text{Re}) = \frac{1}{n}\sum_{i=1}^{n} \frac{\begin{array}{c}\left|\text{AE}(P')\right| - \left|\text{AE}(P_{\text{Raw}})\right| + \left|\text{AE}(\text{ET}')\right| \\ -\left|\text{AE}(\text{ET}_{\text{Raw}})\right| + \left|\text{AE}(Q')\right| \\ -\left|\text{AE}(Q_{\text{Raw}})\right| + \left|\text{AE}(\text{TWSC}')\right| \\ -\left|\text{AE}(\text{TWSC}_{\text{Raw}})\right|\end{array}}{P_0 + \text{ET}_0 + Q_0 + |\text{TWSC}_0|}$$

$$= \frac{1}{n}\sum_{i=1}^{n} \frac{\text{RAE}(P) + \text{RAE}(\text{ET}) + \text{RAE}(Q) + \text{RAE}(\text{TWSC})}{P_0 + \text{ET}_0 + Q_0 + |\text{TWSC}_0|} \quad (27)$$

$$\text{RAE}(P) = \left|\text{AE}(P')\right| - \left|\text{AE}(P_{\text{Raw}})\right| \quad (28)$$

$$\text{AE}(P) = |P - P_0| \quad (29)$$

where, $F(\text{Re})$ represents the ensembled relative error, and RAE refers to the relative value of absolute error, with i denoting the month. The subscript "Raw" corresponds to the raw data of the budget components, the subscript 0 represents the observed data, the superscript "'" denotes the BCC-corrected data for the budget components. The degree of alteration induced by the BCC methods for each budget component are defined based on the value of $F(\text{Re})$, and four intervals are established in 5 % increments: no significant change [0 %–5 %], minor change (5 %–10 %], moderate change (10 %–15 %], and significant change (> 15 %).

3.4 Proposed IWE-Res method for closing water budget

In this section, we propose the IWE-Res method to identify the optimal balance for redistributing ΔRes, minimizing the sum of the introduced error to budget components and the remaining ΔRes error while reducing the negative values introduced by closing the water budget. Unlike existing BCC methods that fully redistribute the ΔRes term in a single step, the IWE-Res method adopts a gradual, iterative redistribution strategy that allows for more consistent correction. Specifically, the method incrementally allocates fractions of ΔRes to P, ET, Q and TWSC, based on fixed percentage steps and guided by existing BCC weighting schemes. At each iteration, the redistribution process seeks to minimize the combined error – defined as the sum of the induced changes in the water budget components and the remaining unexplained ΔRes. This dual-objective criterion ensures that the method balances error reduction while maintaining hydrological plausibility. Importantly, the approach includes a mechanism to avoid introducing implausible negative values. If, during any iteration, the corrected value of a component becomes negative – violating hydrological constraints such as non-negative precipitation or runoff – further redistribution to that component is halted. Subsequent iterations reallocate the remaining ΔRes among the unaffected components. From a hydrological perspective, 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 overcor-

recting 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 Eq. (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 $x \in [0, \Delta\text{Res}]$.

During each redistribution of ΔRes, two error terms are computed: (1) the remaining ΔRes error, defined as $\Delta\text{Res}^* = \Delta\text{Res} - x$, and (2) the error introduced to budget components due to the redistribution of the x error, denoted as IWE (Eq. 31). When these errors are plotted in a coordinate system, two distinct curves emerge (Fig. 2), each representing a different error relationship. For ΔRes^* (Eq. 30), Fig. 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 (Eqs. 31–32). This variability in the IWE curve arises from the nonlinear relationship between the introduced budget component errors and the reduction in ΔRes error.

$$\Delta\text{Res}^* = ax + b = -x + \Delta\text{Res} \quad (30)$$

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

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

where x represents the portion of ΔRes redistributed to the budget components, with a range from 0 to ΔRes. The terms $\varepsilon_P, \varepsilon_{\text{ET}}, \varepsilon_Q, \varepsilon_{\text{TWSC}}$ represent the errors introduced to P, ET, Q and TWSC, respectively, due to the redistribution of x to the budget components. $F(x, \text{RAE})$ denotes the RAE error calculated by the redistribution of the x error to budget components.

During the iterative correction process, if any of the water budget components (P, ET, and Q) becomes negative, the redistribution of water imbalance error to that component is immediately suspended. In subsequent iterations, redistribution is recalculated to ensure that only components with physically meaningful positive values receive the imbalance correction. For example, if ET becomes negative in a given iteration, the imbalance is subsequently redistributed to P, Q, and TWSC only, in accordance with Eq. (33). For TWSC, if a sign reversal occurs during iteration (i.e., from positive to negative or vice versa), the redistribution of the water imbal-

ance error to TWSC is suspended in the following iteration.

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

where F_i denotes the corrected dataset, and X_i denotes the original dataset of budget components. Since ET does not participate in the redistribution of the residual error x based on the example above, the weighting vector is defined as $G = [1, 0, -1, -1]$. The term ε represents the error in budget components estimated using existing BCC methods, as described in Sect. 3.2.

Third, the IWE-Res curve is plotted (Fig. 2) to provide an intuitive comparison between the introduced budget component errors and the remaining water imbalance error. The error calculation results from Eqs. (30) and (31) are presented within the same coordinate system.

The IWE-Res method is illustrated in Fig. 2 using four curves. The x axis represents the percentage of water imbalance error redistributed to budget components using existing BCC methods, while the y axis denotes the percentage of the remaining water imbalance error (ΔRes*) after each iteration. The black dashed line represents the redistributed x-error value among the budget components. The thin blue solid line represents the ΔRes* error curve. Since the sum of redistributed x and remaining ΔRes* equals the total ΔRes error, this curve forms a monotonically decreasing 45° line. The thin green solid line represents the introduced budget component error (IWE) after a given percentage of ΔRes is redistributed (x axis), with its shape varying depending on the redistribution process (Fig. 2 is illustrative). Initially, when no ΔRes is redistributed ($x = 0$), the IWE error is zero. As more ΔRes is redistributed (with increasing x values), IWE increases due to the growing uncertainty introduced. The thin red solid line represents the total error, defined as the sum of ΔRes* and IWE after applying BCC methods. This curve varies depending on the redistribution process, and its minimum value identifies the optimal balance where combined ΔRes* and IWE errors are minimized. The intersection of the ΔRes* and IWE curves indicates only the point at which these errors are equal, not necessarily the optimal balance.

To determine the optimal redistribution of the water imbalance error, we plot the IWE-Res curve (the green solid line) for each basin, identifying the minimum of the red total error curve. We then analyze its patterns across basins with different characteristics to optimize water budget closure and improve the accuracy of budget component datasets.

The IWE error in Fig. 2 also serves as a metric for evaluating the performance of existing BCC methods. If a BCC method perfectly redistributed ΔRes without introducing additional errors, the IWE curve would be a flat line at zero, and the red total error line would coincide with the blue ΔRes* error line. This scenario indicates that full redistribution of water imbalance error achieves the optimal balance, provid-

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

Finally, the optimal balance is identified, enabling the generation of a high-precision dataset that improves water budget closure. The optimal balance corresponds to the minimum of the total error curve (IWE + ΔRes*), where the sum of remaining water imbalance error and introduced budget component errors is minimized. Ideally, both ΔRes* and IWE would reach their minimum values simultaneously, meaning minimal error is introduced while fully redistributing ΔRes. However, since this ideal state may not always be achievable, identifying the point where combined error is minimized is essential. This principle defines the proposed IWE-Res method (Fig. 2).

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 budget components contribute to these discrepancies. Figure 3 illustrates the spatial distribution of monthly ΔRes errors across various data combinations. To prevent the cancellation of positive and negative values, the absolute values of monthly ΔRes errors were first computed for each basin and then averaged.

As shown in Fig. 3, ΔRes values vary significantly across basins. Most basins in Africa, South America, and Europe exhibit high ΔRes values, typically exceeding 20 mm. In North America, ΔRes values generally range from 15 to 45 mm. Due to inconsistencies among budget component datasets, substantial differences in ΔRes also emerge across different data combinations. In combinations where only P data varied while other budget component datasets remained constant (combinations in Fig. 3 where the first digit varies while the second and third remain constant), pronounced changes in water imbalance errors were observed in parts of southern Africa, northern Asia, and North America. This suggests substantial estimation errors in P for these regions.

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.
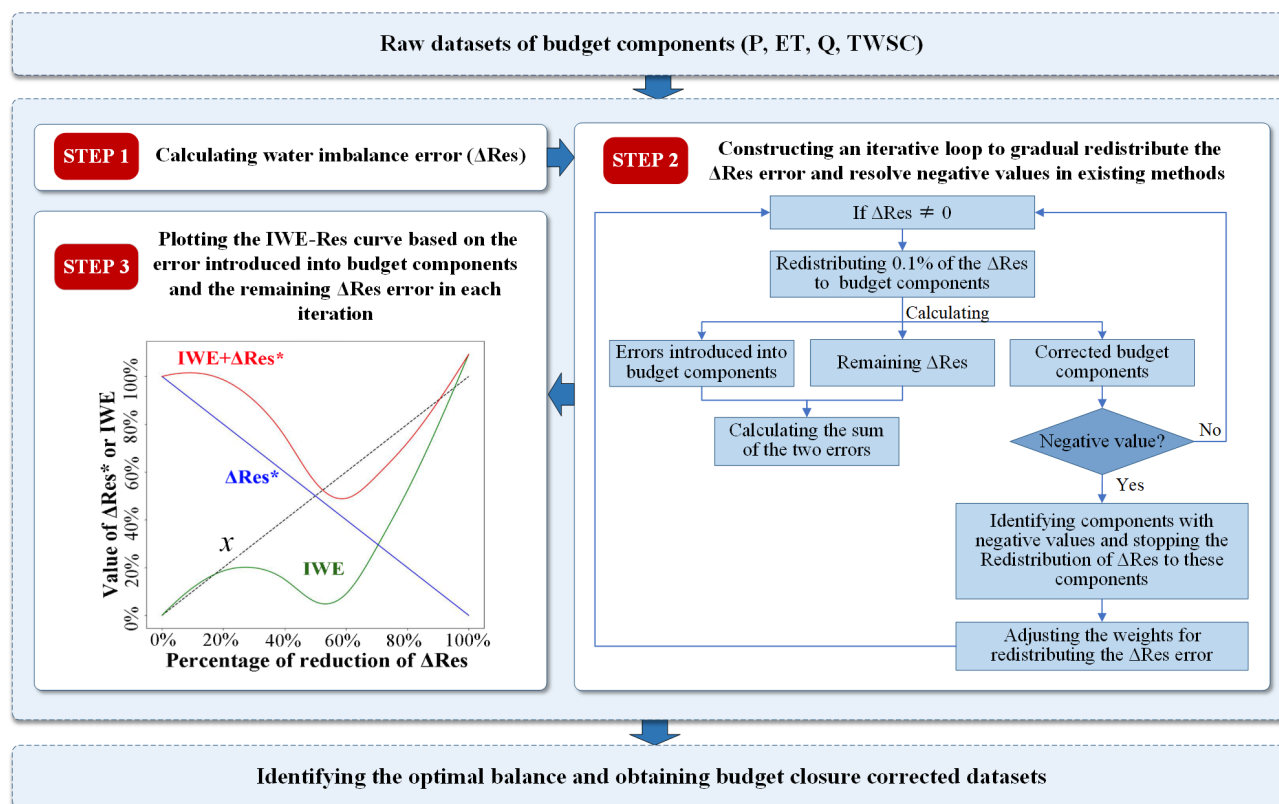
Figure 2. Framework of the IWE-Res method to identify the optimal balance for redistributing the ΔRes error. The x axis represents the proportion of ΔRes redistributed to budget components, while the y axis reflects the proportion of the remaining ΔRes error. The black dashed line represents the redistributed 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.

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., 2023a).

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 (Sect. 3.3).

4.2.1 Errors of individual budget components

Figure 4 presents the relative statistical metrics calculated using Eqs. (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.

Overall, existing BCC methods exhibit notable limitations in enhancing the accuracy of budget components. In particular, for P, nearly all statistical metrics (CC′, NSE′, MAE′, RMSE′) across various basins yield negative values. For instance, under the CKF method, these values are approximately -0.05, -0.15, -3.82, and -8.47 mm, respectively, indicating a significant reduction in the accuracy of the corrected P dataset when BCC methods are applied to enforce water budget closure. Specifically, the accuracy of the corrected P dataset decreases by approximately 6 %, 34 %, 11 %, and 55 %, as reflected in the CC, NSE, MAE, and RMSE metrics, respectively. Analysis of 13 selected basins with sufficient P observations further confirms this decline, showing a reduction in the accuracy of budget-corrected P (Fig. 5). A possible explanation for this decrease is the inherently high accuracy of raw P datasets, supported by advancements in remote sensing technologies, meteorological models, and observational networks. However, when BCC methods are applied, water imbalance errors from other bud-
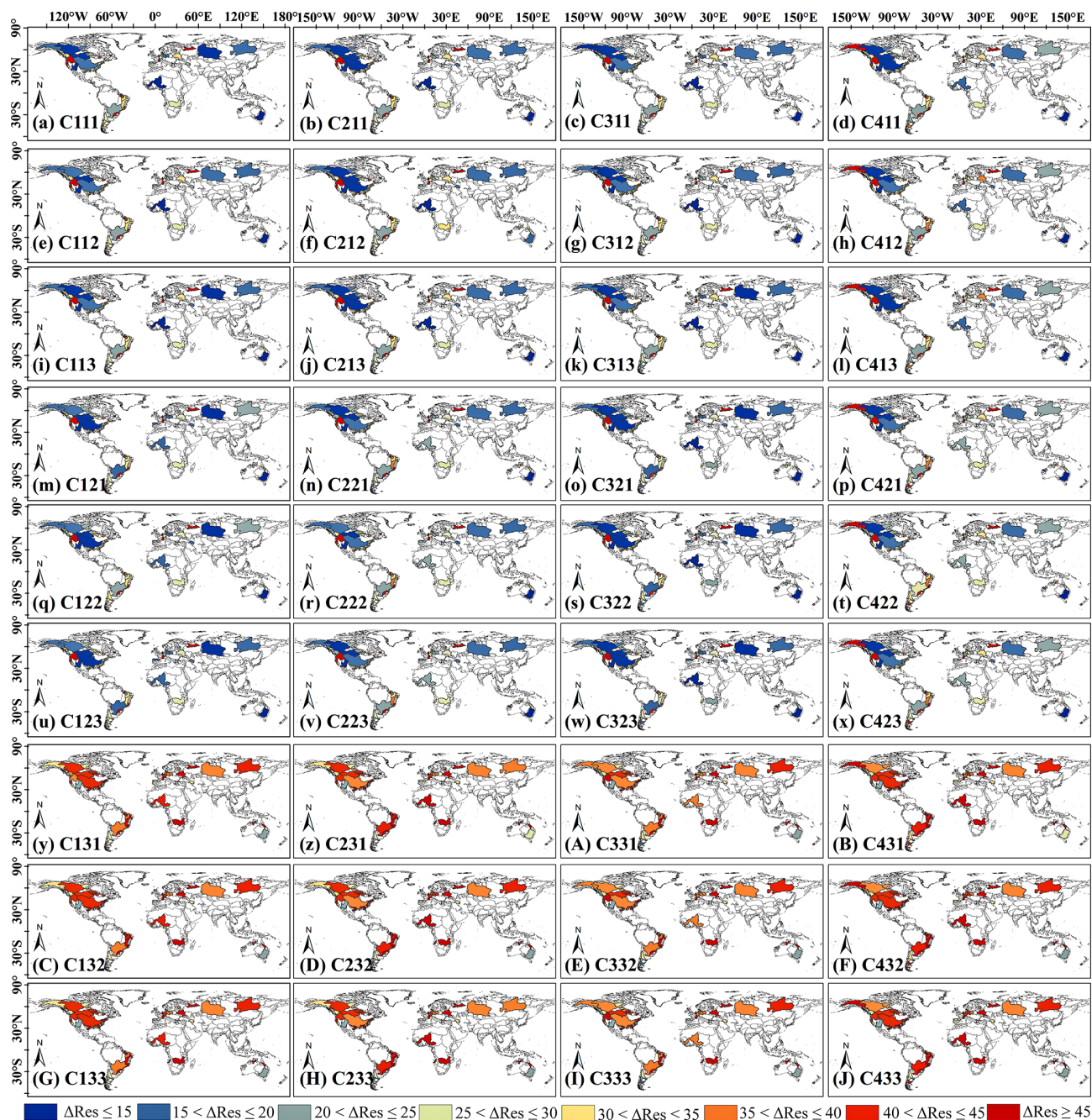
Figure 3. Spatial distribution of the ΔRes error on a monthly scale for different combinations of budget components. The unit of ΔRes is mm. Each subplot represents a distinct combination, where the first digit corresponds to the P product, the second to the ET product, and the third to the TWSC product. The detailed definitions of these combinations are provided in Eq. (3).

get components, such as ET, may be inappropriately redistributed to the corrected P dataset in an effort to enforce overall water budget closure. As a result, while the total water budget is balanced, the accuracy of the corrected P data is compromised.

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 and −20 mm. Consequently, the accuracy of corrected Q declined, with CC, NSE, MAE, and RMSE decreasing by approximately 0.1, 0.2, 3, and 5 mm, respec-
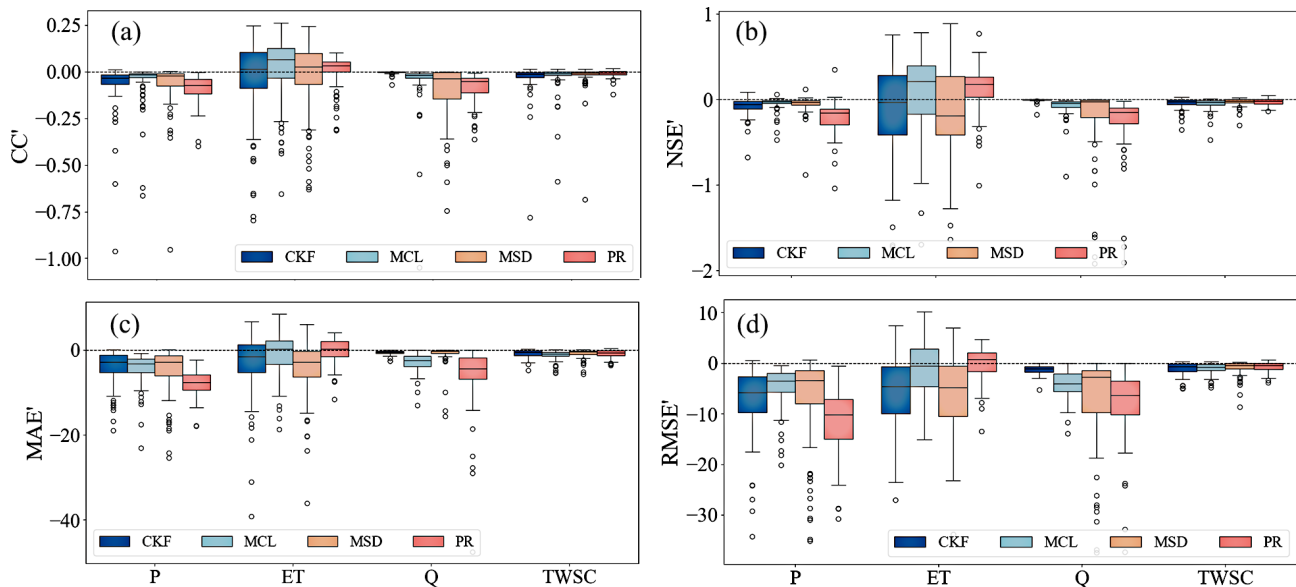
Figure 4. Box plot quantifying the errors introduced into budget components by existing BCC methods when closing the water budget. (a)–(d) represent the results of the CC′, NSE′, MAE′, RMSE′ indicators, respectively. Positive values indicate an improvement in accuracy relative to the reference values after applying existing BCC methods, while negative values indicate a decline. Different colors represent different BCC methods.

tively. 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 errors. Identifying an optimal redistribution strategy for ΔRes errors could help mitigate this issue.

4.2.2 Negative values

This section examines the occurrence of negative values in budget components arising from the application of existing BCC methods to close the water budget. For each budget component, the proportion of months with negative values relative to the total time series was computed (Fig. 6). Overall, the fraction of negative values across 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.

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 and Vogel, 2002; Zhang et al., 2008; Koster and 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.

Figure 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
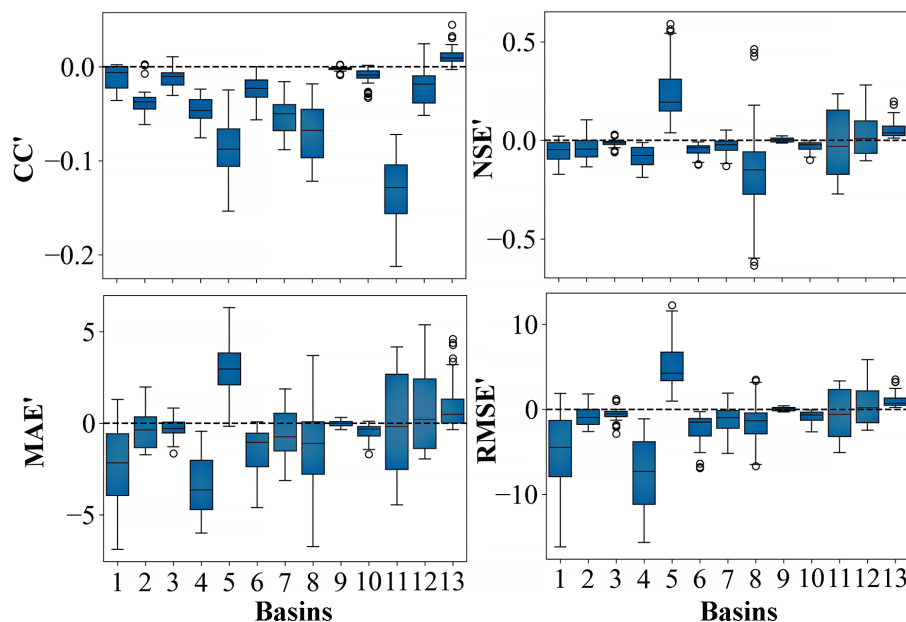
Figure 5. Box plot illustrating precipitation errors introduced by correcting ΔRes using existing BCC methods across 13 basins with sufficient observational precipitation data. The x axis represents the 13 basins in the following order: NIGER, OB, MISSISSIPPI, SACRAMENTO, SAN JOAQUIN, SUSQUEHANNA, BRAZOS, FRASER, NELSON, MURRAY, RIO EBRO, ELBE, and KURA.
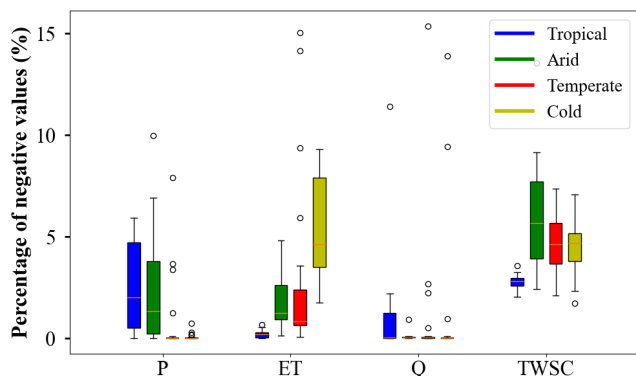

Figure 6. Percentage of negative values for corrected datasets of budget components induced by closing the water budget. Different colors indicate distinct climate classifications.

to arid regions. ET tends to show negative values more frequently in winter and spring, with a lower likelihood in summer and autumn. Except in summer, cold climate zones are most susceptible to negative ET values. Among the four budget components, Q has the lowest occurrence of negative values. Negative TWSC values exhibit no obvious seasonal pattern, with arid regions exhibiting a higher likelihood of negative values throughout the year compared to other climate types. These findings indicate that the occurrence of negative values varies significantly across seasons and climate zones. Future research should account for this seasonal variability to further refine existing BCC methods.

4.2.3 Ensemble errors

Figure 8 presents the ensemble errors in budget components (i.e., $F(\text{Re})$ in Eq. 27) introduced by existing BCC methods (CKF, MCL, MSD, and PR). All four methods exhibit similar spatial distribution patterns. Notably, high ensemble errors ($F(\text{Re}) > 10\%$) are concentrated in the northwestern basins of North America, particularly in Alaska, suggesting substantial variations in budget components in these regions. Basins with minor ensemble errors ($5\% < F(\text{Re}) \leq 10\%$) generally cover larger areas, such as African and Northern Asia. Although these errors are relatively small, they remain non-negligible. Basins with lower ensemble errors ($F(\text{Re}) \leq 5\%$) also cover some basins. Further analysis of ΔRes in basins with higher $F(\text{Re})$ values reveals a strong correlation, as these basins also exhibit larger ΔRes. This finding highlights the limitations of existing BCC methods in effectively redistributing ΔRes errors.

To determine whether the error cost introduced by existing BCC methods in closing the water budget outweighs the reduction in water imbalance error, we analyzed the relationship between the reduction in ΔRes error and the introduced budget component errors (Fig. 9). As shown in Fig. 9, with the exception of the PR method, the basins where |RAE| exceeds |Res| are largely consistent across the other three BCC methods. This discrepancy arises because the PR method redistributes ΔRes based on the magnitude of budget components, whereas the CKF, MCL, and MSD methods redistribute ΔRes according to the estimated errors in budget components.
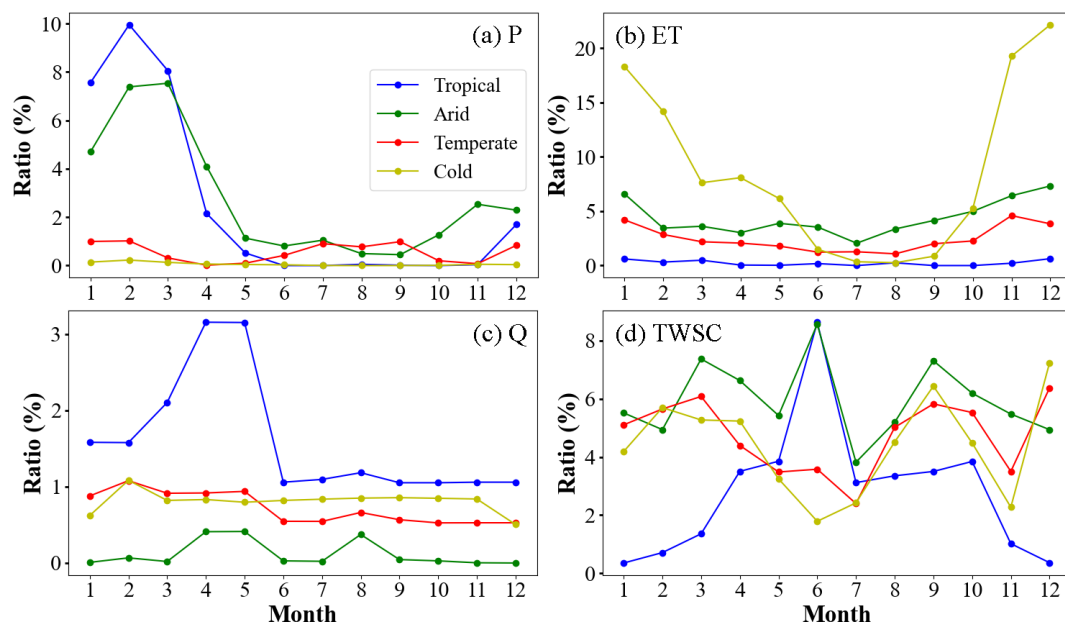
Figure 7. Seasonal cycle of the proportion of negative errors for budget components. Different colors representing various climate types. The Southern Hemisphere data by applying a 6-month shift to align its seasonal phases with those of the Northern Hemisphere.
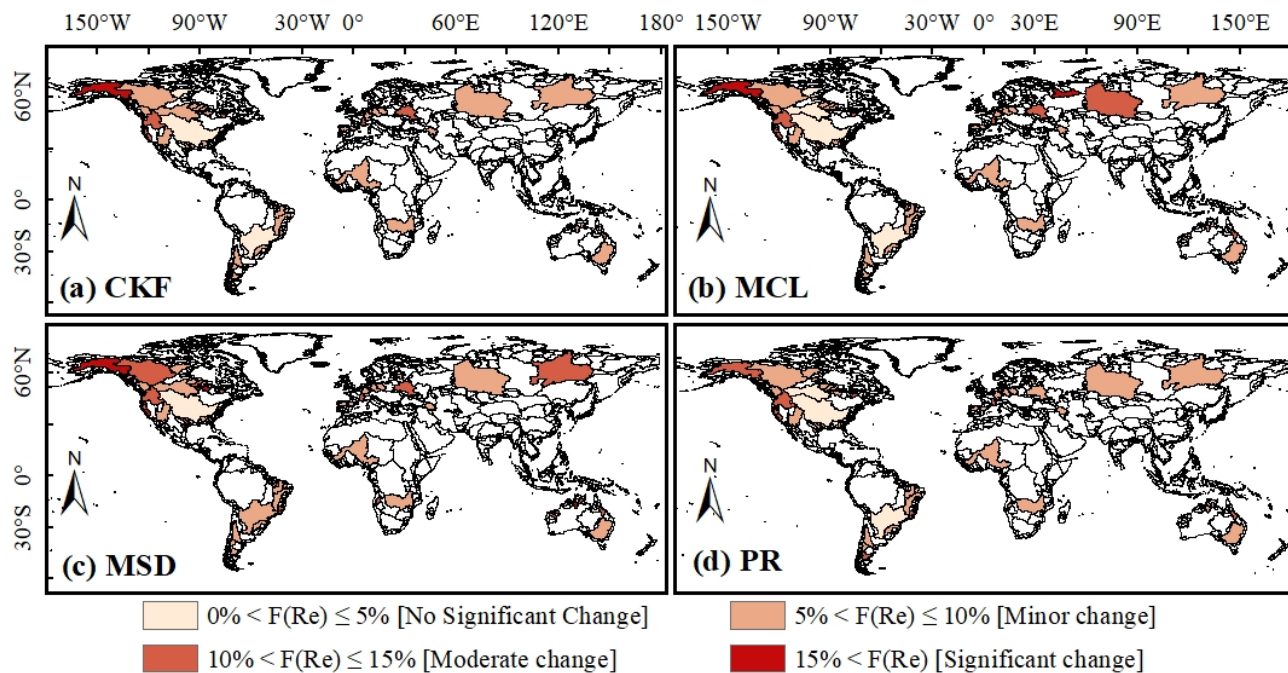
Figure 8. Ensemble errors in budget components introduced by closing the water budget using existing BCC methods.

For the CKF, MCL, MSD, and PR methods, the proportions of basins where |RAE| exceeds |Res| are 44.05 %, 52.44 %, 56.79 %, and 77.38 %, respectively. This indicates that, for all four methods, the introduced |RAE| error in budget components surpasses the reduction in water imbalance error in more than 40 % of the basins. These findings underscore the non-negligible uncertainties introduced by these methods. Striking a balance between reducing water imbalance error and minimizing the impact of budget component errors remains a critical challenge, motivating us to propose the IWE-Res method to identify optimal balance.
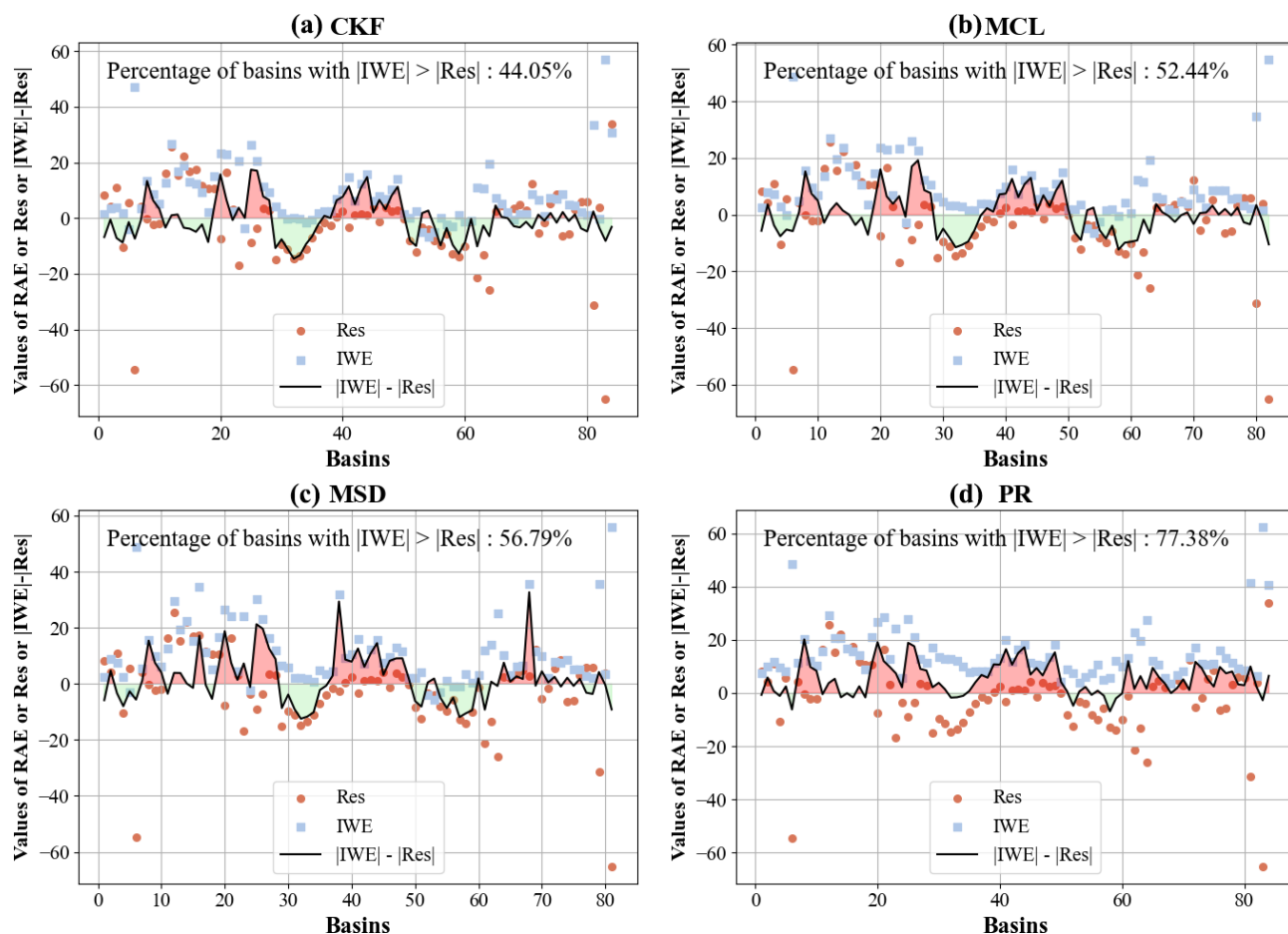
Figure 9. Comparison of relative absolute error (RAE) and residual error (Res) for four BCC methods (CKF, MCL, MSD, PR) across various basins. The black lines in the red shaded area on the upper half of the y axis indicate that the error introduced by the BCC methods for budget components exceeds the reduction in ΔRes error ($|\text{RAE}| > |\text{Res}|$), while the green shaded area on the lower half of the y axis represents cases where the error introduced is less than the reduction in ΔRes error ($|\text{RAE}| < |\text{Res}|$).

4.3 Verifying the accuracy of the proposed IWE-Res method

Based on the error analysis of existing BCC methods in Sect. 4.2, this section assesses the accuracy and reliability of the proposed IWE-Res method. The evaluation is conducted through a comparative analysis with PR, CKF, MCL, and MSD, focusing on three key aspects: the errors of individual budget components, the occurrence of negative values, and ensemble errors.

Figure 10 compares the accuracy of the proposed IWE-Res method with existing PR, CKF, MCL, and MSD methods from the perspective of errors in individual budget components. The red and blue lines represent the IWE-Res method and the existing BCC methods, respectively, while the bars indicate the relative accuracy improvement of the IWE-Res method compared to the BCC methods. As shown in Fig. 10, the proposed IWE-Res method exhibits consistently higher accuracy than all existing CKF, MCL, MSD, and PR meth-

ods 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 %.

Table 2 presents the percentage of negative values observed in the corrected budget components for the proposed IWE-Res method and existing BCC methods. One of the key contributions of the IWE-Res method is its ability to address the critical limitation of negative value generation in existing BCC methods. As a result, the percentage of negative values in the corrected P, ET, Q, and TWSC data using the proposed IWE-Res method is zero. In contrast, the
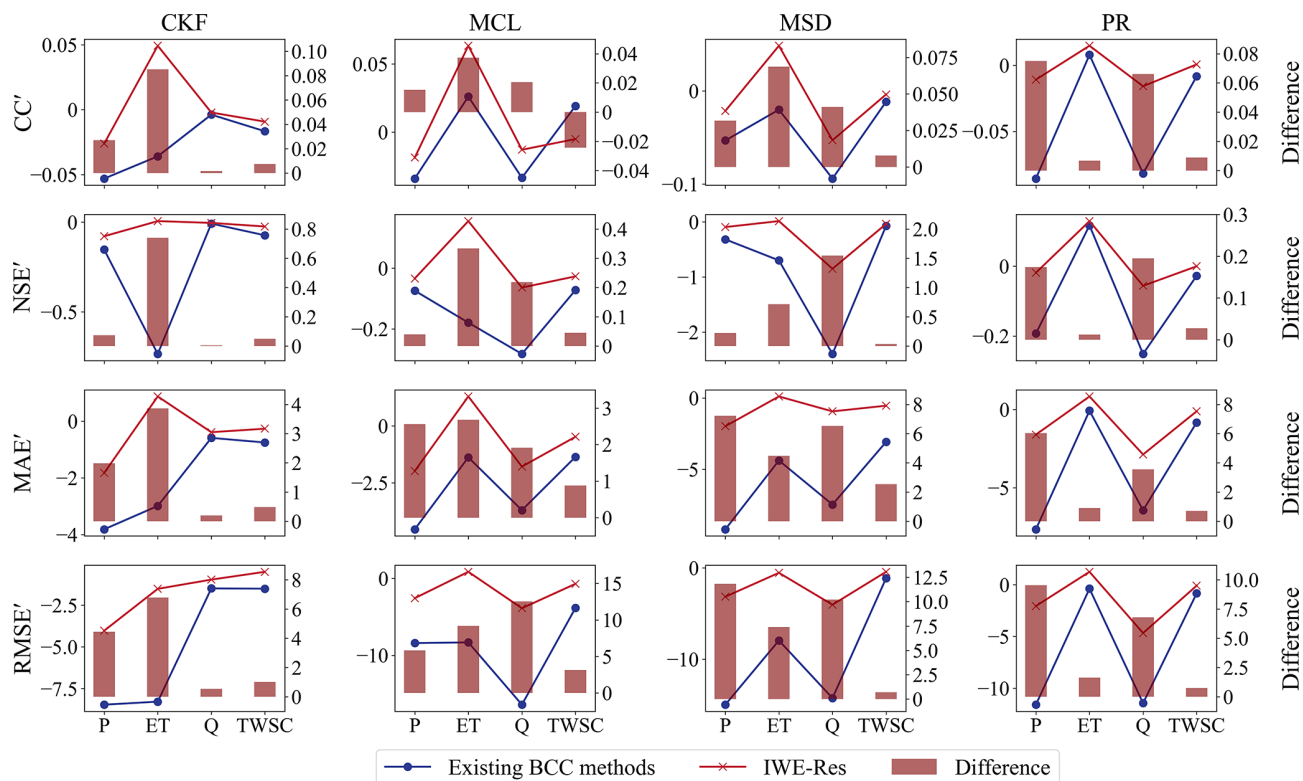
Figure 10. Performance comparison of the proposed IWE-Res method with existing BCC methods in corrected individual budget components. The red and blue lines in the figure represent the average values across all basins considered in this study.

corrected P, ET, Q, and TWSC data obtained from existing BCC methods contain negative values to varying degrees (for a detailed analysis of negative values, see Sect. 4.2.2). These results demonstrate that, in addition to improving the accuracy of budget components relative to observations, the proposed IWE-Res method effectively eliminates the issue of negative values inherent in existing BCC methods.

We further evaluate the accuracy and reliability of the proposed IWE-Res method using the ensemble error metric defined by Eq. (27) (Fig. 11), where lower values indicate better model performance. As shown in Fig. 11, the IWE-Res method significantly reduces ensemble errors compared to existing BCC methods. For 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.
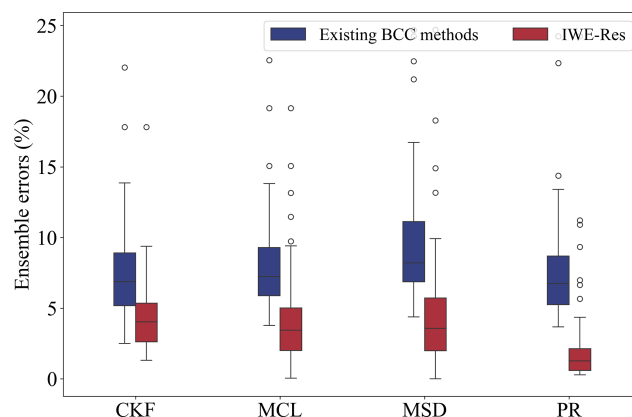
Figure 11. Performance comparison of the proposed IWE-Res method and existing BCC methods based on the ensemble errors of budget components.

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

Table 2. The percentage of months with negative values in the corrected datasets of budget components P, ET, Q, and TWSC for the proposed IWE-Res method and existing BCC methods. The percentages in the table represent the average values 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 %

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 closure. The findings offer a valuable reference for generating high-precision datasets of budget components with a closed water budget in diverse climate regions. When developing the IWE-Res method, we incorporated multiple BCC methods, each based on different principles. As a result, the identified optimal balance results vary across methods. This section presents results for the CKF method only.

Overall, the optimal balance varied among basins located in different climate zones (Figs. 12–15). In most basins within the Tropical, Arid, and Temperate zones, the optimal balance was achieved when only a portion of the water imbalance error, rather than the entire error, was redistributed to budget components. However, this pattern was not observed in the Cold region.

For most basins in the Tropical climate zone (Fig. 12), the optimal balance was reached when 40 %–90 % of ΔRes was reallocated to budget components, suggesting that the corrected budget datasets achieve their highest accuracy within this range. Notably, approximately 20 % of basins attained their optimal balance when 80 %–90 % of ΔRes was redistributed, while about 70 % did so within the 40 %–50 % range. Therefore, in Tropical basins, if sufficient observational data are unavailable to precisely determine the optimal balance, redistributing 40 %–50 % of ΔRes to budget components is recommended to obtain the most accurate dataset.

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.

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.

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.

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 im-
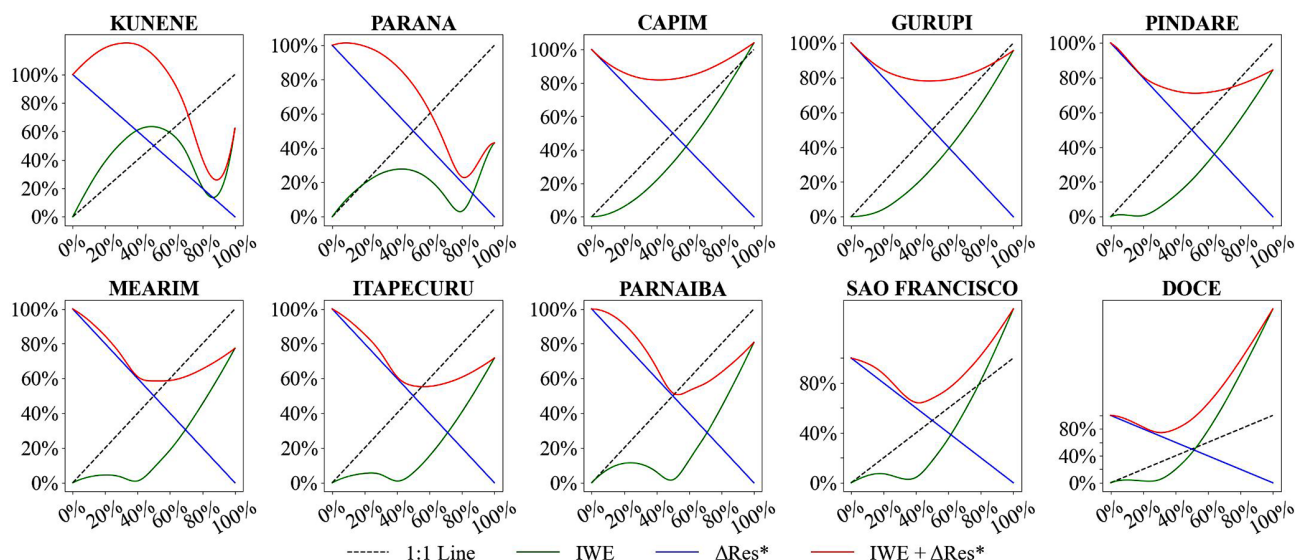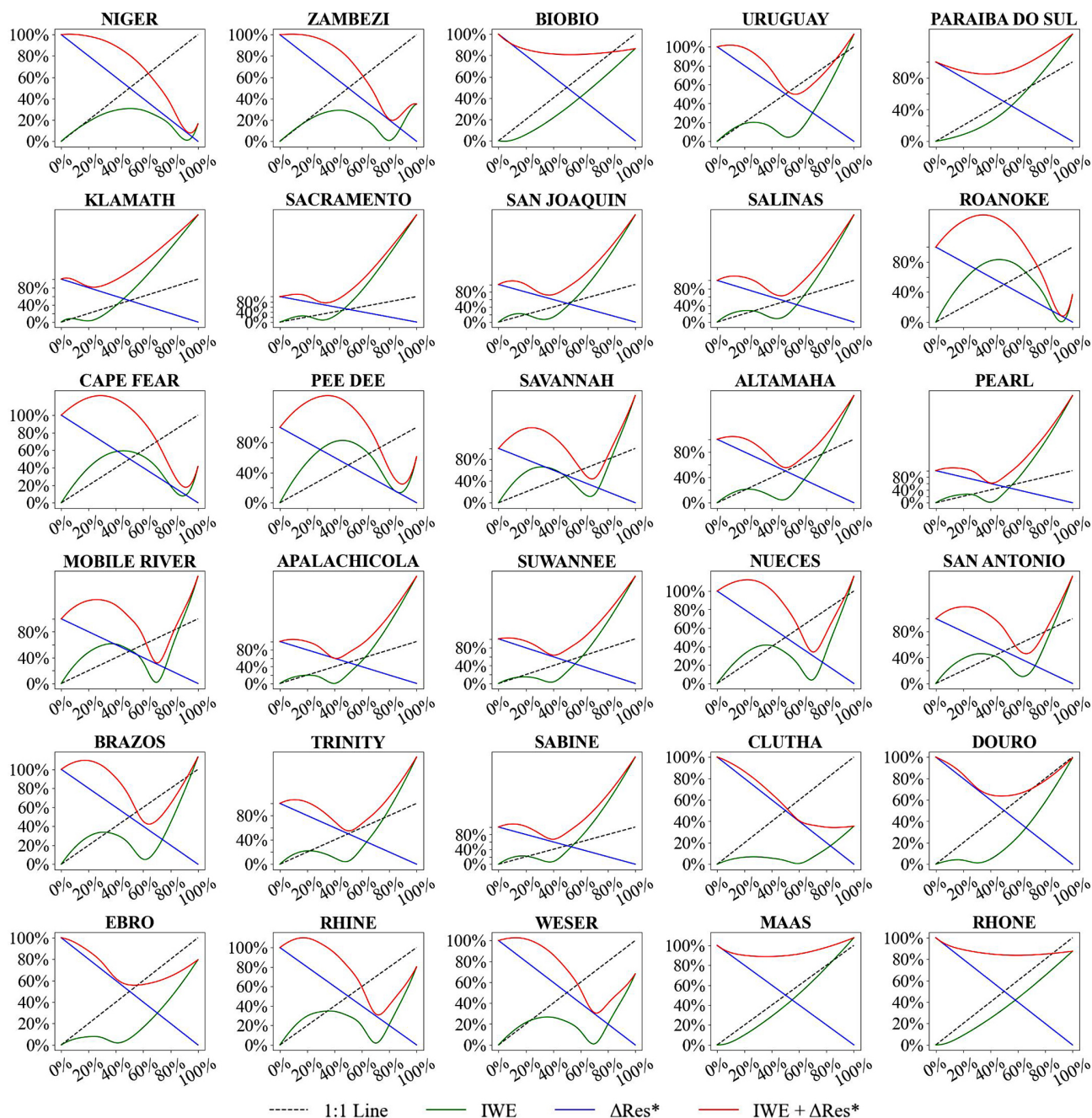
Figure 12. IWE-Res curve in basins of Tropical climate zone for identifying the optimal balance that enhances water budget closure and reduces uncertainty.
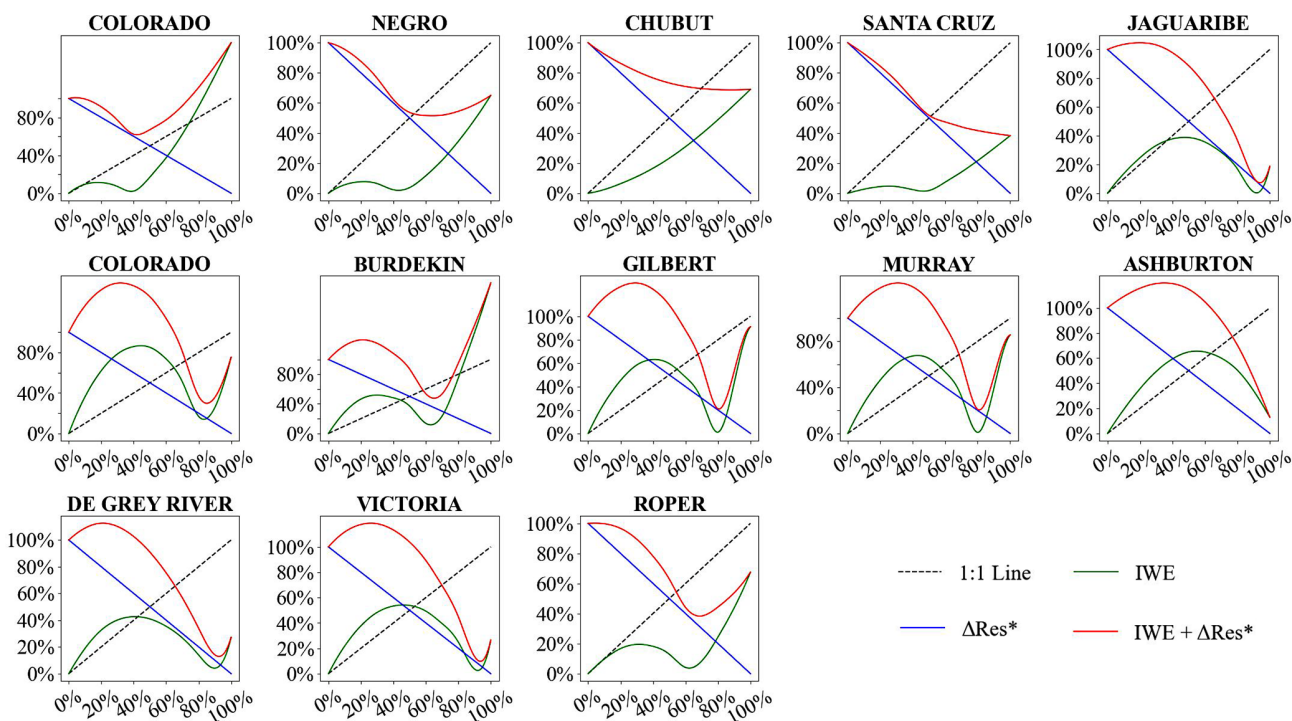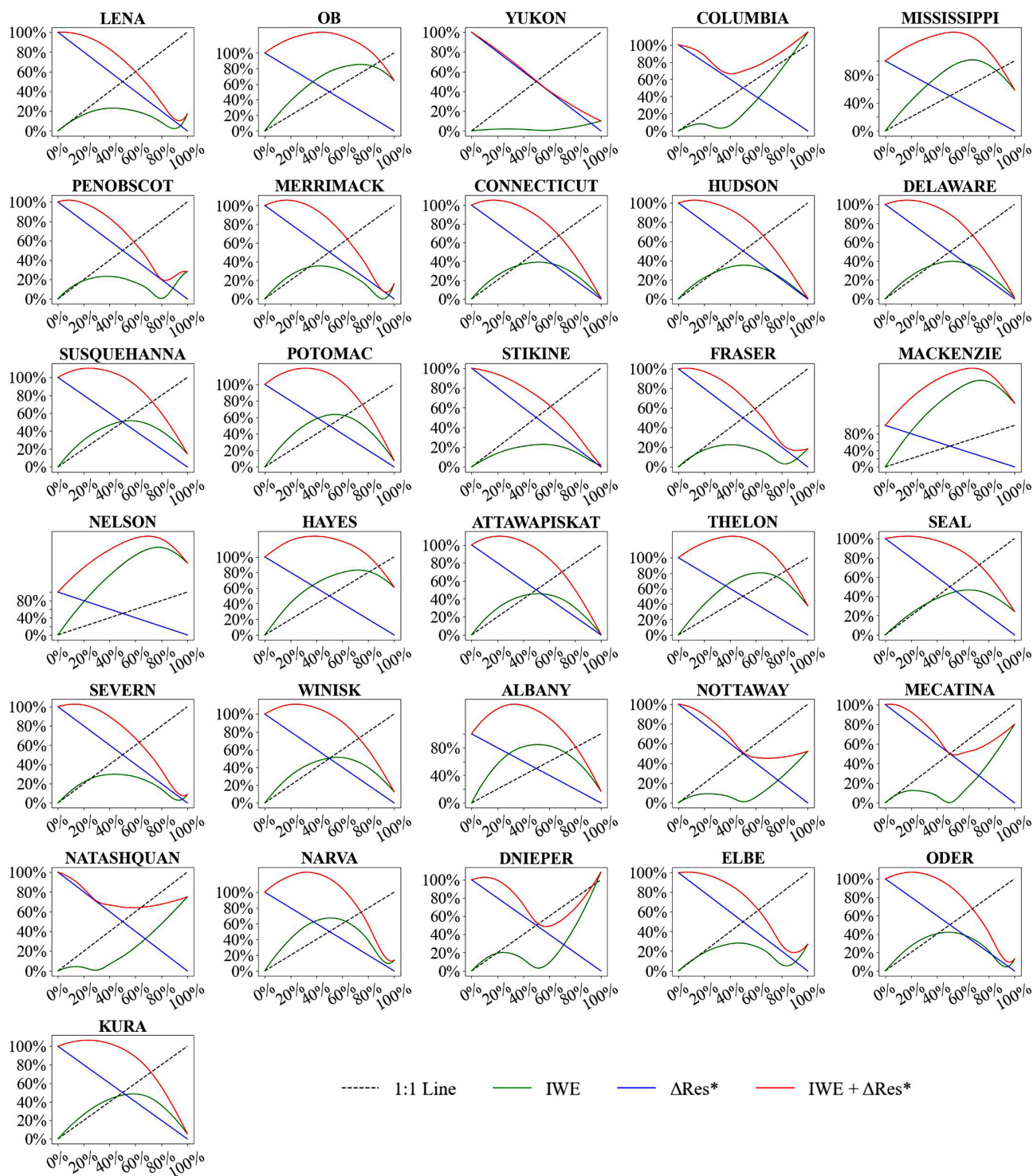

Figure 13. IWE-Res curve in watersheds of Arid climate zone for identifying the optimal balance that enhances water budget closure and reduces uncertainty.

balance errors based on these internally estimated uncertainties (Sect. 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

Figure 14. IWE-Res curve in watersheds of Temperate climate zone for identifying the optimal balance that enhances water budget closure and reduces uncertainty.

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 bal-

anced 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 wa-

Figure 15. IWE-Res curve in watersheds of Cold climate zone for identifying the optimal balance that enhances water budget closure and reduces uncertainty.

ter 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 ap-

proximate 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, 2006; Senan et al., 2022). Following many previous studies (Lehmann et al., 2022; Abolafia-Rosenzweig et al., 2021; Luo et al., 2023c; 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:

1. While applying existing BCC methods reduces water imbalance error, it simultaneously introduces new errors in budget components. For P, a decline in accuracy is observed in most basins. For Q, the corrected data exhibits lower performance than the raw data, with reductions in CC, NSE, MAE, and RMSE of approximately 0.1, 0.2, 3, and 5 mm, respectively. At the basin scale, more than 40 % of basins experience budget component errors greater than the reduction in ΔRes after applying existing BCC methods.

2. The proportion of negative corrected values in each budget component is predominantly within 0 %–5 %. For

https://doi.org/10.5194/hess-29-4607-2025

Hydrol. Earth Syst. Sci., 29, 4607–4635, 2025

ET, negative corrected values are mostly below 5 %, though they reach 9 % in cold climate regions. For P, the proportion is primarily below 5 %, with rare occurrences around 7 %. Q generally exhibits a lower proportion of negative values, mostly below 3 %. In TWSC, negative values are concentrated between 3 % and 10 %.

3. The proposed IWE-Res method improves the accuracy of corrected budget components compared to existing BCC methods. Based on RMSE, it improves the accuracy of corrected P by 29.5 %, corrected ET by 24.7 %, corrected Q by 69.0 %, and corrected TWSC by 6.8 %.

4. Except in cold regions, redistributing 40 %–90 % of ΔRes to budget components yields the optimal balance, minimizing the sum of the remaining ΔRes and the introduced budget component error. In tropical and temperate regions, the optimal balance is typically achieved when 40 %–50 % of ΔRes is redistributed. 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.

References

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

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

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

Aires, F.: Combining datasets of satellite-retrieved products. Part I: methodology and water budget closure, J. Hydrometeorol., 15, 1677–1691, https://doi.org/10.1175/JHM-D-13-0148.1, 2014.

Allen, R. G., Pereira, L. S., Raes, D., and Smith, M.: Crop evapotranspiration – Guidelines for computing crop water requirements, FAO Irrig. Drain. Pap., 56, 300 pp., D05109, https://doi.org/10.4060/cd6621en, 1998.

Aziz, K. M. A., and Rashwan, K. S.: Comparison of different resolutions of six free online DEMs with GPS elevation data on a new 6th of October City, Egypt, Arab. J. Geosci., 15, 1585, https://doi.org/10.1007/s12517-022-10845-5, 2022.

Bai, X., Wu, X., and Wang, P.: Blending long-term satellite-based precipitation data with gauge observations for drought monitoring: considering effects of different gauge densities, J. Hydrol., 577, 124007, https://doi.org/10.1016/j.jhydrol.2019.124007, 2019.

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

Beck, H. E., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A. I. J. M., Weedon, G. P., Brocca, L., Pappenberger, F., Huffman, G. J., and Wood, E. F.: Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling, Hydrol. Earth Syst. Sci., 21, 6201–6217, https://doi.org/10.5194/hess-21-6201-2017, 2017.

Beck, H. E., Pan, M., Roy, T., Weedon, G. P., Pappenberger, F., van Dijk, A. I. J. M., Huffman, G. J., Adler, R. F., and Wood, E. F.: Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS, Hydrol. Earth Syst. Sci., 23, 207–224, https://doi.org/10.5194/hess-23-207-2019, 2019a.

Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. G., Van Dijk, A. I. J. M., McVicar, T. R., and Adler, R. F.: MSWEP V2 global 3-hourly 0.1° precipitation: methodology and quantitative assessment, B. Am. Meteorol. Soc., 100, 473–500, https://doi.org/10.1175/BAMS-D-17-0138.1, 2019b.

Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., Schamm, K., Schneider, U., and Ziese, M.: 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 Syst. Sci. Data, 5, 71–99, https://doi.org/10.5194/essd-5-71-2013, 2013.

Boergens, E., Güntner, A., Sips, M., Schwatke, C., and Dobslaw, H.: Interannual variations of terrestrial water storage in the East African Rift region, Hydrol. Earth Syst. Sci., 28, 4733–4754, https://doi.org/10.5194/hess-28-4733-2024, 2024.

Bormann, H.: Impact of spatial data resolution on simulated catchment water balances and model performance of the multiscale TOPLATS model, Hydrol. Earth Syst. Sci., 10, 165–179, https://doi.org/10.5194/hess-10-165-2006, 2006.

Bouaziz, L., Weerts, A., Schellekens, J., Sprokkereef, E., Stam, J., Savenije, H., and Hrachowitz, M.: Redressing the balance: quantifying net intercatchment groundwater flows, Hydrol. Earth Syst. Sci., 22, 6415–6434, https://doi.org/10.5194/hess-22-6415-2018, 2018.

Burek, P. and Smilovic, M.: The use of GRDC gauging stations for calibrating large-scale hydrological models, Earth Syst. Sci. Data, 15, 5617–5629, https://doi.org/10.5194/essd-15-5617-2023, 2023.

Chen, J., Tapley, B. D., Rodell, M., Seo, K. W., Wilson, C. R., Scanlon, B. R., and Pokhrel, Y.: Basin-scale river runoff estimation from GRACE gravity satellites, climate models, and in situ observations: a case study in the Amazon basin, Water Resour. Res., 56, e2020WR028032, https://doi.org/10.1029/2020WR028032, 2020a.

Chen, S., Liu, B., Tan, X., and Wu, Y.: Inter-comparison of spatiotemporal features of precipitation extremes within six daily precipitation products, Clim. Dynam., 54, 1057–1076, https://doi.org/10.1007/s00382-019-05045-z, 2020b.

Chen, X., Su, Z., Ma, Y., Trigo, I. F., and Gentine, P.: Remote sensing of global daily evapotranspiration based on a surface energy balance method and reanalysis data, J. Geophys. Res.-Atmos., 126, e2020JD032873, https://doi.org/10.1029/2020JD032873, 2021.

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

Crosbie, R. S., Pollock, D. W., Mpelasoka, F. S., Barron, O. V., Charles, S. P., and Donn, M. J.: Changes in Köppen-Geiger climate types under a future climate for Australia: hydrological implications, Hydrol. Earth Syst. Sci., 16, 3341–3349, https://doi.org/10.5194/hess-16-3341-2012, 2012.

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

Dagan, G., Stier, P., and Watson-Parris, D.: Analysis of the atmospheric water budget for elucidating the spatial scale of precipitation changes under climate change, Geophys. Res. Lett., 46, 10504–10511, https://doi.org/10.1029/2019GL084776, 2019.

Dastjerdi, P. A., Ghomlaghi, A., and Nasseri, M.: A new approach to ensemble precipitation estimation: coupling satellite hydrological products with backward water balance models in large-scale, J. Hydrol., 629, 130564, https://doi.org/10.1016/j.jhydrol.2023.130564, 2024.

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

Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K., and Liebert, J.: HESS Opinions "Should we apply bias correction to global and regional climate model data?", Hydrol. Earth Syst. Sci., 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., and Velázquez-Zapata, J. A.: Performance evaluation of global precipitation datasets in northern Mexico drylands, J. Appl. Meteorol. Climatol., 63, 1545–1558, https://doi.org/10.1175/JAMC-D-23-0227.1, 2024.

Famiglietti, J. S., Lo, M., Ho, S. L., Bethune, J., Anderson, K. J., Syed, T. H., Swenson, S. C., De Linage, C. R., and Rodell, M.: Satellites measure recent rates of groundwater depletion in California's Central Valley, Geophys. Res. Lett., 38, L03403, https://doi.org/10.1029/2010GL046442, 2011.

Gao, H., Tang, Q., Ferguson, C. R., Wood, E. F., and Lettenmaier, D. P.: Estimating the water budget of major US river basins via remote sensing, Int. J. Remote Sens., 31, 3955–3978, https://doi.org/10.1080/01431161.2010.483488, 2010.

Guo, W., Hong, F., Yang, H., Huang, L., Ma, Y., Zhou, H., and Wang, H.: Quantitative evaluation of runoff variation and its driving forces based on multi-scale separation framework, J. Hydrol. Reg. Stud., 43, 101183, https://doi.org/10.1016/j.ejrh.2022.101183, 2022.

Hansford, M. R., Plink-Björklund, P., and Jones, E. R.: Global quantitative analyses of river discharge variability and hydrograph shape with respect to climate types, Earth-Sci. Rev., 200, 102977, https://doi.org/10.1016/j.earscirev.2019.102977, 2020.

Hao, X., Zhang, S., Li, W., Duan, W., Fang, G., Zhang, Y., and Guo, B.: The uncertainty of Penman-Monteith method and the energy balance closure problem, J. Geophys. Res.-Atmos., 123, 7433–7443, https://doi.org/10.1029/2018JD028371, 2018.

He, Q., Fok, H. S., Ferreira, V., Tenzer, R., Ma, Z., and Zhou, H.: Three-dimensional Budyko framework incorporating terrestrial water storage: unraveling water-energy dynamics, vegetation, and ocean-atmosphere interactions, Sci. Total Environ., 904, 166380, https://doi.org/10.1016/j.scitotenv.2023.166380, 2023.

Hsu, K.-L., Gao, X., Sorooshian, S., and Gupta, H. V.: Precipitation estimation from remotely sensed information using artificial neural networks, J. Appl. Meteorol. Climatol., 36, 1176–1190, https://doi.org/10.1175/1520-0450(1997)036<1176:PEFRSI>2.0.CO;2, 1997.

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

Huang, C., Hu, J., Chen, S., Zhang, A., Liang, Z., Tong, X., Xiao, L., Min, C., and Zhang, Z.: How well can IMERG products capture typhoon extreme precipitation events over southern China?, Remote Sens., 11, 70, https://doi.org/10.3390/rs11010070, 2019.

Huang, P., Wang, G., Guo, L., Mello, C. R., Li, K., Ma, J., and Sun, S.: Most global gauging stations present biased estimations of total catchment discharge, Geophys. Res. Lett., 50, e2023GL104253, https://doi.org/10.1029/2023GL104253, 2023.

Huffman, G. J., Bolvin, D. T., Braithwaite, D., Hsu, K., Joyce, R., Xie, P., and Yoo, S.-H.: NASA global precipitation measurement (GPM) integrated multi-satellite retrievals for GPM (IMERG), Algorithm theoretical basis document (ATBD) version 4, 26, 2020–2005, 2015.

Jacobs, A. F. G. and De Bruin, H. A. R.: Makkink's equation for evapotranspiration applied to unstressed maize, Hydrol. Process., 12, 1063–1066, https://doi.org/10.1002/(SICI)1099-1085(19980615)12:7<1063::AID-HYP640>3.0.CO;2-2, 1998.

Jiménez, C., Martens, B., Miralles, D. M., Fisher, J. B., Beck, H. E., and Fernández-Prieto, D.: Exploring the merging of the global land evaporation WACMOS-ET products based on local tower measurements, Hydrol. Earth Syst. Sci., 22, 4513–4533, https://doi.org/10.5194/hess-22-4513-2018, 2018.

Jian, J., Ryu, D., Costelloe, J. F., and Su, C. H.: Towards reliable hydrological model calibrations with river level measurements, in: 21st International Congress on Modelling and Simulation, Modelling and Simulation Society of Australia and New Zealand, Gold Coast, Australia, 29 November–4 December 2015, 2325–2331, 2015.

Kaprom, C., Williams, J. A., Mehrotra, R., Ophaphaibun, C., and Sriwongsitanon, N.: A comprehensive evaluation of the accuracy of satellite-based precipitation estimates over Thailand, J. Hydrol. Reg. Stud., 59, 102380, https://doi.org/10.1016/j.ejrh.2025.102380, 2025.

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

Krabbenhoft, C. A., Allen, G. H., Lin, P., Godsey, S. E., Allen, D. C., Burrows, R. M., DelVecchia, A. G., Fritz, K. M., Shanafield, M., Burgin, A. J., Zimmer, M. A., Datry, T., Dodds, W. K., Jones, C. N., Mims, M. C., Franklin, C., Hammond, J. C., Zipper, S., Ward, A. S., Costigan, K. H., Beck, H. E., and Olden, J. D.: Assessing placement bias of the global river gauge network, Nat. Sustain., 5, 586–592, https://doi.org/10.1038/s41893-022-00873-0, 2022.

Landerer, F. W., and Swenson, S. C.: Accuracy of scaled GRACE terrestrial water storage estimates, Water Resour. Res., 48, W04531, https://doi.org/10.1029/2011WR011453, 2012.

Landerer, F. W., Flechtner, F. M., Save, H., Webb, F. H., Bandikova, T., Bertiger, W. I., Bettadpur, S. V., Byun, S. H.,

Dahle, C., Dobslaw, H., Fahnestock, E., Harvey, N., Kang, Z., Kruizinga, G. L. H., Loomis, B. D., McCullough, C., Murböck, M., Nagel, P., Paik, M., Pie, N., Poole, S., Strekalov, D., Tamisiea, M. E., Wang, F., Watkins, M. M., Wen, H.-Y., Wiese, D. N., and Yuan, D. N.: Extending the global mass change data record: GRACE Follow-On instrument and science data performance, Geophys. Res. Lett., 47, e2020GL088306, https://doi.org/10.1029/2020GL088306, 2020.

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

Levison, J., Larocque, M., Ouellet, M. A., Ferland, O., and Poirier, C.: Long-term trends in groundwater recharge and discharge in a fractured bedrock aquifer–past and future conditions, Can. Water Resour. J., 41, 500–514, https://doi.org/10.1080/07011784.2015.1037795, 2016.

Li, L., Dai, Y., Wei, Z., Wei, S., Zhang, Y., Wei, N., and Li, Q.: Enforcing water balance in multitask deep learning models for hydrological forecasting, J. Hydrometeorol., 25, 89–103, https://doi.org/10.1175/JHM-D-23-0073.1, 2024.

Liu, J., Zhang, Q., Singh, V. P., and Shi, P.: Contribution of multiple climatic variables and human activities to streamflow changes across China, J. Hydrol., 545, 145–162, https://doi.org/10.1016/j.jhydrol.2016.12.016, 2017.

Liu, X., Yang, K., Ferreira, V. G., and Bai, P.: Hydrologic model calibration with remote sensing data products in global large basins, Water Resour. Res., 58, e2022WR032929, https://doi.org/10.1029/2022WR032929, 2022.

Lockhoff, M., Zolina, O., Simmer, C., and Schulz, J.: Evaluation of satellite-retrieved extreme precipitation over Europe using gauge observations, J. Climate, 27, 607–623, https://doi.org/10.1175/JCLI-D-13-00194.1, 2014.

Loomis, B. D., Rachlin, K. E., Wiese, D. N., Landerer, F. W., and Luthcke, S. B.: Replacing GRACE/GRACE-FO with satellite laser ranging: Impacts on Antarctic Ice Sheet mass change, Geophys. Res. Lett., 47, e2019GL085488, https://doi.org/10.1029/2019GL085488, 2020.

Lute, A. C., and Abatzoglou, J. T.: Role of extreme snowfall events in interannual variability of snowfall accumulation in the western United States, Water Resour. Res., 50, 2874–2888, https://doi.org/10.1002/2013WR014465, 2014.

Luo, Z., Gao, Z., Wang, L., Wang, S., and Wang, L.: A method for balancing the terrestrial water budget and improving the estimation of individual budget components, Agric. For. Meteorol., 341, 109667, https://doi.org/10.1016/j.agrformet.2023.109667, 2023a.

Luo, Z., Li, H., Zhang, S., Wang, L., Wang, S., and Wang, L.: A novel two-step method for enforcing water budget closure and an intercomparison of budget closure correction methods based on satellite hydrological products, Water Resour. Res., 59, e2022WR032176, https://doi.org/10.1029/2022WR032176, 2023b.

Luo, Z., Yu, H., Liu, H., and Chen, J.: Assessing the water budget closure accuracy of satellite/reanalysis-based hydrological data products over mainland China, Remote Sens., 15, 5230, https://doi.org/10.3390/rs15215230, 2023c.

Lv, M., Ma, Z., Yuan, X., Lv, M., Li, M., and Zheng, Z.: Water budget closure based on GRACE measurements and recon-

structed evapotranspiration using GLDAS and water use data for two large densely-populated mid-latitude basins, J. Hydrol., 547, 585–599, https://doi.org/10.1016/j.jhydrol.2017.02.027, 2017.

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

Masunaga, H., Schröder, M., Furuzawa, F. A., Kummerow, C., Rustemeier, E., and Schneider, U.: Inter-product biases in global precipitation extremes, Environ. Res. Lett., 14, 125016, https://doi.org/10.1088/1748-9326/ab5da9, 2019.

McMahon, T. A., Finlayson, B. L., and Peel, M. C.: Historical developments of models for estimating evaporation using standard meteorological data, WIREs Water, 3, 788–818, https://doi.org/10.1002/wat2.1172, 2016.

Mehrnegar, N., Schumacher, M., Jagdhuber, T., and Forootan, E.: Making the best use of GRACE, GRACE-FO and SMAP data through a constrained Bayesian data-model integration, Water Resour. Res., 59, e2023WR034544, https://doi.org/10.1029/2023WR034544, 2023.

Miralles, D. G., Holmes, T. R. H., De Jeu, R. A. M., Gash, J. H., Meesters, A. G. C. A., and Dolman, A. J.: Global land-surface evaporation estimated from satellite-based observations, Hydrol. Earth Syst. Sci., 15, 453–469, https://doi.org/10.5194/hess-15-453-2011, 2011.

Mourad, R., Schoups, G., Bastiaanssen, W., and Kumar, D. N.: Expert-based prior uncertainty analysis of gridded water balance components: application to the irrigated Hindon River Basin, India, J. Hydrol. Reg. Stud., 55, 2214–5818, https://doi.org/10.1016/j.ejrh.2024.101935, 2024.

Mu, D., Yan, H., Feng, W., and Peng, P.: GRACE leakage error correction with regularization technique: case studies in Greenland and Antarctica, Geophys. J. Int., 208, 1775–1786, https://doi.org/10.1093/gji/ggw494, 2017.

Mueller, B., Seneviratne, S. I., Jimenez, C., Corti, T., Hirschi, M., Balsamo, G., Ciais, P., Dirmeyer, P., Fisher, J. B., Guo, Z., Jung, M., Maignan, F., McCabe, M. F., Reichle, R., Reichstein, M., Rodell, M., Sheffield, J., Teuling, A. J., Wang, K., Wood, E. F., and Zhang, Y.: Evaluation of global observations-based evapotranspiration datasets and IPCC AR4 simulations, Geophys. Res. Lett., 38, L06402, https://doi.org/10.1029/2010GL046230, 2011.

Mueller, D. S.: 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, IEEE, 30–34, https://doi.org/10.1109/CCM.2003.1194278, 2003.

Munier, S. and Aires, F.: A new global method of satellite dataset merging and quality characterization constrained by the terrestrial water budget, Remote Sens. Environ., 205, 119–130, https://doi.org/10.1016/j.rse.2017.11.008, 2018.

Nassaj, B. N., Zohrabi, N., Shahbazi, A. N., and Fathian, H.: Evaluating the performance of eight global gridded precipitation datasets across Iran, Dyn. Atmos. Oceans, 98, 101297, https://doi.org/10.1016/j.dynatmoce.2022.101297, 2022.

Palharini, R. S. A., Vila, D. A., Rodrigues, D. T., Quispe, D. P., Palharini, R. C., de Siqueira, R. A., and de Sousa Afonso, J. M.: Assessment of the extreme precipitation by satel-

lite estimates over South America, Remote Sens., 12, 2085, https://doi.org/10.3390/rs12132085, 2020.

Pan, M., Sahoo, A. K., Troy, T. J., Vinukollu, R. K., Sheffield, J., and Wood, E. F.: Multisource estimation of long-term terrestrial water budget for major global river basins, J. Climate, 25, 3191–3206, https://doi.org/10.1175/JCLI-D-11-00300.1, 2012.

Pan, S., Pan, N., Tian, H., Friedlingstein, P., Sitch, S., Shi, H., Arora, V. K., Haverd, V., Jain, A. K., Kato, E., Lienert, S., Lombardozzi, D., Nabel, J. E. M. S., Ottlé, C., Poulter, B., Zaehle, S., and Running, S. W.: Evaluation of global terrestrial evapotranspiration using state-of-the-art approaches in remote sensing, machine learning and land surface modeling, Hydrol. Earth Syst. Sci., 24, 1485–1509, https://doi.org/10.5194/hess-24-1485-2020, 2020.

Papacharalampous, G., Tyralis, H., Markonis, Y., Máca, P., and Hanel, M.: Features of the Earth's seasonal hydroclimate: characterizations and comparisons across the Köppen–Geiger climates and across continents, Prog. Earth Planet. Sci., 10, 46, https://doi.org/10.1186/s40645-023-00574-y, 2023.

Park, J. and Choi, M.: Estimation of evapotranspiration from ground-based meteorological data and global land data assimilation system (GLDAS), Stoch. Environ. Res. Risk Assess., 29, 1963–1992, https://doi.org/10.1007/s00477-014-1004-2, 2015.

Pellet, V., Aires, F., Papa, F., Munier, S., and Decharme, B.: Long-term total water storage change from a Satellite Water Cycle reconstruction over large southern Asian basins, Hydrol. Earth Syst. Sci., 24, 3033–3055, https://doi.org/10.5194/hess-24-3033-2020, 2020.

Peltier, W. R., Drummond, R., and Roy, K.: Comment on "Ocean mass from GRACE and glacial isostatic adjustment" by DP Chambers et al., J. Geophys. Res.-Sol. Ea., 117, B11, https://doi.org/10.1029/2011JB008967, 2012.

Petković, V., and Kummerow, C. D.: Understanding the sources of satellite passive microwave rainfall retrieval systematic errors over land, J. Appl. Meteorol. Climatol., 56, 597–614, https://doi.org/10.1175/JAMC-D-16-0174.1, 2017.

Potter, N. J., Zhang, L., Milly, P. C. D., McMahon, T. A., and Jakeman, A. J.: Effects of rainfall seasonality and soil moisture capacity on mean annual water balance for Australian catchments, Water Resour. Res., 41, 6, https://doi.org/10.1029/2004WR003697, 2005.

Reager, J. T., Thomas, B. F., and Famiglietti, J. S.: River basin flood potential inferred using GRACE gravity observations at several months lead time, Nat. Geosci., 7, 588–592, https://doi.org/10.1038/ngeo2203, 2014.

Resende, T. C., Longuevergne, L., Gurdak, J. J., Leblanc, M., Favreau, G., Ansems, N., Van der Gun, J., Gaye, C. B., and Aureli, A.: Assessment of the impacts of climate variability on total water storage across Africa: implications for groundwater resources management, Hydrogeol. J., 27, 493–512, https://doi.org/10.1007/s10040-018-1864-5, 2019.

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

Ruhoff, A., de Andrade, B. C., Laipelt, L., Fleischmann, A. S., Siqueira, V. A., Moreira, A. A., Barbedo, R., Cyganski, G. L., Fernandez, G. M. R., Brêda, J. P. L. F., Paiva, R. C. D., Meller, A., Teixeira, A. A., Araújo, A. A., Fuckner, M. A., and Biggs, T.: Global evapotranspiration datasets assessment us-

ing water balance in South America, Remote Sens., 14, 2526, https://doi.org/10.3390/rs14112526, 2022.

Sadeghi, M., Akbari Asanjan, A., Faridzad, M., Afzali Gorooh, V., Nguyen, P., Hsu, K., Sorooshian, S., and Braithwaite, D.: 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 Sens., 11, 2755, https://doi.org/10.3390/rs11232755, 2019.

Sahoo, A. K., Pan, M., Troy, T. J., Vinukollu, R. K., Sheffield, J., and Wood, E. F.: Reconciling the global terrestrial water budget using satellite remote sensing, Remote Sens. Environ., 115, 1850–1865, https://doi.org/10.1016/j.rse.2011.03.009, 2011.

Sankarasubramanian, A. and Vogel, R. M.: Annual hydroclimatology of the United States, Water Resour. Res., 38, 19-1–19-12, https://doi.org/10.1029/2001WR000619, 2002.

Schiavo, M.: The role of different sources of uncertainty on the stochastic quantification of subsurface discharges in heterogeneous aquifers, J. Hydrol., 617, 128930, https://doi.org/10.1016/j.jhydrol.2022.128930, 2023.

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

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

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

Shiklomanov, A. I., Yakovleva, T. I., Lammers, R. B., Karasev, I. P., Vörösmarty, C. J., and Linder, E.: Cold region river discharge uncertainty – Estimates from large Russian rivers, J. Hydrol., 326, 231–256, https://doi.org/10.1016/j.jhydrol.2005.10.037, 2006.

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

Swenson, S. and Wahr, J.: Post-processing removal of correlated errors in GRACE data, Geophys. Res. Lett., 33, L08402, https://doi.org/10.1029/2005GL025285, 2006.

Tan, X., Liu, B., Tan, X., and Chen, X.: Long-term water imbalances of watersheds resulting from biases in hydroclimatic data sets for water budget analyses, Water Resour. Res., 58, e2021WR031209, https://doi.org/10.1029/2021WR031209, 2022.

Tapley, B. D., Bettadpur, S., Ries, J. C., Thompson, P. F., and Watkins, M. M.: GRACE measurements of mass variability in the Earth system, Science, 305, 503–505, https://doi.org/10.1126/science.1099192, 2004.

Trenberth, K. E., Dai, A., Van Der Schrier, G., Jones, P. D., Barichivich, J., Briffa, K. R., and Sheffield, J.: Global warming and changes in drought, Nat. Clim. Change, 4, 17–22, https://doi.org/10.1038/nclimate2067, 2014.

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

Wang, S., McKenney, D. W., Shang, J., and Li, J.: A national-scale assessment of long-term water budget closures for Canada's watersheds, J. Geophys. Res. Atmos., 119, 8712–8725, https://doi.org/10.1002/2014JD021951, 2014.

Wang, S., Huang, J., Yang, D., Pavlic, G., and Li, J.: Long-term water budget imbalances and error sources for cold region drainage basins, Hydrol. Process., 29, 2125–2136, https://doi.org/10.1002/hyp.10343, 2015.

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

Watkins, M. M., Wiese, D. N., Yuan, D. N., Boening, C., and Landerer, F. W.: Improved methods for observing Earth's time variable mass distribution with GRACE using spherical cap mascons, J. Geophys. Res.-Sol. Ea., 120, 2648–2671, https://doi.org/10.1002/2014JB011547, 2015.

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

Wu, C., Yeh, P. J. F., Hu, B. X., and Huang, G.: Controlling factors of errors in the predicted annual and monthly evaporation from the Budyko framework, Adv. Water Resour., 121, 432–445, https://doi.org/10.1016/j.advwatres.2018.09.013, 2018.

Wu, Z., Zhang, Y., Sun, Z., Lin, Q., and He, H.: Improvement of a combination of TMPA (or IMERG) and ground-based precipitation and application to a typical region of the East China Plain, Sci. Total Environ., 640, 1165–1175, https://doi.org/10.1016/j.scitotenv.2018.05.272, 2018.

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

Yao, Y., Liang, S., Li, X., Hong, Y., Fisher, J. B., Zhang, N., Chen, J., Cheng, J., Zhao, S., Zhang, X., Jiang, B., Sun, L., Jia, K., Wang, K., Chen, Y., Mu, Q., and Feng, F.: Bayesian multimodel estimation of global terrestrial latent heat flux from eddy covariance, meteorological, and satellite observations, J. Geophys. Res.-Atmos., 119, 4521–4545, https://doi.org/10.1002/2013JD020864, 2014.

Yeh, P. J. F., Swenson, S. C., Famiglietti, J. S., and Rodell, M.: Remote sensing of groundwater storage changes in Illinois using the Gravity Recovery and Climate Experiment (GRACE), Water Resour. Res., 42, W12203, https://doi.org/10.1029/2006WR005374, 2006.

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

Zhang, L., Potter, N., Hickel, K., Zhang, Y., and Shao, Q.: Water balance modeling over variable time scales based on the Budyko

framework–Model development and testing, J. Hydrol., 360, 117–131, https://doi.org/10.1016/j.jhydrol.2008.07.021, 2008.

Zhang, Y., Pan, M., and Wood, E. F.: On creating global gridded terrestrial water budget estimates from satellite remote sensing, in: Remote Sensing and Water Resources, edited by: Gebremichael, M., Springer, 59–78, https://doi.org/10.1007/978-3-319-32449-4_4, 2016.

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

Zhou, L., Cao, Y., Shi, C., Liang, H., and Fan, L.: Quantifying the Atmospheric Water Balance Closure over Mainland China Using Ground-Based, Satellite, and Reanalysis Datasets, Atmosphere, 15, 497, https://doi.org/10.3390/atmos15040497, 2024.