Response to Review
Reviewer #1:

The manuscript is focused on numerical techniques to 'distribute' residuals of a water balance equation over the contributing terms, avoiding negative values. The manuscript seems focused on the numerical techniques, with limited efforts for a hydrological interpretation. It could be more convincing if authors bring a bit more on the 'explanation' side.

Response: Thank you for your valuable comment. We agree that the manuscript, in its current form, lacks sufficient hydrological interpretation regarding both the redistribution of residuals across water budget components and the prevention of negative values during this process. To address this concern, we revised the manuscript from the following aspects: (1) We provided a hydrological explanation for the existence of residuals in current water balance datasets and emphasized the importance of reducing these residuals; (2) We added a hydrological interpretation of the limitations associated with existing correction methods that fully redistribute the residuals in order to achieve water balance closure; (3) We added more hydrological interpretation to clarify the rationale and validity of the proposed method in this study; (4) Following one of your suggestions below, we conducted a comparison between the monthly-scale results of our proposed method and the annual-scale estimates derived from the physically based Budyko framework. Our specific modifications are as follows:

(1) We have added a hydrological explanation of the origins of residuals in existing water balance datasets and emphasized the importance of minimizing these residuals. The terrestrial water balance, composed of four major hydrological variables—precipitation (P), evapotranspiration (ET), streamflow (Q), and terrestrial water storage change (TWSC)—describes the distribution and movement of water across the Earth's land surface. Notably, these fluxes and storages are dynamically linked, reflecting the actual hydrological processes within a basin. For example, in real-world hydrology, precipitation transforms into ET, Q, and TWSC through physical processes such as infiltration, surface runoff, and evaporation. Therefore, improving the consistency (i.e., closure) among P, ET, Q, and TWSC is essential for advancing our understanding of watershed-scale hydrological dynamics.

However, the data used for water components, whether through direct observations or hydrological modeling, are often imbalanced. Observationally, each component is typically measured independently, and there is a lack of integrated observational systems which are capable of capturing all components simultaneously. From a modeling perspective, structural simplifications, parameter uncertainties, and observational errors in input data often propagate through the system, preventing budget closure.

Unfortunately, achieving a closed water balance remains challenging. In the era of big data, the rapid expansion of remote sensing and reanalysis datasets has significantly improved the availability of hydrological data. However, these datasets also suffer from non-closure in the water balance, mainly because they are independently produced and not physically coupled, making it difficult to reconcile their inconsistencies through modeling. This underscores the urgent need for data-driven approaches to enhance the internal consistency of these datasets—especially since the water cycle variables are inherently interconnected. To address this issue, numerous water budget closure correction methods have been proposed in existing studies.

We acknowledge that our initial manuscript did not provide sufficient hydrological explanation regarding the origin of residuals or the significance of reducing them. To address this, we have added the following paragraph to the revised manuscript:

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

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

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 balance components at the required resolution and accuracy. Each variable is typically derived from independent data sources 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 precipitation 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 precipitation inputs (Wang et al., 2017; Bai et al., 2019; Wu et al., 2018). Satellite-based precipitation 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 using empirical or physically based models (Jacobs et al., 1998; McMahon et al., 2016; Allen et al., 1998). Although these models are generally well calibrated, uncertainties persist due to the complex influence of advection and localized meteorological variability, especially in small catchments. At larger spatial scales, energy balance approaches tend to provide sufficiently accurate estimates (Hua et al., 2020; Hao et al., 2018; Ruhoff et al., 2022). Q measurements typically exhibit low uncertainty when rating curves are well established and regularly maintained (Jian et al., 2015; Krabbenhoft et al., 2022). However, uncertainty can still arise from the delineation of watershed boundaries, particularly in regions where groundwater flow does not align with surface catchment divides (Huang et al., 2023; Bouaziz et al., 2018). This mismatch can result in misrepresentation of actual hydrological contributions. TWSC generally has a negligible impact on water balance calculations over multi-year periods, but can significantly affect short-term (e.g., daily) balances (He et al., 2023; Zhang et al., 2016). A key challenge is defining 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 balance closure using hydrological modeling alone.

In recent years, the rapid expansion of remote sensing and reanalysis datasets has significantly improved global access to water cycle variables, offering new opportunities for data-driven analysis of hydrological processes.".

(2) Hydrological interpretation of the limitations associated with existing correction methods that fully redistribute the residuals.

Existing water budget closure correction (BCC) methods commonly redistribute the entire residual error (ΔRes) among water budget components to enforce strict closure. In the revised manuscript, we explain the limitations of this full redistribution approach from two perspectives: the hydrological origins of ΔRes and the principles underlying the redistribution process used in current BCC methods. ΔRes is a composite error term that includes estimation errors in budget components, systematic biases, and the omission of unmeasured components. However, existing BCC methods typically assume that all of ΔRes arises from estimation errors and then redistribute it according to estimation errors. This assumption oversimplifies the true hydrological causes of imbalance and can lead to unreasonable outcomes—most notably, the appearance of negative values in corrected datasets. We have added the following sentence to the revised manuscript:

"To address this issue, a growing number of studies have adopted water budget closure correction (BCC) methods to reduce water imbalance errors (ΔRes), with the goal of forcing ΔRes from a non-zero value ($\Delta Res \neq 0$) to theoretical closure ($\Delta Res = 0$), where $\Delta Res = P - ET - Q - TWSC$ (Zhou et al., 2024; Munier et al., 2018; Zhang et al., 2016). Common methods include Proportional Redistribution (PR), the Constrained Kalman Filter (CKF), Multiple Collocation (MCL), and the Minimized Series Deviation (MSD) method (Pan et al., 2012; Luo et al., 2023)."

and

"Existing BCC methods redistribute the entire ΔRes error among water budget components to enforce strict water budget closure. This redistribution is typically guided by the relative uncertainties of the individual components, based on the assumption that the entire residual error originates from observational or modeling errors in these datasets. However, this assumption overlooks the fact that ΔRes is not solely the result of measurement or estimation errors in P, ET, Q, or TWSC. Rather, it is a composite residual that also reflects contributions from systematic biases and the omission of unmeasured components. These include deep groundwater exchanges that may cross basin boundaries, snow and glacier storage changes (particularly in high-altitude or high-latitude regions), and anthropogenic influences such as irrigation withdrawals, reservoir operations, and inter-basin water transfers. Because existing BCC methods do not explicitly account for these additional sources of imbalance, forcing strict closure by allocating the entire ΔRes to the measured components can introduce unrealistic uncertainties. As a result, the application of existing BCC methods—despite their goal of improving internal consistency—often leads to limited improvements, or, in some cases, even a decline in the accuracy of the corrected hydrological datasets.

A clear manifestation of this limitation is the occurrence of negative values in corrected budget component datasets when applying existing BCC methods at the monthly scale, such as negative P, ET, and Q. These unrealistic negative values arise when an excessive share of the ΔRes is redistributed to specific components. For instance, if the BCC method overestimates the error in a specific component, it may assign an excessively large portion of ΔRes to that component. When the magnitude of the correction exceeds the component's original value, the result is a negative flux, which is hydrologically incorrect. Beyond introducing negative values, such imbalanced redistribution compromises the integrity of the remaining components. Overcorrecting one variable necessarily reduces the share of ΔRes available for others, potentially degrading their accuracy. Our previous work demonstrated that enforcing water budget closure can, to some extent, reduce the accuracy of individual components and tends to introduce an ET regulation factor to mitigate accuracy loss in ET caused by existing BCC methods (Luo et al., 2023). A more hydrologically sound approach may involve partial closure, whereby only the portion of ΔRes attributable to quantified uncertainties is redistributed, while the residual linked to unmeasured processes is retained.".

(3) We have added further hydrological explanation to justify the design and effectiveness of our proposed IWE-Res method:

"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 residual can be attributed to known components. Some portion of ΔRes may originate from unmeasured or poorly constrained processes. By partially closing the water budget in a controlled and iterative manner, the IWE-Res method reduces the risk of overcorrecting well-characterized components while better preserving the consistency of the entire budget.".

(4) A comparison with the annual-scale results from the physically based Budyko model indicates that our proposed method produces similar outcomes. Both approaches suggest that the water balance is primarily influenced by P and potential ET (Sankarasubramanian & Vogel, 2002; Zhang et al., 2008; Koster & Suarez, 1999). These influences, however, differ across climatic regions. For instance, in tropical and arid zones, P tends to be the dominant factor (Du et al., 2024; Wu et al., 2018; Liu et al., 2017; Guo et al., 2022), whereas in cold regions, the Budyko model shows relatively low accuracy in estimating ET at the annual scale (Lute et al., 2014; Gao et al., 2010; Potter et al., 2005). This comparison further supports the reliability of our proposed method. The following sentence has been added to the revised manuscript accordingly:

"Previous studies based on the Budyko framework (ignoring TWSC) at the annual scale have shown that water balance is primarily governed by P and potential ET (Sankarasubramanian & Vogel, 2002; Zhang et al., 2008; Koster & Suarez, 1999). However, these influences vary across climatic regions. For example, in tropical and arid regions, P tends to be the dominant controlling factor (Du et al., 2024; Wu et al., 2018; Liu et al., 2017; Guo et al., 2022). In cold regions, the Budyko model exhibits relatively limited accuracy in estimating ET at the annual scale (Lute et al., 2014; Gao et al., 2010; Potter et al., 2005). These previous findings at the annual scale provide indirect support for our results derived at the monthly scale.”.

It is not immediately evident to this reviewer that negative values are a problem, especially for the soil storage term (TWSC). In fact one may expect this term to be symmetric around zero, and maybe the same holds for the errors?

Response: Thank you for your careful review. We apologize for not clearly explaining in the original manuscript the meaning of negative values for P, ET, and Q, as well as the treatment of “negative” values in TWSC, which may have led to misunderstanding.

In our manuscript, the term “negative values” refers to instances where the values of P, ET, and Q, which were originally positive, become negative after applying a BCC method. For TWSC, we refer to a change in sign—either from positive to negative or vice versa—after correction. These phenomena are caused by the application of BCC methods and intended to highlight a key limitation of existing BCC methods: when the full ΔRes is redistributed to budget components, it may introduce more uncertainty. Our modifications are as follows:

(1) To avoid any ambiguity, we have revised the manuscript to clarify the meaning of the “negative values” referred to in this study, as follows:

“A clear manifestation of this limitation is the occurrence of negative values in corrected budget component datasets when applying existing BCC methods at the monthly scale, such as negative P, ET, and Q. These unrealistic negative values arise when an excessive share of the ΔRes is redistributed to specific components. For instance, if the BCC method overestimates the error in a specific component, it may assign an excessively large portion of ΔRes to that component. When the magnitude of the correction exceeds the component's original value, the result is a negative flux, which is hydrologically incorrect. Beyond introducing negative values, such imbalanced redistribution compromises the integrity of the remaining components. Overcorrecting one variable necessarily reduces the share of ΔRes available for others, potentially degrading their accuracy. Our previous work demonstrated that enforcing water budget closure can, to some extent, reduce the accuracy of individual components and tends to introduce an ET regulation factor to mitigate accuracy loss in ET caused by existing BCC methods (Luo et al., 2023). A more hydrologically sound approach may involve partial closure, whereby only the portion of ΔRes attributable to quantified uncertainties is redistributed, while the residual linked to unmeasured processes is retained.”.

(2) Regarding TWSC, we fully agree with you that this variable can naturally take both positive and negative values. Indeed, many studies have shown that at annual scales, and with sufficiently long time series, the long-term average of TWSC can be considered negligible. We apologize for not clearly describing how our proposed method handles TWSC during the correction process. In our proposed method, we assume that TWSC observed by the GRACE satellite is reliable. Therefore, if the sign of TWSC changes after correction by the BCC method (e.g., from positive to negative or vice versa), we consider the correction to TWSC to be physically unreasonable, and we suspend further redistribution of ΔRes to TWSC in subsequent iterations. The following sentence has been

added to the revised manuscript:

"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 Equation 33. For TWSC, if a sign reversal occurs during iteration (i.e., from positive to negative or vice versa), the redistribution of the water imbalance error to TWSC is suspended in the following iteration.".

One would assume, based on considerations of the various terms of the water balance that a comparison between the negative (or positive) residuals over time will help identifying which term may be primarily responsible: the soil term can dominate in the short term (e.g. days), but will be small for annual comparisons and may become negligible at decadal scale (except for the long-term desiccation discourse). Spatial patterns are also expected, as frontal rainfall patterns are much easier to represent correctly than thunderstorms (much of tropics and arid zone rainfall) -- indeed your later results (Fig. 6) seem to match this expectation.

Response: Thank you very much for your thoughtful and insightful review. We fully agree with your perspective that analyzing the sign and magnitude of residuals over time can provide valuable clues in identifying the dominant contributors to water imbalance errors at different temporal scales. For instance, at daily scales, soil moisture variations may dominate, while at annual or longer time scales, their influence becomes negligible—except in regions experiencing long-term drying trends. Likewise, spatial patterns of residuals reflect regional climatic features, with frontal precipitation systems generally captured more accurately than convective storms in tropical and arid zones.

However, for the purpose of correcting water budget closure at the monthly scale—which is the focus of this study—this method of using sign analysis of residuals over time is less effective in accurately identifying the dominant error source and quantifying the uncertainty of each component. This is due to two key reasons: 1) TWSC cannot be neglected at the monthly scale, meaning the residuals are not solely attributable to errors in P, ET, or Q; 2) Redistribution of residual errors in BCC methods requires quantitative estimation of monthly uncertainties for each component. Unfortunately, time-series-based sign analysis is generally insufficient for this purpose, especially when precise, month-by-month uncertainty quantification is needed. Nonetheless, we acknowledge that residual pattern analysis over time remains a valuable tool for validating the plausibility of error attribution made at the monthly scale. As such, we view it as a complementary approach for guiding and validating the redistribution weights used in monthly BCC corrections.

Currently, most studies on water budget closure correction focus on the monthly scale, for the following reasons: 1) At daily scales, the uncertainty in remote sensing estimates of water cycle variables is even higher than at the monthly scale, and there is a lack of reliable daily-scale data on water storage change (e.g., GRACE data are available only at the monthly scale), which makes daily water balance studies highly uncertain; 2) At annual scales, the long-term average of TWSC is often negligible, and with relatively accurate observations of Q, the sign of the residual can often indicate the balance between P and ET, enabling identification of the primary sources of error. However, if monthly water cycle variables can be corrected to achieve consistent closure, the resulting annual water balance would also be consistent, with the added benefit of preserving the seasonal variability that is masked in annual-scale analyses. For this reason, the monthly scale remains the primary focus of BCC-related research.

Importantly, we also recognize the potential value of using annual-scale residual analysis as a constraint for monthly correction. For example, within a given basin, the dominant error sources identified at the annual scale should, in principle, align with the redistribution priorities at the monthly scale. To incorporate your suggestions, we have made the following revisions in the manuscript:

(1) Clarified the temporal focus of this study in the introduction: "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.".

(2) Expanded the discussion of how spatiotemporal residual patterns can inform future improvements in BCC methods, especially regarding the potential of using annual-scale insights to guide monthly corrections: "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 precipitation (P) and evapotranspiration (ET) uncertainties become more dominant. Spatial patterns of residuals also reflect the nature of regional precipitation regimes. In the 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.".

Maybe further reference can be made to the 'Budyko' literature that looks at an annual balance, while your current analysis takes a monthly perspective.

Response: Thank you for your constructive suggestions. We have reviewed and summarized relevant literature that applied the Budyko framework at the annual timescale. Many studies have shown that water balance at this scale is primarily influenced by P and potential ET (Sankarasubramanian & Vogel, 2002; Zhang et al., 2008; Koster & Suarez, 1999). These influences, however, differ across climatic regions. For instance, in tropical and arid zones, P tends to be the dominant factor (Du et al., 2024; Wu et al., 2018; Liu et al., 2017; Guo et al., 2022), whereas in cold regions, the Budyko model shows relatively low accuracy in estimating ET at the annual scale (Lute et al., 2014; Gao et al., 2010; Potter et al., 2005). These previous findings indirectly support our results at the monthly scale. The following sentence has been added to the revised manuscript accordingly:

"Previous studies based on the Budyko framework (ignoring TWSC) at the annual scale have shown that water balance is primarily governed by P and potential ET (Sankarasubramanian & Vogel, 2002; Zhang et al., 2008; Koster & Suarez, 1999). However, these influences vary across climatic regions. For example, in tropical and arid regions, P tends to be the dominant controlling factor (Du et al., 2024; Wu et al., 2018; Liu et al., 2017; Guo et al., 2022). In cold regions, the Budyko model exhibits relatively limited accuracy in estimating ET at the annual scale (Lute et al., 2014; Gao et al., 2010; Potter et al., 2005). These previous findings at the annual scale provide indirect support for our results derived at the monthly scale.”.

The abstract could become more attractive to readers if the time unit (monthly balance calculations) is made explicit, as results for daily or annual balance calculations will likely be different.

Response: Thank you very much for your constructive suggestion. We have revised the abstract to clarify that this study calculates water balance at the monthly scale: “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.”.

Details:

The Highlights should be understandable for a non-technical expert -- at the moment they are too full of jargon to attract readers.

Response: Thank you very much for your constructive suggestion. We have revised the Highlights by removing technical jargon and replacing them with plain language descriptions. The updated Highlights clearly convey the key findings and significance of the work in a more understandable and engaging manner. Please see our revised Highlights below.

- 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.

Line 57 Indeed a closed budget gives some confidence in the underlying estimates, but not if the closure is obtained by 'fudging' the data, without 'understanding'. So I disagree that 'closing the budget' helps with 'understanding'.

Response: Thank you very much for your careful review. We sincerely apologize for the inappropriate expression in the original sentence. Although a large number of datasets for individual budget components have been produced, discrepancies such as measurement errors, systematic biases, and unmeasured components prevent the closure of the water budget among these datasets.

In reality, the components of the water budget are interconnected; together, they regulate the exchange of water among the atmosphere, land, and oceans. Many studies have attempted to reduce the imbalance in the water budget among existing datasets by estimating and correcting errors in the individual components, thereby improving the overall consistency.

We apologize for our unclear expression in our original sentence. On the one hand, we changed the word “understanding” to “confidently applying budget components in hydrological studies”; On the other hand, we have expanded this sentence to clearly explain the concept of the terrestrial water balance, the main components it includes, and their interactions. We then emphasize the importance of improving the consistency among these datasets in order to accurately understand hydrological processes, due to the intrinsic interconnections among water budget components in real-world hydrological systems, as follows: “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). It is governed by four interconnected components—precipitation (P), evapotranspiration (ET), streamflow (Q), and terrestrial water storage change (TWSC)—that together regulate the exchange of water among the atmosphere, land, and oceans (Abolafia-Rosenzweig et al., 2021; Sahoo et al., 2011; Chen et al., 2020; Wang et al., 2015). These fluxes and storages are dynamically linked, responding to climatic variability, land surface heterogeneity, and human interventions across a range of spatial and temporal scales. Ensuring a closed water balance is essential for maintaining consistency among these budget components and for confidently applying them in hydrological studies (Li et al., 2024; Mourad et al., 2024).”.

Line 66. Before delving into the details it will be good for the reader to be reminded of the physical aspects of uncertainty in the various terms, as these are of different natures:
P precipitation input -- the typically are fairly reliable point data from rainfall gauge data, often with some need t gap fill missing data. The main uncertainty here is in the spatial distribution and representativeness of rainfall gauges, in relation to rainfall types (for frontal rains the spatial uncertainty is low, for local storms it can be high). The distribution of rainfall gauges is often determined in part by accessibility and convenience, and overall uncertainty of daily rainfall may be easily underestimated. More recent satellite based estimates of rainfall appear to perform well for frontal rains, but not in other rainfall types.
ET Evapotranspiration equations have been fairly well calibrated, but there can be uncertainty over the advection term especially in small catchments. For larger areas energy balance equations may be sufficient.
Q monitoring of outflow can have low uncertainty if ;rating curves' are frequently calibrated. However, the delineation of the watershed (and area used for the calculations) can be off where groundwater flows don't necessarily follow surface catchment delineations and can be underestimated.
TWSC can become negligible if a multi-year balance is considered (verifying the P and Q estimates) but can dominate the balance at a daily time-scale. A major challenge is the depth over which TWSC is to be assessed, as changes in the topsoil can be more easily assessed than that deeper in the soil.
Response: Thank you very much for your thorough review and constructive suggestions, which have been very helpful in improving the quality and readability of our manuscript. According to your suggestion, we have added a paragraph in the revised manuscript to describe the physical aspects of uncertainty in the four water cycle components (P, ET, Q, and TWSC) as follows:

“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 precipitation 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 precipitation inputs (Wang et al., 2017; Bai et al., 2019; Wu et al., 2018). Satellite-based precipitation 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 using empirical or physically based models (Jacobs et al., 1998; McMahon et al., 2016; Allen et al., 1998). Although these models are generally well calibrated, uncertainties persist due to the complex influence of advection and localized meteorological variability, especially in small catchments. At larger spatial scales, energy balance approaches tend to provide sufficiently accurate estimates (Hua et al., 2020; Hao et al., 2018; Ruhoff et al., 2022). Q measurements typically exhibit low uncertainty when rating curves are well established and regularly maintained (Jian et al., 2015; Krabbenhoft et al., 2022). However, uncertainty can still arise from the delineation of watershed boundaries, particularly in regions where groundwater flow does not align with surface catchment divides (Huang et al., 2023; Bouaziz et al., 2018). This mismatch can result in misrepresentation of actual hydrological contributions. TWSC generally has a negligible impact on water balance calculations over multi-year periods, but can significantly affect short-term (e.g., daily) balances (He et al., 2023; Zhang et al., 2016). A key challenge is defining 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.”.

Line 98 Negative ET is possible under 'dew formation' conditions... (be it in only part of a daily temperature cycle)

Response: Thank you for your careful review. We apologize for not clearly explaining in the original manuscript the meaning of negative values for P, ET, and Q, which may have led to misunderstanding. In our manuscript, the term “negative values” refers to instances where the values of P, ET, and Q, which were originally positive, become negative after applying a BCC method. These phenomena are caused by the application of BCC methods and intended to highlight a key limitation of existing BCC methods: when the full ΔRes is redistributed to budget components, it may introduce more uncertainty. Our modifications are as follows:

To avoid any ambiguity, we have revised the manuscript to clarify the meaning of the “negative values” referred to in this study, as follows: “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 unrealistic. Beyond introducing negative values, such imbalanced redistribution compromises the integrity of the remaining components. Overcorrecting one variable necessarily reduces the share of ΔRes available for others, potentially degrading their accuracy. Our previous work demonstrated that enforcing water budget closure can, to some extent, reduce the accuracy of individual components and tends to introduce an ET regulation factor to mitigate accuracy loss in ET caused by existing BCC methods (Luo et al., 2023). A more hydrologically sound approach may involve partial closure, whereby only the portion of ΔRes attributable to quantified uncertainties is redistributed, while the residual linked to unmeasured processes is retained.”.

Line 244 There can be 'bias' (systematic error, e.g. if groundwater flows mean that the basin is not closed and part of outflowing Q is missed; the area of the basin can also be incorrect), part 'measurement error'. As you focus on relatively large basins, the bias term may be relatively small, but for smaller watersheds the bias terms cannot be ignored. Standard techniques such as plotting cumulative Q vs cumulative P give indications, especially if nested Q data exist beyond outflow data.

Response: Thank you very much for your kind reminder. In the revised manuscript, we have added a description regarding systematic biases, as follows: "However, in practice, this balance is rarely achieved due to various sources of error. These include systematic biases (such as missed portions of outflow resulting from unclosed basin boundaries and inaccuracies in catchment area delineation, particularly in small basins), measurement uncertainties, and the omission of unmeasured components. Consequently, each budget component (P, ET, Q, and TWSC) is subject to an associated error term (denoted as ε_P, ε_{ET}, ε_Q, ε_{TWSC}, respectively), leading to a non-closure of the water budget (i.e., Equation 1 becomes Equation 4) (Aires, 2014; Wong et al., 2021).".

Line 702-714, Figure 7 - would it make sense to compensate S Hemisphere data for a 6 month shift in seasons? Or even more flexibly to use a hydrological year concept with a standardized month for maximum P.

Response: Thank you for the insightful suggestion. We agree that seasonal misalignment between the Northern and Southern Hemispheres may obscure underlying hydrological patterns. We have redrawn Figure 7 accordingly. Specifically, based on your recommendation, we adjusted the Southern Hemisphere data by applying a 6-month shift to align its seasonal phases with those of the Northern Hemisphere. This adjustment has been noted in the caption of the revised Figure 7. Finally, we reanalyzed the figure based on the updated version.

"Fig. 7 presents the seasonal cycle of negative values across different climate zones, examining whether these values exhibit significant seasonal patterns. Negative P values predominantly occur in winter and spring, with a higher proportion from January to March in tropical climates compared to arid regions. ET tends to show negative values more frequently in winter and spring, with a lower likelihood in 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.".

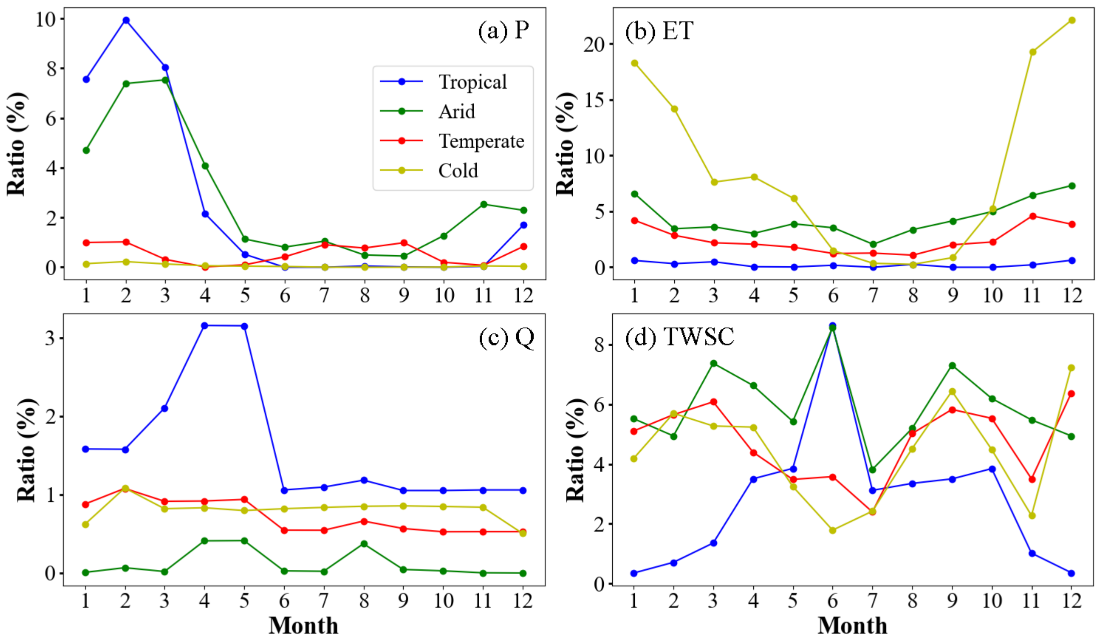

Fig. 7 Seasonal cycle of the proportion of negative errors for budget components. Different colors representing various climate types. The Southern Hemisphere data by applying a 6-month shift to align its seasonal phases with those of the Northern Hemisphere.

Reference:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Huang, P., Wang, G., Guo, L., Mello, C. R., Li, K., Ma, J., & Sun, S. (2023). Most global gauging stations present biased estimations of total catchment discharge. Geophysical Research Letters, 50(15), e2023GL104253. https://doi.org/10.1029/2023GL104253

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

Jian, J., Ryu, D., Costelloe, J. F., & Su, C. H. (2015). 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.

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

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

Lehmann, F., Vishwakarma, B. D., & Bamber, J. (2022). How well are we able to close the water budget at the global scale?. Hydrology and Earth System Sciences, 26(1), 35-54. https://doi.org/10.5194/hess-26-35-2022

Li, L., Dai, Y., Wei, Z., Wei, S., Zhang, Y., Wei, N., & Li, Q. (2024). Enforcing Water Balance in Multitask Deep Learning Models for Hydrological Forecasting. Journal of Hydrometeorology,

25(1), 89-103. https://doi.org/10.1175/JHM-D-23-0073.1
Liu, J., Zhang, Q., Singh, V. P., & Shi, P. (2017). Contribution of multiple climatic variables and human activities to streamflow changes across China. Journal of Hydrology, 545, 145-162.
Lockhoff, M., Zolina, O., Simmer, C., & Schulz, J. (2014). Evaluation of satellite-retrieved extreme precipitation over Europe using gauge observations. Journal of climate, 27(2), 607-623. https://doi.org/10.1175/JCLI-D-13-00194.1
Luo, Z., Gao, Z., Wang, L., Wang, S., & Wang, L. (2023). A method for balancing the terrestrial water budget and improving the estimation of individual budget components. Agricultural and Forest Meteorology, 341, 109667. https://doi.org/10.1016/j.agrformet.2023.109667
Luo, Z., Li, H., Zhang, S., Wang, L., Wang, S., & Wang, L. (2023). A novel two-step method for enforcing water budget closure and an intercomparison of budget closure correction methods based on satellite hydrological products. Water Resources Research, 59(3), e2022WR032176. https://doi.org/10.1029/2022WR032176
Lute, A. C., & Abatzoglou, J. T. (2014). Role of extreme snowfall events in interannual variability of snowfall accumulation in the western United States. Water Resources Research, 50(4), 2874-2888. https://doi.org/10.1002/2013WR014465
Masunaga, H., Schröder, M., Furuzawa, F. A., Kummerow, C., Rustemeier, E., & Schneider, U. (2019). Inter-product biases in global precipitation extremes. Environmental Research Letters, 14(12), 125016. https://doi.org/ 10.1088/1748-9326/ab5da9
Mourad, R., Schoups, G., Bastiaanssen, W., & Kumar, D.N. (2024). Expert-based prior uncertainty analysis of gridded water balance components: Application to the irrigated Hindon River Basin, India. Journal of Hydrology: Regional Studies, 55, 2214-5818. https://doi.org/10.1016/j.ejrh.2024.101935.
McMahon, T. A., Finlayson, B. L., & Peel, M. C. (2016). Historical developments of models for estimating evaporation using standard meteorological data. Wiley Interdisciplinary Reviews: Water, 3(6), 788-818. https://doi.org/10.1002/wat2.1172
Munier, S., & Aires, F. (2018). A new global method of satellite dataset merging and quality characterization constrained by the terrestrial water budget. Remote Sensing of Environment, 205, 119-130. https://doi.org/10.1016/j.rse.2017.11.008
Nassaj, B. N., Zohrabi, N., Shahbazi, A. N., & Fathian, H. (2022). Evaluating the performance of eight global gridded precipitation datasets across Iran. Dynamics of Atmospheres and Oceans, 98, 101297. https://doi.org/10.1016/j.dynatmoce.2022.101297
Palharini, R. S. A., Vila, D. A., Rodrigues, D. T., Quispe, D. P., Palharini, R. C., de Siqueira, R. A., & de Sousa Afonso, J. M. (2020). Assessment of the extreme precipitation by satellite estimates over South America. Remote Sensing, 12(13), 2085. https://doi.org/10.3390/rs12132085
Petković, V., & Kummerow, C. D. (2017). Understanding the sources of satellite passive microwave rainfall retrieval systematic errors over land. Journal of Applied Meteorology and Climatology, 56(3), 597-614. https://doi.org/10.1175/JAMC-D-16-0174.1
Pan, M., Sahoo, A. K., Troy, T. J., Vinukollu, R. K., Sheffield, J., & Wood, E. F. (2012). Multisource estimation of long-term terrestrial water budget for major global river basins. Journal of Climate, 25(9), 3191-3206. https://doi.org/10.1175/JCLI-D-11-00300.1
Potter, N. J., Zhang, L., Milly, P. C. D., McMahon, T. A., & Jakeman, A. J. (2005). Effects of rainfall seasonality and soil moisture capacity on mean annual water balance for Australian catchments. Water Resources Research, 41(6). https://doi.org/10.1029/2004WR003697

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

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

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

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

Wang, S., Huang, J., Yang, D., Pavlic, G., & Li, J. (2015). Long-term water budget imbalances and error sources for cold region drainage basins. Hydrological Processes, 29(9), 2125-2136. https://doi.org/10.1002/hyp.10343

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

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

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

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

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

Zhang, L., Potter, N., Hickel, K., Zhang, Y., & Shao, Q. (2008). Water balance modeling over variable time scales based on the Budyko framework–Model development and testing. Journal of Hydrology, 360(1-4), 117-131. https://doi.org/10.1016/j.jhydrol.2008.07.021

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

Zhou, L., Cao, Y., Shi, C., Liang, H., & Fan, L. (2024). Quantifying the Atmospheric Water Balance Closure over Mainland China Using Ground-Based, Satellite, and Reanalysis Datasets. Atmosphere, 15(4), 497. https://doi.org/10.3390/atmos15040497