

## **Response to Reviewers**

We sincerely thank your constructive and insightful comments, which have significantly helped improve the clarity, scientific rigor, and presentation of our manuscript. We carefully considered all comments and have revised the manuscript accordingly. The primary revisions made in the manuscript are summarized as follows:

(1) The basin subdivision strategy has been clarified and improved, including clearer documentation of data sources and revised ordering of basins in Table S1 to enhance transparency.

(2) Additional basin-scale comparisons with GLDAS CLSM have been included to demonstrate the added value of the proposed reconstruction method. The results have been incorporated into Section 4.2.2 of the revised manuscript, and the section has been extensively revised accordingly.

(3) New validations of daily-scale reconstruction results have been added through comparisons with independent datasets, now included as a new section in the manuscript. The corresponding figure has been included in the revised manuscript.

(4) Following the reviewer's suggestion, we conducted an independent validation experiment by separating training and validation periods. The results have been added as a new section in the manuscript, with supporting figures included in the Supplementary Materials as Figs. S8 and S9.

(5) Following the reviewer's suggestion regarding the physical interpretability of model parameters, we have substantially revised and expanded the derivation and

interpretation of model parameters in the manuscript. The physical meanings of parameters  $a, b, c,$  and  $d$  are now more clearly clarified. The full mathematical derivation and supporting results have been moved to the revised manuscript, with supporting figures included in the Supplementary Materials as Figs. S10–S12.

Please find point-by-point responses on the following pages, where reviewers' comments are shown in black text and our responses are shown in blue text.

Sincerely,

*Pu Xie and Shuang Yi*

February 2026

## **Responses to Major Comments**

### **Major concerns (i): the lack of daily-scale validation**

**Response:** We thank the reviewer for pointing out the importance of validating the daily-scale reconstruction. We conducted basin-scale validations of the reconstructed daily TWSA against two different reference datasets: ITSG-GRACE2018 and GLDAS-2.2 data products (see Fig. R1(a) and Fig. R1(b)). The model performance was quantitatively assessed using the NSE value.

Furthermore, we computed basin-averaged daily TWSA over large regions such as the Mississippi River basin and examined the consistency of sub-monthly variability across the three datasets. As shown in Fig. R1(c), GRACE-REC, ITSG-

GRACE2018, and GLDAS-2.2 exhibit strong agreement in daily variability, indicating the physical reliability of the proposed reconstruction model at sub-monthly timescales.

It is worth noting that in the Indus River basin, as illustrated in Fig. R1(b), the agreement between our reconstruction and GLDAS-2.2 is relatively poor. However, time series analysis in Fig. R2(a) reveals that our reconstructed TWSA peaked in 2015 and declined steadily from 2016 to 2018, which aligns well with the interannual trend of precipitation during the same period. This indicates that the reconstructed TWSA reasonably responds to the temporal dynamics of the meteorological forcing. In contrast, GLDAS-2.2 shows a relatively stable TWSA trend over the same period, failing to capture the observed decrease in precipitation.

In addition, as shown in Fig. R2b, our reconstructed daily TWSA matches well with the GRACE-FO observation in June 2018—the first month of GRACE-FO’s resumed measurements—whereas GLDAS-2.2 clearly overestimates TWSA during this month. This further supports the reliability and physical consistency of our daily reconstruction results.

In the revised manuscript, we have incorporated the comparison between the reconstructed daily TWSA and other datasets into the main text. Specifically, a new Section 4.3 entitled “*Comparison with de-seasonalized and de-trended daily products*” has been added. Figure R1 presented in this response has been included in the revised manuscript as the new Fig. 7, and the numbering of all subsequent figures has been adjusted accordingly. The following text has been added after Line 456 in the original

manuscript:

*To evaluate the reliability of the daily reconstruction, we compared the reconstructed daily TWSA with two different reference datasets, namely the daily ITSG-Grace2018 solution derived using a Kalman smoothing approach and the GLDAS-2.2 Daily product.*

*We first assessed reconstruction performance at the basin scale by computing basin-averaged daily TWSA over major river basins worldwide and comparing the results with both reference datasets. Model performance was quantified using the NSE. As shown in Fig. 7(a–b), NSE values are positive across most large basins. To further investigate sub-monthly variability, we examined basin-averaged daily TWSA time series using the Mississippi River basin as a representative example (Fig. 7c). The results show good agreement among the reconstructed GRACE-REC, ITSG-Grace2018, and GLDAS-2.2 time series.*

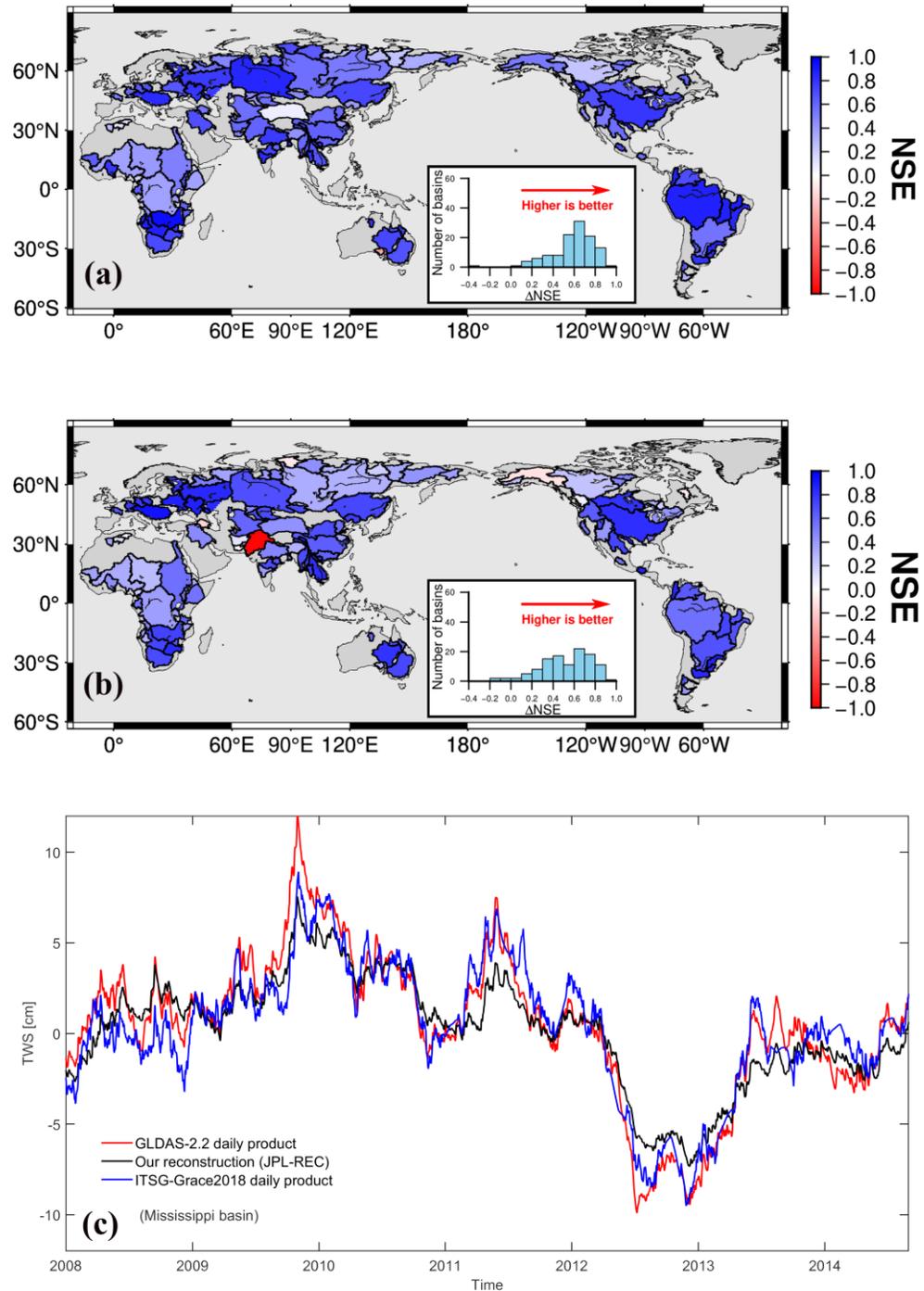


Figure R1: Spatial distribution of the Nash–Sutcliffe efficiency (NSE) of de-seasonalized and de-trended terrestrial water storage anomalies between the reconstructed daily TWSA and (a) ITSG-Grace2018 for the period 2003 to August 2016 and (b) GLDAS-2.2 daily product for the period 2003 to December 2021 across 116 major river basins worldwide. (c) Time-series comparison of the three daily datasets for the Mississippi River basin, focusing on the period 2008–2014 to improve readability of high-frequency fluctuations.

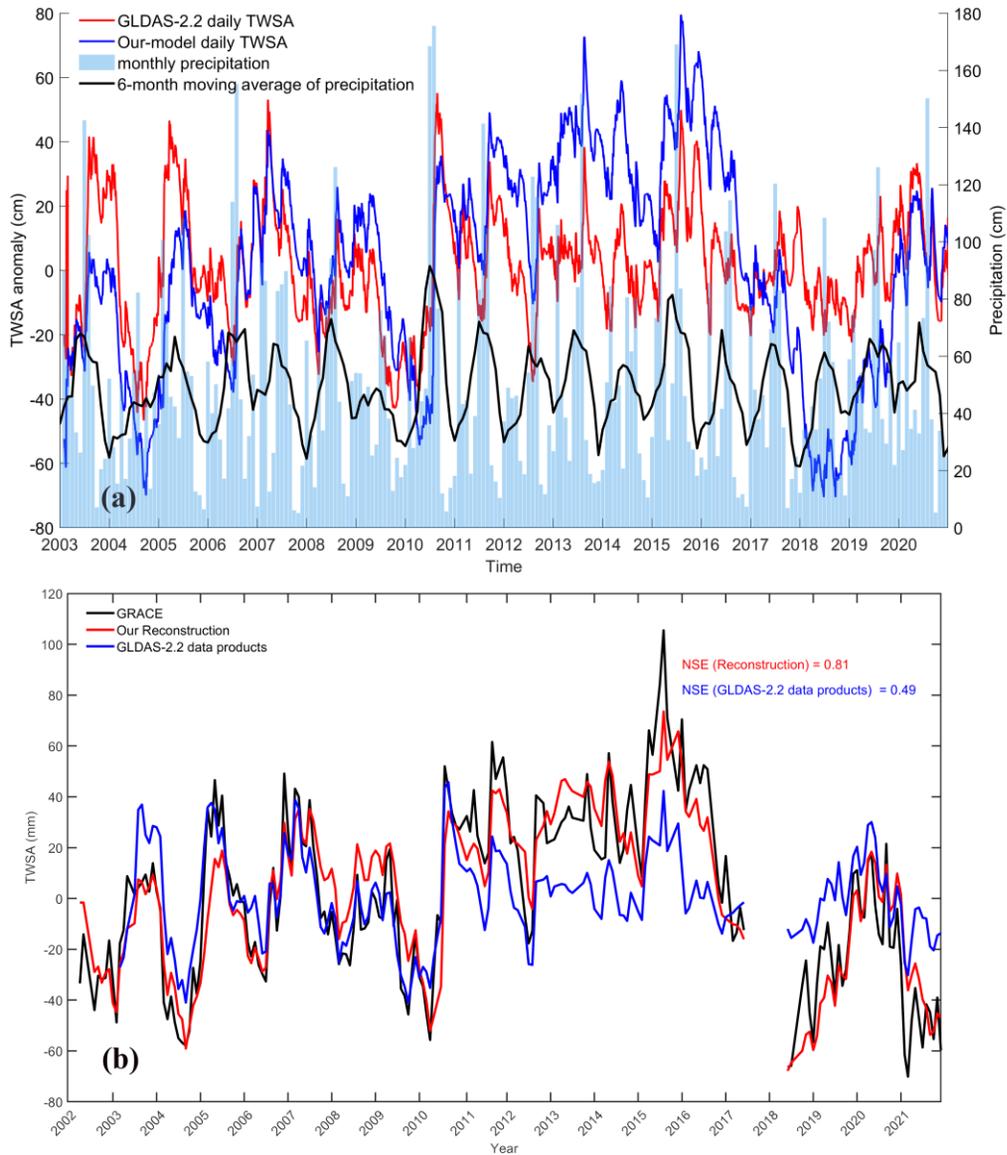


Figure R2: Comparison of daily and monthly TWSA reconstruction in the Indus River basin. (a) Detrended and deseasonalized daily TWSA from GLDAS-2.2 (red line), our model-based daily TWSA reconstruction (blue line), monthly precipitation (light blue bars), and 6-month moving average precipitation (black line); (b) Monthly GRACE TWSA, model-based monthly TWSA reconstruction, and monthly TWSA aggregated from daily GLDAS-2.2 data.

**Major concerns (ii): The arbitrary 116-basin subdivision and global mapping strategy**

**Response:** Thank you for your valuable comments on the basin subdivision strategy adopted in our study. We highly appreciate your concern regarding the “arbitrary 116-basin subdivision and its global mapping strategy” and would like to provide the following clarification:

First, the global basin boundaries used in this study were not arbitrarily defined, but are derived from the publicly available dataset “Major River Basins of the World” provided by the Global Runoff Data Centre (GRDC). This dataset is widely recognized and has been extensively used in global-scale hydrological studies, e.g., Burek and Smilovic (2023), Zhong et al. (2025), Gao et al. (2026). It can be accessed from the official website:

[https://www.bafg.de/SharedDocs/ExterneLinks/GRDC/mrb\\_shp\\_zip.html](https://www.bafg.de/SharedDocs/ExterneLinks/GRDC/mrb_shp_zip.html)

Second, to ensure sufficient spatial representativeness, we followed Zhong et al. (2025), ranked all basins by area, and selected the 116 largest basins with drainage areas greater than  $1 \times 10^5 \text{ km}^2$  for analysis.

Therefore, we believe that the adopted 116-basin subdivision scheme is based on clearly documented data sources and scientifically justified selection criteria, rather than being arbitrarily defined. In addition, we have revised Table S1 in the Supplementary Materials by reordering the listed basins according to drainage area instead of alphabetical order, so that the selection criteria are more transparent and easier to interpret.

**Concern (iii): the insufficient demonstration of added value against physically based assimilation products such as GLDAS CLSM.**

**Response:** We thank the reviewer for highlighting the importance of demonstrating the added value of the proposed reconstruction relative to physically based assimilation products such as GLDAS CLSM. We fully agree that GLDAS CLSM represents a widely used, physically consistent land surface model, and its TWS estimates are of high value for hydrological studies.

We evaluate the basin-averaged monthly TWSA from the proposed reconstruction and GLDAS CLSM against the JPL mascon (JPLM) solution, which serves as a consistent observational reference. Model performance is assessed using the NSE across 116 global river basins (Fig. R3(a), Fig. R3(b)). In addition, we present the spatial distribution of the *NSE* difference ( $\Delta NSE$ ) between the two products to explicitly quantify the added value of the proposed method (Fig. R3(c)).

The results show that the reconstructed TWSA achieves higher NSE values than GLDAS CLSM in most river basins. This is because the TWSA from GLDAS is obtained by explicitly simulating and integrating multiple water storage components (e.g., soil moisture, groundwater, snow, etc.), and its results are inevitably affected by the completeness and precision of the physical processes considered.

In contrast, the reconstruction method proposed in this study does not attempt to simulate each component separately. Instead, it statistically characterizes the total TWS response to precipitation and temperature, using GRACE TWSA as a direct constraint. This approach implicitly integrates all relevant water storage processes,

including those that are difficult to model explicitly in land surface models (e.g., surface water, deep groundwater). Therefore, compared to GLDAS CLSM, our method better captures the temporal dynamics of TWSA in multiple river basins and yields more accurate reconstructions.

These analyses have now been incorporated into the revised manuscript. Specifically, Fig. R3 included in this response has been added to the Supplementary Materials as Fig. S7. In addition, Section 4.2.2 has been revised, with the section title changed from “Comparison with the Zhong Model” to “Basin-scale comparison.” In the original manuscript we inserted a new explanatory paragraph at line 376, and we added a second explanatory paragraph immediately after line 399. The revised section now reads as follows:

*To systematically evaluate model performance at the basin scale, we compare our reconstruction (JPL-REC) with two representative types of products: (i) a previous GRACE-based reconstruction product (Zhong-REC from (Zhong et al., 2025)), and (ii) physically based land surface model simulations forced by meteorological data (GLDAS CLSM).*

*First, to evaluate the relative performance of different statistical reconstruction methods, we adopted the JPLM as a common reference. We compared basin-averaged monthly TWSA from our JPL-REC and Zhong-REC against the JPLM and computed the NSE for each (Fig. R3). By comparing the NSE values for the two approaches, we assessed their relative reconstruction accuracy.*

*Second, to demonstrate the added value of our reconstruction relative to GLDAS*

*CLSM, we compared basin-averaged monthly TWSA from JPL-REC and GLDAS CLSM against the JPLM and computed the NSE for each (Fig. R4). Across 116 major river basins worldwide, NSE values for both products are computed against the JPL mascon solution, and the spatial distribution of the resulting  $\Delta NSE$  is presented (Fig. S7c). The results show that NSE values from our reconstruction are higher in most basins, with  $\Delta NSE$  predominantly positive, indicating that the proposed method more effectively captures temporal TWSA variations at the basin scale.*

*It should be noted that TWSA estimates from GLDAS CLSM are derived from explicit simulations of individual storage components and are therefore influenced by the completeness of the represented physical processes. In contrast, our approach statistically characterizes climate-driven storage responses under direct constraints from GRACE observations, implicitly integrating multiple storage processes. This leads to improved representation of total water storage variations and better agreement across many river basins.*

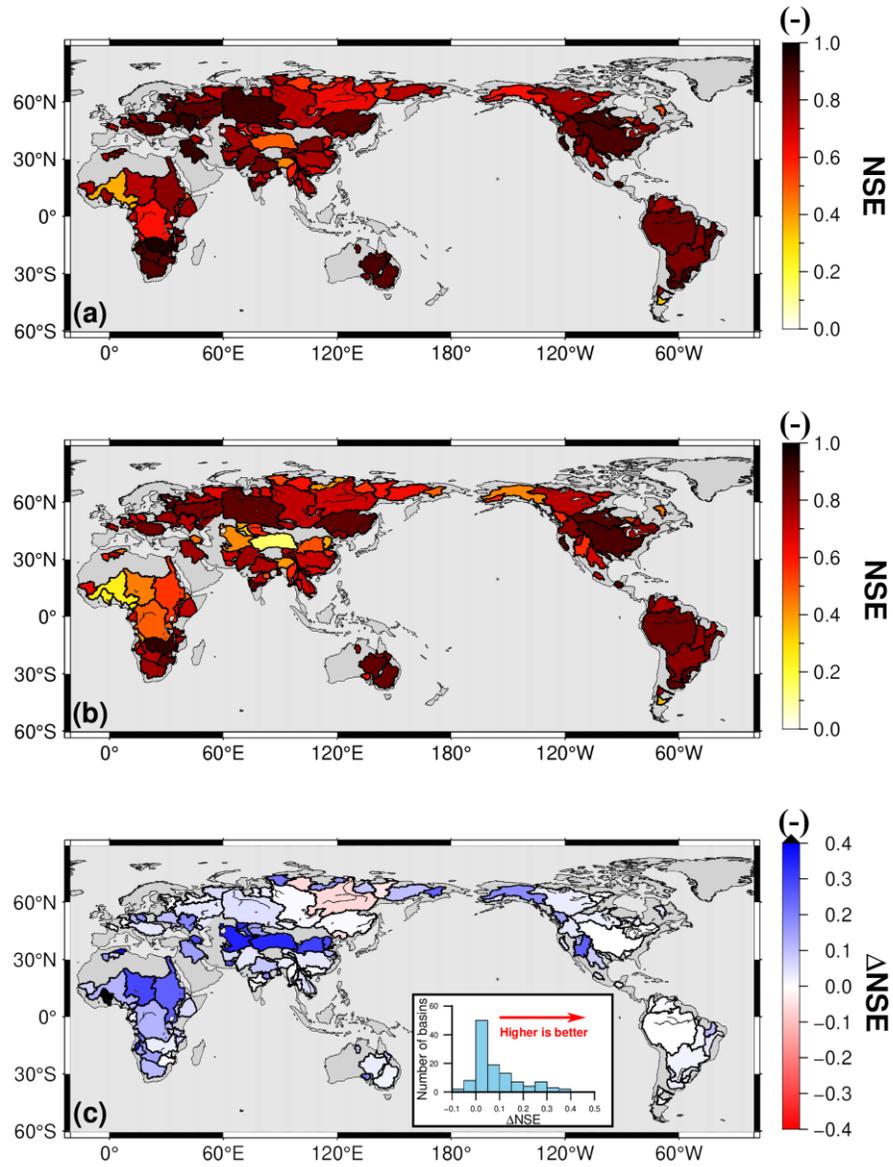


Figure R3: Spatial distribution of NSE (de-seasonalized, de-trended anomalies) between JPLM and two reconstruction models across 116 global river basins for the period 2002 to 2021. The NSE between JPLM and JPL-REC (a); the NSE between JPLM and the reconstruction by Zhong et al. (2025) (b); and their difference ( $\Delta NSE = \text{JPL-REC} - \text{Zhong-REC}$ ) (c). The inset histogram in (c) illustrates the distribution of  $\Delta NSE$  across all basins, where positive values indicate better agreement with GRACE observations by our reconstruction model.

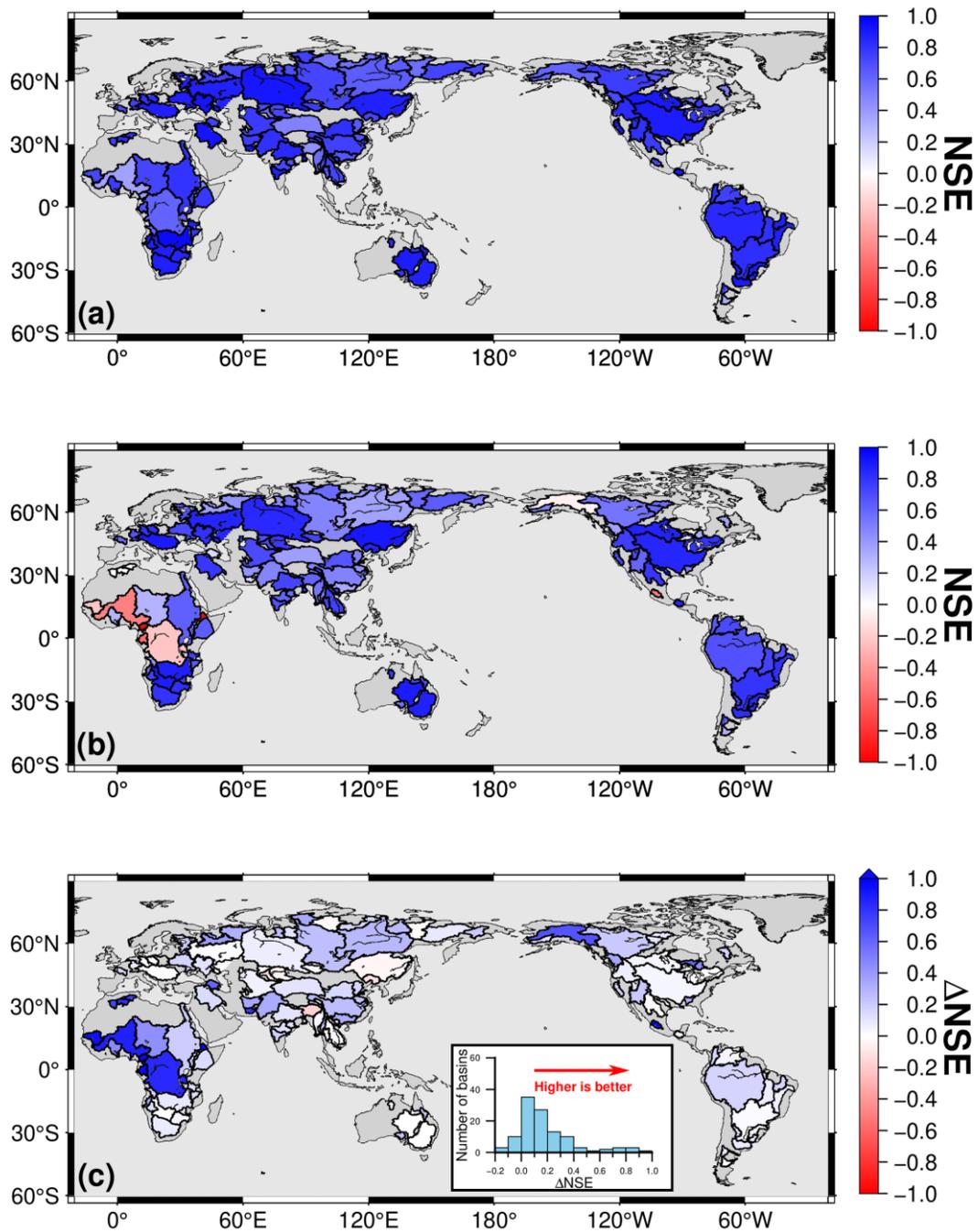


Figure R4: Spatial distribution of the NSE of de-seasonalized, de-trended TWSA between JPLM and (a) our reconstruction model and (b) GLDAS CLSM across 116 global river basins for 2003–2021; (c) their difference ( $\Delta NSE = (a) - (b)$ ). The inset histogram in (c) illustrates the distribution of  $\Delta NSE$  across all basins, where positive values indicate better agreement with GRACE observations by our reconstruction model.

**Comment 1: The model is trained on de-seasonalised and de-trended monthly TWSA from GRACE/GRACE-FO, and daily fields are produced but only monthly datasets are assessed. Although GRACE provides only monthly observations, the daily reconstruction should also be evaluated by comparison with other independent datasets such as model simulations and aridity indexes.**

**Response:** Thank you for highlighting the need for daily-scale validation. This issue has been addressed in full in our response to **Concern (i)** above. In the revised manuscript we have added new daily-scale validation analyses and associated figures; please refer to our Response to **Concern (i)** for details.

**Comment 2: In the work, TWS is reconstructed on land grids, but why do you evaluate  $\beta$  at the basin scale? In my view, you should not use this method to reconstruct grid-scale TWS, because the increase in TWS may be greater than precipitation due to runoff processes, which can lead to  $\beta > 1$ . Actually, Zhong et al. (2025) also apply a similar method to reconstruct basin-scale TWS, which can ignore the impact of runoff.**

**Response:** We thank the reviewer for raising this insightful comment regarding the scale consistency of parameter estimation.

First, concerning the reviewer's question as to why basin-scale calibrated parameter  $b$  values are used in grid-scale reconstruction, we would like to clarify that, in our model, parameter  $b$  is not derived at the basin scale and then applied to

individual grid cells. Instead, parameter estimation is performed independently for each land grid cell during grid-scale reconstruction. Therefore, both the reconstruction results and parameter estimation are spatially consistent at the grid scale. To further illustrate this, we provide a global map of the estimated parameter  $b$  values in this response document (Fig. R5).

Second, we fully agree with the reviewer that parameter  $b$  values may exceed 1 at the grid scale. As correctly pointed out, due to lateral water transfers such as surface runoff convergence and subsurface flow, local increases in terrestrial water storage may exceed local precipitation inputs, resulting in parameter  $b > 1$ . This behavior is physically plausible and does not violate mass conservation. We fully acknowledge this point.

To quantitatively evaluate the impact of the upper bound of parameter  $b$  on reconstruction performance, we conducted an additional sensitivity experiment comparing grid-scale reconstruction results under two parameter constraints: (1)  $0 < b < 1$  and (2)  $0 < b < 2$ . For both scenarios, we calculated the spatial distribution of NSE between the reconstructed results and the JPLM and analyzed the differences (Fig. R6). The results show that, for approximately 95% of global land grid cells, the NSE differences between the two scenarios fall within  $\pm 0.1$ . This indicates that allowing parameter  $b > 1$  results in only marginal improvements in reconstruction performance over most regions. Therefore, we consider that constraining parameter  $b$  to be smaller than 1 remains reasonable for maintaining model stability and physical consistency, while having negligible impact on

reconstruction accuracy.

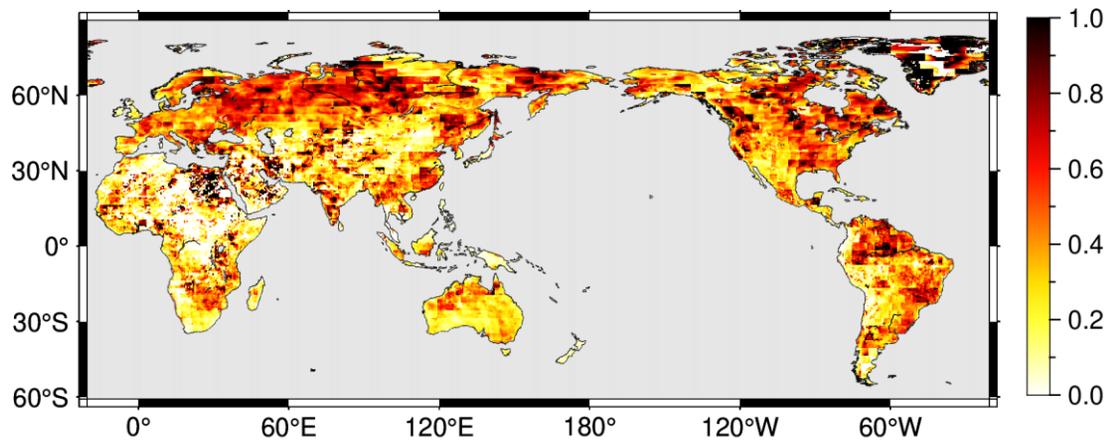


Figure R5. Spatial distribution of the calibrated parameter  $b$  across global land grid points. Parameters calibrated by monthly TWSA from JPLM based on ERA5-Land precipitation and temperature from 2002 to 2021.

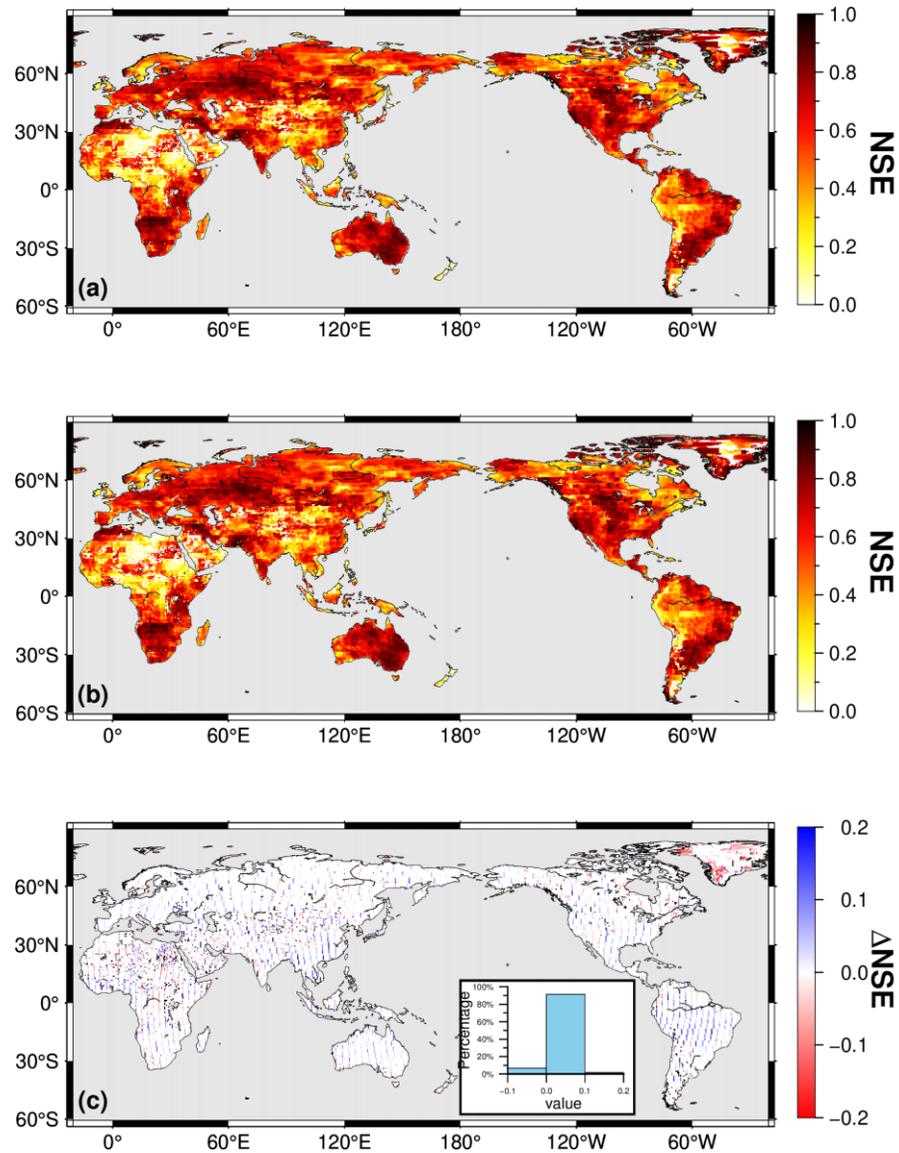


Figure R6. Sensitivity of grid-based reconstruction (JPL-REC) performance to parameter  $b$  constraints. (a) NSE between JPL-REC (with  $0 < b < 1$ ) and JPLM; (b) NSE between JPL-REC (with  $0 < b < 2$ ) and JPLM; (c) Difference in NSE between (a) and (b). Histograms show the distribution of NSE values across all global grid points.

**Comment 3: I think you should divide the datasets into training periods and validation periods, as in machine learning, to avoid the over-fitting risk.**

**Response:** We sincerely thank the reviewer for pointing out the potential risk of

model overfitting and for the valuable suggestion to divide the data into training and validation periods. Following your suggestion, we conducted additional experiments. We used the period from April 2002 to December 2012 as the training phase for parameter estimation, and the period from January 2013 to December 2023 as the independent validation phase. Model parameters were calibrated entirely based on the training data, and data from the validation period were not involved in the parameter estimation process.

Subsequently, we compared the reconstruction with the JPLM in both time periods to evaluate model performance. As shown in Fig. R7(b), most basins maintained relatively high NSE values during the validation period, with 70% of basins achieving  $NSE > 0.5$ . This indicates that the model is not restricted to the training period and has good temporal extrapolation capability. In addition, we further selected four representative basins—the Amazon, Yangtze, Murray, and Mississippi Rivers (Fig. R8)—to validate model performance through time series comparison. The results show that the reconstructed TWSA aligns well with GRACE observations in both the training and validation periods, indicating that the model is able to reproduce water storage variations stably in independent periods, rather than merely fitting the training data.

In the revised manuscript, we have added this analysis as Section 4.2.5: *Independent validation of reconstruction performance*. Figures R7 and 8 in this response document have been added to the Supplementary Materials as Fig. S8 and Fig. S9. Meanwhile, we have inserted the following paragraph into the newly added

Section 4.2.5 (after Line 456 in the original manuscript):

*To further evaluate the robustness of the proposed method and to address the potential issue of overfitting, we conducted an independent validation experiment by dividing the GRACE dataset into separate training and validation periods. Specifically, the period from April 2002 to December 2012 was designated as the training period, while the period from January 2013 to December 2023 was treated as an independent validation period. The model parameters were estimated solely using the training data, and the validation data were not used during the calibration process.*

*We then evaluated the model performance using the NSE metric by comparing the reconstructed TWSA with the JPL Mascon data in both the training and validation periods (Fig. S8). The validation results show that the majority of basins maintained relatively high NSE values during the validation phase, with 60% of basins achieving  $NSE > 0.5$ . This indicates that the model retains stable predictive capability beyond the calibration window and does not exhibit significant overfitting behavior.*

*To provide a more intuitive assessment, we further selected four representative basins with different hydro-climatic characteristics—the Amazon, Yangtze, Murray, and Mississippi River basins (Fig. S9). The time series comparison shows that the reconstructed TWSA during the validation period aligns well with GRACE TWSA, and the amplitude and phase are comparable to those in the training period. These results further confirm the robustness and generalizability of the proposed climate-driven reconstruction model.*

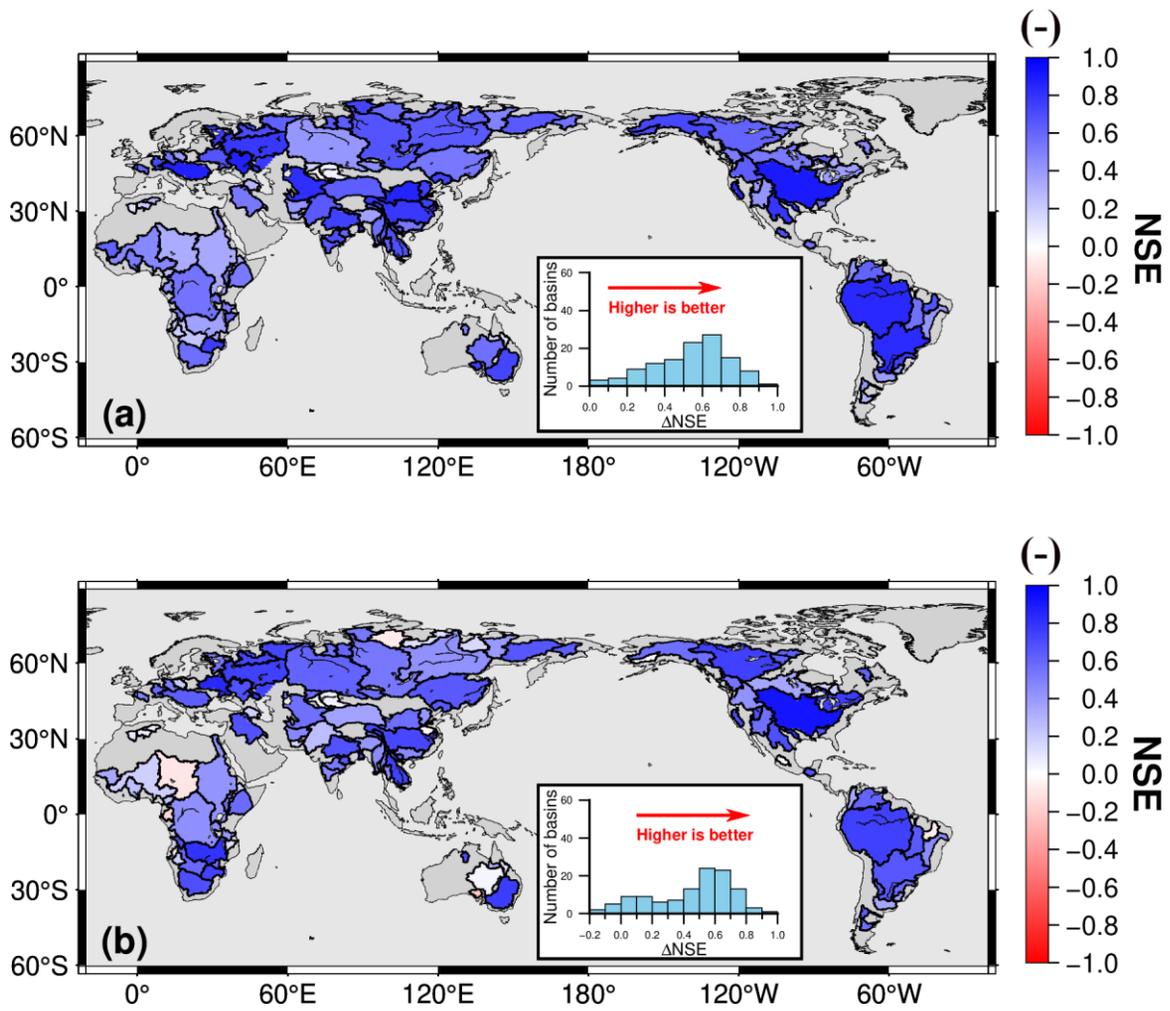


Figure R7. Comparison of reconstruction performance against JPLM data during the training and validation periods. (a) Spatial distribution of the NSE between reconstructed TWSA and JPLM data during the training period (April 2002–December 2012); (b) Spatial distribution of NSE during the validation period (January 2013–December 2023)

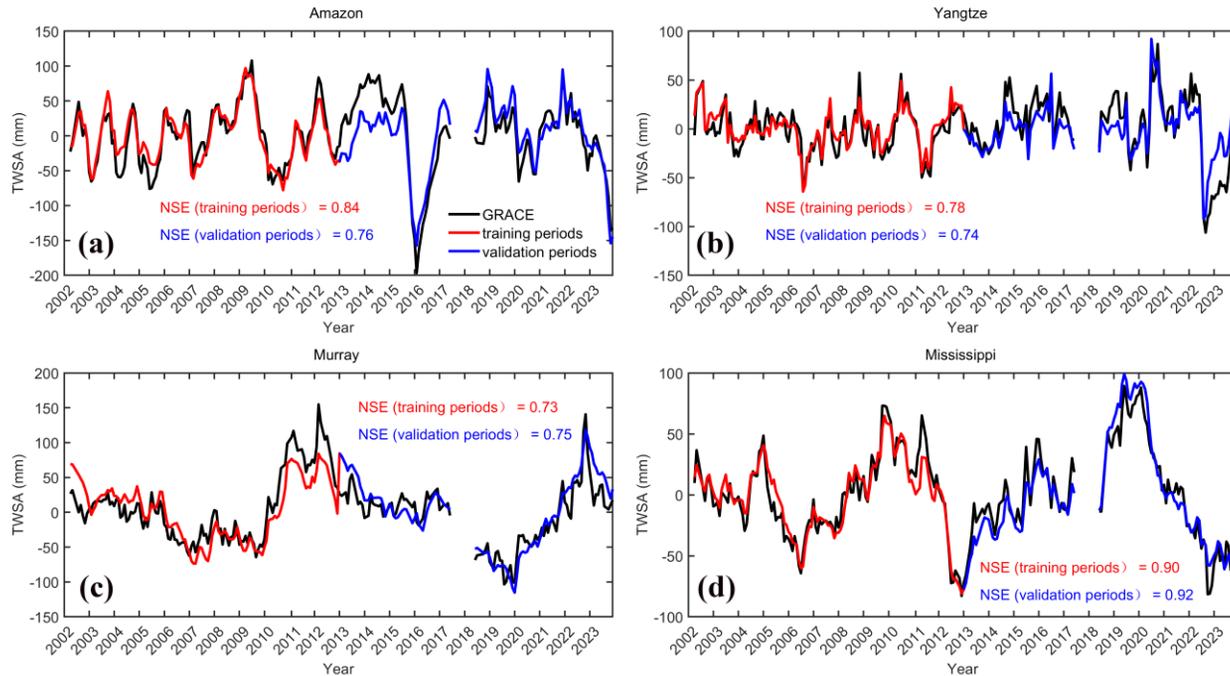


Figure R8. Time series comparison of reconstruction and GRACE/GRACE-FO TWSA in representative river basins during training and validation periods. Amazon, Yangtze, Murray, and Mississippi river basins (a–d). Black lines represent GRACE TWSA, red lines indicate model reconstructions during the training period, and blue lines during the validation period.

**Comment 4: Parameters a, b, c, d are claimed to be "physically meaningful" but lack independent corroboration such as ET/P partitioning or field infiltration measurements. Therefore, you should add scatter plots of a vs. observed ET/P and d vs. estimated groundwater turnover time, and discuss sign mismatches.**

**Response:** We thank the reviewer for the insightful comments regarding the physical interpretability of the model parameters. We fully agree that further assessment of the physical interpretability of the model parameters is important for enhancing the credibility of the proposed reconstruction model.

It should be clarified that the model proposed in this study is essentially a simplified statistical reconstruction model rather than a physically complete

hydrological data assimilation system. Consequently, it is difficult to directly validate individual parameters using existing observational data, as suggested by the reviewer. Nevertheless, to avoid the parameters being interpreted as purely empirical fitting coefficients, we revisited the model formulation starting from the basin-scale water balance equation and provided a systematic clarification of both the derivation and the physical interpretation of the parameters. In addition, independent statistical diagnostics were introduced to support the physical consistency of the parameterization.

We start from the water balance equation:

$$TWS(t) - TWS(t - 1) = P(t) - ET(t) - R(t) = P(t) - ETR(t), \quad (1)$$

where we combine the total loss term as  $ETR$ . First, we assume  $ETR$  is jointly influenced by current precipitation input  $P(t)$  and antecedent water storage  $TWS(t - 1)$ , which can be expressed as

$$ETR_1(t) = x \cdot P(t), \quad (2)$$

$$ETR_2(t) = y \cdot TWS(t - 1), \quad (3)$$

$$ETR(t) = ETR_1(t) + ETR_2(t) = x \cdot P(t) + y \cdot TWS(t - 1). \quad (4)$$

The assumption is validated in Fig. R9, where the  $ETR$  is plotted against precipitation and TWS respectively. Both plots show significant correlations, with  $ETR$  apparently exhibiting a stronger correlation with the current precipitation than the antecedent TWS.

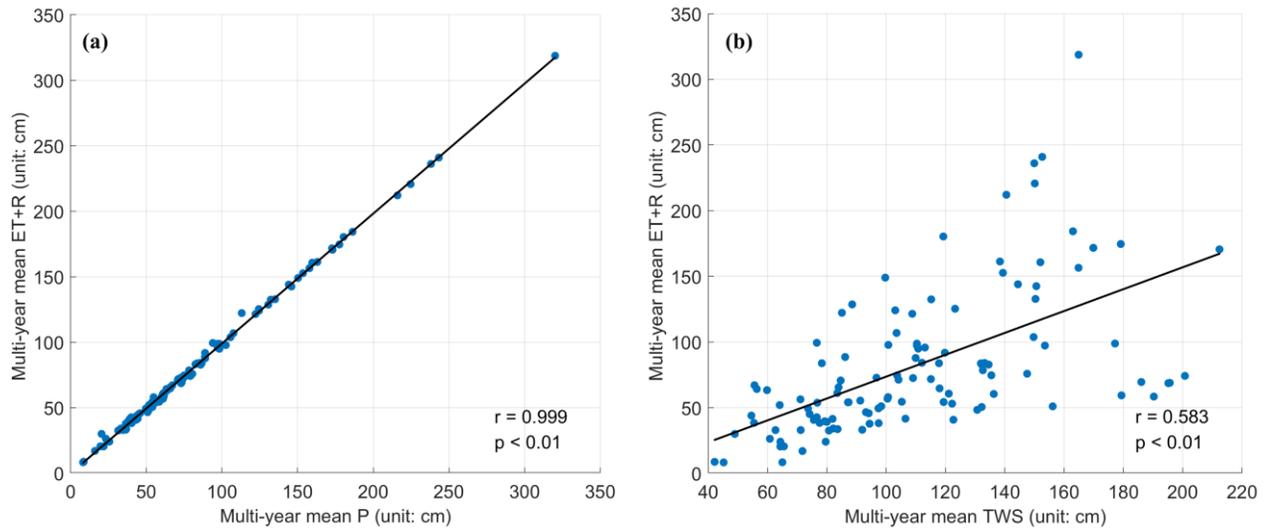


Figure R9. Scatter plot of multi-year mean precipitation versus multi-year mean evapotranspiration plus runoff ( $ET + R$ ) for each basin (a). Scatter plot of multi-year mean TWS versus multi-year mean  $ET + R$  for each basin (b). The black dashed line denotes the linear regression fit. Pearson correlation coefficient  $r$  and the  $p$ -value for the regression slope are annotated on each panel.

We then performed a multiple linear regression analysis to further test this assumption. In this step, temperature modulation was temporarily ignored and the coefficients  $x$  and  $y$  were treated as constants. The results show that all basins pass the F-test at the significance level of  $p < 0.01$ . Moreover, 74 basins (approximately 64% of all basins) exhibit determination coefficients  $R^2 > 0.6$  (Fig. R10(a)), indicating that the loss term  $ETR$  can be statistically explained by the linear combination of precipitation  $P(t)$  and antecedent storage  $TWS(t - 1)$ .

To further examine whether antecedent storage provides additional explanatory power beyond precipitation alone, we compared the full multiple regression model with a reduced model that includes only precipitation  $P(t)$  as the predictor. Figure R10(b) presents the spatial distribution of the  $R^2$  obtained from the precipitation-only regression, while Fig. R10(c) shows the spatial difference in  $R^2$  between the full

model and the reduced model. Although precipitation alone explains a substantial portion of the variability in basin-scale losses, the inclusion of antecedent storage  $TWS(t - 1)$  improves model performance. Across all basins, the incorporation of  $TWS(t - 1)$  increases  $R^2$  by approximately 0.1. Therefore, although  $ETR$  is strongly correlated with  $P(t)$ , the results demonstrate that accounting for  $TWS(t - 1)$  is both statistically meaningful and physically necessary.

On this basis, we consider the modulation effect of temperature on the hydrological cycle by assuming that the proportionality coefficients  $x$  and  $y$  vary with temperature and apply a Taylor expansion yields

$$x = \epsilon \cdot f(T_z) = \epsilon \cdot (1 + \alpha T_z + o(T_z^2)). \quad (5)$$

Retaining only the first-order approximation gives

$$x = a' \cdot T_z + b', \quad (6)$$

where  $T_z$  denotes standardized temperature.

Similarly, we obtain

$$y = c' \cdot T_z + d', \quad (7)$$

Substituting Eqs. (4), (6), and (7) into the water balance equation (Eq. 1) leads to

$$TWS(t) = (1 - a' \cdot T_z - b') \cdot P(t) + (1 - c' \cdot T_z - d') \cdot TWS(t - 1). \quad (8)$$

After merging constant terms simplifying the signs, the recursive formulation adopted in this study is obtained as

$$TWS(t) = (a \cdot T_z + b) \cdot P(t) + (c \cdot T_z + d) \cdot TWS(t - 1). \quad (9)$$

Here, the term  $(1 - b)$  can be interpreted as the fraction of precipitation that

directly contributes to the loss component  $ETR_1(t)$ ; therefore, parameter  $b$  represents the effective proportion of precipitation that contributes to  $TWS$ . Similarly,  $(1 - d)$  represents the fraction of antecedent storage contributing to the loss term  $ETR_2(t)$ , and thus parameter  $d$  quantifies the fraction of previous storage retained in the current storage state. Parameters  $a$  and  $c$  represent temperature modulation effects.

To further support these interpretations, additional independent diagnostics were conducted. Basin-scale scatter analyses indicate a systematic relationship between parameter  $b$  and the observed loss ratio  $(ET + R)/P$  (Fig. R11). The correlation indicates that a larger loss ratio comes with a smaller  $b$ , supporting its interpretation as an effective precipitation conversion coefficient.

We also clarify that parameter  $d$  in our model does not correspond to the hydrological concept of groundwater turnover time. Groundwater turnover time typically refers to the time required to completely renew stored groundwater within an aquifer system, often on centennial to millennial timescales (Befus et al., 2017). In contrast, parameter  $d$  in the reconstruction model represents the fraction of basin-scale water storage retained from one time step to the next after accounting for losses, effectively characterizing the memory or persistence of total water storage variations.

Accordingly, when  $d$  approaches 1, losses associated with antecedent storage become relatively small, and storage variations are primarily driven by current precipitation input. Conversely, when  $d$  deviates significantly from 1, a larger fraction of current losses originates from the release of previously stored water,

reflecting stronger temporal persistence or memory effects in basin-scale storage dynamics.

Furthermore, since *TWS* includes soil moisture, surface water, and groundwater components, and since soil and surface water typically exhibit much faster turnover than groundwater, parameter *d* should be interpreted as representing the integrated memory behavior of multiple storage components rather than being directly comparable to groundwater renewal processes.

In the revised manuscript, we have incorporated the complete derivation of the above equations into the original manuscript at Line 299. Figures R S9–S11 in this response have been included in the Supplementary Materials as Figs. S10–S12. To avoid using excessive space in this response, the detailed manuscript edits are omitted here.

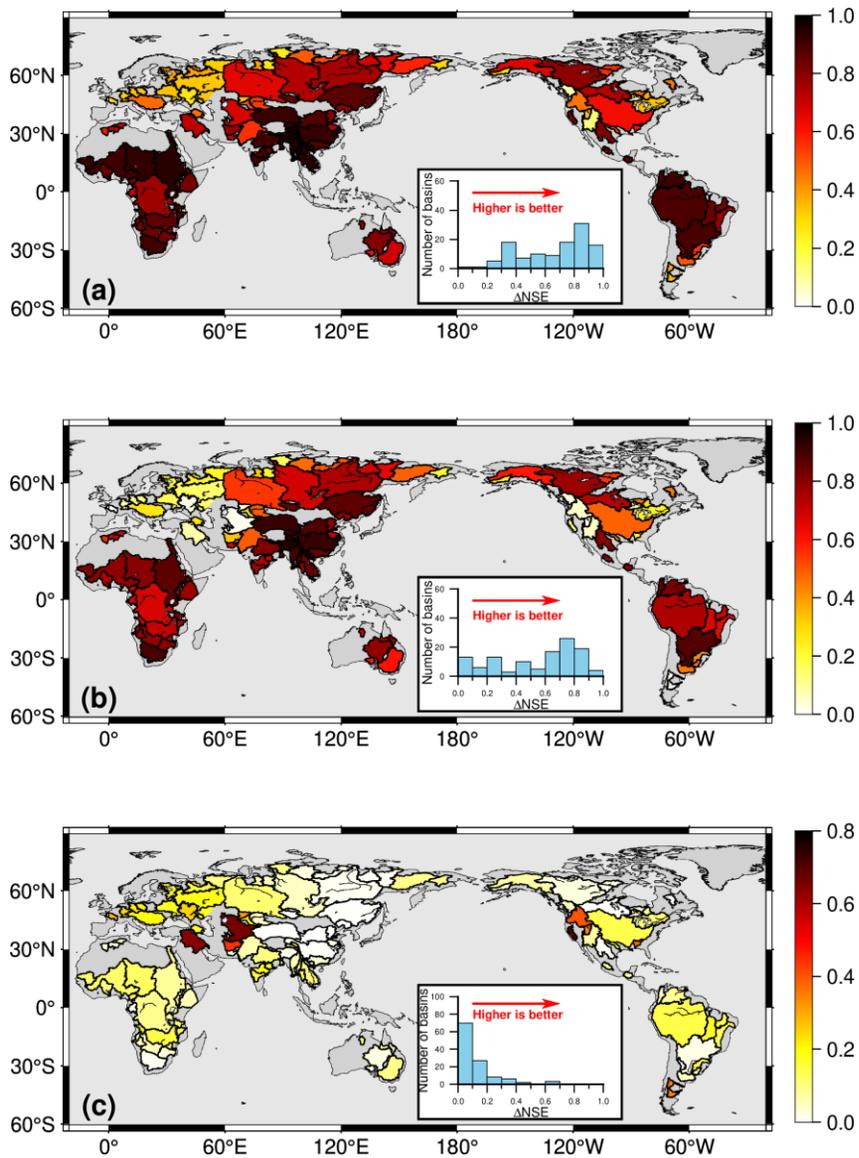


Figure R10. (a) Spatial distribution of the coefficient of determination ( $R^2$ ) from the multiple linear regression model  $ETR = x \cdot P(t) + y \cdot TWS(t - 1)$ ; (b) Spatial distribution of  $R^2$  from the univariate linear regression  $ETR = x \cdot P(t)$ ; (c) Map of the difference between panels (a) and (b). All basins pass the F-test at the 1% significance level ( $p < 0.01$ ).

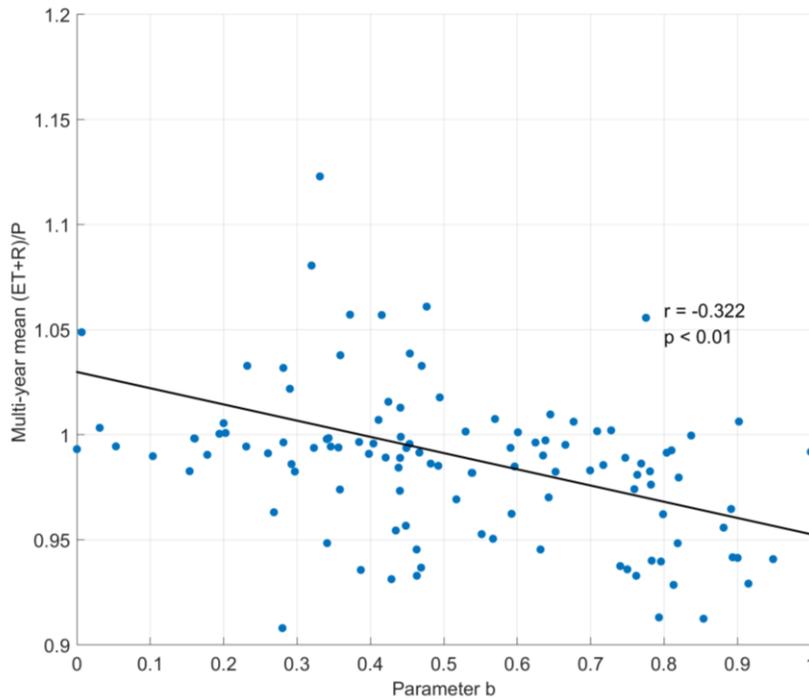


Figure R11. Scatter relationships across 116 global river basins between parameter  $b$  and the long-term mean loss ratio  $(ET + R)/P$ . The black dashed line denotes the linear regression fit. Pearson correlation coefficient  $r$  and the p-value for the regression slope are annotated on each panel.

**Comment 5: As you mentioned, GLDAS CLSM (0.25°, assimilates GRACE) offers a quasi-independent reference. GLDAS provides reasonable TWS estimates. You should compare the results between the reconstruction and GLDAS, and illustrate the advantages of your method.**

**Response:** Thank you for recommending a comparison with GLDAS CLSM. This issue has been addressed above under **Concern (iii)**. In the revised manuscript we added a direct comparison between our monthly reconstruction and GLDAS CLSM; please see our Response to **Concern (iii)** for full details.

## Responses to Minor Comments

1) Line 233: add the missing equation number.

**Response:** Thank you for pointing this out. We have now added the missing equation number to the corresponding equation in Line 233 in the original manuscript.

2) Line 333: "2.81 % of grids with  $NSE > 0.8$ " disagrees with the histogram.

**Response:** We thank the reviewer for identifying this inconsistency. After re-examining the histogram data shown in Fig. 3(b) of the original manuscript, we found that the proportion of grids with  $NSE > 0.8$  is 2.79%. We have corrected the statement to match the actual histogram value in Line 333 of the original manuscript.

3) In Figures 6 and 16 (b–j), the letters (a, b, c, etc.) in panel (a) are preferable to ID numbers.

**Response:** We appreciate the reviewer's helpful suggestion. In the revised manuscript, we have removed the numerical identifiers in panels (b–j) of Figs. 6 and 16 and replaced them with alphabetical labels (a, b, c, etc.) to ensure consistency with panel (a) and to improve overall figure readability. The revised figures are provided in this response as Figs. R12 and R13.

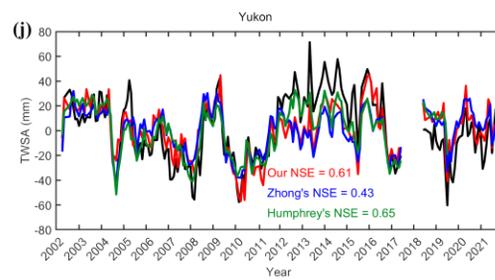
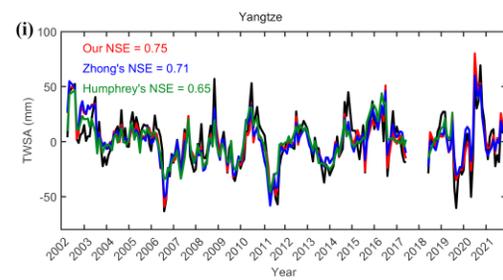
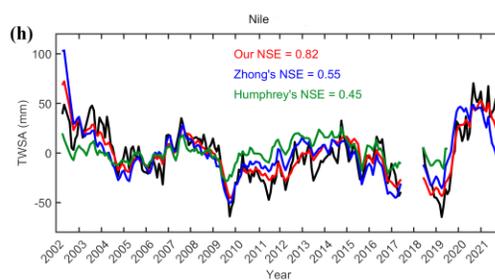
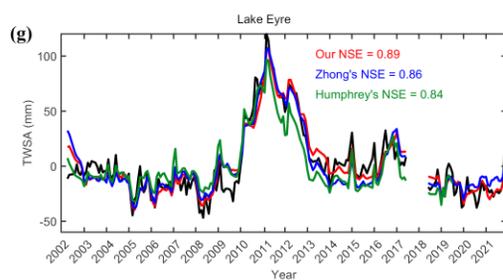
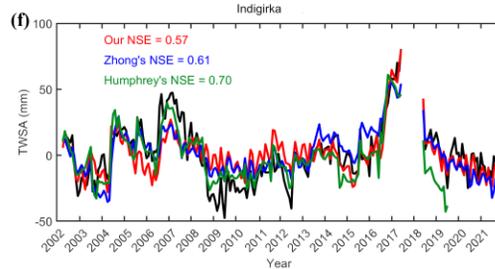
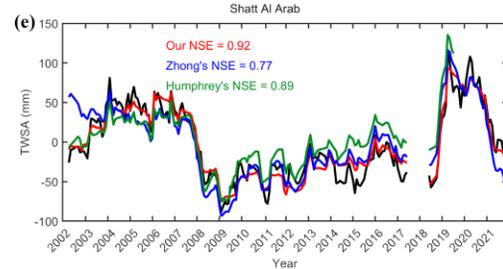
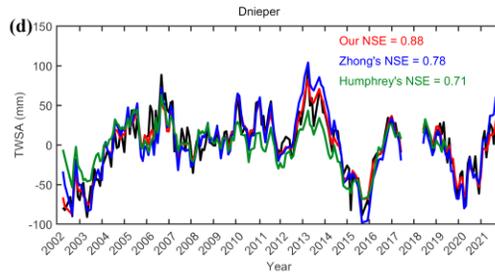
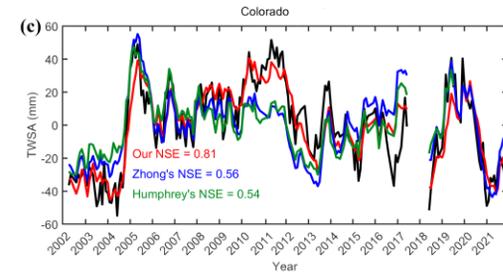
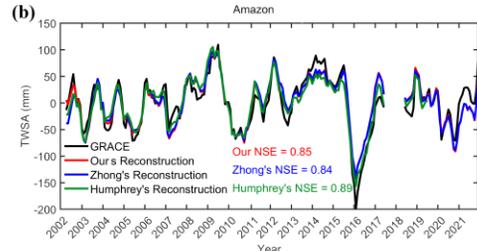
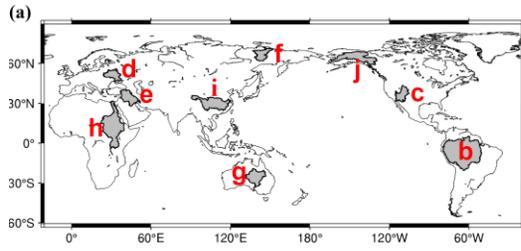


Figure R12. Time series of GRACE/GRACE-FO TWSA and reconstructed TWSA, both de-seasonalized and de-trended, for the nine selected river basins (b-j) from 2002 to 2021. The global distribution of the nine selected river basins (a).

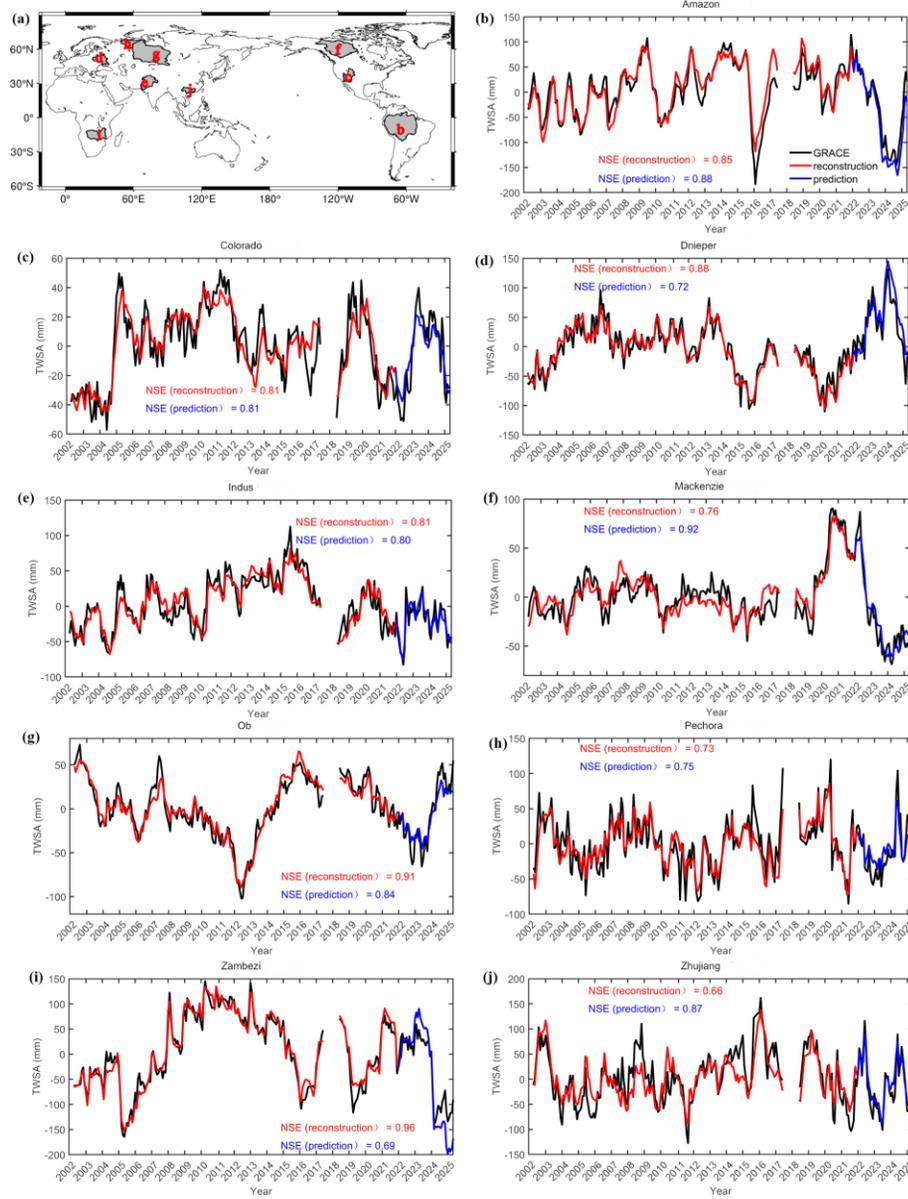


Figure R13: Time series of GRACE TWSA (black), reconstructed TWSA (red), and four-parameter climate-driven forecasts (blue) from January 2022 to May 2025 in nine basins (b-j). The global distribution of the nine selected river basins (a).

## References

References Befus, K. M., Jasechko, S., Lujendijk, E., Gleeson, T., and Cardenas, M. B.: The rapid yet uneven turnover of Earth's groundwater, *Geophys. Res. Lett.*, 44, 5511-5520, 2017.

Burek, P. and Smilovic, M.: The use of GRDC gauging stations for calibrating large-scale hydrological models, *Earth Syst. Sci. Data*, 15, 5617-5629, 2023.

Gao, Y., Luo, Z., Liu, H., Wang, L., Chen, X., and Li, H.: Reconstruction of global long-term daily streamflow dataset using machine learning models for revealing streamflow changes, *Journal of Hydrology: Regional Studies*, 64, 103148, 2026.

Zhong, Y., Tian, B., Kim, H., Yuan, X., Liu, X., Zhu, E., Wu, Y., Wang, L., and Wang, L.: Over 60% precipitation transformed into terrestrial water storage in global river basins from 2002 to 2021, *Commun. Earth Environ.*, 6, 53, 2025.