Response to Editor and Reviewers

We sincerely thank the editor and reviewers for the careful evaluation of our manuscript entitled “Reconstructing Climate-Driven Global Terrestrial Water Storage (2002–2021) Using a Four-Parameter Linear Recursive Model.” We have carefully considered all comments and substantially revised the manuscript. The main revisions are summarized as follows:

(1) Title revision

Following the reviewer’s suggestion, we revised the manuscript title by adding “Using a Four-Parameter Linear Recursive Model” to more clearly reflect the core contribution of the study.

(2) Inclusion of CSR mascon–based reconstructions

To address concerns about the absence of CSR products, we have added CSR-based reconstruction results and incorporated them into the comparative analysis. The corresponding figure has been incorporated into the revised manuscript as Figure 3, and the associated description has been added.

(3) Clarification of temperature forcing datasets (GLDAS-2.2 vs ERA5-Land)

To avoid ambiguity regarding temperature forcing datasets, we have added clarifications in the Methods section to explicitly distinguish GLDAS-2.2 temperature variables from ERA5-Land meteorological forcing data. In addition, new supplementary figures have been included to illustrate spatial and temporal

differences between these datasets. The corresponding figure has been incorporated into the Supplementary Material as Figure S1, and the associated description has been added.

(4) Clarification of basin subdivision strategy

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.

(5) Additional basin-scale comparisons with GLDAS-CLSM

Following the editor's and reviewer's suggestions, 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 corresponding figure has been incorporated into the Supplementary Material as Fig. S5.

(6) New daily-scale validation with independent datasets

Following the editor's and reviewer's suggestions, new validations of daily-scale reconstruction results have been added as a new section in the manuscript. The corresponding figure has been included in the revised manuscript as Fig. 7 and Fig. S7.

(7) Independent training–validation experiment

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 4.2.5 in the manuscript, with supporting figures included in the

revised manuscript as Fig. 8 and Fig. 9.

(8) Substantial revision of parameter derivation and physical interpretation

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. S2–S3 and Fig. S9.

(9) Reorganization of Supplementary figures and updates to the Abstract and Conclusions

We have made substantial revisions to the Supplementary Materials. Specifically, the figures have been reorganized and renumbered according to the order in which they are first cited in the manuscript. Original Figs. S2 and S5 in the previous Supplementary Materials (now Figs. S4 and S8 in the revised Supplementary Materials) have only been repositioned, with no changes made to their content. In addition, Fig. 16 in the original manuscript has been removed. The Abstract and Conclusions have also been revised accordingly to reflect the newly added daily-scale validation, comparison with GLDAS-based products, and the independent training–validation assessment.

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

text. We hope that these revisions adequately address the editor's and reviewers' concerns and significantly improve the clarity of the manuscript.

Sincerely,

Pu Xie and Shuang Yi

March 2026

Response to the Editor

Comment 1: please try to conduct validation at daily scale

Response: Thank you very much for this helpful suggestion. We have carried out additional validation at the daily scale as requested. The new analysis and the corresponding revisions in the manuscript are described in detail in our response to **Referee #2, Concern (i)**.

Comment 2: demonstrate the add value against reanalysis products

Response: Thank you for this important comment. We have followed your suggestion and carried out an explicit comparison between our reconstructed TWSA and GLDAS GLSM products to demonstrate the added value of our approach. The corresponding new analysis and manuscript revisions are described in detail in our response to **Referee #2, Concern (iii).**

Response to Referee #1

Comment 1: The title could be improved to more accurately reflect the core contribution of the paper.

Response: We thank the reviewer for this helpful suggestion regarding the manuscript title. We fully agree that the title plays a critical role in accurately conveying the core contributions of the study. Following your suggestion, we have revised the title accordingly. The proposed new title is: *Reconstructing Climate-Driven Global Terrestrial Water Storage (2002–2021) Using a Four-Parameter Linear Recursive Model*

We believe that this revised title more clearly represents the main objectives and contributions of the study. We would be pleased to further refine the title should the reviewer or editor have additional suggestions.

Comment 2: The authors could provide the rationale for not using the CSR Mascon data product in this study.

Response: We thank the reviewer for raising this important and well-targeted question. We fully acknowledge the value of the CSR Mascon product in terrestrial water storage studies.

Specifically, we have added the reconstruction calibrated against the CSR Mascon product (hereafter CSR-REC) and calculated the NSE between CSR-REC and the CSR Mascon solution (CSRM), as shown in Fig. R1(c) of this response. The

results indicate that the spatial distribution of NSE derived from CSR-REC is generally consistent with those obtained from JPL-REC and GSFC-REC. Most grid cells exhibit NSE values within the range of 0.5–0.8, with relatively high agreement in mid- and high-latitude regions where meteorological observations are denser. This consistency further demonstrates the robustness of the proposed reconstruction framework and suggests that it is not strongly sensitive to the choice of mascon product.

In the revised manuscript, we have incorporated CSR-REC into the overall analysis to provide a more comprehensive evaluation of the reconstruction performance. The original Fig. 3 in the manuscript has been updated to include the newly added CSR-REC results. Fig. R1 in this response letter replaces the original Fig. 3 and will be included as the updated version in the revised manuscript. In addition, we have inserted the following statement after line 333 in Section 4.1 of the original manuscript, which corresponds to Line 413 in the tracked-changes version of the revised manuscript:

We also compared the reconstruction calibrated against the CSRM product (hereafter referred to as CSR-REC) with the CSRM product to further evaluate robustness of the reconstruction method. The spatial distribution of NSE (Fig. 3c) is broadly consistent with those obtained for JPL-REC and GSFC-REC. The NSE distributions derived from the three mascon-based calibrations are largely consistent, indicating that the reconstruction approach exhibits limited sensitivity to the source of the mascon products, which is consistent with findings reported in previous studies

((Humphrey and Gudmundsson, 2019)).

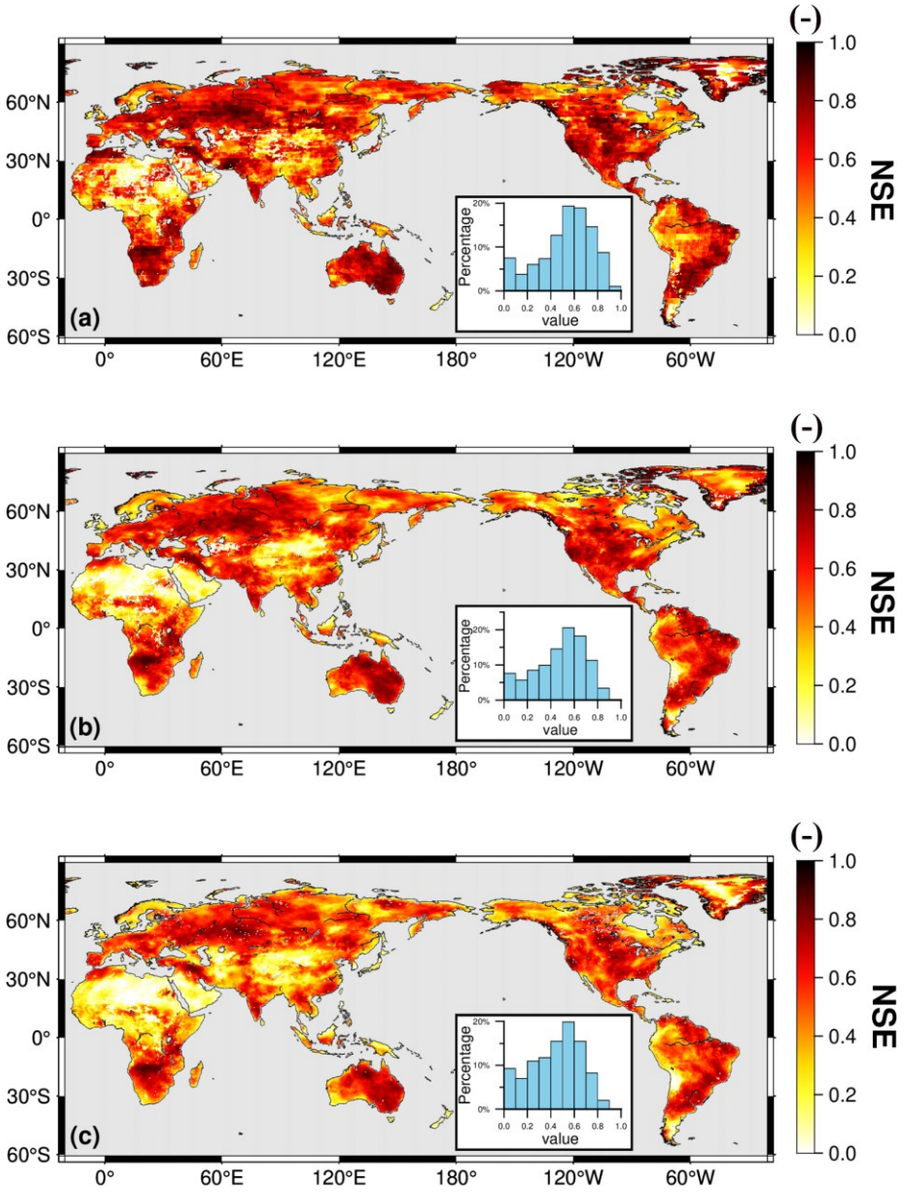

Figure R1: Spatial distribution of NSE (of de-seasonalized, de-trended anomalies) between our reconstruction and (a) JPLM, (b) GSFCM, and (c) CSRM for 2002–2021. Histograms show the distribution of NSE values across all global grid points.

Comment 3: The GLDAS-2.2 data is also forced with the meteorological analysis fields from the operational European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting System. Due to the data agreement with ECMWF, this GLDAS-2.2 daily product does not include the meteorological forcing fields. The GLDAS-2.2 data are archived and distributed in NetCDF format. Thus, the authors should explain the differences between the two temperature datasets.

Response: We sincerely thank the reviewer for this careful and highly professional comment. We fully agree that clarifying the origin and differences of the temperature datasets is essential for avoiding confusion regarding the model forcing variables, and we greatly appreciate the reviewer for pointing out this important issue.

First, we would like to emphasize that the two temperature datasets used in this study are indeed different. Fig. R2(a) presents the basin-scale root-mean-square (RMS) differences between ERA5-Land 2 m air temperature and GLDAS-2.2 surface temperature across 116 global river basins. Please note that we have removed the mean values from each dataset prior to comparison. The figure clearly shows that apparent differences exist between the two datasets at the basin scale. Fig. R2(b) shows one example in the Amazon basin derived from ERA5-Land and GLDAS-2.2. As shown, although the two datasets exhibit generally consistent seasonal variability,

noticeable discrepancies at the daily scale are present.

Regarding the source of these differences, the official documentation further clarifies that ERA5-Land and GLDAS-2.2 are based on distinct atmospheric forcing systems and land surface models, as well as temporal resolution (Fig. R3 and Fig. R4). ERA5-Land is forced by meteorological variables from the ERA5. In contrast, GLDAS-2.2 is forced by the operational ECMWF Integrated Forecasting System (IFS) analysis fields. In addition, ERA5-Land employs the H-TESSEL land surface model, whereas GLDAS-2.2 uses the Catchment land surface model. Finally, ERA5-Land provides hourly temperature fields that retain sub-daily variability (with daily temperature in this study obtained through temporal aggregation), while GLDAS-2.2 provides temperature fields at a daily temporal resolution, further contributing to the observed discrepancies between the two datasets.

To address this issue and avoid ambiguity, we have added a clarification in the Methods section of the revised manuscript to explicitly distinguish between the temperature variables from GLDAS-2.2 and the meteorological forcing temperature provided by ERA5-Land. Specifically, Fig. R2 in this response has been incorporated into the Supplementary Materials of the revised manuscript as Fig. S1. Additionally, we plan to insert the following explanatory text after Line 155 in the original manuscript, which corresponds to Line 167 in the tracked-changes version of the revised manuscript:

It should be noted that, due to the data usage agreement with the European Centre for Medium-Range Weather Forecasts (ECMWF), which prohibits

redistribution of products from the Integrated Forecasting System (IFS), the GLDAS-2.2 daily product does not include the meteorological forcing fields. Instead, GLDAS-2.2 provides land surface variables simulated by the Catchment land surface model, which is forced by meteorological analysis fields from the operational ECMWF IFS.

Although both ERA5-Land and GLDAS-2.2 are ultimately based on meteorological information produced within the ECMWF framework, their temperature datasets exhibit a discrepancy of approximately 1–2 °C (Fig. S6) due to different data assimilating methods and models.

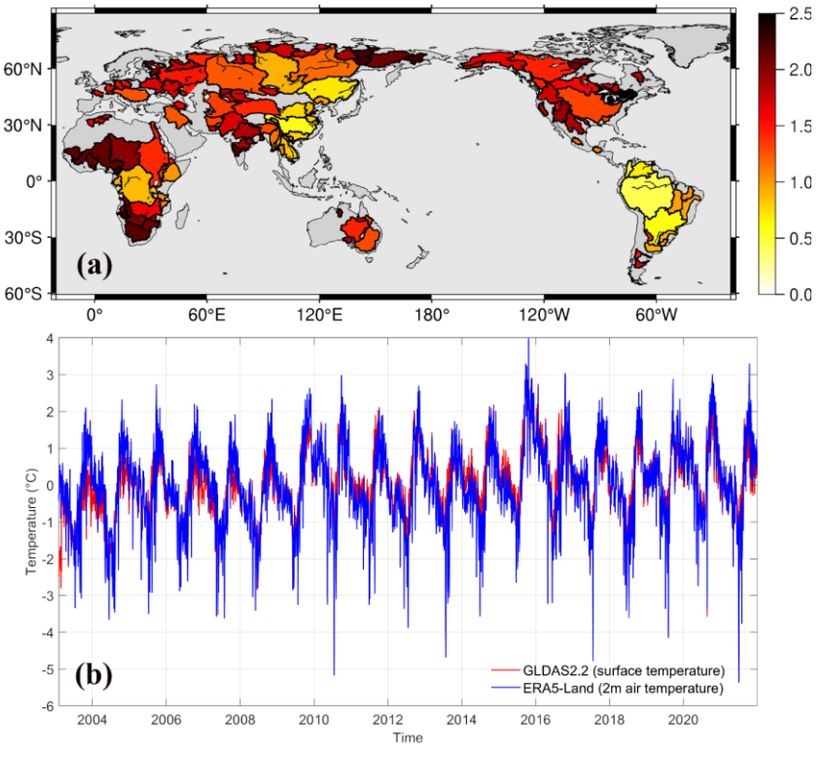

Figure R2: Spatial distribution of RMS differences between ERA5-Land 2 m air temperature and GLDAS-2.2 surface temperature over 116 global river basins (a). Comparison of daily ERA5-Land 2 m air temperature and GLDAS-2.2 surface temperature over the Amazon basin (b).

Here we document the **ERA5-Land** dataset that, in its consolidated version, **covers the period from January 1950 to 2-3 months before the present**. In addition, the **ERA5-Land-T** version **delivers non-checked close to Near-Real-Time (NRT) daily updates.** ERA5-Land-T is synchronized with the close to NRT daily updates provided by the ERA5 climate reanalysis (ERA5T).

ERA5-Land is a replay of the land component of the ERA5 climate reanalysis, forced by meteorological fields from ERA5. Note that ERA5-Land was forced by the ERA5 Preliminary back extension, rather than by the final ERA5 back extension for data older than 1979. ERA5-Land uses forcing fields based on the final release of ERA5 (i.e., expver=0001) for data after 1978. ERA5-Land comes with a series of improvements making it more accurate for all types of land applications. In particular, ERA5-Land runs at enhanced resolution (9 km vs 31 km in ERA5). The temporal frequency of the output is hourly and the fields are masked for all oceans, making them lighter to handle. Click this link here for comparison of the ERA5-Land features against other ECMWF reanalyses.

ERA5-Land is produced under a single simulation, without coupling to the atmospheric module of the ECMWF's Integrated Forecasting System (IFS) or to the ocean wave model of the IFS. **It runs without data assimilation**, making it computationally affordable for relatively quick updates. For example, if significant improvements of the land surface model are implemented, the whole or part of the dataset can be reprocessed in a relatively short period. Also, updates are possible in case improved auxiliary datasets are used as input for the production.

⚠ Observations indirectly influence the simulation through the atmospheric forcing of ERA5. This forcing drives the ERA5-Land single simulation and it has been obtained by assimilating observations through a 4D-VAR data assimilation system and a Simplified Extended Kalman Filter.

The core of ERA5-Land is the Tiled ECMWF Scheme for Surface Exchanges over Land incorporating land surface hydrology (H-TESSEL). It uses version CY45R1 of the IFS.

Currently, ERA5-Land dataset contains only one (9 km) high resolution realisation (HRES). Uncertainty information can currently be used from the reduced resolution ten member ensemble (EDA) of ERA5. The data are available at a sub-daily and monthly frequencies. **For convention and consistency** with the previous ERA-Interim/Land dataset, **the data parameters are labelled as analyses and short** (24 hour) **forecasts** initialised once daily from analyses at 00 UTC. Accumulation parameters are only available from the forecasts and the convention used in ERA5-Land differs from that for ERA5.

Currently, the data **can only be downloaded on a regular latitude/longitude grid of 0.1°x0.1° via the CDS catalogue**. ECMWF member states with access to the ECMWF Meteorological Archival and Retrieval System (MARS) can also retrieve the data in the native grid.

⚠ **NOTE**: Please, note that **since 1st Jan 2020** the new ECMWF Meteorological Interpolation and Regridding interpolation package (**MIR**) has been used to interpolate the atmospheric forcing of ERA5 into the ERA5-Land grid. While this change will be unnoticeable for the overwhelming majority of users, locally and under very limited conditions (some areas with high orography, some coastal points) some fields may suffer of a very small discontinuity this day.

Land Surface Model

H-TESSEL is the land surface model that is the basis of ERA5-Land. The H-TESSEL version used in the production of ERA5-Land corresponds to that of the IFS model documentation CY45R1.

Figure R3. Official documentation of ERA5-Land dataset from ECMWF. (Source: ERA5-Land: data documentation , accessed on 2026-01-22)

1.2.4 GLDAS-2.2

The GLDAS-2.2 Daily Catchment model simulation started on February 1, 2003 using the conditions from the GLDAS-2.0 Daily Catchment model simulation. This simulation was forced with the meteorological analysis fields from the operational European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting System (https://www.ecmwf.int/en/publications/ifs-documentation). The total terrestrial water anomaly observation from Gravity Recovery and Climate Experiment (GRACE) was assimilated (Li et al., 2019). The GRACE RL06 and GRACE Follow-On data were provided by the Center for Space Research at the University of Texas (Save et al., 2012; Save et al., 2016). The Daily Catchment model simulations use the UMD land cover scheme from AVHRR land cover map. Due to the data agreement with the ECMWF that prohibits dissemination of the IFS product, this GLDAS-2.2 Daily product does not include the meteorological fields.

Figure R4. Official NASA GLDAS Version 2 Data Products README provided by the GES DISC at NASA Goddard Space Flight Center. (Source: https://disc.gsfc.nasa.gov/datasets/, accessed on 2026-01-22)

Minor Comment 1: L50, Usage of 'Eq' is not defined before.

Response: Thank you for your advice. We have replaced "Eq." with "equation" in Line 50 of the original manuscript, which now appears as Line 57 in the tracked-changes version of the revised manuscript.

Minor Comment 2: L135-136, the citation format '(Lan and Wenke, 2022)' is wrong. The same for '(Muñoz-Sabater et al., 2021; Muñoz Sabater, 2019)'.

Response: We appreciate the reviewer's careful attention to citation formatting. We have corrected these citation errors in accordance with the journal's style requirements.

Minor Comment 3: L179 and L185, abbreviations with 'where' at the beginning of a sentence are unnecessary.

Response: Thank you for your advice. We have revised the sentences on Lines 179 and 185 in the original manuscript to remove the redundant use of "where" at the beginning of definitions. These sentences now appear on Lines 205 and 211 in the tracked-changes version of the revised manuscript.

Minor Comment 4: L680, 'In some other basins' to 'In other basins'.

Response: Thank you for your advice. The phrase "In some other basins"

appeared in Line 680 of the original manuscript, which was part of Section 5.3. As Section 5.4 has been removed from the revised manuscript, this sentence has also been deleted accordingly.

Response to Referee #2

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. R5(a) and Fig. R5(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. R5(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. R5(b), the agreement between our reconstruction and GLDAS-2.2 is relatively poor. However, time series analysis in Fig. R6(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. R6b, 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 4.2.4 entitled *"Comparison of Daily Time Series"* has been added. Figure R5 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. Figure R6 presented in this response has been incorporated into the Supplementary Material as Fig. S7. The following text has been added after Line 456 in the original manuscript, which corresponds to Line 566 in the tracked-changes version of the revised manuscript:

Although our model is primarily constrained by monthly GRACE observations, the method can also reconstruct daily characteristics in TWS changes based on the daily climate drivers. Here, we evaluated our reconstructed daily TWSA by comparing

it with two different reference datasets, namely the daily ITSG-Grace2018 solution derived using a Kalman smoothing approach and the GLDAS-2.2 Daily product. The comparison with ITSG-Grace2018 evaluates whether the daily reconstruction is consistent with GRACE-based sub-monthly variability, whereas the comparison with GLDAS-2.2 assesses whether the daily reconstruction is consistent with physically based land-surface simulations.

We first assessed reconstruction performance at the basin scale by computing basin-averaged daily TWSA over major river basins worldwide. The difference in the time series was quantified using NSE (Fig. 7a&b). The NSE values are positive across most basins, with over half exceeding 0.5, indicating that our reconstruction results align well with the previous models. Although the reconstruction performs well overall relative to both reference datasets, its agreement with ITSG-Grace2018 is stronger in many basins. Specifically, 63 basins show NSE values greater than 0.6 in comparison with ITSG-Grace2018, whereas the count slightly reduces to 51 when compared with GLDAS-2.2. This is likely because our reconstruction is calibrated under constraints from GRACE/GRACE-FO TWSA, while ITSG-Grace2018 is also a daily product derived from GRACE observations.

To better illustrate sub-monthly variability, we showed daily TWSA time series in the Mississippi River basin in Fig. 7c. The results show similar sub-monthly variabilities among the reconstructed GRACE-REC, ITSG-Grace2018, and GLDAS-2.2 time series. One exception is found in the Indus basin between our GRACE-REC and GLDAS-2.2. The main reason is that GLDAS-2.2 fails to capture a three-year

drought spanning from 2016 and 2018 (see Fig. S7 for details). The reconstructed daily TWSA is consistent with the corresponding reduction in the 6-month moving average of precipitation. By contrast, the GLDAS-2.2 daily TWSA remains comparatively flat during this period, suggesting that it may underestimate the sensitivity of basin-scale storage to interannual hydroclimatic forcing in this basin. Figure S7(b) further shows that our monthly reconstruction remains close to the GRACE/GRACE-FO observations, particularly around the resumption of GRACE-FO observations in June 2018, whereas GLDAS-2.2 tends to overestimate TWSA at that time. This discrepancy highlights the indispensable role of observation-based reconstruction approaches.

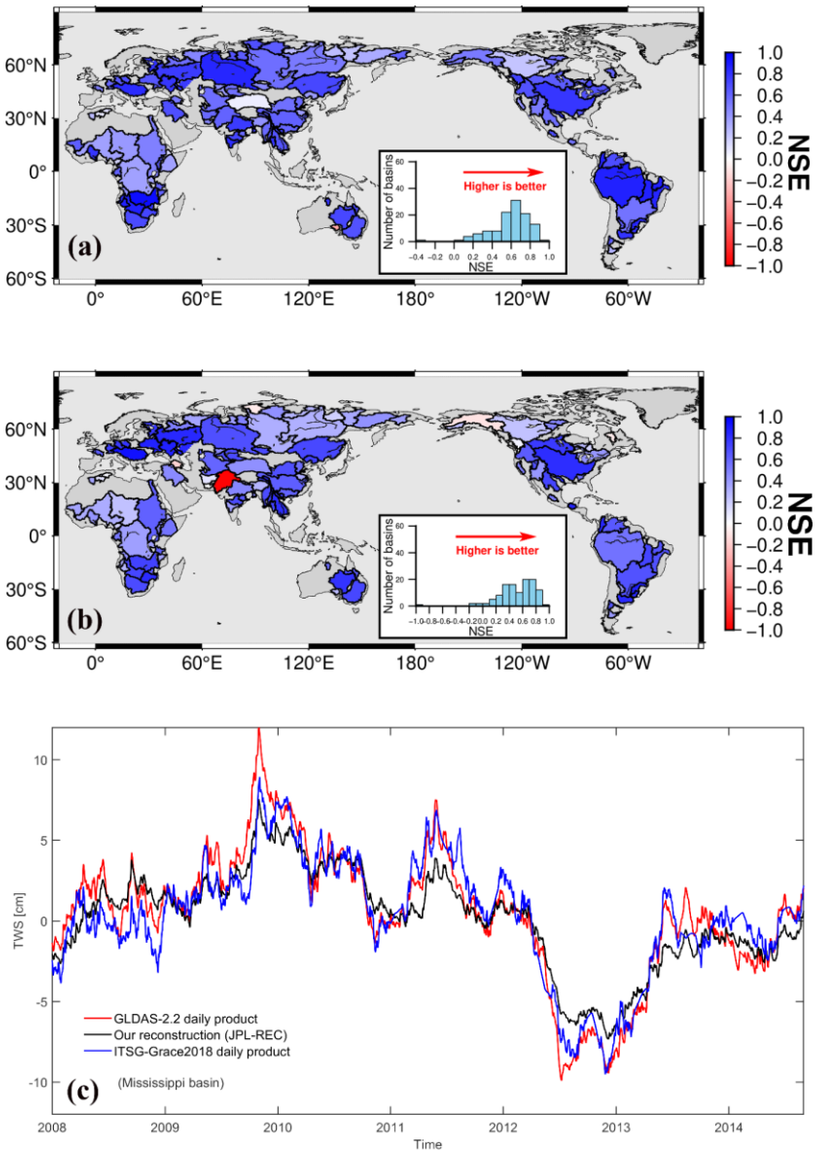

Figure R5: Spatial distribution of the Nash–Sutcliffe efficiency (NSE) of de-seasonalized and de-trended terrestrial water storage anomalies between the reconstructed daily TWSA and (a) ITSG-Grace2018 for the period 2003 to August 2016 and (b) GLDAS-2.2 daily product for the period 2003 to December 2021 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.

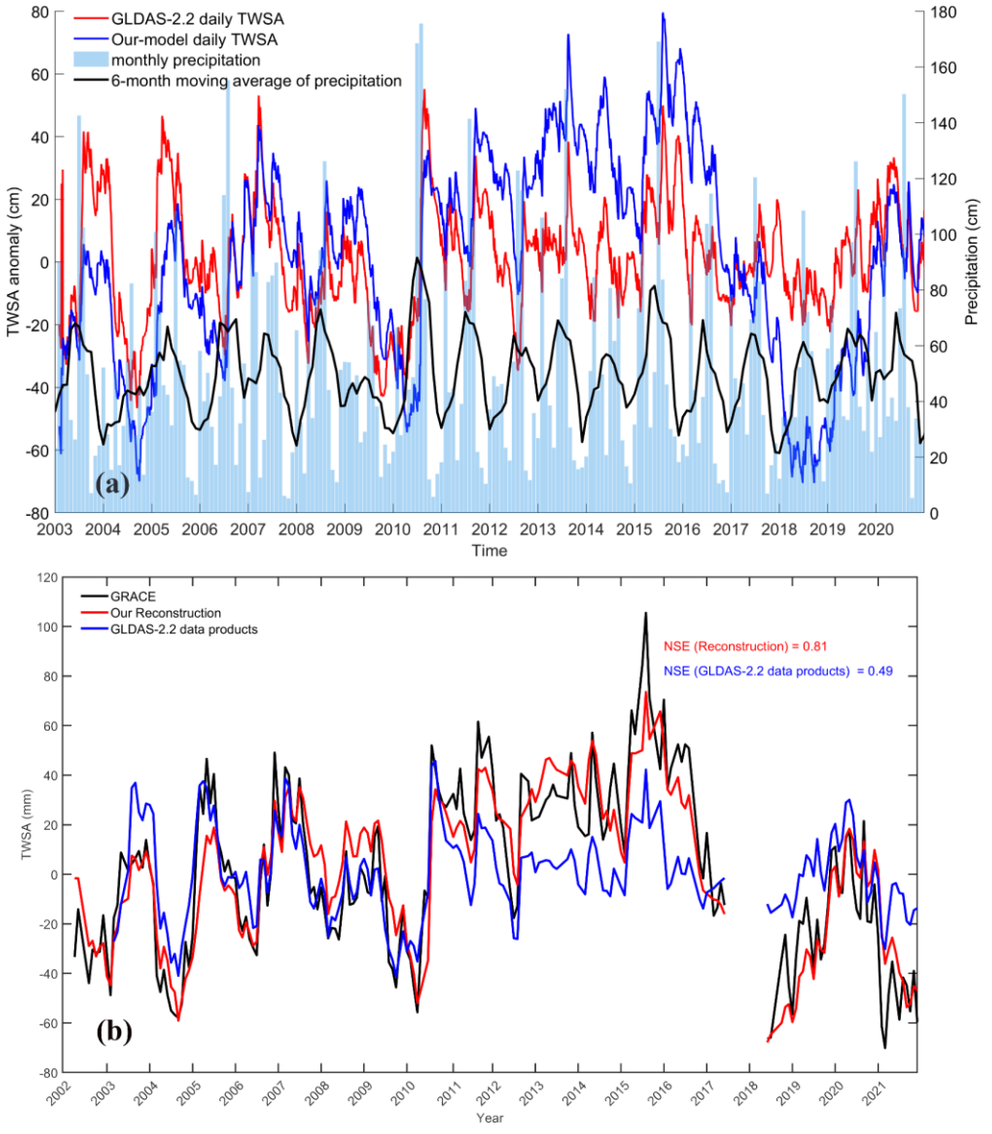

Figure R6: Comparison of daily and monthly TWSA reconstruction in the Indus River basin. (a) Detrended and deseasonalized daily TWSA from GLDAS-2.2(red line), our model-based daily TWSA reconstruction (blue line), monthly precipitation (light blue bars), and 6-month moving average precipitation (black 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

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 $1 \times 10^5\ km^2$ 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. The following text has been added after Line 161 in the original

manuscript, which corresponds to Line 181 in the tracked-changes version of the revised manuscript:

This study employs basin boundaries from the publicly available dataset "Major River Basins of the World" provided by the Global Runoff Data Centre (GRDC). This dataset has been widely used in large-scale hydrological studies (e.g., Burek and Smilovic (2023); Zhong et al. (2025); Gao et al. (2026)). We followed the strategy of Zhong et al. (2025) and ranked all basins according to drainage area to ensure spatial representativeness and avoid subjective selection. The 116 largest basins with drainage areas exceeding $10^5\ km^2$ *were selected for analysis.*

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. R8(a), Fig. R8(b)). In addition, we present the spatial distribution of the NSE difference (ΔNSE) between the two

products to explicitly quantify the added value of the proposed method (Fig. R8(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. R8 included in this response has been added to the Supplementary Materials as Fig. S6. In addition, Section 4.2.2 has been revised, with the section title changed from "*Comparison with the Zhong Model*" to "*Basin-scale comparison*." The first and second paragraphs of the revised content were inserted after Line 376 in the original manuscript, which now corresponds to Line 463 in the tracked-changes version of the revised manuscript, whereas the third and fourth paragraphs were inserted after Line 399 in the original manuscript, which now corresponds to Line 495

in the tracked-changes version of the revised manuscript. The revised content is as follows:

We compare our reconstruction (JPL-REC) with two representative types of products to systematically evaluate model performance at the basin scale: (1) a previous GRACE-based reconstruction product (Zhong-REC from (Zhong et al., 2025)), and (2) physically based land surface model simulations forced by meteorological data (GLDAS CLSM).

First, we adopted the JPLM as a common reference to evaluate the relative performance of different statistical reconstruction methods. We compared basin-averaged monthly TWSA from our JPL-REC and Zhong-REC against the JPLM and computed the NSE for each (Fig. 5). By comparing the NSE *values for the two approaches, we assessed their relative reconstruction accuracy.*

Second, we compared basin-averaged monthly TWSA from JPL-REC and GLDAS CLSM to highlight the improvement of our reconstruction method relative to the land surface model (Fig. S6). Across 116 major river basins worldwide, NSE *values for both products are computed against the JPLM, and the spatial distribution of the resulting* ΔNSE *is presented (Fig. S6(c)). The results show that* NSE *values from our reconstruction are higher in most basins, with* ΔNSE *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 and accuracy 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.

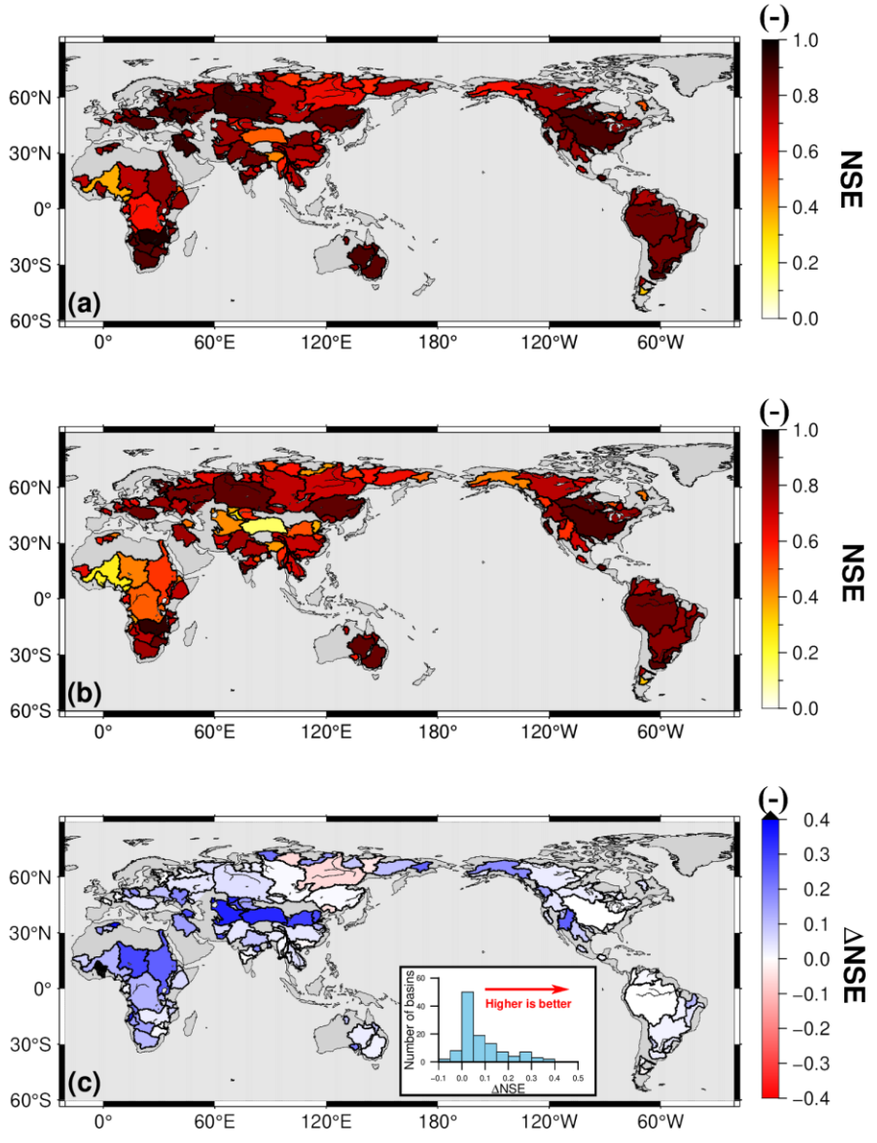

Figure R7: Spatial distribution of NSE (de-seasonalized, de-trended anomalies) between JPLM and two reconstruction models across 116 global river basins for the period 2002 to 2021. The NSE between JPLM and JPL-REC (a); the NSE between JPLM and the reconstruction by Zhong et al. (2025) (b); and their difference (ΔNSE =JPL-REC − Zhong-REC) (c). The inset histogram in (c) illustrates the distribution of ΔNSE across all basins, where positive values indicate better agreement with GRACE observations by our reconstruction model.

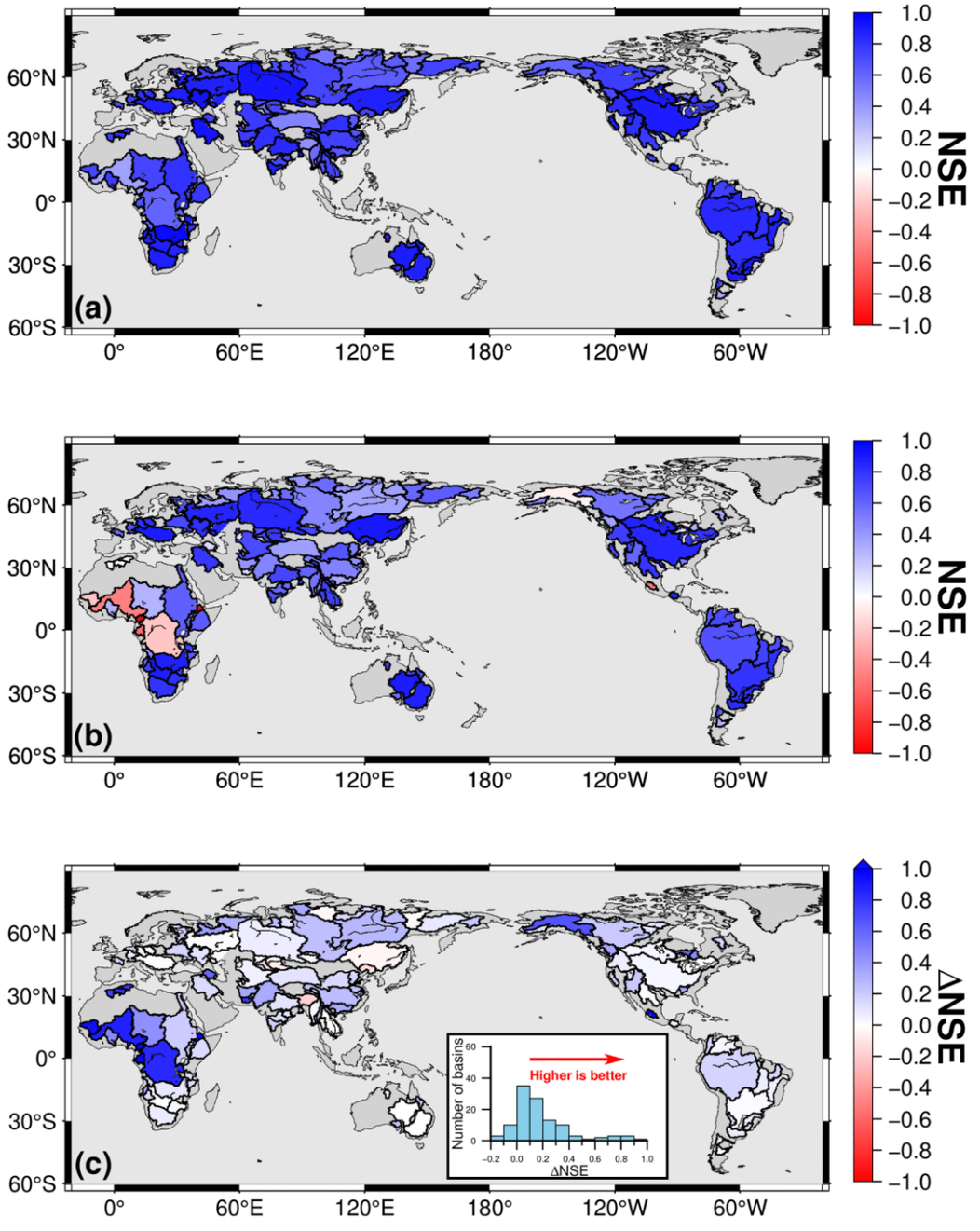

Figure R8: Spatial distribution of the NSE of de-seasonalized, de-trended TWSA between JPLM and (a) our reconstruction model and (b) GLDAS CLSM across 116 global river basins for 2003–2021; (c) their difference ($\Delta NSE = (a) - (b)$). The inset histogram in (c) illustrates the distribution of ΔNSE 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 β 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 β > 1. Actually, Zhong et al. (2025) also apply a similar method to reconstruct basin-scale TWS, which can ignores 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. R9).

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. R10). The results show that, for approximately 95% of global land grid cells, the NSE differences between the two scenarios fall within ± 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.

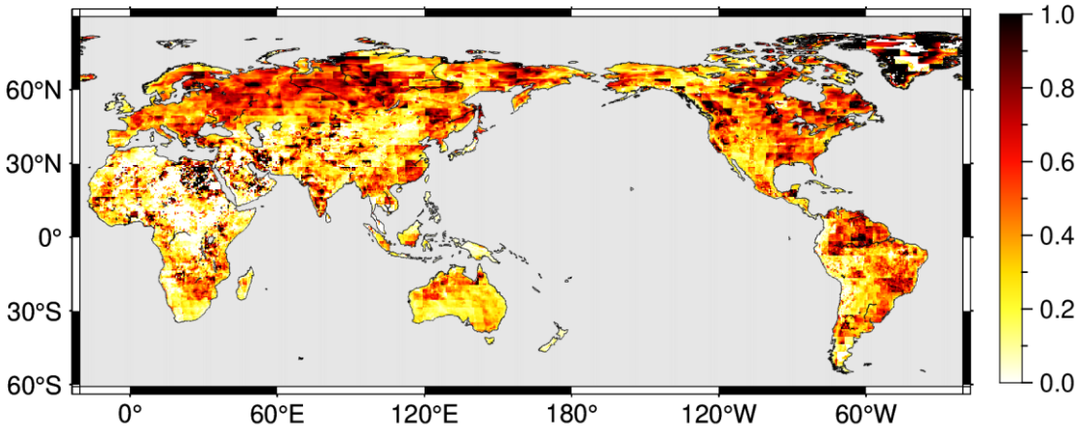

Figure R9. Spatial distribution of the calibrated parameter b across global land grid points. Parameters calibrated by monthly TWSA from JPLM based on ERA5-Land precipitation and temperature from 2002 to 2021.

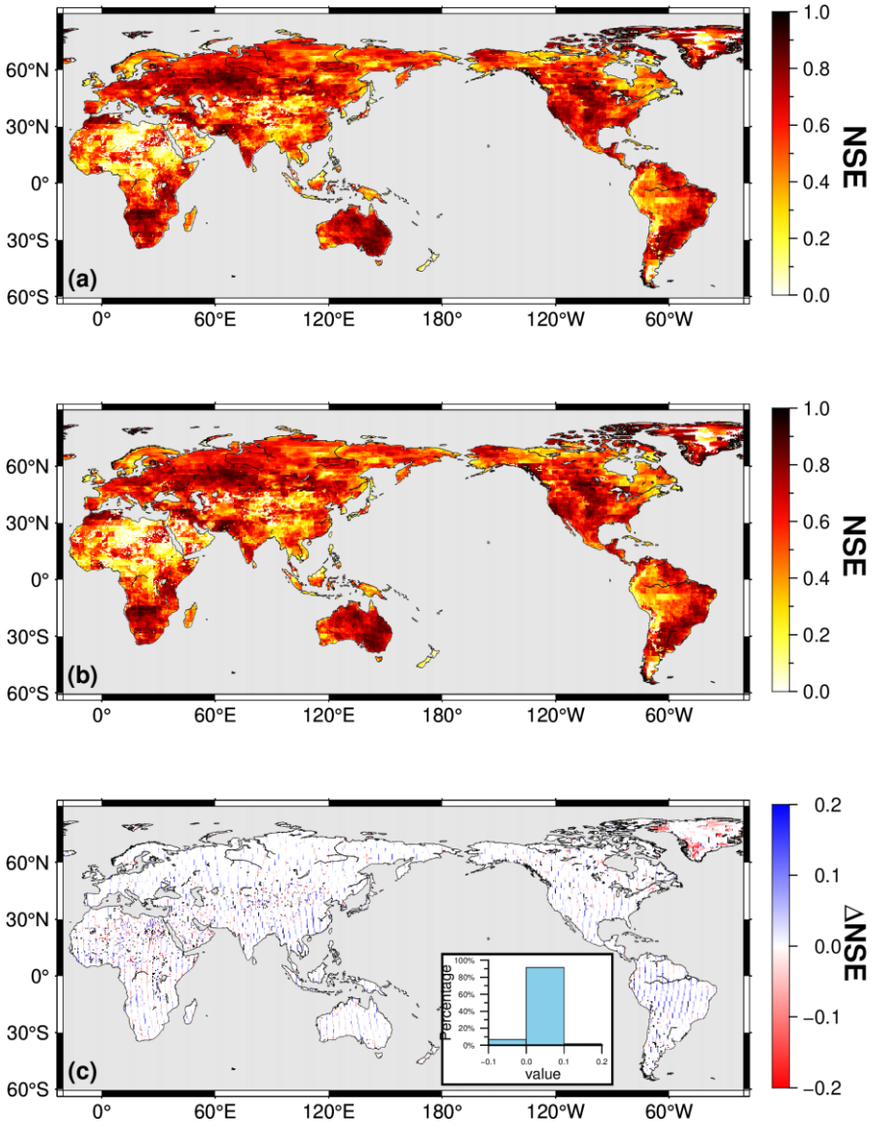

Figure R10. Sensitivity of grid-based reconstruction (JPL-REC) performance to parameter b constraints. (a) NSE between JPL-REC (with $0 < b < 1$) and JPLM; (b) NSE between JPL-REC (with $0 < b < 2$) and JPLM; (c) Difference in NSE between (a) and (b). Histograms show the distribution of NSE values across 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. R11(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. R12)—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 R11 and 12 in this response document has been included in the revised manuscript as Figure 8 and Figure 9. Meanwhile, we have inserted the following paragraph into the newly added Section

4.2.5 after Line 456 in the original manuscript, which now corresponds to Line 605 in the tracked-changes version of the revised manuscript:

We conducted a self-validation experiment to further evaluate the robustness of the proposed method by dividing the JPLM product into separate training and validation periods. Specifically, the period from April 2002 to December 2012 was designated as the training period and the model parameters were estimated, while the period from January 2013 to December 2023 was treated as an independent validation period.

We evaluated the model performance using the NSE by comparing JPL-REC with the JPLM in both the training and validation periods (Fig. 8). During the training period (Fig. 8a), most basins exhibit high reconstruction skill, with NSE values exceeding 0.6 over large parts of North and South America as well as Eurasia. A very similar spatial pattern is obtained for the validation period (Fig. 8b), with only a modest reduction in overall skill. Approximately 60 % of the basins retain NSE values above 0.5 in the validation period. This indicates that the model retains stable predictive capability beyond the calibration window and does not exhibit significant overfitting behavior.

The spatial pattern of the results during validation period reveals distinct regional characteristics. High NSE values are mainly found in large, humid basins with relatively dense observations, such as the Amazon, Mississippi, Yangtze River basins. In these regions, variability in TWS is strongly controlled by climate forcing. Consequently, our model can reliably reproduce the TWSA signals observed by

GRACE/GRACE-FO.

We further selected eight representative basins spanning high-latitude cold regions, tropical humid regions and typical monsoon-affected areas (Fig. 9). For each basin, the validation period TWSA (blue line) was compared with the reconstructed values during the training period (red line) and GRACE observations (black line). The results show that in certain basins, such as the Amazon, Yangtze, Mississippi and Murray, model performance during the extrapolation period remained comparable to or even exceeded that during the training period. This indicates that under these hydroclimatic conditions, the fixed-parameter model can reliably capture interannual TWSA variability driven by precipitation and temperature. In contrast, basins such as Chad and Khatanga show lower NSE during the validation period and larger discrepancies in certain years. This decline may be attributed to the limitations imposed by parameter nonstationarity and errors in the forcing data. These results suggest that when only precipitation and temperature forecasts are available, the proposed recursive model can be directly applied for estimating TWS changes during missing periods or into the future. Nevertheless, its applicability varies with hydroclimatic conditions, and for regions requiring high prediction accuracy, periodic parameter updates or the inclusion of additional forcing information may still be necessary.

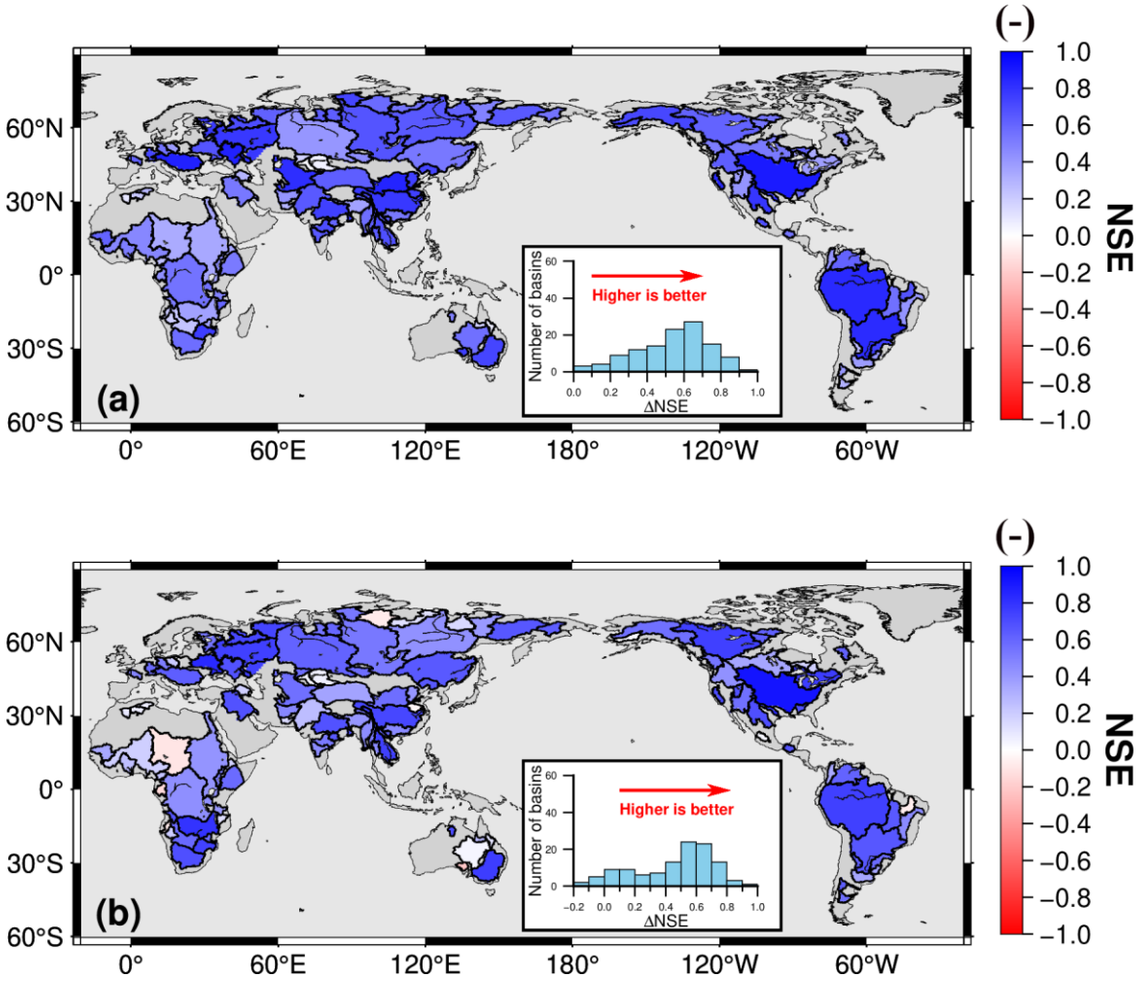

Figure R11. Comparison of reconstruction performance against JPLM data during the training and validation periods. (a) Spatial distribution of the NSE between reconstructed TWSA and JPLM data during the training period (April 2002–December 2012); (b) Spatial distribution of NSE during the validation period (January 2013–December 2023)

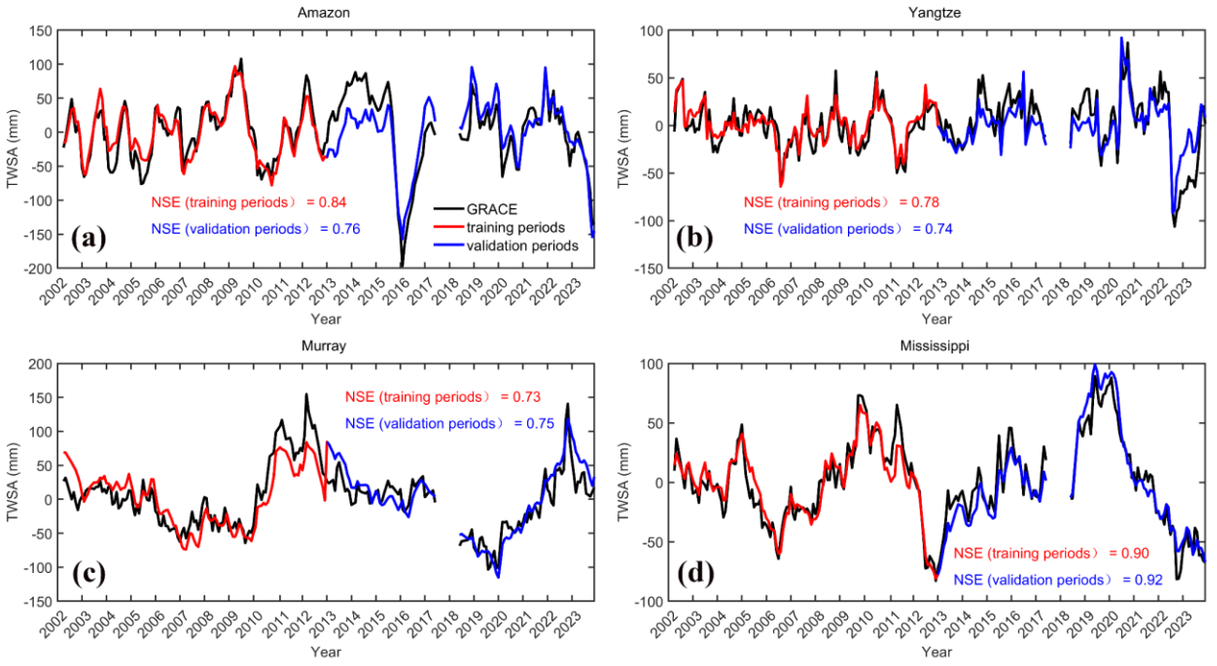

Figure R12. Time series comparison of reconstruction and GRACE/GRACE-FO TWSA in representative river basins during training and validation periods. Amazon, Yangtze, Murray, and Mississippi river basins (a–d). Black lines represent GRACE TWSA, red lines indicate model reconstructions during the training 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. R13, 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.

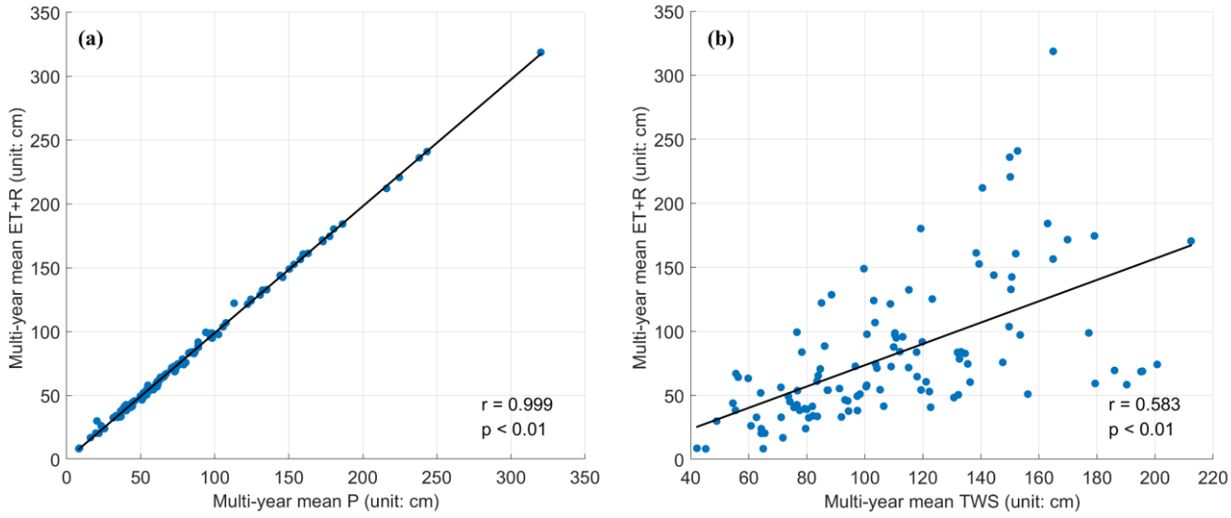

Figure R13. Scatter plot of multi-year mean precipitation versus multi-year mean evapotranspiration plus runoff ($ET + R$) for each basin (a). Scatter plot of multi-year mean TWS versus multi-year mean $ET + R$ for each basin (b). The black dashed line denotes the linear regression fit. Pearson correlation 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. R14(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 R14(b) presents the spatial distribution of the R^2 obtained from the precipitation-only regression, while Fig. R14(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

$$x = \epsilon \cdot f(T_Z) = \epsilon \cdot \left(1 + \alpha T_z + o(T_z^2)\right). \tag{5}$$

Retaining only the first-order approximation gives

$$x = a' \cdot T_z + b', \tag{6}$$

where T_z denotes standardized temperature.

Similarly, we obtain

$$y = c' \cdot T_z + d', \tag{7}$$

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

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

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

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

Here, the term $(1 - \mathrm{b})$ 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. R15). 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 revised manuscript. In the revised manuscript, Figures R13 and R14 in this response have been incorporated into the Supplementary Materials as Figs. S2 and S3, respectively, whereas Figure R15 has been included as Fig. S9. The following text has been added after Line 227 in the original manuscript, which corresponds to Line 253 in the tracked-changes version of the revised manuscript:

However, the parameters a and b in Eq. (1) and (7) are obtained by calibration, and they are less directly linked to specific water-balance components. In this study, our model is derived directly from the basin-scale water balance equation and provided a systematic clarification of both the derivation and the physical interpretation of the parameters.

We start from the water balance equation:

$$TWS(t) - TWS(t-1) = P(t) - ET(t) - R(t) = P(t) - ETR(t), \quad (9)$$

where $P(t)$, $ET(t)$, and $R(t)$ denote precipitation, evapotranspiration, and runoff at time step t, respectively, and $ETR(t) = ET(t) + R(t)$ represents the total

loss term. We assume that ETR *consists of two components: (1) the contemporaneous precipitation loss* ETR_1*, i.e., the direct loss occurring as precipitation is converted into water storage; and (2) the storage-release loss* ETR_2*, i.e., the loss arising from the depletion of antecedent water storage* $TWS(t-1)$ *in subsequent periods, which can be expressed as*

$$ETR_1(t) = x \cdot P(t), \tag{10}$$

$$ETR_2(t) = y \cdot TWS(t-1), \tag{11}$$

$$ETR(t) = ETR_1(t) + ETR_2(t) = x \cdot P(t) + y \cdot TWS(t-1). \tag{12}$$

The assumption is validated in Fig. S2, 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*.*

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. S3(a)), indicating that the total loss term* ETR *can be statistically explained by the linear combination of precipitation* $P(t)$ *and antecedent storage* $TWS(t-1)$*.*

We compared the full multiple regression model with a reduced model that includes only precipitation $P(t)$ *as the predictor to further examine whether antecedent storage* $TWS(t-1)$ *provides additional explanatory power beyond*

precipitation alone. Figure S3(b) presents the spatial distribution of the R^2 *obtained from the precipitation-only regression, while Fig. S3(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 proportional coefficients x *and* y *vary with temperature and apply a Taylor expansion, which yields*

$$x = \epsilon \cdot f(T_z) = \epsilon \cdot \left(1 + \alpha T_z + o(T_z^2)\right). \tag{13}$$

Retaining only the first-order approximation gives

$$x = a' \cdot T_z + b', \tag{14}$$

where T_z *denotes standardized temperature.*

Similarly, we obtain

$$y = c' \cdot T_z + d'. \tag{15}$$

Substituting Eqs. (12), (14), and (15) into the water balance equation (Eq. 9) leads to

$$TWS(t) = (1 - a' \cdot T_z - b') \cdot P(t) + (1 - c' \cdot T_z - d') \cdot TWS(t-1). \tag{16}$$

After merging constant terms simplifying the signs, the recursive formulation

adopted in this study is obtained as

$$TWS\,(t) = (a \cdot T_z + b) \cdot P(t) + (c \cdot T_z + d) \cdot TWS\,(t-1), \qquad (17)$$

which is the reconstruction model adopted in this study. In this form, parameters a *and* c *represent temperature modulation effects. 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 *after basin losses. Similarly,* $1-d$ *represents the fraction of antecedent storage contributing to the loss term* $ETR_2(t)$*, and thus parameter* d *represents the fraction of previous storage retained in the current storage state. We will continue to discuss these parameters in subsequent sections.*

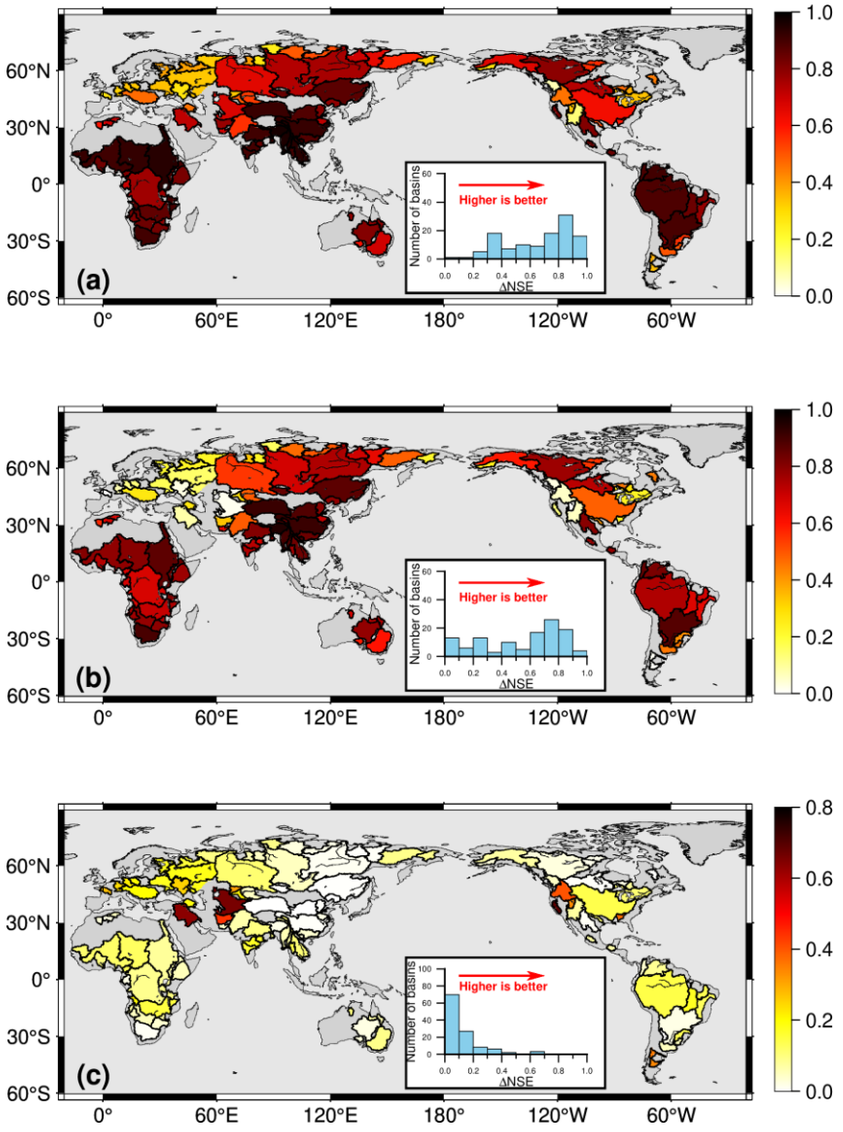

Figure R14. (a) Spatial distribution of the coefficient of determination (R^2) from the multiple linear regression model $ETR = x \cdot P(t) + y \cdot TWS(t-1)$; (b) Spatial distribution of R^2 from the univariate linear regression $ETR = x \cdot P(t)$; (c) Map of the difference between panels (a) and (b) .All basins pass the F-test at the 1% significance level ($p < 0.01$).

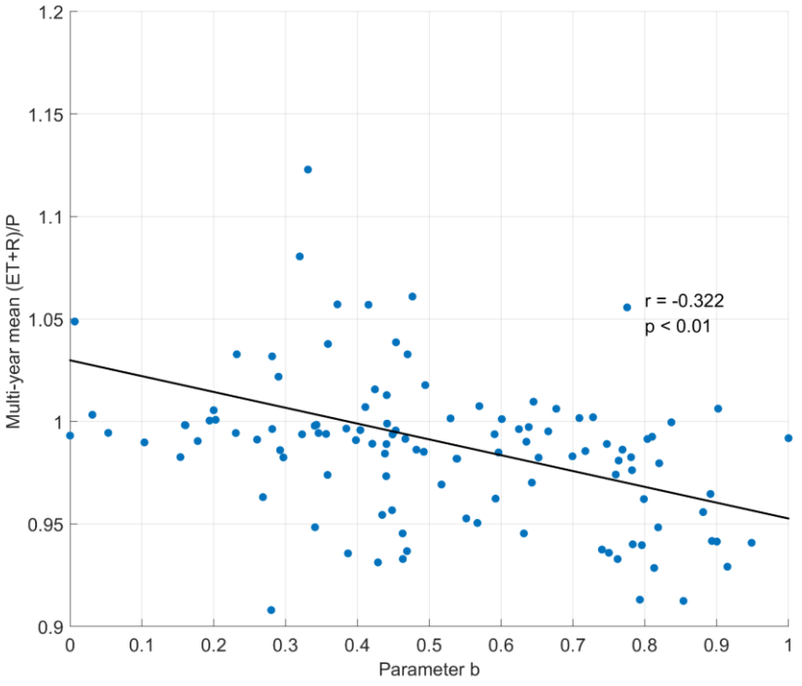

Figure R15. Scatter relationships across 116 global river basins between parameter b and the long-term mean loss ratio $(ET + R)/P$. The black dashed line denotes the linear regression fit. Pearson correlation coefficient r and the p-value for the regression slope are annotated on each panel.

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.

Minor Comment 1: Line 233: add the missing equation number.

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.

Minor Comment2: 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, which corresponds to Line 413 in the tracked-changes version of the revised manuscript.

Minor Comment 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 Fig. 6 and replaced them with alphabetical labels (a, b, c, etc.) to ensure consistency with panel (a) and improve overall figure readability. The ID numbers shown in Figs. 8, 11, and 14 have been removed to improve figure readability. The revised versions of these figures are provided in this response as Figs. R16–R19. In addition, Sect. 5.3 of the

original manuscript, which contained Fig. 16, has been deleted in the revised manuscript; accordingly, Fig. 16 has also been removed.

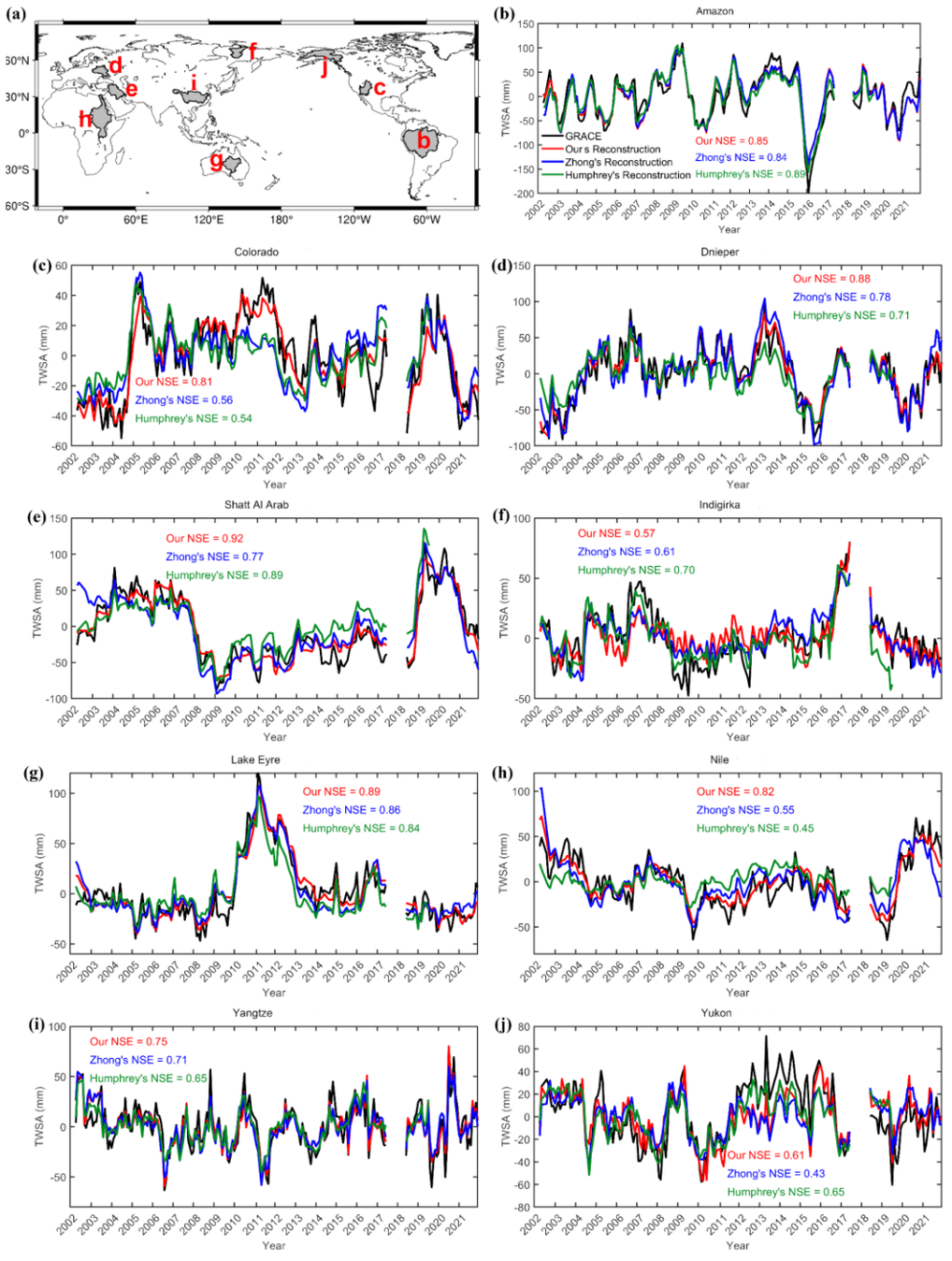

Figure R16. Time series of GRACE/GRACE-FO TWSA and reconstructed TWSA, both de-seasonalized and de-trended, for the nine selected river basins (b-j) from 2002 to 2021. The global distribution of the nine selected river basins (a).

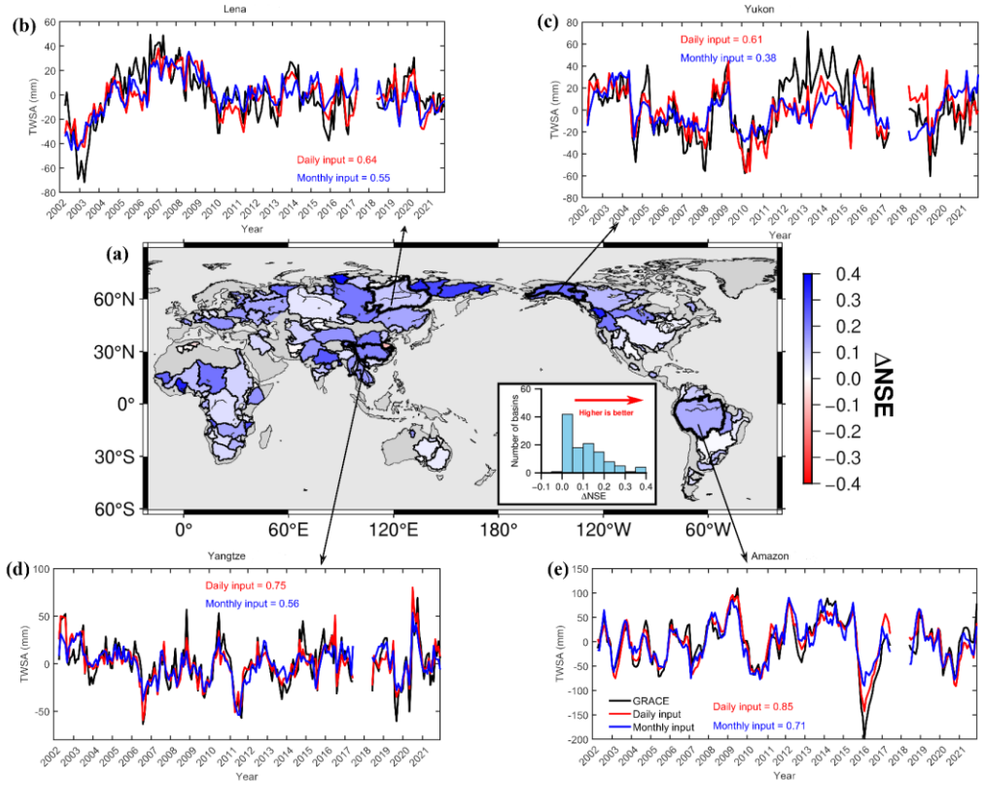

Figure R17: Spatial distribution of the difference in NSE between TWSA reconstructions forced by daily and monthly meteorological data, with respect to GRACE JPLM from 2002 to 2021 (a). Time series of GRACE/GRACE-FO TWSA and reconstructed TWSA (de-seasonalized and de-trended) using daily and monthly meteorological data in the (b) Lena, (c) Yukon, (d) Yangtze, and (e) Amazon basins.

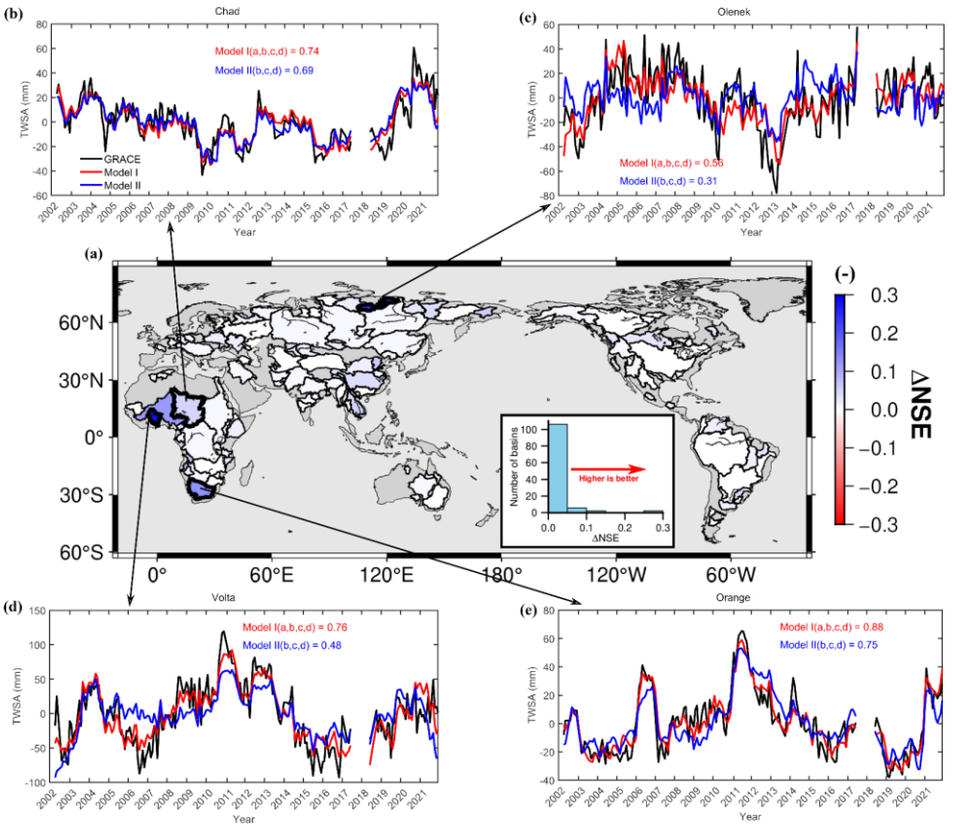

Figure R18: Spatial distribution of the NSE difference between TWSA reconstructions from a four-parameter and a three-parameter (excluding parameter a) daily recursive model with respect to JPLM across 116 global river basins during 2002–2021 (a). Time series of GRACE/GRACE-FO TWSA and reconstructed TWSA (de-seasonalized and de-trended) in the (b) Chad, (c) Olenek, (d) Volta, and (e) Orange basins.

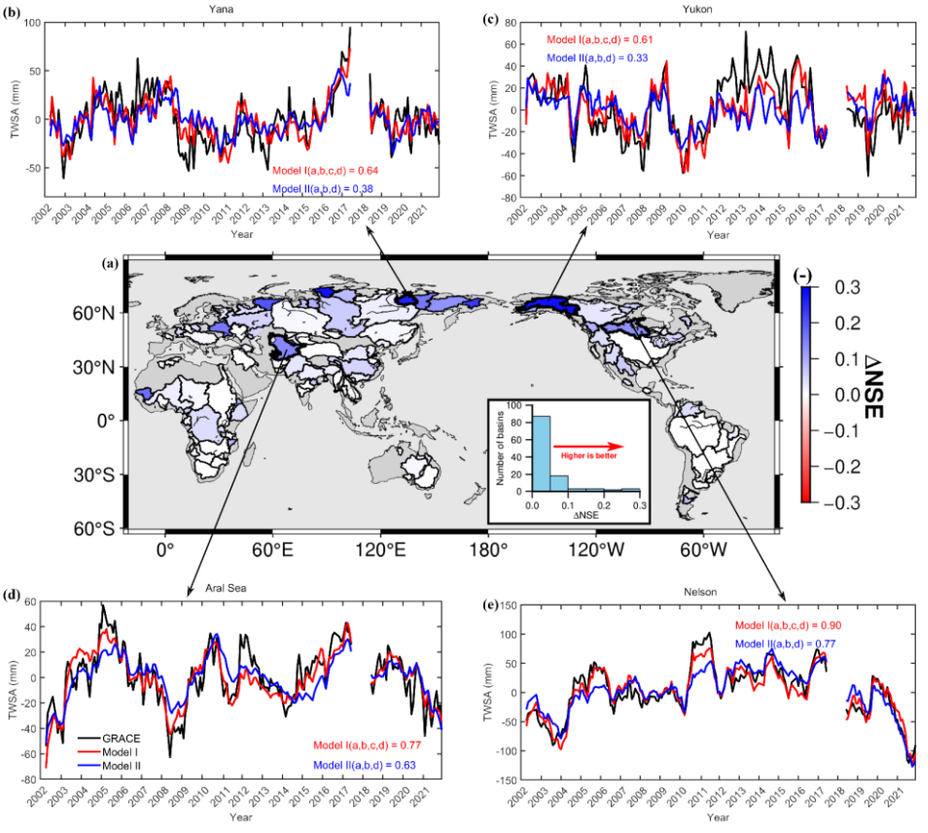

Figure R19: Spatial distribution of the NSE difference between TWSA reconstructions from a four-parameter and a three-parameter (excluding parameter c) daily recursive model with respect to JPLM across 116 global river basins during 2002-2021 (a). Time series of GRACE/GRACE-FO TWSA and reconstructed TWSA (de-seasonalized and de-trended) in the (b) Yana, (c) Yukon, (d) Aral Sea, and (e) Nelson basins.

References

References Befus, K. M., Jasechko, S., Luijendijk, 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.

Humphrey, V. and Gudmundsson, L.: GRACE-REC: A reconstruction of climate-driven water storage changes over the last century, Earth Syst. Sci. Data, 11, 1153-1170, 2019.

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.