SYSU TWSA v1.0: Global High-Resolution Terrestrial Water Storage Anomalies via Satellite Gravimetry

Reply to Referee 1

Yuhao Xiong, Wei Feng, Jun Huang, Hongbing Bai, Guangyu Jian, and Min Zhong

Note: the referees' comments are shown in blue, the authors' responses in black

1 Referee 1 comments (RC1)

In this study, Xiong et al. applied a joint inversion downscaling method to downscale GRACE/GFO products to 0.5 degrees. The method extracts spatial information from hydrological simulations (WGHM) and applies spatial corrections to glaciers and lakes to account for components that WGHM insufficiently models. Then, the temporal consistency with GRACE/GFO is ensured to keep the mass conservation. The authors provide comprehensive evaluations in different aspects and compare them with some other downscaled/data assimilation products, and demonstrate the quality of SYSU TWSAv1.0.

This paper is generally well written and provides valuable contributions to the community. The special care of inland glaciers and large lakes is especially appreciated. In the end, it shows remarkable improvements in the glaciated regions compared to Gou and Soja (2024), who clearly stated that such regions are their product's main limitation. No major issues exist for this approach and dataset, but some descriptions are a bit too brief and need to be elaborated more for the audience to understand the product/approach better. Therefore, I would recommend publication of this paper after some rather minor revisions. The specific comments are as follows:

Response: We thank the Referee 1 for the positive assessment of our manuscript and for recognizing the value of the dataset.

About Section 2 Data:

1. Line 87: There are multiple institutions that provide GRACE L2 data. Please specify which one you used.

Response: Thank you for this helpful comment. We have revised the manuscript to specify that the GRACE/GRACE-FO Level-2 spherical harmonic coefficients used in this study were obtained from CSR (Center for Space Research at the University of Texas) (Lines 96-97).

2. Lines 97-98: Please specify if you used WaterGap v2.2d or 2.2e. If they are forced by GSWP3-W5E5, maybe you used v2.2e, right? In that case, please provide the reference to 2.2e (Müller Schmied et al., 2024).

Response: Thank you for this helpful and careful comment. You are correct that the WGHM dataset forced by GSWP3-ERA5 corresponds to WaterGAP v2.2e. We have clarified this in the revised manuscript and added the corresponding reference to Müller Schmied et al. (2024) (Lines 106-107).

3. Lines 121: Are all the 5-yr records continuous? Or some gaps are allowed in these records. If some gaps exist, how did you correct some jumps that may exist?

Response: Thank you for this comment. The 5-yr requirement does not mean that each well record is fully continuous. Because groundwater well records with no missing months are very limited at the global scale, and imposing such a strict criterion would have substantially reduced the number of wells available for validation. Therefore, we allowed a limited number of missing months and did not interpolate missing values. In addition, to reduce the impact of obvious outliers or short-term disturbances (e.g., pumping-test-related anomalies), we further applied a threshold-based screening to exclude anomalous monthly values.

About Section 3 Joint Inversion Downscaling Framework

This section is important for the audience to understand the whole approach. Although I understand the method is based on their earlier publication (Xiong et al. 2025), I highly recommend that the authors expand this section to describe the methods in more detail. For example:

1 Line 145: Did you apply ICA to different WGHM components, such as groundwater and surface water, separately?

Response: Thank you for this helpful comment. Yes, ICA was applied separately to each WGHM water compartment to extract component-specific spatial patterns. Specifically, groundwater, soil moisture, snow water equivalent, reservoirs, wetlands, and rivers were decomposed individually, and the resulting dominant spatial patterns were used as spatial basis functions in the inversion (Lines 169-173).

2. Lines 147-148: What do "larger singular values" mean? How many leading values did you use?

Response: Thank you for pointing this out. We agree that the original expression “larger singular values” was ambiguous and could be misleading in the context of ICA. What we intended to convey is that we retained the dominant spatial modes whose temporal evolutions represent the trend, seasonal, and interannual variations in each compartment, rather than selecting modes according to a fixed number of singular values. In practice, these retained modes generally correspond to the major variance contributions of each compartment and therefore capture the main temporal behaviors of monthly water storage changes. Retaining only these dominant modes also reduces the number of unknown parameters and helps improve inversion stability.

3. How did you apply the least-squares adjustment to all the spatial patterns to ensure mass conservation? Please describe them in more detail.

Response: Thank you for this important comment. We agree that the original manuscript described the least-squares adjustment too briefly. In the revised manuscript, we have expanded Section 3 to clarify how the observation equation was constructed and how all spatial patterns were jointly fitted. Specifically, we first clarified that TWSA can be represented as the sum of multiple water storage compartments, and that, in our framework, each compartment is further expressed as a linear combination of spatial basis functions and their temporal evolution coefficients. To ensure consistency with the effective spatial resolution of GRACE/GFO, all high-resolution spatial basis functions were subjected to the same spherical harmonic truncation and DDK3 filtering as the GRACE/GFO fields. Based on these filtered basis functions, we established the observation equation in which GRACE/GFO-derived TWSA field is expressed as the sum of all filtered basis functions weighted by their temporal coefficients, plus a residual term. The temporal coefficients associated with all retained basis functions were then estimated by least squares. Finally, the downscaled TWSA fields were reconstructed by combining the adjusted temporal evolution with the unfiltered (high-resolution) spatial basis functions. We have revised the manuscript accordingly and added Eqs. (1)–(2) and the related explanations in Section 3.

About Section 4 Results

1 Section 4.3: Did you remove other components, like soil moisture, from your downscaled TWSA products before comparing with in-situ groundwater data? If yes, please mention which datasets you used to remove them.

Response: Thank you for this important comment. We did not remove other components, such as soil moisture, from the downscaled TWSA before comparing them with the in-situ groundwater observations. The main reason is that the contribution of non-groundwater components varies substantially across regions, and a globally consistent removal would require additional model-based estimates of soil moisture, snow, surface water, and other compartments, which would introduce extra uncertainties and reduce the independence of the validation. In particular, using model-simulated non-groundwater components for correction could propagate regionally varying model errors into the comparison. Therefore, in this study, we directly compared the downscaled TWSA with in-situ groundwater observations to assess whether the product captures groundwater-related variation, rather than attempting to isolate pure groundwater storage changes using additional model corrections.

2 Line 335: It is good to see that SYSU has more variability than Gou and Soja (2024), which shows the possibility of involving more small-scale signals (also in Fig. 9). However, how would you argue if they include

more signals or more noise? Moreover, GLWAS seems to have lower magnitudes in Fig. 8 than SYSU but rather similar degree variances in Fig. 9. What would be the reason behind this? As the data description paper is open for the data users from the hydrological community, more explanations about the spectral domain would be appreciated, so that the audience does not need much geodetic background knowledge to understand it.

Response: Thank you for this important comment.

We agree that this is a difficult question, and that it cannot be answered conclusively from the spectral results alone. Specifically, based only on the spectral amplitudes shown in Figs. 8 and 9, we cannot strictly determine whether the larger high-degree variability in SYSU reflects more recovered small-scale signals or more amplified noise. Our intention here was therefore not to interpret the stronger high-degree content as signal only, but rather to state more cautiously that SYSU retains more high-degree variability than Gou and Soja (2024), which may suggest the preservation of more small-scale signals, while also implying a higher noise level. To make the spectral-domain analysis more accessible to readers from the hydrological community, we added a brief clarification in the revised manuscript stating that, in the spectral domain, low degrees correspond to longer-wavelength, large-scale mass variability, whereas high degrees correspond to shorter-wavelength variability.

Regarding why GLWS seems to have lower magnitudes than SYSU in Fig. 8 but rather similar degree variances in Fig. 9, the main reason is that the two figures characterize different aspects of spectral behavior. Specifically, Fig. 8 shows the RMS of individual spherical harmonic coefficients over 2003–2019, whereas Fig. 9 shows the degree variance spectrum for a single month (October 2005). In this particular month, GLWS and SYSU exhibit similar degree-variance spectra, whereas in many other months SYSU shows larger spectral variability than GLWS (see Fig. R1).

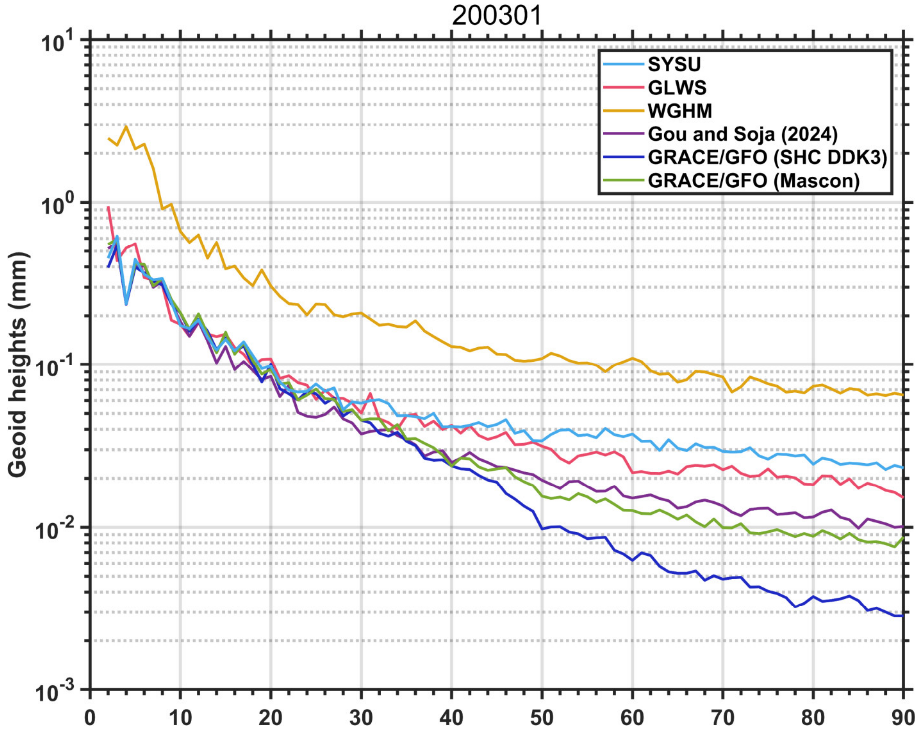


Figure R1. Degree variance for different products in Jan 2003 expressed by geoid heights (mm).

3 Line 390: It's fair that SYSU is closer to SHC solutions while Gou and Soja (2024) is closer to JPL mascons, as SYSU took SHC as inputs while Gou and Soja (2024) took JPL mascons as inputs. Please clarify this point, as the difference is unlikely to be caused by methodologies but rather by data differences.

Response: Thank you for this comment. We agree with you that this difference is more likely attributable to differences in the input GRACE/GFO products than to the downscaling methodologies themselves. Specifically, SYSU used the spherical harmonic coefficients (SHC) as input, whereas Gou and Soja (2024) used the JPL mascon product. We have clarified this point in the revised manuscript (Lines 438-440).

4 15: Did you convert all the datasets back to SHC d/o 90 and apply DDK3 filters to them as you did for Fig. 5? Both approaches are sound, but it's better to keep consistency. Moreover, Fig. 5 and Figs. 12-16 are based on different basin boundaries. Any specific reasons for doing so? Otherwise, I might recommend keeping consistent.

Response: Thank you for this important comment. The comparisons were carried out in a consistent manner.

The reason for using different basin boundaries is that these figures were designed to address different questions. Figure 5 was intended as a more detailed diagnostic analysis of SYSU relative to the raw GRACE/GFO estimates at the basin scale, in order to examine how their differences vary across basins of different sizes and under different hydroclimatic conditions. This analysis shows that SYSU is close to GRACE/GFO in large basins, indicating that the downscaled product preserves consistency with the GRACE/GFO estimates at scales where GRACE/GFO is considered reliable. This indicates that our downscaling method, while improving spatial resolution, still preserves the reliable large-scale observational information contained in GRACE/GFO. In contrast, as basin size decreases, the native spatial resolution of GRACE/GFO becomes increasingly insufficient. After resolution matching through spherical harmonic truncation and the same filtering, it becomes clear that GRACE/GFO is strongly affected by signal attenuation and leakage errors at small scales. In this sense, the high-resolution SYSU has good potential for monitoring water-storage changes in small basins.

In contrast, Figs. 12–15 were intended primarily for inter-product comparison. The purpose there was no longer to examine scale- or aridity-dependent characteristics, but rather to compare the relative performance of different downscaled products and their corresponding evaluation metrics.

5 Starting from line 443 (GLDAS analysis). It is rather similar to the previous analysis and did not provide much more new information. Maybe move to the supplementary information.

Response: Thank you for this helpful suggestion. We agree that the GLDAS analysis is similar in purpose to the preceding comparison and does not add substantial new information to the main text. Following your suggestion, we have moved the GLDAS analysis to the Supplementary Information. In the revised manuscript, we retain only a brief summary in the main text, noting that the comparison with GLDAS provides an additional reference and yields conclusions consistent with the previous analyses, namely that SYSU shows better agreement with GRACE/GFO than the GLDAS-based product.

General comments: Some of the figures have relatively low resolution. Please provide high-resolution figures and ideally vector figures for the final publication.

Response: Thank you for this helpful suggestion. We will provide high-resolution figures in updated revised manuscript.

2 Referee 2 comments (RC2)

I have reviewed the manuscript titled " SYSU TWSA v1.0: Global High-Resolution Terrestrial Water Storage Anomalies via Satellite Gravimetry " authored by Yuhao Xiong et al. The manuscript is generally well written, but several major and minor issues should be addressed to improve readability and reproducibility. Below are some suggestions to enhance the manuscript.

Response: We sincerely thank you for the careful evaluation of our manuscript and for the constructive comments and suggestions. These comments are very helpful for improving the readability, reproducibility, and overall quality of the manuscript. We have carefully revised the manuscript accordingly and provide detailed point-by-point responses below.

Major Comments:

1 The manuscript presents a joint-inversion framework that combines GRACE/GFO constraints with WGHM spatial patterns and predefined mascon groups. However, the level of methodological novelty relative to existing approaches (e.g., data assimilation and machine learning-based downscaling) is not sufficiently clarified. The current presentation makes it difficult to distinguish whether the proposed method represents a fundamentally new approach or an alternative formulation of existing constrained inversion strategies. To improve clarity, the authors are encouraged to:

Provide a clearer conceptual comparison with: 1) EnKF-based assimilation approaches. 2) Machine learning/statistical downscaling approaches.

Explicitly state: 1) What problem existing methods fail to solve. 2) How the proposed method uniquely addresses it.

Response: Thank you for this important comment. We agree that the conceptual novelty of the joint-inversion framework relative to existing downscaling approaches was not sufficiently clarified in the original manuscript. In the revised manuscript, we have expanded the Introduction to provide a clearer comparison with data assimilation and statistical or machine learning downscaling approaches. Specifically, we now clarify that data assimilation can update model states and related hydrological fluxes in a physically consistent way, but its performance remains closely tied to the model structure, state variables, error assumptions, and the allocation of GRACE/GFO-derived increments among individual storage compartments (Lines 51-60 and Lines 73-87). We also clarify that statistical and machine-learning downscaling methods rely on empirical relationships between GRACE/GFO-derived TWSA and high-resolution predictors, and that these relationships may not transfer reliably from the effective GRACE/GFO scale to finer local scales (Lines 61-69).

We further revised the Section 3 to explicitly state the formulation of our method (Lines 137-196). The proposed framework does not train a predictive model or directly update hydrological model states. Instead, it uses WGHM and external datasets to define spatial basis functions, while the temporal evolution of these basis functions is estimated directly from GRACE/GFO observations. This formulation allows missing or poorly represented signals, such as mountain glacier mass changes and selected large-lake storage variations, to be introduced into the inversion through dedicated spatial basis functions, without requiring explicit process modules or additional state variables within the hydrological model.

2 The proposed framework relies heavily on spatial basis functions derived from WGHM (Sect. 3.1), while only the temporal evolution is adjusted using GRACE/GFO observations. This implies that any structural bias in WGHM spatial patterns is directly propagated into the final product. Although briefly acknowledged in the conclusion, this limitation is not quantitatively assessed. This is a critical issue because it challenges the independence and physical reliability of the dataset. The authors should:

Quantify the sensitivity of the results to WGHM spatial patterns

Identify regions where WGHM is known to perform poorly and evaluate SYSU performance there.

Discuss how spatial biases may affect glacier- and groundwater-dominated regions.

Response: Thank you for this important comment. We agree that the dependence of SYSU on WGHM-derived spatial basis functions is a key limitation that should be more clearly explained and assessed.

First, we have clarified the construction principle of the spatial basis functions in the revised manuscript. The spatial patterns were not selected arbitrarily from WGHM. Instead, ICA was applied separately to each WGHM water storage compartment to extract the dominant spatial modes associated with the main temporal behaviors of monthly TWSA, namely long-term trend, seasonal, and interannual variations. This design is motivated by the fact that monthly GRACE/GFO TWSA can generally be represented by trend, seasonal, interannual, and residual components (Lines 170-173). Therefore, once the WGHM outputs and the mode-selection criteria are specified, the spatial basis functions used in the inversion are essentially fixed. GRACE/GFO observations are then used to estimate the temporal evolution coefficients of these predefined basis functions.

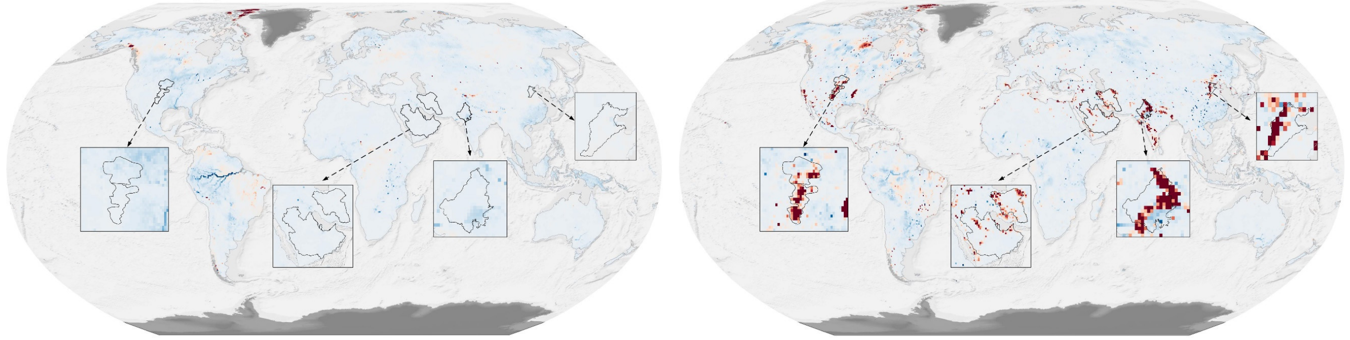
We also clarify that the ICA-derived spatial modes themselves are relatively stable. In our previous work, we tested the sensitivity of the extracted ICA spatial patterns to different time windows and found that the dominant trend, seasonal, and interannual modes remain generally stable at decadal and longer timescales (Xiong et al., 2025a). This suggests that the extraction of the major spatial basis functions is not highly sensitive to the specific temporal window used for ICA decomposition.

Nevertheless, we acknowledge that this does not make SYSU fully independent of WGHM spatial information. Because WGHM does not provide explicit uncertainty estimates for its spatial structures, and because globally distributed high-resolution “true” TWSA observations are unavailable, the structural bias of WGHM spatial patterns cannot be directly quantified. In our framework, WGHM-derived spatial patterns should therefore be understood as spatial priors rather than as ground truth. If a different WGHM version, model setting, or hydrological model were used, the derived spatial basis functions and the final downscaling result could differ.

To further address your concern, we added a sensitivity experiment using two WGHM configurations: one neglecting direct human impacts (WGHM-NHI) and one including direct human impacts (WGHM-HI) to reconstruct the spatial basis functions and compared the resulting solutions with the baseline SYSU product. This experiment provides a quantitative assessment of the dependence of SYSU on WGHM-derived spatial priors. As shown in Fig. S2, WGHM-NHI weakens or misses several human-induced groundwater depletion patterns, especially in groundwater-dominated and heavily managed regions, whereas WGHM-HI better represents these depletion signals. The corresponding joint-inversion downscaling result using WGHM-HI priors therefore better preserves negative trends in regions such as northwestern India and the North China Plain. These results show that SYSU is affected by the choice of WGHM outputs, but the sensitivity is spatially heterogeneous. At the same time, Fig. S3 shows that even when WGHM-NHI is used as the spatial prior, the joint-inversion downscaling result still substantially improves basin-scale consistency with GRACE/GFO relative to the original WGHM-NHI outputs, with lower RMSE and higher NSE and CC across most basins. This indicates that the framework effectively combines high-resolution model-derived spatial information with the reliable large-scale temporal variability constrained by GRACE/GFO, although the regional allocation of groundwater-related signals remains dependent on the quality of the WGHM-derived spatial priors. Therefore, when the model-derived spatial priors more adequately represent the hydrological processes, or when missing processes are supplemented by dedicated basis functions, the downscaled product is expected to provide a closer approximation to the true regional distribution of TWSA. We have added this analysis and expanded the discussion of model-prior dependence in the revised manuscript (Section 5.2).

(a) WGHM-NHI trend

(b) WGHM-HI trend



(c) Joint inversion using WGHM-NHI priors

(d) Joint inversion using WGHM-HI priors

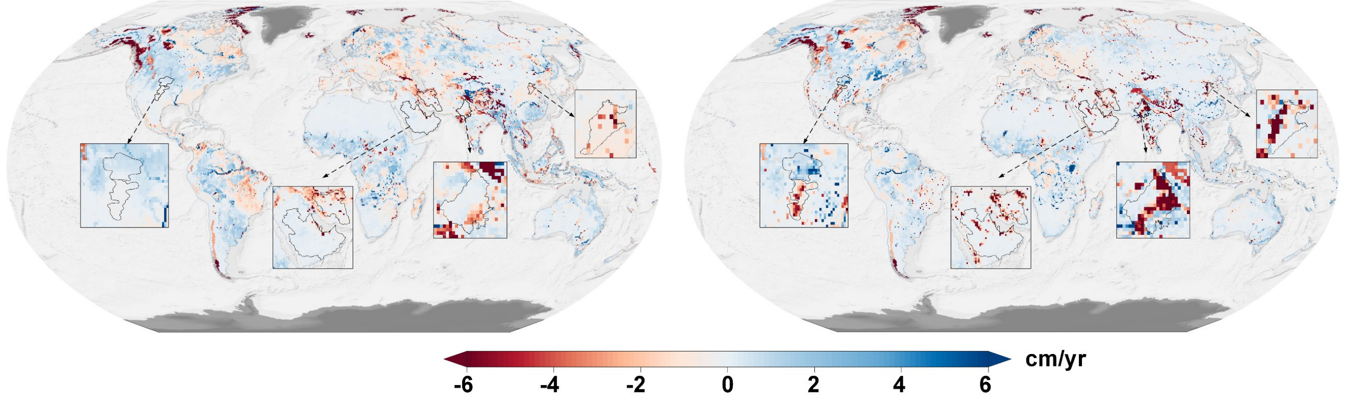
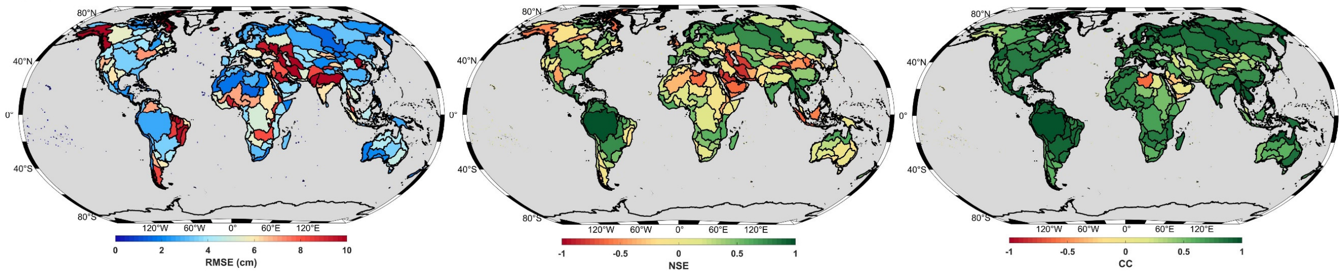


Figure S2: Sensitivity of downscaling TWSA trends to the representation of direct human impacts in WGHM. (a) TWSA trend from WGHM outputs neglecting direct human impacts (WGHM-NHI). (b) TWSA trend from WGHM outputs including direct human impacts (WGHM-HI). (c) Joint-inversion downscaling results using WGHM-NHI-derived spatial priors. (d) Joint-inversion downscaling results using WGHM-HI-derived spatial priors. Insets highlight representative regions with strong human-induced groundwater depletion. Units are cm/yr.

(a) Before joint inversion: WGHM-NHI vs GRACE/GFO



(b) After joint inversion: downscaled TWSA vs GRACE/GFO

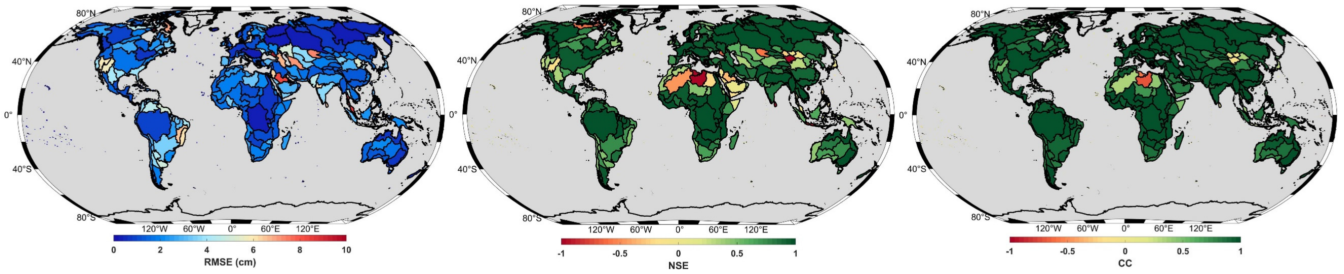


Figure S3: Basin-scale consistency with GRACE/GFO before and after joint inversion using WGHM-NHI spatial priors. The first row shows the comparison between WGHM outputs neglecting direct human impacts (WGHM-NHI) and GRACE/GFO in terms of RMSE, NSE, and CC. The second row shows the corresponding comparison after joint-inversion downscaling result using WGHM-NHI-derived spatial priors, including RMSE, NSE, and CC. RMSE is given in cm.

Finally, we have further discussed the implications for glacier- and groundwater-dominated regions. In glacierized and large-lake regions, WGHM does not adequately represent some key mass-change signals.

SYSU partly alleviates this limitation by incorporating dedicated glacier and lake mascon basis functions, which helps reduce signal distortion in these regions. This is further illustrated in the revised Fig. 11, where SYSU better captures glacier-related mass loss in Alaska and the Tibetan Plateau and better represents storage variations in selected large lakes, such as the North American Great Lakes and Victoria Lake. These results indicate that, when missing processes are supplemented by dedicated spatial basis functions, the joint-inversion framework can partly reduce the limitations of WGHM-derived spatial priors and provide a regional TWSA distribution that is more consistent with the underlying physical processes. In contrast, groundwater-dominated regions remain more dependent on WGHM-derived groundwater spatial priors, and the results in these areas should therefore be interpreted more cautiously. We have revised the manuscript to make these limitations and regional implications clearer.

Reference: Xiong, Y., Feng, W., Bai, H., Chen, W., Jiang, Z., and Zhong, M.: High-Resolution Terrestrial Water Storage Anomalies and Components in China From GRACE/GFO via Joint Inversion Downscaling, Water Resources Research, 61, e2024WR038996, https://doi.org/10.1029/2024WR038996, 2025a.

3 The manuscript states that WGHM outputs forced by the GSWP3-W5E5 dataset were used for the period 2002–2022. However, the standard GSWP3-W5E5 forcing dataset is typically available only up to 2019. It is therefore unclear how WGHM outputs were generated for the period 2020–2022. The authors should clarify whether an extended forcing dataset was used, and if so, provide details on its source and consistency.

Response: Thank you for this careful comment. Referee 1 also noted this issue, and we are grateful to you for identifying this inconsistency. We apologize for the incorrect description in the original manuscript. The WGHM outputs used in this study were not forced by GSWP3-W5E5, but by the GSWP3-ERA5 climate dataset (obtained from https://gude.uni-frankfurt.de/handle/gude/302.2). These WGHM v2.2e outputs cover 2002–2022, so the 2020–2022 fields were obtained from the same WGHM output series rather than from an extension of GSWP3-W5E5. We have corrected the description and made the terminology consistent throughout the revised manuscript (Lines 106-107).

4 A substantial portion of the validation relies on comparisons with GRACE/GFO data, which are already used as constraints in the inversion process. This raises concerns about circularity, particularly in basin-scale comparisons (Sect. 4.1), spectral-domain analysis (Sect. 4.4.1), and parts of the spatial-domain evaluation. While the manuscript acknowledges some of these limitations, the interpretation of results still tends to suggest validation rather than consistency. The authors should:

Clearly distinguish between consistency checks (with GRACE/GFO), and Independent validation

Rephrase conclusions where necessary to avoid overinterpretation.

Strengthen truly independent validation (e.g., groundwater, water balance) with more rigorous analysis

Response: Thank you for this important comment. We agree that comparisons with GRACE/GFO should not be interpreted as fully independent validation, because GRACE/GFO observations are used as constraints in the joint-inversion framework. In the revised manuscript, we have therefore explicitly distinguished between consistency checks and independent comparisons. Specifically, we revised the opening paragraph of Section 4 to clarify that comparisons with GRACE/GFO, including the basin-scale comparison, spectral-domain analysis, and part of the spatial-domain evaluation, should be treated as consistency checks rather than independent validation (Lines 198-203).

We also revised the Section 4.2, Section 4.3, and the Conclusion to avoid overinterpretation. Statements based on GRACE/GFO comparisons were rephrased using terms such as “basin-scale consistency with GRACE/GFO” rather than validation or accuracy claims. For the water-balance analysis, we now describe it as an independent consistency check rather than a strict validation, because uncertainties in P, E, and R and unaccounted human interventions can affect basin closure. For the groundwater-well comparison, we clarified that groundwater levels are not directly equivalent to TWSA and that the correlation-based analysis reflects consistency with groundwater-level variability, not direct validation of groundwater storage changes. These revisions provide a more cautious and balanced interpretation of the evaluation results.

5 The comparison with groundwater wells (Sect. 4.3) is based solely on correlation coefficients between TWSA and groundwater levels. While this approach is understandable given data limitations, it has important shortcomings as: 1) Groundwater levels are not directly equivalent to total water storage. 2) No conversion to storage is performed. 3) Correlation alone does not ensure physical consistency. As a result, the reported improvement (67.7% of wells) may not necessarily reflect improved physical realism. To improve this section, the authors should:

Clarify the limitations of using correlation-based validation more explicitly.

Perform regional or hydrogeological stratification (e.g., by aquifer type or climate zone)

Assess statistical significance of improvements

Moderate claims regarding physical consistency

Response: Thank you for this important comment. We agree that the groundwater-well comparison should be interpreted cautiously. Groundwater wells measure groundwater level variations, whereas SYSU and GRACE/GFO provide total water storage anomaly. A strict conversion from groundwater level to storage would require reliable aquifer parameters, such as specific yield or specific storage. However, such information is not consistently available in the global GGMN dataset, and many wells do not provide sufficient metadata to distinguish between shallow and confined aquifers.

Therefore, we used correlation coefficients only as a consistency metric, not as a direct validation of groundwater storage changes. We have revised Section 4.3 to clarify this limitation and moderated the related interpretation (Lines 330-334 and Lines 339-340). Specifically, the reported 67.7% improvement is now described as improved consistency with local groundwater-level variations rather than evidence of improved physical realism. We also clarified that the statistical significance of the correlations was assessed, and only significant correlations were retained for the comparison.

In addition, because global aquifer-type stratification is not feasible with the available metadata, we provided regional statistics for areas with sufficient well coverage, including the United States, India, Australia, and Europe. These revisions provide a more cautious and transparent interpretation of the groundwater-well comparison.

6 The manuscript does not provide a comprehensive uncertainty analysis of the SYSU TWSA product, which is a major limitation for a data paper. Currently, the evaluation focuses on performance metrics (RMSE, NSE, CC), but does not address: 1) Propagation of GRACE/GFO measurement errors; 2) Uncertainty introduced by: ICA decomposition, Selection of spatial basis functions, and Mascon group definitions; and 3) Spatial variability of uncertainty. To address this, the authors are encouraged to:

Provide uncertainty estimates where feasible (e.g., spatial error patterns or variance fields).

Discuss how uncertainties propagate through the inversion framework

Include sensitivity analyses for key methodological choices

Clarify the expected level of confidence for different regions or conditions

Response: Thank you for this important comment. We agree that uncertainty characterization is essential for a data paper. In the original manuscript, the uncertainty of SYSU TWSA was not sufficiently discussed. In the revised manuscript, we have added an uncertainty and sensitivity discussion and included a spatial uncertainty estimate where feasible.

A fully comprehensive uncertainty propagation is difficult because several sources of uncertainty do not have explicit covariance information. In particular, WGHM does not provide spatial structural uncertainty estimates for its water-storage compartments, and the uncertainties associated with ICA mode selection and glacier/lake mascon definitions cannot be directly propagated using a standard error-propagation framework. GRACE/GFO measurement errors are also spatially correlated and region dependent, and their impact differs with basin size. Therefore, we do not claim to provide a complete total uncertainty budget for SYSU TWSA.

Instead, we added a posterior formal uncertainty estimate derived from the joint-inversion adjustment. Specifically, under fixed spatial basis functions, the posterior covariance of the estimated temporal coefficients was propagated to the reconstructed TWSA fields. The resulting spatial uncertainty pattern is shown in Fig. 16. The map indicates that the formal uncertainty is spatially heterogeneous. Most land areas show relatively low uncertainty, generally below 1–2 cm equivalent water height, whereas higher uncertainties occur in regions with stronger hydrological variability or more complex storage processes, such as parts of North America, northern India, eastern China, central and eastern Africa, Southeast Asia, and Australia (Section 5.1).

We also expanded the discussion of uncertainties that cannot be fully captured by this posterior formal uncertainty. These include WGHM structural biases, ICA-derived basis-function selection, and glacier/lake mascon group definitions (see Comment #2). To address these issues, we added sensitivity analyses for key methodological choices. In particular, we compared reconstructions using WGHM outputs with and without direct human impacts and discussed the role of supplemental glacier and lake mascon basis functions. These analyses clarify that groundwater-dominated and heavily managed regions are more sensitive to WGHM-derived spatial priors, whereas glacierized and selected large-lake regions benefit from dedicated mascon basis functions.

We have revised the manuscript to clarify that the posterior formal uncertainty mainly reflects the uncertainty of the inversion under fixed spatial priors, while the confidence of SYSU in different regions should also be interpreted together with the sensitivity analyses and the known limitations of WGHM and GRACE/GFO.

7 The dataset is provided at 0.5° spatial resolution, but the effective resolution of GRACE/GFO observations remains on the order of ~300 km. While the downscaling enhances spatial detail, the manuscript does not clearly distinguish between nominal grid resolution, and effective information content. This may lead to misinterpretation of fine-scale signals by users. The authors should clearly state the effective resolution of the SYSU product; emphasize limitations of interpreting grid-scale variability; and discuss how much independent information exists at sub-basin scales.

Response: Thank you for this important comment. We agree that the nominal 0.5° grid-cell spacing of SYSU should be clearly distinguished from the effective information content of GRACE/GFO. In the revised manuscript, we have clarified this point in the Conclusion (Lines 564-569). Specifically, we now state that the enhanced spatial detail of SYSU mainly comes from WGHM-derived spatial patterns and the self-defined glacier and lake mascon groups, whereas the temporal evolution of these spatial patterns is constrained by GRACE/GFO observations.

We further clarify that SYSU TWSA should be interpreted as a GRACE/GFO-constrained downscaling product informed by high-resolution spatial priors, rather than as independent GRACE/GFO observations at the 0.5° grid-cell scale. We also emphasize that structural biases in WGHM spatial patterns or in the self-defined mascon groups may propagate into the product. These revisions are intended to help users avoid overinterpreting grid-scale variability and to better distinguish between the nominal spatial resolution of SYSU and the effective gravimetric information content provided by GRACE/GFO.

8 There is insufficient analysis of high-frequency variability (signal vs noise). In the spectral analysis, the SYSU product is shown to retain more high-degree variability compared to some other products (Sect. 4.4.1). However, it is not clear whether this additional variability represents meaningful hydrological signal or amplified noise. This distinction is critical for evaluating the usefulness of the dataset. The authors should:

Discuss whether the high-frequency variability corresponds to physically meaningful features or potential noise

Discuss potential noise amplification due to the inversion process

Provide guidance on how such high-resolution signals should be interpreted

Response: Thank you for this important comment. A similar concern was also raised by referee 1, and we agree that this issue cannot be resolved conclusively from the spectral amplitudes alone. Based only on Figs. 8 and 9, we cannot strictly determine whether the larger high-degree variability retained by SYSU represents recovered small-scale hydrological signals or amplified noise. We have therefore revised the manuscript to use a more cautious interpretation: SYSU retains more high-degree variability than Gou and Soja (2024), which may partly indicate preservation of small-scale variability but may also imply a higher risk of noise amplification.

We also added guidance for users on interpreting high-resolution signals. The fine-scale variation in SYSU is informed by WGHM-derived spatial patterns and supplemental mascon basis functions, while the temporal evolution is constrained by GRACE/GFO at its effective spatial resolution. Therefore, 0.5° grid-scale signals should not be interpreted as fully independent GRACE/GFO observations, and small-scale features should be assessed together with regional process knowledge and sensitivity analyses.

9 Although the manuscript compares SYSU with several products (e.g., GLWS, Gou & Soja 2024, GLDAS CLSM DA), the analysis remains largely metric-based and does not provide sufficient diagnostic insight into: 1) Why certain methods perform better in specific regions. 2) Under which conditions each method fails. To improve this section, the authors should provide regional analysis (e.g., humid vs arid, glacierized vs non-glacierized regions). Analyze differences in trend, seasonal amplitude, phase, and discuss strengths and weaknesses of each approach in specific hydrological regimes

Response: Thank you for this helpful comment. We agree that the original comparison was too metric-oriented and did not sufficiently explain why different products perform differently across regions. In the revised manuscript, we expanded Sect. 4.4.2 to provide a more diagnostic, region-based discussion of product behavior.

Specifically, we added examples from groundwater-dominated, glacierized, lake-dominated, and arid regions. For groundwater-dominated regions, we discuss the northern High Plains of the United States, where some products still inherit the negative WGHM trend, whereas SYSU and GLWS better adjust this WGHM-related bias through GRACE/GFO constraints. For glacierized regions, such as Alaska and the Tibetan Plateau, we clarify that WGHM does not explicitly simulate mountain glacier mass changes, while SYSU improves the representation of these signals by incorporating dedicated glacier mascon basis functions. We also discuss selected large lakes, such as the North American Great Lakes and Lake Victoria, where SYSU benefits from lake mascon basis functions.

We further revised the discussion of annual amplitude and phase to explain that product differences may arise from trend bias, seasonal-amplitude bias, or phase mismatch, and that arid regions with weak hydrological signals require cautious interpretation. In addition, the sensitivity analysis using WGHM outputs with and without direct human impacts further shows that groundwater-dominated and heavily managed regions remain sensitive to the quality of WGHM-derived spatial priors. These revisions provide a clearer explanation of the strengths and limitations of different downscaling approaches under different hydrological regimes.

10 There is inconsistency in GRACE data resolution (degree/order). The metadata of the NetCDF file indicates that the original GRACE data are "CSR SH d/o 60", whereas the manuscript states that Level-2 data up to degree and order 90 were used. This discrepancy should be clarified, as it directly affects the effective resolution and information content of the dataset.

Response: Thank you for this careful check. We apologize for the inconsistency between the manuscript and the NetCDF metadata. The GRACE/GFO Level-2 spherical harmonic data used in this study were processed up to degree and order 90, as stated in the manuscript. The "CSR SH d/o 60" label in the NetCDF metadata was incorrect. We have corrected the metadata to ensure consistency with the manuscript. We have also updated the NetCDF file by adding the posterior formal uncertainty field derived from the joint-inversion adjustment.

Minor Comments:

1 Overall, several sentences are overly long and would benefit from being split for clarity.

Response: Thank you for this helpful comment. We have carefully revised the manuscript to improve readability. Several long sentences have been split or simplified, and the wording has been revised where necessary to improve clarity and flow.

2 The manuscript states that GRACE/GFO Level-2 data up to degree and order 90 were used; however, the data source (e.g., CSR, GFZ, or JPL) is not clearly specified in Section 2.1. Although CSR RL06.2 is mentioned later in the data availability section, this information should be explicitly stated in the methodology. In addition, given that higher-degree coefficients (beyond ~60) are typically noise-dominated, the rationale for using degree/order 90 should be clarified.

Response: Thank you for this helpful comment. We have revised Sect. 2.1 to explicitly state that the GRACE/GFO Level-2 spherical harmonic coefficients were obtained from CSR and processed up to degree and order 90 (Lines 96-97). We agree that higher-degree coefficients, especially beyond about degree 60, are generally more noise-dominated. But we used degree/order 90 to retain potentially useful mid- to high-degree GRACE/GFO information.

3 There is incorrect description of spatial resolution. The dataset is described in the manuscript as having a spatial resolution of 0.5°, which is consistent with the grid dimensions (360 × 720). However, the NetCDF metadata specifies “Spatial_resolution = 50 km”, which is not strictly correct for a global latitude–longitude grid and may be misleading. This should be revised to “0.5 degree” or clarified appropriately.

Response: Thank you for this helpful comment. We agree that “50 km” is not an appropriate description for a global latitude–longitude grid and may be misleading. We have revised the NetCDF metadata and changed the spatial resolution description to 0.5° to make it consistent with the manuscript and the grid dimensions.

4 The NetCDF metadata does not clearly specify the physical units of the TWSA variable. This information is essential for data usability and should be explicitly included.

Response: Thank you for this helpful comment. We have revised the NetCDF metadata to explicitly specify the physical unit of the TWSA variable. The unit is now given as cm (equivalent water height), consistent with the manuscript and figures.

5 The manuscript refers to the application of a GIA correction using the ICE-6G model, whereas the NetCDF metadata specifies “ICE 6G-D”. It is unclear whether these refer to the same model variant, and the naming should be made consistent and explicitly clarified.

Response: Thank you for this careful comment. We have revised both the manuscript and the NetCDF metadata to use a consistent “ICE 6G-D” throughout.

6 The manuscript refers to the use of GRACE/GFO Level-2 data, which are provided as spherical harmonic coefficients rather than gridded fields. However, the dataset is described at a 0.5° resolution. The authors should clearly distinguish between the native resolution of Level-2 data and the derived gridded product after processing and downscaling.

Response: Thank you for this helpful comment. We have clarified this distinction in Sect. 2.1. Specifically, we now state that the GRACE/GFO Level-2 data were provided as spherical harmonic coefficients, which were subsequently post-processed and synthesized onto a 0.5° grid. We also explicitly note that, although the resulting surface mass changes are represented on a 0.5° grid, their effective spatial resolution remains approximately 330 km.

In addition, we clarified that the final SYSU downscaled product is provided at 0.5° resolution because the reconstruction inherits fine-scale spatial information from WGHM-derived spatial basis functions and the self-defined mascon groups, whereas the temporal evolution of these spatial patterns is constrained by GRACE/GFO observations. This clarification distinguishes the native Level-2 spherical harmonic data, the synthesized GRACE/GFO gridded fields, and the final downscaled SYSU product.

7 The dataset metadata clearly lists missing GRACE/GFO months, which is appreciated. However, the manuscript does not explicitly describe how these gaps are handled in the dataset (e.g., removed, interpolated, or retained as discontinuities). Clarifying this aspect is important, particularly for time-series analyses and derivative-based quantities such as TWF.

Response: Thank you for this helpful comment. We have clarified how missing GRACE/GFO months were handled in the revised manuscript. The missing months were not interpolated or gap-filled. They were retained as missing records in the dataset and are listed in the NetCDF metadata. For analyses requiring continuous time derivatives, such as the TWF calculation, we used only periods without missing GRACE/GFO months (Lines 97-98).

8 The methodology combines spherical harmonic representations with mascon-based spatial groupings. However, the consistency between these two representations is not fully explained. The authors should clarify how mascon groups are integrated within the spherical harmonic inversion framework.

Response: Thank you for this helpful comment. We have clarified in the revised methodology how the mascon groups are incorporated into the spherical harmonic inversion framework. The self-defined mascon groups are first represented on the 0.5° grid as spatial basis functions. They are then transformed into spherical harmonic coefficients, truncated to the same maximum degree and order as the GRACE/GFO fields, and filtered using the same DDK3 filter. This ensures consistency with the effective GRACE/GFO resolution (Lines 149-151). This scale matching is essential for the subsequent least-squares fit, because the temporal coefficients can be robustly estimated only when the observations and basis functions share a consistent effective spatial resolution (Lines 153-155).

9 (Page 1, lines ~17–25) The use of strong claims such as “strong agreement” and “competitive overall accuracy” appears somewhat overstated given the validation limitations. Consider moderating the language (e.g., “generally good agreement”, “demonstrates competitive performance”).

Response: Thank you for this helpful suggestion. We have revised the wording in the abstract by replacing terms such as “strong agreement” and “competitive overall accuracy” with more cautious expressions, including “generally good basin-scale consistency” and “demonstrates competitive performance” (Lines 20 and 25).

10 (Page 2, lines ~52) The statement that “models still exhibit large uncertainties” would benefit from more precise qualification or additional references.

Response: Thank you for this helpful comment. We have revised the statement and added a supporting reference to qualify this point more clearly. Specifically, we now cite Scanlon et al. (2018) when discussing the large uncertainties of global models in representing interannual to long-term TWS variations (Line 54).

11 (Page 6, lines ~144–148) The term “dominant patterns” (ICA decomposition) is not sufficiently defined. Please specify:

- how many components were retained;

- the selection criterion used.

Response: Thank you for this helpful comment. We have revised the methodology to define the “dominant patterns” more clearly (Lines 171-175). Specifically, we now explain that ICA was applied separately to each

WGHM water storage compartment, and that the retained components were those whose temporal coefficients represent the main temporal behaviors of monthly TWSA, namely trend, seasonal, and interannual variations. We also clarified that retaining these dominant modes reduces the number of unknown parameters, avoids an over-parameterized inversion, and improves the stability of the solution.

12 (Page 7, lines ~163–170) The selection of specific lake groups (e.g., Great Lakes, Lake Victoria, Tibetan Plateau lakes) should be justified more explicitly.

Response: Thank you for this helpful comment. We have revised the methodology to justify the selection of the lake mascon groups more explicitly. These lakes were selected because their storage variations are large enough to be detected by GRACE/GFO, but they are not explicitly simulated in WGHM (Lines 195-196).

13 (Page 8, lines ~180–188) While limitations are acknowledged, it would help to explicitly clarify that basin-scale comparison represents a consistency check rather than independent validation.

Response: Thank you for this helpful comment. We have revised the text to explicitly state that the basin-scale comparison with GRACE/GFO should be interpreted as a consistency check rather than an independent validation (Lines 198-203).

14 Figures 4, 11, 12 are overly dense and difficult to interpret. Consider simplifying or splitting panels, or increase the font sizes.

Response: Thank you for this helpful suggestion. We have revised these figures to improve readability. Specifically, we increased the font size in Figs. 4 and 12. For Fig. 11, we added representative glacierized regions, including Alaska and the Tibetan Plateau, to make the regional comparison clearer and to better highlight the performance of SYSU in glacier-dominated areas.

15 In multi-panel figures (e.g., Figure 11), the large number of panels combined with small font sizes reduces readability. It is difficult to clearly distinguish spatial patterns and labels at standard viewing scale. 1) Suggested improvements: Increase font size for legends and annotations. 2) Consider splitting large composite figures into multiple figures. 3) Provide zoomed-in regional examples for key areas (e.g., glacierized or arid regions)

Response: Thank you for this helpful suggestion. After careful consideration, we revised Fig. 11 by adding zoomed-in regional examples for key glacierized areas, including Alaska and the Tibetan Plateau. These insets help improve readability and make the regional differences among products clearer. We also increased the font sizes of labels, legends, and annotations where appropriate.

16 In several spatial maps (e.g., Figures 6, and 11), the color contrast is insufficient to clearly distinguish variations in low-signal regions, particularly in arid areas (e.g., northern Africa and the Middle East). The current color scale compresses small-magnitude variations, making it difficult to visually interpret differences between products. The authors are encouraged to use a diverging color scale centered at zero (for anomaly fields); Adjust the color range to better resolve low-amplitude signals; and Consider using non-linear scaling or percentile-based limits to enhance contrast in low-variance regions. Ensure that color choices remain perceptually uniform and colorblind-friendly.

Response: Thank you for this helpful suggestion. We have revised the color settings of the relevant spatial maps to improve readability, especially in low-signal regions. For Fig. 6, we adopted a diverging color scale centered at zero, with white representing near-zero values. The color scale was adjusted based on the publicly available PiYG10 (https://github.com/djoshea/matlab-utils/tree/master/libs/othercolor), following a similar visual design to Gou and Soja (2024), and the revised scheme has been checked for colorblind accessibility. For Fig. 11, we added zoomed-in regional examples for glacierized regions to improve the readability of spatial differences among products. In addition, after checking the figures for colorblind accessibility, we also revised the color schemes of Figs. 9 and 10 to improve visual clarity.

17 The manuscript includes a large number of figures (16 Figures) but only one table. This creates an imbalance

between visual and quantitative presentation. The authors are encouraged to include additional tables summarizing key metrics (e.g., RMSE, NSE, CC, R² across basin sizes and regions) to facilitate clearer comparison and reproducibility.

Response: Thank you for this helpful suggestion. We agree that quantitative summaries are important for comparison and reproducibility. In the revised manuscript, we have carefully checked that the key metrics, including RMSE, NSE, CC, and R^2, are explicitly reported in the text and figures for the main comparisons. Because many of the evaluations are spatially distributed and region dependent, we believe that the current figures provide a clearer representation of the spatial variability than an additional table would. Adding another table would largely duplicate the information already presented in the figures and text without providing substantial new insight. Therefore, we retained the figure-based presentation but revised the text to make the key quantitative results and regional differences clearer.