Response to Reviewer #2.

We would like to thank Reviewer #2 for the insightful comments provided. We believe that the various suggestions provided have helped to improve the quality of the study and the manuscript. In what follows, our replies to the comments are presented point by point.

General comment:

This manuscript presents a methodologically innovative contribution to large-scale hydrological data assimilation by integrating GRACE/GRACE-FO Terrestrial Water Storage (TWS) anomalies into the W3RA land surface model using an Ensemble Kalman Filter (EnKF) framework with a novel **model space mixed localization** scheme. The study is applied to the Brahmaputra basin, a hydrologically complex and data-scarce region, and demonstrates improvements in both TWS and streamflow simulations at daily temporal resolution. The multi-variable validation approach, encompassing TWS, streamflow, and surface soil moisture (SSM), strengthens the credibility of the results.

The paper addresses a relevant scientific problem and makes genuine contributions to the field. However, several major concerns must be addressed before the manuscript can be considered for publication in the Hydrology and Earth System Sciences journal. I believe it needs “major revisions”. Some suggestions for revisions are as follows.

Reply: Thanks for the acknowledging the significance of the study. In what follows, we address the comments.

Major comments:

1. The authors employ an ensemble of N=30 members in the EnKF framework. Given that the W3RA model state vector encompasses a large number of variables across the Brahmaputra basin, this ensemble size appears insufficient to adequately sample the model's uncertainty space and may lead to significant sampling errors and filter degeneracy. The authors provide no explicit justification for this choice, nor do they present a sensitivity analysis demonstrating that N=30 yields stable and converged results. It is strongly recommended that the authors either:

- Provide a formal justification based on the effective degrees of freedom of the system, or;
- Include a sensitivity analysis showing key metrics (e.g., NSE, RMSE, ensemble spread) as a function of ensemble size (e.g., N = 20, 30, 50, 100).

Reply: We acknowledge that the results of the DA experiment can be sensitive to the ensemble size. In this study, an ensemble of 30 members was chosen following previous TWS and SSM studies (De Lannoy et al., 2022; Springer et al., 2026). However, following the reviewer’s suggestion, an experiment with N=100 will be performed to test the sensitivity of the results to the ensemble size. Due to the limited time available to respond to the reviewer comments, this experiment will be performed later at the manuscript revision stage. The results of this experiment will be described in Appendix D of the manuscript.

2. The Brahmaputra basin contains significant glaciated areas, and GRACE/GRACE-FO TWS signals include a long-term trend associated with glacial mass loss. The authors appear to assimilate the full TWS signal and subsequently subtract the glacial component *a posteriori*. This approach raises important questions:

- Does the assimilation of the glacial trend introduce systematic biases into the model state variables (soil moisture, groundwater) that are not physically meaningful?

Reply: Thanks for the question. The answer is yes. As mentioned in the manuscript, the long-term trends caused by glacier retreat are found to be introduced into the groundwater component, and are accounted for a posteriori (lines 538-542). A sentence will be added earlier in line 177 (methods) to clarify this aspect.

- How does this *a posteriori* correction compare to an approach where the glacial signal is removed *prior* to assimilation? The authors should provide a more explicit discussion of the implications of this methodological choice and, if possible, a quantitative comparison between both approaches.

Reply: An experiment involving an a-priori glacier trend removal has been carried out and evaluated following the same procedures described in the main manuscript. The results are described in what follows, and will be added to Appendix D of the manuscript. Additionally, a sentence will be added in the main manuscript summarizing the results and referring to the Appendix.

In the main manuscript, the assimilated TWS observations include glacier-induced trends, and these trends are removed a posteriori to perform the validation. A second multivariate DA experiment has been performed where the glacier trends were removed prior to the DA experiment. In general lines, the conclusions of the study do not vary when the trend is removed prior to DA. Details are provided in what follows.

Evaluation against the assimilated data: After multivariate DA, the RMSD difference between the modelled and assimilated observations is reduced below 10 mm (Table 1), reaching values very similar to the experiments reported in the main manuscript with a maximum difference of 2.2 mm in RMSD between the two experiments. The time series in Fig. R1, which is analogous to Fig. 3 of the main manuscript, shows that the DA modifies the model SSM and TWS estimates in terms of timing and magnitude, bringing them closer to the observations.

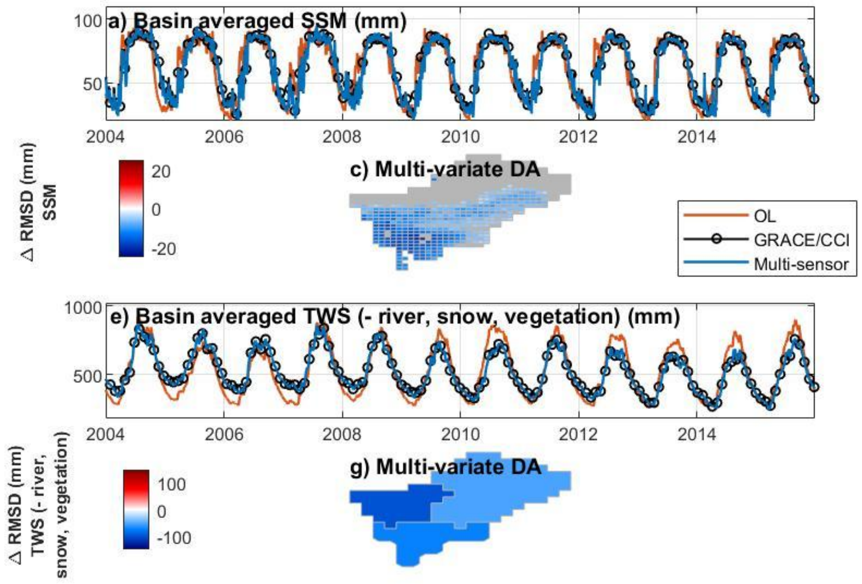

Figure R1. Basin averaged time series of SSM and TWS (a,e) and improvement in the RMSD of tile-averaged SSM (c) and sub-basin averaged TWS (g) of the model, after multivariate assimilation with when excluding glacier trends from TWS time series. ΔRMSD is defined as the RSMD of the DA product minus the RMSD of the OL, where the RMSD is computed by taking the assimilated data as a reference. For (c), ESA CCI tiles where no observation was assimilated have been shaded (gray area). **This figure is analogous to Fig. 3 of the main manuscript, and alphabetic tags have been preserved for comparison.**

	RMSD (mm)	
	OL	Multivariate DA (glacier removed a-priori)
TWS East	58.7	4.9
TWS West	108.0	7.3
TWS South	80.6	6.9
SSM (average)	13.6	3.9
SSM (min)	4.9	1.3
SSM (max)	24.9	9.1

Table R1. RMSD of estimates of OL and multi-variate DA when excluding glacier trends from TWS observations, with respect to assimilated data. SSM min and SSM max refer to the minimum and maximum RMSD values across the 0.25 × 0.25 degree tiles. **This table is analogous to Table 3 of the main manuscript.**

Evaluation of vertical and horizontal disaggregation: The vertical and horizontal disaggregation obtained when removing the glacier trends a priori are similar to those obtained when glacier trends were included (Fig. 4 of main manuscript). While the intensity of the updates is notably lower probably due to the smaller systematic differences between OL and the assimilated TWS data, the updates follow similar spatial patterns for the three variables (Fig. R2b, f and j). A different feature observed in this experiment is that slightly larger SSM updates are introduced around mountainous areas. This could likely be related to differences in ensemble statistics caused by the reduction of systematic model-observation biases, which is caused by the removal of glacier trends before DA.

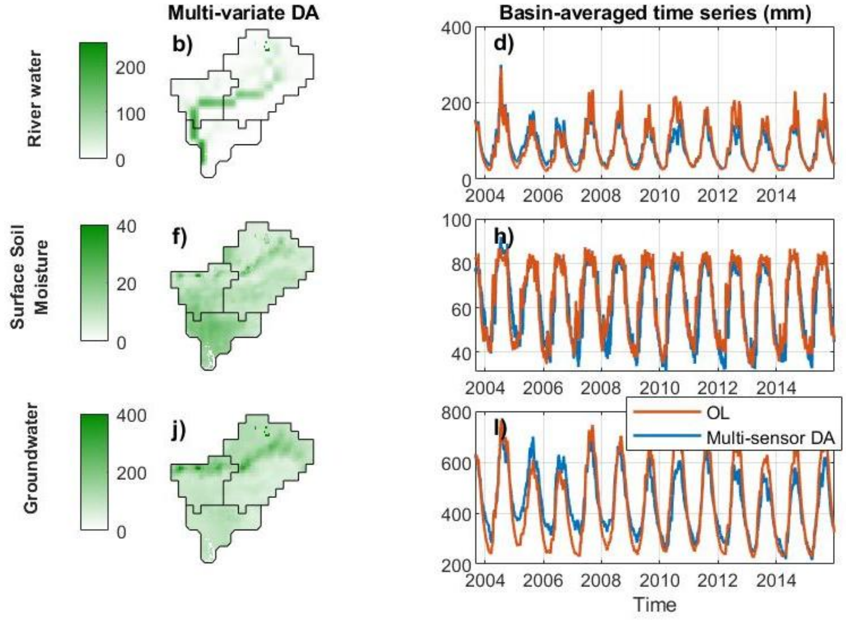

Figure R2. Standard deviation of DA minus OL on spatially distributed (0.1◦ × 0.1◦ grid) model estimates for river water (b), SSM (f) and groundwater (j), obtained for the multivariate DA experiment assimilating TWS without glacier trends. The standard deviation is expressed in mm. The column in the right (d, h, l) shows basin-averaged time series for thesame variables. **This figure is analogous to Fig. 4 of the main manuscript, and the alphabetic tags have been preserved for comparison.**

Comparison with SPEI: the comparison of groundwater estimates with SPEI leads to similar results as those presented in the main manuscript, as shown by the statistics and time series represented in Fig. R3. The time series are very similar, and the statistics are identical for the West and South sub-basin and leads to a slight change of 0.04 correlation points in the East sub-basin. Given this similarity, and the similar spatial distribution observed for the South sub-basin in Fig. R2, the results are also expected to be similar when evaluating against in-situ groundwater storage observations.

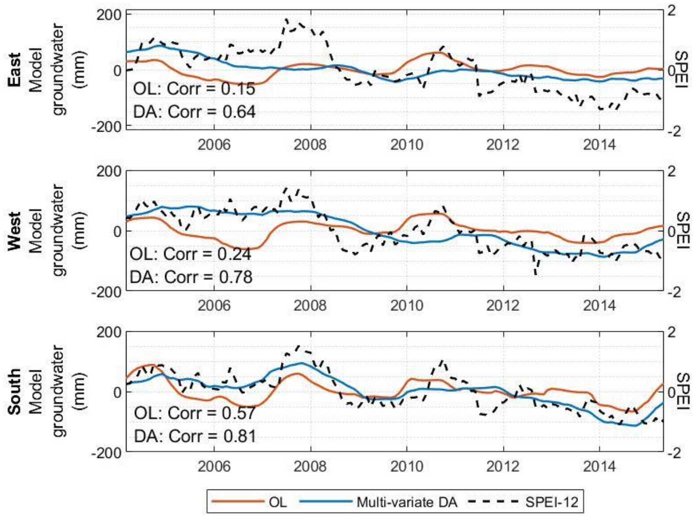

Figure R3. Inter-annual groundwater variability of OL and multi-variate DA; SPEI time (rescaled, on the right axis). **This figure is analogous to Fig. 6 of the main manuscript.**

Evaluation of SSM variability: the impact of the experiment in the SSM estimates is shown in Fig. R4, in terms of improvement in correlation coefficient with respect to WaterGAP SSM estimates. The improvements present a similar spatial distribution and intensity with respect to those represented in Fig. 8c of the main manuscript. Small differences can be observed along the northern edge of the area were SSM observations were assimilated (black outline in Fig. R4): the present experiment seems to slightly degrade SSM in this region. However, the degradation is not strong enough to significantly deteriorate sub-basin averaged SSM estimates (see, e.g., Fig. R2h), unlike the observation space localization results analysed in Fig. 5c, g, k, m and o of the main manuscript, which lead to significant SSM degradations. Therefore, the degradations observed in Fig. R4 can be considered minor and do not seem to have a major impact on sub-basin averaged water storage dynamics.

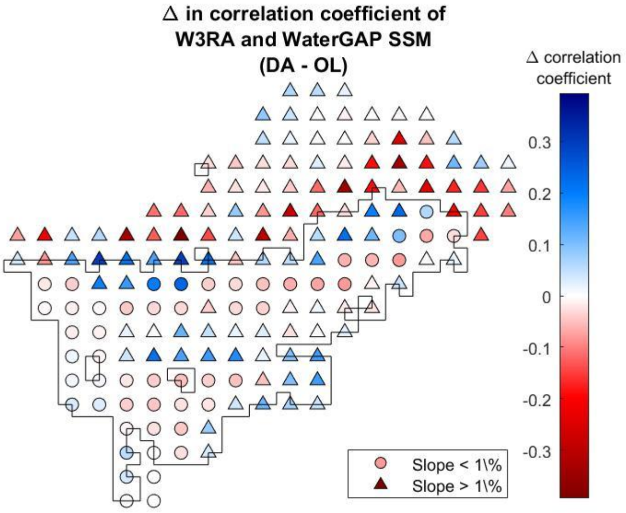

Figure R4. Distributed map of change in correlation coefficient between W3RA and WaterGAP after multivariate DA when assimilating TWS observations without glacier trends. The black outline indicates the area where SSM observations were assimilated. **This figure is analogous to Fig. 8c of the main manuscript.**

Evaluation of river water storage variability: The evaluation of river storage variability in the West and South sub-basin outlet indicates that a negative trend is introduced in the river storage time series (Fig. R5), although the trend is less pronounced than that observed in the main manuscript (Fig. 9 of main manuscript) likely because the introduction of glacier trends in the groundwater component was avoided in this setting.

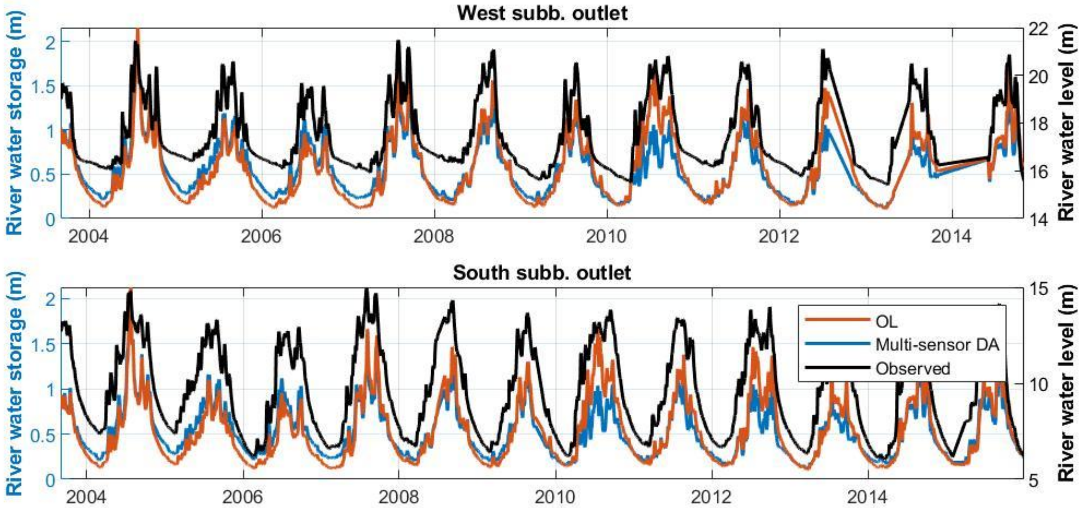

Figure R5. River water storage time series derived from OL (red) and multi-variate DA (blue), as well as in situ water level observations (black, with y axis on the right). The time series correspond to the two sites shown in Fig. 1**. This figure is analogous to Fig. 9 of the main manuscript.**

As a summary, except for slight differences in some aspects including horizontal SSM update distribution and river water storage trends, the estimates obtained by removing the glacier trends prior to DA and after DA are overall very similar.

3. While the assimilation improves TWS estimates, the manuscript reports a non-trivial reduction in Nash-Sutcliffe Efficiency (NSE) for streamflow at certain gauging stations. This result is concerning and warrants a more thorough investigation. Specifically:

- What are the dominant mechanisms driving this degradation? Possible causes include cross-variable error covariances, inadequate localization, or structural limitations of the W3RA baseflow parameterization.
- Is the W3RA model's baseflow representation adequate for the Brahmaputra's complex glacial and snowmelt-driven hydrological regime?
- Have the authors considered recalibrating the model after assimilation to mitigate this effect?

A dedicated subsection analyzing the sources of streamflow degradation would substantially strengthen the manuscript.

Reply: We thank the reviewer for this comment. From the way the comment is formulated, we interpret that the reviewer is referring to the results presented in Appendix D (Fig. D2). This figure does not represent the general findings of the main experiment but refers to a side-experiment where the surface water component was allowed to be updated during the DA process. As expressed in the manuscript, this resulted in degradations of the river water storage in the timescales below 30 days and therefore motivated the exclusion of the river water storage component from the DA update process in the main experiments (lines 739-745). We understand that the fact that Fig. D2 looks very different from Fig. 9 suggests that these are new results. To avoid confusion, Fig. D2 of the manuscript will be adapted as shown below (Fig. R6) so that it is more similar to Fig. 9.

The results of the experiments presented in the main manuscript suggest that the DA process has very little impact on the river water storage estimates (except for a long term trend), and therefore no significant degradation is observed in terms of NSE. Therefore, we interpret that the comments posed by the reviewer do not refer to the main results.

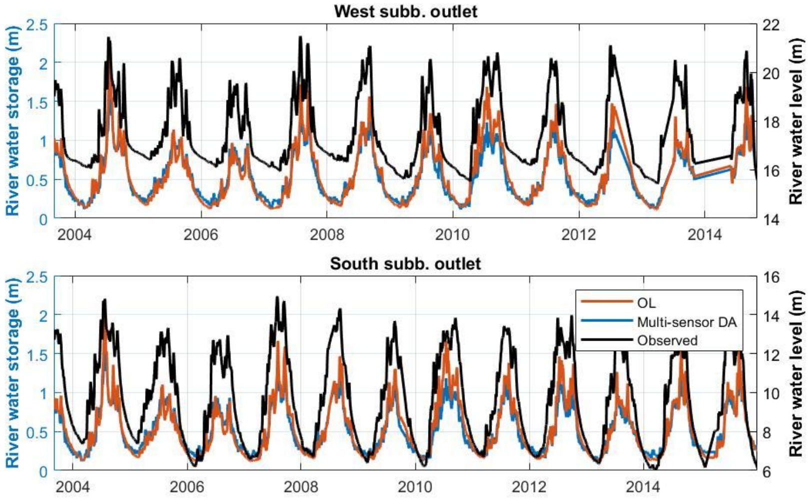

Figure R6. River water storage time series derived from OL (red) and multi-variate DA when river water storage is included in the state vector to be updated (blue), as well as in situ water level observations (black, with y axis on the right). The time series correspond to the two sites shown in Fig. 1.

4. The validation of surface soil moisture relies on WaterGAP model outputs as a reference dataset rather than on independent observational data. This introduces a

circular dependency, as WaterGAP is itself a model subject to its own structural and parametric uncertainties. The authors should:

- Quantify the uncertainty associated with WaterGAP SSM estimates and discuss how this uncertainty propagates into the validation conclusions.
- Consider supplementing the validation with observational datasets such as ESA CCI Soil Moisture, SMAP, or in-situ measurements where available.

Reply: We agree that the ideal situation would be to validate SSM against independent data. However, no in-situ measurements are available for the study region. Additionally, the assimilated SSM product (ESA CCI Combined SSM) contains data from various active and passive satellites, in such a way that validation against any of these satellite SSM retrievals cannot be considered independent (lines 182-186).

Given this situation, the results of the experiments are compared against the WaterGAP model. This model is chosen because it is calibrated against streamflow over the Brahmaputra River basin and has a more elaborate representation of snow melt.

Despite the chosen strategy, we acknowledge that the WaterGAP model might present its own biases and limitations, and this comparison approach cannot be equated to a validation against independent observations. An elaboration will be added at the end of the comparison section to clarify this aspect.

5. The proposed model space mixed localization scheme is a central methodological contribution of this work. However, the manuscript does not present a systematic sensitivity analysis of the localization radii applied to the different model state variables. Given that localization is known to critically influence EnKF performance, the authors should demonstrate:

- How sensitive are the results to the chosen localization radii?

Reply: A few experiments have been conducted to test the sensitivity of the results to the localization radius but were not shown in the manuscript for conciseness. More specifically, an alternative experiment has been performed by using a radius of 3 degrees for groundwater. Tests with additional localization radii will be carried out. Due to the limited time available to respond to the reviewer comments, these experiments will be performed later at the manuscript revision stage. The results and a brief discussion will be added in Appendix D ("Evaluating the impact of additional DA setting parameters") in the revised manuscript.

- What criteria were used to select the final parameter values?

Reply: Thanks for the comment, we acknowledge that information regarding the radius choice is lacking in the main manuscript. The SSM localization half-radius (0.5 degrees) was chosen based on the spatial variability that can be expected from the observations, as model estimates were found to be overly smooth in the area. The groundwater localization half-radius was set to 10 degrees to avoid unrealistically sharp updates and anomalies in the study area (see also Retegui-Schiettekatte et al., 2025). A few sentences will be added specifying this in the Model run and DA section.

Were alternative localization configurations tested?

Reply: Yes. A more conventional observation space localization approach and non-localized DA were tested. A comparison of the impact of each localization approach on the DA results is provided in Section 4.2 of the main manuscript. A sentence will be added in the Introduction to make the reader aware about the fact that this comparison was performed.

6. The manuscript reports quantitative estimates of land water loss in the Brahmaputra basin. However, no uncertainty bounds (e.g., confidence intervals derived from the ensemble spread, or sensitivity to model assumptions) are provided for these estimates. Given the policy relevance of these figures, particularly in the context of climate change and transboundary water management, the authors must include a rigorous uncertainty analysis accompanying all reported water loss values.

Reply: We acknowledge that a more complete discussion of possible uncertainty sources is lacking in Section 4.4. First, the DA process provides quantitative ensemble-based uncertainties that should be analysed. In the revised manuscript, these uncertainties will be computed and a paragraph will be added to analyse and evaluate their magnitude.

Second, the validation of groundwater component in this manuscript is mainly carried out in terms of correlation coefficients, which validates the overall groundwater variations but cannot confirm their magnitude. Therefore, although a clear negative water storage trend seems to exist over the whole basin, its magnitude should be interpreted with caution. A few sentence will be added recalling this fact and to make the reader aware of this possible uncertainty source.

Minor comments:

1. **Table 1:** The notation used to describe model state variables is inconsistent with the notation employed in the equations. A unified nomenclature should be adopted throughout the manuscript.

   Reply: We acknowledge the reviewer's point. The reason for this is that, in the EnKF update equations, all model variables are included in the model state variables X- and

X+. A line will be added prior to the equations, specifying the water storage variables that X contains.

2. **Equation 6:** There appears to be a typographical error in the formulation of the observation operator. The authors should verify the consistency of this equation with the surrounding text and correct it accordingly.

Reply: We thank the reviewer for this comment. After reviewing the formulation, the symbol representing the observation operator in Eq. 6 (A) seems to be consistent with the symbol used in the other equations as well as in the text. However, we acknowledge that the symbol "H" is more usually used to refer to the observation operator, which might have caused this confusion. To avoid future confusion, the observation operator will be referred to as "H".

3. **Definition of SSM:** The manuscript uses the term "surface soil moisture" inconsistently, at times referring to the top soil layer and at others to a vertically integrated quantity. A precise and consistent definition should be provided at first use.

Reply: We thank the reviewer for pointing out this issue. The definition of these variables will be revised to ensure that they are properly defined and consistently used.

4. **Asynchronous Assimilation:** The authors propose a hypothesis regarding the effects of asynchronous assimilation on streamflow performance. This hypothesis, while plausible, remains speculative and is not supported by direct evidence within the manuscript. It should either be tested explicitly or clearly framed as a working hypothesis requiring future investigation.

Reply: We thank the reviewer for this comment. We interpret that the reviewer refers to lines 508-514, which state: "*A likely reason could be the sequential or asynchronous assimilation of TWS and SSM observations in their approaches, meaning that the TWS and SSM observations are not (always) assimilated simultaneously. The fact that either TWS or SSM is unconstrained during the DA update could have enhanced the cross-influence between them, leading to problems when the observations introduce conflicting or anti-correlated updates. Although this explanation is plausible, further investigation is needed to confirm it and better understand how the cross-variable influence works in asynchronous DA*". We think that, in its current form, the manuscript clarifies that this is a hypothesis that will require further research.

5. **Relegation of Results to Appendices:** Several results that appear central to the evaluation of the proposed localization scheme are presented in appendices. The

authors should consider whether these results merit inclusion in the main body of the manuscript to improve readability and scientific transparency.

Reply: We acknowledge that many results are presented in the Appendices. However, we think that these results are not central to the study, but rather describe side-experiments or quantitative results supporting what can be seen from the visual results presented in the main manuscript. Based on the last comment of reviewer #3, which suggests that the "Results" section is rather hard to follow, we will abstain from adding more content to the main body of the manuscript to keep the results clear and to the point. Nevertheless, we will revise the text to make sure that all the appendices are appropriately referenced where relevant to ensure transparency.

6. **Model Recalibration:** Given the known structural limitations of W3RA in glacially influenced basins, the authors should discuss whether a recalibration of the model prior to assimilation was considered and, if not, justify this decision.

Reply: We thank the reviewer for this comment. A calibration of the model was not considered due to two reasons: (i) this study mainly focuses on the improvement of land water storage, including groundwater and SSM, rather than streamflow; and (ii) the W3RA model does not represent glaciers, in such a way that calibration against streamflow would hardly help account for this factor in the streamflow. In any case, a calibration of the model prior to DA is likely not to change the lack of sensitivity of streamflow to the multivariate DA experiment, which hinders the usefulness of such a calibration. A few lines on this will be added in the Methods section of the paper.

7. **GRACE Spatial Resolution:** The coarse spatial resolution of GRACE/GRACE-FO (~300 km) relative to the spatial heterogeneity of the Brahmaputra basin may limit the physical interpretability of the assimilated signal at sub-basin scales.

   The authors should explicitly discuss this limitation and its potential impact on the results.

Reply: We acknowledge that satellite-based TWS, especially sub-monthly TWS products, might have a limited resolution. However, in this study, the sizes of the three sub-basins where TWS is assimilated (100 000 - 250 000 km2) were chosen to be consistent with the approximate GRACE resolution. Therefore, the reduced resolution is not expected to be an issue at sub-basin scale. A sentence will be added in Section 2.1 ("Study area") to clarify this choice.

Data Availability:

The Zenodo DOI provided for data availability appears to be reserved but not yet publicly accessible. The authors should confirm that all data and code necessary to reproduce the results will be made available upon publication, in accordance with the journal's open science policy.

Reply: We confirm that all the data and code necessary to reproduce the results will be made available upon acceptance of the manuscript.

Conclusion:

This manuscript makes a genuine and timely contribution to the field of hydrological data assimilation in data-scarce, high-mountain basins. The proposed localization scheme is innovative, the validation framework is multi-variable, and the application to the Brahmaputra basin is scientifically relevant. Nevertheless, the major concerns outlined above, particularly regarding ensemble size justification, glacial signal treatment, streamflow degradation, and uncertainty quantification, must be thoroughly addressed before the manuscript can be recommended for publication.

Reply: Thanks for the encouraging comments. We would like to thank the reviewer for the useful and specific comments, and we believe they will help to improve the submitted manuscript.

References

De Lannoy, G.J.M., Bechtold, M., Albergel, C., Brocca, L., Calvet, J.-C., Carrassi, A., Crow, W.T., de Rosnay, P., Durand, M., Forman, B., Geppert, G., Girotto, M., Hendricks Franssen, H.-J., Jonas, T., Kumar, S., Lievens, H., Lu, Y., Massari, C., Pauwels, V.R.N., Reichle, R.H., Steele-Dunne, S., 2022. Perspective on satellite-based land data assimilation to estimate water cycle components in an era of advanced data availability and model sophistication. Front. Water 4. https://doi.org/10.3389/frwa.2022.981745

Retegui-Schiettekatte, L., Schumacher, M., Madsen, H., Forootan, E., 2025. Assessing daily GRACE Data Assimilation during flood events of the Brahmaputra River Basin. Science of The Total Environment 975, 179181. https://doi.org/10.1016/j.scitotenv.2025.179181

Springer, A., De Lannoy, G., Rodell, M., Ewerdwalbesloh, Y., Gerdener, H., Khaki, M., Li, B., Li, F., Schumacher, M., Tangdamrongsub, N., Tourian, M.J., Nie, W., Kusche, J., 2026. A review of current best practices and future directions in assimilating GRACE/-FO terrestrial water storage data into numerical models. Hydrology and Earth System Sciences 30, 985–1022. https://doi.org/10.5194/hess-30-985-2026