

Dear editor and reviewers,

We would like to thank you for the critical and useful comments provided in the first review round. In what follows, we have tried to answer to each of the inquiries. Reviewer comments are shown in black, and the answers are shown in blue. Note that line numbers refer to the updated manuscript, except when otherwise indicated.

Response to Reviewer #1.

We would like to thank Reviewer #1 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.

Comment 1: Retegui-Schietekatte et al. address the problem of observation localization in multivariate ensemble data assimilation. The application here is integration of both GRACE TWSA maps and satellite SSM maps in hydrological modelling for the Brahmaputra basin.

The Brahmaputra basin (BB here) represents a hydrologically complex region with significant long-term changes due to warming climate in the Himalayas, above-average sea level rise with increased river blocking and salinization, and socio-economic processes such as land use change, population pressure and water diversion, but also with frequent large-scale flooding caused by cyclone landfall. Since the hydrological ground infrastructure is sparse, improved understanding of hydrological processes with the aid of satellite data is welcome. The manuscript addresses thus an important topic with broad relevance.

Reply: Thanks for the acknowledging the significance of the study.

GRACE data assimilation has been often evaluated with sparse surface soil moisture data, and only few papers address joint assimilation. As most frameworks nowadays seem to rely on some localization technique, a systematic study such as provided here is welcome. The proposed model space mixed localization appears new to me in this context. The results are interesting and support the new idea. This would certainly warrant publication. The authors may want to elaborate somewhat more on the reason for the need for localization, is it due to spurious model ensemble correlations and if yes why, is it due to the observation error characteristics or is it both?

Reply: We agree that additional contextual information on the purpose and advantages of covariance localization would be beneficial. [The following paragraph was added to the Introduction section \(lines 97-103 of revised manuscript\):](#)

One of the methodological aspects in which previous studies diverge is the implementation of covariance localization. Covariance localization is widely used in hydrological DA, as well as in other geoscientific disciplines, to mitigate spurious ensemble correlations or sampling errors arising from limited ensemble sizes and inherent constraints in ensemble perturbation methods. Especially, distance-based localization helps to confine the influence of observations to relevant spatial domains, thereby increasing their impact on the relevant model variables (Houtekamer et al., 1998; Hamill et al., 2001).

Comment 2: However, the manuscript in present form suffers from two major drawbacks, in my view: (1) The hydrological model is not well motivated and described, and it is thus not clear at all what the baseline for the data assimilation is. (2) The evaluation of results appears not very thorough. Here are more details:

The hydrological model that is used in this study is neither mentioned in the abstract nor in the introduction, which raises a red flag. The BB hydrology is really complex and the choice of a model should be motivated by the study purpose; are we interested in river discharge forecasting, in flood potential monitoring on timescales of days of weeks, in agricultural forecast, in water management at which level, water resources, or in climate change assessment on decadal timescales? The authors need to clarify this and explain why their model is fit for purpose, and how it compares to other global or regional models that have been used to describe the BB hydrology.

Reply: We acknowledge that this study design choice should be clarified. The objective of the study is to generate improved *water storage* estimates in the sub-seasonal to decadal timescale for the Brahmaputra River basin, with the aim of supporting future investigations concerning the water storage variability and availability (in the inter-annual to decadal timescale) as well as the role of land water in flood generation mechanisms (in sub-seasonal to inter-annual timescale). [A sentence was added in the Introduction section to specify this \(lines 107-110\).](#)

Especially, we aim to improve model water storage estimates in the sub-seasonal to decadal timescale, with the aim of supporting future investigations concerning the water storage variability and availability (in the inter-annual to decadal timescale) as well as the role of land water in flood generation mechanisms (in sub-seasonal to inter-annual timescale).

Given this study purpose, the W3RA model was chosen based on previous works suggesting its good performance at representing water storage variability when compared to other state-of-the-art models (Mehrnegar et al., 2020). In addition, the result evaluations performed in this study also suggest a reasonable agreement between the model and satellite observations prior to DA in terms of water storage. GRACE and modelled TWS present a similar seasonality and interannual variability in terms of timing and magnitude, and mainly differ in the decadal variability and trend, which is a feature that many hydrological models fail to represent accurately (Scanlon et al., 2018; Fig. 3a of original manuscript). Regarding SSM, model estimates and satellite observations show similar dynamics from the sub-seasonal to the inter-annual timescale, when averaged over the basin (Fig. 3e of original manuscript). This suggests that the W3RA provides a reasonable representation of the large-scale water storage variability in the Brahmaputra River basin and constitutes a strong baseline for the DA experiments. We acknowledge that this information was missing in the paper and we added it in the section “Data and Model” (lines 163-168). Additionally, we now also mention the use of W3RA both in the Abstract as well as in the Introduction.

Previous studies suggest that W3RA offers a good performance at representing water storage variability when compared to other state-of-the-art models (Schellekens et al., 2017; Mehrnegar et al., 2020). The evaluations performed in this study also suggest a reasonable agreement between the model and satellite observations prior to DA in terms of water storage (see Section 4.1), suggesting that W3RA provides a reasonable representation of the large-scale water storage variability in the Brahmaputra River basin and constitutes an appropriate baseline for the DA experiments.

It should be noted that the methods explored in this paper can be applied to any hydrological model. A sentence noting this has been added to the Conclusions section (lines 828-830).

Comment 3: The authors mention some limitations of W3RA, e.g. the model has not been calibrated over the BB region, they use ERA5/ERA5 land forcing but there seems no bias-correction.

Reply: Thank you for this comment. The main reason for not applying model calibration or precipitation bias correction is that we expect the DA process to address possible

limitations of model parameters and forcing data. [A sentence has been added to specify this in the “Data and Model” section \(lines 190-194\).](#)

The model parameterization for the Brahmaputra River basin was performed through a regionalization approach (van Dijk et al., 2013). No further calibration or correction of forcing data was implemented since possible errors resulting from these limitations are expected to be addressed through the DA process. A prior calibration of the model against streamflow would be unlikely to modify the main conclusions of this study due to the little sensitivity of streamflow estimates to water storage dynamics (later analyzed in Section 4.3.4).

Comment 4: Even if the authors must rely on this model due to the assimilation framework, they should compare the open-loop total water storage, surface soil moisture and river discharge to the results from some published global models such as GLDAS, WaterGAP, PCR-GLOBWB in the BB.

Reply: We agree that addition of such comparison would be interesting. However, we think that the realism of the water storage representation within the W3RA is well justified based on the arguments provided in the answer to Comment 3. We consider that an inter-comparison of the OL model performance with other available models is out of scope of this article.

Comment 5: Since TWSA and SSM are assimilated, the authors should also explain the W3RA parameterization of processes relevant to these observables such as crops, water demand and anthropogenic water withdrawal, groundwater, surface water storage beyond river storage (lakes, floodplanes, reservoirs), and soils (which soil maps are used). How is evapotranspiration computed in W3RA? There will be severe limitations in the W3RA representation of these processes, as it is for other models, but whether assimilation is useful depends on our understanding of such error sources.

Reply: We agree that the manuscript could benefit from more details on the representation of various water balance processes in the W3RA model, which would also help interpret the results. [A more detailed description of the model representation of processes affecting TWS and SSM variables has been added in the “Data and model” section \(lines 170-189\):](#)

The W3RA model represents natural land water and energy exchange processes. The soil profile is divided into four layers: topsoil, shallow soil, deep soil, and groundwater (Table 2). In this study, topsoil and shallow soil are treated jointly as “SSM”, and deep soil and groundwater are treated jointly as “groundwater” (a detailed explanation of the reason behind this aggregation is provided in Section 3.1). Water exchange processes within the soil are described through the representation of infiltration, drainage, root water uptake, and capillary rise. Soil and surface water evaporation is computed as a linear function of available energy, and groundwater discharge is estimated through a linear reservoir model (van Dijk, 2010; van Dijk et al., 2013). Heterogeneities in soil composition are not considered, and anthropogenic water use activities, such as groundwater pumping or agricultural irrigation, are not modeled.

Vegetation is modeled in two categories, shallow-rooted short vegetation (grass and shrubs) and deep-rooted tall vegetation (trees), which are represented as two different Hydrological Response Units (HRUs). The proportion of vegetation belonging to each category for each grid cell is determined through the satellite-based tree cover fraction map by Hansen et al. (2003). Vegetation transpiration is computed as the minimum of maximum root water uptake and optimum transpiration rate, where the latter is estimated using the Penman-Monteith equation. Root water uptake is modeled through a linear ramp function (as a function of soil water content) (van Dijk, 2010; van Dijk et al., 2013). Differences in crop types or seasonal variations of vegetation are not considered. More details on the explicit formulation of all these processes can be found in (van Dijk, 2010; van Dijk et al., 2013).

W3RA also has a snow module that represents snow accumulation and melt based on a degree day approach (van Dijk et al., 2013). Glaciers are not represented within the snow module. To compute surface water routing, a grid-based model was coupled with W3RA where each grid cell is represented as a linear reservoir (Mizukami et al., 2016). Lakes, reservoirs and floodplanes are not represented within the routing model, and evaporation of water routed in streams and rivers is not represented either.

Additionally, possible uncertainty sources coming from the model structure are now discussed after the TWS evaluation (lines 634-638) and SSM evaluation (lines 731-735):

TWS: The disagreement could partly originate from model limitations, such as a limited representation of evaporation in surface water bodies, or vegetation and transpiration. However, it likely also reflects a more general difficulty of hydrological models to represent decadal TWS variability and trend related to climate variations and human water use, as reported in a multi-model experiment by Scanlon et al. (2018).

SSM: An important consequence of these results is that the assimilation of SSM could be accounting for the limited performance of the W3RA model at estimating SSM variability in mountain areas, but it is not clear if improvements could be expected from the assimilation of SSM products if a better adapted base model was used (e.g., WaterGAP itself). Evaluating the added value of the satellite-based SSM product with respect to different hydrological models is not within the scope of this study, but should be considered in future satellite SSM DA experiments, especially regarding its operational implementation.

Comment 6: Put in other words, if the baseline model is poor then the assimilation will almost always succeeded in „bringing the model closer to reality“ (page 25).

Reply: We absolutely agree with the statement of the of the reviewer, as DA in a poorly performing model will almost always lead to an improvement. However, we do not think that the W3RA constitutes a poor baseline: at the contrary, we think the model water storage estimates are realistic when compared to observations and other models, as exposed in the answer to Comment 2. We believe that the inclusion of more details in the manuscript concerning the general performance of W3RA compared to other models (see comment 2), the similarity of its estimates with the satellite observations (see comment 2), and the parameterization and structure of the W3RA model (see comment 5) have helped clarify the baseline from which this study departs and convince the reader of its adequacy.

Comment 7: Validation in the BB is challenging, as the authors are aware. This raises the question why the study did not focus on Australia where the model was calibrated and the monitoring infrastructure appears much more complete, or some other region like the US.

Reply: We acknowledge that alternative study regions such as the United States or Australia could have facilitated a more comprehensive validation of our results due to the

availability of extensive open-access hydrological datasets. However, these regions have already been widely investigated in the context of TWS DA and multivariate DA frameworks (Australia: Tian et al., 2017; Khaki et al., 2019, 2020; Tangdamrongsub et al., 2020; U.S.: Girotto et al., 2019; Khaki et al., 2019, 2020).

Our selection of the Brahmaputra River basin was motivated by two primary considerations. First, we sought to evaluate multivariate data assimilation techniques in a hydrological environment that differs substantially from those examined in previous studies. Second, the basin exhibits a pronounced hydrological variability which is difficult to track precisely due to the lack of abundant in situ observations, leading to significant challenges for water resources management and hydrological risk mitigation. As such, this region stands to benefit from refined and more accurate hydrological water storage estimates. Despite the limited availability of ground-based hydrological observations, we were able to collect in-situ groundwater and river water level measurements for portions of the basin. These datasets provide a foundation for validating a subset of our model outputs.

We added a paragraph to the Study area section to clearly articulate the rationale behind the selection of the study area (lines 146-153).

The Brahmaputra River basin was selected as a study area for two reasons: (i) we sought to evaluate multi-variate data assimilation techniques in a hydrological environment that differs substantially from those examined in previous studies (e.g., Australia or the US, see Table 1); and (ii) the basin exhibits a pronounced hydrological variability which is difficult to track precisely due to the lack of abundant in situ observations, leading to significant challenges for water resources management and hydrological risk mitigation. As such, this region stands to benefit from refined and more accurate hydrological water storage estimates. Despite the limited availability of ground-based hydrological observations in the region, we were able to collect in situ groundwater and river water level measurements for portions of the basin. These datasets provide a foundation for validating a subset of our model outputs.

Comment 8: Comparison against the SPEI appears doubtful to me – W3RA also uses precipitation and radiation input and computes evapotranspiration and I wonder why then the simple SPEI could be considered as a „truth“ for groundwater change?

Reply: We thank the reviewer for this comment and we acknowledge that SPEI relies on a model and uncertain input data, and therefore is affected by its own uncertainties. Comparison against SPEI in this study is not meant as a validation, but rather as a consistency check in sub-basins where no other data is available. In regions where in-situ groundwater data was available, validation was performed against in-situ data. [A](#)

[sentence has been added to clarify this at the stage where the SPEI comparison is described \(lines 630-632\).](#)

[Because SPEI is derived from climatic water balance rather than direct observations of storage, the comparison is interpreted as a consistency check rather than a strict validation.](#)

Comment 9: Many studies have shown that accumulated precipitation (minus evapotranspiration) corresponds well to TWSA but this does not prove anything about groundwater, in a region where surface water dynamics plays a BIG role.

Reply: We completely agree with the reviewer and acknowledge that surface water has a large contribution to TWS variability in the Brahmaputra River basin (see, e.g., Retegui-Schiettekatte et al., 2025). More specifically, river water level observations suggest that, while no significant decadal trend is observed in river water storage, it can be affected by a strong inter-annual variability. To account for this factor, [in the revised manuscript the SPEI is compared to modelled TWS rather than modelled groundwater](#). Additionally, the glacier trends are also included in this comparison, since glacier mass changes can also be equated to temporally aggregated precipitation minus evapotranspiration dynamics. [A section describing the comparison has been added in the Results section \(lines 630-644\).](#)

Comment 10: In this reviewer's opinion, the authors need to validate groundwater change against quality-controlled well data with documented specific yield information (it wasn't clear to this reviewer whether this was available for the wells in Fig. 7). Or they validate against specific groundwater 3D modelling that has been calibrated against wells. If this is not possible, this kind of validation cannot be performed.

Reply: We agree that groundwater validation is a critical part of this study and should be well supported with quality-controlled data and specific yield information. The groundwater level data used to validate the results have been quality controlled by the collecting institution (the Bangladesh Water Development Board) and curated by Shamsudduha et al. (2022). [A sentence has been added in the "Data and Model" section of the manuscript to clarify this \(line 225\).](#)

[The groundwater level data have been quality controlled by the BWDB and curated by Shamsudduha et al. \(2022\).](#)

In the original manuscript, no specific yield data was used to perform the groundwater validation. Following the reviewer's suggestion, specific yield data has now been used to convert groundwater levels to groundwater storage. The distributed specific yield information was originally obtained from around 200 pumping tests and interpolated using ground lithology maps by Shamsudduha et al. (2011). The specific yield values range from 1% to 20% with a mean value of 6%. The groundwater storage changes were computed by multiplying groundwater level change (Δh) with the specific yield (S_y), $\Delta S_{GW} = \Delta h \times S_y$. This information has been added in the "Data and Model" section of the manuscript (lines 226-227).

Specific yield information was originally obtained from around 200 pumping tests and interpolated using ground lithology maps by Shamsudduha et al. (2011).

The use of specific yield data allowed to evaluate the magnitude of model groundwater variability. The results are presented in the Results section of the manuscript (lines 645-681).

Comment 11: And the situation is similar for SSM, the authors need to work with in-situ data from the representative depth range, potentially with models informed by in-situ data. WaterGAP as a conceptual model has known limitations in soil moisture representation, and even disregarding this, comparing soil moisture across different models with differing layers, model physics, data integrated is highly problematic. If there is no ground data, no validation is possible.

Reply: We acknowledge that comparing the DA results with WaterGAP model outputs does not constitute a rigorous validation, but rather provides a consistency check, similar to the comparison with SPEI. A sentence was added prior to the comparison to clarify this aspect (lines 709-711).

Although model outputs might be affected by limitations such as errors in forcing data, process simplification, or missing representation of certain processes, they can provide an additional perspective to perform a consistency check and assess the realism of the DA results.

Unfortunately, openly accessible in-situ soil moisture observations are not available for the study region. A common approach in previous studies has been to compare the DA results with independent satellite observations (e.g., Soltani et al., 2024). However, the assimilated SSM product (ESA CCI Combined) integrates both passive and active soil moisture satellite retrievals; consequently, any comparison against these satellite datasets would also lack independence. We therefore recognize that the absence of independent SSM validation represents a limitation of this study. A sentence has been

[added to the manuscript to explicitly disclose this limitation in the Conclusions section \(lines 836-839\).](#)

A limitation of this study is the absence of independent SSM observations to validate the model dynamics. To overcome this, a consistency-check was performed by comparing the estimates to those obtained by the WaterGAP model, which is in principle better adapted to represent high-mountain environments.

Comment 12: The authors could still compare their results to measured streamflow. Of course the evaluation should be tailored to the purpose of the study, i.e. looking at either short or long timescales.

Reply: We acknowledge the relevance of this suggestion. The improvement of streamflow is not one of the primary objectives of this study (see reply to Comment 2), but it is true that a comparison against streamflow at the basin outlet would help validate the water storage changes, which would be relevant given the absence of other independent validation data such as in-situ soil moisture observations. However, comparisons against river water levels provided in Section 2.2 of the original manuscript denote a lack of sensitivity of model river water storage estimates to changes in soil and groundwater storage. This lack of sensitivity equally affects streamflow estimates, and therefore a comparison against streamflow would not lead to any conclusive results.

Comment 13: Last, I noticed the authors compared river water storage from the assimilation to gauge observations, this requires an explanation. While stage-discharge (rating curve) relations are typically derived empirically from simultaneous measurements with ADCPs and gauges, it is less clear how the routed model simulation can be mapped to observable water level change as this would require river cross-section assumptions.

Reply: As the reviewer suggests, we acknowledge that river water storage cannot strictly be compared to water levels without assumptions on the river cross section. Due to this reason, model estimates of river water storage were rescaled to match the observed river level data before validation. In other words, we assume a rectangular cross-section and we set the width of the rectangle to obtain a maximum agreement between observed and modelled water levels. We acknowledge that this approach has many limitations. However, in any case, the final conclusion of the streamflow comparison is the little sensitivity of the model estimates to water storage updates in the seasonal to sub-seasonal timescale. Therefore, the simplified the cross-section assumption does not hinder any of the conclusions. [This assumption is now mentioned in Section 3.2 \(lines](#)

474-479) and a sentence has been added in Section 4.3.3 to recall the approach adopted (lines 743-746).

Section 3.2 (Assessment of the DA performance)

Comparing river level to river storage requires assumptions on the river cross-section. Due to the absence of this information, we here perform a simple transformation by rescaling the modeled water storage estimates to match the variance of the observed river level data. This is equivalent to assuming a rectangular cross-section and we set the width of the rectangle to obtain a maximum agreement between observed and modeled water levels. Although this approach has many limitations, it does not affect the final conclusion regarding little sensitivity of the model estimates to water storage updates in the seasonal to sub-seasonal timescale (Section 4.3.4).

Section 4.3.4:

Comparing river level to river storage requires assumptions on the river cross-section. Due to the absence of this information, the modeled water storage estimates were rescaled to match the variance of the observed river level data before validation (Section 4.2). Although this approach has important limitations, it does not affect the main conclusions described in what follows.

Final reply: We would like to thank the reviewer for the exhaustive and critical comments. We believe that the multiple suggestions made to improve the result evaluation and validation have helped to strengthen our study and the manuscript.

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: In this study, an ensemble of 30 members was chosen following previous TWS and SSM DA studies (De Lannoy et al., 2022; Springer et al., 2026). We acknowledge that the use of an ensemble size of $N=30$ can lead to sampling errors and spurious correlations. To address this, covariance localization was implemented to help avoid spurious correlations between distant model variables. [A sentence was added to highlight this aspect \(lines 386-388\):](#)

An ensemble size of $N_e=30$ members was chosen to perform the experiments, following previous TWS and SSM DA (De Lannoy et al., 2022; Springer et al., 2026), together with covariance localization techniques to avoid possible spurious correlations between distant grid cells.

Although testing the sensitivity of results to different ensemble sizes could be interesting, implementing these experiments would entail an increased computational demand and increased complexity of the manuscript. Given the large number of experiments already presented in the manuscript, we have decided to omit this sensitivity analysis.

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, the long-term trends caused by glacier retreat are found to be introduced into the groundwater component, and are accounted for *a posteriori*. [This is now highlighted in lines 528-530 of the manuscript.](#)

Note that groundwater is the only land water storage component that exhibits negative trends, and therefore this component is likely absorbing the negative trends induced by glacier decline in the East and West sub-basins.

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. 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.

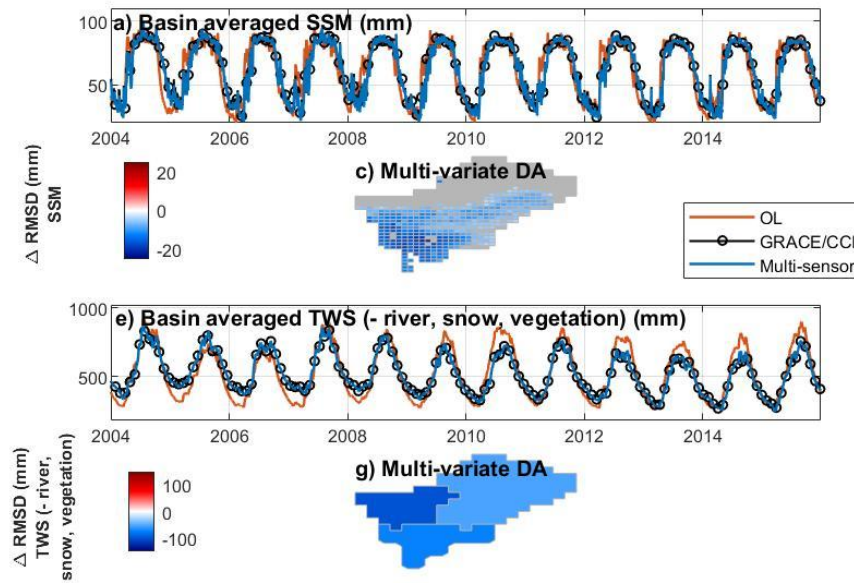


Figure R1. Basin averaged time series of SSM and TWS (a,e) and improvement in the RMSD of tile-averaged SSM (c) and sub-basin averaged TWS (g) of the model, after multivariate assimilation with when excluding glacier trends from TWS time series. Δ RMSD is defined as the RSMD of the DA product minus the 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.

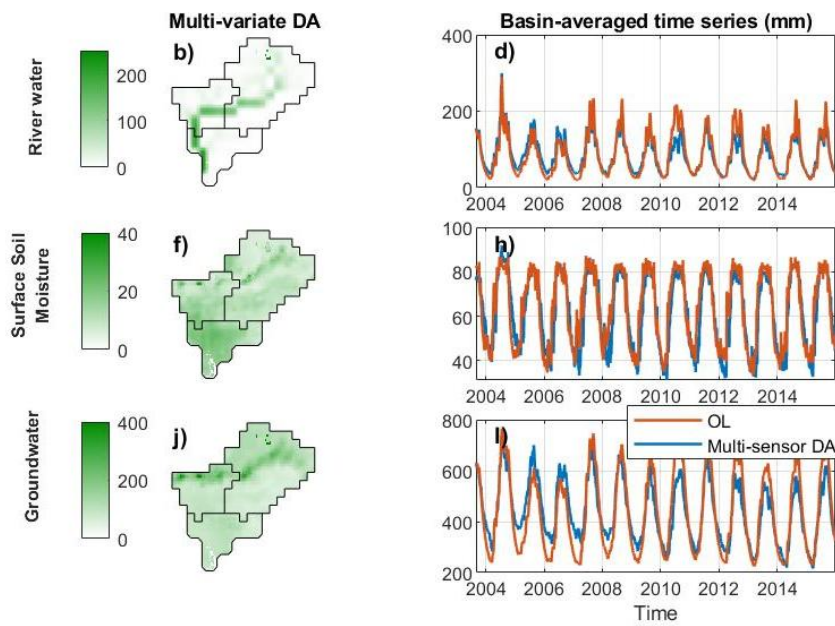


Figure R2. Standard deviation of DA minus OL on spatially distributed ($0.1^\circ \times 0.1^\circ$ grid) model estimates for river water (b), SSM (f) and groundwater (j), obtained for the multivariate DA experiment assimilating TWS without glacier trends. The standard deviation is expressed in mm. The column in the right (d, h, l) shows basin-averaged time series for the same 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.

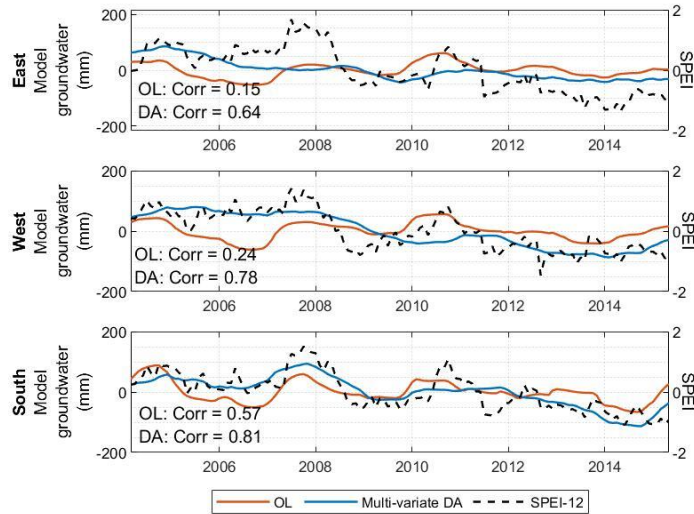


Figure R3. Inter-annual groundwater variability of OL and multi-variate DA; SPEI time (rescaled, on the right axis). **This figure is analogous to Fig. 6 of the main manuscript.**

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 where 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.

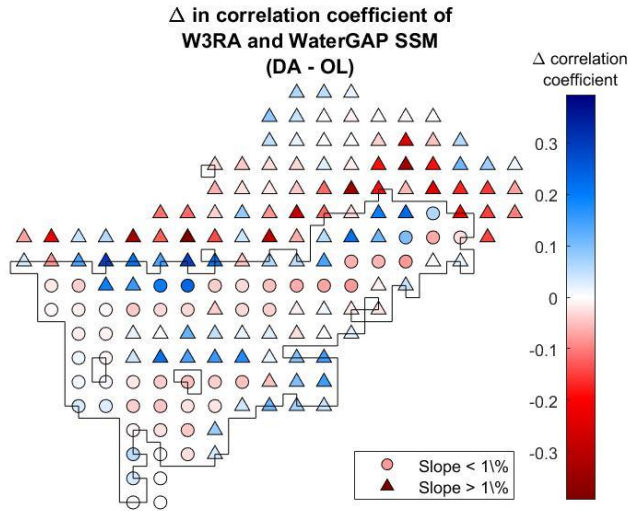


Figure R4. Distributed map of change in correlation coefficient between W3RA and WaterGAP after multivariate DA when assimilating TWS observations without glacier trends. The black outline indicates the area where SSM observations were assimilated. **This figure is analogous to Fig. 8c of the main manuscript.**

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.

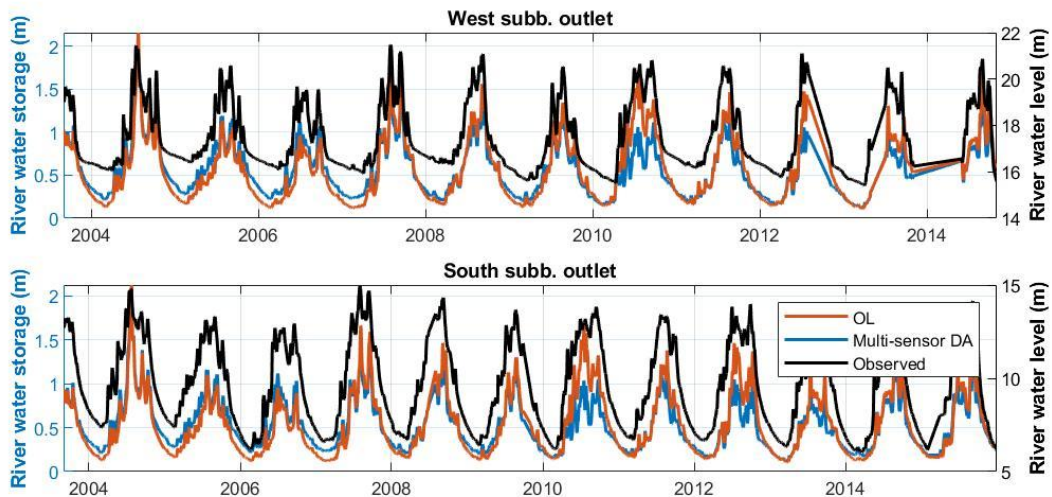


Figure R5. River water storage time series derived from OL (red) and multi-variate DA (blue), as well as in situ water level observations (black, with y axis on the right). The time

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.

For conciseness, a summary of these results has been added to Appendix D of the manuscript. Additionally, a sentence has been added in the Section 3.1 referring to the Appendix (lines 404-407).

In the main experiments presented in this study, we do not apply this reduction to equate the experimental conditions to regions where no robust glacier retreat estimates are available. As part of the evaluation, we assess how the W3RA accommodates these trends. Additionally, a side-experiment involving an a-priori glacier trend removal is described and evaluated in Appendix D.

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:

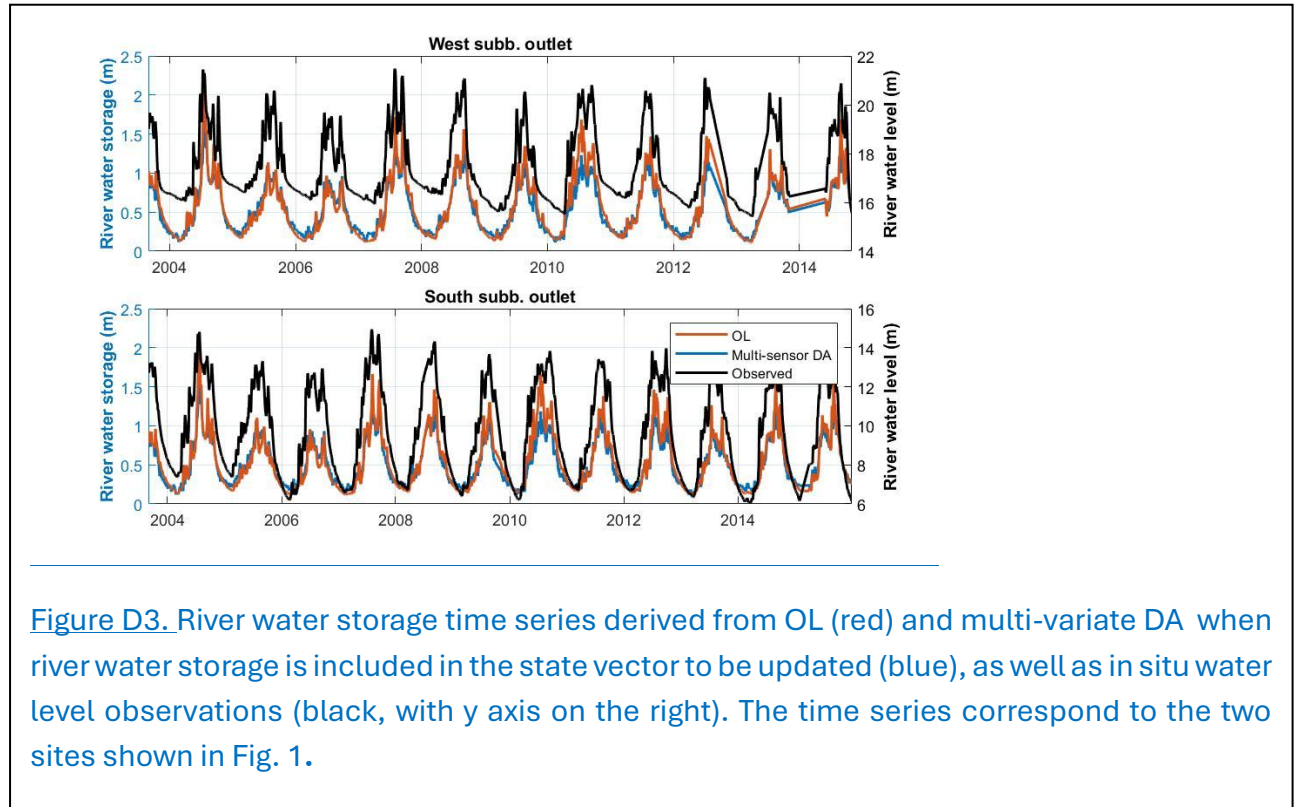
- What are the dominant mechanisms driving this degradation? Possible causes include cross-variable error covariances, inadequate localization, or structural limitations of the W3RA baseflow parameterization.
- Is the W3RA model's baseflow representation adequate for the Brahmaputra's complex glacial 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 of the original manuscript). 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 of original manuscript). 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 has been adapted as shown below 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.



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 247-249).

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 has been added at the beginning of the comparison section to clarify this aspect (lines 709-711).

Although model outputs might be affected by limitations such as errors in forcing data, process simplification, or missing representation of certain processes, they can provide an additional perspective to perform a consistency check and assess the realism of the DA results.

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. Tests with additional groundwater localization radii of 0.5, 4, and 20 degrees as well as no localization have been now carried out and are described to Appendix D, together with a discussion on their main implications.

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 have been added specifying this in the Model run and DA section (lines 434-444).

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 has been added in the Introduction to make the reader aware about the fact that this comparison was performed (lines 110-112).

To reach the first objective, we test and compare different covariance localization settings used in previous literature and propose a novel approach, here called “model space mixed localization”, to maximize the performance of the multi-variate DA.

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. These uncertainties have been computed and a paragraph has been added to evaluate their realism (lines 796-902).

The DA process provides uncertainty values for model estimates, which are computed based on ensemble statistics. The statistics suggest an average standard deviation value of 15 mm for TWS, with a maximum value of 36 mm in periods with a high water storage variability, which is rather low compared to the large seasonal variability of TWS. However, these values are likely underestimated, since the uncertainty of daily GRACE TWS used in this study was approximate due to the absence of exact estimates, and possible temporal autocorrelation in observations were not accounted for (Section 2.2). Future studies quantifying the magnitude and structure of sub-monthly TWS errors will enable more robust DA frameworks that are able to represent model variable uncertainties with increased realism.

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 sentences will be added recalling this fact and to make the reader aware of this possible uncertainty source in the TWS evaluation section (lines 633-634).

Although SPEI is standardized and is not suitable for validating the magnitude of TWS variability, it reveals clear negative trends for the three sub-basins in the period 2004-2014.

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 has been added prior to the equations, specifying the water storage variables that X contains (lines 260-262).

the model state $x_k^{(i)}$ represents the water storage values predicted by the model for each grid cell and each water storage component in that timestep (see a list of these components in Section 3.1)

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 carefully reviewing the formulation, we could not find any inconsistencies in the use of the symbol representing the observation operator in Eq. 6 (A). However, we acknowledge that the symbol " H " is more typically used to refer to the observation operator, which we think might have caused this confusion. To avoid future confusion, the observation operator will be referred to as " H ".

2. **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 has been revised to ensure that they are properly defined and consistently used. The definition can now be found in lines 170-173:

The soil profile is divided into four layers: topsoil, shallow soil, deep soil, and groundwater (Table 2). In this study, topsoil and shallow soil are treated jointly as “SSM”, and deep soil and groundwater are treated jointly as “groundwater” (a detailed explanation of the reason behind this aggregation is provided in Section 3.1).

3. **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 of the original manuscript, 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.

4. **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 have revised the text to make sure that all the appendices are appropriately referenced where relevant to ensure transparency.

5. **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 (lines 191-195).

No further calibration or correction of forcing data was implemented since possible errors resulting from these limitations are expected to be addressed through the DA process. A prior calibration of the model against streamflow would be unlikely to modify the main conclusions of this study due to the little sensitivity of streamflow estimates to water storage dynamics (later analyzed in Section 4.3.4).

6. **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 was added in Section 2.1 (“Study area”) to clarify this choice (lines 157-160).

For the assimilation of TWS and interpretation of the results, the basin was divided into the East, West and South sub-basins (Fig. 1), with areas of 100 000 - 250 000 km² chosen to be consistent with the approximate daily GRACE TWS resolution. Therefore, the reduced resolution of daily observations is not expected to be an issue at sub-basin scale.

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 have helped to improve the submitted manuscript.](#)

Response to Reviewer #3.

[We would like to thank the Brahma Dutt Vishwakarma 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.](#)

Summary: The manuscript improves assimilation of satellite-derived TWS and SSM products into a hydrological model. They explore various data assimilation strategies, including univariate and multi-variate, to obtain better representation of water storage dynamics over the Brahmaputra River basin. They claim to improve TWS, groundwater, and soil moisture representation in the model. The manuscript is well written and addresses a significant challenge/problem in the field.

Reply: [We want to thank the reviewer for the encouraging comments.](#)

Major comments:

1. From a methodological development point of view, the manuscript tries to add information from TWS and SSM in a multivariate DA setting., which is an advance over usual two step DA schemes. The authors also discuss a potential challenge in their approach, which is deciding the weightage given to one variable relative to the other. They discuss tuning approaches and propose observation space mixed localization as a solution. The idea seems interesting. However, see the next point.

Reply: [Thanks for acknowledging the methodological advancement.](#)

2. They fix a smaller region of influence for SSM while a larger for TWS. This appears to be driven by data resolution. However, it would be nice to have some discussion and possibly analysis based on spatial wavelengths of processes responsible for changes in SSM or TWS. Groundwater has a larger wavelength than SSM but the current space localization scheme have these wavelengths at least an order of magnitude higher and I believe it could be two order of magnitudes for the time scale at which the assimilation is done, i.e. daily. Around line 270, the arguments presented are again appearing to be motivated by data resolution rather than from process spatial variability.

Reply: We acknowledge that the criteria followed to select the localization radii could use some more details in the manuscript. We agree with the reviewer that in general SSM should have a smaller localization radius than groundwater storage due to its stronger spatial variability. A spatial semivariogram of the OL SSM estimates suggest that daily model SSM estimates become decorrelated at around 4 degrees of distance (Fig. R1). This large spatial coherence likely occurs due to the smoothness of the model parameters, as well as the large spatial coherence of precipitation occurrence in tropical areas (in the order of 100 km, Moron et al., 2007). However, it is likely that satellite-based SSM observations can reflect spatial heterogeneities that are not reflected in the model parameters. To allow the DA process to introduce these heterogeneities, a shorter localization radius needs to be selected. Therefore, and following previous works (Giroto et al., 2019), a half-radius of 0.5 degrees was selected.

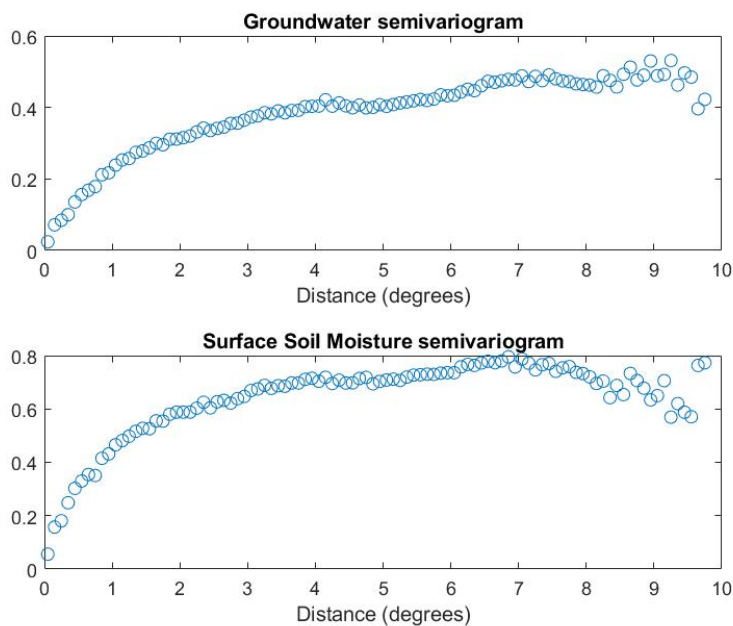


Figure R1. Semivariogram of groundwater and SSM variability in the study area.

As far as groundwater storage is concerned, due to the horizontal independence of model grid cells in the W3RA model, groundwater is found to have a similar semivariogram to that of SSM (Fig. R1). However, the use of small localization radii (3-4 degrees) for groundwater storage was found to lead to very sharp groundwater updates, which degraded the distributed groundwater estimates and reduce the realism of the results (see also Retegui-Schiettekatte et al., 2025).

Therefore, from a practical standpoint, a larger localization radius was finally used for the groundwater storage localization (10 degrees).

A paragraph explaining this has been added to the section Model run and DA (lines 434-444). For conciseness, the figure with the variograms has been omitted in the manuscript.

Localization radii in a model-space localization approach should in principle follow the spatial wavelength of water storage variability represented by the model for each component. When analyzing the spatial SSM variability in W3RA, a very smooth variability was found (variations correlated within distances of a few hundreds kilometers), probably due to the coarse parameterization of the model. Since satellite-based SSM observations are likely capable to reflect spatial heterogeneities that are not represented in the model parameters, a localization half-radius 0.5° was applied to SSM (as well as to the covariance between SSM and groundwater) based on the spatial variability that can be expected from SSM observations (e.g., Giroto et al., 2019), to allow introducing the observed spatial patterns. Regarding groundwater, the model displayed a decorrelation length of around 300 km, so a localization half-radius of 3° was initially deemed appropriate (based on Retegui-Schiettekatte et al., 2025a). However, this was found to produce very sharp anomalies in groundwater estimates, and finally, a half-radius of 10° was adopted to enhance smoother groundwater variability patterns. In both cases, the values are in the range of TWS localization degrees reported in the literature (see Table 1).

3. Equation 4, suggests independence between TWS and SSM, while it is known to be strongly correlated with the TWS budget equation defining causality. Would it be better to not have off diagonals as zero. This issue also haunts the framework, where TWS and SSM are ingested while information from SSM is a subset of TWS, although at different spatial resolution. A philosophical as well as analytical explanation would help a lot.

Reply: We agree that this point requires further clarification. As noted by the reviewer, correlations between TWS and SSM are expected within the hydrological system. These physical correlations are represented in the off-diagonal elements of the model error covariance matrix $C(X_k^-)$ in Eq. (2).

In contrast, the covariances contained in the observation error covariance matrix Σ_{y^o} (Eq. 4) do **not** describe correlations between hydrological variables. Instead, they represent possible correlations between the *empirical errors* of the GRACE TWS observations and satellite-derived SSM observations. Such empirical observation errors arise from factors including, but not limited to, instrumental noise, GRACE orbit determination uncertainties, and post-processing errors. Given the fundamentally different measurement principles and processing chains underlying TWS and SSM products, it is theoretically unlikely that correlated empirical observation errors would exist between the two.

From a practical standpoint, there is no evidence that incorporating cross-correlations between these observational errors would improve the performance of the DA system. Moreover, estimating the magnitude of such cross-covariances would be highly challenging. For these reasons, and consistent with previous studies (e.g., Wongchuig et al., 2024), the off-diagonal elements of the observation error covariance matrix were set to zero. [A more detailed explanation will be added to the main manuscript to make this point clearer \(lines 286-291\).](#)

Note that the matrix $\Sigma_{y^o_{\text{multi-variate}}, k}$ represents the empirical observation errors of the GRACE TWS observations and satellite-derived SSM observations. Such empirical errors arise from factors including, but not limited to, instrumental noise, GRACE orbit determination uncertainties, and post-processing errors. Given the fundamentally different measurement principles and processing chains underlying TWS and SSM products, it is theoretically unlikely that correlated empirical observation errors would exist between the two, and it is therefore realistic to set the extra-block-diagonal elements in $\Sigma_{y^o_{\text{multi-variate}}, k}$ to zero.

4. The validation could be improved a lot. SPEI-12 is very smooth. With simulations and DA done at daily scale, SPEI-3 itself is good enough. Majority of drought studies use either SPEI-6 or 3.

Reply: We thank the reviewer for this comment. A comparison of water storage changes (smoothed at scales of 3 and 6 months) with SPEI-3 and SPEI-6 is represented in Figs. 1 and 2, respectively (see below). The results show that DA improves the model performance in terms of correlation coefficient in both timescales. However, SPEI-3 and SPEI-6 seem to respond very strongly to sub-seasonal anomalies and do not always capture the seasonality reflected in water storage estimates (see, e.g., beginning 2012 or 2010), leading to overall lower correlation coefficient with both the OL and DA. In contrast, SPEI-12 presents much smoother dynamics and overall higher correlation coefficients with both OL and DA (Fig. 3).

These results are consistent with those obtained in previous studies, where a higher correlation between water storage and SPEI or SPI (Standardized Precipitation Index) was found at timescales of 12 months rather than 3 or 6 months (Zhang et al., 2018; Cammalleri et al., 2019; Satish Kumar et al., 2019). This can be explained by the fact that the SPEI represents anomalies in water fluxes (precipitation minus evapotranspiration, has no memory of past anomalies), while water storage is a time integration of these flux anomalies (has a memory of past states). The strong temporal smoothing applied to SPEI-12 effectively works as a temporal integration and therefore makes it more adequate for comparison with water storage estimates (Cammalleri et al., 2019).

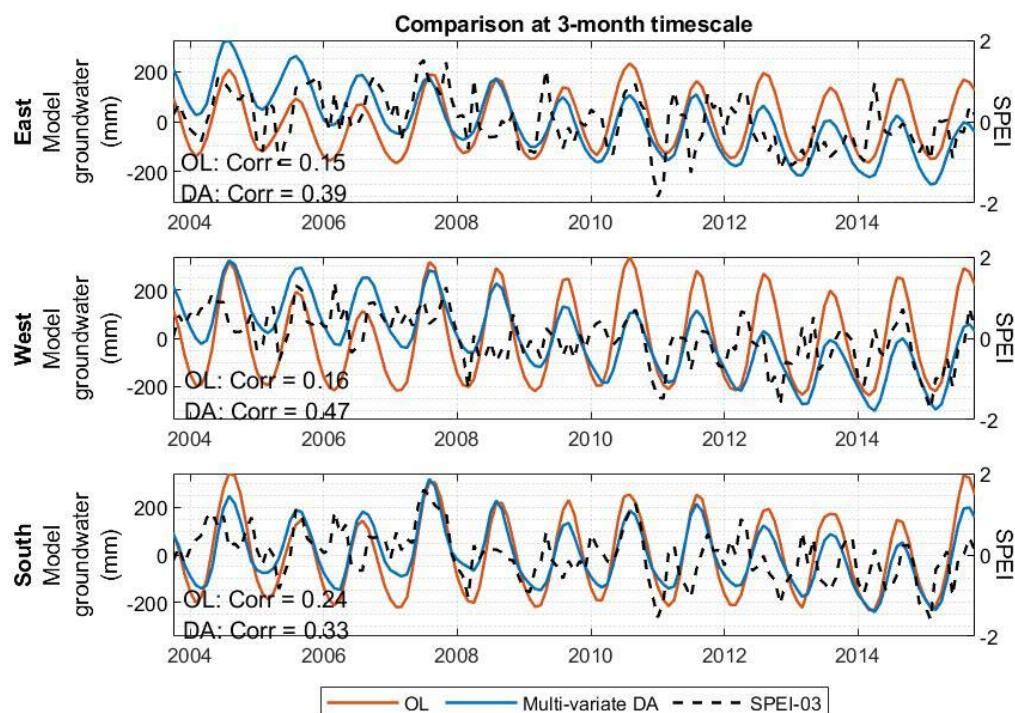


Figure 1. Comparison of OL and multivariate DA groundwater estimates with SPEI at a **3-month** timescale.

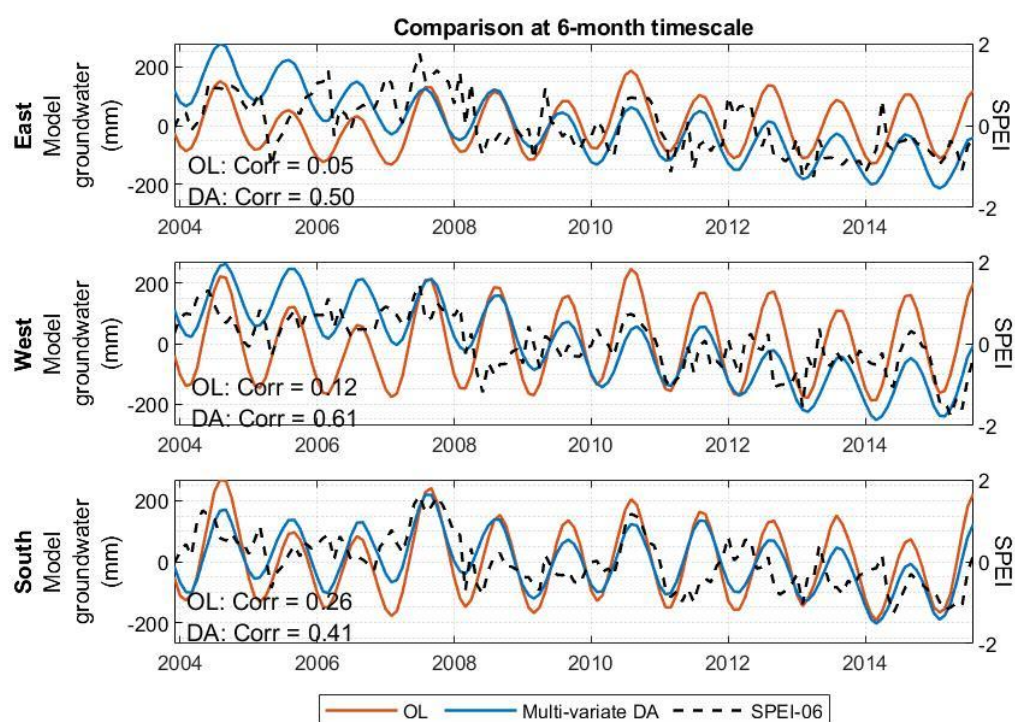


Figure 2. Comparison of OL and multivariate DA groundwater estimates with SPEI at a **6-month** timescale.

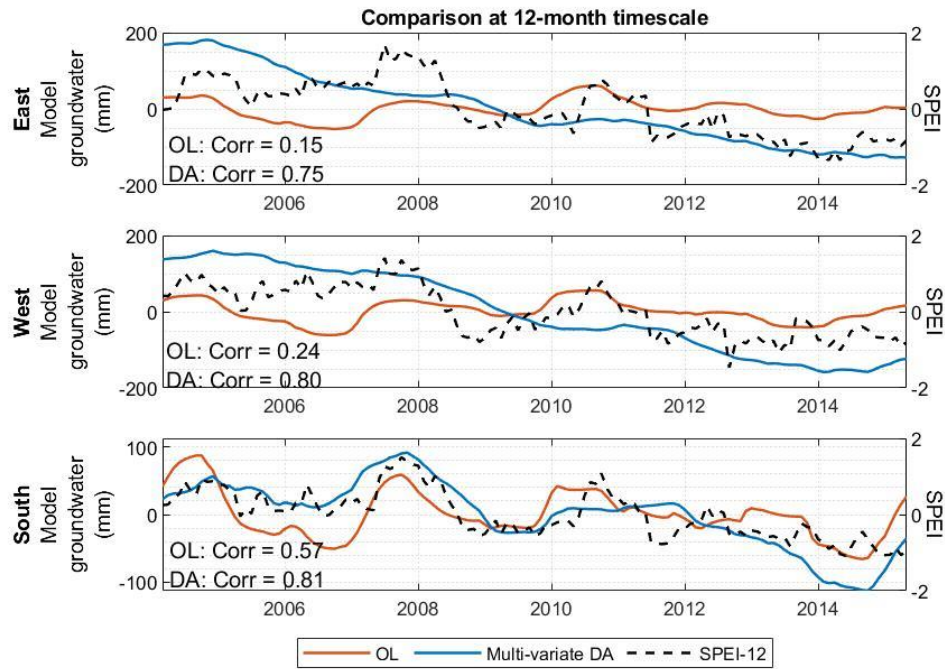


Figure 3. Comparison of OL and multivariate DA groundwater estimates with SPEI at a **12-month** timescale.

Based on the results observed and the theoretical background explaining the poor agreement at the 3- and 6-months timescales, we will keep the comparison with SPEI-12. For conciseness, the comparison with SPEI-3 and SPEI-6 have not been explicitly included in the manuscript, but a few sentences have been added in the results section to justify the comparison with SPEI-12 and briefly report the results obtained in the comparisons with SPEI-6 and SPEI-3 (lines 234-237).

Although the SPEI reflects variability of net water fluxes rather than water storage, previous literature suggests that the aggregation of SPEI over time windows of 12 months or longer effectively works as a temporal integration, leading to higher correlations with TWS (Zhang et al., 2018; Cammalleri et al., 2019; Satish Kumar et al., 2021). Based on this literature, the SPEI-12 was chosen for comparisons in this study.

Secondly, the well observations are compared in term of their correlation in Figure 7. The GW storage from MV-DA appears to do poorly in comparison to OL in c and f subplots (RMSE). This correlation plot also suggests that groundwater spatial-variability is not very large (in line with point 2 above).

Reply: The reviewer suggests that the multivariate DA performs worse than the OL in Fig. 7c and Fig. 7f.

Regarding Fig. 7f, we agree that the multivariate DA underperforms relative to the OL. This outcome is in fact one of the key points highlighted in our analysis. The figure represents the average groundwater time series from a limited number of wells located predominantly along the Brahmaputra River. As discussed in lines 573–578, these locations do not exhibit declining

groundwater trends, likely due to continuous river-induced groundwater recharge. When the negative TWS trend is assimilated in the multivariate DA experiment, a corresponding negative trend is introduced at these sites, which degrades the correlation with the reference groundwater measurements. To clarify this, we will add an explanatory note to the figure caption.

However, for the remaining sites shown in Fig. 7c, we believe it is not accurate to conclude that the multivariate DA performs poorly relative to the OL. At these locations, groundwater levels show a pronounced declining trend that the OL fails to capture but the multivariate DA successfully introduces. Certain short-term anomalies, such as the positive anomaly in 2010–2012, may appear better represented in the OL solution, posing a need for further investigation in future studies, particularly to assess potential links to sub-decadal glacier- or snow-related anomalies. Conversely, other anomalies, such as the 2007–2009 event, are more accurately reproduced by the multivariate DA.

To improve clarity, additional explanatory lines have been incorporated into the manuscript (690-693). Unfortunately, the comparison cannot be performed in terms of RMSD due to the scale difference between groundwater storage and groundwater levels. Therefore, correlation coefficient values will be left as part of the figure.

5. SSM relation with TWS is tricky in irrigated regions. For example, in rainy season, a change in TWS is positively correlated with SSM but when groundwater irrigation is done, SSM is anti-correlated to TWS change. This has a significant impact on water budget also, which can lead to phase shifts in ET (for example, see Goswami et al., 2025). How is the assimilation approach able to deal with such scenarios?

Reply: We definitely acknowledge the reviewer's point. As shown by Goswami et al. (2025), SSM and TWS can be anti-correlated in areas with widespread irrigation practices, such as the Ganges basin. This is likely also the case in some areas of the Brahmaputra with a high presence of irrigation, and should therefore be discussed. The following paragraph has been added in the Introduction section (lines 87-93):

However, conflicts between datasets can appear in regions with the presence of irrigation, as is the case of the Brahmaputra River basin, where negative TWS and groundwater trends caused by water pumping can be strongly anti-correlated with increasing SSM values caused by expanding irrigation practices (Goswami et al., 2025). If the influence of TWS DA in the SSM component is not adequately constrained, this can result in the introduction of long term trends in SSM and evapotranspiration, degrading the representation of these variables within the model (Giroto et al., 2017). In multivariate DA scenarios, Giroto et al. (2019) showed that when the updates introduced by the different observations are anti-correlated, they can end up canceling each other.

Other comments:

- Section 2.1: here authors acknowledge groundwater irrigation and thus point 5 above can be discussed.

Reply: indeed, as the reviewer suggests, there is an important presence of irrigation in some parts of the Brahmaputra basin.

- Daily GRACE products are generated using a KF, which introduces a strong autocorrelation in time. How would this impact the quality of assimilation?

Reply: We acknowledge the reviewer's point. The ITSG-Grace2018 daily TWS solution is computed using a Kalman smoother, which can be expected to introduce temporal autocorrelations in the daily to weekly timescale. In the present study, these temporal correlations were not accounted for, for two main reasons. First, our analysis focuses on monthly, seasonal, and inter-annual water storage variability. At these longer timescales, the temporal autocorrelation present in the daily ITSG-18 product is expected to be minimal, or at least not greater than that inherent in standard monthly GRACE TWS products. Therefore, explicitly accounting for these correlations is not essential for the objectives of this study (see, e.g., Zhao and Yang, 2018). Second, no quantitative estimate of the magnitude or temporal structure of these autocorrelations has yet been provided. Deriving such an estimate would require a dedicated methodological effort beyond the scope of this work.

For future DA experiments aimed at improving sub-monthly water storage dynamics, incorporating these temporal correlations could indeed be beneficial. However, further research is needed to characterize the temporal error structure of the ITSG-18 daily product. Moreover, the inclusion of temporally correlated observation errors in DA systems would require more sophisticated assimilation approaches.

To improve clarity, we have added a few lines to the manuscript disclosing this aspect of our experimental design in the Data and Model section (line 207-213).

It should be mentioned that the Kalman Smoother approach used to generate the daily TWS product can be expected to introduce temporal autocorrelations in the daily to weekly timescale. In the present study, these temporal correlations were not accounted for because (i) the temporal autocorrelation is expected to be weaker in the seasonal to inter-annual timescale considered in this study; and (ii) the magnitude or temporal structure of these autocorrelations have not been quantified yet, and computing them would require a whole dedicated study. Accounting for temporal autocorrelations could be relevant for future studies exploring water storage estimates in weekly to sub-weekly timescales.

Additionally, some reflections have been added on the implications of this study design aspect on the uncertainty estimates of the DA products (Results section, lines 796-802):

The DA process provides uncertainty values for model estimates, which are computed based on ensemble statistics. [...]. However, these values are likely underestimated, since the uncertainty of daily GRACE TWS used in this study was approximate due to the absence of exact estimates, and possible temporal autocorrelation in observations were not accounted for (Section 2.2). Future studies quantifying the magnitude and structure of sub-monthly TWS errors will enable more robust DA frameworks that are able to represent model variable uncertainties with increased realism.

- Line 170: Well data should also be checked for errors (repeating values, negative values, large gaps, see Kuruva et al., 2025). In South-east Asia, the dataset quality check is necessary sometimes. If that has been done, please mention.

Reply: we appreciate the reviewer's comment. The groundwater level data used to validate the results have been quality controlled by the collecting institution (the Bangladesh Water Development Board) and curated by Shamsudduha et al. (2022). [A sentence has been added in the "Data and Model" section of the manuscript to clarify this \(line 225\).](#)

- Since the covariance localization approach uses damped-spatial correlation based on distance b/w observations and model, do we expect observations to be at a higher spatial resolution than model? If yes, then it must be clarified how a coarse resolution GRACE product was used? Same values for all grids in a sub-basin?

Reply: We thank the reviewer for this comment. In all of the experiments, the TWS observations were assimilated as sub-basin averaged fields, and ensemble-statistics were used to disaggregate the spatially averaged observations into the model grid cells.

[We acknowledge that this aspect is not clearly explained in the manuscript. An explanation has been added in lines 427-429.](#)

Spatially, TWS was assimilated as a sub-basin averaged field, and SSM was assimilated as an average over the $0.25^\circ \times 0.25^\circ$ ESA CCI tiles (Fig. 1). The observations were downscaled to the model resolution within the DA process through ensemble statistics.

- L 265: Smooth distant, should be replaced by smooth distance .

Reply: [Thanks, has been applied.](#)

- Line 299 and 303: use comma, not full stop (33.808 -->33,808)

Reply: [Thank you for the comment, it has been applied.](#)

- Result section was slightly harder to follow. Maybe improve the writing aspect.

Reply: [Thanks for this comment. We have tried to revise the writing of the results section to make the results easier to follow.](#)

Best,

Bramha

Reply: We appreciate the useful and specific comments of the reviewer, and we believe they have contributed to improving the manuscript.

References – Reviewer 1

- Cammalleri, C., Barbosa, P., Vogt, J.V., 2019. Analysing the Relationship between Multiple-Timescale SPI and GRACE Terrestrial Water Storage in the Framework of Drought Monitoring. *Water* 11. <https://doi.org/10.3390/w11081672>
- Giroto, M., De Lannoy, G.J.M., Reichle, R.H., Rodell, M., Draper, C., Bhanja, S.N., Mukherjee, A., 2017. Benefits and pitfalls of GRACE data assimilation: A case study of terrestrial water storage depletion in India. *Geophysical Research Letters* 44, 4107–4115. <https://doi.org/10.1002/2017GL072994>
- Giroto, M., Reichle, R.H., Rodell, M., Liu, Q., Mahanama, S., De Lannoy, G.J.M., 2019. Multi-sensor assimilation of SMOS brightness temperature and GRACE terrestrial water storage observations for soil moisture and shallow groundwater estimation. *Remote Sensing of Environment* 227, 12–27. <https://doi.org/10.1016/j.rse.2019.04.001>
- Goswami, S., Rajendra Ternikar, C., Kandala, R., Pillai, N.S., Kumar Yadav, V., Abhishek, Joseph, J., Ghosh, S., Dutt Vishwakarma, B., 2024. Water budget-based evapotranspiration product captures natural and human-caused variability. *Environ. Res. Lett.* 19, 094034. <https://doi.org/10.1088/1748-9326/ad63bd>
- Khaki, M., Hendricks Franssen, H.-J., Han, S.C., 2020. Multi-mission satellite remote sensing data for improving land hydrological models via data assimilation. *Sci Rep* 10, 18791. <https://doi.org/10.1038/s41598-020-75710-5>
- Moron, V., Robertson, A.W., Ward, M.N., Camberlin, P., 2007. Spatial Coherence of Tropical Rainfall at the Regional Scale. *Journal of Climate* 20, 5244–5263. <https://doi.org/10.1175/2007JCLI1623.1>
- 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>
- Satish Kumar, K., Venkata Rathnam, E., Sridhar, V., 2021. Tracking seasonal and monthly drought with GRACE-based terrestrial water storage assessments over major river basins in South India. *Science of The Total Environment* 763, 142994. <https://doi.org/10.1016/j.scitotenv.2020.142994>
- Shamsudduha, M., Taylor, R.G., Longuevergne, L., 2012. Monitoring groundwater storage changes in the highly seasonal humid tropics: Validation of GRACE measurements in the Bengal Basin. *Water Resources Research* 48. <https://doi.org/10.1029/2011WR010993>
- Tian, S., Tregoning, P., Renzullo, L.J., van Dijk, A.I.J.M., Walker, J.P., Pauwels, V.R.N., Allgeyer, S., 2017. Improved water balance component estimates through joint assimilation of GRACE water storage and SMOS soil moisture retrievals. *Water Resources Research* 53, 1820–1840. <https://doi.org/10.1002/2016WR019641>
- Wongchuig, S., Paiva, R., Siqueira, V., Papa, F., Fleischmann, A., Biancamaria, S., Paris, A., Parrens, M., Al Bitar, A., 2024. Multi-Satellite Data Assimilation for Large-Scale Hydrological-Hydrodynamic Prediction: Proof of Concept in the Amazon Basin. *Water Resources Research* 60, e2024WR037155. <https://doi.org/10.1029/2024WR037155>

Zhang, Y., Yu, Z., Niu, H., 2018. Standardized Precipitation Evapotranspiration Index is highly correlated with total water storage over China under future climate scenarios. *Atmospheric Environment* 194, 123–133. <https://doi.org/10.1016/j.atmosenv.2018.09.028>

References – Reviewer 2

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., Giroto, 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>

References – Reviewer 3

Giroto, M., Reichle, R.H., Rodell, M., Liu, Q., Mahanama, S., De Lannoy, G.J.M., 2019. Multi-sensor assimilation of SMOS brightness temperature and GRACE terrestrial water storage observations for soil moisture and shallow groundwater estimation. *Remote Sensing of Environment* 227, 12–27. <https://doi.org/10.1016/j.rse.2019.04.001>

Hamill, T.M., Whitaker, J.S., Snyder, C., 2001. Distance-Dependent Filtering of Background Error Covariance Estimates in an Ensemble Kalman Filter. *Monthly Weather Review* 129, 2776–2790. [https://doi.org/10.1175/1520-0493\(2001\)129%253C2776:DDFOBE%253E2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129%253C2776:DDFOBE%253E2.0.CO;2)

Houtekamer, P.L., Mitchell, H.L., 1998. Data Assimilation Using an Ensemble Kalman Filter Technique. *Monthly Weather Review* 126, 796–811. [https://doi.org/10.1175/1520-0493\(1998\)126%253C0796:DAUAEK%253E2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126%253C0796:DAUAEK%253E2.0.CO;2)

Khaki, M., Hoteit, I., Kuhn, M., Forootan, E., Awange, J., 2019. Assessing data assimilation frameworks for using multi-mission satellite products in a hydrological context. *Science*

of The Total Environment 647, 1031–1043.
<https://doi.org/10.1016/j.scitotenv.2018.08.032>

Khaki, M., Hendricks Franssen, H.-J., Han, S.C., 2020. Multi-mission satellite remote sensing data for improving land hydrological models via data assimilation. *Sci Rep* 10, 18791. <https://doi.org/10.1038/s41598-020-75710-5>

Mehrnegar, N., Jones, O., Singer, M.B., Schumacher, M., Bates, P., Forootan, E., 2020. Comparing global hydrological models and combining them with GRACE by dynamic model data averaging (DMDA). *Advances in Water Resources* 138, 103528. <https://doi.org/10.1016/j.advwatres.2020.103528>

Qu, W., Zhang, P., Chen, P., Li, J., Gao, Y., 2024. Spatiotemporal Variations and Sustainability Characteristics of Groundwater Storage in North China from 2002 to 2022 Revealed by GRACE/GRACE Follow-On and Multiple Hydrologic Data. *Remote Sensing* 16. <https://doi.org/10.3390/rs16071176>

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>

Rodell, M., Chen, J., Kato, H., Famiglietti, J.S., Nigro, J., Wilson, C.R., 2007. Estimating groundwater storage changes in the Mississippi River basin (USA) using GRACE. *Hydrogeol J* 15, 159–166. <https://doi.org/10.1007/s10040-006-0103-7>

Scanlon, B.R., Zhang, Z., Save, H., Sun, A.Y., Müller Schmied, H., van Beek, L.P.H., Wiese, D.N., Wada, Y., Long, D., Reedy, R.C., Longuevergne, L., Döll, P., Bierkens, M.F.P., 2018. Global models underestimate large decadal declining and rising water storage trends relative to GRACE satellite data. *Proceedings of the National Academy of Sciences* 115, E1080–E1089. <https://doi.org/10.1073/pnas.1704665115>

Shamsudduha, M., Taylor, R.G., Ahmed, K.M., Zahid, A., 2011. The impact of intensive groundwater abstraction on recharge to a shallow regional aquifer system: evidence from Bangladesh. *Hydrogeol J* 19, 901–916. <https://doi.org/10.1007/s10040-011-0723-4>

Shamsudduha, M., Taylor, R.G., Haq, M.I., Nowreen, S., Zahid, A., Ahmed, K.M.U., 2022. The Bengal Water Machine: Quantified freshwater capture in Bangladesh. *Science* 377, 1315–1319. <https://doi.org/10.1126/science.abm4730>

Shamsudduha, M., Taylor, R.G., Longuevergne, L., 2012. Monitoring groundwater storage changes in the highly seasonal humid tropics: Validation of GRACE measurements in the Bengal Basin. *Water Resources Research* 48. <https://doi.org/10.1029/2011WR010993>

Soltani, S.S., Ataie-Ashtiani, B., Al Bitar, A., Simmons, Craig.T., Younes, A., Fahs, M., 2024. Assimilating multivariate remote sensing data into a fully coupled subsurface-land

surface hydrological model. *Journal of Hydrology* 641, 131812. <https://doi.org/10.1016/j.jhydrol.2024.131812>

Sun, A.Y., Green, R., Rodell, M., Swenson, S., 2010. Inferring aquifer storage parameters using satellite and in situ measurements: Estimation under uncertainty. *Geophysical Research Letters* 37. <https://doi.org/10.1029/2010GL043231>

Tangdamrongsub, N., Han, S.-C., Yeo, I.-Y., Dong, J., Steele-Dunne, S.C., Willgoose, G., Walker, J.P., 2020. Multivariate data assimilation of GRACE, SMOS, SMAP measurements for improved regional soil moisture and groundwater storage estimates. *Advances in Water Resources* 135, 103477. <https://doi.org/10.1016/j.advwatres.2019.103477>

Tian, S., Tregoning, P., Renzullo, L.J., van Dijk, A.I.J.M., Walker, J.P., Pauwels, V.R.N., Allgeyer, S., 2017. Improved water balance component estimates through joint assimilation of GRACE water storage and SMOS soil moisture retrievals. *Water Resources Research* 53, 1820–1840. <https://doi.org/10.1002/2016WR019641>