

Responses to Comments of Reviewer #1

Comments:

1. I believe authors should enhance the readability of the text, improve the length of the paragraphs, and elevate the quality of the figures. Additionally, the authors could conduct another experiment using all available satellite observations to fully explore the potential of remote sensing data in data assimilation.

(a) Regarding readability and paragraph structure, we have revised some sections of the manuscript for clarity and flow (Introduction, results and discussions sections).

(b) Regarding figure quality, we have ensured that all figures are clear and readable, and that they have clear labels, legends and axes. The quality of Figure 1 depends on the source material from Collischonn et al. (2007) and Decharmes et al. (2019). Figures 2, 3 and 15 have sufficient resolution to present their key information. Figures 4 to 6 were previously enhanced in response to minor comment 21 from the first review. We chose not to include the (DA–OL) metric values in Figures 5 and 6 to maintain readability. Instead, we provide a supplementary table with detailed statistics (median, minimum, maximum, first and third quartiles) for each experiment in both basins (see response to comment #9). Figures 7 through 14 were enhanced where necessary: Figures 8 and 13 were edited again, Figures 7, 9, 10 and 12 improved, and Figures 11 and 14 had already been revised in the previous round.

(c) Concerning the suggestion to conduct an additional assimilation experiment using all available satellite observations simultaneously, we appreciate the reviewer's interest in this direction. However, as discussed in our response to Major Comment 4 during the first round of review, the primary objective of this study was to clearly isolate and quantify the standalone impact of each long-term ESA CCI product (Qalti, Qmultispec, and WSE anomalies) on large-scale hydrological simulations, across contrasting models and basins.

Simultaneous assimilation of multiple datasets would confound the attribution of improvements to specific products, making it difficult to determine their individual contribution and suitability for climate-oriented hydrological modeling. Such a design would also require a different methodological setup, including the tailored treatment of observation uncertainties, tuning of cross-variable error covariances, and possibly more advanced filtering techniques.

Our aim in this first-phase study was not to explore the full potential of remote sensing data in general, but to assess the interpretability and relevance of each CCI product independently, in the context of climate-oriented hydrological modeling.

We fully agree that multi-product assimilation is a promising direction and a logical next step. However, it goes beyond the scope of this first-phase study, which was already considered overly too long during the first review (35 pages + 9 pages of supplementary material). Adding such an experiment would significantly increase the complexity of the analysis and risk diluting the main message.

We do acknowledge the limitations of our separate-product approach in the Discussion section and we have added a perspective on multi-source assimilation in the Conclusion following discussions on first round review. Such developments are already planned as part of the second phase of our project over the

coming year, which will include updated CCI datasets and explore the joint assimilation of more diverse satellite-derived variables.

If any conclusions appear overstated in light of our current experiment design, we would be happy to reconsider or rephrase them to better reflect the scope and limits of the present study. If, however, the reviewer considers that the study in its current scope does not sufficiently contribute to the scientific community, we fully respect that judgment.

2. The author should improve the abstract by adding more numerical data about the results.

We thank the reviewer for this suggestion. Numerical results have now been added to the abstract to better reflect the quantitative improvements obtained through the data assimilation experiments.

Below the updated abstract :

“Large scale hydrological models like CTRIP and MGB are essential for simulating river dynamics and supporting large-scale climate studies. Their accuracy can be significantly improved through satellite data assimilation. This study leverages 20 years of high-resolution discharge data (2000–2020) from the ESA Climate Change Initiative (CCI) to enhance CTRIP and MGB models via ensemble Kalman Filter frameworks (HyDAS and HYFAA). Applied to the Niger and Congo basins, the models assimilate discharge data derived from altimetry and multispectral imagery, alongside water surface elevation (WSE) anomaly data, to evaluate their impact on model performance.

In the Niger basin, discharge assimilation outperformed WSE anomaly assimilation by providing a more direct and physically consistent correction. In MGB, assimilating altimetry-derived discharge led to the strongest improvement, increasing the median Nash-Sutcliffe Efficiency (NSE) to 0.83 and the correlation coefficient to 0.94. WSE anomaly assimilation was also beneficial, particularly when the baseline simulation was poor and when observed and simulated rating curves were well aligned.

Temporal data density in discharge data assimilation emerged as a key driver of performance gains. Assimilating high-frequency discharge data from multispectral imagery significantly reduced bias, from 1.2 to near 1 in MGB, and from 2.3 to 1.75 in CTRIP (median values), which can support hydrological assessments related to long-term variability. Furthermore, the higher temporal resolution allowed for better capture of flow variability, with Kling-Gupta Efficiency γ approaching 1.0 in MGB, which is crucial for both seasonal climate studies and short-term predictions, such as extreme hydrological events.

The study also emphasizes the trade-offs between data resolution and quality, particularly evident in the Congo Basin where improvements were more modest. Future advancements include merging altimetry and multispectral discharge data, improving discharge retrieval algorithms using SWOT data, and refining data assimilation techniques to support climate studies and river system modeling in complex, climate-impacted basins.”

3. L103: I believe the first key question has already been answered by many previous studies (e.g., Paiva et al., 2013, Michailovsky et al., 2013, Emery et al., 2020, Clark et al., 2008).

We thank the reviewer for the comment. However, we believe there may be a misunderstanding regarding the scope and intention behind our first research question.

Our study is not aimed at exploring the general potential of WSE versus discharge assimilation across all available satellite data, an area indeed addressed by the earlier studies mentioned, but is specifically focused on the long-term ESA Climate Change Initiative (CCI) discharge and WSE products. These

long-term datasets are unique as they are derived from a harmonized multi-mission altimetry record (for both WSE and one of the Q products) and from multispectral imagery (for the other Q product), and they offer consistent daily time series over two decades at shared virtual stations. Our experimental setup was built to assess their performance in large-scale hydrological data assimilation (DA), in a climate-study context, and not to explore the general potential of WSE vs. discharge assimilation across all available satellite data.

Within this defined framework, the first research question aims to compare the assimilation of altimetry-derived discharge and WSE data under strictly equivalent spatial and temporal conditions : same stations, same period, same resolution. This is a key distinction that, to our knowledge, has not been done in previous studies.

Indeed, while earlier studies (e.g., Clark et al., 2008; Paiva et al., 2013; Michailovsky et al., 2013; Emery et al., 2018) have explored assimilation of WSE or discharge data, they differ significantly in scope and design. Clark et al. (2008) and Paiva et al. (2013) focused on in-situ discharge assimilation, while Michailovsky et al. (2013) looked at satellite-based WSE only. Emery et al. (2018) did assimilate an altimetry-derived discharge product, but did not compare it to WSE in the same framework. Among these, Paiva et al. (2013) is the only one to attempt both, but their comparison involved different spatio-temporal characteristics and came from different data sources and time periods (in situ discharge vs satellite-based WSE).

Our study was specifically designed to fill this gap. We set up a controlled, side-by-side comparison of satellite-derived discharge and WSE assimilation, under strictly identical spatial coverage, time period, and temporal resolution. We believe this fills a gap in the literature by isolating the effect of the observation type itself, rather than differences in data availability or density, on model performance. The comparison is further reinforced by using two different hydrological models (CTRIP and MGB) and basins (Niger, Congo). Given the climate focus of the CCI program, the use of over 20 years of consistent satellite observations is a central feature of this study, and we now emphasize this aspect more clearly in the revised introduction (Lines 95–116, see response to comment #4).

To avoid confusion, we have kept the research question in its original form but we have ensured that the product-specific scope is now clearly stated in the paragraph immediately preceding the questions, so that they are read and interpreted within the appropriate context.

4. L105: To access the second key question, the authors should evaluate all satellite altimetry locations in WSE assimilation.

As stated in response to comment #3, there may have been a misunderstanding regarding the scope of our research questions, and particularly the second one.

Our study does not aim to evaluate the full potential of altimetry-based WSE assimilation using all available satellite tracks. It focuses on how the characteristics and limitations of the specific ESA CCI products, particularly their spatial and temporal resolution, uncertainty, and derivation method, affect data assimilation outcomes in large-scale hydrological models.

The second research question specifically addresses the trade-offs between different types of CCI discharge products: one derived from altimetry, the other from multispectral imagery. These datasets differ significantly in spatial and temporal resolution, spatial coverage, and uncertainty, all of which come from their distinct derivation methods. The goal is to understand how these differences influence the

effectiveness of assimilation, within the specific context of using long-term satellite products for hydrological reanalyses in climate-related applications.

On the other hand, both the CCI WSE and altimetry-based discharge products are derived from the same harmonized multi-mission altimetry record. They share the same spatio-temporal resolution and coverage, and are therefore directly compared in the first research question, which focuses on how the nature of the variable (WSE vs. Q), and its level of uncertainty, impacts assimilation outcomes when all other conditions are held equal.

The experimental setup is therefore designed to reflect the structure and constraints of the CCI products: daily, multi-decadal records at fixed virtual stations. This allowed us to conduct a controlled comparison and offer guidance on how to use these products in assimilation frameworks. Evaluating all possible altimetry-based WSE locations would constitute a different study, aimed at maximizing observational density and DA performance, and falls outside the scope of our current work. That said, we agree that such an investigation would be of interest, and it is indeed planned as part of the next phase of our project.

To better reflect this scope, we have revised lines 95–116 of the manuscript. The revised text now reads:

“In the context of data assimilation, both water surface elevation (WSE) and discharge observations present distinct advantages and limitations. WSE is directly measurable from altimetry and typically less uncertain, but is often misaligned with the simplified river geometry used in hydrological models (which generally rather consider a conceptual river depth). In contrast, discharge estimates align closely with model outputs but are indirectly derived—either from altimetry or multispectral imagery—introducing additional uncertainties through empirical rating curves and processing chains (Gal et al., 2024; ESA CCI ATBD, 2023).

In this study, we focus specifically on the long-term satellite-derived ESA CCI WSE and discharge products, which are designed for climate applications and offer consistent, multi-decadal daily time series over a shared network of virtual stations. By comparing these products in data assimilation experiments over the Niger and Congo basins, we aim to examine their respective potential and limitations in improving large-scale hydrological simulations through data assimilation.

Within this defined framework, we seek to answer three key questions:

- 1. Does the assimilation of higher-uncertainty discharge data improve model performance more than the assimilation of lower-uncertainty WSE data?*
- 2. How do trade-offs between spatial coverage, temporal resolution, and the quality of these long-term satellite-derived products impact the effectiveness of data assimilation for large-scale hydrological modeling?*
- 3. How do these impacts vary between a calibrated conceptual model (MGB) and a physically-based, uncalibrated routing model (CTRIP), and between basins with differing hydrological characteristics (Congo vs. Niger)?*

This analysis aims to provide insights into the added-value of long-term, satellite-based data assimilation for improving model accuracy and variability representation, particularly in the context of developing reliable hydrological reanalyses for climate-related applications. While this study does not explicitly

address climate attribution or trend detection, it contributes to climate-oriented hydrology by evaluating the reliability and value of these products for multi-decadal model improvement.“

5. L107: What do you mean by “levels of calibration”? Does this refer to the DA method? If so, it would be better to use clearer wording for that.

We agree that the phrasing of the third question could be misleading. The intention was not to refer to differences in the DA method, but to the calibration approach of the hydrological models themselves. MGB is a conceptual model that is explicitly calibrated at the basin level, while CTRIP is a physically-based global routing model without any calibration, since it relies on global datasets and physical equations.

To improve clarity, we have rephrased the third question as :

“How do these impacts vary between a calibrated conceptual model (MGB) and a physically-based, uncalibrated routing model (CTRIP), and between basins with differing hydrological characteristics (Congo vs. Niger)?”

6. L327: Does “free run of the model.” refer to open-loop?

Yes indeed, we are referring to an open-loop simulation, but not to an ensemble run. It is a single-member reference simulation run over 20 years. We used the term free run to describe this long historical model simulation, used exclusively to compute background spatial correlations for the localization step.

We have clarified this point in the manuscript by rephrasing the sentence as follows :

“...based on a long-term (20-year) reference simulation, used to estimate spatial correlations.”

7. Why don't the authors investigate SWOT like discharge, where discharge can be derived everywhere in the river network?

We agree that SWOT like discharge is highly promising and we mention the potential of SWOT mission in the Perspectives section. However, using the SWOT derived discharge falls outside the scope of this study.

Our aim here was to evaluate the standalone impact of the ESA CCI products, 20-year daily WSE and discharge datasets derived from altimetry and multispectral imagery, under a controlled setup using long, consistent time series at fixed virtual stations where these products are available. SWOT like discharge is a different type of product and it's not available as a long-term dataset.

Bringing it into the current framework would require a completely different experiment setup, and would blur the main message of the study, which is to guide the use of current CCI products for climate-oriented hydrological modeling. This is we didn't include SWOT-like discharge here, though we clearly see its value and plan to explore it in future work.

8. Why does CTRIP-HyDAS perform so poorly, while Emery et al. (2018) and Emery et al. (2020) show much better assimilation in the Amazon basin?

We agree that the performances of the CTRIP model over the two selected basins look quite poor, especially when compared to those obtained over the Amazon basin as cited by the reviewer. There are different reasons.

First, the hydrological dynamics is particularly difficult to capture by global scale river routing models, namely because of processes specific to these basins and generally neglected (such as evaporation over the Niger Inner Delta and the Congo's Cuvette Centrale). On the contrary, CTRIP usually captures quite well the dynamics in the Amazon basin.

Second, the model configuration is a bit different in our study, since we use the version developed by Munier and Decharme (2022) at 1/12° resolution, with a slightly different method to compute the model parameters.

Third, as global scale river routing models are generally developed in the context of Earth system models or climate models, they have to simulate river discharge over the whole continents, even in regions without any observation usable for calibration purposes. Although some of such models calibrate their parameters (e.g. roughness coefficient) wherever observations exist, this is not the case for the CTRIP model for which it is preferred to remain independent from river gauge observations, then being able to transpose the performances from gauged regions to ungauged ones. Besides, the performances of CTRIP over the Niger and the Congo basins are comparable to those obtained over other basins (see e.g. Fig. 4 of Decharme and Colin, 2024) or from similar other global river routing models.

Fourth, a key difference between this study and those by Emery et al. lies in the spatial sampling and data assimilation strategy. Emery et al. assimilated a dense network of virtual altimetric stations in the Amazon using multi-mission tracks from Jason and Envisat and derived discharge using model-calibrated rating curves (based on the MGB model). By contrast, our study uses ESA CCI discharge and WSE products, which are more sparsely distributed but provide consistent daily time series over a period of more than 20 years. This longer temporal record aligns with the CCI programme's objective of prioritising long-term, homogeneous datasets suitable for climate applications.

Finally, it is precisely over poorly simulated basins, such as those studied here, that the model could benefit the most from data assimilation, particularly for long term reanalyses for climate studies. Our results show how CCI-derived observations can help correct structural model biases and improve simulations in basins that have not yet been extensively explored in global modelling initiatives.

Decharme, B., & Colin, J. (2024). Influence of Floodplains and Groundwater Dynamics on the Present-Day Climate simulated by the CNRM Model. *EGUsphere*, 2024, 1-33.

9. Figure 5: The authors can provide a relative metric compared to the open-loop. They can determine whether that enhances discharge performance.

We thank the reviewer for the suggestion. To address this suggestion while maintaining the readability of the figures, we propose not adding the annotations directly to Figures 5 and 6, as this would make them visually too dense. Instead, we have provided a supplementary table presenting detailed relative performance statistics for each DA experiment in both basins. These statistics include the median, minimum, maximum, and the first and third quartiles of the difference in each metric ($\Delta = \text{DA} - \text{Open-Loop}$), computed across all stations in both basins.

10. Figure 7: The hatched effect observed in the assimilated values may be related to the number of observations in the local patch. If there were more observations within the local patch, the hatch effect

could be diminished. The same effect can be reproduced if the authors used a zero patch (Revel et al., 2021). To test this more effectively, the authors should incorporate additional satellite altimetry data within each local patch. In the current setup, I assume there is only one observation per local patch area, as shown in the Ansongo and Malanville figures. I am also not entirely convinced by the response to Major Comment #6.

We thank the reviewer for their valuable comment. The hashed effect observed after assimilation results from multiple interacting factors, and we agree that our previous response to Major Comment #6 may have been too focused on the role of rating curve mismatch alone. We now provide a more complete explanation of this effect.

The hashed pattern does not arise only when assimilating WSE; it can also appear when assimilating discharge.

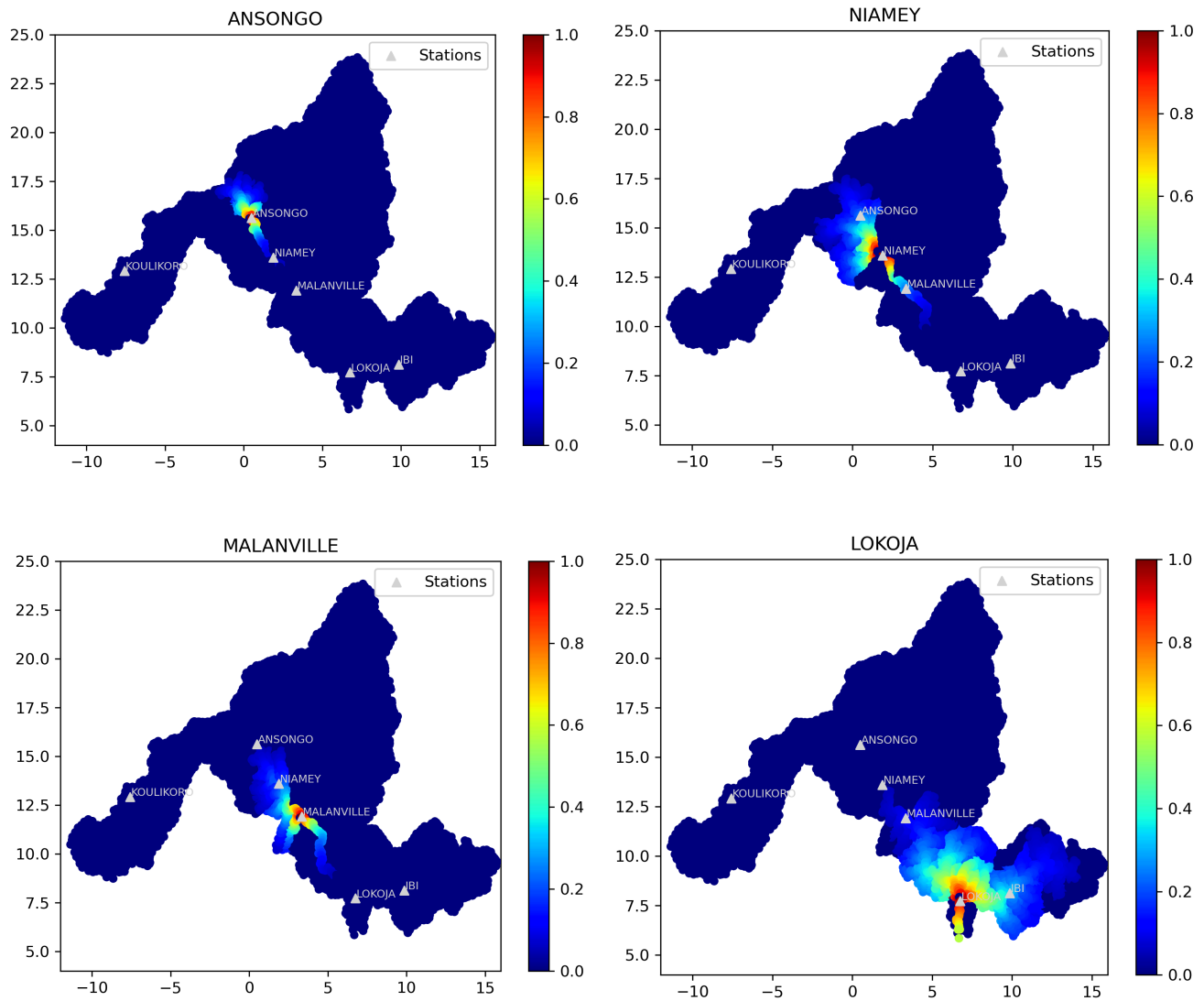
The hashed pattern primarily arises because the assimilation correction is not continuously sustained between observation times. When observations are temporally sparse, the model tends to revert toward its open-loop trajectory between updates. Several mechanisms can help reduce this hashed effect: Increasing the temporal density of observations, which limits the divergence between model and observations between updates; Increasing the spatial density of observations, but this is mainly effective if it also increases temporal coverage (the added observations do not fall on the same dates as existing ones); Correcting model forcings, which can be particularly effective when the issue is a simple bias, for example by applying a multiplicative correction factor as in Munier et al. (2015); Assimilating more upstream observations, which can improve runoff and inflow inputs and smooth downstream behavior.

However, the hashed effect is amplified when assimilating WSE at stations where there is a strong mismatch between observed and simulated rating curves. In such cases, the hashed effect can also stem from the conversion of assimilated observed WSE into the model discharge state when the model does not follow the same rating curve, which can create larger jumps in discharge after each assimilation step.

The hashed effect is influenced by the innovation term (observation - model prior state) which is linked to the assigned uncertainties, to the type of variable assimilated (since WSE and Q have different uncertainty levels) and to each model open-loop position, all of which affect the Kalman gain. It is also affected by how observations along the local patch help prolong the effect of the correction. Using localization allows the model state upstream of the observed pixel to be corrected, and this correction then propagates downstream in the following time steps, which can indirectly increase the persistence of the assimilation effect.

We acknowledge the relevance of Revel et al. (2021), which demonstrates the benefit of merging altimetry tracks to improve spatial consistency in assimilation. While this study does not specifically analyze hashed patterns, it supports the broader conclusion that denser observation networks improve robustness.

In our specific case, the assumption that each local patch contains only one observation is not valid for both basins. While this is true for the Congo basin, the Niger basin shows more overlaps. For instance, as shown in the figure below, the local patch for Malanville includes Niamey; Lokoja is influenced by both Malanville and Ibi; and Niamey includes Ansongo and Malanville. Table 2 in the manuscript shows that, within the simulation years, several EO observations were available and assimilated simultaneously in neighboring stations. As also visible in the figures, the impact of neighboring stations within a local patch is weaker, which is due to the limited spatial density constrained by the availability of the CCI products.



Localisation for Ansongo, Niamey, Malanville and Lokoja based on adaptive distance (semi-variogram) (Niger Basin)

During method development, as stated in our response to Major Comment #6 in the first round, we tested three localization strategies in the Niger basin using MGB-HYFAA with WSE assimilation: (i) no localization (i.e., zero patch); (ii) constant-distance localization (300 km); and (iii) adaptive localization based on semi-variograms. Among these, adaptive localization improved spatial consistency and visibly reduced hashed effects in discharge compared to the other approaches, despite the limited spatial density of the current WSE observations. However, we observed that hashed patterns persisted in stations where significant mismatches existed between observed and simulated rating curves.

In such cases, the presence of additional assimilated observations within the local patch helped attenuate the amplitude of hashed effects, but this was not sufficient to eliminate them when structural inconsistencies in the model remained—particularly in the conversion between observed WSE and simulated discharge, and this is what we see in Figure 7 in Ansongo.

We agree with the reviewer that future experiments incorporating spatial densification, such as assimilating additional altimetry datasets, as well as exploring smoother assimilation techniques would be highly valuable. These are part of the directions we intend to pursue. However, the current study was

designed to evaluate the standalone impact of long-term water surface elevation (WSE) and discharge products from the ESA CCI. These products have their own limitations in terms of spatial coverage, and their evaluation is based on a specific configuration involving 20 continuous years of data available for assimilation at continuous high temporal resolution. While adding non-overlapping altimetry missions could help mitigate hashed effects during overlapping periods, such an approach would not resolve gaps across the full 20-year time series.

To better align the manuscript with our response to this comment and improve clarity regarding the hashed effects and their causes, we have revised the subsection 4.2.2 of the Results section. Below are the proposed modifications:

(a) We clarified the origin of the hashed effect (linked to Figure 7) by replacing the original paragraphs :

“As shown in Figure 7, at Ansongo and Malanville, discharge assimilation produces smoother results, while WSE introduces noise due to rating curve discrepancies between MGB and CCI observations. The hashed effect seen with dH assimilation is evident in several stations but absent in others like Ibi and Lokoja. This absence is due to the combined effect of low observation uncertainties and compatible rating curves between the model and observations at these stations.

This discrepancy is particularly evident at Ansongo (Figure 7), where MGB's rating curve diverges significantly from the CCI observations. Figure 8 highlights this finding by comparing observed and simulated rating curves across both Niger Basin stations. This mismatch between the two curves at Ansongo implies additional errors when converting WSE anomalies to discharge, leading to noisy results. In contrast, Qalti, as a directly modelled variable, fits more naturally to the hydrological model.”

With the revised version :

“As shown in Figure 7, at Ansongo and Malanville, discharge assimilation produces smoother results, while WSE assimilation introduces noise, with visible jumps in the simulated discharge after each update. These hashed effects arise when corrections are applied intermittently and the model gradually returns to its prior state between updates. They are especially pronounced when observations are temporally sparse and when there are large discrepancies between observed and simulated values.

The hashed effect is further amplified when assimilating WSE at stations with a strong mismatch between observed and simulated rating curves. At Ansongo, for example, this mismatch (Figure 8) leads to conversion errors when using WSE anomalies to update discharge, amplifying the hashed effect. In contrast, these are much less pronounced at stations like Malanville (Figure 7), Ibi, and Lokoja, where the model and observation rating curves from the model and observation are more consistent (Figure 8, panel b). These stations also benefit from overlapping local patches that include observations from nearby stations (e.g., Malanville includes Niamey; Lokoja includes Malanville and Ibi), which helps sustain corrections between assimilation steps. However, the overall effectiveness of the local patch remains constrained by the limited spatial and temporal coverage of the CCI products.

To provide further context, the Supplementary Material includes an illustration of the local patches for the six stations in the Niger basin (Figure S8), along with rating curve comparisons between MGB (Open-Loop) and CCI observations (dH, Qalti) at Ibi and Lokoja.

Finally, because Qalti is a directly modeled variable, it integrates more consistently into the hydrological model equations. This contributes to the smoother corrections observed across all stations during its

assimilation, as illustrated for example at Ansongo and Malanville (panels 2b and 2d, Figure 7), and similarly seen the rest of the stations.”

(b) We reformulated the sentence describing structural limitations of dH assimilation, by replacing paragraph :

“This mismatch between modeled and observed rating curves, also noted in the Niger Basin (Ansongo, Figure 8), illustrates a broader structural limitation of dH assimilation in MGB when rating curve slopes differ significantly.”

With :

“This mismatch between modeled and observed rating curves, also noted in the Niger Basin (Ansongo, Figure 8), illustrates a broader structural limitation of dH assimilation in MGB : when rating curves slopes differ significantly, converting WSE anomalies to discharge can introduce significant errors.”

We hope these elements adequately address the reviewer’s concerns and help clarify the origin and propagation of hashed effects and instabilities related to WSE assimilation in our modeling framework. To complement the additional information and explanations now included in the manuscript, we have also added supplementary material featuring the local patches for six stations in the Niger basin (Figure S8), along with rating curve comparisons between MGB (Open-Loop) and CCI observations (dH, Qalti) at Ibi and Lokoja.

References :

Munier, S., Polebistki, A., Brown, C., Belaud, G., & Lettenmaier, D. P. (2015). SWOT data assimilation for operational reservoir management on the upper Niger River Basin. *Water Resources Research*, 51(1), 554–575. <https://doi.org/10.1002/2014WR016157>

Revel, M., Ikeshima, D., Yamazaki, D., & Kanae, S. (2021). A framework for estimating global-scale river discharge by assimilating satellite altimetry. *Water Resources Research*, 57, e2020WR027876. <https://doi.org/10.1029/2020WR027876>

11. The effect of the spatial distribution of assimilation stations was obvious in Figure 10, OUESSO gauge. In addition, the authors should have examined the locations where there is no local observation for assimilation.

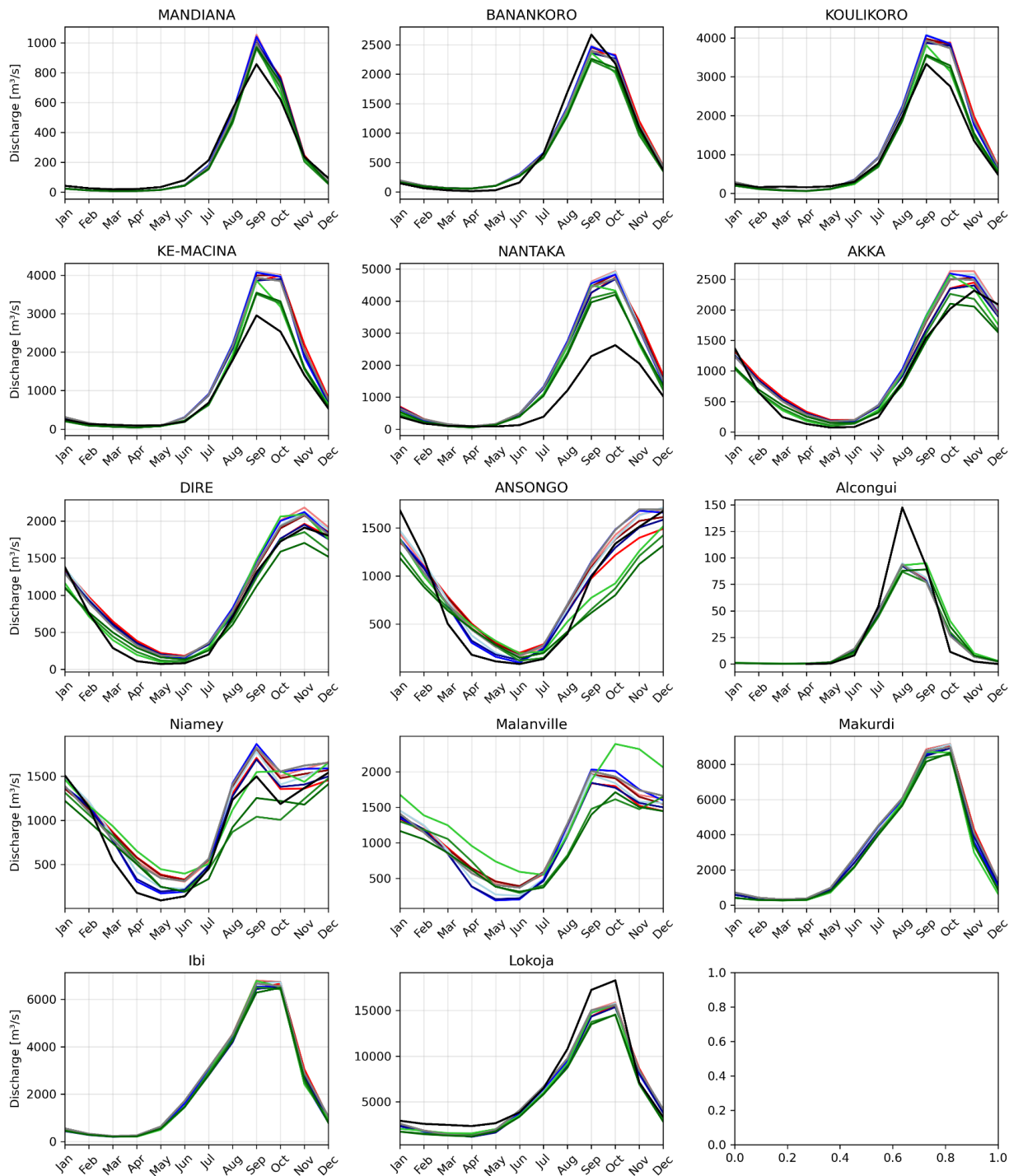
We thank the reviewer for this comment.

In the Congo basin, our ability to analyze locations without local assimilation was limited by the availability of in-situ discharge data. Only three stations (Ouessou, Bangui, and Kinshasa) provide independent observations (Table 2), and assimilation was performed at Chembe-Ferry, Bangui, and Kinshasa. We were therefore unable to assess effects in ungauged regions of the Congo.

In the Niger basin, more in-situ stations were available. However, Mandiana and Banakoro are located upstream of Koulikoro, where assimilation was performed, and were therefore not relevant to include in the main figures. Alcongui lies on an intermittent tributary. Ibi and Makurdi, although located on the main

Benue tributary of the Niger, showed negligible impact from assimilation. They were among the least affected stations, and thus there was no significant result to highlight or discuss further.

Rather than including full time series for these sites, we propose to include a figure from the CCI technical report as Supplementary Material. This figure presents discharge climatologies at all available stations in the Niger basin, sorted from upstream to downstream, and includes both the open-loop and assimilation experiments from MGB-HYFAA. This addition should help clarify the spatial performance of data assimilation, even for stations not analyzed in detail in the manuscript.



Discharge climatologies at all available stations in the Niger Basin, sorted from upstream to downstream: in-situ observations (black), MGB open-loop (grey), and data assimilation experiments (dH in red, Qalti in blue, Qmultispec in green) with varying uncertainty levels represented by different shades following same colour coding (MGB-HYFAA).

We will refer to this figure in the Results section (L485). Finally, as stated in the Perspectives section, future work will aim to improve the spatial distribution of assimilated observations. However, this lies beyond the scope of the current study, which focuses on evaluating the standalone value of ESA CCI products under current observational limitations.

12. L434-445: Has any model been calibrated or corrected?

MGB is a conceptual model and has been explicitly calibrated, as mentioned in line 239. when presenting both models implementations over the Niger and Cong basin (Section 2.2.3). CTRIP follows an uncalibrated, physically based approach.

13. Figure 9: Is this for MGB-HYFAA? necessary to include it in the figure caption?

We thank the reviewer for pointing this out. Yes, Figure 9 corresponds to MGB-HYFAA, and we have updated the caption accordingly to explicitly state this.

The revised caption now reads:

“Comparison of dH and Q assimilation at Ansongo, Niger, over the 2010-2020 period, within MGB-HYFAA: Top panels show time series of (a) dH assimilation with 0.4m uncertainty and (b) Qalti assimilation with 30% uncertainty, with open-loop, analysis, and observed data. The middle and lower panels show the corresponding innovation and residuals, with the mean innovation and residuals represented by dashed lines, illustrating the model’s performance and adjustments after assimilation.”

The same clarification has been applied to Figure 10, which also shows results from MGB-HYFAA. The updated caption reads :

“Comparison of dH (1) and Qalti (2) Assimilation in Ouesso, Bangui, and Kinshasa, within MGB-HYFAA : OL Simulation in black, Assimilation Results in blue and Observed CCI Data in red.”

14. Figure 10: What error happened around 2016-07 in the panel [1]?

We thank the reviewer for highlighting this issue and apologise for the confusion caused. The anomaly visible in Panel 1b for Bangui was caused by a plotting error. We have verified the simulation outputs and we corrected this error in the revised version of the paper.

15. L675: What do the authors mean by “additional satellite observations”? Is it related to more temporal coverage or spatial coverage?

We thank the reviewer for the comment. As mentioned at the beginning of the paragraph and further explained in the Conclusion and Perspectives section (lines 739–753), when we refer to “additional satellite observations,” we are mainly referring to greater spatial coverage achieved by integrating other

WSE datasets from satellite altimetry. These datasets are more spatially dense than the CCI products used in this study. The aim is to improve the localization process, particularly in areas with few or no in-situ stations, and to test multi-sensor assimilation setups. Combining different WSE datasets with discharge products will also be explored, particularly in areas where spatial coverage is limited. Using WSE datasets for their wider spatial density and discharge products for their physical consistency, this approach should help to increase the number of assimilated points and improve the results. We will also explore using slope data from SWOT to improve model structure and assimilation consistency.

We have clarified the sentence in the manuscript as follows (L701-703):

“We plan to explore how these products can be integrated with additional satellite observations, particularly higher-density WSE datasets from different altimetry missions, to further improve model performance.”

16. L692-694: Is this correct for both models?

We thank the reviewer for the question. This sentence refers specifically to CTRIP in the Niger basin, where the open-loop performance was particularly poor due to the lack of evaporation representation in the model structure, which led to a strong overestimation of discharge. In this specific case, the EnKF gave more weight to WSE anomalies (which have lower uncertainty than discharge), leading to more effective corrections. This is not a general statement about CTRIP, but a context-specific observation. MGB, on the other hand, was calibrated and performed better in open-loop, so discharge assimilation remained more effective.

We have clarified this point in the revised text to make the context more explicit :

“However, in poorly performing models with low open-loop skill and well-aligned rating curves, WSE anomalies can become more effective than Q, as their lower observation uncertainty leads to more impactful corrections. This was the case for CTRIP in the Niger basin, where structural limitations led to large discharge biases. ”

17. L12: Two commas

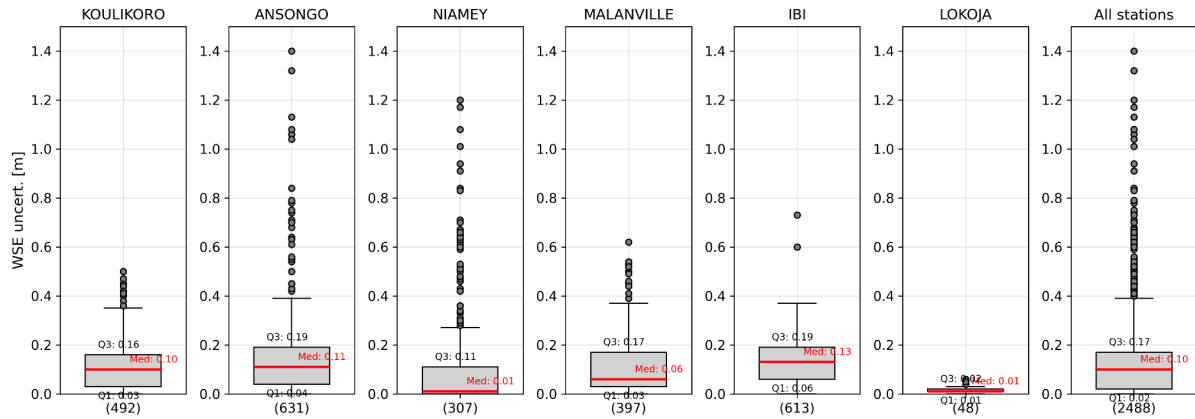
The typo has been fixed, thank you.

18. L413-416: The authors should explain the reasoning behind using these errors.

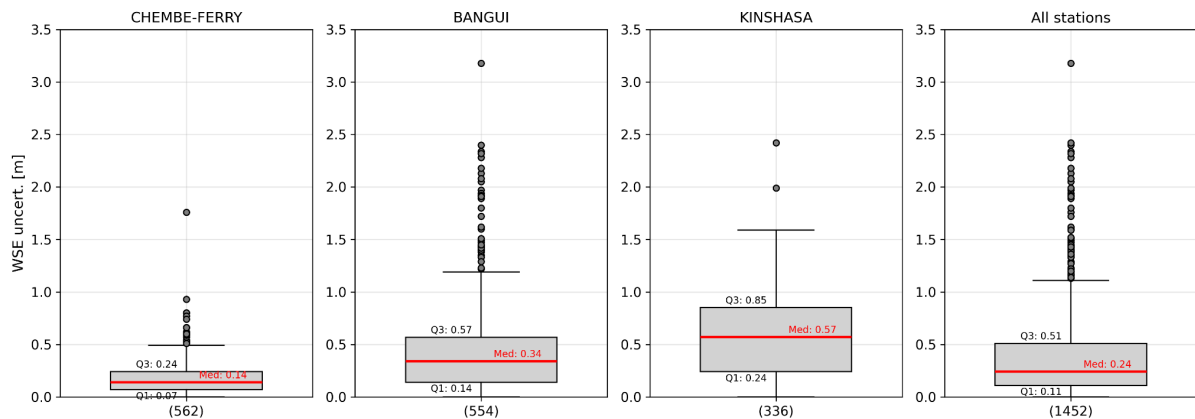
We thank the reviewer for this relevant point. The uncertainty levels adopted in our assimilation experiments were based on a combination of (i) error estimates computed directly from the ESA CCI products for WSE anomalies and Qalti, and (ii) values reported in the literature.

For WSE anomalies (dH), CCI-derived uncertainty distributions across the Niger and Congo basins typically range from 0.01 to 0.85 m. Median values are around 0.10 m at most stations in the Niger basin, reaching 0.24 m in the Congo basin. In Niger, the third quartile (Q3) is mostly below 0.20 m, whereas in the Congo, some stations show heavier-tailed distributions with a Q3 of up to 0.85 m (Kinshasa). To define representative test values for data assimilation, we selected 0.2 m as a baseline level, as this corresponds to the median in the Congo basin and the upper range of the third quartile in the Niger basin. A higher level of 0.4 m was chosen to represent the upper bound of most observed distributions and to account for

noisier or more uncertain locations in both basins. This is twice the baseline level. We also ensured that the chosen values fell within the [0.1–0.5 m] uncertainty range commonly used in previous altimetry-based data assimilation studies (Andreadis et al., 2007; Biancamaria et al., 2011; Paiva et al., 2013, 2015; Munier et al., 2015).



Distribution of WSE errors (in meters) for selected Niger stations (Koulikoro, Ansongo, Niamey, Malanville, Ibi, Lokoja) and aggregated results for all stations.

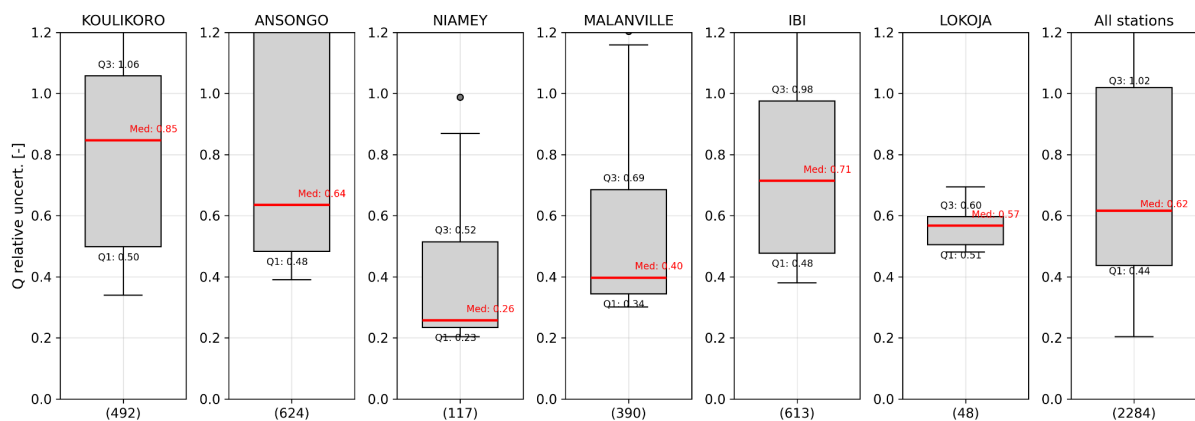


Distribution of WSE errors (in meters) for selected Congo stations (Chembe-Ferry, Bangui, Kinshasa) and aggregated results for all stations.

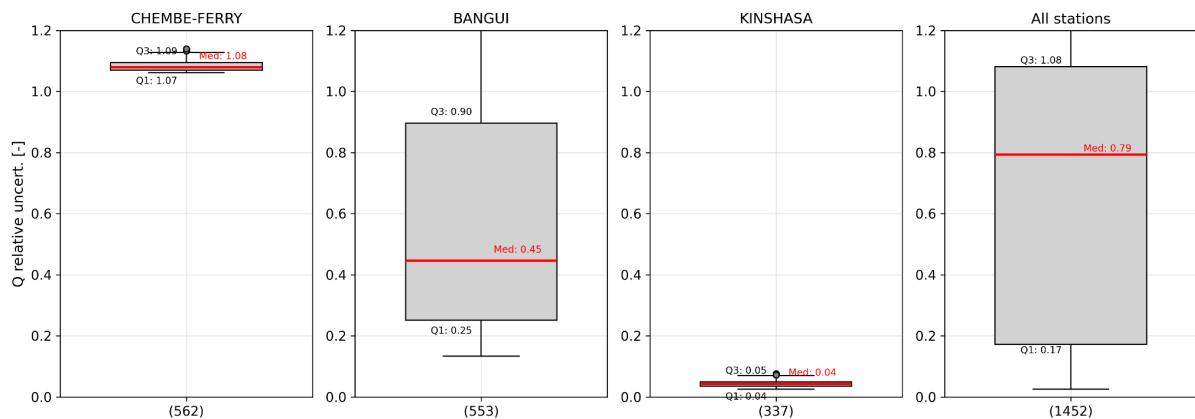
For Qalti, relative uncertainties show a wide spread among stations. In Congo, median values range from 4% (Kinshasa) to over 100% (Chembe-Ferry), and in Niger, they range from 26% (Niamey) to 85% (Koulikoro). At basin-scale, medians are around 62% in Niger and 79% in Congo, although some well-instrumented operational stations (Bangui, Kinshasa, Niamey, and Malanville) show significantly lower uncertainties. To capture this variability without overweighting the model in the filter, we used two levels: 30% (optimistic) and 50% (conservative). We avoided the use of very high uncertainty values (>50%) because it would cause the filter to downweight or even disregard observations. A value of 30% has been adopted in earlier hydrological DA studies (Paiva et al., 2013, 2015), while 50% is consistent medians of station distributions.

For Qmultispec, the current ESA CCI version does not provide station-level uncertainty estimates. We therefore adopted the same levels as for Qalti (30% and 50%). While uncertainties for Qmultispec are likely larger, this approach ensured that comparisons between assimilation experiments were made on the same variable (discharge) with a consistent uncertainty framework.

For both WSE and Q, the choice of values was made to prevent unbalanced weighting in the Kalman filter between model states and observations.



Distribution of relative errors for Qalti across selected Niger stations (Koulikoro, Ansongo, Niamey, Malanville, Ibi, Lokoja) and aggregated results for all stations.



Distribution of relative errors for Qalti across selected Congo stations (Chembe-Ferry, Bangui, Kinshasa) and aggregated results for all stations.

We have clarified the reasoning behind the selection of uncertainty levels by adding a brief paragraph in lines L421–L433 detailing how they were derived :

“The uncertainty levels were selected based on a combination of error distributions derived directly from the ESA CCI products and values commonly adopted in previous hydrological DA studies. For WSE anomalies, the computed uncertainties typically range between 0.01–0.85 m across Niger and Congo stations, with median values around 0.10 m in Niger and 0.24 m in Congo. We therefore retained two representative scenarios (0.2 m and 0.4 m) covering both typical and upper-bound conditions, consistent with uncertainty ranges adopted in previous WSE assimilation studies (0.1–0.5 m) (Andreadis et al., 2007; Biancamaria et al., 2011; Paiva et al., 2013; Paiva et al., 2015; Munier et al., 2015). For Qalti, the relative uncertainties showed a wide spread (4%–100%), with basin-wide medians around 62–79%. However,

several well-instrumented stations (e.g., Bangui, Kinshasa, Niamey, Malanville) exhibit significantly lower values. To capture this variability while maintaining a balanced impact of the observations in the Kalman filter, we selected 30% (optimistic) and 50% (conservative) as representative levels, 30% being in line with values used in previous studies (Paiva et al., 2013; Paiva et al., 2015). Since no station-based uncertainties are currently available for Qmultispec, the same levels as Qalti were adopted for consistency.”

The references have also been integrated.

References

- Andreadis, K. M., Clark, E. A., Lettenmaier, D. P., & Alsdorf, D. E. (2007). Prospects for river discharge and depth estimation through assimilation of swath-altimetry into a raster-based hydrodynamics model. *Geophysical Research Letters*, 34(10). <https://doi.org/10.1029/2007GL029721>
- Biancamaria, S., Andreadis, K. M., Durand, M., Clark, E. A., Rodriguez, E., Mognard, N. M., et al. (2011). Assimilation of virtual wide-swath altimetry to improve Arctic river modeling. *Remote Sensing of Environment*, 115(2), 373–381. <https://doi.org/10.1016/j.rse.2010.09.008>
- Paiva, R. C. D., Collischonn, W., Bonnet, M.-P., de Gonçalves, L. G. G., Calmant, S., Getirana, A., & Santos da Silva, J. (2013). Assimilating in situ and radar altimetry data into a large-scale hydrologic-hydrodynamic model for streamflow forecasts in the Amazon. *Hydrology and Earth System Sciences*, 17, 2929–2946. <https://doi.org/10.5194/hess-17-2929-2013>
- Paiva, R. C. D., Durand, M., & Hossain, F. (2015). Spatiotemporal interpolation of discharge across a river network by using synthetic SWOT satellite data. *Water Resources Research*, 51(1), 430–449. <https://doi.org/10.1002/2014WR015618>
- Munier, S., Polebistki, A., Brown, C., Belaud, G., & Lettenmaier, D. P. (2015). SWOT data assimilation for operational reservoir management on the upper Niger River Basin. *Water Resources Research*, 51(1), 554–575. <https://doi.org/10.1002/2014WR016157>

Responses to Comments of Reviewer #2

Comments:

1. Although the authors argue that the objective of this paper is “rather than comparing model performance, we explore how the same CCI-derived data interact with these different model architectures, spatial scales, and parameterizations”, current analysis presented in the manuscript fails to answer how the different model architectures, spatial scales, parameterizations, and DA method affects the results. For instance, the analysis of open-loop performance fails to answer how the different meteorological forcing, model architectures, parameterizations, and implementation of calibration affect the simulations. Consequently, the subsequent analysis related to the DA performance fails to answer how above uncertainties further affect the DA results. As such, I did not find the benefit to use both CTRIP and MGB models in the study.

We thank the reviewer for this relevant comment and we agree that this phrasing does not reflect the objectives of this study. As stated in the responses to comments #3 and #4 of the first reviewer, we have clarified the scope and objectives of the study and corrected the manuscript accordingly. We have revised this sentence in the manuscript to better reflect the scope of our approach.

Here is a suggested replacement (L86-87):

"Rather than comparing model performance, we explore the robustness, limitations, and potential benefits of using these CCI-derived data products across different modeling contexts with their own assumptions and operational constraints."

We still believe in the relevance of the dual-model approach. Using both CTRIP and MGB allowed us to evaluate the performance of long-term satellite-based data assimilation across two representative modeling contexts. We respectfully believe that including both CTRIP and MGB was critical to the robustness and generalizability of the study's conclusions, and was necessary to uncover how differences in model structure, open-loop performance, process representation, and river geometry shaped data assimilation outcomes. The complementarity of their behaviors allowed us to explore both the conditions where satellite-derived discharge and WSE observations are effective, and the structural limitations that can constrain their assimilation.

CTRIP, a global-scale uncalibrated routing model, showed that WSE assimilation could outperform discharge assimilation. This was not due to WSE being inherently more informative, but because CTRIP's poor open-loop skill after the inner delta in the Niger basin for example caused the filter to favor the lower-uncertainty altimetry signal. CTRIP's open-loop discharge simulations downstream the inner delta had biases in seasonal volume. In stations where the model's simplified geometry didn't conflict with observations (reasonably correct flow-depth relationships), WSE assimilation aggressively corrected the model, improving metrics.

Conversely, MGB, a calibrated regional model, consistently benefited more from discharge assimilation, particularly in stations where model and observed rating curves were misaligned. Assimilating discharge as a directly modeled variable fits more naturally with hydrological models. This enabled observations to be more accurately captured without needing a conversion via a rating curve, thereby avoiding issues tied to simplified geometry.

This contrasting behavior highlights how observation type, model structure, and error characteristics interact to determine assimilation performance. The complementarity between CTRIP and MGB was crucial. Without MGB, the full potential of Q would not have been demonstrated. Without CTRIP, the limitations of Q assimilation and the benefits of dH would not have been visible. This combined approach was essential for identifying when each CCI product is most effective and for highlighting structural factors, such as rating-curve mismatch and model simplifications, that influence their impact.

Furthermore, applying the same observational datasets to both models enabled us to gain a broader understanding of the added value of multispectral CCI discharge products. Despite their differing configurations, the high temporal resolution of Qmultispec systematically improved bias and internal variability in both CTRIP and MGB. The convergence of these results across two different model structures strengthens the robustness and generalisability of our conclusions regarding long-term multispectral-derived discharge products. Additionally, by comparing a high-density, noisy product (Qmultispec) with a lower-frequency, precise product (Qalti) in both models, we were able to highlight the trade-off between data frequency and quality, a nuance that the dual-model analysis was able to generalise.

We believe the manuscript already highlights these differences in model design and their implications for assimilation performance (Sections 2.2 and 4.2.2). Throughout the paper, and especially in the methodology (Sections 2.1–2.2), we emphasized the use of two conceptually distinct models sharing a common DA architecture. This design aimed to ensure that differences in assimilation outcomes could be attributed to model structure, scale, and calibration rather than to differences in DA implementation. The shared principles (localized updates, ensemble-based uncertainty representation, and the assimilation of identical CCI products) provided a controlled framework to isolate the role of model-specific behaviors. This approach was not redundant but necessary to support the paper's main conclusions, which could not have emerged from a single-model setup. The differences in model responses enabled us to better understand the conditions under which long-term satellite-derived observations can improve hydrological modeling.

A paragraph was added to the Conclusions section after discussing the results and conclusions of the study, to provide further clarity on the logic behind using both models (L732-735) :

“By using both CTRIP and MGB, our study strengthened its conclusions and confirmed the added value of assimilating satellite-derived discharge and WSE across differing hydrological modeling frameworks. Improvements observed in both an uncalibrated global model and a calibrated regional model demonstrate the wide-ranging benefits of using these long-term satellite-based datasets for various modelling applications.”

2. Given the fact that calibration or DA based on in-situ discharge measurements are widely used in current studies to improve the discharge simulations, it would be better to compare with these traditional methods to highlight the added value of assimilating satellite-based data.

We agree that comparing with traditional in-situ based calibration or assimilation would be valuable, but it would have significantly expanded the scope of the study. Our goal here was to focus specifically on the standalone impact of the CCI satellite-derived discharge and WSE products in a controlled, satellite-only

setup. Moreover, long-term (20-year) in-situ data are not consistently available at all the stations used in this study, making such a comparison uneven. We have clarified this scope in the revised manuscript.

3. I feel that it's better to divide the introduction of data and model into two separate sections, and the "Experimental Setup" presented in section 3 can be merged into the introduction of the model section.

As mentioned in the initial review, we preferred to keep the introduction to models and observations together in Section 2, as this follows the common structure found in DA studies. Section 2 presents the available material (data and models), while Section 3 describes how this material is used to answer the scientific question. We also believe it is clearer to keep the experimental setup separate from the model descriptions, as it involves the model-DA framework, the use of satellite products and the study basins in the assimilation experiments, the three elements being described in the previous section.