

Responses to Comments of Reviewer #1

Comments:

The authors just try to argue instead of addressing my previous comments. Accordingly, I still suggest a major revision. I will echo my previous important concerns below.

1. The third key question raised by the authors sounds weird.
 - a. Why is it so important to investigate the DA impact between a calibrated conceptual model (MGB) and an uncalibrated physically-based model? If the authors want to know the impact of model physics on the DA, it sounds more reasonable if both the adopted conceptual model and physically-based model are calibrated or uncalibrated. It's not clear how the calibration would affect the calibration. In addition, the input of precipitation for the two models are also different, which inevitably affects the investigation as well.
 - b. From the analyses presented in the results and conclusions, I don't find key findings or summaries related to the third key questions, e.g. what can we learn from the comparisons between two basins with different hydrological characteristics, especially given the fact that different precipitation data and density of discharge gauge networks are utilized.

Response to 1.a. :

We thank the reviewer for this important comment. We agree that the previous formulation of research question #3 could be interpreted as if the study aimed to isolate the respective impacts of model structure, calibration and precipitation forcing through a controlled model-intercomparison experiment. This was not the intended scope of the manuscript, and we recognize that the Introduction did not make this scope sufficiently clear.

We have therefore revised the framing of the Introduction to present the study more explicitly as an application assessment of the ESA CCI WSE and discharge products across two existing hydrological data assimilation systems. The objective is not to determine whether model structure, calibration or forcing explains the differences between MGB and CTRIP. Instead, the objective is to assess whether the added value and limitations of the CCI products remain consistent across realistic modelling contexts representative of different large-scale hydrological applications.

In this revised framing, CTRIP-HyDAS and MGB-HYFAA are treated as two complete modelling/assimilation frameworks, each with its own intended application, forcing strategy, calibration status and DA implementation. Both frameworks have been widely used in the past with distinct objectives. CTRIP is uncalibrated and uses forcings that are available at the global scale so that it can be applied over the whole globe and performances evaluated over gauged basins can be reasonably extrapolated to ungauged basins. On the other hand, MGB is usually applied for regional applications (such as flood forecasting) over some selected basins and calibration and local forcings ensure the best model performances for these applications, but prevent any extrapolation over other basins. Although the contexts in

which the two models are used differ, both could benefit from assimilating CCI-type data to improve their performance. Therefore, the comparison is not designed to attribute differences to one isolated factor, but to evaluate the robustness of the CCI product impacts across these two realistic frameworks, considering these differences.

We also agree that the use of different precipitation forcings can influence the open-loop simulations and therefore the apparent DA impact. For this reason, we now clarify that the models were kept in their native configurations: bias-corrected ERA5 for CTRIP, and GSMaP/IMERG precipitation for MGB in the Niger and Congo basins. This choice reflects the existing configurations of the two systems and their intended applications. Harmonizing the precipitation forcing would require a dedicated sensitivity experiment, including recalibration of MGB, and would address a different question from the present CCI product assessment.

Accordingly, we revised the Introduction, the research questions and the subsection concerning the models implementations over the basins, to better define the scope of the study:

- Within the Introduction, we revised the following lines :

Lines 83-105 : *“In this study, we evaluate long-term discharge and WSE products from the ESA Climate Change Initiative (CCI) in two existing large-scale hydrological models including their assimilation systems framework : CTRIP and MGB. CTRIP is a continental-scale, uncalibrated river routing model designed for use within Earth system modeling chains (Decharme 2019). In contrast, MGB is a basin-scale, calibrated, distributed hydrological model commonly used for regional flood forecasting and water resource assessments (Collischonn 2007, Pontes 2017). Both models have been widely used in the past with distinct objectives. CTRIP is uncalibrated and uses forcings that are available at the global scale so that it can be applied over the whole globe and performances evaluated over gauged basins can be reasonably extrapolated to ungauged basins. On the other hand, MGB is usually applied for regional applications over some selected basins and calibration and local forcings ensure the best model performances for these applications, but prevent any extrapolation over other basins. Rather than comparing model performance, we explore the robustness, limitations, and potential benefits of using these CCI-derived data products across two complete modeling/assimilation frameworks, each with its own assumptions, constraints, and intended applications. The study focuses on two major African basins, the Niger and the Congo, which provide contrasting hydrological and observational contexts. The Niger Basin is characterized by strong seasonal variability, the influence of the Inner Delta, and a comparatively denser in-situ network. In contrast, the Congo Basin has more stable large-scale flow variability, extensive wetlands and tributary storage effects, and a much sparser in-situ observation network. This dual-basin setup allows us to examine whether the assimilation impact of the CCI products remains consistent across different hydrological regimes and data-availability conditions. To focus specifically on the stand-alone value of the ESA CCI satellite products, we deliberately exclude in-situ discharge assimilation and mixed-product assimilation experiments. This design enables a focused evaluation of CCI-derived WSE and discharge products as assimilation inputs, without confounding their impact with additional observation sources. The study is therefore framed as an application-oriented CCI product assessment, testing how assimilating these products*

improve model performance under operational constraints in two distinct hydrological frameworks.”

- The third research question has been revised as follows:

“3- How robust are these impacts across two contrasting hydrological modelling contexts and two basins with different hydrological regimes and observation availability?”

- Within the subsection on model implementation over the Niger and Congo basins (Modeling framework and experimental setup Section), we edited the following paragraphs :

Lines 300-312 : “The use of different precipitation forcings was deliberately retained to preserve the native configurations of the two modelling systems. CTRIP is used with bias-corrected ERA5, consistent with its global and climate-oriented configuration, while MGB is used with the precipitation products adopted in its calibrated basin-scale implementations. This choice is consistent with the scope of the study, which is to assess the ESA CCI products in realistic modelling/assimilation frameworks rather than to conduct a harmonized model-intercomparison experiment.

MGB was calibrated against in-situ discharge and evaluated using altimetry and flood extent data, while CTRIP follows an uncalibrated, physically based approach.

These differences in forcing, calibration strategy, spatial scale, and process representation are treated here as part of the native configuration of each modelling system. The dual-model setup is therefore used to examine whether the impact of long-term satellite-derived WSE and discharge assimilation is consistent across two contrasting hydrological modelling frameworks, rather than to compare absolute model performance or isolate the effect of individual model components.

Simulations are conducted over the period 2000–2020 for both basins and models, ensuring a common evaluation period for the DA experiments.”

- Within the subsection of CTRIP-HyDAS and MGB-HYFAA (Data Assimilation schemes section), we edited the following paragraph :

Lines 355-357 : “Evaluating data assimilation performance in both frameworks provides complementary insight into the potential and limitations of long-term satellite-derived discharge and WSE products under different model configurations, spatial scales, and data-availability conditions.”

Response to 1.b. :

We thank the reviewer for this important comment. We agree that the previous version of the manuscript did not sufficiently synthesize the findings related to the third research question, particularly regarding what can be learned from the comparison between the Niger and Congo basins, the two modelling frameworks, and the different spatial and temporal densities of the assimilated observations. This point is also related to Reviewer 2 Major Comment 5. To address it, we reframed the third research question and added a dedicated

synthesis in the Discussion and Conclusion sections, explicitly distinguishing product-related, basin-specific, and modelling-framework-specific conclusions.

Across the two modelling systems, discharge assimilation is generally more directly representative than WSE anomaly assimilation because discharge is directly represented in the routing models, while WSE anomaly assimilation depends on the consistency between observed and simulated rating curves. However, the results also show that WSE anomalies can still be effective when the open-loop simulation is poor and when the modelled and observed rating curves are sufficiently aligned, because the lower WSE uncertainty gives these observations a stronger weight in the EnKF update.

The revised manuscript also clarifies the combined role of temporal density and product quality. In the Niger Basin, Qmultispec provides high-frequency observations that help reduce bias and improve internal variability, particularly in MGB-HYFAA and, to a lesser extent, in CTRIP-HyDAS. However, this benefit is conditional on product quality. In the Congo Basin, Qmultispec performs less well, especially at Bangui, where the product is noisy and poorly represents the hydrological cycle. This demonstrates that dense observations are beneficial only when their retrieval quality is adequate.

We also added a clearer interpretation of the model-DA framework dependent findings. In MGB-HYFAA, where the model is calibrated and includes a more detailed representation of floodplain and wetland processes, data assimilation mainly refines discharge magnitude, timing, bias, and variability. In CTRIP-HyDAS, used here in its global-scale uncalibrated configuration, data assimilation can partly compensate for larger open-loop errors, with the strongest gains often related to bias reduction, but it cannot fully correct errors caused by missing or simplified physical processes or by water deficits coming from forcings. This is particularly visible in the Niger Basin, where the missing representation of Inner Delta evaporation and hydrodynamics strongly affects CTRIP simulations, and in the Congo Basin, where CTRIP underestimation is partly linked to insufficient rainfall input over the basin, as discussed in the Supplement (2).

About the basin-dependent interpretation : the assimilation impact is clearer in the Niger Basin because the hydrological regime is strongly seasonal, the assimilated stations are more and are distributed along the main river system, and more in-situ stations are available for evaluation. In contrast, the Congo Basin has a more stable hydrological regime, with major tributaries, while the available CCI and in-situ stations provide a more limited spatial representation of the basin. As a result, assimilation improvements are more modest and more product-dependent in the Congo, and spatial coverage becomes a key limitation. The limited influence of the assimilated stations over the Congo tributary system further highlights the need for denser observations in future work.

We would like to clarify that the different precipitation forcings and station densities are not treated as isolated experimental variables in this study. They are part of the native modelling and observational contexts of the two DA systems, the objective being to assess how robust the added value of the available ESA CCI WSE and discharge products is across realistic large-scale hydrological modelling contexts.

Accordingly, we added the following synthesis in the Discussion section:

Lines 722-749 : *“The comparison between CTRIP-HyDAS and MGB-HYFAA, and between the Niger and Congo basins, helps distinguish product-related effects from framework- and basin-dependent responses. A first product-related result is that discharge assimilation is generally more directly representative than WSE anomaly assimilation, because discharge is directly represented in the routing models, while WSE anomalies depend on the consistency between observed and simulated rating curves. However, this result is not absolute: WSE anomalies can still provide substantial corrections when the open-loop simulation is poor when the modelled and observed rating curves are sufficiently aligned, as observed for CTRIP in some cases, because the lower WSE uncertainty gives these observations a stronger weight in the EnKF update. A second robust result concerns the role of temporal density. In the Niger Basin, the high temporal sampling of Qmultispec helps reduce bias and improve internal variability, particularly in MGB-HYFAA and, to a lesser extent, in CTRIP-HyDAS. However, this benefit is conditional on product quality. The Congo experiments show that high temporal density alone is not sufficient: when Qmultispec is noisy or out of phase with the hydrological cycle, as observed at Bangui, assimilation can degrade model performance. Dense observations are therefore beneficial only when their retrieval quality is adequate.*

The model-DA-framework comparison further shows that DA acts differently depending on the baseline model behaviour. In MGB-HYFAA, where the model is calibrated and includes a more detailed representation of floodplain and wetland processes, data assimilation mainly refines discharge magnitude, timing, bias, and variability. In CTRIP-HyDAS, which is used here in its uncalibrated global-scale configuration, assimilation can reduce some open-loop errors, with the strongest gains often related to bias reduction, but it cannot fully compensate for missing or simplified physical processes, such as floodplain storage and evaporation in the Inner Niger Delta, or for forcing-related water deficits such as those contributing to discharge underestimation in the Congo Basin.

The basin comparison finally shows that the assimilation impact is easier to detect in the Niger Basin than in the Congo Basin. In the Niger, the strong seasonal hydrological signal, the availability of several assimilated stations along the main river, and the denser validation network allow the DA impact to be more clearly diagnosed. In the Congo, the hydrological regime is more stable, and the outlet discharge integrates contributions from several large tributaries, while the available assimilated and validation stations provide a more limited spatial representation of the basin. Consequently, improvements are more modest and more dependent on product quality and spatial coverage.

These results suggest that ESA CCI products could support long-term hydrological modelling across different large-scale frameworks. However, their impact should be considered alongside the realism of the models in relation to basin-specific hydrological processes, as well as the products' quality and observation density.”

We also revised the last paragraph of Discussion section :

Lines 764-773 : *“In summary, the effectiveness of data assimilation in large-scale hydrological models depends on the interaction between the assimilated variable, product quality, observation sampling, model behaviour, and basin characteristics. Discharge assimilation is generally more directly consistent with routing models than WSE anomaly*

assimilation, although dH can still provide useful corrections when the modelled and observed rating curves are aligned and the observation uncertainty is low. The results also show that high temporal density is beneficial only when the assimilated product is sufficiently reliable: Qmultispec helps reduce bias and improve internal variability in the Niger Basin, but can degrade the analysis in the Congo Basin when observations are noisy or phase-shifted. Finally, the comparison between the two basins highlights the importance of spatial sampling and hydrological connectivity. The stronger seasonal signal and better station coverage in the Niger Basin make DA impacts easier to diagnose, whereas the sparse sampling of the Congo tributary system limits basin-wide correction and makes the results more dependent on product quality and validation constraints.”

We also revised the following paragraph in the Conclusions to explicitly answer the third research question:

Lines 795-803 : The comparison across CTRIP-HyDAS and MGB-HYFAA and across both basins shows that the main conclusions are not controlled by a single factor. Discharge products are generally more directly assimilated into routing models than WSE anomalies, and temporal density is a key factor for correcting bias and internal variability when product quality is sufficient. However, the magnitude and stability of the improvements depend strongly on the modelling framework and basin characteristics. In calibrated regional configurations such as MGB-HYFAA, DA mainly refines already constrained simulations. In uncalibrated global-scale configurations such as CTRIP-HyDAS, DA helps reduce bias and can partly compensate for larger open-loop errors, but remains limited when relevant basin processes are missing. The Niger Basin illustrates the benefit of denser observations along a strongly seasonal main river system, while the Congo Basin highlights the need for improved spatial coverage across large tributary systems and greater sensitivity to assimilated product quality.”

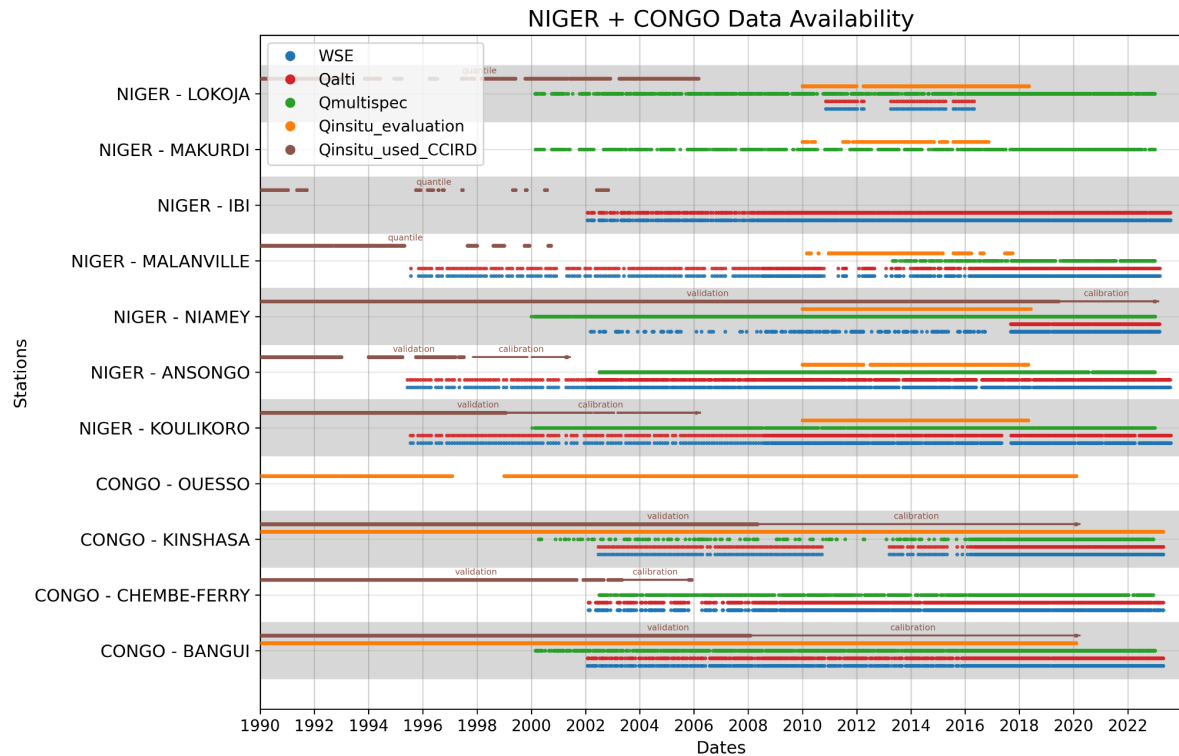
2. I still believe that the structure of Sections 2 and 3 is kind of massy and difficult to follow.

Following the recommendation of the reviewer, we changed the structure of the introduction of data and models. Section 2 now presents the study domain (Niger and Congo basins) and the assimilated satellite products. Section 3 presents the two modeling frameworks and the experimental setup. We agree that the manuscript is clearer that way and we thank the reviewer for this suggestion.

3. The authors should consider assimilating the in-situ discharge data from the stations where both in-situ and satellite-based data are available, and the remaining stations can be used for the validation. As such, the impact of uncertainty of the satellite-based data can be further investigated compared to the performance of in-situ observations.

We thank the reviewer for this suggestion. We considered adding an in-situ discharge assimilation experiment as a baseline, but for most of the stations the available in-situ records are not long enough to provide a fair comparison with the CCI assimilation experiments. And this applies both to the independent in-situ datasets used for validation and to the GRDC records involved in the Qalti rating-curve calibration/validation.

The Figure displayed below shows, for each station and dataset, the temporal availability of WSE, Qalti, Qmultispec, independent in-situ observations, and the in-situ data used for the CCI rating-curve calibration/validation. The figure also distinguishes, when relevant, the calibration and validation periods used for the rating-curve construction.



As shown in the Figure most available independent in-situ records cover only about 8 years, while the CCI products provide close to 20 years of observations over the study period, with only a few exceptions such as Lokoja and Niamey.

For these reasons, we did not include an in-situ assimilation baseline in the present study.

As a result, the performance metrics for an in-situ assimilation baseline would only be computed over the shorter in-situ period, while the CCI assimilation experiments are evaluated over the full long-term period. These scores would therefore not be directly comparable because of the different evaluation periods and the hydrological conditions sampled. For this reason, an in-situ assimilation experiment would not provide a fair baseline for the 20-year CCI assimilation experiments. For the same reason, going back to round #2 of reviewing discussions, we did not add other, more spatially dense WSE datasets to the assimilation experiments since they do not cover the 20 years period, as this would have changed the experimental objective.

In addition, not all selected stations have independent in-situ observations available for such a baseline, for example Ibi and Chembe-Ferry in our setup.

To document this point and to answer also Reviewer #2 3rd major comment, we added Supplementary Fig. S3, which summarizes the temporal availability of the CCI products, and

the in-situ observations (the ones used for evaluation + the in-situ data involved in the Qalti rating-curve calibration/validation) at each selected station. The figure shows that in-situ records do not cover the full 2000–2020 period considered for the CCI experiments.

Responses to Comments of Reviewer #2

This manuscript presents a study to evaluate the effectiveness of assimilating the long-term ESA CCI discharge and WSE data into two contrasting large-scale hydrological models based on model performance. The study seeks to conclude which type of satellite data (noisy flow data vs cleaner height), at which temporal resolution and spatial coverage, provides better model improvements. The study implemented DA into the uncalibrated CTRIP and calibrated MGB models, which allows the readers to evaluate if DA acts as a correction (or calibration) for a poor-physics or a refinement of already good parameters. The paper is well-organized and well-written, but the reviewers have some concerns before this manuscript is considered for acceptance for publication.

Major Comments:

1. Although the authors explicitly state that they did not compare model absolute performance but rather the relative improvement from DA. However, two models are in different scales and driven by meteorological data on different scales. For instance, MGB uses satellite-based precipitation (GSMAP/IMERG) known for better tropical performance, while CTRIP uses ERA5 reanalysis. How do the authors consider the bias in precipitation? If CTRIP performs poorly, is this due to the model physics or because of input driving precipitation data? Why do the authors use the same consistent forcing data to isolate the impact from model structure vs. forcing error? The reviewer is curious if the authors can conduct a sensitivity test where both models use the same precipitation forcing for one sub-basin to see if the differences persist. In the following section, the authors used different perturbation methods for each model. Does this also impact the conclusions the authors made in this paper? The better model performance might be due to a superior DA setup? It is good to see the same localization strategy used for both models.

The authors thank the reviewer for this insightful comment. As explained in the first response to reviewer #1, the main objective of this manuscript is to evaluate the impact of assimilating CCI products into two distinct modeling frameworks that have been used in several studies in the past in different contexts. We agree with the reviewer on the impact of forcings and modeling/DA setup on the performances. Actually, models being approximations of the real physical processes, they always show some limitations and deficiencies. Such limitations generally come from the quality of meteorological forcings and physical processes simplifications. For example, CTRIP does not consider evaporation over the Niger river and floodplain, which is responsible for the high overestimation of the discharge downstream the Niger Inner Delta. In addition, for CTRIP, forcings have to be available at the global scale for global scale river routing simulations, which is the main use of this model. For a specific basin, better forcings may be available and provide better performances, but these results would not be extrapolatable to other regions, which is one the strengths of CTRIP. On the other hand,

MGB could have been forced by the ERA5 reanalysis, but this would have required recalibration of the parameters, which is beyond the scope of this study.

In the context of reanalyses, data assimilation is performed to overcome such limitations by correcting the model state to reduce the differences between model outputs and observations. Here we focus on the impact of assimilating CCI products, which has the particularity to be available over a long time period, within those two distinct modeling frameworks.

Concerning the perturbation methods, the ensemble perturbation strategy can influence the magnitude of the DA impact, since it affects ensemble spread, forecast error covariance, and therefore the Kalman gain. To reduce methodological differences between the two DA systems, we made sure that both CTRIP-HyDAS and MGB-HYFAA use the same hydrological localization strategy, which limits one important source of discrepancy in the spatial propagation of assimilation increments. However, the perturbation methods differ because each framework was used in its native configuration. Since the objective of the study is not to perform a controlled intercomparison of DA setups or to attribute differences between models to one isolated component of the assimilation system, we focused the analysis on assessing the behaviour of ESA CCI WSE and discharge products within two existing hydrological modelling/assimilation frameworks. We have therefore interpreted framework differences cautiously throughout the manuscript, as differences between complete DA systems rather than as evidence that one DA setup is intrinsically superior to the other.

2. The authors mentioned that the floodplain scheme was not activated in the CTRIP, but the Inner Niger Delta and the Congo Basin are described with floodplains. Does the deactivation lead to an over-reliance on DA to simulate water storage

We agree with the reviewer that DA compensates for the fact that floodplains (and evaporation over floodplains) is neglected in CTRIP. Actually, the main purpose of DA is specifically to compensate for model deficiencies, including forcing errors, missing processes and modeling approximations. In both basins considered in this study, floodplains have a clear impact on river flows, but as stated previously (and in section 4.1 Open-loop Performance - Key issues in both models) forcings are not error-free and other approximations (like groundwater flows, dams, etc.) are also responsible for poor performances of the open-loop. This has been reminded in the conclusion section.

3. The discharge data from altimetry and multiple spectral images used for assimilation are derived from rating curves calibrated using in-situ discharge data. For the validation metrics presented (e.g., the NSE of 0.83 in the Niger), did the authors ensure that the in-situ gauges used for validation were independent of those used to calibrate the ESA CCI products? Could the author clarify this? If there is overlap, how can the authors distinguish between true model improvement and the assimilation of pre-calibrated information?

We thank the reviewer for raising this important point. We agree that the link between the in-situ data used to produce the CCI discharge products and the in-situ data used for model evaluation needs to be made clear.

We have checked again the origin of the in-situ data used for the calibration/validation of the rating curves associated with the CCI altimetry-derived discharge product, and compared them with the in-situ data used to validate our assimilation experiments. The situation is different for the Niger and Congo basins.

For the Niger basin, the validation data used in this study are independent from the in-situ data used for the Qalti rating-curve calibration/validation. The rating curves were built using GRDC discharge data, while the validation scores presented in the paper are computed using ABN in-situ observations. Therefore, the validation metrics reported for the Niger basin, including the NSE value mentioned by the reviewer, are not computed against the same in-situ records as those used to calibrate the CCI Qalti product. The Niger results can therefore be considered as an independent evaluation of the assimilation impact.

For the Congo basin, the situation is less independent at some stations. For Kinshasa and Bangui, the Qalti rating-curve calibration/validation relies on SO-HYBAM discharge data, and these SO-HYBAM observations are also the only available in-situ source for validating the model simulations at these stations. There is therefore a partial dependence between the in-situ data used in the CCI product generation and those used for model evaluation. Chembe-Ferry is a different case, since the in-situ data used for the calibration/validation of the Qalti rating curve come from GRDC.

We agree that the lack of fully independent in-situ validation data in the Congo basin limits the interpretation of the validation metrics at these stations. This point is now explicitly acknowledged in the revised manuscript.

To clarify this point, we added a supplementary data-availability figure (please see the Figure displayed in our response to comment #3 of Reviewer #1) showing, for each station and dataset, the temporal availability of WSE, Qalti, Qmultispec, independent in-situ observations, and the in-situ data used for the CCI rating-curve calibration/validation. The figure also distinguishes, when relevant, the calibration and validation periods used for the rating-curve construction. This helps identify whether the in-situ validation data are independent from, partly overlapping with, or identical to the data involved in the CCI discharge product generation.

We have added the following paragraph in the discussion section about the limitation in the interpretation of the validation metrics in Congo basin :

Lines 753-763 : *“A further limitation concerns the independence of the in-situ validation data, which differs between the two basins. In the Niger Basin, the CCI discharge products were generated using rating curves calibrated and validated against GRDC discharge data, while model evaluation was performed using independent ABN observations. The Niger validation metrics can therefore be considered independent from the in-situ data used in the CCI discharge product generation. In the Congo Basin, the validation at Bangui and Kinshasa is less independent, because SO-HYBAM observations are the only available long-term in-situ discharge records and were used both for CCI discharge product generation and for model*

evaluation. The Congo validation scores should therefore be interpreted with more caution than the Niger scores. Nevertheless, they remain useful for analysing the behaviour of the DA experiments, since the metrics are computed from the model discharge simulations after assimilation and not directly from the assimilated observations themselves. Supplementary Fig.~S3 summarizes the temporal availability of the assimilated CCI products, the in-situ observations used for model evaluation, and the in-situ data involved in the CCI discharge rating-curve calibration.”

We also would like to note that obtaining a fully independent validation dataset is often difficult in hydrological data assimilation studies, especially in data-limited basins where the same gauge network may be the only available reference. For the Congo basin, we therefore acknowledge that this partial dependence limits the interpretation of the validation scores. However, the scores are computed on the model simulations after assimilation and propagation through the hydrological model, not through a direct comparison with the data used to build the Qalti rating curves.

4. The open-loop analysis reveals a massive structural bias in CTRIP. Given that the authors assume the model is 'error-free' in their DA setup, isn't the filter essentially being used to compensate for missing physics (evaporation) rather than meteorological uncertainty? Could the authors discuss it? Is DA effectively acting as a bias-correction for missing physics?

As stated in the response to comment #2, we agree that DA is used to compensate for missing physics as well as for errors in forcings. It has to be noted that those errors (forcings and model) are accounted for in the data assimilation algorithm via the error covariance matrix (P) and the use of an ensemble. It is shown in this study that DA has globally more impact with CTRIP than with MGB, in particular because of the model deficiencies visible in the evaluation of the open-loop.

5. Regarding Question 3, which aims to explore the differences between the Congo and Niger basins, the Results and Conclusion sections lack a dedicated synthesis of these findings. Could the authors provide a comparative Table that addresses: how did the inter-annual variability (high in Niger, lower/stable in Congo) affect DA stability? How did the gauge network density influence the localization's effectiveness? What is the generalized takeaway that applies to both, vs. what is basin-specific?"

We agree that the previous manuscript did not sufficiently synthesize the findings related to the third research question (revised). This point has also been raised by Reviewer 1 Major Comment 1.b. Following the reviewer's suggestion, we added a comparative table summarizing how hydrological variability, gauge network density, assimilated products quality, and model-assimilation framework characteristics affected DA impact in the Niger and Congo basins. The table explicitly separates generalized findings from basin-specific and framework-specific interpretations. We also added a synthesis paragraph in the Discussion section, between lines 722-749, and revised the Conclusion to make the main take-away explicit. Please also see our response to Reviewer 1 Major Comment 1.b.

The comparative table was added in the Discussion section as follows:

Dimension	Main take-away	Specific interpretation
Product type	Q assimilation is generally more robust than dH assimilation.	Q is directly represented in routing models; dH depends on rating-curve consistency. However, dH can still perform better in poor open-loop configurations when rating curves are aligned, as in some CTRIP cases
Temporal density / Quality	Dense observations help reduce bias and improve internal variability only if product quality is sufficient.	Qmultispec improves bias/variability as seen in Niger basin, but can degrade results when noisy or phase-shifted, as seen in Congo.
Basin signal/ Spatial Sampling	DA impact is clearer when the dominant hydrological signal is well sampled by hydrologically connected assimilated stations.	The Niger Basin has a strong seasonal signal and several assimilated stations along the main river, favouring coherent upstream/downstream corrections. In the Congo Basin, the discharge signal is distributed across large tributaries, and sparse assimilated stations only weakly constrain basin-wide dynamics.
Validation network and independence	Gauge density affects confidence in evaluation. Validation is more independent in Niger than in Congo.	Niger has a denser independent ABN network; Congo has few validation stations, limiting basin-wide diagnostic confidence, and Bangui/Kinshasa partly relies on SO-HYBAM data also involved in CCI rating-curve calibration/validation.
Model framework	DA response depends on the open-loop error type.	MGB calibrated simulations are mainly refined by adjusting magnitude, timing and variability; in CTRIP, DA can reduce larger open-loop biases but remains limited by missing physical processes and forcing-related errors.

Minor Comments:

1. In line 8, could the author briefly explain why discharge assimilation outperformed WSE?

We thank the reviewer for this comment. We clarified this point in the Abstract. Discharge assimilation outperformed WSE anomaly assimilation mainly because discharge is directly represented in the routing models, while WSE anomaly assimilation depends on the consistency between observed and simulated rating curves. When this consistency is weak, WSE updates can introduce less stable discharge corrections. We therefore revised the Abstract to explicitly state this point.

Revised lines 7-9 : *“In the Niger Basin, discharge assimilation outperformed WSE anomaly assimilation because discharge is directly represented in the routing models, providing a more direct and physically consistent correction, while WSE requires consistency between observed and simulated rating curves to be converted into effective discharge corrections.”*

2. In lines 12-15, it shows that DA works better for the calibrated model than for the uncalibrated model. Could we conclude that it is better to calibrate the model in the first place and then implement DA for further improvements?

We thank the reviewer for highlighting this point. In the calibrated MGB configuration, DA mainly refines discharge magnitude, timing, bias, and variability. In CTRIP, DA can still reduce larger open-loop errors, especially bias, but its impact remains limited when errors are caused by missing physical processes or forcing-related biases. Therefore, our results support the idea that model calibration or baseline realism can improve the effectiveness of DA, but they do not constitute a controlled experiment proving that calibration should always be performed before DA. We clarified this interpretation in the Manuscript and discussed it consistently with our response to Reviewer 1 Major Comment 1.a.

3. In line 17, the author mentioned that the Congo Basin exhibits modest improvements. Could the author clarify the reason? Is it due to the CCI data quality in this region, as it covers wetlands?

We thank the reviewer for this comment. We clarified in the Abstract that the more modest improvements in the Congo Basin are not attributed to a single factor. They result from the combination of sparse assimilated and validation stations, the distributed contribution of large tributaries to the discharge signal, and lower Qmultispec quality at some stations, especially where the product is noisy or poorly represents the hydrological cycle. This point is now also synthesized in the comparative table added to the Discussion in response to Major Comment 5.

Revised lines 18-20 : *“The study also emphasizes the trade-offs between data resolution, spatial sampling, and product quality, particularly in the Congo Basin, where improvements were more modest because of sparse assimilated and validation stations, the distributed contribution of large tributaries and wetlands, and lower quality of discharge products at some stations.”*

4. In line 216, please complete the sentence.

We thank the reviewer for pointing this oversight out. The correct reference link has been included :

Siqueira, V. A., Paiva, R. C. D., Fleischmann, A. S., Fan, F. M., Ruhoff, A. L., Pontes, P. R., et al. (2018). Toward continental hydrologic–hydrodynamic modeling in South America. Hydrology and Earth System Sciences, 22(9), 4815–4842. <https://doi.org/10.5194/hess-22-4815-2018>.

5. In line 245, could the author mention the spin-up period which stabilizes deep soil moisture states before beginning the DA experiments?

We thank the reviewer for this relevant question. The MGB model and the ISBA model (the Land Surface Model used to force CTRIP) only represent the root zone in the soil, generally around 1 m (but depends on the vegetation type). Since it is quite thin, a few weeks or months is generally sufficient to stabilize the soil moisture state. In addition, CTRIP considers aquifers which require a longer spinup to reach a stable state. In this specific case, a long term simulation (100 years) has been done as a preliminary step. Then a 1-year spinup has been designed in all the experiments of this study. This has been clarified in the manuscript.

6. In line 402, could the author clarify whether the temporal lag affects the DA performance?

We thank the reviewer for this comment. We clarified that the temporal lag in Qmultispec can affect DA performance because observations shifted relative to the modelled hydrological cycle may introduce corrections at the wrong phase of the hydrograph. This can reduce the effectiveness of the update, particularly for peak timing and correlation, and may contribute to degraded performance when the product is also noisy.

Revised line (230) : *“In contrast, Qmultispec offers denser temporal coverage but shows greater noise and a temporal lag in capturing hydrological peaks, especially at Niamey and Lokoja, which can affect data assimilation performance by introducing corrections that are out of phase with the modelled hydrological cycle.”*