Point by point author's response

We thank the referees for their feedback, which after revisions, we believe has improved the quality of the manuscript. Below we provide a detailed response for all comments.

Referee comments are in **bold**, our responses are *in italics and blue*, and if corresponds, changes made in the manuscript are in *italics and purple*.

Responses to **Referee 1** start on page 2, and responses to **Referee 2** start on page 19.

Jose P. Teran,

On behalf of the authors

Referee 1

<u>Main comments:</u>

The authors present an interesting study integrating reservoirs and lakes into the CoSWAT global hydrological model. The motivation for this work is clear, and the proposed model improvements appear promising. However, several aspects require further clarification and discussion before publication. The manuscript would benefit from additional analysis of basins where streamflow or storage deviates from observations, and the presentation of some figures could be improved for clarity. A clearer discussion of the model's limitations, including potential sources of uncertainty in representing reservoirs and lakes at the global scale, is needed.

We thank the referee for the positive assessment and the constructive comments. We agree that deeper discussion on poorly performing cases, improved figure clarity, and a clearer discussion of model limitations and uncertainty sources will strengthen the manuscript. We address each of these points in detail below.

<u>Detailed comments:</u>

Abstract

Line 27: Provide more specific information on how reservoirs and lakes affect streamflow and other hydrological variables, including quantitative performance metrics (e.g., stations with improved streamflow, KGE, PBIAS). Shorten the methods section to focus more on results.

We appreciate this suggestion. We will adjust this in the manuscript with more numerical details and shorten the methods. Below is the adjusted section of the abstract (lines 23 to 33):

> *In this study, we address this limitation by combining commonly used reservoir and lake modelling schemes from other global water models and the capabilities of SWAT+, and by implementing a topological approach to estimate irrigation water demands for reservoirs. We also introduce an improved workflow to integrate lakes and reservoirs into the river network, setting up the CoSWAT model at 0.01° resolution on selected regions worldwide, where validation of reservoir and lake storage, inflow, outflow and the impact on streamflow performance was assessed. The results show that the model captures monthly storage dynamics with moderate performance. Overall, 34% of evaluated reservoirs achieved a positive KGE, a performance comparable to other state-of-the-art global hydrological models, with CoSWAT surpassed by only two of five models in the comparison. Moreover, results demonstrate an improvement in performance in streamflow representation on 135 out of 192 (70%) evaluated stations following the integration of lakes and reservoirs. Despite these improvements, persistent streamflow* biases that consequently affect reservoir storage *point to the need for hydrological calibration, which remains a priority for future development.*

Introduction

Line 49: List models like ParFlow that include explicit reservoir and lake representation. An example is the research of Benjamin, https://doi.org/10.5194/hess-29-245-2025, A scalable and modular reservoir implementation for large-scale integrated hydrologic simulations

ParFlow is a good example, and we thank the referee for the suggestion. But it is important to clarify that the purpose of the paragraph is to set the current state of global models, which generally lack that property, as it is, of course, true that many other hydrological models (even at the regional scale, including SWAT+) do represent reservoirs explicitly, not using representative water bodies. Below is the adjusted section (lines 54 to 59):

> *Many hydrological models explicitly represent reservoirs and/or lakes, such as the Soil and Water Assessment Tool (SWAT+; Chawanda et al., 2020a; Sánchez-Gómez et al., 2025), ParFlow (West et al., 2025), or the mesoscale Hydrologic Model (mHM; Kumar et al., 2022; Samaniego et al., 2010; Thober et*

al., 2019), among others. However, most GWMs have a gridded structure which is generally at 0.5° resolution, meaning that lakes and reservoirs are considered as representative water bodies in the model grid when accounted for, not necessarily as individual, explicit elements. As a consequence, they require scaling procedures to analyze them further individually (Ayala et al., 2026).

Line 44-45:"Drying and wetting trends are intensifying…, replace "as is" with "accompanied by" for clarity.

This was corrected on the revised manuscript (lines 47-48):

Drying and wetting trends are intensifying (Bai et al., 2024), accompanied by ecosystem degradation (Grant et al., 2021; La Fuente et al., 2024; Woolway and Merchant, 2019).

Line 51: The sentence includes too many pauses and should be rewritten for clarity.

This line was rewritten in the revised manuscript (lines 56 to 59) as follows:

However, most GWMs have a gridded structure which is generally at 0.5° resolution, meaning that lakes and reservoirs are considered as representative water bodies in the model grid when accounted for, not necessarily as individual, explicit elements. As a consequence, they require scaling procedures to analyze them further individually (Ayala et al., 2026).

2 Methodology

2.1 Global Datasets

Clarify the target resolution to which all data were resampled.

We thank the referee for this comment. It might not be so clear in Table 1. All spatial inputs used for the creation of Hydrologic Response Units (HRUs) were resampled to 0.01° from their native resolution. While the rest (e.g., weather forcings) were not resampled. This was clarified by rewriting lines 130 to 137:

Datasets for model setup provide mappings of the physical properties of land and soil, which are essential for the creation of HRUs. The topography was defined using the Global Aster Digital Elevation Model (DEM) (Abrams, 2016). The reference land cover map was obtained from the ESA CCI Land Cover Product (ESA, 2017), while soil classification and properties were derived from the FAO Harmonized World Soil Database (Fischer et al., 2008). These data were resampled from their native resolution to 0.01°, as shown in Table 1. Finally, the CoSWAT model requires daily precipitation, maximum and minimum temperatures, solar radiation, wind speed, and relative humidity as weather inputs, all of which are available in the GWSP3-W5E5 dataset (Lange et al., 2022) at 0.5°, and were used in this study to perform historical simulations.

Line 145-157: Only highlight the nine major basins instead of listing all basins in detail. Consider merging with Section 2.3 for clarity.

We thank the referee for the suggestions. We believed it was important to mention this as the modelled regions do not all necessarily represent one large system or basin, but can include several independent ones. Nonetheless, we agree with the suggestion, and the section was shortened on the revised manuscript (lines 149 – 157) as follows:

The CoSWAT model is divided into 90 regions (Chawanda et al., 2025) representing one or more large river basins. For this study, 9 of those regions (Figure 1), were selected to apply and evaluate the implementations developed in this study, spanning diverse environmental and operational conditions across nearly all continents, where sufficient data were available for robust evaluation. The selected regions span several climatic zones worldwide, including tropical, arid, temperate, cold, and polar (frost), according to the Köppen-Geiger classification system (Beck et al., 2023). In South America, the modelled region covers the La Plata, Parana and Endorheic Basins. In North America, the Mississippi

River System, the Colorado River Basin, the Grande (or Bravo) River Basin, and the Pacific Ocean Seaboard. In Africa, it encompasses the Nile River and the Orange River basins. In Asia, it covers the Mekong River and Chao Phraya River basins. Finally, two model regions cover most of western and central Europe, encompassing several river basins.

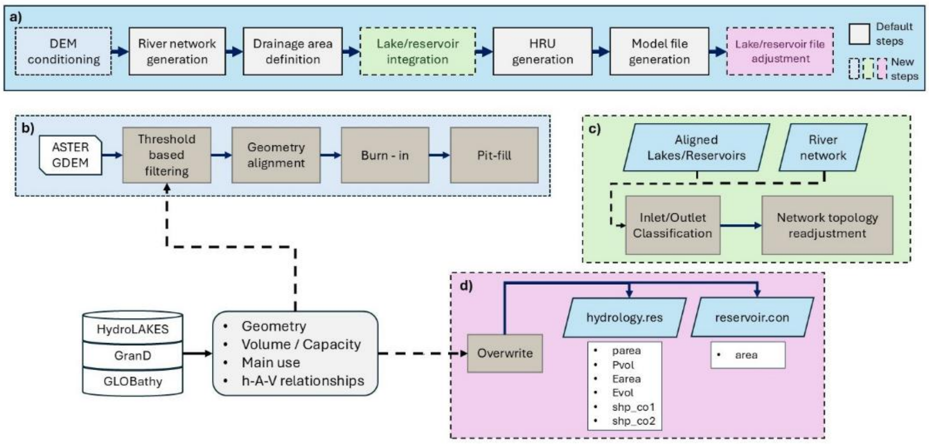
Figure 1: Add the north arrow and scale bar.

We respectfully note that the figure already includes a labeled graticule with latitude and longitude coordinates indicating East and North directions, providing full spatial reference for the map. For a global map in geographic projection, we believed the scale bar was not ideal at this extent as it varies with latitude, and, given the graticule, a north arrow would be redundant. Similar figures in comparable studies (Müller Schmied et al., 2021; Vanderkelen et al., 2022) do not include these elements either. We are happy to adjust if the Editor feels otherwise.

2.3 Reservoir/Lake integration into model network

Figure 2: According to Section 2.3, Figure 2b represents the overall workflow and should appear first. Rename figures to ensure Figure 2b is the first (overall workflow), and 2a follows. Like Figure 2b become 2a, and 2a become 2b.

We appreciate this suggestion. We believe the intended message of the figure may not have been communicated clearly enough. Figure 2a represents the main CoSWAT model setup process built by the CoSWAT Framework (Chawanda et al., 2025), and Figures 2b, 2c, and 2d illustrate the new procedures contributed in this study, showing how they are integrated into that framework. Therefore, Figure 2a intentionally appears first as the overarching context, with the new contributions following. To clarify this, figure 2 was adjusted as follows:



With a better description:

Overview of the a) CoSWAT framework for model structure generation, with three new components introduced in this study: b) DEM conditioning, c) lake and reservoir integration into the network topology, and d) adjustment of model files for reservoirs and lakes using global datasets (HydroLAKES, GranD, GLOBathy).

Moreover, to further clarify, we modify the surrounding text (line 168 to 169):

To overcome these limitations, we integrated new steps into the CoSWAT framework (Figure 2), which allows for an adequate integration of reservoirs and lakes into the model structure.

And lines (170 to 181):

The additions consist of a pre-processing stage (Figure 2b), a lake/reservoir integration stage (Figure 2c), and a model file adjustment stage (Figure 2d). The preprocessing stage has three main steps: (i) threshold-based filtering of water bodies, (ii) alignment of channel-lake intersections to the DEM grid, which prevents the creation of river segments and HRUs smaller than the DEM grid cell, which is required for numerical stability in SWAT+, and (iii) burn-in of lake/reservoir geometries into DEM by 14 meters following Lehner et al. (2008), and a depression correction (i.e. "Pit fill") with the Whitebox Geospatial Analysis Tools (Lindsay, 2016), resulting in a conditioned DEM. After that stage, the river network and drainage areas are generated using the Terrain Analysis Using Digital Elevation Models (TauDEM; Tarboton et al., 2009) tool. With that information, the integration of water bodies into the network proceeds, and it consists of two steps: (i) logical identification of inlets and outlets and (ii) re-adjustment of network topological connections. For this study, water bodies smaller than 20 km^2 were excluded, except when the degree of regulation (i.e., ratio of storage capacity and mean annual outflow) provided by the GranD dataset exceeds 70%, and the area exceeds 10 km^2. After the integration of lakes and reservoirs, the next step is to create the model's HRUs and model file generation.

Line 168-171: Reword to describe "integration of reservoirs and lakes into the model network" rather than "a new lake/reservoir resolution procedure."

This is completely true and we thank the referee for noting it. The wording was not the best. We made adjustments to refer to this as "integration" rather than "resolution" throughout the revised manuscript.

Line 176: how to burn-in of lake/reservoir elevations into DEM?

We thank the referee for posing this question point. Similar to stream burning, a common approach in SWAT+ and similar models, we condition the DEM by reducing the elevation of pixels intersecting the aligned lake/reservoir area by 14 meters. This follows a similar approach to HydroSHEDS (Lehner et al., 2008) and improves the subsequent stream network delineation, ensuring better alignment of river network with lake boundaries. This procedure is now described in lines 173 to 176:

... (iii) burn-in of lake/reservoir geometries into DEM by 14 meters following Lehner et al. (2008), and a depression correction (i.e. "Pit fill") with the Whitebox Geospatial Analysis Tools (Lindsay, 2016) ...

Line 186-188: Clarify why certain variables such as pvol, evol, parea, eaera, shp_col1/shp_col2, which represent reservoir information, point to hydrology.res rather than reservoir.con? Is it not more reasonable to use reservoir.con for reservoir information? Since area refers to the water body surface, is it not more consistent to use hydrology.res for water surface area? Also, why are max/min storage and outflow rates missing?

We thank the referee for this comment. Regarding the file structure, hydrology.res stores hydrological properties of reservoirs (such as volumes, areas, and shape parameters) used in calculations during simulations, while reservoir.con contains connectivity information (e.g., how the reservoir object connects to the channel network). This follows the SWAT+ model structure as documented in the official SWAT+ I/O documentation (https://swatplus.gitbook.io/io-docs), and is therefore the appropriate file for the variables listed. Regarding max/min storage and outflow rates, these are not set as direct parameters in the model files. Instead, they are implicitly handled for each reservoir through decision tables, as described in Section 2.4. A fixed outflow rate is a possible option within decision tables, but since CoSWAT now relies on generalized simulation schemes, specific outflow rates are not prescribed and are therefore not part of the modeling framework. However, the role of each file was clarified in the revised manuscript (lines 187 to 189):

... The relevant files are "hydrology.res", which contains numerical information used for calculations during simulation, and "reservoir.con", which contains information about connectivity with other elements in the model.

2.4 New lake and reservoir simulation scheme

Line 207: provide more details for decision table (e.g.reference).

We thank the referee for noting this, the corresponding reference was added in line 213 of the revised manuscript.

Line 208: Clarify which two parameterization methods are used (e.g., Doll et al., 2003 and H06 scheme).

We appreciate the suggestion, the suggested clarification was done in lines 214 to 215:

> *... We enhanced SWAT+ revision 61.0.2 by modifying reservoir-related subroutines and integrating decision tables with two parametric methods; one based on Döll et al. (2003) for natural lakes, and one based on the H06 scheme (Hanasaki et al., 2006) for regulated lakes and reservoirs, in order to improve global applicability, thereby supporting the CoSWAT GHM.*

Line 208 & 735: Introducing arrays, loops, variable initialization, represents preparatory steps rather than true modifications of the source code. Only changes to formulas/functions should be considered innovations. The third and fifth item in Table A-1 representing a novel contribution.

We thank the referee for this comment and believe the assessment may stem from the descriptions in Table A-1 not being sufficiently clear. The introduction of arrays, loops, and variable initialization is not merely a preparatory step but an integral component of the source code modifications. Specifically, they enable a time-variant implementation of the H06 scheme (Gharari et al., 2024), which is a key contribution of this work. These structural changes are essentially part of the scheme, and it would not be possible to use it without them. The innovations in Table A-1 should therefore be understood as a cohesive set of changes: the array that stores the reservoir "memory" of inflows and demands, updated continuously throughout the simulation, and the equations to calculate outflow based on this, together constitute the novel contribution. We believe the descriptions in Table A-1 may have wrongly suggested that the scheme-related modifications were limited to the res_hydro subroutine, while in fact all listed modifications are relevant and necessary. The following was added to the revised manuscript (lines 812 to 816):

> *The modifications listed in Table A- 1 constitute a cohesive set of changes required for the adequate implementation of the parametric schemes described in Section 2.4. In particular, the implementation of the H06 scheme as a time-variant parametric scheme requires the introduction of arrays and loops to store and continuously update past inflows and demands during the simulation, as well as the irrigation water demand action from HRUs, which conditions water release in irrigation-type reservoirs.*

While table A-1 was adjusted on the "res_hydro" row:

res_hydro	*Reservoir or lake outflow definition in connection with decision tables.*	*Introduced outflow calculation for the corresponding parametric scheme.*

Which in the original manuscript wrongly indicated that modification as the full implementation of the schemes.

Additionally, we believe it is important to describe all changes made in association to new implementations as it ensures transparency and traceability for the scientific community working with SWAT+.

Line 253: Combine sentences for better flow.

We appreciate the suggestion. The following change was made on the revised manuscript (lines 260 to 261):

> *... For the regions applied in this study, the threshold was set to 40%, after which the irrigation water source was identified for irrigated HRUs (Figure 3).*

Figure 3: In subplot (a), the blue area is not labeled, and the brown "Groundwater" layer appears minimal or missing. Please add a north arrow and scale bar.

We thank the referee for this comment. We have improved Figure 3 with better labeling and introduced an additional area where groundwater-sourced and rainfed agricultural HRUs are clearly visible. We have also provided an adequate graticule, a north arrow, and a scale. Below is the corrected figure:

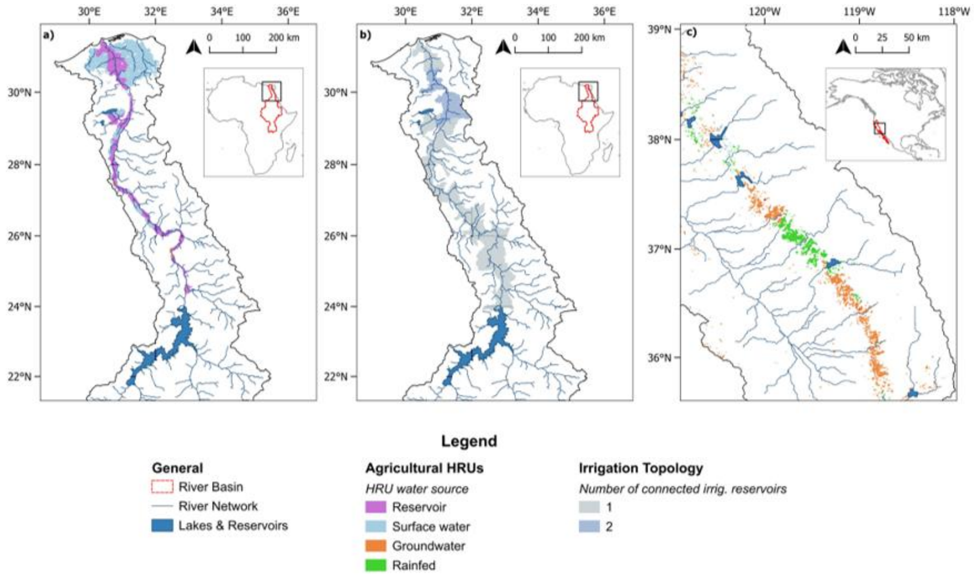


Figure 3: Irrigation water source assignment for agricultural HRUs and irrigation topology, illustrated for two contrasting regions: a) a predominantly irrigated downstream section of the Nile Basin, b) the corresponding irrigation topology showing the number of connected irrigation reservoirs per sub basin, and c) a section of the Pacific Seaboard in North America, where agricultural HRUs rely predominantly on groundwater or are rainfed.

2.6 Simulation setup, model comparison, and evaluation

Line 274 & 295: Ensure uniform resolution for the nine selected basins and all input datasets.

We thank the referee for this comment. We believe this comment may relate to two different datasets serving different purposes. For line 274, the 0.01° resolution refers to the spatial input data used for model configuration, an improvement over the 0.02° resolution of CoSWAT v1. For line 295, the 0.5° grid refers to the ISIMIP model output format, where each cell represents aggregated water storage across all lakes and reservoirs within that cell. These resolutions are therefore not inconsistent, but reflect the different nature of the datasets. We believe this is already made clear in lines 296-298 (in original and revised manuscript):

"For each evaluated reservoir, the storage time series was extracted at the grid cell corresponding to the reservoir outlet. This differs from CoSWAT, whose outputs are not gridded and are produced explicitly for each water body".

We believe that, with the adjustments in section 2.1 clarifying the consistent resolution for spatial inputs, and the text mentioned above, this is clear.

Figure 4: Add the north arrow and scale bar.

Similar to Figure 3, we have added a scale bar and a north arrow.

3.2 Reservoir and lake storage, inflow and outflow evaluation

Figure 5: Why do KGE histograms in the first subplot show a minimum value, but others do not?

As described in Section 2.6, a minimum benchmark value is only established for KGE, following Knoben et al., (2019), who define a threshold below which the model performs worse than the mean of observations. No equivalent benchmark threshold is defined for r or PBIAS, so only the minimum value is shown in the KGE histogram, whereas all other indicators have "satisfactory" and "good" thresholds.

Figure B-1: Why is the spatial distribution presented as a block-type map?

We thank the reviewer for this observation. The block-type appearance of the maps in Figure B-1 results from the location of the centroid of each reservoir/lake for which data are available to evaluate the model. Each location is plotted as a point with a square shape, as plotting each reservoir's actual shape at such a scale would not allow adequate visibility. The figure was updated with a different point shape and size to improve clarity and visibility.

***Most of the Mississippi and South American basins show KGE values below -0.4. The authors should explain the cause of this phenomenon.* | Line 340: Only 34% of basins achieve positive KGE values. What factors contribute to this low performance?**

We thank the referee for pointing out these regions and posing the question about the reasons behind the poor performance. Regarding the poor KGE values in the Mississippi and Parana river basins, the most common cause is the absence of calibration for parameters associated with hydrological processes, which leads to biases in main processes such as surface runoff and consequently produces poor inflow estimates to the water bodies, affecting the storage (As evidenced with specific cases shown). Moreover, alongside the uncalibrated status for general hydrology, specific parameters associated with the implemented schemes to simulate reservoir/lake outflow were also not calibrated, which can have a large influence, especially in cascading reservoirs (Shin et al., 2019; Vanderkelen et al., 2022), a configuration that happens relatively often in the mentioned regions. In that sense, the choice of fixed values for some of the scheme parameters across all water bodies may further contribute to this. Additionally, we cannot rule out contributions from the quality and/or resolution of the atmospheric inputs for simulations, and spatial input data to create HRUs; this could additionally have an impact on hydrological processes, and it is inherent to global applications. That is why we also find similar performance in the other global models CoSWAT was compared with.

This is now explicitly discussed on section 4.2 of the revised manuscript (lines 547 to 561):

> *Based on our comparison with other global water models, and the range of performance reported in similar large-scale evaluations (Hosseini-Moghari and Döll, 2025; Tang et al., 2025; Vanderkelen et al., 2022), the obtained storage skill is broadly comparable in magnitude and spread, indicating that the new implementations yield results as reliable as those of other state-of-the-art models. The relatively low proportion of water bodies achieving positive KGE (34%) is in line with this benchmark and reflects the combined effect of uncalibrated hydrological parameters and coarse spatial and meteorological inputs inherent to global-scale applications. Additionally, while generalized parametric outflow schemes are well-suited to this scale given their simplicity and low data requirements, they carry inherent limitations and are expected to underperform compared to more localized applications, particularly when relying on globally recommended parameter values without a site-specific parameter calibration, such as in this study. Spatially, we identify regions such as the Mississippi River System or the upper La Plata River Basin, where the combination of these factors is particularly evident, with several water bodies with a KGE below the -0.42 threshold. These regions represent clear targets where multi-objective calibration efforts are required to achieve a better representation of storage as well as hydrological processes. As expected, KGE median values remain generally low, but a substantial*

fraction exceed minimum skill thresholds, with many cases achieving a satisfactory PBIAS, reflecting a meaningful representation of storage dynamics that could improve considerably by addressing the issues discussed above.

Line 368: provide more details for these explanations (e.g., figure of a few soil moisture profiles supporting the statement (Figure B-2c).

We thank the referee for this comment, which motivated a more thorough investigation of this cluster of low-elevation reservoirs with significantly negative PBIAS. Further analysis showed that the cluster is not necessarily driven by hydrological overestimation accumulating downstream: While water yield naturally increases with drainage area along the river, neither inflow nor outflow PBIAS shows a statistically significant trend with elevation ($p > 0.05$), as now shown in Figure B-2 on the revised manuscript. Instead, the analysis of representative cases in the cluster points to multiple factors. First, storage capacities derived from GRanD or HydroLAKES drive the simulation scheme toward higher storage states than those actually managed in practice (based on the observational reference datasets). Second, in some cases, the reference observational datasets themselves show substantial variability across sources, introducing uncertainty in the evaluation (As shown in Figure B-3 on the revised manuscript). Furthermore, despite the overestimated storage, inflows actually tend to be underestimated in this cluster, and outflows are suppressed as a consequence, a behavior inherent to the implemented scheme when operating near maximum storage with limited inflow, and pointing to the need for reservoir-specific calibration of scheme parameters and, more importantly, the initial conditions of the simulation. We have modified Figure B-2 accordingly and added Figure B-3 showcasing those representative cases. The manuscript has been revised as follows (lines 377 to 387):

... However, while a clear cluster of low-elevation reservoirs (below 500 masl) exhibits substantial negative storage PBIAS (Figure B-2c), driving the observed elevation relationship, inflow and outflow PBIAS show no statistically significant relationship with elevation nor correspondence with this cluster (Figure B-2d). Further investigation of this cluster reveals three factors behind it; two representative examples are shown in Appendix B (Figure B-3). First, the reference storage capacity from GRanD and HydroLAKES, used to initialize and parametrize the simulation scheme, substantially exceeds the harmonized reference validation data (Table 1), causing simulated storage to fluctuate near this elevated capacity rather than the observed levels. Second, in some cases, the reference validation data itself shows important variability across sources, increasing evaluation uncertainty independently of model performance (Figure B-3a). Third, inflows tend to be underestimated in this cluster, which, combined with the parameterization, leads to inaccurate outflows and keeps storage near the prescribed maximum capacity with limited fluctuation.

Moreover, we extend the discussion in section 4.2 based on these findings (lines 591 – 595):

...Beyond this, the reservoir scheme parametrization itself can introduce systematic biases when reservoirs operate at levels considerably below their reported maximum capacity, as recorded in global datasets, causing the simulation scheme to start the simulation and target unrealistically high storage states, as was shown for the low-elevation cluster of reservoirs with high PBIAS as described in Section 3.2. This is further compounded by the fact that storage capacity values can vary substantially across global datasets, introducing evaluation uncertainty....

Line 374-375: Explain why Berryessa Lake shows underestimation from 1999-2004.

We thank the reviewer for this suggestion. Figure 6 was modified to better show the inflow and outflow time series, which allows for a better explanation. This is now included in the revised manuscript (lines 403 to 408):

Berryessa Lake (Figure 6c) is a regulated water body mainly used for hydroelectricity, and shows the best overall performance: storage is well reproduced except for notable underestimation during 1999-2004 and 2006-2014. The first period is preceded by a large overestimated outflow peak in 1998, causing an abrupt storage decline, after which underestimated inflows combined with overestimated low-season outflows prevent recovery to observed levels. Storage recovers by 2006, but a subsequent high inflow peak triggers renewed outflow overestimation, returning storage to underestimated levels. This case performs well on inflow and satisfactorily on outflow, yet illustrates how punctual but significant over- or underestimations of outflow can substantially affect storage representation for several years.

Line 382: The inclusion of reservoirs in Lake Oahe–Mississippi River appears to decrease streamflow performance, particularly in 2004–2009 when storage is poorly simulated. What causes this, and why do inflow and outflow metrics show large deviations (inflow: KGE = -2.03, PBIAS = -96.9%; outflow: KGE = -1.84, PBIAS = -109.3%)?

We thank the reviewer for this comment, which led to a deeper analysis of this case, and more insights on causes for poor model performance. In the revised manuscript, we have expanded the discussion of Lake Oahe and added Fort Berthold to Figure 6, which now explicitly illustrates the cascading influence of an upstream reservoir on downstream storage and inflow/outflow representation. The mechanisms behind the poor performance metrics, including inflow overestimation propagating from Fort Berthold, and the effect on storage during 2004-2009, are now described in the results section. The revised text reads as follows (lines 391 to 397):

Fort Berthold reservoir (Figure 6a) and Lake Oahe (Figure 6b) are two cascading reservoirs upstream on the Mississippi River system, primarily operated for flood control. Fort Berthold shows overestimated inflows and, consequently, outflows for much of the simulation period, which propagate downstream to Lake Oahe, where similarly overestimated outflows occur, generally causing significant storage fluctuations. During 2004-2009, storage overestimation was persistent, mainly driven by inflow overestimation that exceeded outflow overestimation. These two reservoirs illustrate two important situations: the propagation of errors through cascading reservoir systems, and cases where storage magnitude is reasonably captured but is subject to significant fluctuations due to overestimation of inflows and outflows.

This is further discussed, addressing a comment below on section 4.2 (lines 574 to 580 on the revised manuscript).

3.3 Comparison with other global models

Figure 7: Clarify why CoSWAT's distribution is inconsistent across models. Are the violin plots comparing metrics between CoSWAT and other models or between models and observation data? Also, in subplot (b), the model colors differ. Why are dark and light colors not consistently labeled in the legend to indicate metric ranges? clarify color representations in the legend.

Regarding the varying CoSWAT distribution (Figure 7a), this is a consequence of the evaluation approach: as described in lines 399-403 of the original manuscript:

"... the performance distributions shown in Figure 7 are computed using only the subset of water bodies simultaneously represented by CoSWAT and each respective ISIMIP model, so the CoSWAT sample size varies per comparison rather than always including the full set of 197 reservoirs."

The figure compares the performance of CoSWAT and the ISIMIP models evaluated against observed data. Since the ISIMIP models differ in their reservoir coverage (in terms of reservoirs in the observation data), the CoSWAT sample size varies per comparison. Nonetheless, this was further clarified in the revised manuscript (lines 417 to 427):

As an additional benchmark for the storage evaluation, performance results were compared with five global water models from the ISIMIP Global Water Sector (Figure 7). Across these models, the number of reservoirs evaluated varies widely, from 65 for CWATM to 195 for LPJmL5-7-10-FIRE, with intermediate sample sizes for H08 (78), MIROC-INTEG-LAND (183), and WaterGAP2-2e (96). In contrast, CoSWAT results encompass 197 water bodies, effectively covering the union of those represented across all models. These differences arise because grid cells with all-zero storage time series were excluded, suggesting that the reservoir with observed data is not represented in the model or that the sampled pixel does not precisely match its location.

It is important to note that Figure 7a shows violin plots computed using only the subset of water bodies simultaneously represented by CoSWAT and each respective ISIMIP model. As a result, the CoSWAT distribution differs across comparisons, since the sample varies per model pair rather than always reflecting the full set of 197 reservoirs. The bar charts in Figure 7b, in contrast, report the proportion of water bodies exceeding performance thresholds normalized by the total number of reservoirs represented by each model, thereby reflecting overall coverage and skill.

Regarding the colours in Figure 7b we believe the legend is consistent: for each model, the lighter shade represents the stricter performance threshold and the darker shade the looser one, following the same pattern across all metrics. However, the caption might not make this clear enough, so we have adjusted it as follows:

Figure 7: Performance comparison for mean monthly storage representation between CoSWAT with new implementations and five ISIMIP Global Water Sector models. Box a) shows violin plots with the KDE distribution of CoSWAT (left) and the other models (right) for KGE, PBIAS, and r, considering only the water bodies available in the ISIMIP reference model. A dashed line represents the median, while dotted lines represent the first and third quartiles. Box b) shows, for each model, the percentage of water bodies exceeding the "satisfactory" and "good" performance thresholds for KGE, PBIAS, and r, normalized by the total number of reservoirs represented by that model. For each model, the lighter shade corresponds to the "good" threshold and the darker shade to the "satisfactory" threshold. Outliers were not considered.

Line 395-396: Provide references for the reservoir-related models mentioned.

Thank you for this suggestion. We have added the references in the revised manuscript on lines 301 to 303:

To place model performance in a broader context, monthly reservoir storage was compared against five state-of-the-art global water models from the ISIMIP Global Water Sector (Gosling et al., 2024): CWATM (Burek et al., 2020), H08 (Hanasaki et al., 2018), LPJmL5-7-10-fire (Wirth et al., 2024), MIROC-INTEG-LAND (Yokohata et al., 2020), and WaterGAP2-2e (Müller Schmied et al., 2024).

3.4 Streamflow evaluation

Figure 8: Subplot (a) is redundant with Figure B-1a

Thank you for this comment. Subplot (a) of Figure 8 presents streamflow performance, while Figure B-1a shows reservoir storage performance; these are inherently different variables, and therefore the plots are not redundant.

Subplot (b): Distinguish categories 1-4 with colors or labels.

Regarding subplot (b), skill category labels are already provided on the y-axis, but we have adjusted the figure to increase the font size and weight of those labels for improved visibility, and set the box-plot transparency to change by category to further distinguish them.

Figure 9: Recommend indicating the area of the six sub-basins and adding inset maps showing sub-basin shapes and the locations of reservoirs and lakes in upstream/downstream positions.

We thank the reviewer for this great recommendation, which provides a better spatial and operational context for the figure. We have updated Figure 9 correspondingly:

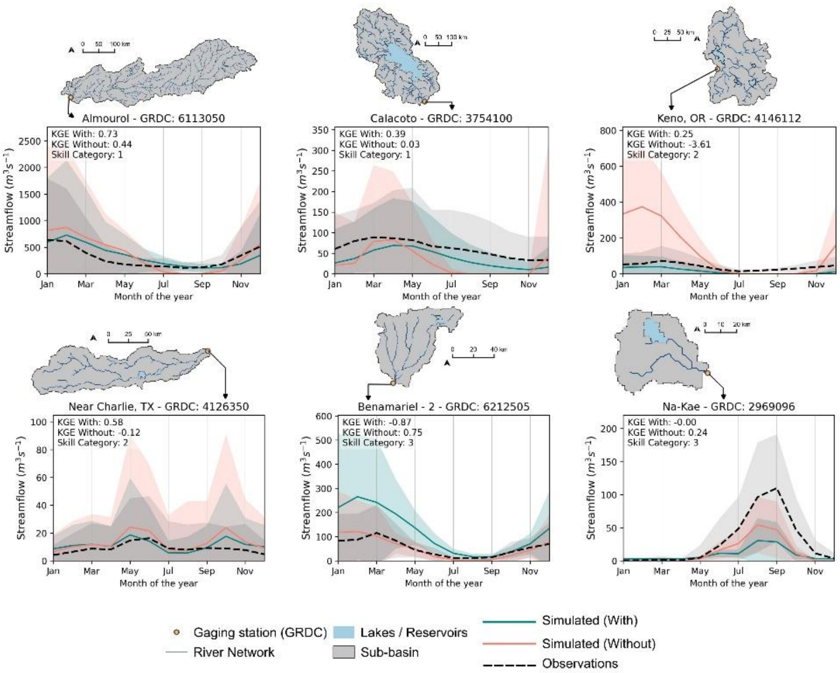


Figure 9: Seasonal monthly streamflow patterns for six selected GRDC stations, contrasting simulations with and without new implementations, and their respective sub-basin, upstream river network, and presence of lakes or reservoirs. Lines show the mean monthly flow across the analysis period, while shaded bands denote ±1.5 standard deviations. Titles indicate the GRDC station name and ID.

Line 451-452: Analyze the reasons for these discrepancies in the Mississippi and Central European regions and propose potential improvements? Figure 8 also seems to reflect this issue.

We thank the reviewer for this observation. The main cause in these regions, including the Nile Basin and specifically La Plata (South America), is yet again, an issue related to the uncalibrated status of the hydrological model, and the fact that the reservoir scheme was used with global "default" parameters, without performing a reservoir-specific calibration. Here, this is very clear because we actually see how the presence of reservoirs is improving streamflow representation, as the clusters on Category 2 are located in the mentioned regions, but performance remains poor. On the other hand, for the case of La Plata and some parts of Western Europe, it can be clearly seen that the reservoir scheme parameterization is the main issue. As illustrated in one of the examples provided in Figure 9 (Benamariel station).

First, these regions are now more clearly highlighted in the results section, trying to maintain this section mainly descriptive, lines 483 to 492:

> *A very common situation is what can be observed in the upper Mississippi River System, overall in the Colorado River Basin, Nile River Basin and most sections of Western Europe: many stations are improving (Category 2), yet their KGE value remains below the minimum threshold, a clear indication that new implementations alone are not sufficient to take performance to satisfactory levels, but an improvement is indeed achieved. In Central Europe, a large number of stations are classified as Category 1 or 2, and, as with the stations mentioned earlier, most of those in Category 2 remain below the satisfactory threshold. There are also a significant number of Category 3 and 4 stations, with the*

> *former remaining satisfactory in KGE despite reductions. The La Plata and Titicaca River Basins in South America have relatively few observations compared with the aforementioned regions, and the results vary significantly. There, upstream areas are generally improving and achieving good performance. The La Plata river basin's downstream stations, which were performing poorly without reservoirs, remain poor and classified as category 4.*

Then, the discussion about this is now more concrete and related to these specific results as part of Section 4.3., where the clear path towards correction of these biases on a large scale is applying Hydrological Mass Balance Calibration (HMBC; Chawanda et al., 2020) and calibration of the reservoir scheme parameters, lines 617 to 637:

> *Despite the improvements on stations in Category 2, the performance of most of these cases remains poor, even with improvements. This pattern is particularly evident in the upper Mississippi River System, western European basins, as well as the upper and lower Nile Basin, which suggests that these regions (dominated by Category 2, Figure 8) still require hydrological calibration to achieve adequate performance in streamflow. In that sense, for these clear clusters of poorly performing stations in which streamflow overestimation occurs, there is a need to correct this bias by adjusting parameters associated with surface runoff generation, which Hydrological Mass Balance Calibration (HMBC; Chawanda et al., 2020) could address at such a scale. .*
>
> *In regions such as the lower La Plata basin or some sections of western Europe, where we find clusters of Category 3 or 4 stations, we additionally identify the need to improve the parameterization of the reservoir schemes, given the over- and underestimation of outflow magnitude or timing, the presence of cascading systems, and the evident negative effect that their inclusion had on streamflow performance. This is shown by the cluster of streamflow stations in Category 3 on these regions (Figure 8), and illustrated clearly on cases such as Lake Oahe (Figure 6) on the Mississippi River System, Benamariel-2 station in the Esla River or Na Kae Station in the Mekong River basin (Figure 9). The case of Benamariel-2 clearly showcases the impact of a poorly parametrized reservoir upstream, as the model without it performs adequately, pointing to an overestimated outflow calculation likely caused by an underestimation of emergency volumes that often lead to conditions 4 and 5 of the simulation scheme (Table 3). This subbasin could be improved by either adjusting its parameterization with alternative data sources, as well as calibration, or by not accounting for the reservoir in the model. On the other hand, Na Kae station showcases the need for hydrological calibration of the model, given the poor performance even without reservoirs, and the further negative impact of upstream reservoirs being poorly parametrized.*

Line 482-483: Clarify whether the overestimation is caused specifically by the reservoir implementation.

We thank the reviewer for pointing this out. This was clarified in the revised manuscript (lines 505 to 507):

> *Nevertheless, this highlights that the underlying issue leading to poor performance at these stations may be an overestimation of surface runoff, present even without reservoir implementation, and therefore related to hydrological parameter calibration rather than the reservoir integration itself.*

Line 485: Justify the choice of the six representative basins.

The six representative cases were selected based on the largest change in KGE (with vs without reservoirs/lakes) within each skill category. This is now clarified on lines 511 to 515 in the revised manuscript:

To illustrate how reservoir implementation affects streamflow behavior, six representative stations were selected to capture the range of positive and negative responses, choosing those with the largest KGE change within each category and a clear, interpretable effect (Figure 9).

Line 495-496: Add figures showing reservoir configurations in the six basins with and without reservoirs for better comparison. *What is the basis for this observed phenomenon (e.g., which figure or data)? provide a deeper analysis of the reason (eg. the modeling mechanisms, the formulation and parameterization choices)*

We appreciate this suggestion. Figure 9 already presents the comparison between simulations with and without new implementations alongside observed GRDC data for each of the six stations, which is the direct basis for the statements made. Following previous comments, we have improved Figure 9, where the subbasin upstream of the gaging station, the river network, and the presence of water bodies are shown. We also expanded the results description for each station , providing more details about the context (i.e., upstream area, type of water bodies upstream). In the case of stations where we find a decrease in performance, we provide better context and reasoning. This is now on lines 515 to 530 of the revised manuscript:

... Almourol station (Tagus river, ~66,000 km^2) falls in category 1, with several hydropower (e.g., Castelo do Bode, Alcantara) and irrigation (e.g., Valdecanas) reservoirs upstream. Performance improved mainly through better representation of dry-season flows, which were heavily underestimated without reservoirs. Calacoto station (Desaguadero river, ~65,000 km^2) shows a similar pattern, here driven by the downstream influence of Lake Titicaca (~8,300 km^2 surface area), whose regulating effect was absent in the simulation without the new implementations.

Keno (Klamath River, ~18,000 km^2) and Near Charlie (Wichita River, ~8,900 km^2) both fall in category 2, showing significant performance improvements. At Keno, the absence of upstream reservoirs, including the regulated Upper Klamath Lake and the Clear Lake irrigation reservoir, leads to a strong overestimation of streamflow that largely disappears once they are included. Near Charlie presents a similar overestimation issue. While the mean climatology appears adequate, the shaded uncertainty band reveals substantial interannual overestimation, reduced considerably after introducing two upstream irrigation reservoirs, including Lake Kemp Dam. In contrast, the two Category 3 stations in Figure 9 illustrate cases where reservoir introduction degraded performance. At Benamariel (Esla river, ~4,000 km^2), overestimated outflows from an upstream irrigation reservoir amplify simulated streamflow, worsening model performance. At Na Kae (Lam Nam Kam river, ~2,400 km^2), the regulation of Nong Han Lake exaggerates an existing underestimation, further suppressing flows that were already too low without reservoirs. In both cases, the current parametrization of the outflow calculation scheme likely drives over- or underestimation of target release and release coefficients, propagating errors downstream.

Moreover, the mechanisms and reasons behind decreased performance are further discussed in section 4.3 (lines 625 to 636), as addressed in a comment below.

Line 499-500: Clarify the basis for the statement in Figure 9.

The basis for these statements is the direct comparison between the "With" and "Without" simulations in Figure 9. For Benamariel, the model already showed reasonable seasonal values without reservoirs, and adding them exaggerates the effect of upstream water bodies, leading to overestimated flows. For Na Kae, the opposite occurs: the model already underestimates streamflow without reservoirs, and the addition of lakes and reservoirs with low outflows further amplifies this underestimation. These points are now addressed in the revised text above.

4 Discussion

4.1 Integration of lakes, reservoirs, and irrigation

Line 518-521: If the irrigation module is included, validation should be provided, or remove the discussion of irrigation and focus on reservoirs and lakes.

We thank the reviewer for this suggestion. We are not including a discussion about this in the revised manuscript, giving focus to the integration of reservoirs and lakes.

4.2 Model performance of simulated reservoirs and lakes

Line 549-550: Discuss why Nasser Lake and Lake Oahe show overestimation and underestimation in certain years.

We thank the referee for the comment. This discussion, reinforced by a better presentation and description of results (Figure 6, Section 3.2), is present in the revised manuscript (lines 573 to 579):

> *The Nasser Lake and Lake Oahe examples (Figure 6) clearly show that the simulation scheme is highly sensitive to inflow representation, with overestimation of inflow leading to misrepresentation of storage and outflow. In both cases, significant storage fluctuations are driven by overestimated outflows, which can be traced back to excess landscape runoff, either arriving directly from the upstream catchment as in Lake Nasser, or through a cascading reservoir system as in Lake Oahe, where Fort Berthold propagates the error downstream. This ultimately reflects the uncalibrated state of the hydrological model and the need to better represent the catchment water balance to reduce over- or underestimation of surface runoff.*

Line 565-568: Given that the model performance does not improve in some cases after introducing reservoirs and lakes, can the authors suggest potential improvements? For example, could the number of reservoirs be adjusted based on basin characteristics (hydropower, flood control, irrigation)? In basins not dominated by hydropower, could some reservoirs or lakes be reduced? Additionally, have the authors considered evaluating the impact of reservoirs and lakes on other hydrological variables beyond streamflow (e.g., soil moisture, ET)?

We thank the reviewer for this comment. Overall, we believe the main pathways to improve model performance are Hydrological Mass Balance Calibration (HMBC; Chawanda et al., 2020), a localized, standalone calibration of reservoir scheme parameters, calibration of lake evaporation parameters (which can greatly affect the water balance in lakes with large surface area), which is mentioned throughout the Discussion section on the revised manuscript. In principle, we believe that while the introduction of reservoirs could lead to a worse streamflow performance downstream in some cases, this could be corrected by following the steps mentioned before. Nonetheless, it is still a good suggestion that we appreciate and will follow in our future work towards improving this model. Filtering out some reservoirs that are not as influential and do not represent a significant storage at the scale we are working is a good solution to reduce the "noise" they can generate. As the aim of the CoSWAT global model, beyond adequately representing streamflow, is to represent freshwater storage robustly, meaning to retain as many water bodies with significant storage as possible, we intent nonetheless balance streamflow performance with storage representation. This was added on Section 4.4 of the revised manuscript (lines 676 to 679):

> *Moreover, a potential avenue could include assessing the selective inclusion of reservoirs based on their relative importance within the basin, for instance, by excluding water bodies with limited disruptive effect and relatively low storage, which may introduce more noise than signal without meaningfully improving freshwater storage representation at the scale of the analysis. This could be done by filtering reservoirs using capacity ratio thresholds, which, with adequate threshold selection, has been shown not to compromise model realism (Shrestha et al., 2024).*

Regarding other variables, although it would be interesting to explore them, we find streamflow and reservoir water budget the most influential, related to new implementations, as we specifically aim to see how the model is able to represent freshwater body storage. This was motivated by the fact that ET was validated for the CoSWAT model previously, showing adequate results (Chawanda et al., 2025), but streamflow was one of the weaknesses, and one of the attributed factors was the non-inclusion of reservoirs. Nonetheless, we believe that exploring how the introduction of these water bodies affects other variables, especially at the global scale, is an important and interesting endeavor that is now mentioned in section 4.4 of the revised manuscript (lines 681 to 683):

> *...Additionally, future work should explore how the integration of lakes and reservoirs affects other hydrological variables such as soil moisture and evapotranspiration, which could provide a more complete picture of the hydrological implications at the global scale.*

Also, following the suggestions provided for Section 3.4, add the discussion about these issues.

We thank the referee for the comment and the suggestion. Based on improvements on Figure 8 and the description of the results around it, these new additions have been made on the revised manuscript (lines 617 to 644) as shown in an already addressed comment above.

Regarding comments and improvements on Figure 9 and its results, the following has been added to the revised manuscript:

Lines 607 to 615:

> *Improved low-flow representation is evident at Almourol (Tagus River) and Calacoto (Desaguadero River), where upstream reservoirs sustain dry-season discharge. At Almourol, this series of managed reservoirs maintains flow from May to October, while at Calacoto, the stable outflow of Lake Titicaca, a large system with a long residence time (~1300 years), is entirely absent from the simulation without the new implementations. Keno (Klamath River) and Near Charlie (Wichita River) stations illustrate the complementary effect on high flows: without the managed reservoirs present upstream, discharge is strongly overestimated. There, observations show a relatively stable year-round regime, without inter-annual high peaks, that the model with new implementations better captures, though for Keno, with some remaining underestimation, pointing again to the need for improved parametrization.*

Lines 628 to 636:

> *This is evidenced by the cluster of streamflow stations in Category 3 on these regions (Figure 8), and illustrated clearly on cases such as Lake Oahe (Figure 6) on the Mississippi River System, Benamariel-2 station in the Esla River, or Na Kae Station in the Mekong River basin (Figure 9). The case of Benamariel-2 clearly showcases the impact of a poorly parametrized reservoir upstream, as the model without it performs adequately, pointing to an overestimated outflow calculation likely caused by an underestimation of emergency volumes that often lead to conditions 4 and 5 of the simulation scheme (Table 3). This subbasin could be improved by either adjusting its parameterization with alternative data sources, as well as calibration, or by not accounting for the reservoir in the model. On the other hand, Na Kae station showcases the need for hydrological calibration of the model, given the poor performance even without reservoirs, and the further negative impact of upstream reservoirs being poorly parametrized.*

4.3 Impacts on streamflow representation

Line 575: Clarify the origin of the 70% and 42% values. Are they based on the sum of categories described in Section 3.4?

This was clarified in the revised manuscript (lines 604 to 605):

> *...Against this background, the inclusion of lakes and reservoirs generally improves simulated streamflow performance (70% of stations; proportion of stations in Category 1 and 2), with a large portion performing poorly (42%; proportion of stations in Category 2)...*

4.4 Implications and future work

Line 599-600: Provide examples of dedicated lake models for coupling with CoSWAT.

We thank the reviewer for this suggestion. Based on it, we have adjusted the manuscript on lines 657 to 660:

> *... The semi-distributed structure of SWAT+ explicitly represents individual lakes and reservoirs rather than aggregated water bodies, making it well suited to produce boundary conditions to lake models, something identified as missing on global simulations by the ISIMIP Lake Sector (Golub et al., 2022). Lake models that account for water balance in their formulation and therefore are natural candidates for this coupling are GOTM (Peng et al., 2025), LAKE (Heiskanen et al., 2015), and GLM (Hipsey et al., 2019), among others. Precedents already exist for SWAT+ being coupled with models such as GPLake-M (Nkwasa et al., 2025b), GOTM (Jiménez-Navarro et al., 2023), and ATLANTIS (Teran et al., 2026).*

Line 613: Consider adjusting reservoir numbers based on basin characteristics in future work.

Thank you for this suggestion. Based on it we have added the following text on lines 675 to 679:

> *Moreover, a potential avenue could include assessing the selective inclusion of reservoirs based on their relative importance within the basin, for instance, by excluding water bodies with limited disruptive effect and relatively low storage, which may introduce more noise than signal without meaningfully improving freshwater storage representation at the scale of the analysis. This could be done by filtering reservoirs using capacity ratio thresholds, which, with adequate threshold selection, has been shown not to compromise model realism (Shrestha et al., 2024).*

5 Conclusion

The model should not be called a "network-resolution approach" since it integrates reservoir and lake areas rather than changing network resolution.

Thank you for noting this. As mentioned before, we have adjusted the manuscript overall in relation to this to change the term to "integration" rather than "resolution". This was also adjusted on the conclusion (lines 690 to 694):

> *This study introduces methodological advances to improve the representation of lakes, reservoirs, and irrigation in the CoSWAT Framework and the global CoSWAT model.* ***A new reservoir/lake integration approach****, irrigation representation, and globally applicable reservoir and lake schemes that combine the capabilities of SWAT+ with state-of-the-art approaches used in global models were implemented.*

Line 625-627: Add quantitative indicators (e.g., performance metrics) to the conclusion.

We appreciate this suggestion. The conclusion was modified to include more quantitative indicators (lines 694 to 699):

Simulations were performed to assess the model's capability with and without the new implementations. The evaluation shows that storage dynamics are simulated with performance comparable to other global water models, with 71% of cases surpassing the minimum KGE threshold (-0.42), but only 34% achieving a positive KGE. With new implementations, streamflow performance improved at 70% of stations where reservoir influence is relevant, with 42% of stations showing clear and considerable improvements that nonetheless remain below satisfactory levels in absolute terms.

Line 628: The statement about improved low-flow control is not supported by the results, as some representative basins (Section 3.2, Figure 6b) still show significant low-flow underestimation. Discuss potential causes.

We appreciate this comment. This statement was partially supported, with cases such as the Almourol and Calacoto in the Tagus River Basin and Lake Titicaca Basin. However, it is true that in other cases there are underestimations. The reasons behind this are now discussed in section 4.3, addressing another comment above, and this statement is not part of the conclusion anymore.

Referee 2

This manuscript presents improvements applied to the global CoSWAT modelling approach, focused on improving the representation of human actions, which emphasis on reservoirs. The manuscript is clear and well structured, and the methods are appropriate and include a comparison with other global models. The main limitation of the study is related to the no calibration of the model, but it is acknowledged and proposed for further work. Improving the inflow calibration will surely lead to better performance for reservoir variables and streamflow, but considering the scale of work, the current advances presented in this manuscript already suppose a novelty and an improvement.

I consider the manuscript suits the journal and it is already in a very good shape, so I can only provide few suggestions/comments to improve it if the authors agree.

We thank you very much for your general assessment. The fact that the model is not yet calibrated is indeed a big limitation, and it is a key objective of our future work. We definitely take into account the comments and suggestions to make improvements in the revised version.

Line 44. Fix the phrase

Thank you for the suggestion. We have adjusted this section on the revised manuscript(lines 47-48):

> *Drying and wetting trends are intensifying (Bai et al., 2024), accompanied by ecosystem degradation (Grant et al., 2021; La Fuente et al., 2024; Woolway and Merchant, 2019).*

Line 96 and 103. You have more recent examples and using SWAT+ (e.g., https://link.springer.com/article/10.1007/s11269-024-04071-9).

This is a great example that we added to the revised manuscript on the suggested lines.

Lines 100 to 101:

> *It has been proven that the implementation of reservoirs, lakes, and management practices significantly affects the performance of hydrological models (Chawanda et al., 2020; Sánchez-Gómez et al., 2025; Zajac et al., 2017)*

Lines 109 to 110k:

> *Although large-scale studies have demonstrated that reservoir operation decision tables can be derived and applied successfully (Chawanda et al., 2020a; Sánchez-Gómez et al., 2025; Wu et al., 2020) these approaches rely heavily on reference data.*

Line 98. The fact that the model has not been calibrated may also explain that.

Thank you for noting that. Indeed, this is true; we have added this to the revised manuscript (lines 104 to 105):

> *This, combined with a lack of representation of water management (e.g., agricultural practices), and the fact that the model has not been calibrated, are causes of poor model performance for streamflow in many locations (Chawanda et al., 2025).*

Line 230. Current storage is also an important factor for this method.

This is a good point, and we thank the referee for noting that. The revised manuscript on this section (lines 235 to 236) was adjusted:

> *...The key to its approach is defining a release target and actual release based on past inflows and/or irrigation demands, reservoir properties, and the simulated storage at each time step.*

Line 277. Is there not available data to initialize the reservoirs more realistically? Starting with the reservoirs almost full may overestimate storage during a long period.

We thank the referee for this comment. In principle, observations from the reference datasets (e.g., ResOpsUs) could be used to derive more realistic initial conditions. However, to our knowledge, SWAT+ does not currently support individually defined initial storage conditions per reservoir. We therefore use the om_water.ini file with volume set to maximum (vol = 1) for all water bodies, following Vanderkelen et al., (2022) who adopted a similar approach. Defining unique initial conditions per reservoir is, nevertheless, a valuable direction for future work, which could allow better initialization.

This was proven to be an issue in some cases after further investigating reservoirs with large biases, fully evidencing the concert with overestimation of storage, now described in the revised manuscript (lines 376 to 387):

> *… However, while a clear cluster of low-elevation reservoirs (below 500 masl) exhibits substantial negative storage PBIAS (Figure B-2c), driving the observed elevation relationship, inflow and outflow PBIAS show no statistically significant relationship with elevation nor correspondence with this cluster (Figure B-2d). Further investigation of this cluster reveals three factors behind it; two representative examples are shown in Appendix B (Figure B-3). First, the reference storage capacity from GRanD and HydroLAKES, used to initialize and parameterize the simulation scheme (Figures B-3a and B-3b), substantially exceeds the harmonized reference validation data (Table 1), causing simulated storage to fluctuate near this elevated capacity rather than the observed levels. Second, in some cases, the reference validation data itself shows important variability across sources, increasing evaluation uncertainty independently of model performance (Figure B-3a). Third, some cases in this cluster have significantly underestimated inflows, which, combined with the parametrization, lead to inaccurate outflows and keep storage near the prescribed maximum capacity with limited fluctuation, leading to the negative values of PBIAS.*

However, again, we found ourselves limited to set up individual initial conditions, and considering that there is also significant variability in the reference datasets used to perform validation, as also mentioned in the revised manuscript (lines 589 to 594):

> *… Beyond this, the reservoir scheme parametrization itself can introduce systematic biases when reservoirs operate at levels considerably below their reported maximum capacity, as recorded in global datasets, causing the simulation scheme to start the simulation and target unrealistically high storage states, as was shown for the low-elevation cluster of reservoirs with high PBIAS as described in Section 3.2. This is further compounded by the fact that storage capacity values can vary substantially across global datasets, introducing evaluation uncertainty.*

The possibility of setting up better initial conditions in the model is something that can greatly improve the model performance, but could also mean reducing the warm-up period, particularly when thinking of calibration efforts. This not only applies to reservoirs, but also to all components in the system. This is now integrated in Section 4.4 of the revised manuscript (lines 671 to 675):

> *… Another proposed development is extending SWAT+ to support prescribed initial conditions for individual water bodies, as the current implementation applies a single common initial condition across all elements. This would allow better initialization of reservoir storage states, addressing some of the issues identified in this study, while also reducing the warm-up period required for other model elements, which could considerably reduce computation time during calibration efforts*

Line 280. These thresholds are questionable.

The KGE threshold of -0.42 follows Knoben et al. (2019), who define this as the point at which a model outperforms the mean of observations. We use this as a minimum acceptable value, meaning the model must at least outperform that benchmark, but we set a positive KGE as the objective for satisfactory performance and 0.4 as good performance, as done in similar applications (Gutenson et al., 2020; Hosseini-Moghari & Döll, 2025). We acknowledge that these thresholds might differ from those commonly used in smaller-scale applications, but we believe they are appropriate given the scale and inherent limitations of a global model with coarse input data, as well as the context and goal of the study.

Figure 5. Inflow simulation and its overestimation clearly affects the simulation of outflow and storage. In part, because the method you are using consider the inflow magnitude to determine the release. I wonder if the decision of running the warm-up period with all the reservoirs as natural lakes is in part responsible for this, especially for cascading reservoirs.

We thank the referee for this insightful comment. Running the warm-up period with reservoirs as natural lakes is a necessary step to populate the reservoir "memory" of past inflows and demands before the H06 scheme begins operating, which then gets continuously updated as the simulation advances. To illustrate an interesting case, we provide an example (Three "chained" reservoirs in the downstream area of the Mississippi River System) on Figure 1R, where all of those systems are considered regulated and therefore start as natural lakes during the warm-up period and transition to the H06 scheme afterward.

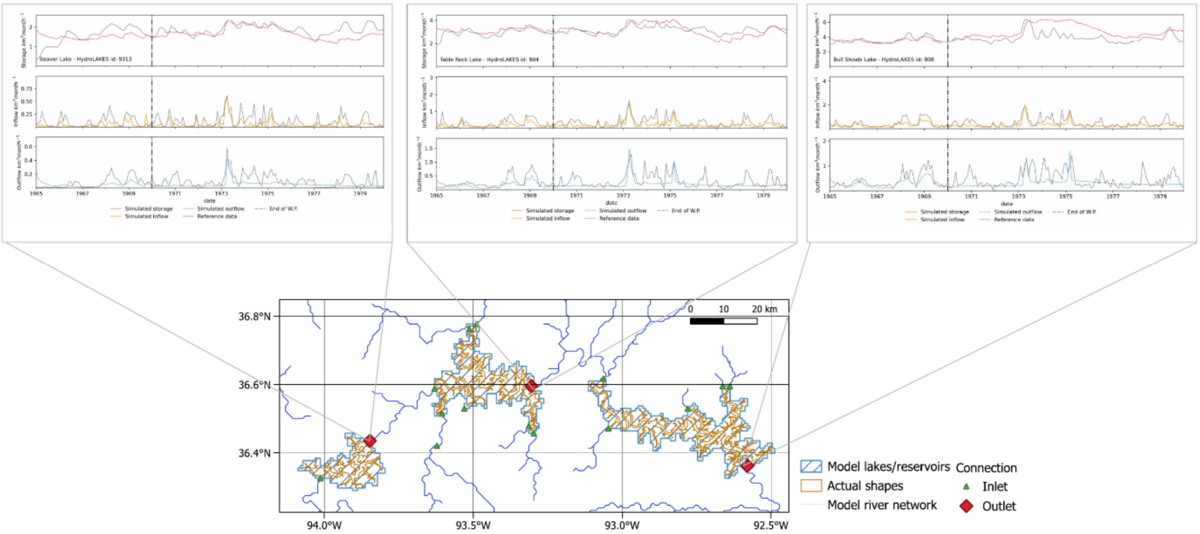


Figure 1R: Storage, inflow, and outflow results, including the warm-up period (W.P.) for three consecutive reservoirs in the Mississippi River System.

What can be seen is that the simple natural scheme produces relatively stable storage that remains close to long-term average conditions (based on the initial storage set based on the global datasets used as inputs), which is precisely what we aim for, so that the H06 scheme initializes at an appropriate level. Looking more closely at the relationships between reservoirs, an inadequately represented outflow from one reservoir after the warm-up period (i.e., with the H06 scheme) can have a comparable impact to simulating it as a natural lake during warm-up. This points back to the broader limitation of the model: beyond general hydrological parameters, individual reservoir scheme parameters are not calibrated, which is likely consequential, especially for cascading reservoirs (Shin et al., 2019; Vanderkelen et al., 2022). *We therefore consider the current initialization method our best available option, with possible refinements to the initial storage conditions at the very beginning of the simulation using observations, as noted in a previous comment. Parameter calibration is the path to follow for significant improvements.*

What is important is that this can bring us back to a previous comment regarding initial conditions. If the reservoir fluctuates in a more stable way near the initial condition (in principle, not an issue), but the initial conditions are far from actual operational levels, that could be problematic and spread errors along cascading systems. The new discussion about this issue was integrated into the revised manuscript, as mentioned in the corresponding comment above.

Line 438. I would consider changing 'remains similar' with 'was slightly reduced'.

Thank you for noting this. That is true and more transparent. This was adjusted accordingly (line 462):

… while the number of stations with satisfactory performance (KGE > 0) slightly reduces (44 to 40) …

References

Chawanda, C. J., Arnold, J., Thiery, W., & Van Griensven, A. (2020). Mass balance calibration and reservoir representations for large-scale hydrological impact studies using SWAT+. Climatic Change, 163(3), 1307–1327. https://doi.org/10.1007/s10584-020-02924-x

Chawanda, C. J., Van Griensven, A., Nkwasa, A., Teran Orsini, J. P., Jeong, J., Choi, S.-K., Srinivasan, R., & Arnold, J. G. (2025). CoSWAT Model v1: A high-resolution global SWAT+ hydrological model. Hydrology and Earth System Sciences, 29(23), 6901–6916. https://doi.org/10.5194/hess-29-6901-2025

Gharari, S., Vanderkelen, I., Tefs, A., Mizukami, N., Kluzek, E., Stadnyk, T., Lawrence, D., & Clark, M. P. (2024). A Flexible Framework for Simulating the Water Balance of Lakes and Reservoirs From Local to Global Scales: mizuRoute-Lake. Water Resources Research, 60(5), e2022WR032400. https://doi.org/10.1029/2022WR032400

Gutenson, J. L., Tavakoly, A. A., Wahl, M. D., & Follum, M. L. (2020). Comparison of generalized non-data-driven lake and reservoir routing models for global-scale hydrologic forecasting of reservoir outflow at diurnal time steps. Hydrology and Earth System Sciences, 24(5), 2711–2729. https://doi.org/10.5194/hess-24-2711-2020

Hosseini-Moghari, S.-M., & Döll, P. (2025). The value of observed reservoir storage anomalies for improving the simulation of reservoir dynamics in large-scale hydrological models. Hydrology and Earth System Sciences, 29(17), 4073–4092. https://doi.org/10.5194/hess-29-4073-2025

Knoben, W. J. M., Freer, J. E., & Woods, R. A. (2019). Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. Hydrology and Earth System Sciences, 23(10), 4323–4331. https://doi.org/10.5194/hess-23-4323-2019

Lehner, B., Verdin, K., & Jarvis, A. (2008). New Global Hydrography Derived From Spaceborne Elevation Data. Eos, Transactions American Geophysical Union, 89(10), 93–94. https://doi.org/10.1029/2008EO100001

Müller Schmied, H., Cáceres, D., Eisner, S., Flörke, M., Herbert, C., Niemann, C., Peiris, T. A., Popat, E., Portmann, F. T., Reinecke, R., Schumacher, M., Shadkam, S., Telteu, C.-E., Trautmann, T., & Döll, P. (2021). The global water resources and use model WaterGAP v2.2d: Model description and evaluation. Geoscientific Model Development, 14(2), 1037–1079. https://doi.org/10.5194/gmd-14-1037-2021

Sánchez-Gómez, A., Arnold, J. G., Bieger, K., Čerkasova, N., Sammons, N. B., Martínez-Pérez, S., & Molina-Navarro, E. (2025). Modelling Water Management using SWAT+: Application of Reservoirs Release Tables and the New Water Allocation Module in a Highly Managed River Basin. Water Resources Management, 39(5), 2357–2399. https://doi.org/10.1007/s11269-024-04071-9

Shin, S., Pokhrel, Y., & Miguez-Macho, G. (2019). High-Resolution Modeling of Reservoir Release and Storage Dynamics at the Continental Scale. Water Resources Research, 55(1), 787–810. https://doi.org/10.1029/2018WR023025

Vanderkelen, I., Gharari, S., Mizukami, N., Clark, M. P., Lawrence, D. M., Swenson, S., Pokhrel, Y., Hanasaki, N., Van Griensven, A., & Thiery, W. (2022). Evaluating a reservoir parametrization in the vector-based global routing model mizuRoute (v2.0.1) for Earth system model coupling. Geoscientific Model Development, 15(10), 4163–4192. https://doi.org/10.5194/gmd-15-4163-2022

Zajac, Z., Revilla-Romero, B., Salamon, P., Burek, P., Hirpa, F. A., & Beck, H. (2017). The impact of lake and reservoir parameterization on global streamflow simulation. Journal of Hydrology, 548, 552–568. https://doi.org/10.1016/j.jhydrol.2017.03.022