Response to the referee comments (RCs)

Anonymous Referee #4

This study presents an extensive database of reservoir storage timeseries for the Southeast Asian reservoirs from 1985-2023. Indeed, this paper is methodologically rigorous and can be integrated with hydrological models for the validation of their reservoir operations. However, the methodological framework section of this article needs to be significantly modified before publication to improve readability. Therefore, overall, the present version is not acceptable for publication, and I recommend a major revision for this manuscript.

A: We thank the reviewer for the positive feedback. We will carefully address all your comments to strengthen the manuscript.

Below are the comments/questions for the authors.

Abstract

1. R: Line 15 – 16: "*The 185 reservoirs collectively store around 175 km³ (140 km³ – 210 km³) of water, covering an aggregated area of 8,700 km² (6,500 km² – 10,000 km²)*". What I understood from the manuscript is that the present total water storage (year 2023) from 185 reservoirs is 175 km^3. Please reflect the year in the sentence, otherwise it is confusing that which year we are referring to. Further, what is this 140 km^3 and 210 km^3 range, which I couldn't find in the manuscript? Same about the reservoir aggregated area. The area values are not described in the manuscript, especially the area range (6,500 km² – 10,000 km²). Please explain all these clearly to avoid confusion.

   A: The range inside the brackets represents the minimum and maximum values for the total storage and area in the year 2023. We will clearly mention this in our revised manuscript to avoid any further confusion.

2. R: Line 17: "*average reservoir storage has increased from 70 km³ to 160 km³ (+130%) from 2008 to 2017*". Why the reservoir storage change from 2008 to 2017 has been considered instead of the timeframe of the database? Any specific reason for that consideration? If no, it is better to show the change from 1985 to 2023.

   A: Here, we highlighted the most significant change in reservoir storage, which occurred during the period 2008-2017, when the majority of large dams were built. Nonetheless, we will also show the change from 1985 to 2023.

3. R: Line 18 – 21: "*Our in-situ validation provides a good match between estimated storage and in-situ observations, with 60% of the validation sites (12 out of 20) showing an R² > 0.65 and an average nRMSE < 15%. The indirect validation (based on altimetry-converted storage) shows even better results, with an R² > 0.7 and an average nRMSE < 12% for 70% (14 out of 20) of the reservoirs*". For in-situ validation, reference R^2 value was 0.65, whereas for indirect validation, the R^2 value was kept at 0.7. Please unify the reference R^2 value as 0.7 as like in the main text so that 10 out 20 stations (50% of the validation sites) show good agreement with in-situ observations. Rewrite the entire sentence accordingly.

A: Thank you for your suggestion. We will rewrite these sentences accordingly, keeping the R^2 value of 0.70 and the nRMSE value of 20% as the reference.

4. R: Line 21: "*2019-2020 drought event*". Where has this drought happened? Please specify in the sentence.

   A: The drought covered the entire region of Mainland Southeast Asia, albeit with different intensities. We will mention this point in the revised manuscript.

5. R: Line 25 – 26: "*possibility of applications in other parts of the world*". Briefly explain how this dataset will be applicable in other parts of the world in the Results section.

   A: We would like to clarify that the possibility of global applicability was in the context of the method to estimate the storage time series and not the derived dataset. We will clarify this aspect in the revised manuscript.

6. R: Line 26: "*MSEA-Res database*". This abbreviation is using for the first time in the abstract. It should be clearly explained in the earlier part of the manuscript.

   A: We will mention it clearly in the revised manuscript.

7. R: Overall: The manuscript title and abstract does not clarify regarding the methodology used to estimate the storage timeseries in MSEA. The readers need to proceed to further sections of the manuscript to understand that they have used remote sensing data to estimate the reservoir storage. It should be further emphasized in the title and abstract.

   A: Thanks for the suggestion. We will modify the title and abstract accordingly.

Introduction

1. R: Line 29 – 30: "*influencing the redistribution of water*". It should be rewritten as "influencing the distribution of water".

   A: We will modify the sentence accordingly.

2. R: Line 44: The term "Mainland Southeast Asia" has been abbreviated as MSEA in the abstract. Hence, when it appears for the first time in the Introduction, it should be again fully spelled and abbreviated, which has not been done. The MSEA abbreviation can be seen in line number 93 for the first time in Introduction without fully spelled. Also, it has been presented as "Mainland Southeast Asia" many times although it has been abbreviated. Please maintain the consistency throughout the manuscript by defining abbreviation in the first place when they appear.

   A: We will spell out the full form, 'Mainland Southeast Asia,' when introducing the abbreviation 'MSEA' for the first time in each section. While 'MSEA' is used throughout the manuscript, we have occasionally opted to use the full term 'Mainland Southeast Asia' within paragraphs to enhance readability. This approach helps prevent readers from losing track of the abbreviation, avoiding the need to revisit earlier sections for its full form, which can be distracting. In the revised manuscript, we will ensure that the use of the full form and abbreviation is consistent and systematic.

3. Line 51: "*where a few large rivers flowing in the region originate*". What the authors meant by saying "in the region originate"? Please rewrite to make it clear.

   A: Here, we meant the large rivers that originate and flow in Mainland Southeast Asia, such as the Mekong, Salween, Red, etc. We will clarify the sentence in the revised manuscript.

4. R: Line 55: Add Hanasaki et al. (2008), which is one of the pioneering works to include reservoir operation in global hydrological models. https://doi.org/10.1016/j.jhydrol.2005.11.011.

   A: Thanks. We will add the suggested reference in the revised manuscript.

5. R: Line 97 – 105: The additional data that the authors offer from Hou et al. (2024) is some reservoirs in the MSEA region with 3-year extra data from 2020. How significantly different this work is from Hou et al. (2024)? I recommend the authors to validate the storage database against the ones available in Hou et al. (2024) and discuss further whether they are comparable or not. If different, please explain why it has happened. If similar, again explanations are needed on why this new dataset is relevant.

   A: We have mentioned briefly in the introduction section on the need for our dataset over Hou et al. (2024). Please see the following paragraph:

   "*L87-95: Although Hou et al. (2024) cover the entire globe by providing a comprehensive dataset for large-scale assessments, it has a few limitations for the reservoirs located in Mainland Southeast Asia. First, the model parameters (used in the storage estimation) strongly depend on mean depth (extrapolating the surrounding topographical slope towards the centre of the lake to estimate lake depth), the surface area of the lake (derived from Landsat satellite images), and average slope (derived from DEM). Therefore, uncertainties in the estimates of reservoir storage may be generated by the estimation of depth, slope, and other model coefficients. Second, GloLakes does not include some of the largest reservoirs in MSEA, including Nuozhadu (22 km3), Xiaowan (15 km3), Xe Kaman 1 (4 km3), and Lower Seasan 2 (6 km3), which play a significant role in water redistribution and hydropower generation (Ang et al., 2024; Galelli et al., 2022; Vu et al., 2022).*"

   Nonetheless, we will validate the storage database against the ones available in Hou et al. (2024) and include them in the revised manuscript.

6. R: Line 107: 2019-2020 drought has been mentioned here as well. Clearly specify where this drought has happened.

   A: The drought covered the entire region of Mainland Southeast Asia, but with different intensities.

Water reservoirs in Mainland Southeast Asia

1. R: Table 1: the GDAT and dams in the Mekong links are not working. Please check it.

   A: Thanks for pointing this out. We will check and update this in the revised manuscript.

2. R: Figure 1: From Fig. 1(a), what I understood is that both Bhumibol and Sirikit reservoirs in the Chao Phraya basin have a reservoir storage ranging between 9000-13000 km^3. However, this is not true for Bhumibol, whose is storage is greater than 13000 km^3. Be careful about the storage capacities of all the reservoirs. Besides, Pasak dam is missing in the same basin that is included in the GRanD database. Any reason for leaving out this reservoir?

   A: We will reclassify the storage ranges in Figure 1a for better representation. You are right, the storage capacity of Bhumibol is 13,462 MCM, which is slightly higher than 13,000 MCM.

   Thank you for bringing this important detail to our attention. We sincerely appreciate your observation and will include the Pasak Dam in our revised database. It seems we may have inadvertently overlooked it during the data download process, even though it is part of the GRanD database (Lehner et al., 2011). We are grateful for your feedback and will address this in the updated version.

3. R: Figure 1 caption (Line 146): "*Basin-wise distribution of dam location (red dots), stream network, and order*". It should be rewritten as "Basin-wise distribution of dam location (red dots), stream network in the respective catchments, and stream order".

   A: Thanks for the suggestion. We will modify the caption in the revised manuscript.

Methodological Framework

1. R: Line 156: What the authors meant by sub-monthly here? How many times the data is available in a month? Please describe the minimum and maximum based on all the catchments.

   A: With the expression "sub-monthly" we referred to a 10-day interval period. Therefore, the data are available three times (i.e. three images) per month. We will clarify this point in the revised paper.

2. R: Line 156: It has been mentioned that the authors have used both Landsat and Sentinel-2 satellite data. However, the spatial resolution of Sentinel-2 is 10m and hence uncertainty and accuracy attribution will be very different for both imageries. How did the authors solve this issue? What treatment has done for the Sentinel-2 data to match its spatial resolution as that of Landsat?

   A: We used the surface reflectance bands from Landsat and Sentinel-2 to estimate NDWI, which are atmospherically-corrected (and ready-to-use) by NASA and ESA, respectively. As for the spatial resolution, we have re-gridded (with bilinear interpolation) the 10 m Sentinel-2 images to 30 m resolution, to make them consistent with the Landsat resolution. All the images (including DEM) are processed at 30 m resolution. There might be some uncertainty and accuracy mismatch between reflectance bands in Landsat and Sentinel-2.

   R: Figure 2: The reservoir maximum extent is not an input data, instead, a derived product after processing. What about the NDWI images? Is it acquired or derived? In some parts of the manuscript, it is mentioned that the NDWI has been acquired, while in some parts it says as derived. Please clearly state which way has been used to get

NDWI images. If the study uses both acquired and derived NDWI images, clearly state that when and where the derived products has used. Change Fig. 2 accordingly. Also show the steps of methodological framework in Fig. 2 for better understanding.

A: We derived the reservoir's maximum extent and NDWI images from the GEE environment. We will clearly mention in our revised manuscript that the maximum water extent, frequency map, and NDWI images are derived, whereas DEM is acquired (downloaded) using GEE. Since these three datasets are primarily required for storage estimation, we kept them in the input data ('Inputs') category.

3. R: Line 164 – 165: What is the meaning of time series satellite images? Does it mean a series of imageries?

   A: You are right. Here 'time series satellite images' means a (time) series of satellite images in a given period. We will revise the text accordingly.

4. R: Line 169 – 170: What is water frequency raster and maximum water extent raster? What is their meaning? In the later stage of manuscript, I understood that these are derived products. Then, why the authors have mentioned them as acquired input dataset?

   A: For each reservoir, the water frequency (FREQ) raster is a 30 m gridded image that indicates the probability of the presence of water (%) at each grid (or pixel) in a given period. The maximum water extent (EXT) raster is a 30 m gridded image indicating the maximum spread of water in the reservoir. Please note that the FREQ, EXT, DEM, and NDWI images are of the same dimension and geographically referenced to the same reservoir.

   We will correct the term from 'acquired' to 'derived' for FREQ, EXT, and NDWI.

5. R: Line 175 – 176: Need further information that which satellites products have used to create NDWI, water frequency raster and maximum water extent raster. Add this information in Table 2.

   A: Thanks for the suggestion. We mentioned in Lines 180-183 that Green and NIR bands are used to estimate NDWI followed by water frequency raster and maximum water extent raster. We will now add the band specifications for different satellite products in Table 2.

6. R: Line 180: The bands (green, red, and NIR) changes with satellite products and hence the NDWI formula. How to generalise these bands for all sensors? Please rewrite correctly.

   A: The NDWI formula is based on Green and NIR reflectance bands (wavelength range), which are generalized in nature. Since we are specifically mentioning the type of wavelength range (i.e. Green and NIR in this case) and not the band number, the NDWI formula still holds true and works for different satellite sensors.

7. R: Line 185: Choosing images with cloud coverage less than 80% is not a good idea. What if the cloud coverage is 89% and is exactly over the reservoir extent? What information the authors can access from such an image and how do you treat the

image further? I suggest the authors to further mask the satellite image for the reservoir extent and apply the cloud coverage threshold, preferably below 20%.

A: Thanks for the suggestion. We actually did something similar to what you suggested. The 80% cloud threshold is applied to the entire image. In the image pre-processing step, we further masked the satellite images (less than 80% cloud-affected area) for the reservoir extent and applied the cloud coverage threshold below 20% for estimating the frequency raster. We also filled the cloud-masked portion of the satellite image (NDWI) before applying the water classification algorithm. We better described these steps in the revised manuscript.

8. R: Line 186: The revisit time of Landsat is generally 16 days. So, how did the authors make composites at 10-day intervals? Is it achieved by combing Landsat 9 and Landsat 8, whose combined temporal resolution is 8 days at the mid-latitudes. Such explanations are missing in the manuscript. I presume there could be some months without any data. How did the authors generate data for those months?

   A: Landsat has a 16-day revisit time; however, more than one Landsat mission has been active in the time domain (except for the pre-1999 period). For instance, 2013 will have active sensors from the Landsat-7 ETM+ and Landsat-8 series of satellites, making it possible to achieve image composite at an interval of 10 days.

   Yes, you are right. There could be some months without any data (Level-0 and Level-1); we generated data for those months by interpolation (Level-2). We will further elaborate on this in the revised manuscript.

   R: Line 193: How the composite NDWI has been created? For example, if we have three NDWI images with a grid cell having values of 0, 1, and 0. What will be the composite value? Is the FREQ for that particular grid 33.3, which means one-third of the grid is covered with water? Please clearly explain this in the main text or in the supplement. Further, I do not understand the EXT layer calculation. How is it calculated? How to derive the largest extent of ones from binary NDWI images? What I understood is a single FREQ and EXT maps are created for the entire period (2013-2023). Is it true? All such technical details should be clarified in the text with further details.

   A: Let us clarify this point: The NDWI composite is the average of NDWI images in a given time interval (10 days in our case). For example, if we have three NDWI images with a grid cell having values of 0, 1, and 0, then the NDWI value in the composite image will be 0.33. Please note that there can be a maximum of three composite images in each month.

   Yes, you are right. The FREQ value of the grid cell having values of 0, 1, and 0 will be 33.3. Please note that there can be only one FREQ raster (image), which is derived by averaging all the binary NDWI images (cloud percentage <20%) available over the reservoir.

   On the other hand, the EXT layer is created by taking the largest extent of ones in all binary NDWI images available between 2013 and 2023. For example, if we have three NDWI images with a grid cell having values of 0, 1, and 0, then the EXT value will be 1 for that grid. We will clearly mention these technical details in the revised manuscript.

9. R: Line 194: To generate the NDWI, the authors used a cloud coverage threshold below 80%. But in later stages, a threshold of 20% was used to derive FREQ and EXT maps from NDWI images. Why it has to be different? Further, how the NDWI images have cloud coverage because I suppose it is already a processed image after cloud coverage removal?

   A: The NDWI images with 80% or less cloud coverage are being processed for filling and correction before estimating the water surface area. In contrast, only high-quality NDWI images (with a cloud coverage threshold of 20% or less) have been used to derive the FREQ and EXT maps.

   Removing cloud coverage requires masking the cloud pixels with a no-data value. In the subsequent steps, the cloud-marked (no-data) pixels are classified as either water or non-water (binary) using the FREQ and EXT rasters.

10. R: Line 197: What is scene-based NDWI image?

    A: We estimated the water surface area using NDWI images, which represent individual scenes from satellite imagery. This is why we refer to them as scene-based NDWI images.

11. R: Line 190: Clearly rewrite the entire paragraph.

    A: We will rewrite the paragraph for greater clarity, particularly incorporating the following explanation in the revised manuscript:

    "*The NDWI composite represents the average of NDWI values over a specified time interval (10 days in our case). For example, if three NDWI images have a grid cell with values of 0, 1, and 0, the NDWI value in the composite image will be 0.33, and the FREQ value for that grid cell will be 33.3%. On the other hand, the EXT layer is created by identifying the largest extent of '1s' across all binary NDWI images. For instance, if three NDWI images have a grid cell with values of 0, 1, and 0, the EXT value for that grid cell will be 1.*"

12. R: Line 200: Did the authors compare the derived area-elevation-storage curve against observed curve? I believe it is very crucial to validate these curves because they form the heart of this study.

    A: Unfortunately, no observed data are available for the area-elevation-storage curve. However, since these curves are derived from the DEM, we can still confidently rely on them, particularly for reservoirs constructed after the acquisition of the DEM (i.e., on or after the year 2000). Note that about 70% of the reservoirs fall in this category. We also note that it is common practice to derive the area-elevation-storage curve using DEM (e.g., Vu et al. 2022; Li et al., 2023). On the other hand, some uncertainty is expected with the reference area-elevation-storage curve obtained from Hao et al. (2024), which is used for reservoirs built before the year 2000 and accounted for only 30% of the reservoirs.

13. R: Line 213: Why have the authors taken the A-E-S curves from Hao et al. (2024)? Does this dataset have any specific advantages over other existing datasets?

    A: We specifically used the A-E-S curves from Hao et al. (2024) for reservoirs built before the year 2000, as DEM data were not available prior to 2000. Before Hao et al. (2024), reservoir bathymetry (and thus A-E-S curves) was typically derived using geostatistical modeling approaches based on simplified geometrical assumptions and/or higher-order extrapolation techniques (Hou et al., 2024; Khazaei et al., 2022). These methods were more challenging and prone to greater uncertainty, especially when applied to a group of reservoirs with varying characteristics. Another option could be to extrapolate the hypsometric curve below the water surface (Schaperow et al., 2019; Liu et al., 2020), which may not be very reliable. The other option is to use other datasets of bathymetry, but Hao is the only one covering all dams in the GranD. Therefore, we adopted a more robust hypsometric database derived using deep learning-based bathymetry reconstruction (Hao et al., 2024).

14. R: Line 217: Is this the water surface area when the reservoir is full or area timeseries?

    A: This is the time series of water surface area derived from different NDWI images. There may be some instances when the water surface corresponds to the full reservoir level.

15. R: Line 222: Why CLAHE operates in a small region? How to choose this operational window?

    A: We selected CLAHE and determined the size of its operational window (8 x 8) based on the literature (Asghar et al., 2023), which suggests that CLAHE enhances the contrast and texture features of water, thereby improving the visualization of satellite images. This enhancement facilitates the classification of water and non-water pixels.

    We will also add the above texts in the revised manuscript as a justification for using CLAHE.

16. R: Line 223: How is the surface area calculated? How the k-means clustering is useful here? It is not clear. Further explanations are needed.

    A: We used an unsupervised classification technique (k-means clustering) to classify water pixels from NDWI images. This method does not require any training data for operation. During the pre-processing steps, each image was masked (assigned an arbitrary value, such as -1) to represent the maximum reservoir extent. In the NDWI images, higher (brighter) values indicate water. By setting the number of clusters (k) to three, the clusters correspond to water (highest cluster mean), non-water, and no data (cluster mean = -1). The area of the cluster with the highest cluster mean represents the water surface area of the reservoir. We will include additional details in the manuscript to explain how the surface area is calculated. You can also refer to Vu et al., (2022) for more details, which is the reference for developing our methodology.

    Reference: Vu, D. T., Dang, T. D., Galelli, S., and Hossain, F.: Satellite observations reveal 13 years of reservoir filling strategies, operating rules, and hydrological *alterations in the Upper Mekong River basin, Hydrology and Earth System Sciences, 26, 2345–2364, https://doi.org/10.5194/hess-26-2345-2022, 2022.*

17. R: Line 245: How the Level-2 data is generated? It has been mentioned that using a trend-preserving interpolation technique. What is it?

    A: Level-2 data are generated using the interpolation technique. Instead of using linear interpolation, we have used the 'spline' interpolation technique, which fits a non-linear function to the data, thus preserving the curvature between two points (trend-preserving interpolation technique). To avoid potential confusion, we will remove the term 'trend-preserving' and refer to it simply as the 'spline' interpolation technique.

18. R: Overall: The overall clarity and logical flow are missing in this section. Further rewriting with clear explanations on the technical details are needed.

    A: Thank you. We will improve the overall clarity of this section, and will add the suggested technical details in our revised manuscript.

Results

1. R: Line 290: In Table 3, it has been mentioned as area-level-storage, while in text wrote as area-elevation-storage. Please unify.

   A: Thank you for pointing this out. We will uniformly use the term 'area-elevation-storage' in our revised manuscript.

2. R: Line 291: The mean sea level the authors are referring to is a common datum or different for different regions.

   A: Here, the mean sea level is a common datum (i.e. WGS 1984).

3. R: Figure 4: What is the maximum storage capacity of Sirikit? In my knowledge it is blow 10000 km^3. Then, why the y-axis of Fig. 4(e) shows a maximum above 15000 km^3? The figure caption says the relationship is based on their maximum storage capacity, which is not true for at least Sirikit reservoir.

   A: The curve shown in Figure 4 also includes values beyond the maximum reservoir level (outside the reservoir). These higher values were deliberately retained to assess the uncertainty in storage estimates in case values exceeding the maximum storage capacity were encountered. You are correct that the full reservoir capacity of Sirikit is approximately 10 km^3 (or 10,000 MCM), which is achieved at an elevation of ~174 m above MSL. We will update the figure caption accordingly.

4. R: Figure 5: Why the storage pattern is different for Longjing, Son La, and Nuozhadu reservoirs? They show storage fluctuations before commissioning unlike the Xe Kaman 1 reservoir.

   A: Yes, you are correct. The storage fluctuations in the Longjing, Son La, and Nuozhadu reservoirs before their commissioning are due to the limitations of the outlier removal algorithm, which should ideally align with the pattern observed in Xe Kaman 1. However, since our primary focus is on the storage time series after the reservoirs' commissioning, our results remain reliable for drawing meaningful insights into reservoir storage dynamics and their impacts. Moreover, we provide level-1 and 2 data to allow users to clean the raw time series with the tools that seem more useful to their applications.

5. R: Figure 5: Is this a daily timeseries of sub-monthly? Which product has been used (level-1 or level-2) to plot this figure. Please mention it in the figure caption.

   A: We have used level-1 (orange and green circle) and level-2 (black-line) to plot Figure 5. We will update the Figure 5 caption accordingly in the revised manuscript.

6. R: Figure 5 caption: What is scene-based reservoir storage (Line 322)? "dynamic components" should be rewritten as "dynamic component" (Line 321).

   A: 'Scene-based reservoir storage' is the storage corresponding to the area estimated using individual scenes from satellite imagery. Since we have shown three (Level-0, Level-1, and Level-2) types of storage time series, we have used 'dynamic components' instead of 'dynamic component'.

7. R: Line 304: How to say that the Longjing (2010) reservoir was filled in roughly one year from Fig. 5(a)? What is the full capacity of that reservoir?

   A: The storage volume of the Longjing Reservoir (commissioned in 2010) was close to zero before 2010. After about a year, the reservoir began operating normally, with storage fluctuations ranging between ~0.45 km³ and ~1.25 km³, as shown in the storage time series (Figure 5a). Note that it has not dropped below 0.4 km³ since then. This indicates that the Longjing Reservoir was filled in approximately one year. The full capacity of the reservoir is 1.22 km³.

8. R: Line 327: Why to use level-1 data when the authors have a level-2 data? The authors use level-1 data in the subsequent sections as well. Any reason for this?

   A: Level-2 data are interpolated to a daily timescale, which introduces additional uncertainty when used for analyses where average storage statistics are sufficient. Therefore, we used Level-1 data instead of Level-2 data for reservoir analysis in the subsequent sections.

9. R: Line 345: It is not the volume reduction in Chao Phraya basin. Instead, the storage has substantially reduced due to persisting drought conditions and both Bhumibol and Sirikit reservoirs showed a continuous decline in storage.

   A: Thank you for your point. We will rewrite this accordingly. Specifically, we will include the following lines in the revised manuscript.

   "*In fact, the storage volume in Chao Phraya has been found to be substantially reduced by ~15% in the post-2010 due to persisting drought conditions during which both Bhumibol and Sirikit reservoirs showed a continuous decline in storage (Fig. S2b, Fig 5e)*".

10. R: Please clearly mention the temporal scale of all timeseries figures (Fig. 5, Fig. 6, Fig. 7, and Fig. 8).

A: We will clearly mention the temporal scale in the Figure captions of Fig. 5, Fig. 6, Fig. 7, and Fig. 8.

11. R: Line 371: We need to refer Fig. S2, not Fig. S3.

    A: Thank you for pointing this out. We will replace Fig. S2 with Fig. S3 at Line 371.

12. R: Table S1: How the authors can explain the underperformance of Bhumibol, Rajaprbha, and Bang Lang reservoirs? Also, what is the allowable level of nRMSE to be good?

    A: There are several possible explanations for the underperformance observed in the Bhumibol, Rajaprabha, and Bang Lang reservoirs. One key factor is the quality of the NDWI images used to estimate the water surface area of the reservoirs. Additionally, the presence of shallow or muddy water during dry years or when reservoir levels are low can affect the accurate classification of water pixels, leading to unrealistic estimates of reservoir storage. While this is somewhat subjective, we can consider an nRMSE of less than 20% to be acceptable, which really depends on the specific application at hand.

13. R: Line 412: Why has the reference period set between 2017 and 2023? Drought is a slow process sometimes persists for decades. Hence, the reference period has to be changed.

    A: From Figure 6h, we can see that only after the 2017 the total storage reached a somewhat 'stable' value—a fact simply explained by the slowdown in the construction of new reservoirs. For this reason, we selected the period 2017–2023 to analyze the impact of drought on reservoir storage under a 'stable reservoir system' across the basins.

14. R: Line 418 – 419: storage conditions were worsened in 2020 because of the combined effect of reduced precipitation and reduced storage levels in the dams in 2019.

    A: Thanks. We will rewrite the sentence accordingly in the revised manuscript.

Conclusions

1. R: Line 438: Aggregated storage capacity of nearly 175 km^3 was observed by the year 2023. Please mention it.

   A: We will modify the sentence accordingly in the revised manuscript.

2. R: Careful about the usage of abbreviations such as SRTM, DEM, GDAT, SAR, NASA SWOT, etc.

   A: Thank you. We will mention the full form before using any abbreviations such as SRTM, DEM, GDAT, SAR, NASA SWOT, etc.

3. R: Line 459: ", flood control" should be rewritten as "and flood control".

   A: We will rewrite the sentence accordingly in the revised manuscript.

References:

1. Hao, Z., Chen, F., Jia, X., Cai, X., Yang, C., Du, Y., and Ling, F.: GRDL: A New Global Reservoir Area-Storage-Depth Data Set Derived Through Deep Learning-Based Bathymetry Reconstruction, Water Resources Research, 60, e2023WR035781, https://doi.org/10.1029/2023WR035781, 2024.
2. Hou, J., Van Dijk, A. I. J. M., Renzullo, L. J., and Larraondo, P. R.: GloLakes: water storage dynamics for 27000 lakes globally from 1984 to present derived from satellite altimetry and optical imaging, Earth System Science Data, 16, 201–218, https://doi.org/10.5194/essd-16-201-2024, 2024.
3. Ang, W. J., Park, E., Pokhrel, Y., Tran, D. D., and Loc, H. H.: Dams in the Mekong: a comprehensive database, spatiotemporal distribution, and hydropower potentials, Earth System Science Data, 16, 1209–1228, https://doi.org/10.5194/essd-16-1209-2024, 2024.
4. Galelli, S., Dang, T. D., Ng, J. Y., Chowdhury, A. F. M. K., and Arias, M. E.: Opportunities to curb hydrological alterations via dam re-operation in the Mekong, Nat Sustain, 5, 1058–1069, https://doi.org/10.1038/s41893-022-00971-z, 2022.
5. Asghar, S., Gilanie, G., Saddique, M., Ullah, H., Mohamed, H. G., Abbasi, I. A., & Abbas, M. (2023). Water Classification Using Convolutional Neural Network. IEEE Access.
6. Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., Magome, J., Nilsson, C., Robertson, J. C., Rödel, R., Sindorf, N., and Wisser, D.: High-resolution mapping of the world’s reservoirs and dams for sustainable river-flow management, Frontiers in Ecology and the Environment, 9, 494–502, https://doi.org/10.1890/100125, 2011.
7. Khazaei, B., Read, L. K., Casali, M., Sampson, K. M., and Yates, D. N.: GLOBathy, the global lakes bathymetry dataset, Sci Data, 9, 36, https://doi.org/10.1038/s41597-022-01132-9, 2022.
8. Schaperow, J. R., Li, D., Margulis, S. A., & Lettenmaier, D. P. (2019). A curve-fitting method for estimating bathymetry from water surface height and width. Water Resources Research, 55(5), 4288-4303.
9. Liu, K., Song, C., Wang, J., Ke, L., Zhu, Y., Zhu, J., ... & Luo, Z. (2020). Remote sensing-based modeling of the bathymetry and water storage for channel-type reservoirs worldwide. Water Resources Research, 56(11), e2020WR027147.
10. Li, Y., Zhao, G., Allen, G. H., and Gao, H.: Diminishing storage returns of reservoir construction, Nat Commun, 14, 3203, https://doi.org/10.1038/s41467-023-38843-5, 2023.
11. Vu, D. T., Dang, T. D., Galelli, S., and Hossain, F.: Satellite observations reveal 13 years of reservoir filling strategies, operating rules, and hydrological alterations in the Upper Mekong River basin, Hydrology and Earth System Sciences, 26, 2345–2364, https://doi.org/10.5194/hess-26-2345-2022, 2022.