Response to the referee comments (RCs)

Anonymous Referee #3

R: This manuscript employs Landsat and Sentinel-2 imagery to calculate the Normalized Difference Water Index (NDWI) and estimate changes in reservoir water surface area across Mainland Southeast Asia. The authors then use hypsometric curves to estimate absolute water storage dynamics for these reservoirs. These storage estimates are validated against several in-situ datasets. Using this dataset, the authors demonstrate the impact of the recent 2019-2020 drought on reservoir storage in the region. Overall, the manuscript is well-written. However, I think the use of NDWI to map surface water and the application of established hypsometric curves for storage estimation do not contribute any significant methodological innovation. I also believe it is unacceptable for a manuscript to lack a Discussion section. I also have a few major concerns outlined below:,

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

We agree that mapping water surface area using NDWI is not novel, but we would also like to note that, in this study, we have not simply used NDWI images (which might be affected by cloud coverage). Rather, we corrected the NDWI images using a novel approach to address the cloud-affected areas and thus get the complete boundary of water reservoirs (please refer to Section 3). Moreover, we have integrated the enhanced NDWI images with a novel bathymetry dataset (Hao et al., 2024) to infer the reservoir's absolute storage. Finally, we would like to stress that Earth System Science Data focuses "on original research data (sets), furthering the reuse of high-quality data of benefit to Earth system sciences", rather than novel methodologies. When introducing our contribution, we thus focussed more on gaps pertaining to the existing datasets (instead of the methodologies with which they were designed); hence the reduced emphasis on our methodological approach.

As for the Discussion, please note that our discussion is embedded in Section 5 ("Conclusions"), since the journal does not provide specific guidelines on where the discussion should be placed. In the revised manuscript, we will separate 'Discussion' and 'Conclusions' and further expand the former by highlighting and discussing the comparison of our NDWI-based derived maps (i.e., maximum water extent map and frequency map – key dataset for estimating storage) with the Global Surface Water Dataset (GSWD) (Pekel et al., 2016).

R: Further validation of the surface water estimates, and hypsometric methods is necessary. I recommend comparing your surface water estimates with the Global Surface Water Dataset (GSWD) and/or other published reservoir datasets. Since you calculated water frequency rasters and maximum water extent, these can also be compared against GSWD water occurrence data to strengthen your results. Additionally, many studies have focused on developing hypsometric curves for reservoirs; it is essential to clarify why your approach is advantageous compared to others. Given the significant uncertainties in using DEMs to derive hypsometric curves, I suggest addressing these limitations in your study.

A: As suggested, we compared our water surface estimates against the ones provided by the GSWD. In particular, we began by comparing the estimates for the Sirikit and Shringarind reservoirs, which are part of the direct validation exercise (Figure 7 in the main manuscript). As shown below, the results of this comparison show a good agreement between our maps and the GSWD ones. For the revised manuscript, we will extend the comparison to all reservoirs within our database. The figures reported below (plus the additional ones we will generate) will be added to the supplementary material.

Please also note that our comparison is carried out in terms of maximum reservoir extent since one-third of the GSWD dataset is affected by gaps and is also available at a monthly frequency (Hao et al., 2024). This contrasts against our dataset, which has a sub-monthly resolution.

Finally, we will also clearly discuss the limitations of developing the hypsometric curves for reservoirs using DEMs. Specifically, we will add the following text to the revised manuscript.

"Developing hypsometric curves using DEM data is constrained by the acquisition date of the DEM, with the earliest widely available dataset being the SRTM DEM (30 m) (year 2000). Consequently, for approximately 30% of reservoirs (constructed before 2000), we utilized the recently released Global Reservoir Area-Storage-Depth Database (GRDL; Hao et al., 2024), which employs deep learning-based bathymetry reconstruction. This database provides reliable bathymetry information for 7,250 GRanD reservoirs (Lehner et al., 2011) worldwide, offering an alternative to traditional methods based on simplified geometric assumptions (Hou et al., 2024; Khazaei et al., 2022; Yigzaw et al., 2018).

While GRDL demonstrates superior performance compared to earlier hypsometric curve methods, its accuracy depends heavily on the size and quality of the training dataset, introducing potential uncertainties in storage estimation. Furthermore, the reproducibility of GRDL's deep learning-based results remains a challenge, limiting opportunities for further refinement and development. In contrast, geometric assumption-based methods, though less precise, offer greater flexibility and transparency for modification and advancement."

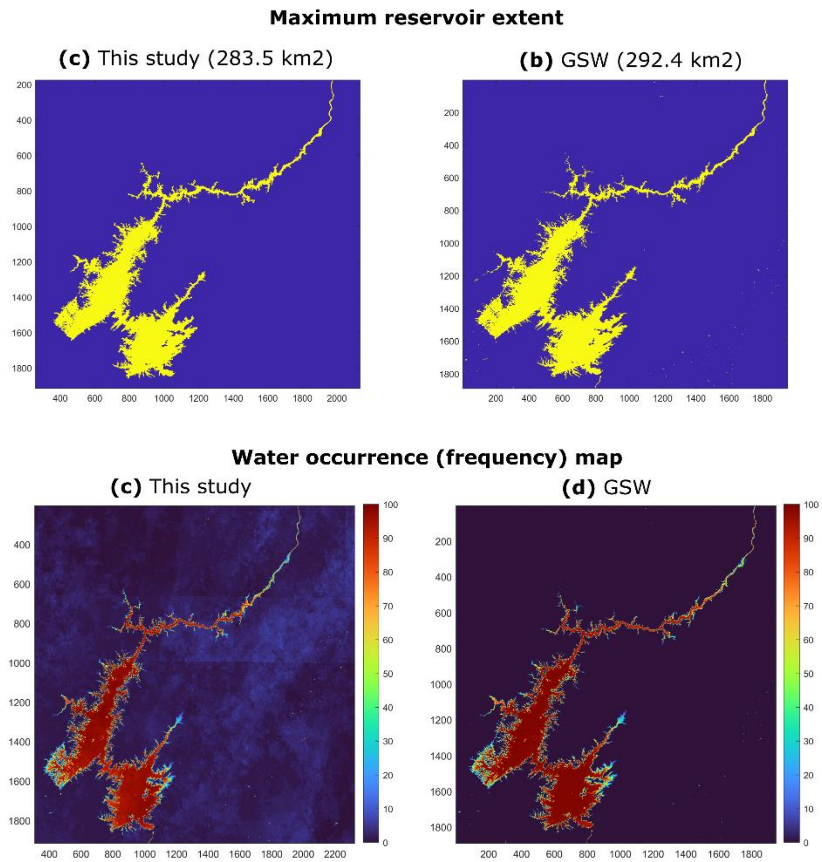

Figure S4. *Comparison of maximum water extent and frequency maps with Global Surface Water Dataset (GSWD) for Sirikit reservoir.*

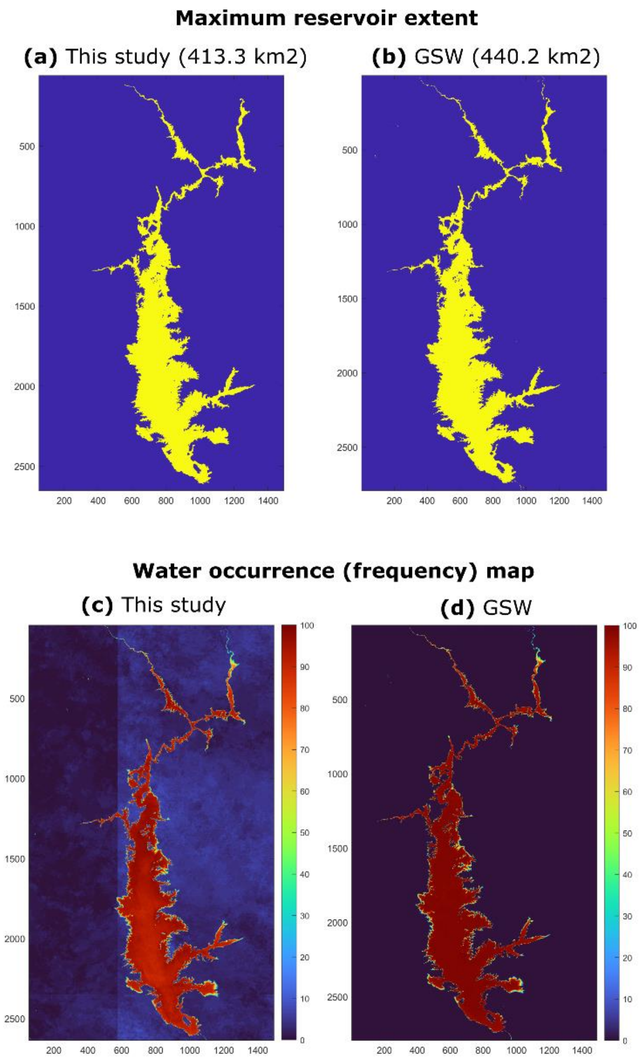

Figure S5. *Comparison of maximum water extent and frequency maps with Global Surface Water Dataset (GSWD) for Srinagarind reservoir.*

R: The use of NDWI alone may be too simplistic and may lack the accuracy needed to effectively map surface water dynamics. Without additional processing, NDWI can be prone to misclassification, especially in areas with mixed land-water pixels or seasonal vegetation cover. Additionally, factors such as high turbidity, shadows, or the presence of aquatic vegetation can further impact the accuracy of surface water mapping. Addressing these limitations is essential, and the authors might consider discussing alternative or supplementary approaches to enhance the reliability of water detection across diverse environmental conditions.

A: NDWI can be prone to misclassification, especially in areas with mixed land-water pixels or seasonal vegetation cover. This is why we have applied locally-adjusted Contrast Limited Adaptive Histogram Equalization (CLAHE, Reza, 2004) to enhance the NDWI images before classification (Section 3.3), which accounts for reducing the mixed pixel effect from the original NDWI images. Please refer to Section 3.3.

In the revised version of the manuscript, we will further discuss about techniques and opportunities to improve the reliability of the water surface estimates.

R: Combining Landsat and Sentinel-2 data should enhance the observation frequency for monitoring surface water dynamics. However, the manuscript does not highlight this potential benefit. You mention that sub-monthly surface water observations are achievable with these datasets, but it seems likely that an even higher frequency could be attained by fully leveraging both satellite sources. I recommend clarifying the observation frequency achieved in this study and discussing how the combined use of Landsat and Sentinel-2 could improve temporal resolution, potentially down to a weekly or even more frequent basis, which would provide greater detail on surface water changes.

A: Thank you for your insightful thought. We have also thought of combining a series of individual Landsat and Sentinel-2 images to generate surface water maps potentially up to a bi-weekly time period. However, many reservoirs do not fit in a single tile of Landsat and/or Sentinel-2, because of the shape and location of the reservoir. This leads to many no-data (missing) pixels over the reservoir, making image enhancement and gap-filling more challenging, and sometimes unrealistic. Therefore, we decided not to go for individual tiles of Landsat and Sentinel-2. Rather, we compromised slightly with the temporal frequency of the images and opted for using the image composites at 10-day intervals.

We will also discuss on the selection of image composite instead of individual Landsat and Sentinel-2 images in our revised manuscript.

Specific Comments:

R: Abstract: The abstract needs to be revised. It does not clearly convey that this study utilizes remote sensing data to estimate reservoir water area dynamics. Instead, it reads more like a compilation of reservoir data in Mainland Southeast Asia.

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

R: L87: The GloLakes database provides absolute water storage data from 1984 to the present, rather than just up to 2020.

A: Thanks for spotting this inconsistency, which we will correct in the revised manuscript.

R: L101: By combining Landsat and Sentinel-2 data, it is possible to derive sub-weekly reservoir dynamics time series, offering higher temporal resolution than the sub-monthly intervals mentioned in your study.

A: We have clarified above the reason for combining Landsat and Sentinel-2 data to get sub-monthly data instead of processing individual tiles to achieve a sub-weekly time-series data.

R: L102: Why did you choose the hypsometric curves developed by Hao et al. (2024)? What advantages does this database offer over those from other studies?

A: As explained in the Introduction, the dataset developed by Hao et al. (2024) is, at this stage, the only dataset providing hypsometric curves for all 7,250 reservoirs within the GRanD. Naturally, this dataset has its own limitations, but adopting these curves is certainly a more robust approach than adopting other techniques, such as inferring the curves from a DEM and then extrapolating the curves below the water surface (Schaperow et al., 2019; Liu et al., 2020). Please also note that we did not use this dataset for all reservoirs, but only for the ones constructed before the SRTM DEM V3 was made available. That corresponds to ~30% of the reservoirs. We will strengthen this explanation in the revised manuscript.

R: Table 1: The GDAR link is not working; please check it. Additionally, the link for "Dams in the Mekong" appears to point to the GRanD database instead.

A: Thank you for spotting these errors, which we will fix.

R: Figure 1: Please ensure the volume units are consistent throughout the manuscript. "Km^3" was used previously, whereas "million m^3" is used here. Consider standardizing to one unit for clarity.

A: We will adopt million m^3 (MCM) throughout the revised manuscript.

R: L169-170: Instead of saying you "acquire" water index, water frequency, or maximum water extent, it's more accurate to state that you "derive" or "calculate" these data.

A: We will use the suggested expressions in the revised manuscript.

R: L182: The term "optical images" should not refer exclusively to "Green (G) and Near-Infrared (NIR)" bands.

A: We will rephrase the sentence in the revised manuscript as follows:

"Shorter wavelength bands, Green (G) and Near-Infrared (NIR) can be affected by the presence of clouds – especially on rainy days – and so, NDWI."

R: L185: Since Sentinel-2 provides a cloud cover product, have you considered using it to filter out cloudy images?

A: Yes, we have applied the cloud information to mask the raw NDWI images before making a composite in both Landsat (cloud information taken from its metadata file) and Sentinel-2.

R: L185: Images with even 5-20% cloud coverage can still significantly impact the accuracy of surface water extent measurements. This level of cloudiness may obscure key areas or introduce errors, making it essential to account for even minimal cloud presence in your analysis.

A: A 5-20% cloud coverage can still have a significant impact on the accuracy of surface water extent measurements. Therefore, to enhance the pixel accuracy, we used NDWI image composites to increase the likelihood of detecting water pixels.

R: L186: Given that Landsat has a 16-day revisit time, how are you compositing 16-day images into a 10-day interval?

A: Landsat satellites have a 16-day revisit time; however, multiple Landsat missions have often operated simultaneously (except before 1999). For example, in 2013, sensors from the Landsat-7 ETM+ and Landsat-8 series were active, enabling the creation of image composites at 10-day intervals.

R: L190-195: You need to classify these NDWI pixels before calculating water frequency and maximum water extent.

A: To clarify, we use a threshold slightly above zero (e.g., 0.1) to classify water and non-water pixels in the NDWI image. In general, a positive value (>0) indicates a water pixel, and using a higher threshold (e.g., 0.1) increases the likelihood of identifying water pixels accurately. While some water pixels with NDWI values between 0 and 0.1 might be misclassified as non-water, this effect is negligible when creating composites. By averaging more than 200 images from the Landsat and Sentinel collections (2013–2023), we estimate water frequency and maximum water extent maps with high reliability.

R: L203: "for": after?

A: Thanks for pointing this out. We will correct it.

R: L240-244: I do not understand why you need to generate level-0 data.

A: Level-0 data are the initial set of results, which are expected to have some outliers because of the uncertainty in the input NDWI images. Yes, we can only supply Level-1 and Level-2 data and not Level-0; however, the idea behind providing Level-0 data as well is to allow the users to make their own Level-1 data using other outlier removal algorithms (if required at all). We will clarify this point in the revised manuscript.

R: L246: Please specific "trend-preserving interpolation technique".

A: We intended to convey that the reservoir storage time series exhibits an increasing trend, particularly during the filling period. In such cases, linear interpolation may not perform well, especially when there are steep slopes, curvatures, or seasonal variations between data points. To address this, we employ a more robust interpolation technique, such as spline or LOESS interpolation, which fits a polynomial to the data and preserves the underlying trend.

R: Table 3: remove low dash after the words (e.g., "Level_m_")

A: We will remove them in the revised manuscript.

R: Figure 5: Do the solid lines represent Level-2 data, or are they a combination of Sentinel and Landsat data?

A: The black solid line represents the Level-2 data. We will correct the legend in our 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. Reza, A. M.: Realization of the Contrast Limited Adaptive Histogram Equalization (CLAHE) for Real-Time Image Enhancement, The Journal of VLSI Signal Processing-Systems for Signal, Image, and Video Technology, 38, 35–44, https://doi.org/10.1023/B:VLSI.0000028532.53893.82, 2004.
4. Pekel, J. F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. Nature, 540(7633), 418-422.
5. 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.
6. 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.
7. Yigzaw, W., Li, H.-Y., Demissie, Y., Hejazi, M. I., Leung, L. R., Voisin, N., and Payn, R.: A New Global Storage-Area-Depth Data Set for Modeling Reservoirs in Land Surface and Earth System Models, Water Resources Research, 54, 10,372-10,386, https://doi.org/10.1029/2017WR022040, 2018.
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.