We thank the reviewer for spending time and effort reviewing our work. Below the reviewer's comments are in **black** and the replies in **blue.**

This manuscript evaluates the use of SWOT L2 HR PIXC data to derive relative water surface elevation (WSE) time series for small Danish lakes and compares the results with the LakeSP product. The topic is timely and the validation dataset is potentially valuable. However, in its current form, the paper reads more like a processing/engineering report than a scientific study, and the broader significance of the proposed workflow is not yet convincingly demonstrated. Several methodological choices also require clearer definition and stronger justification. Below I outline major issues that should be addressed before the manuscript can be considered for publication.

Reply: The authors would like to thank the anonymous reviewer for their time and effort in providing their expert views and comments on the manuscript during the review process. We believe addressing them will help us strengthen and clarify the manuscript. We respond to each of the points raised below.

Major comments

1) Novelty and scientific contribution relative to LakeSP are not sufficiently established.

At present, the manuscript mainly documents a processing workflow and reports summary metrics for Danish lakes. The improvement over the filtered LakeSP product appears modest (median RMSE 5.34 cm vs. 6.05 cm; median PCC 0.91 vs. 0.89), which raises the question of whether the contribution is a conceptual advance or primarily an alternative filtering/aggregation strategy. The authors should explicitly clarify:

Reply: Thank you for pointing out this unclarity in our manuscript. We think our approach offers distinct advantages over LakeSP and will answer your points below.

What is fundamentally new in the proposed approach (beyond parameter choices and filtering rules)?

Reply: Fundamentally, we use the same data source as LakeSP – namely *SWOT PIXC* – to aggregate a single WSE value per lake and timestamp. Two distinct differences are that (1) LakeSP uses observations labelled as 'dark water' and 'water near land' besides 'open water' whereas we only use the 'open water' class. (2) Lake SP doesn't use a static lake mask for observation extraction – the initial lake mask from the PLD is dynamically adapted based on SWOT observation classification.

In ideal circumstances, this will yield the same results. However, 'dark water' observations tend to have significantly higher noise associated with them and can dominate the 'good' signal. This can cause a singular timestamp of LakeSP to be overly 'noisy'. Regarding the dynamic extraction mask: SWOT is sensitive to wet soil etc. which can cause misclassifications of land into the 'water' classes. If this happens close to a lake, the lake mask expands causing the faulty observations to be included in the LakeSP. There could be other causes of wrongly expanding lake masks. In Figure A1F we present and example.

With our approach, we circumvent these scenarios. Hence, we retain more valid (*good*) timestamps in the lake timeseries.

Why this constitutes a meaningful advancement over LakeSP in terms of methodological design, robustness, or interpretability?

Reply: As mentioned above, the main advantage of our approach is that we retain roughly double the amount of valid (*good*) timeseries points for the lakes we investigated for the same observation period (fx. Line 11). We agree that the performance in terms of RMSE and PCC are not large, which only highlights the groundbreaking capabilities of SWOT. However, we firmly believe that in an observational frame ca 2 years (July 23 to Mai 25) a difference of 35 retained observations for LakeSP vs 65 observations for our approach constitutes a meaningful advantage (line 195-198).

Under what conditions the proposed method is expected to outperform LakeSP (e.g., lakes with small seasonal amplitudes? high-latitude multi-pass coverage?), and why.

Reply: we expect our approach to outperform LakeSP for lakes with highly variable WSE given that we retain double the amount of timeseries points compared to filtered LakeSP. Generally, in all contexts that necessitates as many 'good' observations as possible, fx.:

- Floodings
- Dam operations

Without a clearer positioning, the contribution currently appears incremental.

Reply: We agree that we should strengthen our argument in the manuscript, why our approach constitutes a significant advantage for some lakes. In the revised manuscript, we plan to strengthen out argument with the listed points above. Thank you for pointing this out.

2) Limited geographic scope and weak discussion of transferability/scalability.

The study focuses exclusively on Danish lakes. While this is a reasonable testbed, the manuscript does not clearly explain how the findings generalize to other hydroclimatic settings (e.g., arid regions, vegetated shorelines, tropical systems)

Reply: Agreed, we will expand the manuscript to include a broader discussion.

- the **geomorphology** can impact our approach twofold. (i) tall flanking topography to the sides of the lake can cause layover → this should be masked by the SWOT intrinsic DEM. It will affect the number of observations per lake and time stamp which inturn may cause our filter to reject more observations as the threshold might not be reached. (ii) significant elevation differences along the lake extent will cause problems for using a singular WSE aggregate for lake representation. Depending on where retained SWOT observations are sampled on the lake, a bias will be introduced. The impact on small lakes should be small however, but extremely elongated lakes in mountainous terrain should likely be excluded from out approach.
- For **ice cover** we expect a degradation of coherence leading to fewer reliable observation points. The general point observation quality should also degrade compared to open water. We don't know the scattering mechanisms for Ka band SAR over ice and snow. More research is required. (https://doi.org/10.21203/rs.3.rs-8454201/v1 - accepted and soon to be published in final version)
- For **submerged vegetation** and **turbid** water, we expect no impact on the WSE measurements. Ka band SAR is expected to not penetrate the water surface.
- **Surrounding vegetation** will hinder Ka band penetration to the water surface, but its impact should remain small as it is limited to the immediate shoreline. We would expect its impact to be analogous to a slightly too small lake polygon.

- **Wind** conditions will impact backscatter intensity. Initially we expect an increase in backscatter intensity for winds up to ca. 3 m/s which will fall again for wind speeds exceeding this threshold (doi: 10.1109/TGRS.2013.2258402) (interplay of surface roughness, Ka band wavelength and satellite look angle). Hence in high wind regime environments we expect worse data quality.
- Changing **hydrological regimes** may impact SWOT measurement. The static lake mask could cause error by including land. However, we expect the impact to be small as we (i) use SWOT's intrinsic water classification by using 'open water' only and (ii) using the median to aggregate a unified WSE measure per lake and time stamp – dampening the impact of outliers.
  At the same time, if seasonal WSE signal is highly variable, higher temporal resolution increases in importance, which is where our approach retains more time stamps than filtered Lake SP.

or to regions with different viewing geometries and sampling densities.

Reply: we partially agree. The lakes in Denmark are well spaced out and are observed by the ~10 SWOT passes covering Denmark in all positions of the swath. Therefore, different viewing geometries and sampling densities are covered. We can expand on that in the data section.

In particular, it is unclear whether key selection criteria (e.g., ≥200 points per acquisition, strict geo-flag handling) are transferable to areas with fewer valid observations per pass, smaller lakes, or more heterogeneous shorelines.

Reply: we partially agree. The purpose of our study is to investigate how SWOT works on small tempered lakes. While we agree that trying our approach globally would be an interesting study, this would be out of scope for this manuscript. We will expand on the above listed discussion:

- **Fewer observations per pass** will degrade our approach as we set a strict limit of 200 observations. The Danish lakes sample the whole swath; thus, we don't see this to impact our approach negatively when transferring. The main hindrance we see are lakes in steep terrain (see geomorphology)
- **Heterogeneous shorelines** can affect our approach just as LakeSP. With an accurate lake mask, we do not expect significant degradation.
- **Smaller lakes** will intrinsically have fewer observations. If one wants to look at lakes smaller than the scope of this study, the parameters could be tuned to either relax the 200 observations threshold or to include 'water near land'.

3) Methodological description needs clearer formalization and stronger justification.

Despite the technical focus, several core steps are not described with sufficient precision to ensure reproducibility or to interpret why performance improves:

Reply: thank you for your comment. We will aim to improve the clarity of our statements in the revised manuscript.

The WSE aggregation procedure should be formally defined (e.g., summary statistic used, whether any weighting is applied, outlier handling rules, spatial screening, and how multiple water bodies within a mask are treated).

Reply: We will aim to improve our description of our workflow. Our workflow constitutes of the following steps:

a) For each SWOT observation
    (i) Extract SWOT observations by lake mask (line 107)

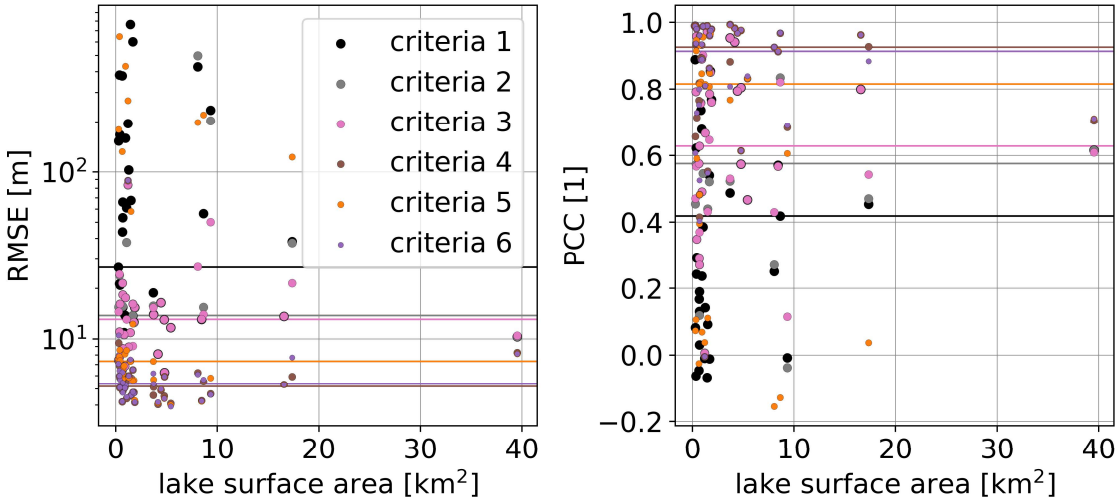
b) For each lake extracted from the SWOT observation
   (i) Reject observations outside 3 sigma confidence interval (line 107-110)
   (ii) Take the median of points – check Table 1 to see on which (line 112-116)
   (iii) Discard timestamp if criteria are fulfilled – check Table 1(line 117-119)

We only partially agree that our approach is not formally defined. Everything we do is described in Section 3.3. We do not perform weighting, spatial screening besides the lake mask, and our governmental authorative lake masks include by definition only one water body per polygon.

However, we hope, that a restructuring will improve the clarity of our approach.

The sensitivity of results to the ≥200-point threshold should be assessed more explicitly (e.g., dependence on lake area, swath edge vs. near-nadir sampling, and seasonal/ice conditions).

Reply: We agree, this could be expanded over what is presented in Table 2. We plan to add a figure like this to more clearly show the impact of our selection criteria on lakes with different surface areas.



to the revised manuscript. We hope this addresses your point adequately. Seasonal variations of lakes in Denmark are small, and lakes are rarely ice covered. For near nadir the SWOT pixels have a larger footprint, hence less observations would be present for a lake in that swath location compared to if it was in the far range. This is a clear limitation to our approach which we will clearly state in the revised version of our manuscript.

The rule to exclude all geo-flagged observations (unless the point count drops below 200) needs a clearer statistical rationale. Why is this threshold optimal or conservative, and how robust is it?

Reply: We do not exclude geo-flagged observations (unless the point count drops below 200). We exclude always exclude geo-flagged observations (criteria 5 & 6) retain them (criteria 1,2,3), retain them but discard the observation if the geo-flagged points are 20% or more of all points within the lake (criterion 4), Table 1. The threshold N=200 is set arbitrarily but was chosen to minimize outliers in the final lake timeseries. It is conservative in the sense that we chose to set the threshold knowingly high, to reduce the risk of remaining outliers. We deliberately accept less retained timestamps to trade off for '*no*' outliers in our timeseries. We chose to evaluate the robustness based on the absence of outliers in the timeseries. Our results show that we successfully supress outlier occurrences in the timeseries.

In addition, the roll-error discussion remains largely qualitative. If residual roll error is emphasized as affecting ~18% of acquisitions, the manuscript should provide a more systematic quantification approach (e.g., uncertainty envelopes, swath-position dependence, cycle-to-cycle variability), and discuss the implications for inland-water monitoring beyond Denmark.

Reply: We partially agree with the point. We agree that a systematic quantification and analysis of the residual roll error would be relevant and certainly worth a study. However, we don't think that this is within scope of our study.

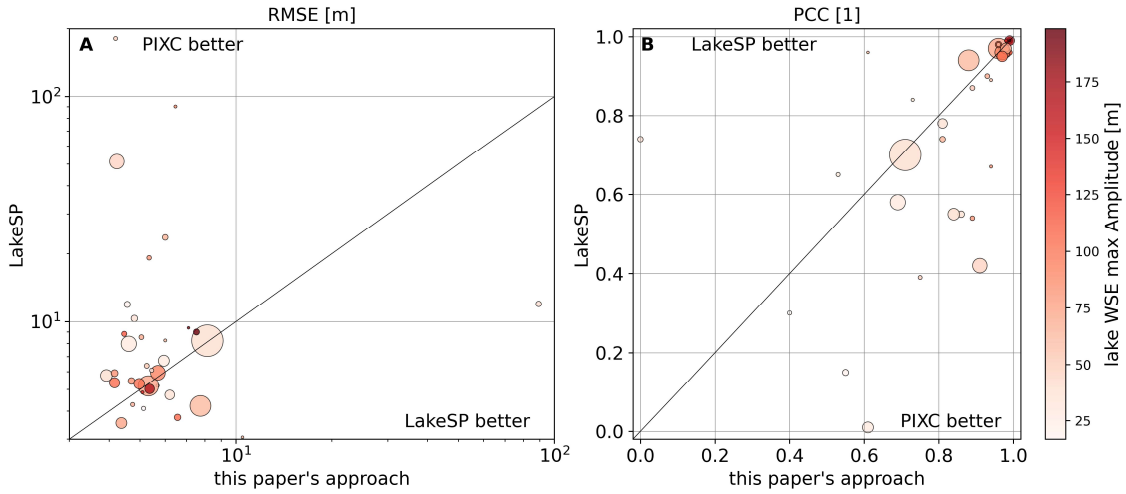
We are not trying to correct or quantify the error but only want to highlight its existence and explain the errors that we are seeing. We will revise our wording on that matter and adapt to better represent out purpose for the revised manuscript.

4) LakeSP comparison is not sufficiently diagnostic to support strong conclusions.

The manuscript argues that the proposed approach performs slightly better while retaining more timestamps, but the comparison does not yet isolate *why* differences occur. In particular:

The evaluation should include a more systematic statistical comparison (e.g., lake-by-lake paired differences in RMSE/PCC, confidence intervals/bootstrapping, performance stratified by lake size, WSE amplitude, swath position, and presence/absence of roll error).

Reply: We agree that these comparisons add value to the manuscript. We presented them in Table A1. The data can be visualized to improve understanding, and we plan to add a figure like below to the revised manuscript. It shows RMSE and PCC for our approach vs filtered LakeSP with scatter point size scaled to lake surface area and colour coded by maximum WSE amplitude (date from Table A1).



Each lake is covered by ~3 passes per cycle and will be positioned in different sections of the swath for each of them. As argued in a preceding section, we expect that our approach may unnecessarily reject timestamps of lakes in a near nadir position given it will have less observations. Given our approach retaining more timestamps than filtered LakeSP, however, we do not see that this as a problem. Otherwise, we expect the lake's position in the swath to impact our method and LakeSP equally. Analogously for the residual roll error.

The trade-off between operational simplicity (LakeSP) and configurability (proposed workflow) should be framed explicitly: in what use cases does the proposed method justify additional complexity?

Reply: We agree. Our argument we tried to make in Line 207ff should be revised and sharpened. We aim to improve it in the revision along with a general overhaul of the results and discussion sections.

Generally, we see LakeSP to be the preferred data source for global / large scale applications, for users for whom ease of use is most important, and for applications where temporal resolution is not the highest priority. On the other hand, we see our approach (or similar) should be used for applications where temporal resolution is of the highest priority or for localised studies.

As written, the marginal performance gains make it difficult to justify the proposed workflow as a replacement or substantive supplement to LakeSP.

Reply: We partially agree, the performance metrics of RMSE and PCC alone would not justify our approach as a supplement to LakeSP. Yet 50% higher temporal resolution for the same period, at least for particular use cases, does. We will make sure to highlight this better in the revised manuscript.

5) Structure and framing should be revised toward a hypothesis-driven scientific narrative.

The manuscript would benefit from substantial restructuring to move from a workflow description to a clearer scientific contribution. Specifically:

Reply: we agree.

State explicit research questions/hypotheses early (e.g., “Can PIXC-derived relative WSE time series for small lakes achieve cm-level accuracy, and what selection criteria are required?”).

Reply: we agree. Thank you for pointing this out. We will incorporate this in the revised version. We are sure it will improve the manuscript.

Separate methodology development, validation results, and implications for broader SWOT inland-water applications.

Reply: we agree.

Strengthen the discussion to emphasize what the Danish case study reveals about SWOT capabilities/limitations for small lakes globally.

Reply: We hope that incorporating the comments we made here and in reply to reviewer 1 will address your concern adequately.

Summary recommendation

The dataset and topic are valuable, and applying SWOT to small lakes is an important research direction. However, the current manuscript does not yet demonstrate a sufficiently strong methodological innovation or convincingly argue global relevance. I recommend **Major Revision**, focusing on (i) clarifying novelty relative to LakeSP, (ii) formalizing and justifying methodological choices, and (iii) providing a clearer framework for transferability and scalability beyond Denmark.

Reply: Thank you for your comments. We will address them in the revised manuscript and particularly focus on a rewriting and restructuring of the paper to improve the reading flow and expanding on the discussion on SWOT capabilities, transferability, and design choice.