This paper provides an evaluation of passive microwave (PM) sea ice concentration (SIC) estimates using classified airborne visual imagery from NASA's Operation IceBridge (OIB) and introduces a new ICESat-2-based Linear Ice Fraction (LIF) dataset. The comparison includes classified imagery from four satellite imagery scenes (Sentinel-2 and WorldView). The study is a follow-up to a previous submission, which I also reviewed. In response to that feedback, the authors have now split the original manuscript into two parts to allow for a more focused discussion. This is a good approach, but I am somewhat uncertain about how well it has worked in practice.

We thank the reviewer for their thoughtful and constructive comments. We appreciate the time and effort they took to read and evaluate our manuscript. The suggestions and feedback have helped us clarify and improve the presentation and interpretation of our results. Below, we provide detailed responses to each comment and describe the corresponding changes made to the manuscript.

In this Part 1 study, much of the paper is dedicated to presenting classified airborne imagery from Operation IceBridge, which is used to highlight biases in PM SIC data. However, we already know from past research (e.g., Kern et al.) that PM SIC products contain biases, and the comparisons here mostly serve to confirm those findings. While it is valuable to have more insight into these biases, the OIB imagery is not actually used to assess the new ICESat-2 LIF dataset, even though LIF appears to be the main focus of the study. More than half of the paper is therefore spent reaffirming known PM SIC biases, rather than contributing directly to the LIF validation. My impression from the earlier review was that the authors planned to expand the comparisons with coincident imagery, ideally incorporating OIB data that overlaps with ICESat-2 tracks. There were indeed OIB cal/val flights in 2019 and a summer calibration campaign in 2022 that could have been utilized for this.

Thank you for your comment. Validation of the ICESat-2 surface type classification is not the purpose of this paper. We refer the reviewer to other studies that have conducted such validation and assessment (Kwok et al., 2020, Petty et al., 2021, Tilling et al., 2020). The novelty of this study is understanding PM biases with the OIB imagery and development of the LIF product from ICESat-2.

Kwok, R., Petty, A. A., Bagnardi, M., Kurtz, N. T., Cunningham, G. F., & Ivanoff, A. (2020). Refining the sea surface identification approach for determining freeboards in the ICESat-2 sea ice products. *The Cryosphere Discussions*, *2020*, 1-18.

Petty, A. A., Bagnardi, M., Kurtz, N. T., Tilling, R., Fons, S., Armitage, T., ... & Kwok, R. (2021). Assessment of ICESat-2 sea ice surface classification with Sentinel-2 imagery: Implications for freeboard and new estimates of lead and floe geometry. *Earth and Space Science*, *8*(3), e2020EA001491.

Tilling, R., Kurtz, N. T., Bagnardi, M., Petty, A. A., & Kwok, R. (2020). Detection of melt ponds on Arctic summer sea ice from ICESat-2. *Geophysical Research Letters*, *47*(23), e2020GL090644.

The second half of the paper presents the LIF analysis, where the authors introduce four Sentinel-2 and WorldView scenes to evaluate ICESat-2-derived LIF estimates. However, they only show one example in detail, and then provide a summary table, which makes the evaluation of this new dataset feel quite limited. A major advantage of working with a small number of high-resolution scenes should be the ability to explore different surface types, environmental conditions, and classification performance in depth, but this aspect is underdeveloped. For example, it would be interesting to analyze how different ATL07 classification types influence LIF retrievals, how well the drift correction works, or how dark leads, which have known retrieval issues and are no longer included in the sea surface height retrievals (Kwok et al., 2021), affect the results. Similarly, while drift correction is applied, the fact that only four scenes are used means that manual adjustments could have been done instead. Recent studies such as Koo et al. (2023) and Liu et al. (2025) analyzed 17-18 coincident Sentinel-2 scenes and manually adjusted them, so this approach should be considered. Having at least one summer scene and scenes from other parts of the Arctic and the Southern Ocean would be highly beneficial.

Thank you for the comment. We agree that a more in depth study of how ICESat-2 responds to a variety of sea ice conditions would be beneficial. However, the purpose of this study is to demonstrate the PM biases relative to the optical imagery sea ice concentration and to suggest that we may take advantage of ICESat-2's narrow footprint to better resolve leads and reduce some uncertainty and biases of PM SIC in high ice concentration areas. The three papers mentioned above can provide information on the accuracy of the sea ice surface classification. We have added a sentence in the conclusion:

"We acknowledge that the LIF is a derived product and thus dependent on the accuracy of the ICESat-2 surface type classification."

As a point of clarification on the ATL07 product- dark leads are no longer included in the sea surface height retrieval, but are still included as leads and thus fall in the category of open water in the LIF determination.

Another aspect that could be explored in more detail is the definitional differences between PM SIC, optical imagery SIC, and ICESat-2 LIF. This is only briefly mentioned in L219, where the authors state that "new ice that appears gray in color is considered ice for SIC and LIF calculations." However, this is a significant issue that deserves more discussion. How do different ice types (open water, leads, gray ice) compare across these datasets? Addressing this would improve the interpretation of the results.

We agree that this is an important topic to discuss. We have added a paragraph to discuss this:

"An important consideration when comparing SIC estimates across different sensors is the definition of what constitutes "ice." Thin ice emits microwave radiation at levels intermediate between open water and thick, snow-covered ice, making it difficult to distinguish using standard PM SIC retrieval algorithms (Comiso and Sullivan, 1986). In high-resolution imagery, new ice is often visually distinct: it appears darker than first-year or multi-year ice, significantly brighter than open water, and with a near-infrared reflectance higher than that of melt ponds. These spectral and brightness differences make it relatively straightforward to develop algorithms that distinguish new ice from other surface types. While thin ice is generally classified as ice in ICESat-2 data (Figure 4), the radiometric properties of thin and thick ice remain challenging to distinguish. This study finds that passive microwave products generally overestimate SIC, where the potential underestimation caused by misclassifying thin ice as open water is offset by the overestimation resulting from the inability of coarse-resolution sensors to resolve narrow leads. These opposing biases can obscure the true impact of thin ice on SIC retrievals.”

Specific Comments:

L142: The authors state that they examine images where MPF ≤ 50% in summer to avoid outliers and misclassified images in the unsupervised analysis. However, wouldn’t it be more informative to include scenes with high melt pond fractions? Why are these considered outliers?
This wording was also pointed out by reviewer 1. We have corrected this sentence for clarification.

By “outliers,” we are referring to images where the very high MPF (MPF>50%) may not be representative of actual surface conditions- due to either small image footprints that are not representative of the whole PM footprint area or potential misclassifications in the image processing routine. This is 2% of images. We have clarified this in the text.

We have edited the text in the manuscript:

In the summer, we restrict the analysis to images with MPF<= 50% to reduce the influence of potentially misclassified images that may produce unrepresentatively high melt pond fractions.

L151: The paper states that PM SIC products on average overestimate SIC, but there is significant spread.

We changed this sentence to: “PM products overestimate SIC on average, but there is some spread.”

L160: It is mentioned that PM SIC products have a bias on average, but they are highly variable. However, OSI SAF and ASR do not appear to be biased on average—can this be clarified?

They are biased on average–3.5% and 5.5 % are their average biases. See Table 2. We have added a reference to Table 2 into the text.

L218: The authors mention “other pixels” in the text, but in Figure 4, the term “new ice” is used instead. This should be consistent.

Thank you for pointing this out. In this case, the term *“other pixels”* refers specifically to *“new ice”* that appears gray in optical imagery, as noted in the manuscript. To improve clarity and maintain consistency with Figure 4, we have added a clarification indicating that these "other" pixels correspond to "new ice:"

“Following Buckley et al., (2023), we classify the WorldView and Sentinel-2 image pixels into surface types: open water, ice, and other (new ice).”

L225: How well does the drift correction work? This was not very clear in the methods section. Given that only four scenes are used, why wasn't a manual adjustment tested as an alternative?

In this study, we visually verified the alignment between the corrected ICESat-2 tracks and corresponding lead and floe features in the optical imagery (Figure 4 and 5). In the four cases presented, the drift correction consistently improved alignment and no additional manual adjustment was necessary. We have clarified this in the methods section and now note that visual inspection was used to assess correction quality. We added a clarification sentence:

We visually verified the effectiveness of the drift correction by ensuring that leads identified in the ICESat-2 surface type classifications aligned with corresponding leads in the optical imagery.

L236: The statement that the May 7, 2022 image represents an area of highly fractured sea ice that four PM SIC products classify as completely ice-covered is interesting. It would be useful to include a visual example of this to illustrate the discrepancy.

We agree! We have now added the May 7, 2022 image, showing the highly fractured sea ice along with the ICESat-2 ground track, to illustrate the discrepancy between the observed surface conditions and the PM SIC estimates as Figure 5.

Figure 3: This figure is hard to interpret, and a better way to display this data should be considered.

We have added text to better describe this figure and aid with interpretation:

Figure 3 illustrates the relationship between increasing MPF and PM-SIC bias across products. The delineated interquartile ranges emphasize that not only does the bias increase with MPF, but the variability across scenes also grows, underscoring the challenge of accurately estimating SIC under ponded conditions due to spectral variability in melt pond signatures.

ASI Data in Figure 3: Why is ASI data only shown in the 0-5% SIC interval? Shouldn't all products be included in every bin?

We added this to the Figure caption to clarify:

Note that the average melt pond fraction (MPF) within a PM grid cell is typically greater than 5%; only within the smaller 6.25 km ASI grid cells is the average OIB-derived MPF below 5% for some cells.

All Sentinel-2/WorldView scenes: Since there is space, why not show all four Sentinel-2/WorldView comparisons with ICESat-2 LIF? The Figure 4 scene also looks quite small—is this the full scene, or just a zoomed-in version?

We now include all four Sentinel-2 and WorldView scenes with the ICESat-2 tracks overlaid, as requested as Figure 5. As noted in the manuscript, the WorldView images are smaller in spatial extent than the Sentinel-2 scenes. Specifically, "The ICESat-2 tracks transect the 14 km × 17 km WorldView image for 14.2 km." Figure 4 shows the full extent of the WorldView image, not a zoomed-in subset.

Comparison with MODIS/Landsat PM Evaluations: The authors reference previous Kern et al. studies but do not clearly compare their results. How do the biases in this study compare with those found in MODIS and Landsat SIC evaluations? Are there any new insights gained from the OIB comparisons?

Thank you for this comment. We have added to our discussion of results in relation to MODIS-based comparisons in Section 3.2 where we had noticed similar bias patterns across algorithms (e.g., higher biases for Bootstrap and NT products). We added further discussion on how PM algorithms interpret summer melt surface conditions, incorporating relevant findings from the Kern et al. studies. We further clarify the comparison across studies, we have now added a sentence to the conclusion summarizing how our OIB-based results compare with evaluations:

"These findings are generally consistent with previous PM studies including comparisons with MODIS (Kern et al., 2020), Landsat (Kern et al., 2022), and ship-based observations (Kern et al., 2019). However, the OIB-based comparisons in this study reveal generally smaller absolute biases and provide new insights into how PM SIC may not capture the smallest-scale sea ice features seen in high-resolution imagery."

Methods and Results Organization: The methods and results are somewhat mixed together, which makes it harder to follow. It may be better to fully separate them, even if this shortens the results section.

We present the results and methodology alongside one another because there are two themes of this paper: first we do a comparison of the PM SIC and OIB SIC, and second we demonstrate the ability of LIF. Presenting the methodology and results for each together allows for greater clarity within each section and helps readers follow the logic of the analyses. We have reviewed the text and made small edits to improve clarity and better separate the methodology and results within each section.

References

Kern, S., Lavergne, T., Notz, D., Pedersen, L. T., Tonboe, R. T., Saldo, R., and Soerensen, A. M.: Satellite Passive Microwave Sea-Ice Concentration Data Set Intercomparison: Closed Ice and Ship-Based Observations, The Cryosphere, pp. 1–55, https://doi.org/10.5194/tc- 2019-120, 2019.

Kern, S., Lavergne, T., Notz, D., Pedersen, L. T., and Tonboe, R.: Satellite passive microwave sea-ice concentration data set inter-comparison for Arctic summer conditions, The Cryosphere, 14, 2469–2493, https://doi.org/10.5194/tc-14-2469-2020, 2020.

Koo, Y., Xie, H., Kurtz, N. T., Ackley, S. F., and Wang, W.: Sea ice surface type classification of ICESat-2 ATL07 data by using data-driven machine learning model: Ross Sea, Antarctic as an example, Remote Sensing of Environment, 296, 113726, https://doi.org/10.1016/j.rse.2023.113726, 2023.

Kwok, R., Petty, A. A., Bagnardi, M., Kurtz, N. T., Cunningham, G. F., Ivanoff, A., and Kacimi, S.: Refining the sea surface identification approach for determining freeboards in the ICESat-2 sea ice products, The Cryosphere, 15, 821–833, https://doi.org/10.5194/tc-15-821-2021, 2021.

Liu, W., Tsamados, M., Petty, A., Jin, T., Chen, W., and Stroeve, J.: Enhanced sea ice classification for ICESat-2 using combined unsupervised and supervised machine learning, Remote Sensing of Environment, 318, 114607, https://doi.org/10.1016/j.rse.2025.114607, 2025.