Thank you for your thoughtful and constructive review on our manuscript, *Evaluation of climatic predictors of surface ponding on Antarctic ice shelves*. Your comments are in gray text below, with our responses in black. Line numbers refer to the submitted version of the manuscript.

Summary

Glazer and Tinto evaluate climate predictors for surface ponding on Antarctic ice shelves, testing the theoretical meltwater-over-accumulation (MOA) threshold of 0.7 against a calibrated threshold based on observational surface melt products, air temperature and RACMO across three spatial resolutions (2, 11 and 27 km). They further apply the best empirical MOA value and a grounding line proximity index against the theoretical one to project surface melt ponding until 2100 under three climate scenarios, demonstrating a substantial underestimation of surface ponding when relying solely on the theoretical threshold. Depending on the emission scenario, the theoretical threshold fails to capture a significant amount ponded pixels relative to the empirical threshold. This underscores the necessity of calibrating the MOA threshold to avoid systematic and potentially drastic underestimation of surface ponding on Antarctic ice shelves in future projections. The manuscript is well written and the findings are of considerable scientific significance. The following comments are intended to assist the authors in clarifying certain methodological choices and strengthening the overall presentation.

General Comments

Title: The title is currently quite broad. Consider refining it to explicitly reference MOA, grounding line proximity and air temperature thresholds, so that the scope and contribution of the study are immediately apparent to the reader.

We agree that the current title is broad, but incorporating MOA, GLPI, and temperature into the title would make it quite long. We suggest “Evaluation of meltwater-over-accumulation ratio as a predictor of surface ponding on Antarctic ice shelves”, since MOA is the primary metric being tested. We will change the title in our revised manuscript.

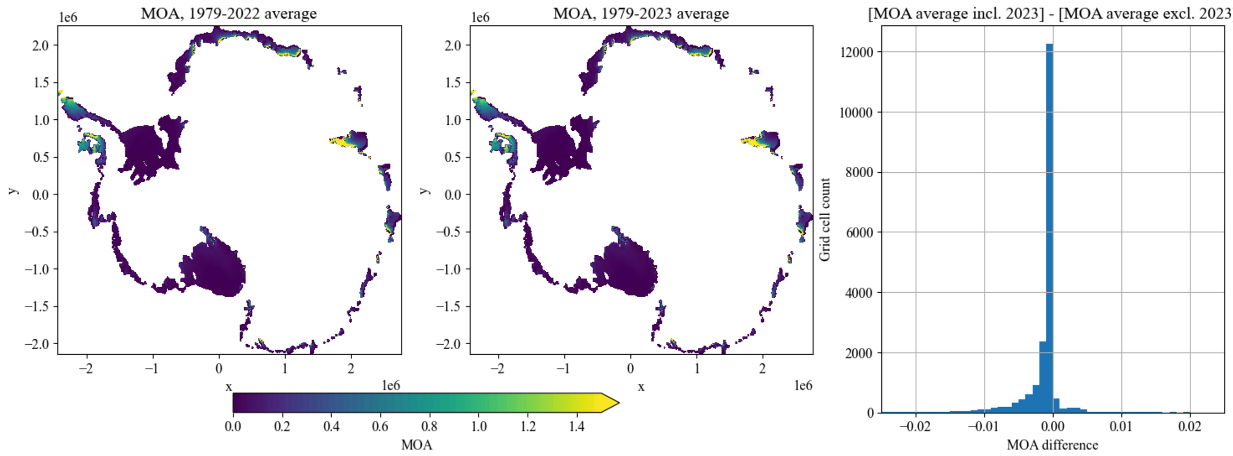
Temporal aggregation of datasets and used variables: It is not entirely clear over which time spans the averages were calculated. For example, Morris & Vaughan report the −9°C isotherm based on the mean annual air temperature for the year 2000. It would be helpful to specify whether air temperature was calculated for each lake dataset over its respective observation period or over the full RACMO time span. The selection of air temperature over skin temperature would also benefit from explicit justification, and it may be worth considering whether a parallel analysis using skin temperature could be conducted to assess sensitivity to this choice.

We will add a statement at line 86 that says “Average air temperature and annual average MOA were calculated over the full RACMO time span, from 1979 to 2022 or 2023, depending on the model iteration (see Table 1).” The column “RACMO temporal coverage” in Table 1 (line 100) contains the date ranges over which both air temperature and MOA averages were calculated.

We tested both air temperature and skin temperature, and found that the results did not significantly differ. We chose to focus on air temperature in the manuscript, for consistency with the methods of both Morris & Vaughn (2003) and Cook & Vaughn (2010). We will add both of these explanations to Appendix B1, by altering lines 515-516 as follows: “In order to determine temperature and calculate MOA, we used outputs of 2m air temperature, meltwater production, total precipitation, snowfall, rain, and sublimation from the hydrostatic regional climate model RACMO at 27 km, 11 km, and 2 km resolutions. We also examined skin temperature in comparison with 2m air temperature, and found the relationship between F1 scores and optimal thresholds to have similar characteristics. We chose to focus on air temperature for this study to maintain consistency with the work of Morris & Vaughan (2003) and Cook & Vaughan (2010).”

Taking into account strong intra-annual variability of surface ponding on Antarctic ice shelves: As understood, lake ponding was aggregated based on the number of observed lakes on a pixel basis for the Dell and Tuckett datasets. However, e.g. RACMO covers the more recent year 2023, during which a pronounced peak in ponded surface meltwater occurred along the Antarctic Peninsula that is not captured by the observational products. Given that these datasets span different time periods, it would be important to clarify how this temporal inconsistency was accounted for, as it may substantially influence the results and how the strong intra- and inter-annual variability of surface ponding was accounted for.

Thank you for raising this point. We investigated the consequences of using the full temporal extent of the RACMO data as opposed to truncating it to the final year of available lake observations. We find that this choice does not substantially influence the results. Please see below plots of the annual average MOA from RACMO2.4p1 (11 km), with and without inclusion of 2023, as well as a histogram of the differences:



Our analysis is based on multidecadal averages of climate variables. As discussed in the manuscript (lines 435–436, 523–524), this reflects the timescale over which firn-air depletion preconditions the ice shelf for surface ponding. This means that edge cases or extreme years, such as the 2023 melt season, have limited influence on the mean climate variables used in our analysis.

We apply similar reasoning to the lake observations. While warm temperature and high melt extremes can promote ponding, the presence of surface lakes also depends on firn conditions, which evolve over multiple years. Accordingly, we aggregate lake data across all available years to capture this longer-term preconditioning rather than year-to-year variability.

We will add this reasoning to the Methods section of our paper at line 98 (right before Table 1), as follows: "We use the full temporal extents of both the RACMO climate variables and the lakes datasets to capture processes relevant to both seasonal surface ponding and prior firn preconditioning."

Specific Comments

L22: It should be noted that not all parts of an ice shelf exert buttressing forces. See Fürst et al. 2016.

We will add the following statement to line 23: "This buttressing effect is complex and often nonlocal. Removing some ice shelf regions would have little to no effect on upstream ice, while removing others could accelerate upstream glaciers by up to eight times their current speed (Fürst et al., 2016)."

L28: Consider acknowledging the role of blue ice areas and rock outcrops in meltwater formation, given their characteristically lower albedo relative to surrounding snow and ice surfaces.

We will amend the sentence beginning on line 28 as follows: "Surface meltwater is widespread across Antarctica, occurring mainly at the lower latitudes and altitudes of ice shelves (Stokes et al., 2019; Tuckett et al., 2025) and near areas of blue ice and exposed rocky outcrops, where albedo is lower relative to surrounding snow and ice surfaces (Kingslake et al., 2017)."

L37: The causal interpretation for Larsen B warrants more careful treatment. Leeson et al. 2020 raise the question of whether lake drainage constitutes a cause or a consequence of ice shelf collapse, which should be acknowledged here.

To acknowledge this nuance, we will amend the sentence starting on line 36 as follows: "Similarly, the Larsen B Ice Shelf collapsed within days in February 2002, in an event associated with synchronized drainage of over 2,750 supraglacial lakes (Banwell et al., 2013; Cook & Vaughan, 2010; MacAyeal et al., 2003; Scambos et al., 2003). While it remains uncertain whether the lake drainage was a cause or a consequence of collapse (Leeson et al., 2020), extensive surface meltwater ponding is widely thought to have preconditioned the shelf for disintegration through hydrofracture.”

Figure 1: Please indicate the time period over which the average air temperature was calculated, and specify in the legend that 2 m air temperature is shown. It would also be helpful to clarify why the MOA color bar ends at 0.5 while the theoretical threshold is 0.7. Consider enlarging the maps for Dronning Maud Land, as the current size makes it difficult to resolve the gridded datasets.

In the Figure 1 caption, we will add, “Temperature and MOA are outputs from RACMO2.3p2 averaged over the period 1979-2022,” to make this more clear.

The MOA colorbar only goes to 0.5 because this maximizes the contrast across the range of MOA values present over most of the domain. However, the point is well taken that it would be helpful to indicate the areas that do reach the theoretical threshold, so we will extend the colorbar to 0.7. Values exceeding 0.7 are of course still included in the analysis; we will make this more clear in the figure by changing the colorbar shape to indicate that values extend beyond what is shown.

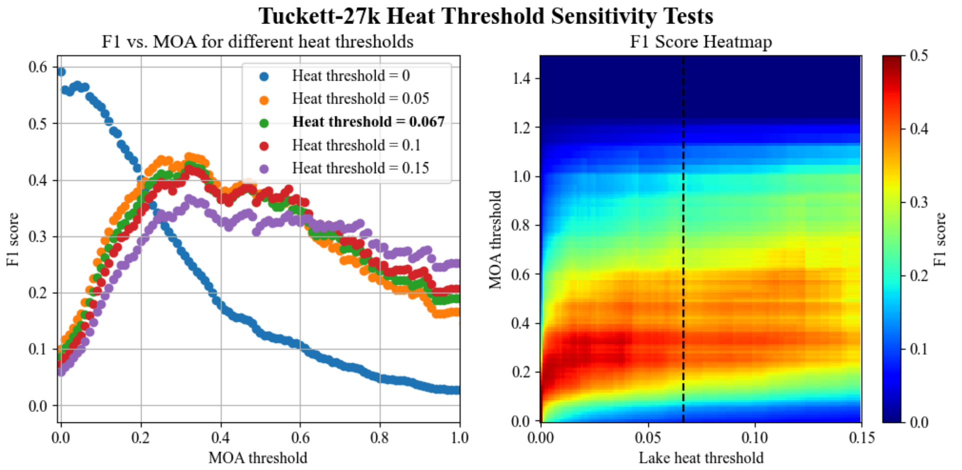
We will also make the Dronning Maud Land maps larger.

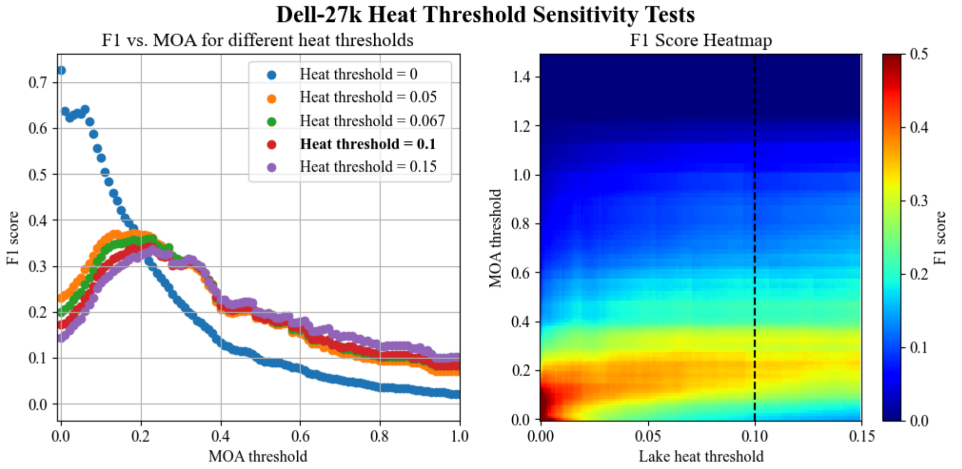
L112: Please clarify which year's ice shelf shapefiles were used for clipping. E.g. whether they predate or postdate the collapsed ice shelves when clipping RACMO. Depending on the observed time span some very important pixels might be missing.

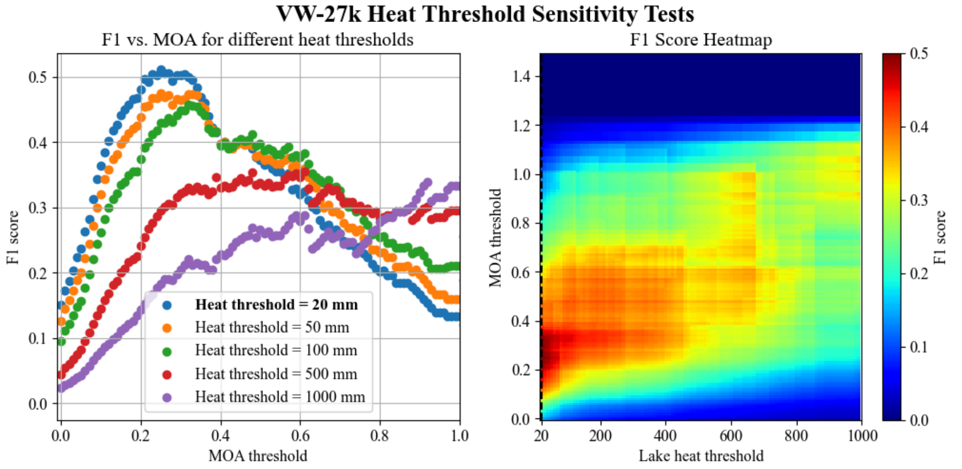
The shapefiles used for clipping were the combined maximum extent of shapefiles from MEaSUREs v.2 (Mouginot et al., 2017) and Dell et al. (2024a), described in Appendix B2, lines 587-590. The MEaSUREs product was created based on satellite data from 2007-2009, and Dell et al. (2024a) used vector polylines of the Antarctic coastline compiled from various mapping sources in 2020. None of these sources include the collapsed ice shelves that we discussed (Larsen A and Larsen B), as can be seen in the maps of the Antarctic Peninsula. We will add the following sentence to line 113 to clarify this: “Note that these shapefiles postdate the collapses of Larsen A and B ice shelves, and so these are not included in our analysis.”

L118: A sensitivity analysis systematically testing a range of thresholds and their respective effects on the results would help quantify the uncertainty associated with what is acknowledged to be a somewhat subjective choice.

We performed a sensitivity analysis similar to what you are describing, some of the results of which are shown in the figures below:







For Tuckett-27k and Dell-27k, with a heat threshold of 0 (i.e., the presence of a single lake pixel makes a grid cell a “wet” grid cell), we can see that the F1 score is optimized with a MOA of 0, since nearly all grid cells are now considered wet. While this gives us the highest possible F1 scores, it is not a physically meaningful result. For nonzero heat thresholds in all three scenarios, the F1 vs. MOA curves generally follow a similar pattern, although they peak at slightly different MOA's. Because of this, we are comfortable making a somewhat subjective choice, based on the logic described in the text (see lines 595-603 in Appendix B2).

We propose to not include this level of methodological detail in the manuscript.

Figure 4: The maps for Dronning Maud Land should be enlarged, as spatial differences in the gridded datasets are currently very difficult to see.

Agreed, we have made the maps for DML larger.

Figure 5: Please indicate the time span over which the average air temperature shown was calculated.

We will modify the Figure 5 caption to specify "average temperature" and "annual average MOA", instead of just "temperature" and "MOA" (line 176) , and add a sentence to the end of the caption that says "Temperatures and annual MOA's are averaged over the period 1979-2022 or 2023, as indicated in Table 1." (line 179).

Section 3.1. Average Air Temperature

L194: The characterisation "vastly underpredicts" may require revision in light of Table 2, which shows that the F1 score for the Tuckett-2k dataset is second highest at F1-score of 0.454, corresponding to an optimal temperature threshold of −9.2°C.

The F1 score of 0.454 corresponding with a temperature threshold of -9.2 °C is in reference to the result for the Antarctic Peninsula only. For the whole ice sheet, the Tuckett-2k scenario has an optimal temperature threshold of -10.8 °C, and the theoretical threshold of -9 °C corresponds with an F1 score of just 0.178. We will amend the text at lines 194-195 to make this more clear, as follows: "However, using T ≥ -9 °C as a ponding threshold vastly underpredicts present-day lakes, resulting in F1 scores less than 0.2 for all five comparison scenarios where the whole ice sheet was being considered."

L197: The finding that no single temperature threshold constitutes an effective predictor merits further discussion. It would be worthwhile to examine whether skin temperature, maximum temperatures, or mean temperatures over the summer months might offer greater predictive utility.

This finding corroborates the findings of Van Wessem et al. (2023), who noted that variable temperature thresholds for ponding on Antarctic ice shelves are dictated by the amount of snowfall accumulation in a given region. We will add this context to line 200 as follows "This broadly corroborates the findings of van Wessem et al. (2023), who noted that variable temperature thresholds for ponding are correlated with the amount of snowfall in a given region."

In response to your second comment of this review, we added an explanation of our choice to use air temperature instead of skin temperature (and the fact that the results do not qualitatively differ) to Appendix B1, lines 515-516.

It's certainly possible that maximum temperatures or summer-only averages would be better predictors. However, we consider this to be beyond the scope of this work, which is focusing on the utility of simple annual averages. We discuss these limitations in the discussion, Section 4.4: Limitations of a MOA threshold. On lines 435-438, we say: "However, a more accurate predictor would need to incorporate other climatic variables. We used multi-decadal average MOA values to represent the potential timescale of firn air depletion, but this long-term averaging deemphasizes warm temperature extremes and positive degree days, both of which have been tied to increased melting on the Antarctic Peninsula (Barrand et al., 2013; Trusel et al., 2015; Vaughan, 2006; Zheng et al., 2023)."

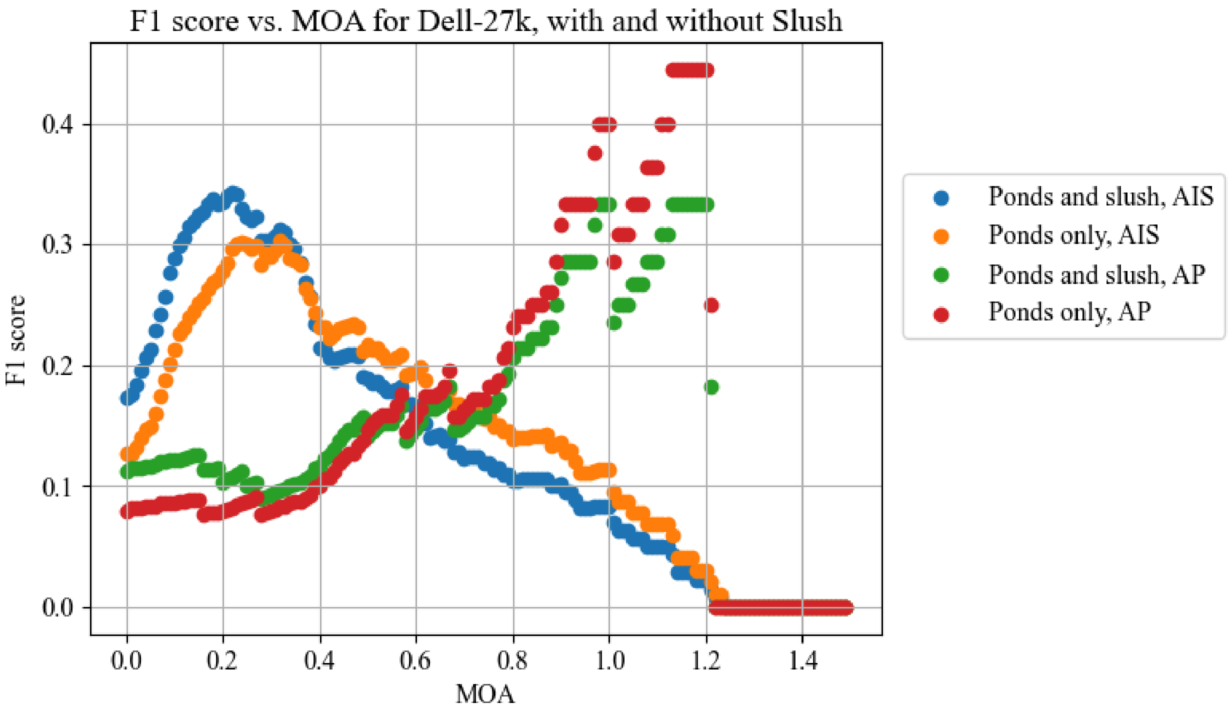
Discussion: A notable gap in the discussion is the absence of a section addressing the need for region-specific thresholds. A more physically motivated distinction between the Antarctic Peninsula and the East Antarctic Ice Sheet would likely yield more robust and actionable recommendations than the current AP versus non-AP split.

Exploration of the need for region-specific thresholds would indeed strengthen our discussion. We will add a paragraph at the end of section 4.4 (Limitations of a MOA threshold), starting on line 460, as follows:

"Due to the multitude of processes contributing to surface pond formation, region-specific climate thresholds are likely necessary for accurate ponding predictions. The role of downslope winds in driving surface melt varies across regions, depending on local topography and prevailing atmospheric conditions. East Antarctica experiences a high proportion of wind-driven surface melt (Laffin et al., 2023), so likely corresponds with a higher dependence on GLPI or a similar index than West Antarctica. East Antarctic ice shelves contain more widespread ice lenses within the firn than other regions (Tuckett et al., 2025), and both East Antarctic and Antarctic Peninsula shelves are predicted to experience increased ice lens formation between now and 2100 (Veldhuijsen et al., 2024). This suggests lower future MOA thresholds in these regions. As ice lens formation intensifies under future warming, these regional differences in ponding susceptibility may become increasingly pronounced."

Section 4.1: It should be noted in the discussion that the Dell et al. 2024 dataset includes areas of slush in addition to open water. Could this be a reason why in Figure 5 the Dell dataset is slightly different from the other datasets? Furthermore, the implications of the differing temporal coverages of the respective datasets for the comparability of results should be addressed.

The incorporation of slush is indeed an important factor in the Dell et al. (2024) dataset, but we examined the F1 scores and optimal MOA thresholds with and without the slush included and found that it didn't change the conclusions or the pattern of the results. The main results of this analysis are shown below:



For the continent-wide analysis (blue and orange scatter points), the Dell et al. (2024) data without the slush included (orange) has a slightly lower F1 peak at a slightly higher MOA, but the general shapes of the curves are the same. For the Peninsula only (green and red scatter points), we see that the F1 score increases almost monotonically with the MOA threshold, until around MOA=1.2. This unusual shape, relative to the other subplots in Figure 5, indicates that the Dell et al. (2024) data does include more melt pixels on the Peninsula than the other datasets, but this difference does not appear to be due to the inclusion of slush. We will add the following sentence in a new paragraph after line 370: "The Dell et al. (2024) dataset was the only one to include slush in its assessment of wet areas. We also performed the Dell-27k analysis with the slush excluded, and found that this did not alter the conclusions."

We agree that the RACMO models and lakes datasets spanning different time periods could influence comparability. In our analysis, the model outputs and data are used to characterize long-term spatial patterns in surface ponding and its climatic controls, rather than to resolve temporally matched events. In response to your earlier suggestion, we will add an explanation of why we used the entirety of the available extents to our Methods section at line 98, right before Table 1: "We use the full temporal extents of both the RACMO climate variables and the lakes datasets to capture processes relevant to both seasonal surface ponding and prior firn preconditioning." This choice allows us to maximize representation of all potential ponding locations, as opposed to restricting all data products to their overlapping period (2015-2021).

Handling data with different time spans and temporal resolutions is an important issue that is inherent to remote sensing. In this work, we addressed this issue by choosing unique lake coverage thresholds (for binarizing each grid cell as "lake" or "not lake") for each lakes data product, as discussed in Appendix B2 (lines 595-603).

L375: I am not a modelling expert but I wonder if you can actually do a proper comparison over different spatial scales when not all gridded datasets are based on the same RACMO version.

This is a good point, and we will clarify section 4.2 ("Comparison of grid resolutions"). On lines 375-376, we state, "The improved performance at 11 km likely reflects improvements in the underlying climate model", and then go on to discuss these improvements. We will make the limitations of the comparison more clear at the beginning of the paragraph starting at line 390, as follows: "However, the performance of MOA as a predictor depends not only on the underlying RACMO version, but also on the spatial resolution of analysis."

L414: The discussion would benefit from incorporating recent findings on the role of foehn winds in Dronning Maud Land from Mahagaonkar et al. 2025, rather than restricting this aspect of the discussion to the Antarctic Peninsula.

We will amend the sentence starting on line 413 as follows: "Foehn winds are especially prominent on the Antarctic Peninsula and in regions of East Antarctica." After the sentence where we detail the impact on the Peninsula, at the end of line 416, we will add: "In Dronning Maud Land, both katabatic and foehn winds have been shown to increase the presence of lakes near grounding zones through snow erosion, scouring, and surface warming (Mahagoankar et al., 2025). Similar influence of downslope winds on melt has been observed elsewhere in East Antarctica, including the Shackleton Ice Shelf (Saunderson et al., 2022) and the McMurdo Dry Valleys (Hofsteenge et al., 2022)." To tie this back to our results, on line 419, we will reference Figure 6 (A) (ii, iii), where we can see the difference in predictive performance between MOA alone and MOA with GLPI. We will say, "We also see that in Dronning Maud Land, the incorporation of GLPI removes some false negatives at the grounding line and some false positives further from the grounding line (Fig. 6 (A) (ii, iii)). However, the threshold still misses some lakes near the grounding line in the west side of the region."

Conclusion: The manuscript presents scientifically valuable findings that merit a more confident presentation in the conclusion. In particular, the key result from the future projections section, namely that substantially greater ponding is projected under the empirical threshold relative to the theoretical one, is currently underrepresented and should be stated more prominently. I would suggest to provide explicit guidance on region-specific thresholds for the Antarctic Peninsula and

EAIS/non-AP regions, and to state the optimum threshold values directly so that readers can readily reference them.

We will rewrite the conclusion (lines 460-474) to be more confident and actionable, as follows:

“Accurately predicting the distribution of surface meltwater on Antarctic ice shelves is critical for assessing areas that could be vulnerable to hydrofracture, particularly in regions of high extensional stress (Lai et al., 2020). In this paper, we quantified the performance of average air temperature, annual average MOA, and a weighted sum of MOA and GLPI as predictors of surface ponding for three different lakes datasets and three climate model grid resolutions. We find that the predictive power of MOA depends strongly on both climate forcing and region, and that it can be improved by incorporating GLPI to account for grounding line-adjacent processes that exert an important control on pond formation.

Empirically calibrated thresholds consistently outperform the theoretical threshold of MOA ≥ 0.7. For the VW-27k comparison scenario under emissions pathway SSP1-2.6, applying the optimized continent-wide threshold (F1 = 0.587) predicts 2.3 times more lake coverage in 2100 than the theoretical threshold (F1 = 0.261). Across data products, future projections using empirically calibrated thresholds yield substantially greater surface ponding than projections based on the theoretical threshold, suggesting that reliance on MOA ≥ 0.7 may systematically underestimate future meltwater extent on Antarctic ice shelves. Optimal threshold values differ between the Antarctic Peninsula and the remainder of the ice sheet, reflecting regional differences in melt processes, firn conditions, and the influence of downslope winds. For the VW-27k scenario, we identify optimal thresholds of MOA + 0.73(GLPI) ≥ 0.92 for the Antarctic Peninsula (F1 = 0.698) and MOA + 0.34(GLPI) ≥ 0.42 for the rest of the continent (F1 = 0.626).

Although MOA relies on simplifying assumptions about firn properties and does not explicitly resolve processes such as ice lens formation, it remains a simple, physically interpretable, and computationally efficient framework for predicting meltwater presence. If MOA is to be used as a diagnostic tool, we recommend it be applied at relatively coarse spatial scales (tens of kilometers), and that the threshold value be calibrated to the specific climate forcing using depth-weighted present-day lake observations. We also suggest that MOA be augmented to account for grounding line-adjacent melt processes, including downslope winds. Used in this way, MOA provides a practical tool for identifying ice-shelf regions susceptible to surface ponding and potential collapse.”

L585: The data aggregation procedure should be described in greater methodological detail. Downscaling from 30 m to 27 km resolution can introduce substantial errors depending on the approach taken, and a transparent account of the methodology is essential for reproducibility.

We will expand the description of the data aggregation procedure when it first comes up, in regards to the Tuckett et al. (2025) dataset on line 572. We will amend lines 572-574 as follows: “We aggregated this frequency map onto 2 km, 11 km, and 27 km grids to match the grids of our three RACMO products. This aggregation was done by assigning each high-resolution pixel to a RACMO grid cell based on its coordinates. For each RACMO cell, frequency values from all contributing 30 m pixels were summed, and the number of contributing pixels was recorded. We divided the summed frequency values by the total number of ice-shelf pixels contained within each grid cell to produce a normalized lake presence value for comparison.” Then, on line 585, we will reference this procedure by simply adding “using the same binning method as described above for the Tuckett et al. (2025) data.”

References mentioned:

Fürst, J. J., Durand, G., Gillet-Chaulet, F., Tavard, L., Rankl, M., Braun, M., and Gagliardini, O.: The safety band of Antarctic ice shelves, Nature Climate Change, 6, 479–482, 2016.

Leeson, A. A., Forster, E., Rice, A., Gourmelen, N., and Van Wessem, J. M.: Evolution of Supraglacial Lakes on the Larsen B Ice Shelf in the Decades Before it Collapsed, Geophysical Research Letters, 47, e2019GL085591, https://doi.org/10.1029/2019GL085591, 2020.

Mahagaonkar, A., Moholdt, G., Glaude, Q., and Schuler, T. V.: Katabatic and foehn winds control the distribution of supraglacial lakes in Dronning Maud Land, Antarctica, Earth and Planetary Science Letters, 666, 119482, https://doi.org/10.1016/j.epsl.2025.119482, 2025.

Morris, E. M. and Vaughan, D. G.: Spatial and Temporal Variation of Surface Temperature on the Antarctic Peninsula And The Limit of Viability of Ice Shelves, in: Antarctic Peninsula Climate Variability: Historical and Paleoenvironmental Perspectives, American Geophysical Union (AGU), 61–68, https://doi.org/10.1029/AR079p0061, 2003.

Thank you again for your comments. Please don’t hesitate to reach out with additional questions or concerns.

Best,
Emily Glazer and Kirsty Tinto