Dear Dr. Tuckett,

Thank you for your review. We respond to each comment in full below, with your original text reproduced in black and our responses in blue.

General Comment

Glen et al. provide a detailed analysis of surface slush extent across the Greenland Ice Sheet between 2016-2024. The dataset is novel and demonstrates the impressive scale of data processing achievable through cloud-computing and machine learning approaches. While the base method is not novel (the study builds upon the methodological framework of Dell et al.2022, 2024) and similar work has been conducted in Antarctica, this is the first ice-sheet wide study of slush in Greenland that I am aware of. The dataset itself is the strongest contribution of the work, providing a baseline for future process-based studies.

We thank the reviewer for this positive assessment of our dataset and the manuscript.

The manuscript is well-written, thoroughly referenced, and has a clear and logical structure. The methodological approach appears sound, although I think more clarity is needed regarding the modifications made to the Dell et al. (2022, 2024) workflow – in particular, the reason for using Sentinel-2 data instead of Landsat, and the methodological implications of this switch.

We agree that greater clarity is needed regarding our methodological modifications and will revise the methods section accordingly.

The results are well described and are supported by clear and well-designed figures. While the results are original due to the novelty of the dataset, they are not particularly surprising; they largely show the same interannual and seasonal patterns that have already been documented for supraglacial meltwater features (i.e. supraglacial lakes and streams) around Greenland. Similarly, the reported relationships between air temperature, snowmelt and slush extent are useful to show, but they offer limited new physical insight. The most interesting result is the relative area comparison between slush and surface lakes/streams, highlighting the abundance of slush in comparison to the more widely studied supraglacial meltwater features.

We acknowledge and agree with this point. We anticipate that the additional RACMO analyses described in response to '*Suggestion 1*' (below) will provide the deeper mechanistic insight that is currently lacking.

The main conclusions of the manuscript are that slush is widespread, significantly more abundant than other surface meltwater features (lakes and streams), and that slush could become an increasingly important factor for future ice sheet mass balance. The first two points are well evidenced by the data presented, but I think the manuscript is lacking in evidence to support this last point. The comparisons between slush area and RACMO snowmelt and air temperature do not provide much process-based insight, and there is no quantitative analysis linking slush to albedo, firn air content or surface energy absorption (as done in Dell et al., 2024).

We are pleased that the reviewer has acknowledged that our conclusions are well supported by the data presented. We agree that the third conclusion is currently underdeveloped relative to the evidence presented. We will revise the manuscript to either provide stronger supporting analysis for this claim or to moderate it appropriately where needed. We address the reviewer's specific suggestion regarding additional RACMO variables and process-based analysis with our response to '*Suggestion 1*' below. In terms of albedo and surface energy uptake, we will also consider whether this can be more effectively discussed using existing published frameworks scaled to the slush extents derived here, suggested by Reviewer 2.

Linked to this, I think the authors should compare their results with the results of Dell et al. (2024), highlighting differences in slush extent/variability etc. between Greenland and Antarctica, and the likely reasons for this. The 'Implications' section has the potential to be the most impactful part of the paper, but it is currently the shortest and weakest part of the discussion.

We agree that the '*Implications'* section requires development and that a comparison with Dell et al. (2024) would strengthen the manuscript. We address both points in response to '*Suggestions 2 and 3*' below.

Overall, I do not have any major issues with what the authors have submitted, and believe the dataset is a valuable contribution to The Cryosphere. My main criticism is that the data analyses and subsequent discussion are rather surface-level, meaning the paper currently lacks particularly insightful results or discussion. I think some further analysis of different RACMO datasets (e.g. snowfall precipitation & albedo), consideration of lags, and a more detailed discussion of the broader implications of the work would help to strengthen the impact of the paper.

We thank the reviewer for recognising the value of our work. We acknowledge that the current analysis is primarily descriptive and will address this by incorporating additional analyses and extending the discussion, as suggested.

Suggestions:

1) Compare your slush data with other RACMO variables, such as snowfall precipitation, albedo and firn air content (from a coupled firn densification model). Also consider temporal lags as part of your analyses and discussion.

We agree that incorporating additional RACMO variables would substantially strengthen the mechanistic interpretation of our results. Specifically, we will examine spatial and temporal relationships between slush extent and RACMO snowfall accumulation (as a control on snowpack water storage capacity) and firn air content from IMAU FDM (as a key determinant of the capacity of the firn column to absorb meltwater before saturation occurs). We will also explicitly examine potential temporal lags in these analyses (for example, examine whether antecedent snowfall or melt conditions in preceding months are more strongly correlated with slush extent than same-period values).

2) Compare your results with those from Dell et al. (2024), and discuss the potential glaciological and climatological reasons for the differences in slush proportions between the two ice sheets - e.g. why does slush in Antarctica account for ~50% of the total meltwater area, whereas in Greenland this percentage is much higher?

We intend to add a comparison with the results of Dell et al. (2024) to the revised Discussion, noting the contrast in slush proportions between the two ice sheets and offering qualitative discussion of the likely glaciological and climatological reasons for this difference, for instance varying climate, firn air content, surface mass balance, and slope.

3) Broaden your discussion section to incorporate new analyses, and provide a more nuanced discussion of the wider implications of your results.

We agree that the Discussion requires expansion and intend to address this through the addition of a new section, provisionally titled *'Controls on slush formation and persistence'*, as well as a revised '*Implications'* section. This new section will draw on the analyses described under '*Suggestion 1'* to move the paper beyond a purely descriptive account of slush distribution and towards a process-informed understanding of the climatic and glaciological factors governing where and when slush forms across the GrIS.

Specific comments:

Line 53: It is misleading to suggest that meltwater routing to the bed will always promote dynamic ice loss. Meltwater injection can cause both ice speed-up and slow-downs depending on the configuration of the subglacial drainage system. Suggest changing to 'influencing ice dynamics'.

The text will be revised in the revised manuscript.

Line 72: Given that you go on to show that slush area in Greenland can be 4-9 times greater than the area of surface meltwater features, this 50% statistic from Antarctica seems surprisingly low. You should come back to this in the discussion – why is there so much more slush area relative to surface meltwater area in Greenland than Antarctica (ice shelves)?

We agree this discrepancy warrants explicit discussion, and will address it as outlined in response to *'Suggestion 2'.*

Line 74: 'Total meltwater storage' implies Dell et al. (2024) comment on water volumes, whereas I believe they just look at slush/surface meltwater area.

This will be corrected to '*area*' in the revised manuscript.

Line 75: Be specific about what these climatological similarities are, as there are some key differences between the regions that should also be mentioned – e.g. elevation, moisture availability, FAC etc. (specifically, if comparing Greenland vs Antarctica slush as suggested, it is important to state these differences).

We will revise this sentence.

Line 101: I think in general, you need to be more specific about which parts of the method of Dell et al. (2022) you follow directly, and which elements you change. One key difference is that Dell et al. (2022, 2024) use Landsat 8 imagery, whereas this study uses Sentinel-2 – this difference is not explicitly stated. What is your reasoning for using Sentinel-2 and not Landsat- 8? What difference does this have on the method, especially the selection of thresholds during the masking and NDWI stages? Given the number of different thresholds (for both L8 and S2) used in both Moussavi et al. (2020) and Corr et al. (2022), I think you need to explicitly state the threshold values you applied to avoid confusion.

The methods section will be revised to clearly describe which elements of the Dell et al. (2022) workflow were followed directly and which were modified. We will explicitly state that this study uses Sentinel-2 rather than Landsat-8 and provide justification for this choice. All threshold values applied during the masking and NDWI stages will be explicitly stated, with the reader directed to Glen et al. (2025) where the Sentinel-2-specific thresholds are described in full.

Line 120: Why remove images with >25% cloud cover if you mask out cloud anyway? Could you not be missing lots of useful imagery?

This threshold is consistent with approaches adopted in comparable supraglacial meltwater mapping studies (e.g., Corr et al. 2022; Glen et al. 2025; Tuckett et al., 2025). We nonetheless acknowledge the reviewer's concern and will add a brief justification to the revised methods.

Line 172: How was this 'manual' review stage conducted? Was it one individual or did several people do this, with an average then taken? Distinguishing slush from other surface meltwater features is very subjective, and two people could classify an image in quite different ways.

The manuscript will be updated to clarify that this stage was conducted by a single expert analyst, consistent with the approach of Dell et al. (2022). We acknowledge that the distinction between slush and ponded water is inherently subjective, as the boundary between these classes is not always well defined, and this will be explicitly noted in the revised manuscript, with the addition of a supplementary figure to improve transparency around the classification workflow.

Line 203: I suggest you provide data in the Supplement to validate the statement that 'no year was disproportionately affected by cloud or temporal gaps'. Presumably some sectors had more frequent S2 images that others, and similarly, some sectors will have been more influenced by cloud cover than others. It would be good to show metrics for this, such as average cloud cover per sector and the average number of images used within monthly/annual mosaics, per sector.

We will add supplementary material presenting image availability and cloud cover statistics per sector, including the average number of images contributing to monthly and annual mosaics, allowing readers to assess whether any sectors or time periods were disproportionately affected by data gaps.

Line 221: These slush area values need some kind of uncertainty values reporting.

Uncertainty estimates derived from the RF classifier will be reported in the revised manuscript, with mapped slush area carrying an estimated uncertainty of ~0.2-1.3% based on classifier performance metrics, and will be reported alongside all slush area values in the revised Results.

Line 223: Why do you think this is? You don't really explain this in the discussion – given its southerly latitude and the high number of firn aquifers, it is slightly surprising that the SE sector has the least slush. Is there a climatological reason (e.g. more snowfall) for this?

We agree this is currently unexplained and will address it in the revised Discussion. We anticipate that higher snowfall accumulation in the SE sector may maintain a deeper snowpack that is less susceptible to full saturation, and will draw on our RACMO snowfall data and IMAU FDM firn air content to examine this more rigorously.

Line 226: I'm confused by the result here compared to the one stated in the next paragraph…. In Figure 3, there are dark orange areas in the NE showing slush recurrence in all 9 years, as stated in the following paragraph. So why here do you say that it only reappeared in the same location in four out of nine years?

The statement at line 226 refers to the mean persistence across each sector, whereas the following paragraph refers to the maximum pixel-level persistence, specifically the area in which slush was present in all nine melt seasons. We will clarify this distinction in the revised manuscript to ensure the two statements are not misleading when read in conjunction.

Line 248: What happens if you account for temporal lags with your correlations? There is likely a lag effect with higher air temperatures translating to increased surface slush, due to the time taken for a snowpack to saturate with melt. This likely explains the low R2 for this result. This section would be much stronger if you could quantify these time lags (e.g. do you get a stronger correlation if you correlate air temperature with slush the following month?). At a minimum, the potential for lag effects should be incorporated more into the discussion.

We agree and will explicitly test lagged correlations in the revised manuscript, examining whether antecedent temperature and snowmelt conditions yield stronger associations with slush extent than same-period comparisons.

Line 305: Again, there is no mention of lag effects here. Furthermore, you suggest that air temperature is not the sole driver of meltwater production (which is certainly true), but you do not explore this with other datasets. In particular, spatial and temporal variations in snowfall accumulation are not considered, which is likely crucial in determining meltwater storage capacity (and hence how likely it is to become fully saturated, resulting in slush). Your results would be much more impactful if you included some other variables provided by RACMO within your analyses.

This section will be substantively revised in conjunction with the additional analyses described under *'Suggestion 1'.*

Line 330: Some of the colours for the six sectors in Figure 7 are quite hard to distinguish. I suggest using more distinct colours.

The colour scheme in Figure 7 will be revised.

Line 332: It would be better to report the actual p-values (here and elsewhere), rather than simply stating p<0.05.

We report statistical significance using standard thresholds (e.g., $p < 0.05$), as the exact p-values do not affect the interpretation of the results.

Line 341: I think this whole section is missing a 'so what?' element to it. A greater surface area of the ice sheet is slush than lakes/stream, but why is this important? Is slush important for ice dynamics as with lake drainage events? Do lakes and streams still account for the majority of surface water volume? I know you can't necessarily answer this last question with your data, but some discussion of it is required.

We agree that the broader implications of slush dominance need to be more clearly articulated in the Discussion. We will revise this section to better address why the greater areal extent of slush matters, including its potential implications for meltwater storage, runoff generation, firn evolution, and possible links to ice dynamics, while also acknowledging the current limitations of our dataset in resolving relative water volumes.

Line 357: I don't think this statement is particularly helpful, because the same could be said for lakes and streams – e.g. sub-surface lakes, englacial channels. These comparisons also assume that surface area extent is the most important aspect, but there is no mention of water volume.

This statement will be removed from the revised manuscript.

Line 382: Can you quantify these links with ice slabs and firn aquifers? I think more discussion of specific examples related to this is required - E.g. Why are some ice slabs (e.g. in the NW sector) not associated with slush? Why is there such high persistence in some parts of the NE and NO where there isn't an ice slab?

This section will be revised to distinguish between slush occurring over bare ice, ice slabs, and firn aquifer regions, in response to comments raised by Reviewer 2. This revised framing will allow us to more directly address specific spatial patterns, including the limited slush associated with ice slabs in the NW sector and the high slush persistence observed in parts of the NE and NO in the apparent absence of ice slabs.

Line 402: Do you mean September 2022? You state in the methods that mapping was only conducted between May-September each year.

We will change this to '*late-September*' in the revised manuscript

Line 411: Could you not estimate long-term trends in slush based on the relationships you have derived with RACMO snowmelt? Or do you not think these relationships would hold going back in time?

We will consider the feasibility of using the derived relationships to estimate historical slush trends in the revised manuscript. However, key caveats would limit such an approach, including the potential non-stationarity of relationships over longer timescales as ice surface properties and firn structure evolve. We will propose this as a direction for future work.

Line 426: This final sentence is too sensationalist, and isn't supported by any of the data or results you have presented.

This sentence will be removed from the revised manuscript.

Line 431: Why haven't you done some of these steps? Wouldn't they be relatively simple to implement given you already have the RACMO data?

We will revise the manuscript to incorporate additional analyses of key RACMO variables, including snowfall accumulation, firn air content, as outlined under '*Suggestion 1',* and expand the Discussion accordingly. We note, however, that a fully comprehensive treatment of all controls on slush formation remains beyond the scope of the present study and represents an important direction for future work.

Technical corrections:

Line 46: Close bracket.

Line 47: Fullstop needed after 'laterally'.

Line 228: Grammar – 'one' not 'once'

All technical corrections will be made in the revised manuscript.

Dear Dr. Vandecrux,

Thank you for your review. We respond to each comment in full below, with your original text reproduced in black and our responses in blue.

The manuscript presents and analyses nine years of slush mapping based on annual and monthly (May to September) Sentinel-2 mosaics. The resulting dataset is very valuable, and the manuscript is clearly written, well illustrated, and suitable for publication in The Cryosphere. My only major concerns are as follows.

We thank the reviewer for this positive assessment of our dataset and the manuscript.

1. Although the uncertainty of the classification algorithm is currently provided, it is unclear how this uncertainty scales when the algorithm is applied to the entire ice sheet. Beyond the uncertainty quantification itself, all results are presented without uncertainty ranges, making it difficult to assess whether reported trends and interannual differences exceed the algorithm's intrinsic uncertainty. I wonder whether the random forest (RF) classifier outputs class probabilities that could be used to derive upper and lower bounds on slush extent, and potentially an uncertainty metric to be distributed with the dataset.

We agree that the presentation of uncertainty alongside the reported results is important for interpreting interannual variability and trends. We will incorporate uncertainty estimates derived from the RF classifier. Specifically, mapped slush area will be stated to have an associated estimated uncertainty of +1.3/−0.2%, based on classifier performance metrics. Uncertainty ranges will be reported alongside all slush area values in the revised Results.

I am also not fully convinced by the use of a K-means clustering algorithm to segment slush and non-slush areas during the training and validation process. To my understanding, unsupervised classification groups pixels based on spectral characteristics rather than physical meaning. Depending on the scene, the sharpness of the spectral difference (and transition) between slush and non-slush, and the image-specific number of classes, there is no guarantee that clusters will align with physical boundaries. I assume that the manual labeling step also revealed some ambiguity, with some slush clusters matching the visible slush extent better than others depending on the image. These ambiguities in the labeling and the manual curation process that follows therefore appear insufficiently documented: for example, where is the boundary drawn between slush and channels within slush areas? This combination of unsupervised classification and undocumented manual correction currently limits confidence in the uncertainty assessment. Providing illustrative examples of the workflow, including class maps and labeled images in the supplement, would help build that confidence.

While we agree that unsupervised clustering does not inherently correspond to physically meaningful classes, we note that the use of K-means clustering as an intermediate step to support manual labeling is consistent with previous studies, including Dell et al. (2022, 2024) and Halberstadt et al. (2020). We will revise the Methods section to clarify that K-means clustering is used to guide the identification of spectrally distinct regions, rather than to directly define physical classes. In practice, clusters are interpreted in the context of the underlying imagery and subsequently grouped into physically meaningful classes (e.g., slush and non-slush) through manual inspection.

We will also revise the manuscript to clarify that this stage was conducted by a single expert analyst, consistent with Dell et al. (2022). We acknowledge that the distinction between slush and ponded water is inherently subjective, as boundaries between these classes are not always well defined, and this will be explicitly recognised in the revised manuscript. To further improve transparency, we will include illustrative examples of the workflow in the Supplement to demonstrate how ambiguous cases are handled in practice.

> Finally, comparing expert delineations of “uncertain slush” areas with the RF-derived uncertainty metric would help demonstrate that the proposed uncertainty quantification is realistic. These points are intended as suggestions, and alternative solutions consistent with the authors’ methodology are of course welcome.

We will consider incorporating an additional supplementary figure, or similar supporting material, consistent with our existing methodology.

2. The analysis of the slush maps is limited to their spatio-temporal patterns (with some repetition between regional and ice-sheet-wide behavior) and rather superficial comparison with RACMO estimates of melt and air temperature. Deriving additional insights on the formation of slush and/or impact of the slush on the energy balance, would greatly increase the manuscript's relevance. For instance, a potentially valuable analysis would be to study the role of surface topography and annual snowfall in controlling slush formation and its persistence from year to year. It could also support more informed discussion of future evolution: for example, whether slush can form everywhere on the ice sheet, and whether increased melt could shift its occurrence to higher elevations. Additionally, although the impact of slush on surface albedo and energy uptake is not directly quantified, it could be discussed using previously published work scaled to the slush extents derived here.

We agree that, in its current form, the analysis is primarily descriptive and that further process-based interpretation would strengthen the manuscript and its broader relevance.

We will revise the manuscript to incorporate additional analyses aimed at improving mechanistic understanding of slush formation and persistence. In particular, we will assess the role of firn air content from IMAU FDM, snowfall accumulation and using RACMO outputs, and consider how previously published albedo-energy balance frameworks may be applied to the slush extents derived here. We will also consider topographic controls using the ArcticDEM. These analyses will allow us to better constrain the factors governing where slush forms and persists across the ice sheet, and to move beyond purely spatio-temporal description. Please also see our response to a similar comment made by Reviewer 1.

Addressing these points, together with the minor comments below, would improve the robustness and depth of the study and make it a valuable contribution to The Cryosphere.

We thank the reviewer for recognising the value of our work, and will address the points raised.

l.19 "maximum" -> "minimum"?

Maximum is correct, as maximum slush areal extent is the metric used (which we will ensure is clearer)

l.46 missing parenthesis after "Machguth et al., 2022" / l.47 missing period after "laterally"

Both typographical errors will be corrected.

l.127–130 How does the use of the "high-end" ice slab extent impact the statistics presented later on? Why use this instead of a more reasonable mid-range estimate?

We will revise the manuscript to include both high- and low- end estimates of ice slab extent.

l.253 "trend": do you mean temporal trend? Please check for other instances where "temporal" might be missing before "trend".

We confirm that 'trend' refers to temporal trend, and this will be clarified throughout the manuscript.

l.319–320 This statement seems counterintuitive. One would expect the slush area to be located at lower elevation in the northern regions. Could you elaborate?

We agree that this statement is unclear and does not add substantial value. It will be removed from the revised manuscript.

l.365–370 SW is also the region where ice flow, melt rates, and surface topography lead to the longest bare ice duration, which favors bare ice darkening (Feng et al., 2024). Do these same settings also shape the slush area above the bare ice zone? Reading this paragraph, I am also wondering about areas that may be mapped as slush early in the melt season and later turn into bare ice after depletion of the snowpack. This could be briefly discussed, as it may introduce noise in the slush extent versus elevation relationship: when slush reaches its maximum elevation, low-elevation slush areas may already have transitioned to bare ice.

We agree that the same controls that promote prolonged bare ice exposure in the SW (i.e. high melt rates, ice flow, and surface topography) are also likely to influence slush distribution, particularly by limiting its persistence at lower elevations. This will be clarified in the revised manuscript, with reference to Feng et al. (2024).

We also clarify that our elevation analysis is based on the maximum annual elevation of slush, rather than its simultaneous distribution, and therefore captures the highest inland extent reached during the melt season.

l.374 "30% of Greenland's slush area lies above ice slabs" Here, as in many other places, I miss the uncertainty range associated with this value, both from your slush maps and from the ice slab maps.

We agree that uncertainty should be reported and we will revise the manuscript to include uncertainty ranges for this (and other similar) value(s), incorporating uncertainties from both the slush classification and the ice slab dataset.

l.378 "regions lacking underlying low-permeability layers" There are several issues with this framing. (1) There is substantial slush mapped in the bare ice area in Fig. 3. These should be counted separately and definitely have an underlying low-permeability layer: the ice surface. (2) Firn aquifer regions are known for deep meltwater infiltration enabled by temperate firn and the absence of refreezing. Even if these areas do present ice layers (e.g., Miller et al., 2020, Fig. 6), they should not be grouped with ice slabs, which are defined by their low permeability. Finally, there is evidence of physical limits to meltwater infiltration in Greenland firn in the absence of massive ice layers, driven instead by contrasts in grain size and density (Humphrey et al., 2021). By documenting slush outside bare ice and ice slab areas, you may be providing the first observation of meltwater ponding due to minor ice lenses combined with density and grain-size contrasts, as studied experimentally by Humphrey et al. (2021).

We agree that this framing requires revision. We will separate slush occurring over bare ice (i.e. a low-permeability surface) from slush occurring in firn areas, using the firn area as delineated by Vandecrux et al. (2019), to distinguish these zones explicitly. We additionally agree that firn aquifer regions should not be grouped with ice slab areas given their fundamentally different hydrological behaviour, and will ensure this distinction is clearly made in the revised text. Finally, we agree that slush documented outside bare ice and ice slab areas may represent observational evidence of meltwater ponding driven by minor ice lenses, as studied experimentally by Humphrey et al. (2021), and will discuss this possibility in the revised manuscript.

l.382–384 Could this paragraph be more quantitative? If there is topographic control on slush formation, slush areas should be largely persistent for years with similar melt. I find non-persistent slush areas in unexpected zones such as the firn aquifers suspicious at first glance. It would be helpful to provide either an RF-derived uncertainty estimate for these cases or at least an expert assessment of the reliability of these non-persistent slush events.

This section will be substantively revised in response to your related comment above and will include a more quantitative treatment of these patterns.

References:

All following references have been added to the manuscript.

Feng, S., Cook, J. M., Naegeli, K., Anesio, A. M., Benning, L. G., & Tranter, M. (2024). The impact of bare ice duration and geo-topographical factors on the darkening of the Greenland Ice Sheet. Geophysical Research Letters, 51, e2023GL104894. https://doi.org/10.1029/2023GL104894

Humphrey, Neil F., Joel T. Harper, and Toby W. Meierbachtol. “Physical Limits to Meltwater Penetration in Firn.” Journal of Glaciology 67, no. 265 (2021): 952–60. https://doi.org/10.1017/jog.2021.44.

Miller, O., Solomon, D. K., Miège, C., Koenig, L., Forster, R., Schmerr, N., et al. (2020). Hydrology of a perennial firn aquifer in southeast Greenland: An overview driven by field data. Water Resources Research, 56, e2019WR026348. https://doi.org/10.1029/2019WR026348