The authors thank the reviewer for their insightful comments and suggestions that have helped us to improve the manuscript. We have carefully considered all the points raised and have revised the manuscript accordingly. In the following, we provide a point-by-point response to the reviewers' comments.

Original comments from the reviewers are shown in gray.

Authors replies are shown in bold.

Excerpts from the manuscript are show in italic.

I would recommend the authors to consider to address 2 points: (1) the potentials and possible challenges if one would extend this work to a wider time frame with stronger spatial variability which can have a bigger potential to impact the surface energy budget?

Of course extending the method to a wider area is in general possible if the general characteristics of the surface types in this area are similar to our training data. A stronger variability of the same surface types does not affect the algorithm. However, adjustments, refinements and an extension of the training data will be needed when additional surface types, e.g., melt ponds, are present. Similar the application to observations in other seasons will be challenging. E.g. in summer when skin temperatures of sea ice and sea water converge to -1.7°C a distinction from thermal IR observations becomes impossible. This would require the integration of data from other sensors (e.g., hyperspectral or microwave sensors) to improve classification accuracy. We addressed this points together with the second comment in Section 5 of the manuscript.

(2) the classification scheme is a great start and provides an inspiring direction for the arctic community, but i wonder if it is realistic to upscale this method to provide a pan-arctic insight?

We thank the reviewer for this question. While direct upscaling of our airborne measurement technique to a pan-Arctic scale is not feasible due to its limited spatial and temporal coverage, the methodology and findings offer a crucial pathway for improving pan-Arctic analyses. Our high-resolution data could serve as "ground truth" for calibrating and validating satellite-based products. As shown in our analysis, our dataset is increasingly dominated by snow-covered ice further away from the sea-ice edge. This makes it difficult for a model trained solely on this data to learn the full heterogeneity of the marginal ice zone (MIZ), reinforcing the

challenge posed by the limited spatial and temporal coverage of a single airborne campaign.

Ln 323: Extending this analysis to different seasons will be crucial for capturing processes like melt pond evolution, though this will require multi-sensor data fusion to resolve the increased complexity of surface types. Therefore, the primary value of this high-resolution methodology lies in providing "ground truth" for calibrating satellite retrievals and refining sub-grid-scale parameterizations in pan-Arctic models.