Supplement of The Cryosphere, 20, 4345–4365, 2026
https://doi.org/10.5194/tc-20-4345-2026-supplement

The Cryosphere
Open Access

Supplement of

A nine-year record of slush on the Greenland Ice Sheet

Emily Glen et al.

Correspondence to: Emily Glen (emily_glen@psu.edu)

Supplementary Sections

S1. Spectral assessment of cluster labels

To clarify and quantify the basis of the expert labels, cluster assignments were made through visual interpretation of the S2 true-colour images, supported by $NDWI_{ice}$ and the spatial position of each cluster relative to visible slush fields. No single $NDWI_{ice}$, hue, or saturation threshold was used because the spectral characteristics of slush varied between images and overlapped with those of neighbouring non-slush surfaces.

We extracted the distributions of $NDWI_{ice}$, hue, and saturation for all pixels assigned to each cluster across the 18 training images. Hue and saturation were calculated by converting the S2 red, green, and blue bands into HSV colour space. Hue represents the colour, while saturation describes how strong that colour is compared with a grey or white surface. Across clusters containing valid observations, clusters labelled as slush had pixel-weighted pooled mean ± standard deviation values of 0.177 ± 0.079 for $NDWI_{ice}$, 0.584 ± 0.011 for hue, and 0.295 ± 0.095 for saturation. The corresponding values for non-slush clusters bordering the slush fields were 0.221 ± 0.089, 0.578 ± 0.012, and 0.354 ± 0.106, respectively.

The mean ± standard-deviation intervals overlapped over $NDWI_{ice}$ values of 0.132–0.257, hue values of 0.573–0.590, and saturation values of 0.248–0.390. This overlap shows that $NDWI_{ice}$ or colour alone cannot provide a universal distinction between slush and non-slush and identifies the ranges within which spatial context and expert interpretation were most important. Cluster-level statistics and pixel counts are reported in Table S2.

S2: Comparison of RF classifier to thresholding methods

To benchmark the RF classification against established threshold-based approaches, we compared a subset of the RF-classified slush results with classifications produced using two threshold-based methods. All images included in this comparison were pre-processed using the same masking approach described in Section 2.2.1 of the main text.

S2.1. Normalised Difference Water Index

As stated in the main text (Section 2.2.1), two NDWI-based indices are often used to detect meltwater features on glacier ice: $NDWI_{standard}$ (e.g. McFeeters, 1996; Box and Ski, 2007) and $NDWI_{ice}$ (Equation 1; main text), an adaptation optimised for ice- and firn-covered surfaces (e.g. Yang and Smith, 2012). $NDWI_{standard}$ is calculated from the green and near-infrared bands, as defined in Equation S1:

$$NDWI_{Standard} = \frac{B3 - B8}{B3 + B8} \tag{S1}$$

Slush has been detected using an $NDWI_{ice}$ threshold alone, which is generally lower than those applied to open water and lakes. For example, Williamson et al. (2018) identified lakes using a threshold of > 0.25, whereas slush is typically detected at lower $NDWI_{ice}$ values. For the GrIS, Yang and Smith (2012) first proposed classifying shallow water and slush using a threshold range of $0.12 < NDWI_{ice} < 0.14$ for both classes, which were later adopted by Bell et al. (2018) for the AIS. More recently, Glen et al. (2025) showed that combining both types of NDWI indices with respective thresholds, applying $NDWI_{standard} > 0.14$ alongside $NDWI_{ice} > 0.15$, improved slush detection on the GrIS while reducing misclassification with other meltwater features. A similar benefit of combining NDWI indices was also reported by Corr et al. (2022) in their mapping of supraglacial lakes on the GrIS.

In this study, we compared our RF classification results to NDWI threshold-based methods following Glen et al. (2025). Specifically, slush was identified using the combined thresholds of $NDWI_{ice} > 0.15$ (Equation 1) and $NDWI_{standard} > 0.14$ (Equation S1). All agreement metrics presented below (and in Figure S5) therefore relate to this combined-threshold method.

S2.2. Greenness Index

We also compared our RF results to the Greenness Index (G_{ind}) developed by Covi (2022). This index enhances green reflectance relative to blue and red, while reducing cloud contamination through the inclusion of the shortwave infrared band. High values of G_{ind} indicate strong green reflectance compared to other visible bands, which produces a blue hue in true-colour composites and is characteristic of surface water (Covi, 2022). The shortwave infrared (SWIR; B12) is used to filter out clouds: since clouds are highly reflective in SWIR, high B12 values lower G_{ind}, reducing the likelihood of clouds being misclassified as slush.

G_{ind} is defined in Equation S2:

$$G_{ind} = \frac{3 \cdot B3}{B2 + B3 + B4} - 2 \cdot B12 \tag{S2}$$

In our study, we applied a threshold of $G_{ind} > 1$ to classify slush. This value was determined through iterative testing against Sentinel-2 imagery. We found that > 1 provided the best match to visually identifiable slush areas.

S2.3. Analysis of agreement metrics

Agreement between our RF classifier and the NDWI and G_{ind} methods was assessed through analysis of spatial overlap and classification metrics (Tables S5 and S6). For these comparisons, the RF classification was treated as the reference classification, and precision, recall, and F1-score were macro-averaged across the slush and non-slush classes. Overall, RF showed stronger correspondence with NDWI than with G_{ind}. Spatial overlap between RF and NDWI was highest in the SW (733 km^2) and the CW (41 km^2), while overlap between RF and G_{ind} was lower, peaking at 436 km^2 in the SW and just 11 km^2 in the CW. Classification metrics reflected similar trends (Tables S5 and S6). For example, F1-scores were generally higher for RF compared to NDWI (0.56–0.76), indicating stronger overall agreement than RF compared to G_{ind} (0.51–0.68).

Visual comparisons between the three methods supported the quantitative results described above. In the CW region (Figure S5a), RF and NDWI were closely aligned spatially, while G_{ind} generally underestimated areas of slush and misclassified lake edges. However, in the NO (Figure S5b), NDWI overestimated areas of slush due to cloud contamination, whereas the RF classification was less affected by cloud contamination in this example. In the SW (Figure S5c), the RF classifier identified more extensive slush than either threshold method. G_{ind} again showed lower agreement, including some misclassification of shallow lakes compared to both RF and NDWI, while NDWI occasionally misclassified bright surfaces as melt. Overall, RF provided more spatially consistent results, particularly in noisy or complex conditions.

S3. Limitations of our RF classifier

Our results shown in Section S1 above as well as the main paper highlight the strength of RF in capturing the spectral expression of slush. That said, our approach has limitations. For example, we observe occasional overestimation in areal extents of slush (e.g., in the SW in 2019; Figure S5c), and some subjectivity in the manual clustering step (see Section 2.3). Despite cloud and terrain masking, commission errors may occur in certain conditions (e.g., cloudy scenes and regions with complex topography) and low sun angles can obscure melt features, especially in early and late summer. Even so, relative to previous studies, our RF approach offers a major step forward in reliably mapping slush across Greenland, with strong potential for continued refinement and wider application. Improving cloud masking, incorporating time-series filtering, and exploring multi-sensor approaches offer promising paths to enhance future slush detection and better constrain the full extent of meltwater saturation.

Supplementary Figures

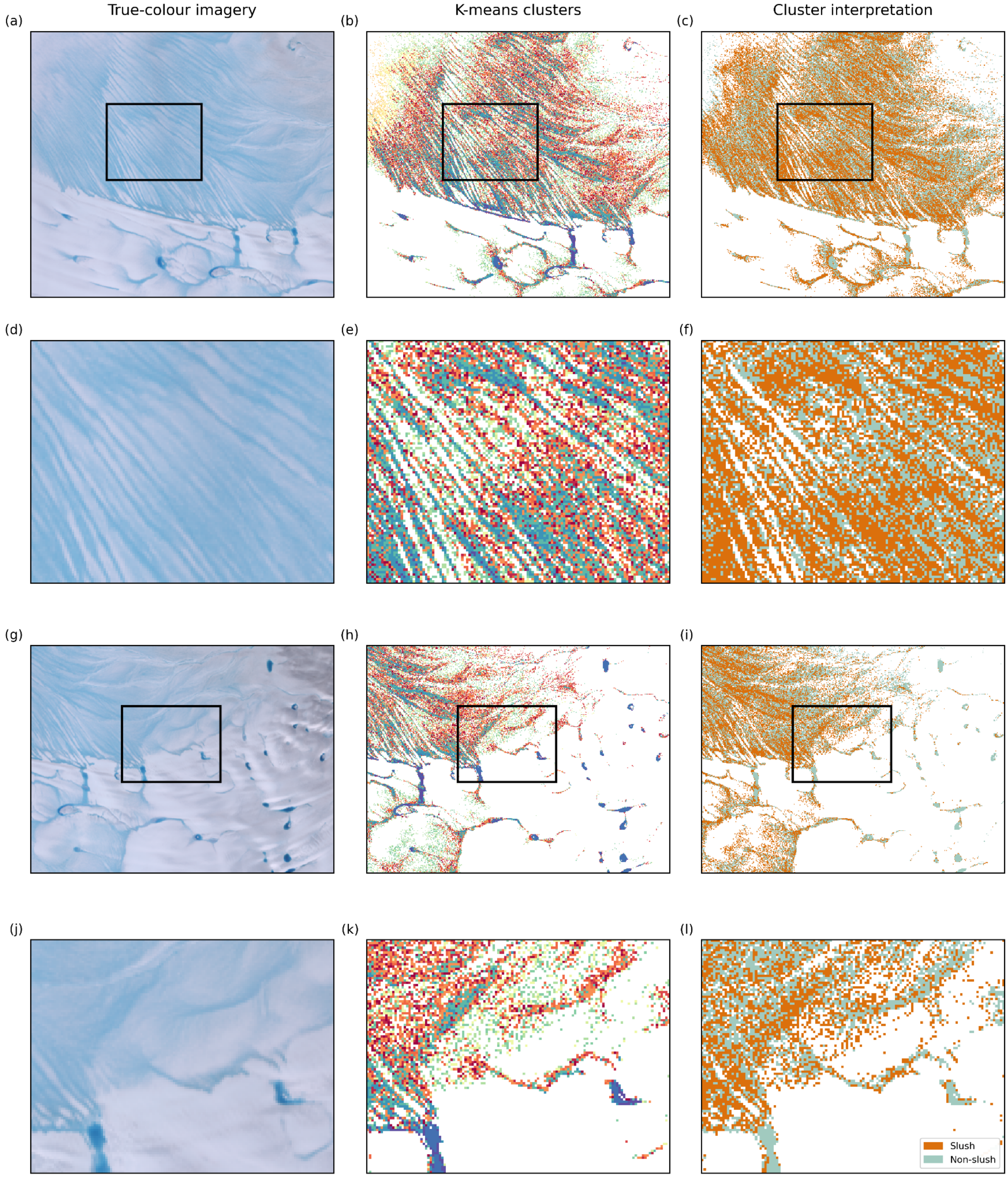

Figure S1. *Examples of the k-means-based training data generation workflow used for slush classification. True-colour Sentinel-2 imagery (left), corresponding k-means spectral clusters (centre), and manually assigned cluster labels used to define slush and non-slush classes (right) are shown for two representative examples in panels (a–c) and (g–i). Black rectangles indicate the areas enlarged in panels (d–f) and (j–l). Colours in the k-means panels represent arbitrary spectral clusters and do not correspond directly to surface classes. Orange denotes slush and turquoise denotes non-slush.*

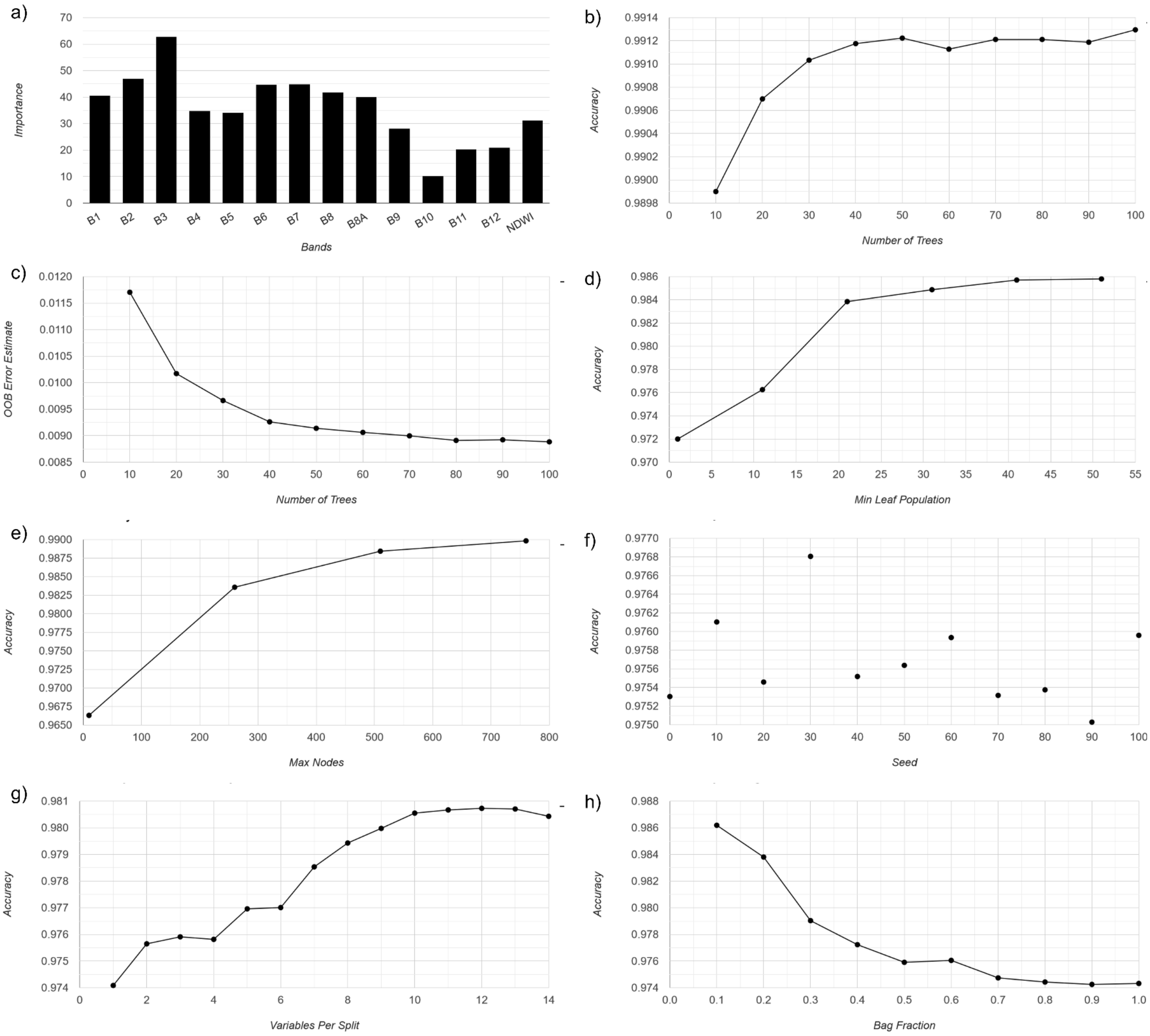

Figure S2. *Hyperparameter tuning results for the RF classifier. (a) Feature importance for the selected bands used in classification. (b) The effect of number of trees on model accuracy. (c) OOB error estimate as a function of number of trees. (d) Model accuracy as a function of minimum leaf population. (e) Model accuracy as a function of maximum number of nodes per tree. (f) The effect of random seed value on model accuracy. (g) Model accuracy as a function of number of variables considered per split. h) Model accuracy as a function of bag fraction.*

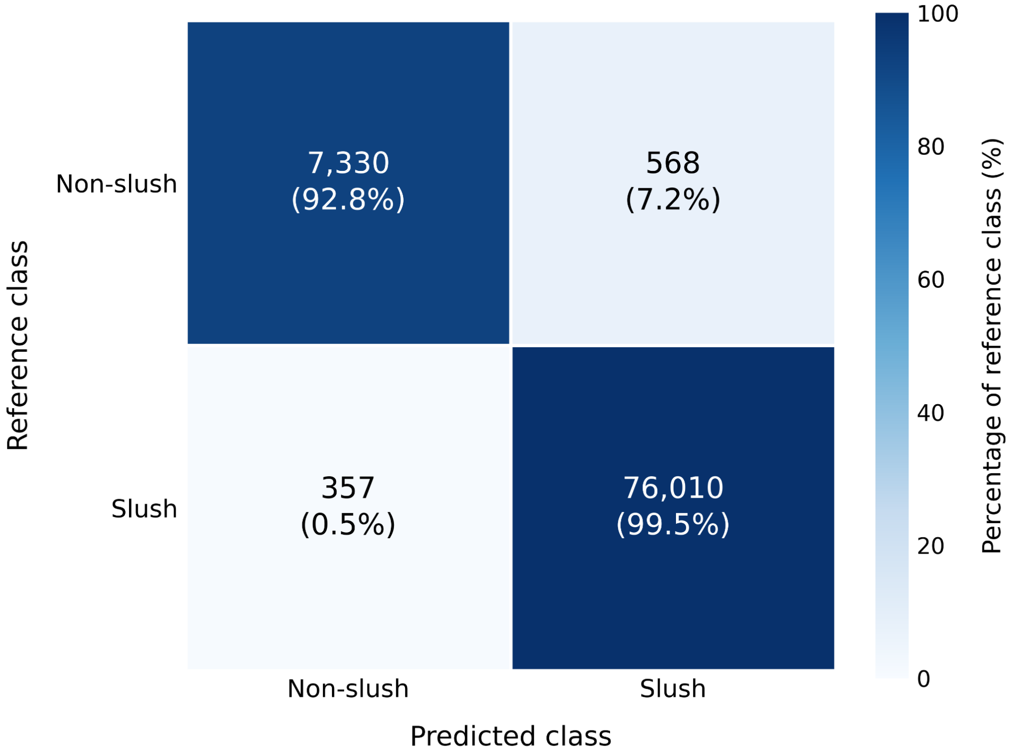

Figure S3. *Confusion matrix showing the internal validation performance of the RF classifier using the withheld 20% subset of labelled pixels. The matrix compares predicted classes with expert-derived reference labels for the slush and non-slush classes. The vertical axis represents the reference labels, while the horizontal axis represents the predicted classes. Cell annotations show the number and percentage of observations within each reference class.*

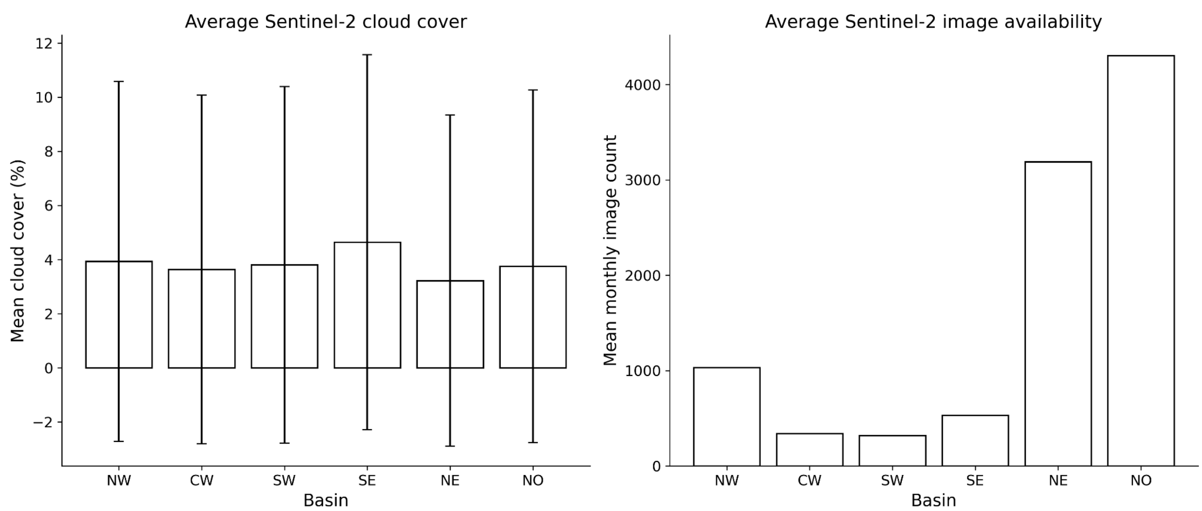

Figure S4. *Mean S2 cloud cover (%) and average monthly image availability across GrIS drainage basins during the 2016–2024 melt seasons. Error bars show ±1 standard deviation* *in cloud cover. Results indicate consistently low cloud contamination and sustained image availability across all basins throughout the study period.*

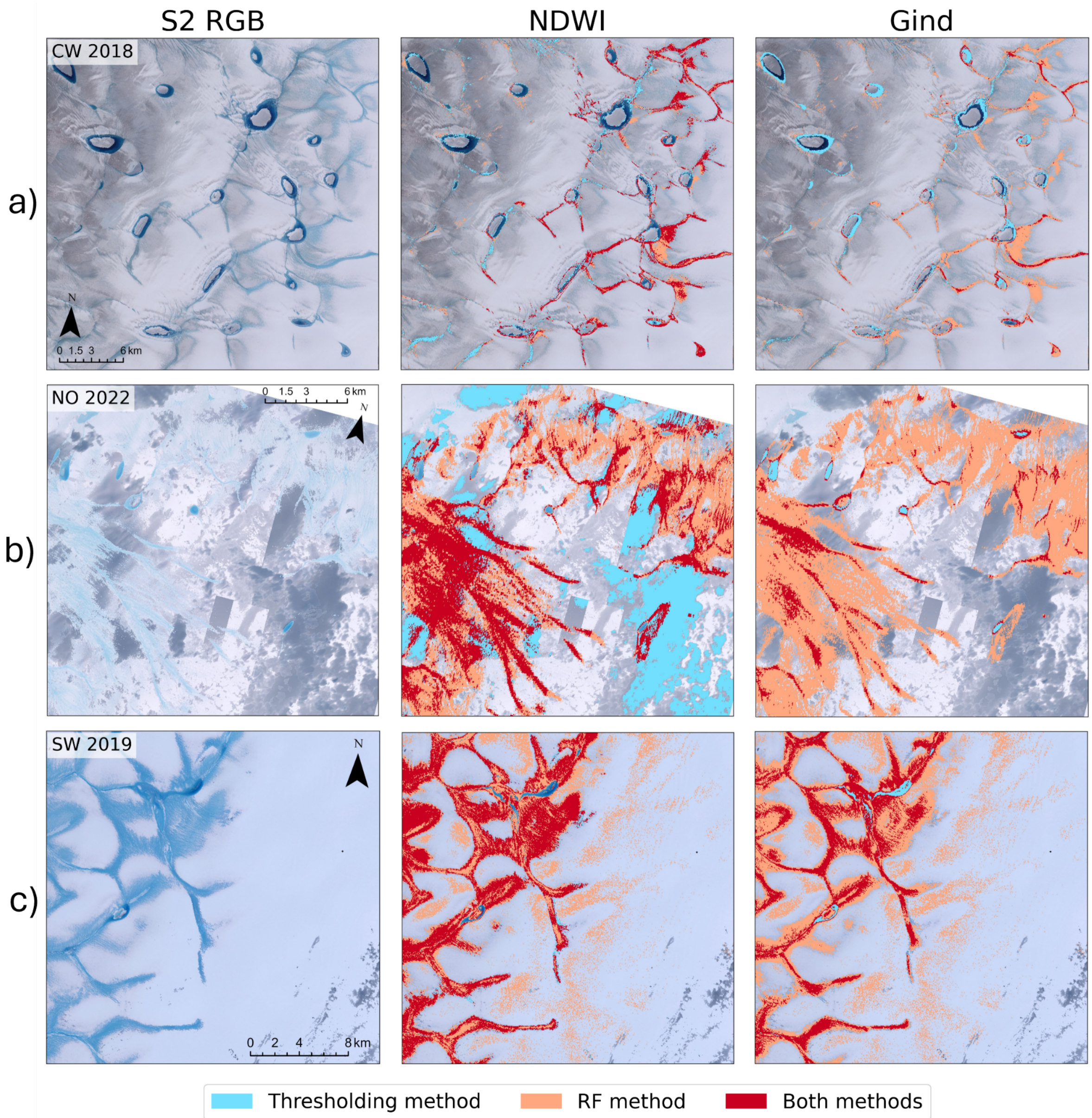

Figure S5. *Visual comparison of our RF method with both NDWI ($NDWI_{standard}$ and $NDWI_{ice}$ combined method; Glen et al., 2025) and G_{ind} (Covi, 2022) thresholding methods for slush detection, applied to S2 optical imagery across three GrIS* basins *and years: (a) CW in 2018, (b) NO in 2022, (c) SW in 2019 (Figure 2; main text). The left column displays the original S2 RGB images for reference. The middle column shows NDWI-classified slush compared to RF-classified slush, and the right column shows G_{ind}-classified slush versus RF-classified slush. Red indicates slush detected by both RF and the respective thresholding method, orange represents slush identified only by RF, and blue indicates slush identified only by the thresholding method. Of note, the NO 2022 image was intentionally selected as a complex* *case due to significant cloud cover and artefacts such as visible seams and spectral discontinuities between adjacent scenes resulting from the mosaicking process.*

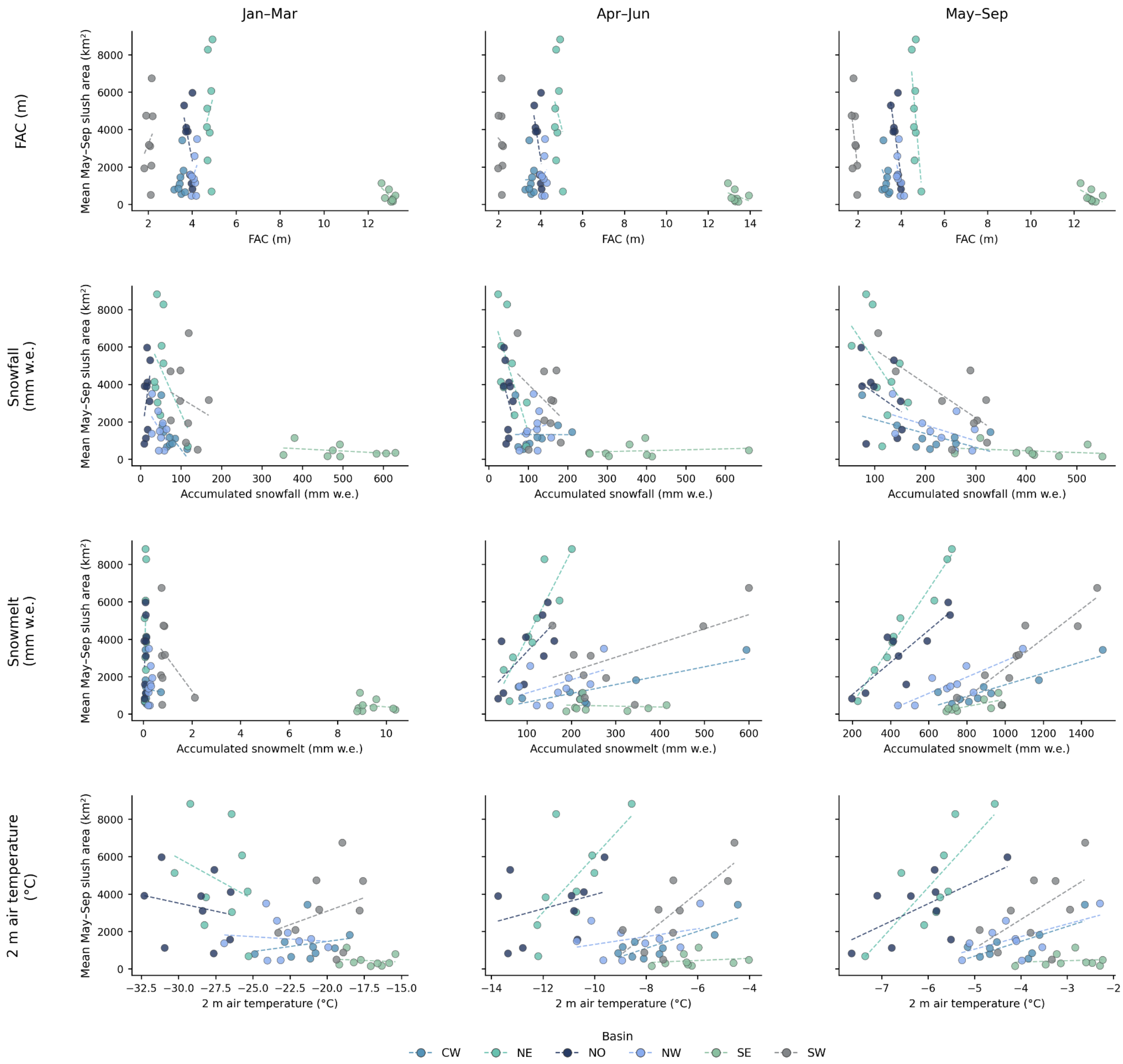

Figure S6. *Basin-wise Spearman rank relationships between mean May–September slush area and seasonal firn and climatic variables. Predictor variables were averaged across the potential slush zone within each basin and include mean FAC and 2 m air temperature, and accumulated snowfall and snowmelt, calculated over January–March, April–June, and May–September. Points are coloured by basin, with dashed lines showing basin-specific linear trends. Correlations were calculated separately for each basin using nine annual observations, or eight for FAC.*

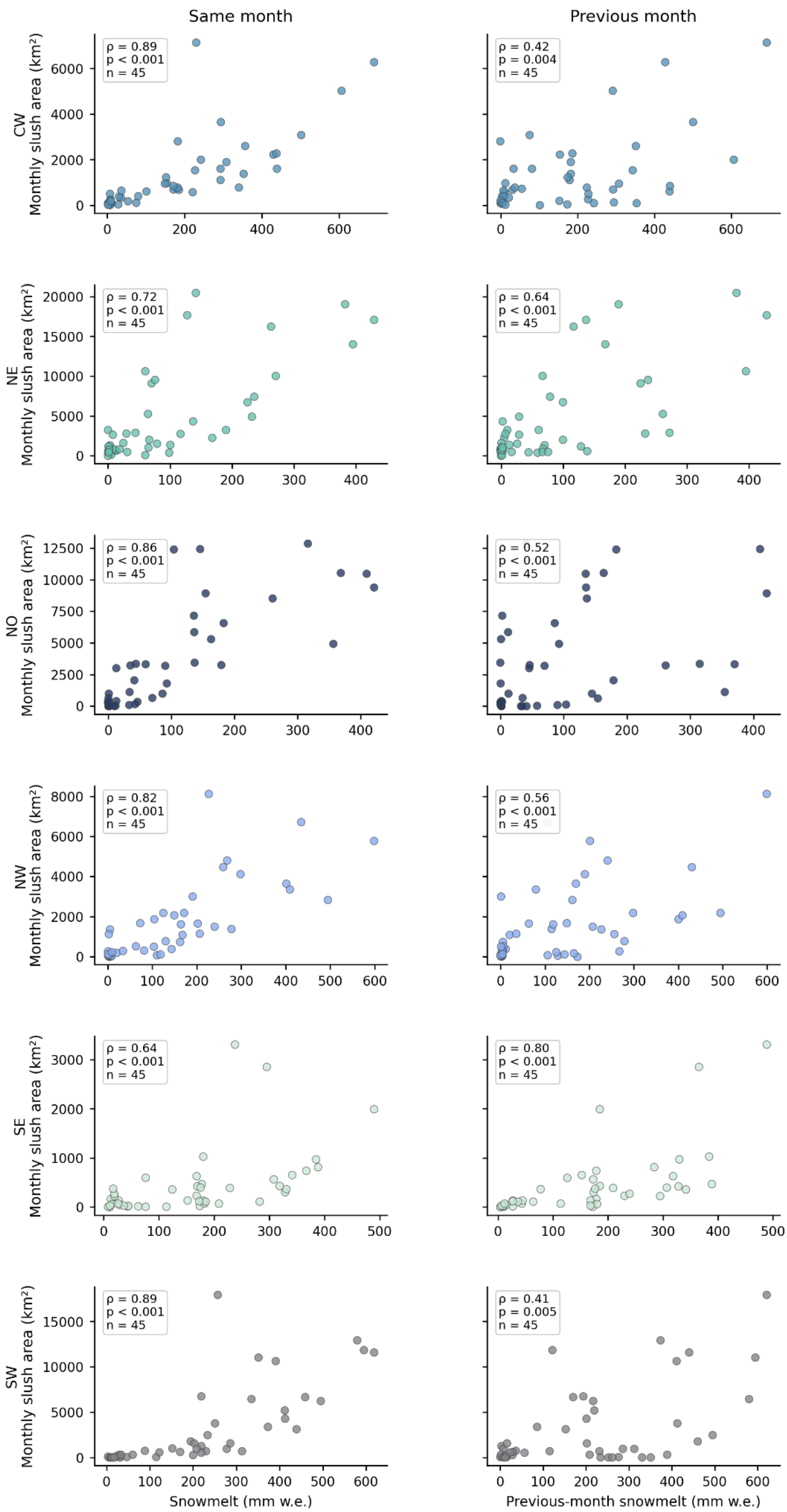

Figure S7. *Spearman rank relationships between monthly slush extent and same-month or previous-month snowmelt at the intra-annual basin-wise scale.*

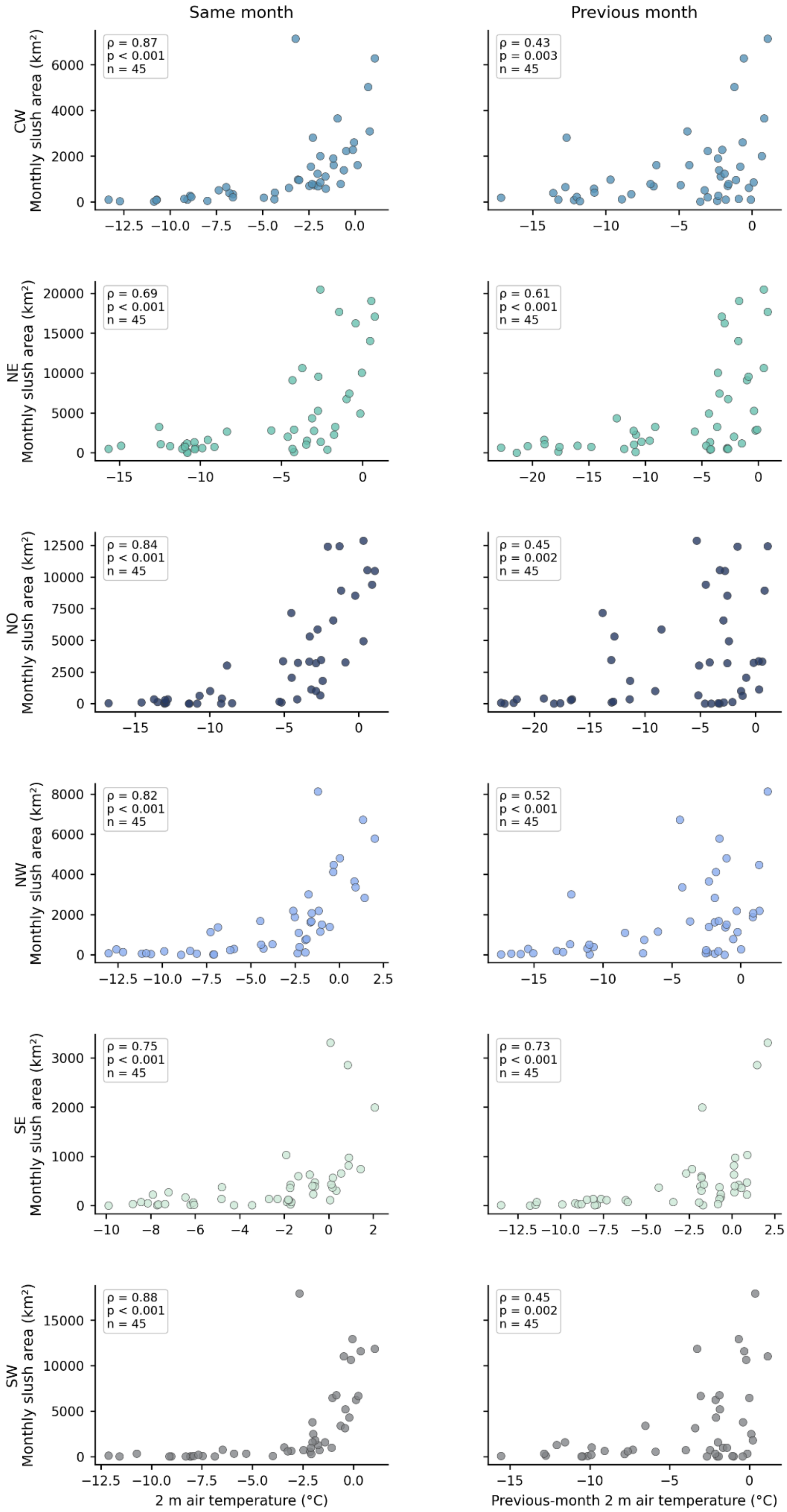

Figure S8. *Spearman rank relationships between monthly slush extent and same-month or previous-month 2 m air temperature at the intra-annual basin-wise scale.*

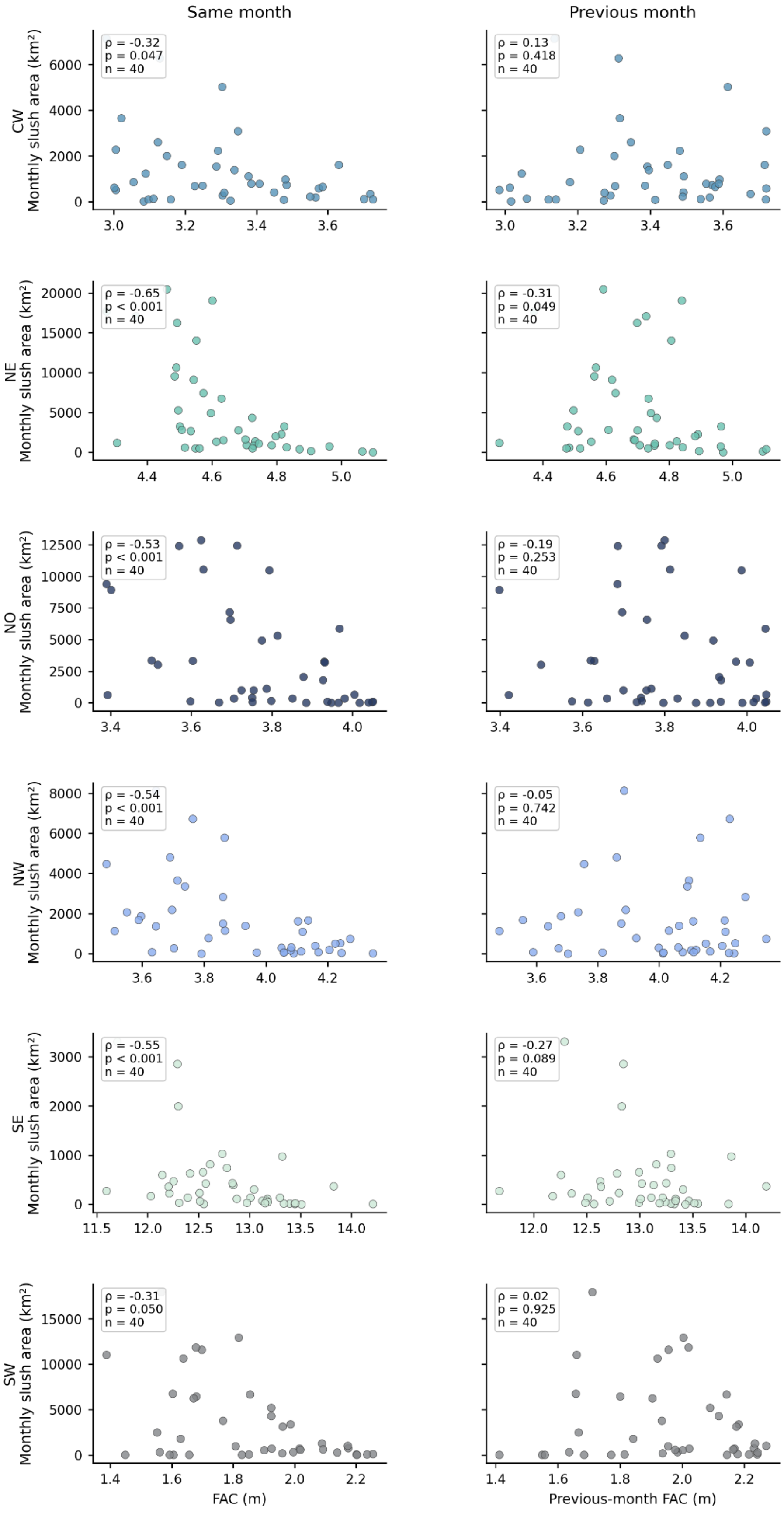

Figure S9. *Spearman rank relationships between monthly slush extent and same-month or previous-month FAC at the intra-annual basin-wise scale.*

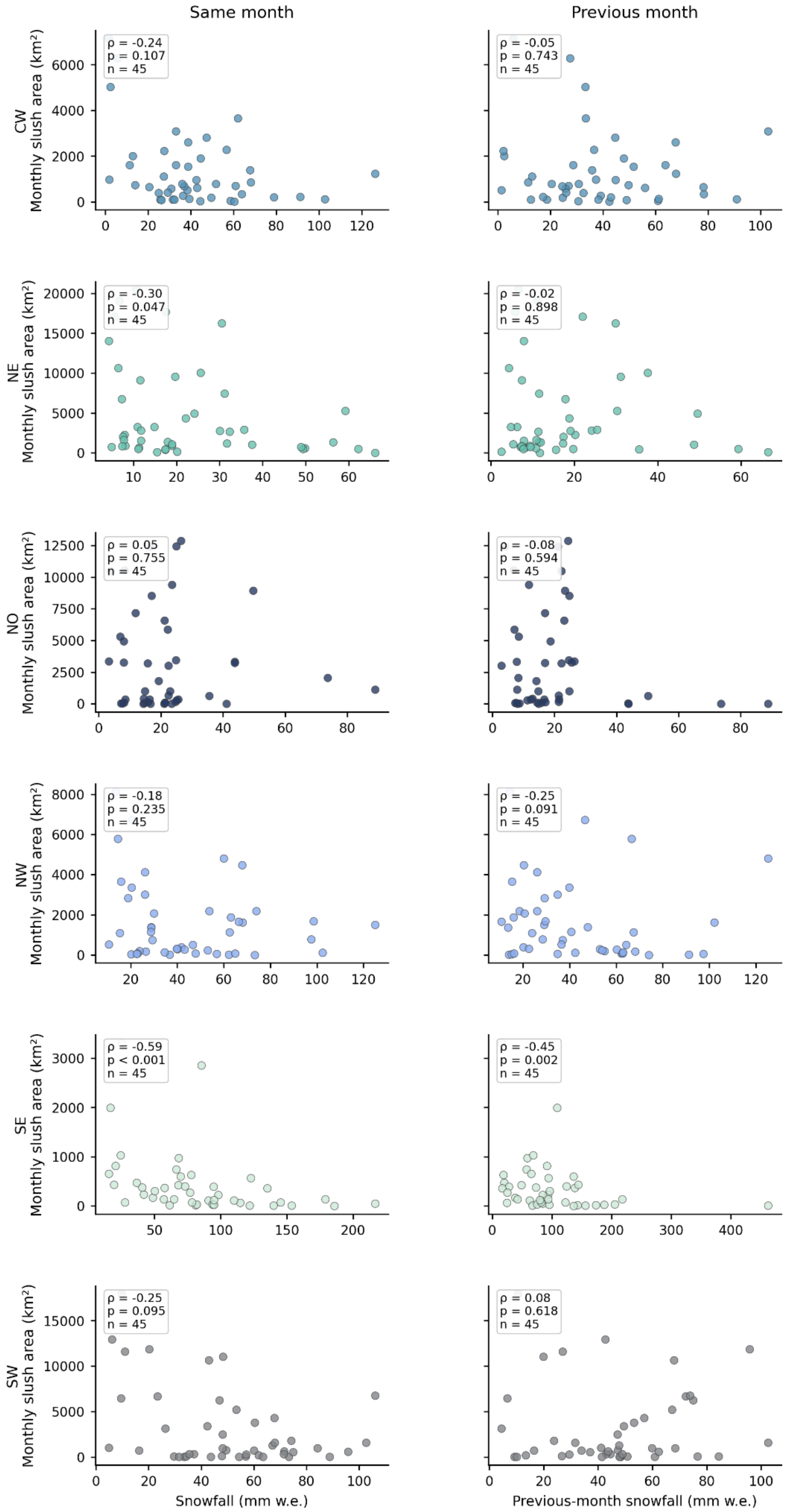

Figure S10. *Spearman rank relationships between monthly slush extent and same-month or previous-month snowfall at the intra-annual basin-wise scale.*

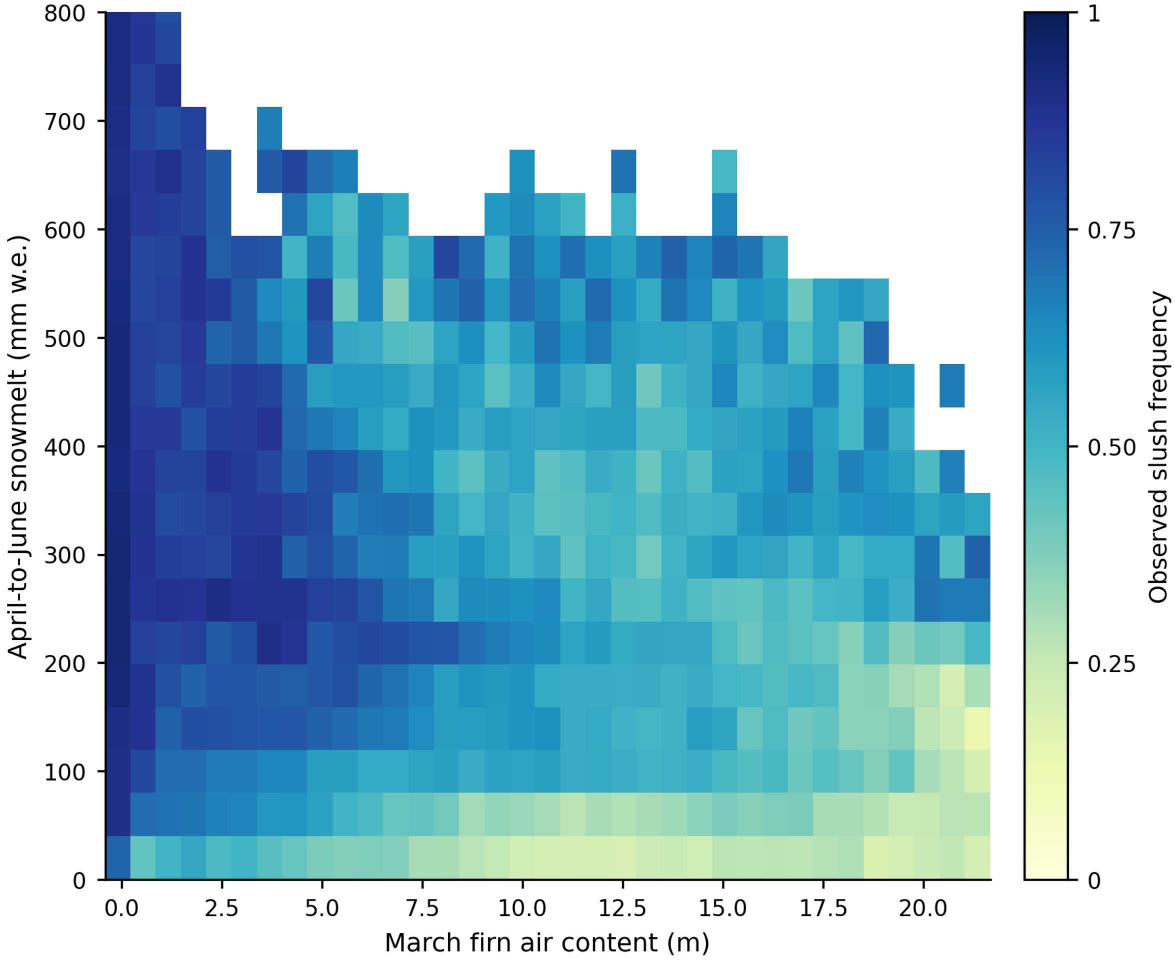

Figure S11. *Observed slush frequency for different combinations of March FAC and April–June snowmelt within the potential slush zone. Colours show how often slush was mapped under each combination, while white areas contain fewer than 100 observations.*

Supplementary Tables

Table S1. *List of selected S2 images used for training and validation across basins. Image IDs correspond to GEE asset identifiers from the COPERNICUS/S2 collection (Gorelick et al., 2017).*

Image Number	***Image Date***	***Sentinel 2 Image ID***
1	2019-08-01	COPERNICUS/S2/20190801T144759_20190801T144759_T22WFV
2	2019-07-13	COPERNICUS/S2/20190713T150921_20190713T150916_T26XNP
3	2023-07-03	COPERNICUS/S2/20230703T161831_20230703T161832_T21XWC
4	2020-07-26	COPERNICUS/S2/20200726T180919_20200726T180916_T20XMN
5	2021-08-05	COPERNICUS/S2/20210805T135729_20210805T135732_T25WFS
6	2023-07-02	COPERNICUS/S2/20230702T150801_20230702T150959_T22WEC
7	2018-06-04	COPERNICUS/S2/20180604T143919_20180604T143920_T26XNJ
8	2020-06-13	COPERNICUS/S2/20200613T161829_20200613T161831_T21XWD
9	2016-05-03	COPERNICUS/S2/20160503T180920_20160503T231352_T20XMP
10	2023-05-01	COPERNICUS/S2/20230501T142741_20230501T142742_T23WPN
11	2017-06-16	COPERNICUS/S2/20170616T151911_20170616T152036_T22WED
12	2021-06-21	COPERNICUS/S2/20210621T144749_20210621T144900_T22WES
13	2021-08-22	COPERNICUS/S2/20210822T143931_20210822T144327_T22WFS
14	2018-09-13	COPERNICUS/S2/20180913T145911_20180913T145910_T26XNM
15	2023-08-15	COPERNICUS/S2/20230815T171859_20230815T171900_T20XML
16	2022-08-16	COPERNICUS/S2/20220816T164851_20220816T164854_T25XDL
17	2021-08-23	COPERNICUS/S2/20210823T141011_20210823T141007_T24WWV
18	2017-08-15	COPERNICUS/S2/20170815T151911_20170815T152131_T22WFB
19	2016-05-16	COPERNICUS/S2/20160516T145922_20160516T150205_T22WEA
20	2019-07-21	COPERNICUS/S2/20190721T151809_20190721T151812_T23WMS
21	2019-08-02	COPERNICUS/S2/20190802T164901_20190802T164904_T26XMP
22	2022-07-04	COPERNICUS/S2/20220704T140739_20220704T140740_T25WDS
23	2022-08-25	COPERNICUS/S2/20220825T180919_20220825T181113_T19XEH
24	2023-07-16	COPERNICUS/S2/20230716T185919_20230716T185920_T21XWK

Table S2. *Spectral characteristics of clusters labelled as slush and neighbouring non-slush across the 18 Sentinel-2 training images. Values show the number of valid pixels and the pixel-weighted mean ± standard deviation of NDWI$_{ice}$, hue, and saturation for each class.*

Image	Region	Date	Class	Pixels	NDWI$_{ice}$	Hue	Saturation
1	SW	2019-08-01	Slush	11,464,089	0.161 ± 0.066	0.584 ± 0.008	0.273 ± 0.075
			Non-slush	9,232,650	0.230 ± 0.057	0.573 ± 0.007	0.370 ± 0.071
2	NE	2019-07-13	Slush	3,995,741	0.211 ± 0.073	0.580 ± 0.011	0.343 ± 0.090
			Non-slush	3,504,859	0.157 ± 0.033	0.587 ± 0.007	0.270 ± 0.043
3	NW	2023-07-03	Slush	181,643	0.207 ± 0.072	0.572 ± 0.010	0.337 ± 0.097
			Non-slush	231,633	0.213 ± 0.113	0.574 ± 0.010	0.338 ± 0.137
4	NO	2020-07-26	Slush	298,036	0.192 ± 0.090	0.582 ± 0.014	0.313 ± 0.113
			Non-slush	739,550	0.162 ± 0.075	0.587 ± 0.011	0.272 ± 0.091
5	SE	2021-08-05	Slush	73,023	0.192 ± 0.059	0.569 ± 0.014	0.318 ± 0.080
			Non-slush	186,556	0.446 ± 0.159	0.563 ± 0.013	0.600 ± 0.153
6	CW	2023-07-02	Slush	93,132	0.200 ± 0.083	0.571 ± 0.010	0.326 ± 0.109
			Non-slush	159,507	0.257 ± 0.107	0.568 ± 0.007	0.399 ± 0.127
7	NE	2018-06-04	Slush	0	-	-	-
			Non-slush	1,681	0.137 ± 0.019	0.603 ± 0.010	0.241 ± 0.028
8	NW	2020-06-13	Slush	401	0.133 ± 0.012	0.600 ± 0.012	0.234 ± 0.019
			Non-slush	6,371	0.183 ± 0.061	0.593 ± 0.014	0.305 ± 0.082

9	NO	2016-05-03	Slush	0	-	-	-
			Non-slush	974	0.157 ± 0.032	0.610 ± 0.011	0.271 ± 0.045
10	SE	2023-05-01	Slush	5,071	0.149 ± 0.016	0.614 ± 0.006	0.259 ± 0.024
			Non-slush	7,618	0.160 ± 0.033	0.611 ± 0.009	0.274 ± 0.048
11	CW	2017-06-16	Slush	28,679	0.270 ± 0.081	0.573 ± 0.013	0.419 ± 0.100
			Non-slush	93,756	0.269 ± 0.186	0.579 ± 0.012	0.395 ± 0.200
12	SW	2016-05-16	Slush	386,772	0.145 ± 0.023	0.579 ± 0.008	0.253 ± 0.034
			Non-slush	308,442	0.291 ± 0.121	0.569 ± 0.011	0.438 ± 0.137
13	SW	2019-08-22	Slush	2,430,812	0.177 ± 0.067	0.584 ± 0.012	0.296 ± 0.089
			Non-slush	1,299,390	0.289 ± 0.134	0.571 ± 0.008	0.434 ± 0.144
14	NE	2018-09-13	Slush	846,968	0.148 ± 0.051	0.608 ± 0.010	0.255 ± 0.069
			Non-slush	599,202	0.181 ± 0.062	0.603 ± 0.013	0.302 ± 0.079
15	NW	2023-08-15	Slush	76,510	0.180 ± 0.104	0.586 ± 0.010	0.294 ± 0.130
			Non-slush	200,774	0.178 ± 0.056	0.582 ± 0.010	0.299 ± 0.077
16	NO	2022-08-16	Slush	87,988	0.276 ± 0.074	0.588 ± 0.010	0.427 ± 0.096
			Non-slush	103,218	0.246 ± 0.076	0.589 ± 0.012	0.389 ± 0.096
17	SE	2021-08-23	Slush	26,082	0.269 ± 0.133	0.573 ± 0.013	0.408 ± 0.155
			Non-slush	11,359	0.194 ± 0.054	0.575 ± 0.011	0.322 ± 0.074
18	CW	2017-08-15	Slush	275,927	0.416 ± 0.208	0.585 ± 0.012	0.557 ± 0.204

			Non-slush	541,136	0.315 ± 0.157	0.583 ± 0.013	0.459 ± 0.175

Table S3. *Cluster assignments for training images used in the Random Forest classification. Each row represents a k-means cluster number (0–39) and each column represents one training image (see Table S1). Blue cells indicate clusters assigned to the slush class; grey cells indicate clusters assigned to the non-slush class.*

	Image no.																	
Cluster no.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
0																		
1																		
2																		
3																		
4																		
5																		
6																		
7																		
8																		
9																		
10																		
11																		
12																		
13																		
14																		
15																		
16																		
17																		
18																		
19																		
20																		
21																		
22																		

23																		
24																		
25																		
26																		
27																		
28																		
29																		
30																		
31																		
32																		
33																		
34																		
35																		
36																		
37																		
38																		
39																		

Table S4. *Hyperparameter settings for the RF classifier, detailing our chosen values for key parameters.*

Hyperparameter	Value	Description
Number of Trees	50	Controls the number of trees in the forest. More trees improve stability and accuracy but increase computational demand.
Bag Fraction	0.1	Fraction of the training data used for each tree. A lower value increases variance in the predictions made by individual trees, while a higher value reduces it but may introduce bias.
Minimum Leaf Population	20	Minimum number of samples required at a leaf node to prevent overfitting by ensuring sufficient data per leaf.
Maximum Nodes	500	Limits the number of nodes per tree, balancing complexity and simplicity.
Seed	30	Ensures reproducibility by initialising the random number generator used for tree building.
Variables per Split	10	Controls the number of randomly chosen variables that are considered at each split in a tree.

Table S5. *Agreement metrics (precision, recall, and F1-score) and spatial overlap areas for the comparison between RF and NDWI classifications across all six GrIS basins. RF was treated as the reference classification, and precision, recall, and F1-score were macro-averaged across the slush and non-slush classes.*

RF VS. NDWI										
Image	Region	Date	Overall Agreement	Kappa	Precision	Recall	F1-Score	RF Area (km²)	NDWI Area (km²)	Overlap Area (km²)
1	CW	Aug-18	0.96	0.52	0.79	0.73	0.76	86.04	67.82	41.33
2	SW	Aug-19	0.68	0.35	0.79	0.68	0.64	2019.85	758.16	733.17
3	NW	Aug-21	0.82	0.31	0.66	0.65	0.66	134.74	126.68	54.77
4	NO	Aug-22	0.77	0.29	0.66	0.64	0.65	353.22	299.01	142.12
5	SE	Aug-23	0.85	0.16	0.76	0.55	0.56	400.55	71.90	47.15
6	NE	Aug-24	0.9	0.25	0.71	0.59	0.62	430.4	176.25	88

Table S6. *Agreement metrics (precision, recall, and F1-score) and spatial overlap areas for the comparison between RF and G_{ind} classifications across all six GrIS basins. RF was treated as the reference classification, and precision, recall, and F1-score were macro-averaged across the slush and non-slush classes.*

RF VS. G_{ind}										
Image	Region	Date	Overall Agreement	Kappa	Precision	Recall	F1-Score	RF Area (km²)	G_{ind} Area (km²)	Overlap Area (km²)
1	CW	Aug-18	0.94	0.16	0.64	0.56	0.58	86.04	33.19	11.06
2	SW	Aug-19	0.61	0.21	0.75	0.6	0.54	2019.85	461.64	435.64
3	NW	Aug-21	0.87	0.38	0.78	0.65	0.68	134.74	66.21	44.5
4	NO	Aug-22	0.79	0.1	0.84	0.53	0.51	353.22	27.97	25.03
5	SE	Aug-23	0.85	0.1	0.77	0.53	0.52	400.55	39.26	27.14
6	NE	Aug-24	0.91	0.14	0.78	0.54	0.56	430.4	61.76	39.98