Supervised classification of slush and ponded water on Antarctic ice shelves using Landsat 8 imagery

Rebecca L. Dell[1,3*], Alison F. Banwell[2,1], Ian C. Willis[1], Neil S. Arnold[1], Anna Ruth W. Halberstadt[4], Thomas R. Chudley[5], Hamish D. Pritchard[3]

[1] Scott Polar Research Institute, Lensfield Road, Cambridge, CB2 1ER, United Kingdom.
[2] Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, 80309, United States.
[3] British Antarctic Survey, High Cross, Madingley Road, Cambridge, CB3 0ET, United Kingdom.
[4] Department of Geosciences, University of Massachusetts, Amherst, 01003, United States.
[5] Byrd Polar & Climate Research Centre, The Ohio State University, 1090 Carmack Rd, Columbus, 43210, United States.

Corresponding author: Rebecca Dell (rld46@cam.ac.uk)

ABSTRACT

Surface meltwater is becoming increasingly widespread on Antarctic ice shelves. It is stored within surface ponds and streams, or within firn pore spaces, which may saturate to form slush. Slush can reduce firn air content, increasing an ice-shelf's vulnerability to break-up. To date, no study has mapped the changing extent of slush across ice shelves. Here, we use Google Earth Engine and Landsat 8 images from six ice shelves to generate training classes using a k-means clustering algorithm, which are used to train a Random Forest Classifier to identify both slush and ponded water. Validation using expert elicitation gives accuracies of 84% and 82% for the ponded water and slush classes respectively. Errors result from subjectivity in identifying the ponded water/slush boundary, and from inclusion of cloud and shadows. We apply our classifier to the Roi Baudouin Ice Shelf for the entire 2013 to 2020 Landsat 8 record. On average, 64% of all surface meltwater is classified as slush and 36% as ponded water. Total meltwater areal extent is greatest between late January and mid-February. This highlights the importance of mapping slush when studying surface meltwater on ice shelves. Future work will apply the classifier across all Antarctic ice shelves.

1 INTRODUCTION

Surface meltwater is present on the majority of Antarctica's ice shelves (e.g., Langley and others, 2016; Kingslake and others, 2017; Macdonald and others, 2019; Stokes and others, 2019; Arthur and others, 2020a; Dell and others, 2020; Banwell and others, 2021). It can act as a key control on ice-shelf stability (Lai and others, 2020) and thus the contribution of Antarctica's grounded ice to global sea level rise (Rignot and others, 2004; Berthier and others, 2012; Furst and others, 2016). Surface meltwater is stored either in ponds within topographic depressions on top of impermeable ice surfaces (Bell and others, 2018; Banwell and others, 2019) or in firn pore spaces (Dunmire and others, 2020; Montgomery and others, 2020). When firn pore spaces become saturated, slush is formed and this may be particularly likely where firn overlies former blue ice areas or refrozen lakes, or where refreezing of infiltrated water has formed extensive ice layers at depth within the firn. Melting and refreezing of slush promotes firn air content depletion, thereby increasing its density

and increasing an ice shelf's vulnerability to fracture (Hubbard and others, 2014; Kuipers Munneke and others, 2014; Alley and others, 2018). Ponded water has been shown to drive ice-shelf collapse events through hydrofracture (Banwell and others, 2013; Banwell and MacAyeal, 2015; Robel and Banwell, 2019; Scambos and others, 2003, 2004) and therefore several studies have mapped the changing extent of ponded water on ice shelves (e.g. Arthur and others, 2020; Dell and others, 2020; Spergel and others, 2021). Despite the potential role of water as slush in driving hydrofracture, there has been very little work investigating the changing extent of slush on ice shelves. This means that previous work will not only have underestimated total surface meltwater areas on Antarctic ice shelves, but also underestimated their potential vulnerability to hydrofracture and collapse.

Across Antarctic ice shelves, areas of ponded water and slush are more commonly observed near to grounding lines (Kingslake and others, 2017; Lenaerts and others, 2017). Here, katabatic and/or foehn winds facilitate snow erosion, exposing widespread areas of blue ice and lowering the surface albedo, which in turn amplifies surface melting (Kingslake and others, 2017; Lenaerts and others, 2017). The extent of surface melting is expected to increase as air temperatures rise throughout the 21st Century (Trusel and others, 2015; IPCC, 2019), as demonstrated across the northern George VI Ice Shelf during the 2019/2020 melt season, when sustained periods of warm air temperatures above 0^{o}C led to 32-year record-high melting (Banwell and others, 2021). It is therefore crucial to quantify the area and volume of surface meltwater on the surface of ice shelves, and to evaluate the potential impacts of this surface meltwater, including slush, on ice-shelf stability. Furthermore, producing time series of surface meltwater across ice shelves will allow current surface mass balance models to be validated, potentially leading to improved projections of future meltwater evolution.

Remotely-sensed data can be used to track surface water bodies (i.e. ponds and streams) across space and over time. At present, two key methodologies are used to map water bodies on Antarctic ice shelves: threshold-based mapping (e.g. Banwell and others, 2014; Dell and others, 2020; Moussavi and others, 2020) and machine

learning (ML) (e.g. Dirscherl and others, 2020, 2021; Halberstadt and others, 2020). The former identifies water bodies where pixels exceed a reflectance threshold in specific bands or band combinations. Whilst most threshold-based approaches rely solely on an Normalized Difference Water Index of ice ($NDWI_{ice}$) threshold (e.g. Dell and others, 2020; Williamson and others, 2018), Moussavi and others (2020) employs a multiple threshold approach to map surface lakes more accurately on a pan-Antarctic scale, achieving accuracies of > 95% and > 97% for Landsat 8 and Sentinel-2 respectively.

Despite the significance of slush for firn-air depletion and as a possible precursor to the formation of surface water bodies, little is known about its spatial-temporal trends across Antarctic ice shelves on intra-seasonal and inter-annual timescales. Previous research on the Nansen Ice Shelf utilised a threshold-based approach on cloud-free imagery to identify areas of slush as those with an $NDWI_{ice}$ between 0.12 and 0.14 (Bell and others, 2017). This approach built upon the work of Yang and Smith (2013), who used $NDWI_{ice}$ thresholds to map surface streams on the southwestern Greenland Ice Sheet. Yang and Smith (2013) commented on the difficulties of using remote sensing to distinguish between water and slush on the ice sheet surface, as the high liquid water content of slush results in similar spectral reflectance values to water. However, Yang and Smith (2013) found that a low $NDWI_{ice}$ threshold of 0.12 identified all water pixels, and a moderate $NDWI_{ice}$ threshold of 0.14 helped to eliminate slush. Whilst this approach may perform well in particular locations, it cannot necessarily be applied across all Antarctic ice shelves given the spectral similarities of slush to surface water, blue ice and shaded snow (Moussavi and others, 2020). As such, thresholds that are suitable in one scene may not be suitable in other scenes, and variable thresholds would be needed if this approach were to be applied across many scenes.

ML offers an alternative to the threshold-based approach, and typically utilises more spectral information than single or multi-band methodologies as ML methods can automatically determine which spectral information is valuable for making classification decisions. Whilst ML is more computationally expensive, cloud-based

geoprocessing platforms such as Google Earth Engine (GEE) have made possible its application on a pan-Antarctic scale, without the need for local, high-performance computing clusters. Overall, ML has been shown to produce similar results to the threshold method when mapping surface water bodies on Antarctic ice shelves (Halberstadt and others, 2020). However, it has not been applied to the mapping of slush, and therefore the total area of all surface meltwater across Antarctic ice shelves remains underestimated.

This study, therefore, aims to use a machine learning (ML) methodology to develop a supervised classifier within GEE capable of detecting, and differentiating between slush and ponded surface water across all Antarctic ice shelves. To do this, we: (1) Train a supervised classifier capable of lake and slush identification on six different Antarctic ice shelves; (2) Validate the classifier by investigating the agreement with manual classification by a set of experts; and (3) apply the final classifier to the Roi Baudouin Ice Shelf (RBIS) for the period 2013 to 2020 to identify spatial patterns and temporal variability in slush and ponded surface water.

2 MATERIALS AND METHODS

Here we introduce the study areas used to train and validate the classifier. We also describe the steps taken to select and pre-process the Landsat 8 Level 1 images used by the classifier. We then describe the methods used to build the classifier, before explaining how we validate it. Finally, we discuss how we apply the validated classifier to the RBIS.

2.1 Study areas

We trained and validated our methods on six individual ice shelves (Figure 1); (i) Nivlisen, (ii) Roi Baudouin, (iii) Amery, (iv) Shackleton, (v) Nansen, and (vi) George VI (Figure 1, Table S.1). These ice shelves are characterised by a range of surface melt conditions and features, resulting in a wide variety of surface spectral characteristics. Additionally, all six ice shelves experience snow erosion driven by katabatic winds, which leads to the formation of extensive areas of blue ice at their grounding lines. The key information for each of these ice shelves is presented in Table 1.

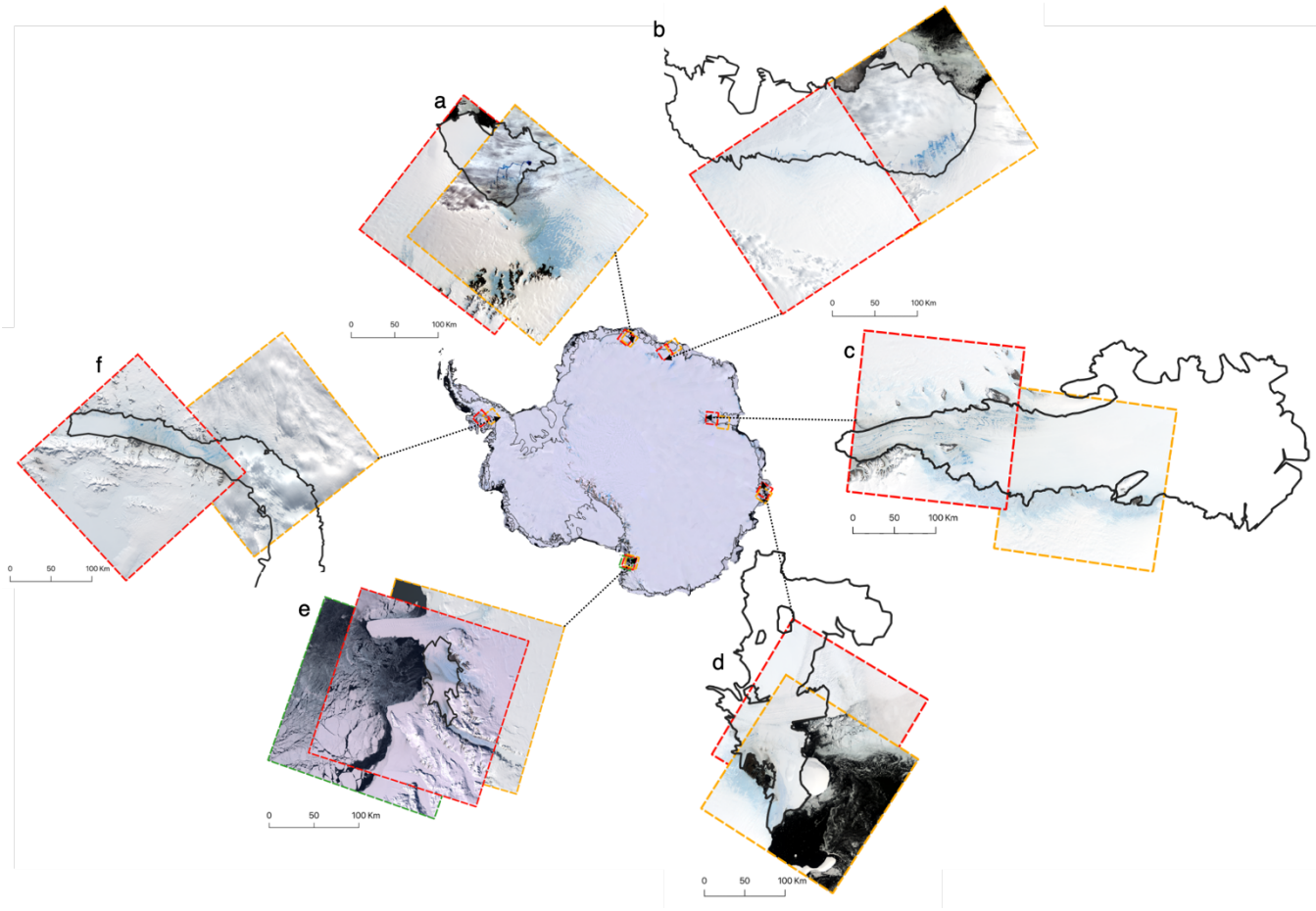

Figure 1: Study area figure showing the six ice shelves selected for use in the unsupervised k-means clustering algorithm. Dashed coloured boxes indicate the location of the surrounding Landsat-8 images for a) Nivlisen, b) Roi Baudouin, c) Amery, d) Shackleton, e) Nansen, f) George VI. The different coloured boxes indicate different path/rows used for each study site. Ice-shelf boundaries (from the SCAR Antarctic Digital Database (Gerrish and others, 2021) are marked by a solid black line on both the main and subset images. The central map of Antarctica is the Centre-Filled LIMA Mosaic (Bindschadler and others, 2008).

Table 1: Study area details for the six ice shelves used in the unsupervised k-means clustering algorithm.

Ice Shelf	Latitude	Longitude	Ice-sheet Region	Area (km^2)	General Surface Water Characteristics	Key Citations
Nivlisen	70.7° S	11.7° E	Dronning Maud Land, East Antarctica	7380	Elongate surface lakes expand towards ice shelf's calving front as melt season progresses	Dell and others (2020)
Roi Baudouin	69.9° S	32.6° E	Dronning Maud Land, East Antarctica	33,200	Extensive melt near the grounding line in addition to buried lakes	Dunmire and others (2020); Lenaerts and others (2017)
Amery	73.1°S	67.3° E	East Antarctica, Pyrdz Bay, East Antarctica	61,800	Surface drainage system comprised of surface channels and lakes	Fricker and others (2020); Spergel and others (2021)
Shackleton	66.4°S	100° E	Queen Mary Land, East Antarctica		Surface lakes near the grounding line	Arthur and others (2020b)
Nansen	74.9°S	162.8° E	Victoria Land, East Antarctica	2270	Large surface river exports surface meltwater into ocean via a 130 m wide waterfall	Bell and others (2017); Frezzotti (1993)
George VI	70.7°S	68.2° W	South-west Antarctic Peninsula	30,300	Extensive ponding in northern region since early 1940s	Banwell and others (2021); Reynolds (1981); Wagner (1972)

2.2 Scene selection and pre-processing

Identical criteria and methods were used to select and pre-process suitable Landsat 8 scenes across both the training and validation steps of this methodology (Figure 2). We first identified suitable image scenes for each study site by searching the Landsat 8 Level 1 image collection from 2013 to 2020, filtering for images with < 40% cloud cover and > 20° solar elevation (Halberstadt and others, 2020). Solar elevations > 20° only were used to reduce the impact of shadowing (Halberstadt and others, 2020). 14 training images (two for each ice shelf, and an extra two for Nansen; see section 2.3 for further explanation), and six separate validation images (one for each ice shelf)

were then selected for the purpose of training and validating the classifier respectively (Table S1). When choosing suitable training and validation images, we aimed to select a range of images that spanned the full austral melt season (1 November to 31 March) and were acquired at a range of solar elevations (20.9° to 36.6°) (Table S1). This approach ensured that we were training and validating the classifier using images with a wide range of spectral characteristics.

Scenes were pre-processed by converting to per-pixel top-of-atmosphere (TOA) values (Dell and others, 2020), and by clipping to the ice shelf-boundaries (from the SCAR Antarctic Digital Database (Gerrish and others, 2021). A rock mask was then applied to each scene, following the method of Moussavi and others (2020). This mask was then buffered by 1 km to ensure full removal of rock and rock shadow from each scene (Halberstadt and others, 2020). Clouds (including cirrus) and cloud shadows were identified and masked using the Landsat 8 Quality Assessment Bands, with a 4 km buffer applied to ensure full removal.

Finally, all pixels with an $NDWI_{ice} > 0.1$ were selected for further analysis. We note that in previous studies, to identify slush in addition to shallow and deep water, a threshold of 0.12 has been used (Bell and others, 2017; Yang and Smith, 2013). However, in our study, we lowered the $NDWI_{ice}$ threshold to 0.1 to include more potentially wet pixels, which were then categorised as 'slush', 'water' or 'other' by the classifier at a later stage. $NDWI_{ice}$ was calculated using Landsat 8 bands 2 (blue) and 4 (red):

$$NDWI_{ice} = (\text{Blue Band} - \text{Red Band}) / (\text{Blue Band} + \text{Red Band}). \quad [1]$$

2.3 Training data generation and supervised classification

To generate training data and to train a supervised classifier, we followed the general methodology of Halberstadt and others (2020), which we briefly summarise here. Training data were generated by applying an unsupervised k-means clustering algorithm (Arthur and Vassilvitskii, 2007) in GEE, which identifies clusters of spectrally-distinct pixels across a set of 14 scenes from bands 1-7 (Figure 3c). The k-means clustering algorithm, which is the only supervised classification algorithm

available in GEE, is widely-used by the community and is robust, and for these reasons chosen for this study. Initial training data were generated using two image scenes per ice shelf. Our initial trained classifier produced significant misclassification errors over 'dirty ice' (i.e. ice that contains debris and/or sediment) regions; the inclusion of two additional Nansen Ice Shelf training scenes added 'dirty ice' training data and improved classifier performance.

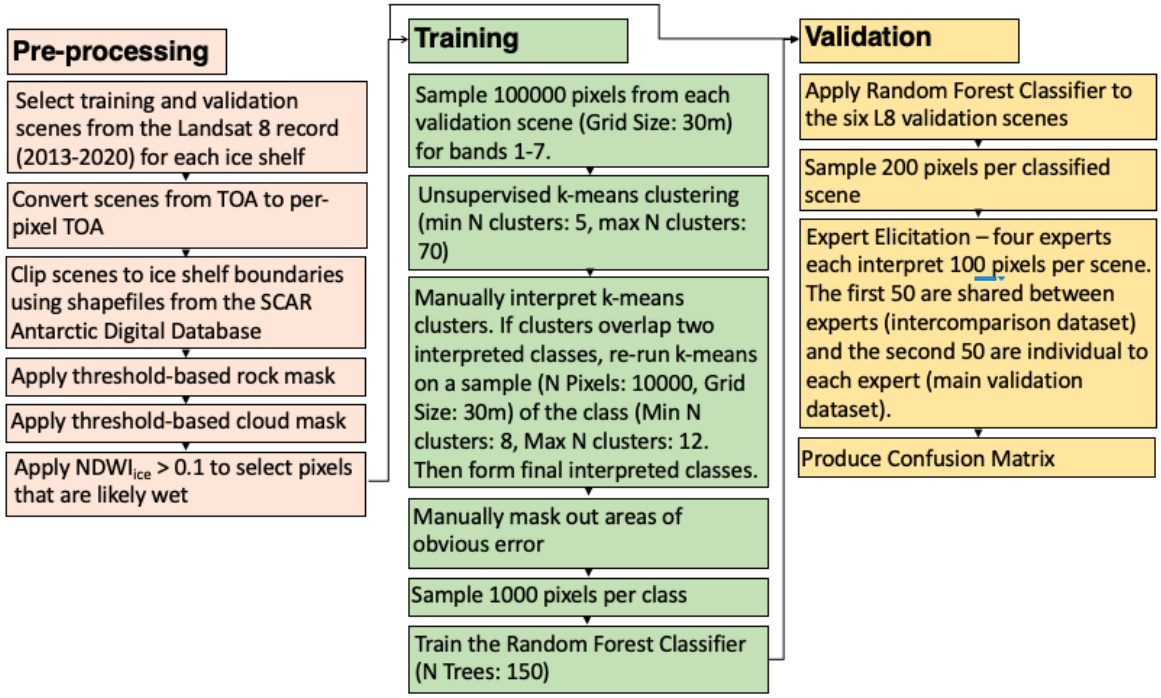

Figure 2: Workflow detailing the pre-processing, training, validation, and application steps for creating and using a supervised classifier to map slush and ponded water across Antarctic ice shelves using GEE

The k-means clustering algorithm was executed by sampling 100,000 pixels from each image at the Landsat 8 native grid size of 30 m. We specified a minimum of 5 and a maximum of 70 clusters when running the k-means clustering algorithm. This maximum value was manually determined, and increasing the value further did not have an impact on the output of the clusterer, as the cluster typically returned no more than ~ 20 clusters. We then manually interpreted the resulting clusters and grouped

them into interpreted classes: ponded water, slush, and other (including, but not limited to, blue ice, snow, and dirty ice). The boundary between slush and ponded water was determined by the developer of the classifier, however the transitional and subjective nature of this distinction should be noted, and this boundary is therefore imperfect. In some cases, clusters identified using the k-means algorithm overlapped two interpreted classes. These clusters were therefore further subdivided using k-means (sampling 10,000 pixels at a grid size of 30 m, and specifying a minimum of 8 and maximum of 12 clusters) and the sub-clusters were assigned to an interpreted class. Once the final interpreted classes were formed, areas of mis-classification error were manually masked from the training data. We then randomly sampled 1000 pixels from each interpreted class, to form the final training dataset for all ice shelves combined. These data were then used to train a Random Forest Classifier, implemented in Google Earth Engine. Random Forest Classifiers use numerous tree predictors to generate a most-likely outcome (Breiman, 2001). The number of trees for this classifier was set to 150. The relative importance of each band within the Random Forest Classifier was also determined within GEE.

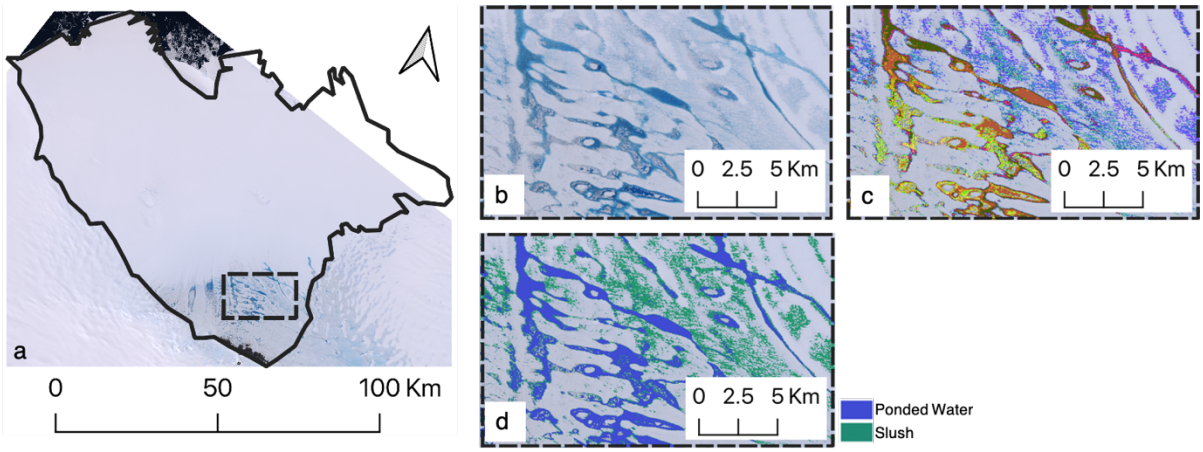

Figure 3: An example workflow for the k-means clustering algorithm over the Nivlisen Ice Shelf (Landsat 8, 2016-12-27). a) Base image of the Nivlisen Ice Shelf, the solid black line marks the ice-shelf area, the dashed box shows the zoomed area featured in b, c, and d. b) True colour composite, c) k-means clusters (shown as different colours), d) interpreted ponded water and slush classes, identified from the k-means

clusters in (c). In total, 10 k-means clusters were combined to form the ponded water class, and 10 k-means clusters were combined to form the slush class.

2.4 Validation

The performance of the supervised classifier was validated using the validation dataset, which included one image scene for each of the six study areas. For each of the six validation scenes, the Random Forest Classifier was applied (Figure 4), and 250 classified pixels were randomly sampled from each scene. We then used expert elicitation (Bamber and Aspinall, 2013), where four glaciologists, who we call 'experts', were each asked to manually interpret a total of 100 pixels for each image scene, classifying them as either 'ponded water', 'slush', or 'other'. Experts viewed each pixel within its surrounding spatial context, and were permitted to zoom in and out of the image. Furthermore, the experts were all familiar with looking at ice sheet/shelf surface hydrology using medium-resolution optical data, and were not directly involved with training the classifier. Experts were not given direction for the interpretation of pixels, to ensure that their interpretations were not biased by the individual who developed the classifier. Of the 100 pixels per image interpreted by each expert, the first 50 pixels for each of the six images were identical. These 300 pixels (the 'intercomparison dataset') were used to compare expert opinions to highlight the subjectivity of manually identifying slush and ponded water in satellite imagery. The second 50 pixels per image were unique to each expert, and comprised the 'main validation dataset' (i.e. 1200 pixels in total).

For each pixel, in addition to providing an interpretation, each expert assigned a confidence score to reflect the certainty of their manual interpretation. The confidence score values were assigned as either: (1) low-confidence, (2) medium-confidence, or (3) high-confidence (Bamber and Aspinall, 2013). These confidence scores provided a way to identify pixels that were likely classified with less accuracy by the experts, due to their uncertainty.

Finally, we present true positives and negatives, as well as false positives (errors of commission) and false negatives (errors of omission) as a confusion matrix (Stehman,

1997) to calculate the classifier accuracy (compared to the expert interpretations) for all pixels, as well as just for the high-confidence pixels. The overall classifier accuracy was calculated by summing all correctly classified pixels (true positive and true negatives) and dividing this sum by the total number of pixels sampled.

2.5 Application on the Roi Baudouin Ice Shelf

Once validated, the classifier was applied to the entire RBIS for Landsat 8 images from 2013 to 2020 to test how well the method upscaled through space and time. We filtered only for images with a solar elevation > 20°, but accepted any level of cloud cover in order to utilise as much of the available imagery as possible, thereby increasing data coverage through space and time. These selected images were then pre-processed using the same steps that were applied in the training and validation phases (see section 2.2). However, rather than processing individual scenes as we did previously, we created 15-day (bi-monthly) mosaiced products from the available scenes to maximise spatial coverage prior to applying the $NDWI_{ice} > 0.1$ filter. Each 15-day mosaiced product was produced using the 'quality mosaic' function in GEE, which used the pixel with the greatest $NDWI_{ice}$ value for locations where pixels overlapped. For each melt season, the products start on 1 November, and continue in blocks of exactly 15 days until 31 March (or until 1 April for leap years). The supervised classifier was applied to each 15-day product, and the total areas of both slush and ponded water were calculated. For 15-day periods that did not have complete data coverage across the RBIS, we scaled slush and ponded water areas to the full ice-shelf area by calculating the area of each 15-day product as a fraction of the full ice-shelf area, and then multiplying this fraction by the full ice-shelf area (Williamson and others, 2018; Banwell and others, 2021). In addition to 15-day products, for each melt season, we compiled maximum melt extent products (Williamson and others, 2018) to show each pixel that was covered by either slush, ponded meltwater, or both slush and ponded meltwater.

3 RESULTS

3.1 Classification accuracy based on expert elicitation

316 Table 2a shows the results from the intercomparison dataset for each scene in the
317 validation dataset, which were interpreted by all four experts. The data shown include
318 all interpreted pixels regardless of the associated confidence scores. Overall, the
319 accuracy of the ponded water class is 78%, and the accuracy of the slush class is
320 71%. For the ponded water class, the experts all produced similar accuracy scores for
321 the RBIS (8% spread), and more dissimilar scores for the Nansen Ice Shelf (30%
322 spread), with an mean spread across all 6 ice shelves of just 6%. For the slush class,
323 the experts are in closest agreement over the George VI Ice Shelf (11% spread), and
324 in least close agreement over the Nansen Ice Shelf (79% spread). As with the water
325 class, these discrepancies tend to cancel out between experts giving an overall mean
326 spread across all ice shelves of just 5%. Table 2b shows the same data as Table 2a,
327 but only for the pixels for which the experts had 'high-confidence' in their
328 interpretation.

Table 2a: Accuracy scores for the intercomparison dataset (the 50 pixels shared by all experts for each ice-shelf validation image), listed by expert.

	Ponded Water							Slush						
	Roi B	**Nansen**	**Nivlisen**	**Shackleton**	**GVI**	**Amery**	**Mean**	**Roi B**	**Nansen**	**Nivlisen**	**Shackleton**	**GVI**	**Amery**	**Mean**
Expert 1	88%	91%	71%	84%	65%	64%	77%	70%	85%	71%	65%	67%	63%	70%
Expert 2	90%	80%	88%	88%	80%	69%	82%	76%	52%	92%	73%	76%	69%	73%
Expert 3	96%	94%	80%	78%	65%	50%	77%	88%	94%	83%	60%	65%	37%	71%
Expert 4	92%	64%	85%	88%	68%	58%	76%	68%	15%	82%	79%	70%	93%	68%
Mean	91%	82%	81%	84%	70%	60%	78%	76%	61%	82%	69%	69%	65%	71%

Table 2b: High-confidence accuracy scores for the intercomparison dataset (the 50 pixels shared by all experts for each ice-shelf validation image), listed by expert.

	Ponded Water							Slush						
	Roi B	**Nansen**	**Nivlisen**	**Shackleton**	**GVI**	**Amery**	**Mean**	**Roi B**	**Nansen**	**Nivlisen**	**Shackleton**	**GVI**	**Amery**	**Mean**
Expert 1	83%	80%	83%	88%	89%	67%	82%	59%	86%	83%	64%	89%	67%	75%
Expert 2	94%	50%	94%	89%	88%	75%	82%	76%	33%	100%	89%	88%	67%	76%
Expert 3	100%	92%	83%	88%	91%	40%	82%	87%	92%	89%	70%	100%	40%	80%
Expert 4	100%	100%	100%	100%	77%	62%	90%	50%	25%	100%	100%	91%	57%	71%
Mean	94%	81%	90%	91%	86%	61%	84%	68%	59%	93%	81%	92%	58%	75%

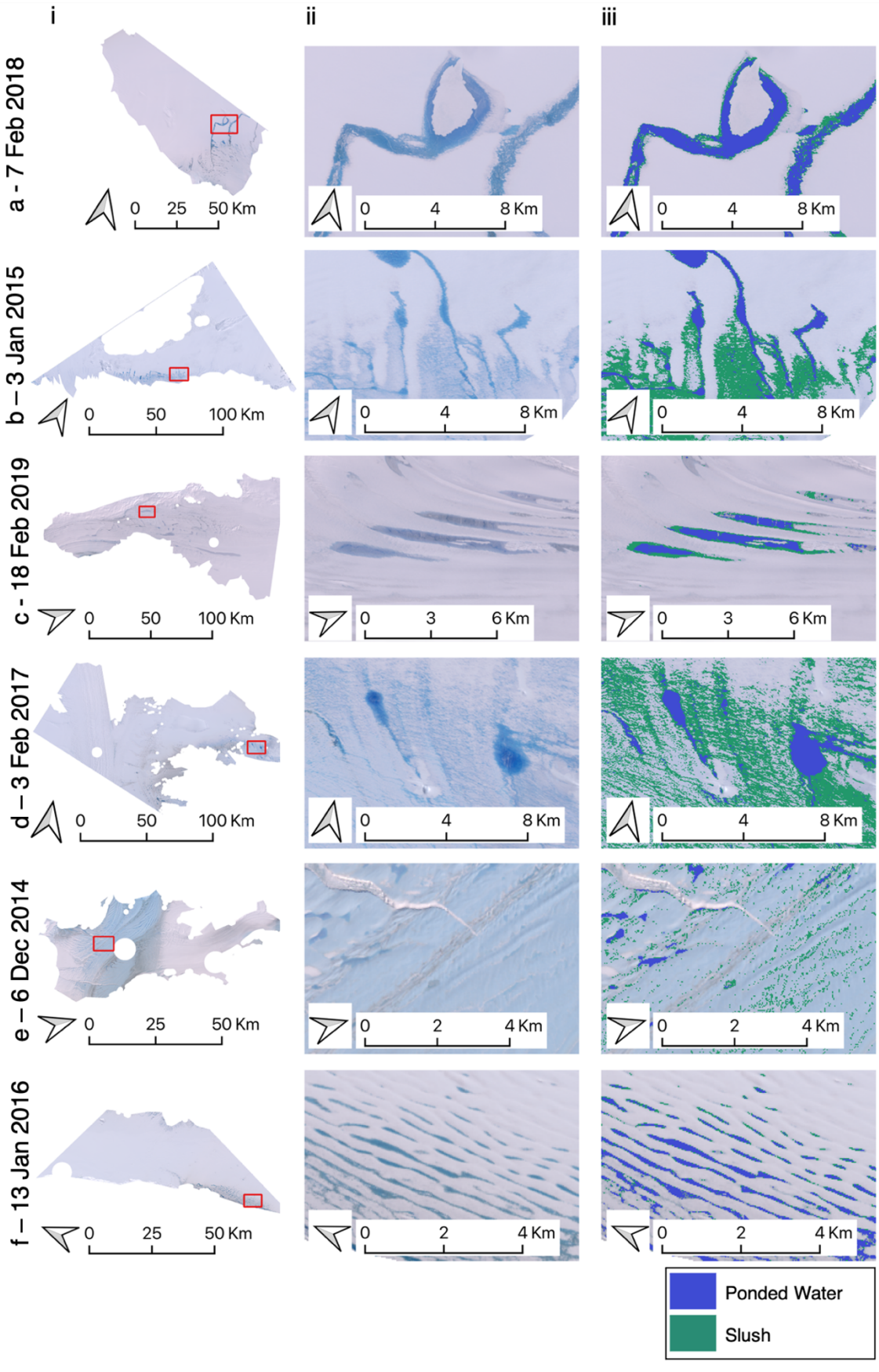
i
ii
iii
a - 7 Feb 2018
b – 3 Jan 2015
c - 18 Feb 2019
d – 3 Feb 2017
e – 6 Dec 2014
f – 13 Jan 2016
0 25 50 Km
0 50 100 Km
0 4 8 Km
0 3 6 Km
0 2 4 Km
Ponded Water
Slush

Figure 4: Preliminary outputs from the supervised classifier, as applied to six Landsat 8 validation images for a) Nivlisen Ice Shelf, b) Roi Baudouin Ice Shelf, c) Amery Ice Shelf, d) Shackleton Ice Shelf, e) Nansen Ice Shelf, f) George VI Ice Shelf. Panels in column i) show the pre-processed Landsat 8 RGB images to be classified, with the red boxes delineating close-up areas shown in panels in columns ii) and iii). Panels in column ii) show the close up areas in RGB, and panels in column iii) show the results for these areas produced by the supervised classifier, with blue = ponded water and green = slush.

Table 3a shows the accuracy results for the classifier over the main validation dataset (where each expert interpreted 50 different pixels per ice-shelf). The accuracy for the ponded water class is 78% and for the slush class is 70%; these values are very similar to those produced by the intercomparison data set. The classifier is most accurate at identifying ponded water for the Shackleton Ice Shelf (91%) and least accurate for the Amery Ice Shelf (61%). In contrast, the classifier is most accurate at identifying slush for the Nivlisen Ice Shelf (80%) and least accurate for the Nansen Ice Shelf (60%). The percentage of low confidence pixels ranges from 13% (Nivlisen and George VI ice shelves) to 28% (Shackleton Ice Shelf).

Table 3b shows the accuracy results for the main validation dataset using high-confidence pixels only. The mean accuracy for the lake class is 84% and for the slush class is 82%. Agreement between the experts and the classifier is greatest for ponded water over the Shackleton Ice Shelf (96%) and for slush over the Nivlisen Ice Shelf (92%). This agreement is lowest for ponded water over the Amery Ice Shelf (65%) and for slush over the RBIS (72%).

For the ponded water class, Expert 2 had the lowest agreement with the classifier. This was due to the classifier designating certain pixels as ‘other’ (e.g. non-wet surface facies), whilst the expert interpreted the pixels to be ponded water. For the slush class, Expert 4 had the lowest agreement with the classifier, which classified certain pixels as ‘other’ that were interpreted to be slush by the expert.

3.2 Relative importance of input bands

The relative importance of each band within our supervised classifier was determined within GEE using the '.explain()' function', and the results show that all bands contribute towards the classification of slush and ponded water (Table 4). However, band 5 (near-infrared) is of greatest importance for the supervised classifier, with an importance score of 20% (Table 4). Bands 1-4 (visible) and 6-7 (short-wave infrared 1 and 2) all have similar weightings, with importance scores ranging between 12% and 15%.

Table 3a: Accuracy scores for the main validation dataset (the 250 individual pixels (50 per expert) for each ice-shelf validation image) for the ponded water and slush classes separately. The percentage of pixel confidence scores for each ice shelf are also given.

	Ponded Water Accuracy	Slush Accuracy	Low Confidence Pixels	Medium Confidence Pixels	High Confidence Pixels
Nivlisen	80%	80%	13%	48%	40%
Roi Baudouin	87%	65%	19%	32%	50%
Amery	61%	64%	15%	59%	27%
Shackleton	91%	75%	28%	46%	26%
Nansen	81%	60%	22%	47%	31%
George VI	70%	74%	13%	52%	36%
Mean	*78%*	*70%*	*18%*	*47%*	*35%*

Table 3b: High-confidence accuracy scores for the main validation dataset (the 250 individual pixels (50 per expert) for each ice-shelf validation image) for the ponded water and slush classes separately.

	Ponded Water	Slush
Nivlisen	92%	92%
Roi Baudouin	86%	72%
Amery	65%	73%
Shackleton	96%	88%
Nansen	80%	74%
George VI	86%	91%
Mean	*84%*	*82%*

Table 4: Relative importance of each of the Landsat 8 bands used by the supervised classifier.

B1	B2	B3	B4	B5	B6	B7
14%	13%	14%	15%	20%	12%	12%

3.3 Application to the Roi Baudouin Ice Shelf

After applying the supervised classifier to the Roi Baudouin Ice Shelf, two key datasets are produced; a raw (unscaled) dataset, and a scaled dataset. The scaled dataset is produced to provide a better estimate of the total ice-shelf surface water area, as for many dates in this study, there is incomplete area-of-interest coverage (Figure 5). Of the 48 15-day periods presented In Figure 5, 14 have a percentage AOI coverage below 50 %. For the remainder of this paper, the scaled values only will be presented, however readers should remain aware of the potential for error when scaling up values across a full ice-shelf, because, for example, unscaled data with incomplete AOI coverage could already represent 100% of the total surface meltwater on the ice shelf surface. Unscaled data are presented in Figure S.1.

The maximum areas of slush and ponded water are reached between 15 January - 29 January 2016 (3.5×10^9 m^2) and 30 January - 13 February 2017 (1.9×10^9 m^2) respectively (Figure 5). In contrast, the lowest summer maximum areas of slush and ponded water occur between 15 January - 29 January 2019 (slush), and 14 February - 28 February 2019 (ponded water), reaching values of 5.7×10^8 m^2 and 2.9×10^8 m^2 respectively. For all seven melt seasons, the total area of slush and ponded water is greatest in either January or February. Furthermore, for all melt seasons except 2018/2019, the greatest areas of slush and ponded water are observed in the same 15-day periods within each melt season. However, for the austral summer of 2018/2019, the greatest total area of slush is recorded approximately a month prior to the greatest total area of ponded water (Figure 5).

Overall, the absolute difference between the greatest areas of slush for each melt season is larger than the absolute difference between the greatest areas of ponded

meltwater for each melt season, whilst the percentage change in ponded water is slightly greater than the percentage change in slush. Slush ranges from 5.7 x 10^8 m^2 between 15 January and 29 January 2019, to 3.5 x 10^9 m^2 between 15 January and 29 January 2016 (a 521% change in area), whereas ponded water varies from 2.9 x 10^8 m^2 between 14 February 2019 and 28 February 2019, to 1.9 x 10^9 m^2 between 30 January 2017 and 13 February 2017 (a 559% change in area) (Table S.2). Overall, slush dominates the total melt area across the RBIS, making up over half of the total melt on 39 of the 48 15-day periods investigated, and on average accounts for 64% of the total meltwater area (Table S.2). From the 2014/2015 melt season onwards, the percentage slush on the RBIS is greatest between 16 November and 30 December, when it accounts for between 84% and 96% of the total meltwater area.

Of the seven melt seasons investigated, the 2016/2017 melt season has the greatest recorded total meltwater area, reaching 5 x 10^9 m^2 between 30 January and 13 February 2017. Of this total area, 62% is slush, and 38% is ponded water (Table S.2). Conversely, the melt season that had the lowest total meltwater area is 2019/2020, with 7.5 x 10^8 m^2 between 15 January and 29 January 2019. Of that total area, 76% is slush and 24% is ponded water (Table S.2).

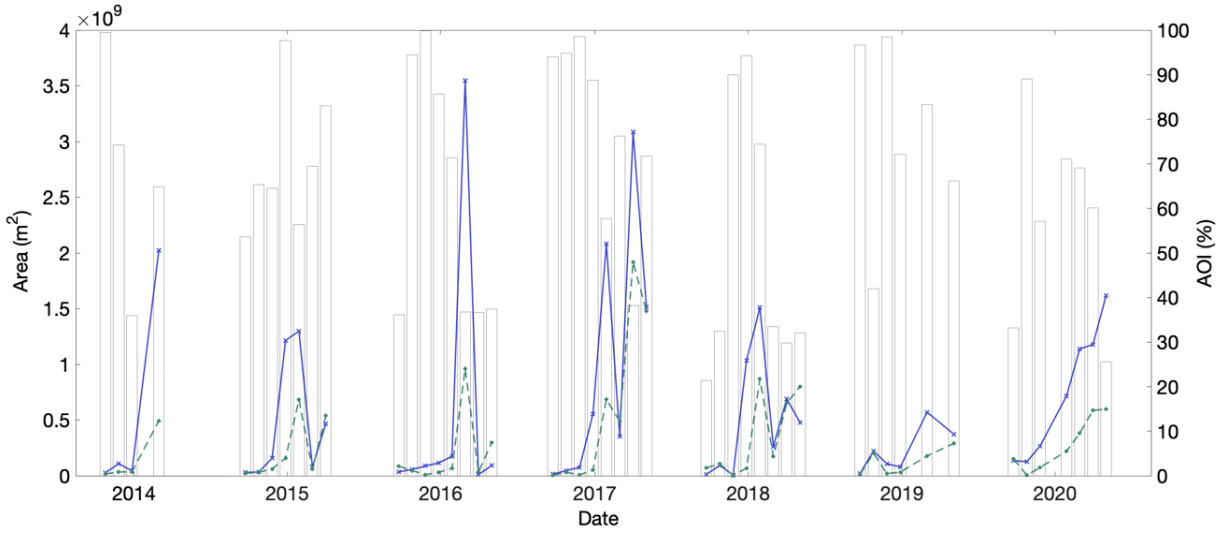

Figure 5: Time series data for slush and ponded water across the Roi Baudouin Ice Shelf. Grey bars show the % AOI coverage for each 15-day period plotted. Lines show scaled areas of slush (blue line) and ponded water (green line) on the RBIS from

2013/2014 to 2019/2020, derived from supervised classification of 15-day Landsat 8 mosaic products created in GEE (see section 2.5). Data are only plotted where ≥ 20% coverage of the RBIS is met. X axis date labels indicate 1 January for each year.

Figure 6 shows each of the 15-day data products that were produced within GEE for the 2016/2017 melt season over the RBIS. In these 15-day products, we manually inspected each image and ignored errors of commission (false positives) across the central and distal regions of the ice shelf. Therefore, the following results focus on the true positive results for the 2016/2017 season, which show meltwater in proximity to the ice shelf's grounding line. Little meltwater is detected between 1 November and 15 December 2016. However, from 16 December - 30 December 2016 onwards, areas of slush begin to develop near the grounding line in both the south-east and central southern parts. By early January (31 December 2016 - 14 January 2017) ponded water also begins to form amongst the areas of slush, and the areas of both classes increase until 30 January - 13 February 2017, after which the areas of both classes begin to decrease (Figures 5 and 6). A number of the 15-day products for this melt season have data gaps resulting from cloud masking, or a lack of image scenes covering the area of interest. The percentage ice-shelf area coverage by imagery for the 2016/2017 melt season ranges from 38% (30 January - 13 February 2017) to 99% (1 December - 15 December 2016) (Table S.2).

Data were combined to produce maximum melt extent across the RBIS for each melt season (1 November - 31 March) from 2013/2014 to 2019/2020 (Figure 7). In every melt season, both slush and ponded water are present predominantly in the south-east of the ice shelf, towards the grounding line. This area of slush and ponded water is the most spatially extensive in 2016/2017 and 2017/2018 (Figure 7d, e), when it extends approximately 47 km from the grounding line towards the ice-shelf front. In this region, slush is more spatially extensive than ponded water. Ponded water is typically observed towards the northern edge of the melt zone (i.e., closer to the ice front) each year, and is often surrounded by slush (Figure 7). Between 2013 and 2020, we find that 26% of all pixels that are covered by surface water are covered by both slush and ponded water at least once.

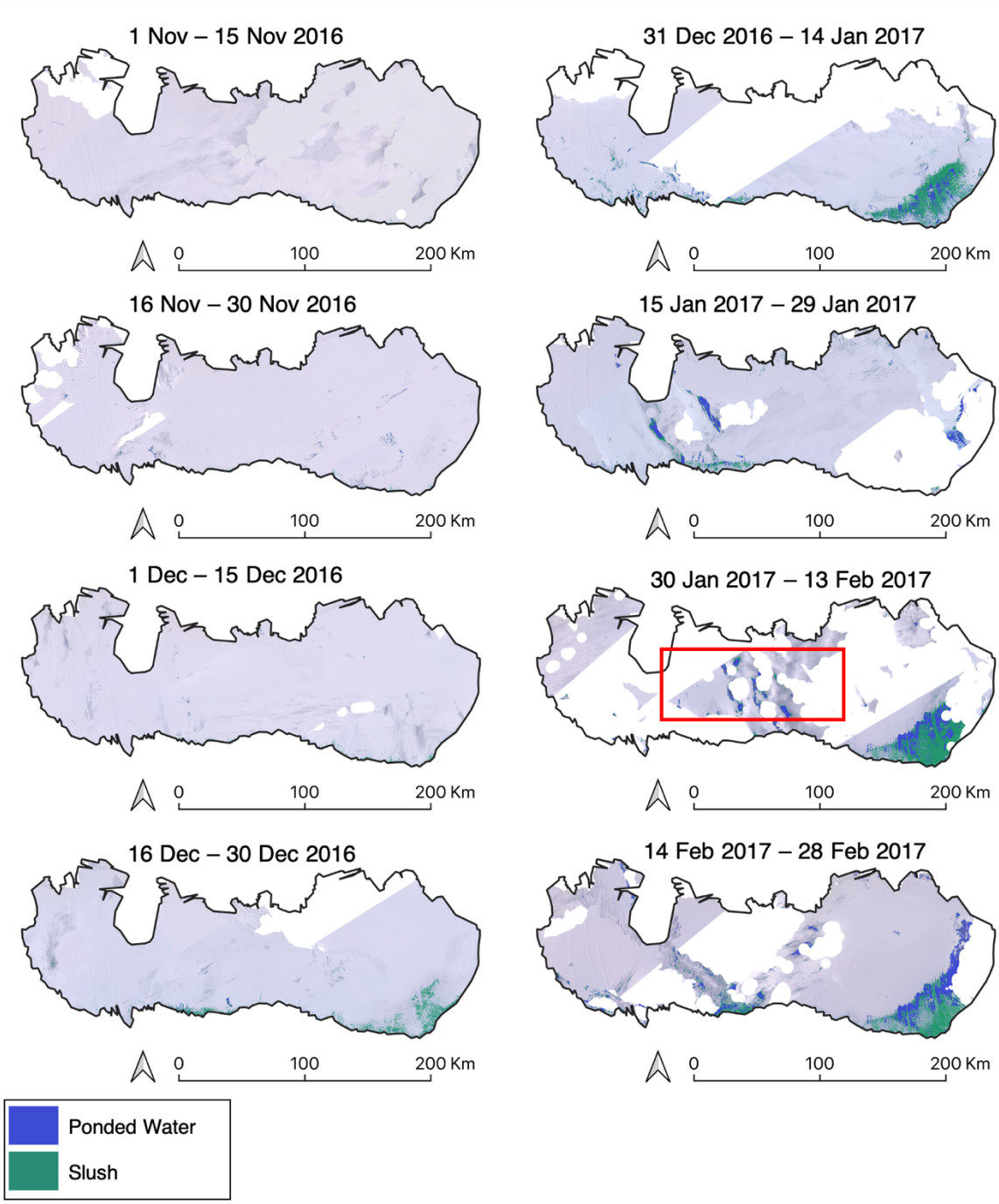

Figure 6: 15-day melt products for the 2016/2017 melt season across the Roi Baudouin Ice Shelf. White areas are areas that have either been masked out or were not covered by imagery in the first instance. The red box in the 30 Jan 2017 – 13 Feb 2017 panel roughly denotes the area where errors of commission due to cloud and cloud shadows are generally found.

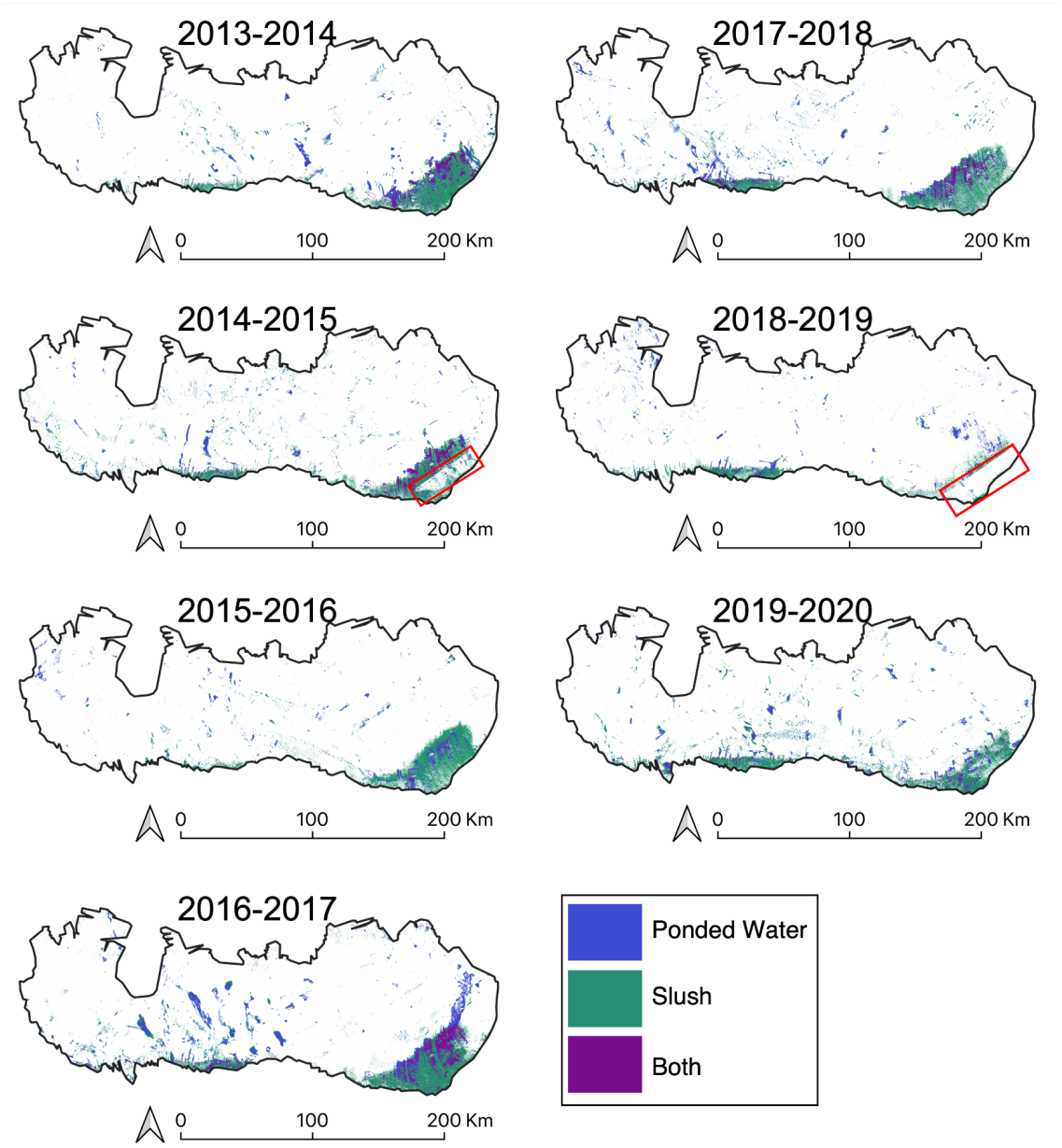

Figure 7: Maximum melt extent plots for each melt season, calculated by mosaicing all 15-day melt products for each melt season. Maximum areas of slush, ponded water, and both (where both slush and ponded water are identified within the melt season) are mapped. Red boxes roughly delineate areas affected by data gaps in the 2014/2015 and 2018/2019 melt seasons.

4 DISCUSSION

4.1 Classifier Accuracy

The mean accuracies across all ice shelves of the ponded water and slush classes were 84% and 82% respectively when comparing the classifier's outputs to high-confidence expert interpretations (which comprised 35% of all pixels within the main validation dataset) (Table 3b). Over all ice shelves, the percentage of pixels that were classified with high confidence did not exceed 50% (Table 3a), highlighting that even 'experts' are unable to classify all pixels with total confidence. Thus, although we use expert opinion to assess the accuracy of our classifier, each expert may be no more accurate than the classifier output itself. A solution to this would be to use ground based multi- or hyper-spectral data from ice shelves as ground truth data. However, to the authors' knowledge, no such data currently exist.

By collecting four expert interpretations, we aimed to minimise the effects of bias that each expert may have, and to get a more holistic set of expert interpretations for each ice shelf. The need for this approach was indicated by the spread between high-confidence pixels classified by experts for each ice shelf in the intercomparison dataset (Table 2b). For example, on the Nansen Ice shelf, agreement between the experts and the classifier ranged from 50% to 100% for ponded water, and from 25% to 86% for slush. Whilst the accuracy assessment attempts to best mimic ground-truthing through the use of multiple experts, it should be noted that the classifier is trained predominantly by a single person (separate to the experts used to validate the classifier), and so the classifier may reflect the biases of that individual. In addition, whilst experts are able to interpret a pixel within its surrounding spatial context, including both the immediate surrounding pixels as well as those elsewhere on the ice shelf, the classifier assesses the spectral characteristics of the pixel alone. This difference could be overcome by using object-based image analysis, however Halberstadt and others (2020) found such methods had a lower overall accuracy in comparison to pixel-based methods for the classification of ponded water. In the future, work should look to collect ground-based multi- or hyper-spectral data across ice shelves, which would facilitate a more robust assessment of this classifier's accuracy.

As previously mentioned, the main validation dataset for high-confidence pixels returned accuracy scores of 84% for ponded water and 82% for slush. Similar work for supervised classification of surface lakes only (i.e. not including slush) on Antarctic ice shelves achieved a mean pixel-based accuracy score of 93% (Halberstadt and others, 2020). Our slightly lower scores likely reflect the incorporation of slush into the classifier, in addition to the fact that we used a wider range of training sites. Furthermore, our validation techniques were different, as we validated the classifier against multiple expert opinions, as opposed to just one expert in Halberstadt and others (2020).

In our study, agreement between the classifier and the expert interpretations for high-confidence pixels was greatest for ponded water over Shackleton (96%) and for slush over Nivlisen (92%). However, the classifier accuracy was lowest over Amery, achieving 65% accuracy for ponded water and 73% for slush. The majority of the classification errors on the Amery Ice Shelf in particular appear to have resulted from topographic shadows being incorrectly classified as either slush or ponded water (Figure 4). Additionally on the validation image for the Amery Ice Shelf, there were examples of ponded water covered by a thin ice layer (Figure 4). The classifier tended to classify these areas as slush, as the thin ice layer adjusted the spectral properties of each pixel, whereas the experts differed in their interpretations and often interpreted them as ponded water or other.

Another source of classifier error was subjectivity when defining the slush/ponded-water boundary. Whilst the classifier utilised training data to determine the slush/ponded-water boundary, comparing classifier results with expert interpretations revealed some disagreement. However, we note that this disagreement is likely no greater than disagreement between the experts themselves, resulting from individual subjectivity, as neither the experts or the classifier were consistently more or less conservative when marking the slush/ponded-water boundary. Again, considering future work, without ground based multi- or hyper-spectral data it would be difficult to further improve such estimations of the slush/ponded-water boundary.

A final source of classifier-error was errors of commission resulting from cloud and cloud shadows and this is discussed separately in section 4.4.

4.2 Comparison to $NDWI_{ice}$

Whilst threshold-based methods have been used for the identification of deep surface meltwater bodies (e.g. surface lakes and streams) on Antarctic ice shelves (e.g. Banwell and others, 2014; Bell and others, 2017; Kingslake and others, 2017; Stokes and others, 2019; Dell and others, 2020; Moussavi and others, 2020), no prior studies have also attempted to map slush across an entire ice shelf for multiple melt seasons. Upscaling slush identification through space and time using simple threshold-based mapping approaches would lead to significant errors of omission and commission, owing to the spectral similarities between slush and other surface facies (e.g. lakes, blue ice, dirty ice) (Figure 8). For example, we found that applying $NDWI_{ice}$ thresholds of > 0.12 and ≤ 0.14 for slush and > 0.14 for ponded water (following Yang and Smith, 2013 and Bell and others, 2017) over the Shackleton Ice Shelf led to large errors of omission for slush when compared to the classifier output, due to confusion between water and slush (Figure 8). In contrast, applying these $NDWI_{ice}$ thresholds over the Nansen Ice Shelf led to errors of commission for slush, due to confusion between blue ice and slush (Figure 8). On the George VI Ice Shelf, however, the differences between the threshold method and the classifier output were similar, although even here the threshold method tended to underestimate slush area compared to the classifier (Figure 8).

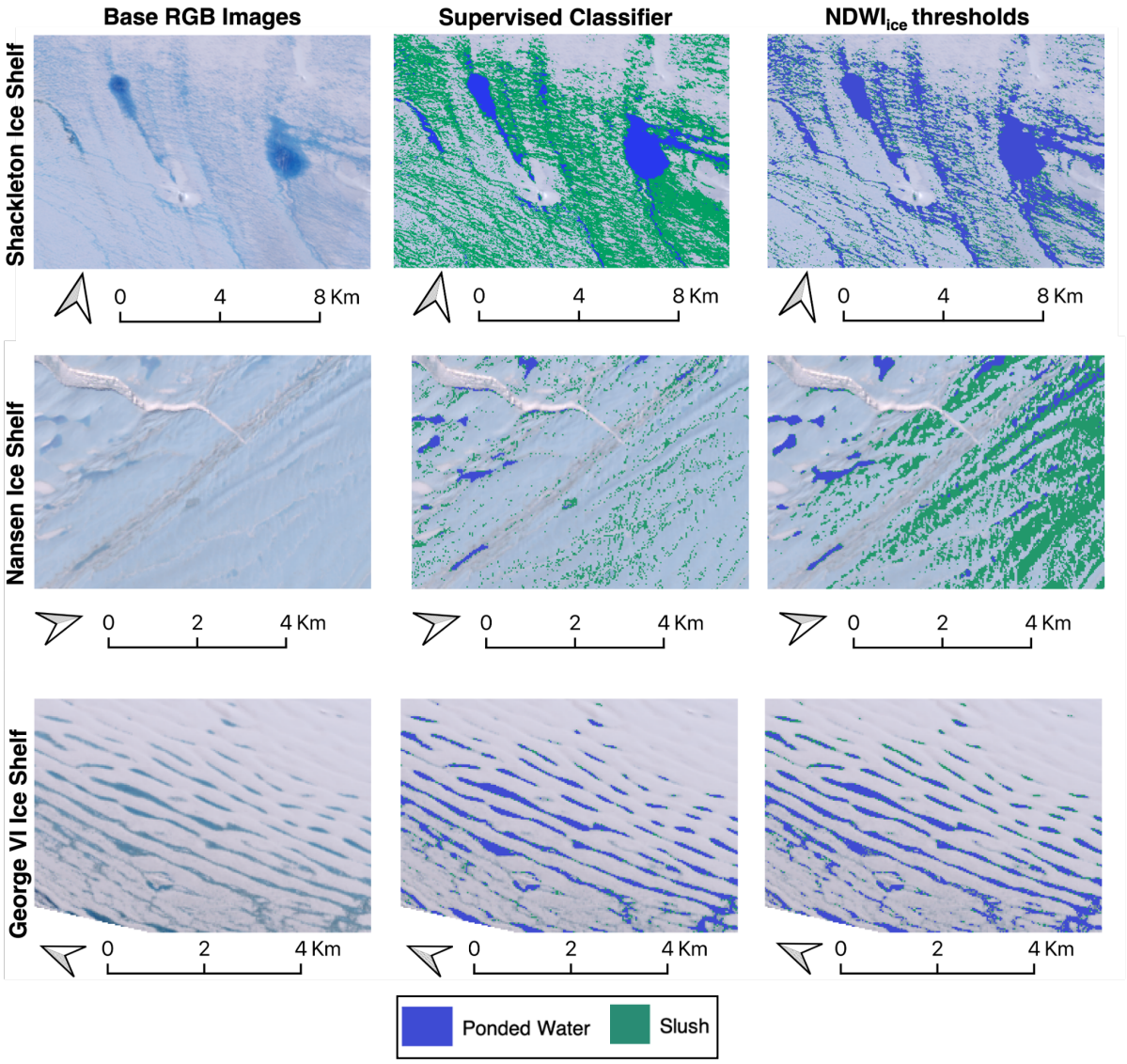

Figure 8: Outputs from the supervised classifier and from $NDWI_{ice}$ thresholding applied to sections of the validation images (as shown in Figure 4) for Shackleton, Nansen, and George VI ice shelves. Panels show the base RGB images, the area classified using the supervised classifier developed in this study, and the area classified using $NDWI_{ice}$ thresholds, where slush is < 0.12 and ≤ 0.14 and ponded water is > 0.14.

The limitations of the $NDWI_{ice}$ method that we have described above were overcome through our supervised classifier, as it was trained using seven Landsat 8 bands (bands 1-7) as opposed to just two (bands 2 and 4) for $NDWI_{ice}$, and it was therefore better able to distinguish between surface classes using a broader range of spectral

information. For our classifier, the near infrared band (band 5) was found to be the most important when distinguishing between classes (Table 4). This is likely related to the low reflectivity of water in near-infrared wavelengths (Work and Gilmer, 1976; Yang and others, 2011). Overall, whilst simple threshold-based methods seem capable of accurately classifying ponded meltwater on ice shelves, classifying surface facies such as slush, which have similar spectral properties to much of their surroundings, requires more spectral information. Whilst threshold-based approaches do not exclude the use of more spectral information, the manual selection of each threshold is arduous. ML overcomes this as it is able to determine which spectral information is of value for each classification based upon the training data.

4.3 Evolution of slush and ponded water over the Roi Baudouin Ice Shelf

To demonstrate the capability of our supervised classifier for pan-Antarctic identification of slush and ponded water over time, we applied it across the RBIS for the Landsat 8 images between 2013 and 2020. Of the seven melt seasons investigated (2013/2014 to 2019/2020), the greatest total meltwater extent (5.0×10^9 m^2) was recorded between 30 January and 13 February 2017. This observation is corroborated by Halberstadt and others (2020) who classified surface lakes on the RBIS over a number of image scenes between 2013 and 2018, and found peak melt area on the 25th February 2017. Furthermore, our findings align with studies on the Amery Ice Shelf, where threshold-based methods (Moussavi and others, 2020) and ML methods (Halberstadt and others, 2020) were used to calculate the area of surface lakes over a single path/row. Similarly to Moussavi and others (2020), whilst we identified marked inter-annual variability in both slush and ponded melt areas, we found the intra-seasonal trends for inferred meltwater storage to be fairly consistent.

As slush (which may be saturated firn or saturated snow overlying blue ice or refrozen lakes or extensive ice layers of refrozen previously infiltrated water) accounted for an average of 64% of the total meltwater area on the RBIS over the full study period, our findings highlight the importance of accurately mapping slush extent in addition to ponded water extent when investigating surface meltwater on Antarctic ice shelves. Most work until this point has focussed on meltwater stored in surface lakes, owing to

their significance for potential hydrofracture-induced ice-shelf collapse. For example, work by Stokes and others (2019) identified >1300 km^2 of surface meltwater held in surface lakes across East Antarctica in January 2017. Based on our findings, in January 2017, the mean proportion of slush on the RBIS was 59%. Whilst the proportion of slush on other East Antarctic ice shelves has not yet been quantified, our observations of the proportion of slush across the RBIS highlight the need to account for slush when calculating total surface meltwater areas, and it is likely that the total area of meltwater across East Antarctica far exceeds the 1300 km^2 of ponded meltwater that has been reported by Stokes and others (2019).

We found that the proportion of slush relative to ponded meltwater across the RBIS was greatest between 16 November and 30 December each melt season (excluding 2013/2014, when it was greatest between 15 January and 29 January 2014). Whilst no previous literature has mapped the extent of slush on an interannual timescale, Bell and others (2017) used a simple $NDWI_{ice}$ threshold to identify slush on a small area of the Nansen Ice Shelf in the 2013/2014 melt season. They found the area of slush was greatest on 26th December 2013 and then gradually declined throughout early January 2014 (Bell and others, 2017). Whilst this trend contradicts our findings for the 2013/2014 season on the RBIS, it corroborates the trends we identify through the remaining six melt seasons (2014/2015 to 2019/2020). Bell and others (2017) suggested that the expansive slush identified on the Nansen Ice Shelf in December coalesced to form ponded meltwater by early January. We propose that a similar transition occurred across the RBIS, as the percentage of the total meltwater on the ice shelf held in slush generally fell from the end of December and into early January, and an increasing amount of melt was therefore held in water bodies.

For surface meltwater to pond, the underlying surface needs to be impermeable, and is likely, therefore, to be either blue ice or saturated firn (slush). Based on the results presented here (Figure 7) many pixels that are classified as ponded water are also classified as slush at least once in the melt season. Over the full study period (2013-2020), 26% of all water-covered pixels are occupied by slush and ponded water at least once. In these locations, therefore, it is likely that as melt increases throughout

667 the melt season, the firn layer becomes increasingly saturated and water can no longer
668 percolate into the firn pack, which results in ponding at the surface, and lateral transfer
669 of meltwater across the ice shelf surface. However, we also note that some pixels are
670 classified as only ponded meltwater during a melt season, and were therefore not
671 preceded by slush (Figure 7). Evidence for this is seen in all melt seasons and is
672 particularly prominent towards the central grounding line. We postulate that these
673 areas of ponded melt are filling depressions within blue ice surfaces or are forming on
674 top of melt ponds which may have refrozen.
675
676 Exposed blue ice surfaces have been identified previously in proximity to the Roi
677 Baudouin grounding line, and result from katabatic winds which cause snow erosion
678 and an increase in near-surface temperatures as winds cause mixing in the stable
679 boundary layer and adiabatic warming (Vihma and others, 2011; Lenaerts and others,
680 2017). Lenaerts and others (2017) attributed a doubling in summer surface melt at the
681 grounding line to the katabatic winds, and they also noted that the exposed blue ice
682 surfaces will contribute to further melt, as they have a lower surface albedo than snow-
683 covered surfaces. These processes help to explain the main patterns of ponded
684 meltwater that we observe across the RBIS, as ponded meltwater is clustered near to
685 the grounding line (Figures 6 and 7).
686
687 **4.4 Errors arising from cloud and cloud shadows**
688 In both the validation dataset and the larger Roi Baudouin dataset, errors of
689 commission due to cloud and cloud shadows are evident (Figures 4-6), which
690 highlights a limitation of our classifier. For example, from 31 December 2016 to 14
691 January 2017, and through to the end of the melt season, errors of commission are
692 identified over the central and distal regions of the RBIS (e.g. see red panel in Figure
693 6). Similar errors are identified within the maximum melt extent products (Figure 7).
694 This limitation has also been found in similar previous work (e.g. Halberstadt and
695 others, 2020), with errors resulting from imperfect cloud masking methods.
696
697 The transient nature of cloud and cloud shadows mean that these errors of
698 commission will have a low persistence over an entire melt season. This is

699 demonstrated by Figure 9, which shows the number of times over the full study period
700 that a pixel is classified as either slush or ponded water over the RBIS. The errors of
701 commission in the central and distal regions of the ice shelf have a persistence score
702 of one (Figure 9, grey pixels), meaning that each pixel was classified as water at only
703 a single point in time. In contrast, areas of extensive meltwater towards the southeast
704 and central southern grounding line generally have higher persistence values (Figure
705 9). Therefore, a potential solution to errors of commission resulting from cloud and
706 cloud shadows when looking at maximum melt products for each melt season would
707 be to filter out pixels with a persistence of one. However, this would lead to the removal
708 of some true positives, where water has been correctly classified at its maximum
709 extent for the melt season, but for only a single point in time. Future work is needed to
710 develop methods to reduce the errors of commission introduced by clouds, either at
711 the pre-processing stage prior to classifier development, or post classifier application.
712 Meanwhile, our maximum melt extents (Figure 7) are likely to be overestimates.
713
714
715

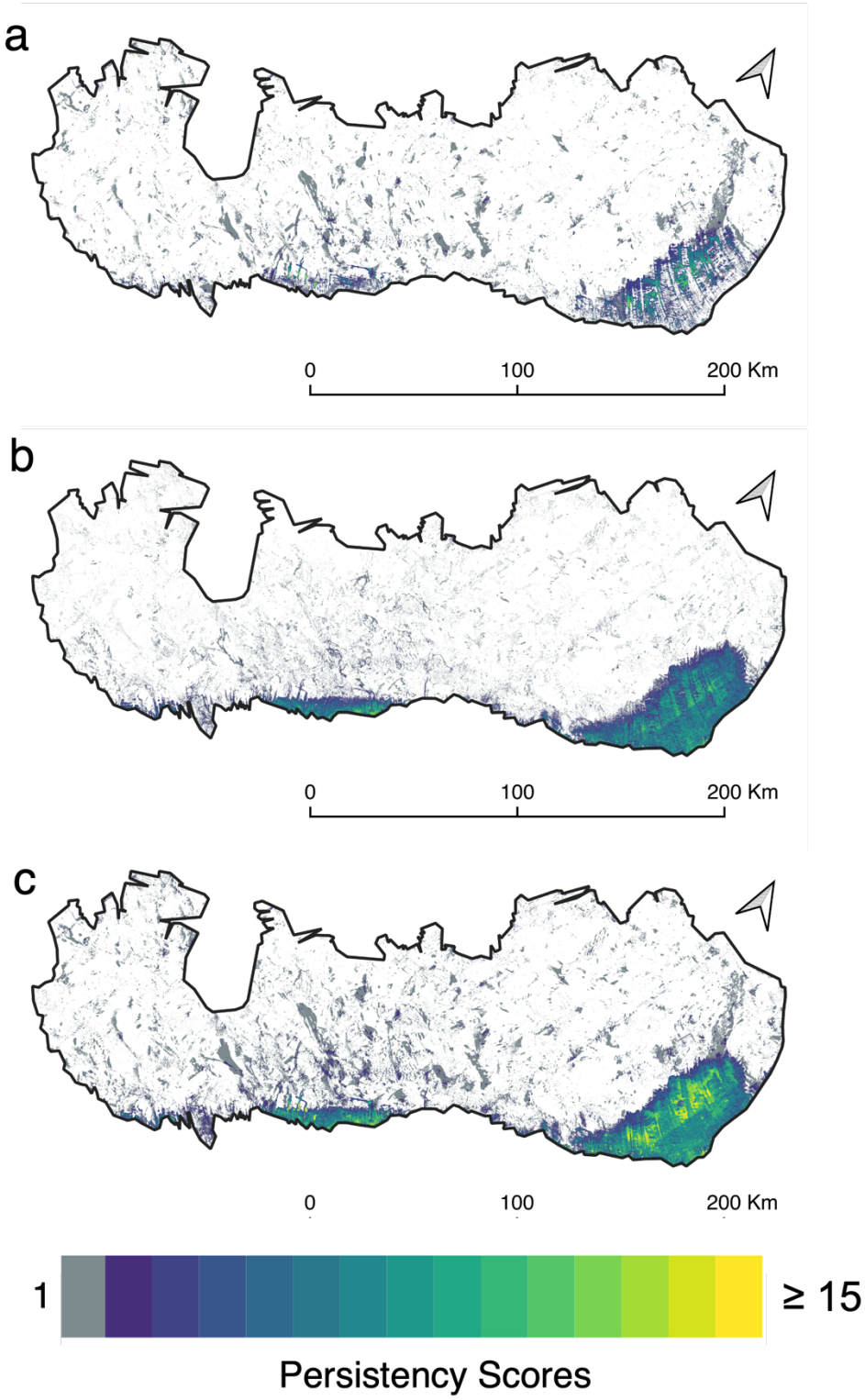

Figure 9: Heatmap showing the number of times (i.e. persistency scores) each pixel is classified as a) slush, b) ponded water and c) either slush or ponded water over all of the 15-day products produced for the full study period (2013 to 2020).

5 CONCLUSIONS

We have presented a ML method that is capable of accurately classifying slush and ponded water across Antarctic ice shelves using the Landsat 8 record from 2013 to 2020. This is achieved by using a Random Forest Classifier, which is trained using spectral data from six different ice shelves around the continent. The classifier performs well across all ice shelves throughout multiple melt seasons, achieving mean accuracies of 84% for ponded water and 82% for slush. Whilst the classifier encounters errors when defining the slush/ponded-water boundary, we also find that experts disagree on where this boundary should lie, and it is therefore likely that the extent of slush cannot be more accurately mapped without the collection of ground-truthed data. Errors of commission caused by cloud and cloud shadows are the main source of error associated with this method. Future work should look to improve cloud masking approaches before applying the classifier, or to develop a means of filtering out false positives caused by clouds after the classifier has been applied. In this way, it will be possible to produce accurate time series of slush and ponded meltwater extent across all Antarctic ice shelves.

Finally, we applied the classifier to the RBIS for the 2013/2014 to 2019/2020 melt seasons in order to produce a time series of slush and ponded melt extent. For each melt season, many of the pixels classified as ponded water were also classified as slush; an observation that likely captures the saturation of firn and subsequent formation of surface ponds as the melt season progresses. The mean average of slush over time is about two thirds of the total meltwater extent (between 2013 and 2020). This highlights the need to map slush in addition to ponded water on ice shelves over a pan-Antarctic scale, to ensure we do not underestimate the area of surface meltwater. The accurate time series data produced by this method, which captures all surface meltwater across Antarctic ice shelves should be used to validate and improve surface mass balance models.

ACKNOWLEDGEMENTS

We thank Mahsa Moussavi for her guidance throughout this project. Rebecca L. Dell was funded by a Natural Environment Research Council (NERC) Doctoral Training

Partnership Studentship (CASE with the British Antarctic Survey, grant no. NE/L002507/1). Alison F. Banwell received support from the U.S. National Science Foundation (NSF) under award no. 1841607 to the University of Colorado Boulder. Ian Willis was supported by the UK Natural Environment Research Council under NE/T006234/1 awarded to the University of Cambridge. Ice-shelf boundaries were downloaded from the SCAR Antarctic Digital Database (Gerrish and others, 2021). Key code and derived data from this study will be made available in the Apollo – University of Cambridge Repository.

REFERENCES

Alley, K. E., Scambos, T. A., Miller, J. Z., Long, D. G., & MacFerrin, M. (2018). Quantifying vulnerability of Antarctic ice shelves to hydrofracture using microwave scattering properties. *Remote Sensing of Environment*, *210*, 297–306. https://doi.org/10.1016/j.rse.2018.03.025

Arthur, D., & Vassilvitskii, S. (2007). K-means++: The advantages of careful seeding. *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms*, *07-09-Janu*, 1027–1035. Stanford.

Arthur, J. F., Stokes, C., Jamieson, S. S., Carr, J. R., & Leeson, A. A. (2020a). Recent understanding of Antarctic supraglacial lakes using satellite remote sensing. *Progress in Physical Geography*, 030913332091611. https://doi.org/10.1177/0309133320916114

Arthur, J., Stokes, C., Jamieson, S., Carr, J. R., & Leeson, A. (2020b). Distribution and seasonal evolution of supraglacial lakes on Shackleton Ice Shelf, East Antarctica. *The Cryosphere Discussions*, 1–36. https://doi.org/10.5194/tc-2020-101

Bamber, J. L., & Aspinall, W. P. (2013). An expert judgement assessment of future sea level rise from the ice sheets. *Nature Climate Change*, *3*(4), 424–427. https://doi.org/10.1038/nclimate1778

Banwell, A. F., Willis, I. C., Macdonald, G. J., Goodsell, B., & MacAyeal, D. R. (2019). Direct measurements of ice-shelf flexure caused by surface meltwater ponding and drainage. *Nature Communications*, *10*(1), 1–10.

https://doi.org/10.1038/s41467-019-08522-5

Banwell, A. F., Datta, R. T., Dell, R. L., Moussavi, M., Brucker, L., Picard, G., Shuman, C. A., and Stevens, L. A. (2021). The 32-year record-high surface melt in 2019/2020 on the northern George VI Ice Shelf, Antarctic Peninsula, The Cryosphere, 15, 909–925, https://doi.org/10.5194/tc-15-909-2021

Banwell, A. F. & MacAyeal, D. R. 2015, Ice-shelf fracture due to viscoelastic-flexure stress induced by fill/drain cycles of supraglacial lakes. *Antarctic Science*, doi: 10.1017/50954102015000292.

Banwell, A. F., Caballero, M., Arnold, N. S., Glasser, N. F., Cathles, L. Mac, & MacAyeal, D. R. (2014). Supraglacial lakes on the Larsen B ice shelf, Antarctica, and at Paakitsoq, West Greenland: A comparative study. *Annals of Glaciology*, *55*(66), 1–8. https://doi.org/10.3189/2014AoG66A049

Banwell, A. F., MacAyeal, D. R., & Sergienko, O. V. (2013). Breakup of the Larsen B Ice Shelf triggered by chain reaction drainage of supraglacial lakes. *Geophysical Research Letters*, *40*(22), 5872–5876. https://doi.org/10.1002/2013GL057694

Bell, R. E., Chu, W., Kingslake, J., Das, I., Tedesco, M., Tinto, K. J., … Lee, W. S. (2017). Antarctic ice shelf potentially stabilized by export of meltwater in surface river. *Nature*, *544*(7650), 344–348. https://doi.org/10.1038/nature22048

Bell, R. E., Banwell, A. F., Trusel, L. D., & Kingslake, J. (2018). Antarctic surface hydrology and impacts on ice-sheet mass balance. *Nature Climate Change*, *8*(12), 1044–1052. https://doi.org/10.1038/s41558-018-0326-3

Berthier, E., Scambos, T. A., & Shuman, C. A. (2012). Mass loss of Larsen B tributary glaciers (Antarctic Peninsula) unabated since 2002. *Geophysical Research Letters*, *39*(13), n/a-n/a. https://doi.org/10.1029/2012GL051755

Dell, R., Arnold, N., Willis, I., Banwell, A., Williamson, A., Pritchard, H., & Orr, A. (2020). Lateral meltwater transfer across an Antarctic ice shelf. *The Cryosphere*, *14*(7), 2313–2330. https://doi.org/10.5194/tc-14-2313-2020

Dirscherl, M., Dietz, A. J., Kneisel, C., & Kuenzer, C. (2020). Automated Mapping of Antarctic Supraglacial Lakes Using a Machine Learning Approach. *Remote Sensing*, 1–27. https://doi.org/10.3390/rs12071203

Dirscherl, M., Dietz, A. J., Kneisel, C., & Kuenzer, C. (2021). A novel method for automated supraglacial lake mapping in antarctica using sentinel-1 sar imagery

814 and deep learning. *Remote Sensing*, *13*(2), 1–27.
815 https://doi.org/10.3390/rs13020197
816 Dunmire, D., Lenaerts, J. T. M., Banwell, A. F., Wever, N., Shragge, J., Lhermitte, S.,
817 … Keenan, E. (2020). Observations of buried lake drainage on the Antarctic Ice
818 Sheet. *Geophysical Research Letters*, *47*(15).
819 https://doi.org/10.1029/2020gl087970
820 Frezzotti, M. (1993). Glaciological study in Terra Nova Bay, Antarctica, inferred from
821 remote sensing analysis. *Annals of Glaciology*, *17*, 63–71.
822 https://doi.org/10.3189/s0260305500012623
823 Fricker, H. A., Arndt, P., Brunt, K. M., Datta, R. T., Fair, Z., Jasinski, M. F., …
824 Wouters, B. (2020). ICESat-2 melt depth retrievals: application to surface melt
825 on Amery Ice Shelf, East Antarctica. *Geophysical Research Letters*,
826 e2020GL090550. https://doi.org/10.1029/2020gl090550
827 Fürst, J. J., Durand, G., Gillet-Chaulet, F., Tavard, L., Rankl, M., Braun, M., &
828 Gagliardini, O. (2016). The safety band of Antarctic ice shelves. *Nature Climate*
829 *Change*, *6*(5), 479–482. https://doi.org/10.1038/nclimate2912
830 Gerrish, L., Fretwell, P., & Cooper, P. (2021). High resolution vector polylines of the
831 Antarctic coastline (7.4) [Data set]. UK Polar Data Centre, Natural Environment
832 Research Council, UK Research & Innovation. https://doi.org/10.5285/e46be5bc-
833 ef8e-4fd5-967b-92863fbe2835
834 Halberstadt, A. R. W., Gleason, C. J., Moussavi, M. S., Pope, A., Trusel, L. D., &
835 DeConto, R. M. (2020). Antarctic supraglacial lake identification using landsat-8
836 image classification. *Remote Sensing*, *12*(8), 1–29.
837 https://doi.org/10.3390/RS12081327
838 Hubbard, B., Luckman, A., Ashmore, D. W., Bevan, S., Kulessa, B., Munneke, P. K.,
839 ... & Rutt, I. (2016). Massive subsurface ice formed by refreezing of ice-shelf
840 melt ponds. *Nature communications*, *7*(1), 1-6.
841 IPCC. (2019). *Special Report on the Ocean and Cryosphere in a Changing Climate -*
842 *Technical Summary (Final Draft)*. Retrieved from https://www.ipcc.ch/srocc/
843 Kingslake, J., Ely, J. C., Das, I., & Bell, R. E. (2017). Widespread movement of
844 meltwater onto and across Antarctic ice shelves. *Nature*, *544*(7650), 349–352.
845 https://doi.org/10.1038/nature22049

Kuipers Munneke, P., Ligtenberg, S. R. M., Van Den Broeke, M. R., & Vaughan, D. G. (2014). Firn air depletion as a precursor of Antarctic ice-shelf collapse. *Journal of Glaciology*, *60*(220), 205–214. https://doi.org/10.3189/2014JoG13J183

Lai, C. Y., Kingslake, J., Wearing, M. G., Chen, P. H. C., Gentine, P., Li, H., … van Wessem, J. M. (2020). Vulnerability of Antarctica's ice shelves to meltwater-driven fracture. *Nature*, *584*(7822), 574–578. https://doi.org/10.1038/s41586-020-2627-8

Langley, E. S., Leeson, A. A., Stokes, C. R., & Jamieson, S. S. R. (2016). Seasonal evolution of supraglacial lakes on an East Antarctic outlet glacier. *Geophysical Research Letters*, *43*(16), 8563–8571. https://doi.org/10.1002/2016GL069511

Lenaerts, J., Lhermitte, S., Drews, R., Ligtenberg, S. R. M., Berger, S., Helm, V., … Pattyn, F. (2017). Meltwater produced by wind – albedo interaction stored in an East Antarctic ice shelf. *Nature Climate Change*, *7*, 58–63. https://doi.org/10.1038/NCLIMATE3180

Macdonald, G. J., Banwell, A. F., Willis, I. A. N. C., Mayer, D. P., Goodsell, B., & Macayeal, D. R. (2019). Formation of pedestalled, relict lakes on the McMurdo Ice Shelf, Antarctica. *Journal of Glaciology*, *65*(250), 337–343. https://doi.org/10.1017/jog.2019.17

Montgomery, L., Miège, C., Miller, J., Scambos, T. A., Wallin, B., Miller, O., … Koenig, L. (2020). Hydrologic Properties of a Highly Permeable Firn Aquifer in the Wilkins Ice Shelf, Antarctica. *Geophysical Research Letters*, *47*(22), e2020GL089552. https://doi.org/10.1029/2020GL089552

Moussavi, M., Pope, A., Halberstadt, A. R. W., Trusel, L. D., Cioffi, L., & Abdalati, W. (2020). Antarctic Supraglacial Lake Detection Using Landsat 8 and Sentinel-2 Imagery: Towards Continental Generation of Lake Volumes. *Remote Sensing*, *12*(1), 134. https://doi.org/10.3390/rs12010134

Reynolds, J. M. (1981). Lakes on george vi ice shelf, antarctica. *Polar Record*, *20*(128), 425–432. https://doi.org/10.1017/S0032247400003636

Rignot, E., Casassa, G., Gogineni, P., Krabill, W., Rivera, A., & Thomas, R. (2004). Accelerated ice discharge from the Antarctic Peninsula following the collapse of Larsen B ice shelf. *Geophysical Research Letters*, *31*(18), L18401.

https://doi.org/10.1029/2004GL020697
Robel, A. A., & Banwell, A. F. (2019). A Speed Limit on Ice Shelf Collapse Through Hydrofracture. *Geophysical Research Letters*, *46*(21), 12092–12100. https://doi.org/10.1029/2019GL084397
Scambos, T. A., Bohlander, J. A., Shuman, C. A., & Skvarca, P. (2004). Glacier acceleration and thinning after ice shelf collapse in the Larsen B embayment, Antarctica. *Geophysical Research Letters*, *31*(18), L18402. https://doi.org/10.1029/2004GL020670
Scambos, T., Hulbe, C., & Fahnestock, M. (2003). Climate-induced ice shelf disintegration in the Antrctic Peninsula. In *Antarctica Peninsula climate variability: a historical and paleo-environmental perspective* (pp. 79–92). https://doi.org/10.1029/AR079p0079
Spergel, J. J., Kingslake, J., Creyts, T., van Wessem, M., & Fricker, H. A. (2021). Surface meltwater drainage and ponding on Amery Ice Shelf, East Antarctica, 1973–2019. *Journal of Glaciology*, 1-14.
Stehman, S.V., 1997. Selecting and interpreting measures of thematic classification accuracy. *Remote sensing of Environment*, *62*(1), pp.77-89.
Stokes, C. R., Sanderson, J. E., Miles, B. W. J., Jamieson, S. S. R., & Leeson, A. A. (2019). Widespread distribution of supraglacial lakes around the margin of the East Antarctic Ice Sheet. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-50343-5
Trusel, L. D., Frey, K. E., Das, S. B., Karnauskas, K. B., Kuipers Munneke, P., Van Meijgaard, E., & Van Den Broeke, M. R. (2015, December 1). Divergent trajectories of Antarctic surface melt under two twenty-first-century climate scenarios. *Nature Geoscience*, Vol. 8, pp. 927–932. https://doi.org/10.1038/ngeo2563
Vihma, T., Tuovinen, E., & Savijärvi, H. (2011). Interaction of katabatic winds and near-surface temperatures in the Antarctic. *Journal of Geophysical Research: Atmospheres*, *116*(D21).
Wagner, A. C. (1972). Flooding of the Ice Shelf in George VI Sound. *Antarctic Survey Bulletin*, *28*, 71–74.
Williamson, A. G., Banwell, A. F., Willis, I. C., & Arnold, N. S. (2018). Dual-satellite

(Sentinel-2 and Landsat 8) remote sensing of supraglacial lakes in Greenland. *The Cryosphere*, *12*(9), 3045–3065. https://doi.org/10.5194/tc-12-3045-2018

Work, E. A., & Gilmer, D. S. (1976). UTILIZATION OF SATELLITE DATA FOR INVENTORYING PRAIRIE PONDS AND LAKES. *Photogrammetric Engineering and Remote Sensing*, *42*(5), 685–694.

Yang, H., Wang, Z., Zhao, H., & Guo, Y. (2011). Water body extraction methods study based on RS and GIS. *Procedia Environmental Sciences*, *10*(PART C), 2619–2624. https://doi.org/10.1016/j.proenv.2011.09.407

Yang, K., & Smith, L. C. (2013). Supraglacial streams on the Greenland Ice Sheet delineated from combined spectral – shape information in high-resolution satellite imagery. *IEEE Geoscience and Remote Sensing Letters*, *10*(4), 801–805. https://doi.org/10.1109/LGRS.2012.2224316