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

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

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

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 2013–20 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 research 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 ponding (Kuipers Munneke and others, 2014; Hubbard and others, 2016; Alley and others, 2018). Ponded water has been shown to drive ice-shelf collapse events through hydrofracture (Scambos and others, 2003, 2004; Banwell and others, 2013; Banwell and MacAyeal, 2015; Robel and Banwell, 2019) and therefore several studies have mapped the changing extent of ponded water on ice shelves (e.g. Arthur and others, 2020a, 2020b; Dell and others, 2020; Spergel and others, 2021). Despite the potential role of water as slush in driving hydrofracture, there has been very little research investigating the changing extent of slush on ice shelves. This means that previous research 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 föhn 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 2019/20 melt season, when sustained periods of warm air temperatures above 0°C led to a 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. Although most threshold-based approaches rely solely on an normalised difference water index of ice ($NDWI_{ice}$) threshold (e.g. Dell and others, 2020; Williamson and others, 2018), Moussavi and others (2020) employ 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 is built upon the study 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. Although 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. Although 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 an 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 2013–20 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 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 (Fig. 1); (i) Nivlisen, (ii) Roi Baudouin, (iii) Amery, (iv) Shackleton, (v) Nansen and (vi) George VI (Fig. 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.

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 (Fig. 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). Fourteen 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 S.1). 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 that were acquired at a range of solar elevations (20.9° to 36.6°) (Table S.1). 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 values (Dell and others, 2020), and by clipping to the ice-shelf boundaries (from the SCAR Antarctic Digital Database). 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 (Yang and Smith, 2013; Bell and others, 2017). 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', 'ponded water' or 'other' by the classifier at a later stage. $NDWI_{ice}$ was calculated using Landsat 8 bands 2 (blue) and 4 (red):

$$NDWI_{ice} = (\text{Blue Band} - \text{Red Band}) / (\text{Blue Band} + \text{Red Band}) \quad (1)$$
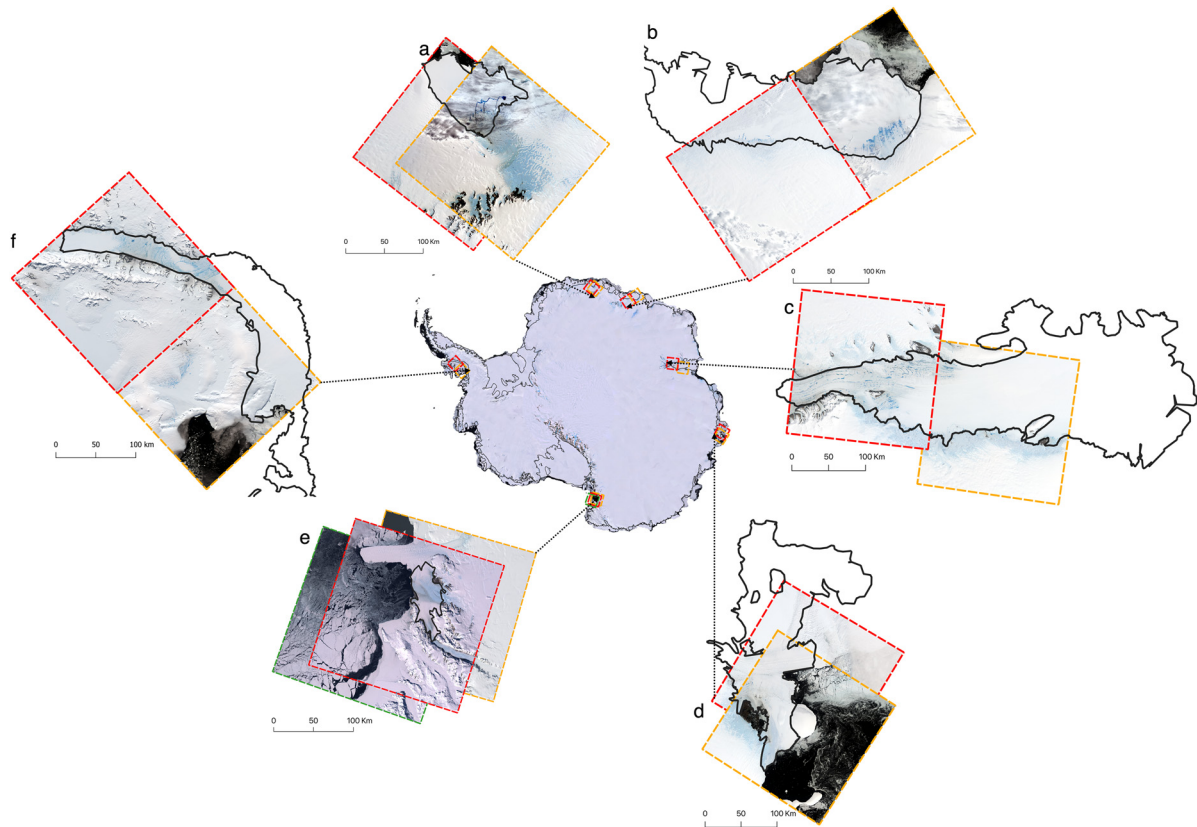
Fig. 1. Study area figure showing the six ice shelves selected for use in the unsupervised *k*-means clustering algorithm. Dashed coloured boxes indicate the location of the Landsat-8 training images for (a) Nivlisen, (b) Roi Baudouin, (c) Amery, (d) Shackleton, (e) Nansen and (f) George VI, with different coloured boxes indicating different paths/rows. Whilst images used in this Figure show training image locations, they are not necessarily the training images themselves, a record of images used is provided in Tabel S.1. Ice-shelf boundaries (from the SCAR Antarctic Digital Database) 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).

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 (Fig. 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.

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 several others (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 misclassification 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 GEE. 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.

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 (Fig. 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

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 (km²)	General surface water characteristics	Key citations
Nivlisen Roi	70.7° S 11.7° E	Dronning Maud Land, East Antarctica	7 380	Elongate surface lakes expand towards ice shelf's calving front as melt season progresses	Dell and others (2020)
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 (2021); Spergel and others (2021)
Shackleton	66.4° S 100° E	Queen Mary Land, East Antarctica	28 600	Surface lakes near the grounding line	Arthur and others (2020b)
Nansen	74.9° S 162.8° E	Victoria Land, East Antarctica	2 270	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)
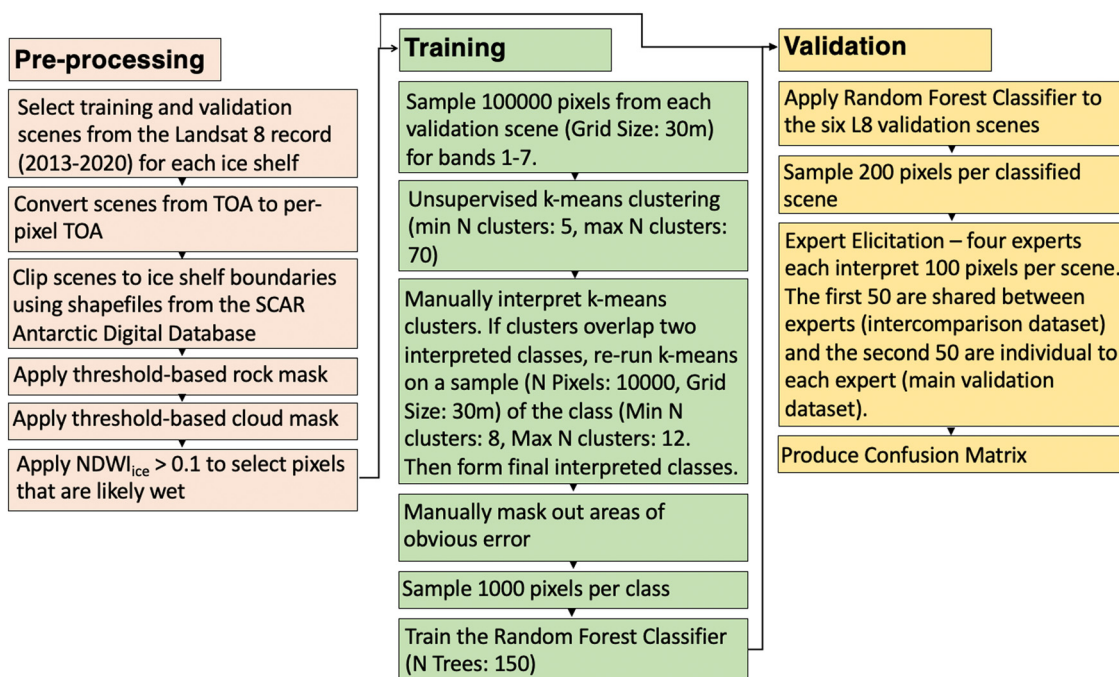
Fig. 2. Workflow detailing the pre-processing, training, and validation steps for creating a supervised classifier to map slush and ponded water across Antarctic ice shelves using GEE.
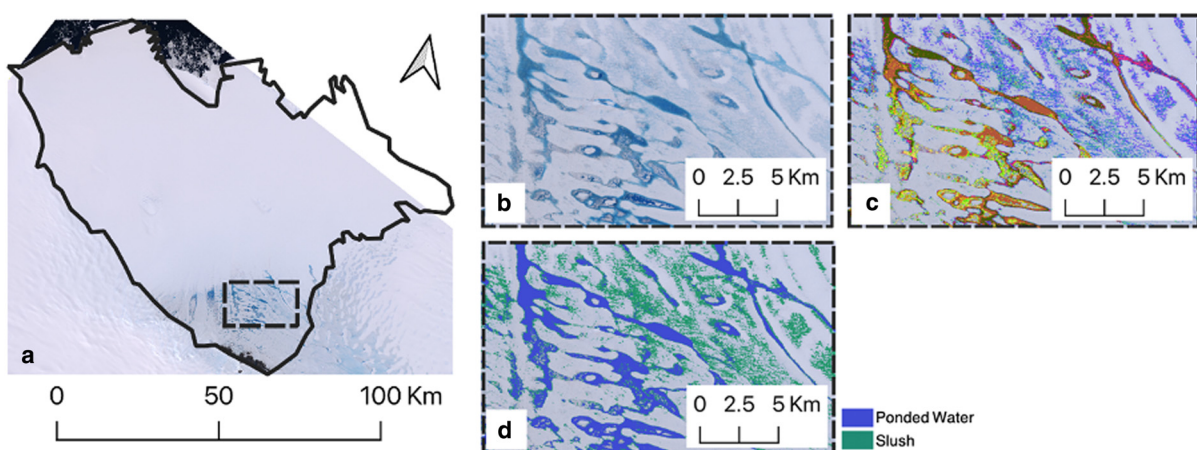
Fig. 3. Example workflow for the *k*-means clustering algorithm over the Nivlisen Ice Shelf (Landsat 8, 2016-12-27). (a) Base image of the Nivlisen Ice Shelf, the solid black line marks the ice-shelf area (from the SCAR Antarctic Digital Database), the dashed box shows the zoomed area featured in (b), (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, ten *k*-means clusters were combined to form the ponded water class, and ten *k*-means clusters were combined to form the slush class.
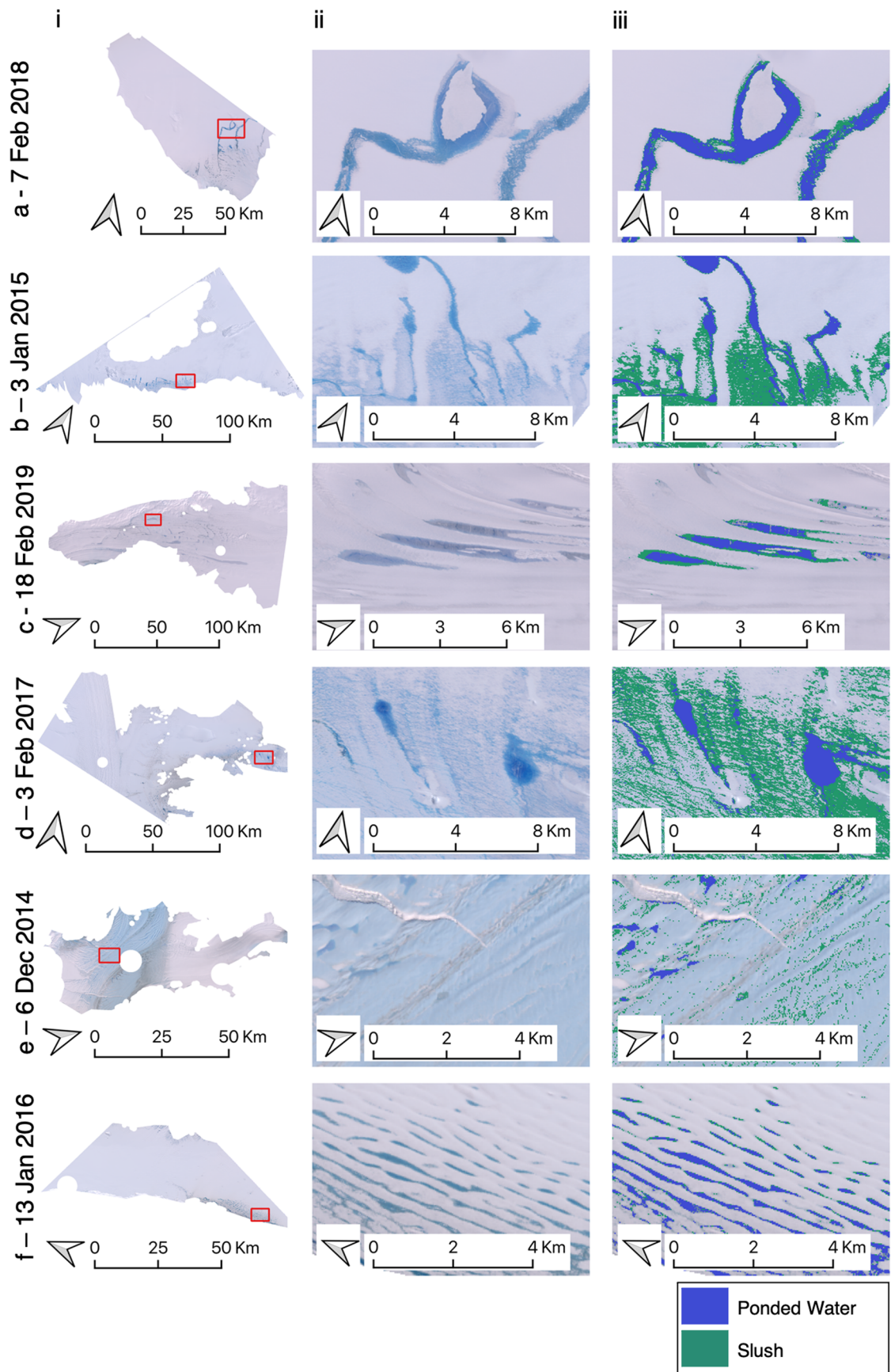
Fig. 4. Preliminary outputs from the supervised classifier, as applied to six Landsat 8 validation images for the (a) Nivlisen Ice Shelf, (b) RBIS, (c) Amery Ice Shelf, (d) Shackleton Ice Shelf, (e) Nansen Ice Shelf and (f) George VI Ice Shelf. Panels in column (i) show the pre-processed Landsat 8 RGB 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 2. 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 3. High-confidence accuracy scores for the intercomparison dataset , 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

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 of 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 slush or ponded water found within each 15-day product as a fraction of the visible ice-shelf area of each 15-day product, and then multiplying this fraction by the full ice-shelf area (Williamson and others, 2018; Banwell and others, 2021). In addition to the 15-day products that we exported from GEE, we compiled maximum melt extent maps for each meltseason in MATLAB (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

Table 2 shows the results from the intercomparison dataset for each scene in the validation dataset, which were interpreted by all four experts. The data shown include all interpreted pixels regardless of the associated confidence scores. Overall, the accuracy of the ponded water class is 78%, and the accuracy of the slush class is 71%. For the ponded water class, the experts all produced similar accuracy scores for the RBIS (8% spread), and more dissimilar scores for the Nansen Ice Shelf (30% spread), with a mean spread across all six ice shelves of just 6%. For the slush class, the experts are in the closest agreement over the George VI Ice Shelf (11% spread), and in least close agreement over the Nansen Ice Shelf (79% spread). As with the ponded water class, these discrepancies tend to cancel out between experts giving an overall mean spread across all ice shelves of just 5%. Table 3 shows the same data as Table 2, but only for the pixels for which the experts had 'high-confidence' in their interpretation.

Table 4 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

Table 4. Accuracy scores for the main validation dataset (the 200 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 5. High-confidence accuracy scores for the main validation dataset 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 6. 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%

to those produced by the intercomparison dataset. 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% (the Shackleton Ice Shelf).

Table 5 shows the accuracy results for the main validation dataset using high-confidence pixels only. The mean accuracy for the ponded water 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), whereas 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 6). However, band 5 (near-infrared) is of greatest importance for the supervised classifier, with an importance score of 20% (Table 6). Bands 1–4 (visible) and 6–7 (shortwave infrared 1 and 2) all have similar weightings, with importance scores ranging between 12 and 15%.

3.3 Application to the Roi Baudouin Ice Shelf

After applying the supervised classifier to the RBIS, 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 (AOI) coverage (Fig. 5). Of the 48 15-day periods presented in Fig. 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 Fig. 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 (Fig. 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 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 for 2018/19, 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 2018/19, the greatest total area of slush is recorded approximately a month prior to the greatest total area of ponded water (Fig. 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, whereas the percentage change in ponded water is slightly greater than the percentage change in slush. Slush ranges from 5.7×10^8 m^2 between 15 January and 29 January 2019, to 3.5×10^9 m^2 between 15 January and 29 January 2016 (a 521% change in area), whereas ponded water varies from 2.9×10^8 m^2 between 14 February 2019 and 28 February 2019, to 1.9×10^9 m^2 between 30 January 2017 and 13 February 2017 (a 559% change in area) (Table S.2). Overall, slush dominates the total meltwater area across the RBIS, making up over half of the total meltwater area 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 2014/15 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 2016/17 melt season has the greatest recorded total meltwater area, reaching 5×10^9 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 2019/20, with 7.5×10^8 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 6 shows each of the 15-day data products that were produced within GEE for the 2016/17 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
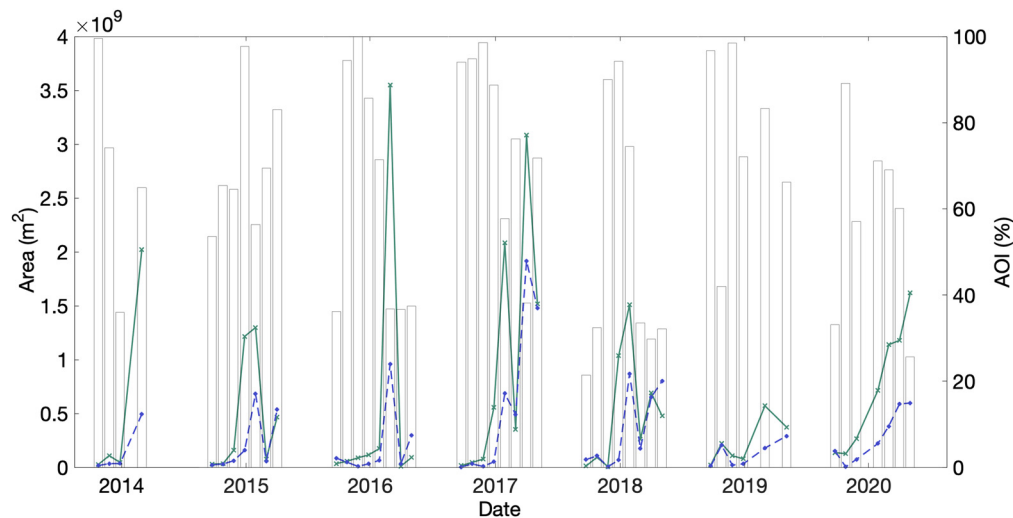
Fig. 5. Time series data for slush and ponded water across the RBIS. Grey bars show the % AOI coverage for each 15-day period plotted. Lines show scaled areas of slush (green line) and ponded water (blue line) on the RBIS from 2013/14 to 2019/20, derived from supervised classification of 15-day Landsat 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.

focus on the true positive results for the 2016/17 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 to 30 December 2016 onwards, areas of slush begin to develop near the grounding line in both the southeast and central southern parts. By early January (31 December 2016–14 January 2017) ponded water also begins to form among 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 (Figs 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 2016/17 melt season ranges from 38% (30 January–13 February 2017) to 99% (1 December–15 December 2016) (Table S.2).

Data products from GEE were combined in MATLAB to produce maximum melt extents across the RBIS for each melt season (1 November–31 March) from 2013/14 to 2019/20 (Fig. 7). In every melt season, both slush and ponded water are present predominantly in the southeast of the ice shelf, towards the grounding line. This area of slush and ponded water is the most spatially extensive in 2016/17 and 2017/18 (Figs 7d, e), when it extends ~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 (Fig. 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.

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 of 35% of all pixels within the main validation dataset) (Table 5). Over all ice shelves, the percentage of pixels that were classified with high confidence did not exceed 50% (Table 4), 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 3). 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. Although 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, although 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 with pixel-based methods for the classification of ponded water. In the future, research should look at collecting 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 research 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
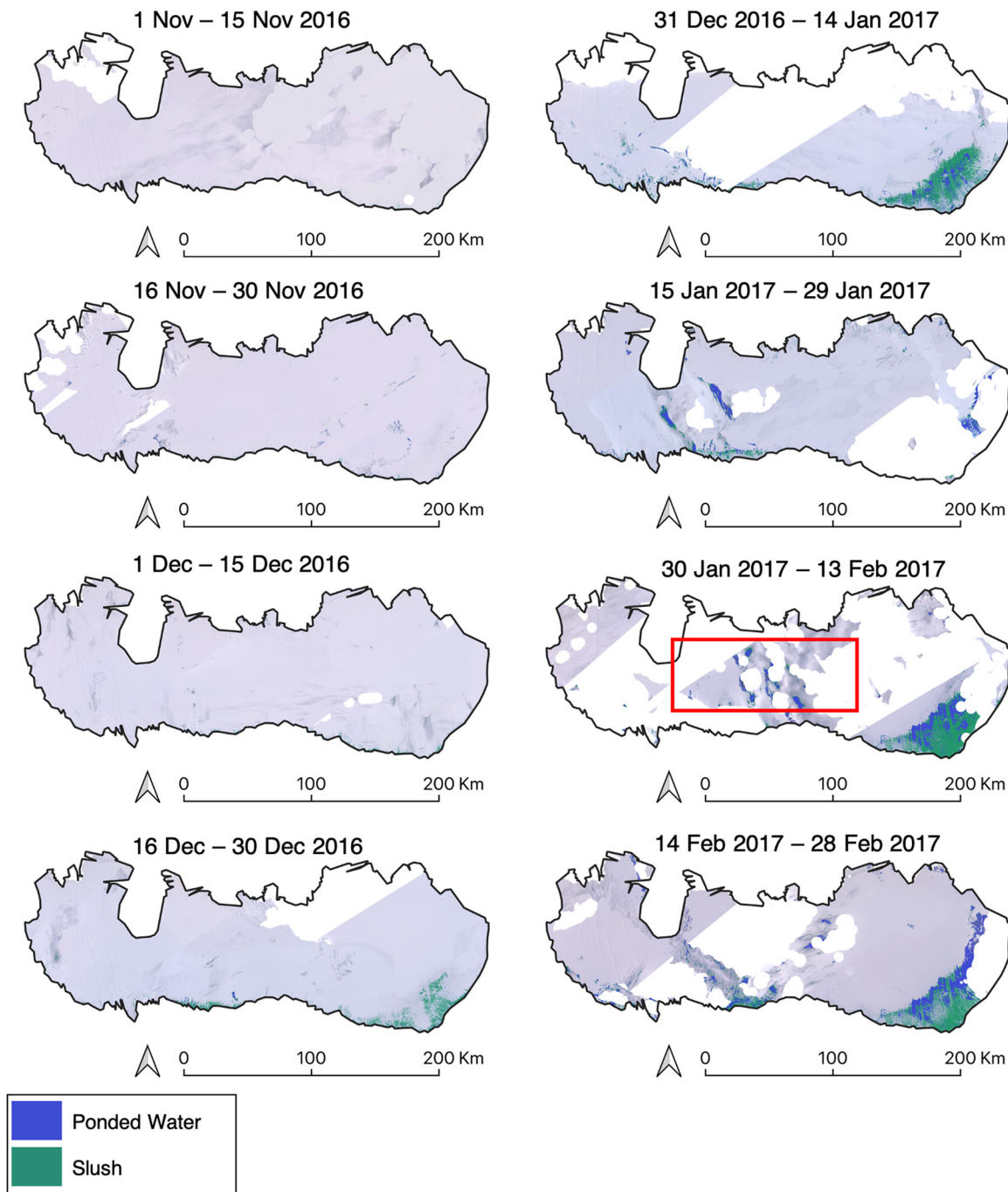
Fig. 6. 15-day melt products for the 2016/17 melt season across the RBIS. White areas are areas that have either been masked out or were not covered by imagery in the first instance. The red box in the 30 January 2017–13 February 2017 panel roughly denotes the area where errors of commission due to cloud and cloud shadows are generally found.

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 (Fig. 4). Additionally, on the validation image for the Amery Ice Shelf, there were examples of ponded water covered by a thin ice layer (Fig. 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. Although 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 nor the classifier were consistently more or less conservative when marking the slush/ponded-water boundary. Again, considering future research, 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.
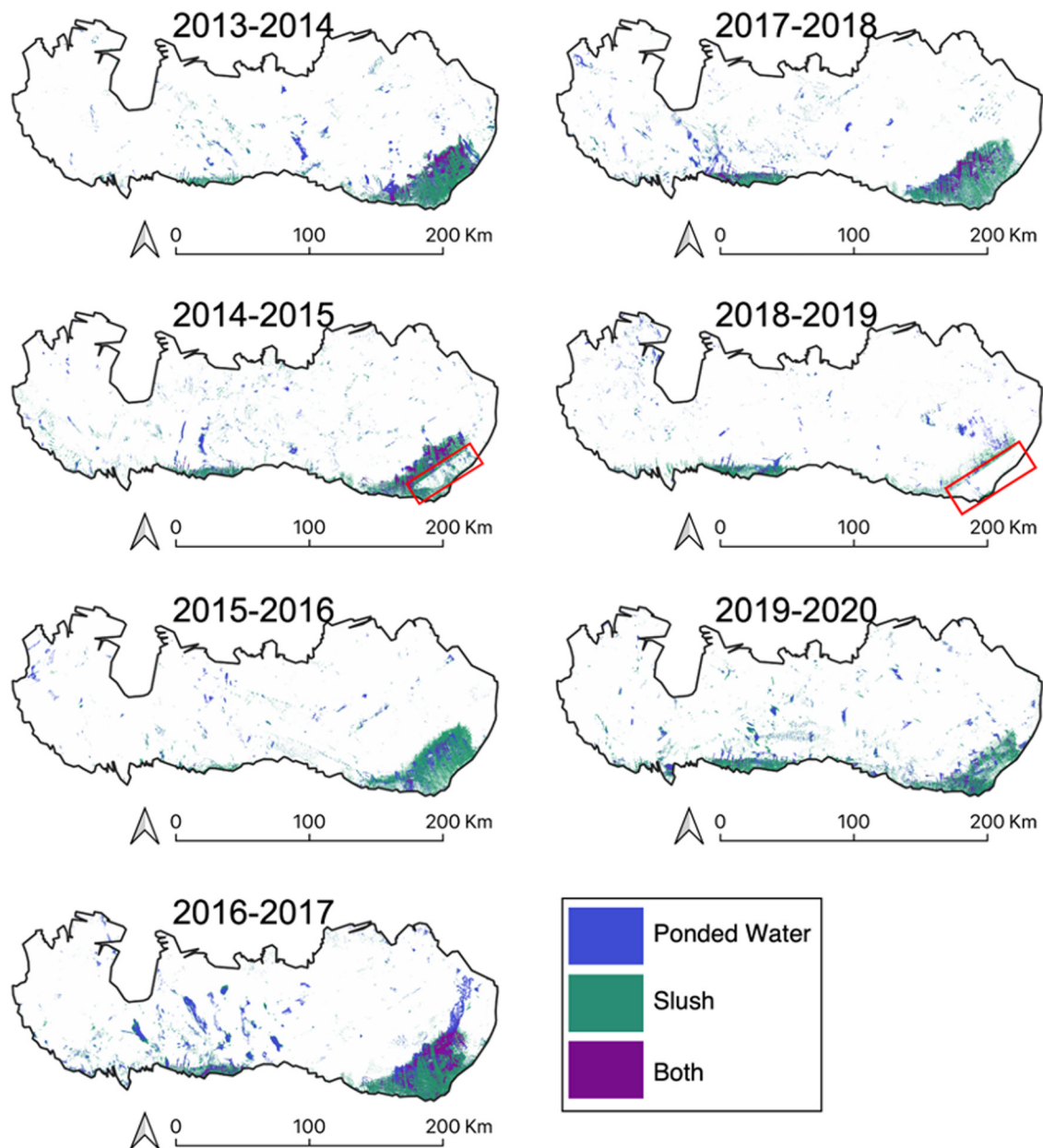
Fig. 7. Maximum melt extent plots for each melt season, calculated by mosaicking all 15-day melt products for each melt season in MATLAB. Maximum areas of slush, ponded water and both (where both slush and ponded water are identified within the melt season) are mapped. Red boxes roughly delineate areas affected by data gaps in the 2014/15 and 2018/19 melt seasons.

4.2 Comparison to NDWI$_{ice}$

Although 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 and dirty ice) (Fig. 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 ponded water and slush (Fig. 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 (Fig. 8). On the George VI Ice Shelf, however, the differences between the threshold method and the classifier output were small, although even here the threshold method tended to underestimate slush area compared to the classifier (Fig. 8).

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 6). This is likely related to the low reflectivity of water in near-infrared wavelengths (Work and Gilmer, 1976; Yang and others, 2011). Overall, although simple threshold-based methods seem capable of accurately classifying ponded meltwater on ice shelves, classifying surface facies such as slush, which have
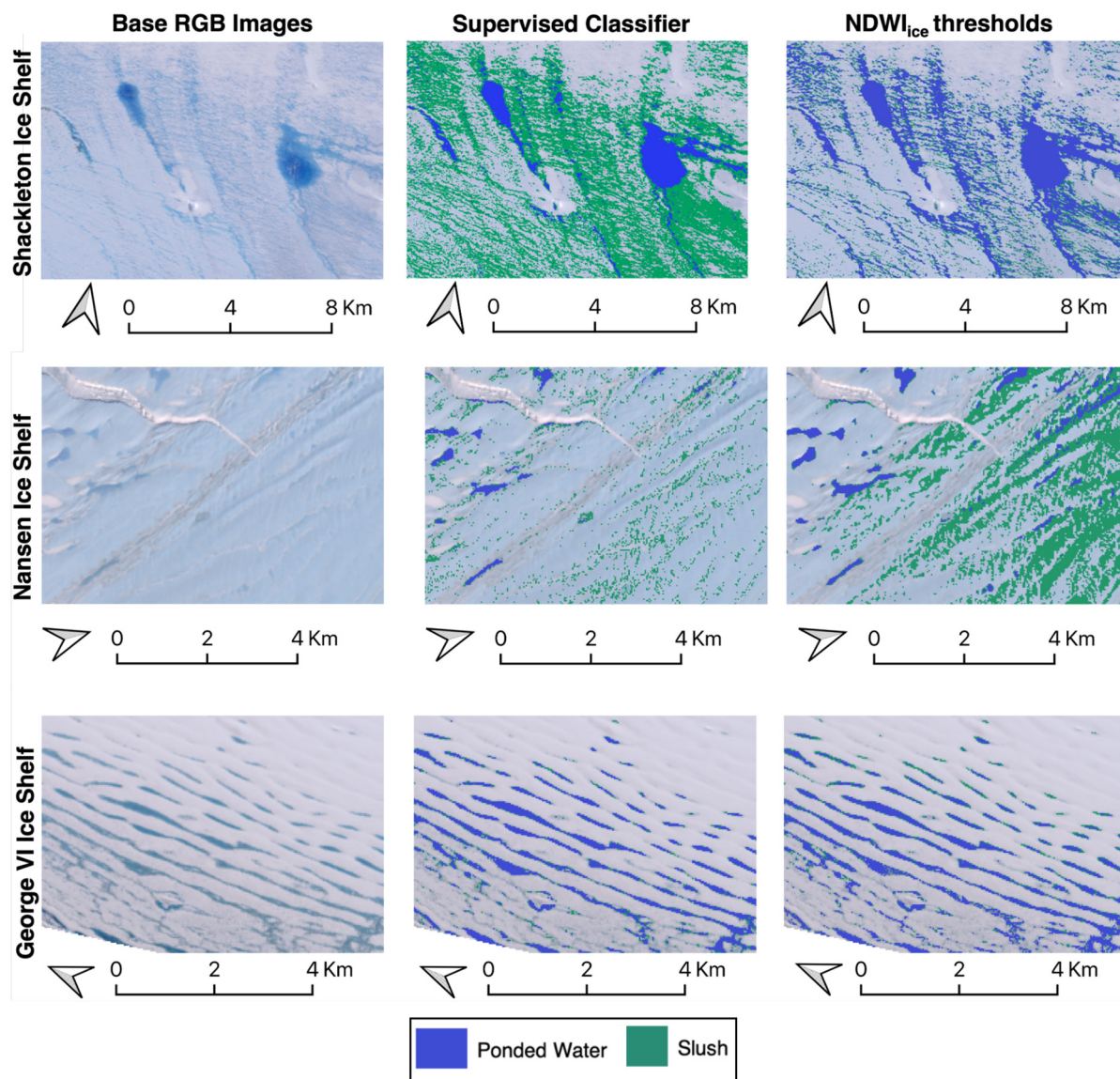
Fig. 8. Outputs from the supervised classifier and from NDWI$_{ice}$ thresholding applied to sections of the validation images (as shown in Fig. 4) for Shackleton, Nansen and George VI ice shelves. Panels show the base RGB images, the area classified using the supervised classifier developed in this study and the area classified using NDWI$_{ice}$ thresholds, where slush is >0.12 and ⩽ 0.14 and ponded water is > 0.14.

similar spectral properties to much of their surroundings, requires more spectral information. Although 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 potential 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 (2013/14 to 2019/20), the greatest total meltwater extent (5.0 × 10^9 m^2) was recorded between 30 January and 13 February 2017. This observation is broadly 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), although we identified marked inter-annual variability in both slush and ponded water 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, 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 research until this point has focused on meltwater stored in surface lakes, owing to their significance for potential hydrofracture-induced ice-shelf collapse. For example, a study 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%. Although 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
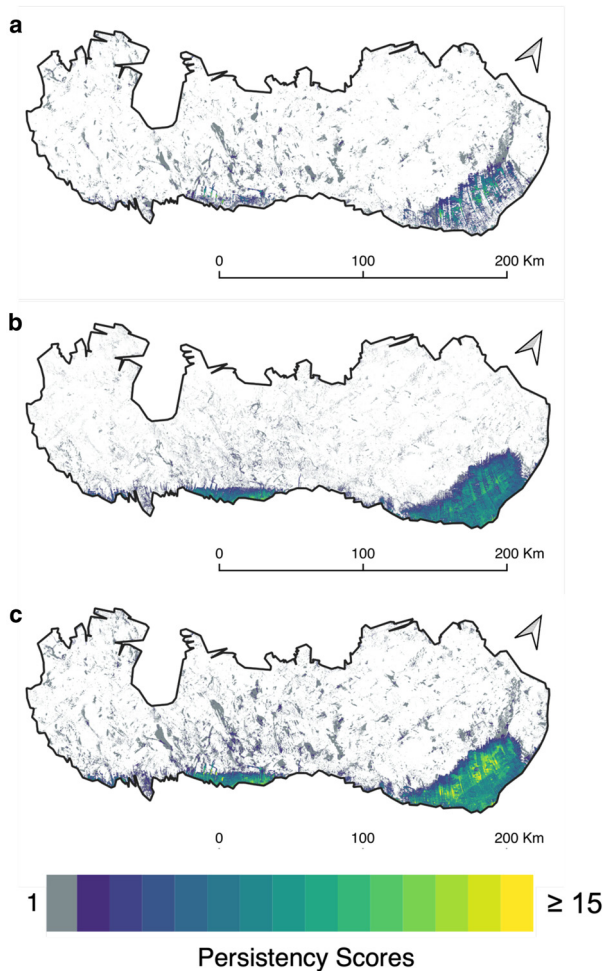
Fig. 9. Heatmap showing the number of times (i.e. persistency scores) each pixel is classified as (a) slush, (b) ponded water and (c) either slush or ponded water over all of the 15-day products produced for the full study period (2013–20).

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 2013/14, when it was greatest between 15 January and 29 January 2014). Although 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 2013/14 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). Although this trend contradicts our findings for the 2013/14 season on the RBIS, it corroborates the trends we identify through the remaining six melt seasons (2014/15 to 2019/20). 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 (Fig. 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 the melt season, the firn layer becomes increasingly saturated and water can no longer percolate into the firn pack, which results in ponding at the surface, and lateral transfer of meltwater across the ice shelf surface. However, we also note that some pixels are classified as only ponded water during a melt season, and were therefore not preceded by slush (Fig. 7). Evidence for this is seen in all melt seasons and is particularly prominent towards the central grounding line. We postulate that these areas of ponded water are filling depressions within blue ice surfaces or are forming on top of melt ponds which may have refrozen.

Exposed blue ice surfaces have been identified previously in proximity to the Roi Baudouin grounding line, and result from katabatic winds which cause snow erosion and an increase in near-surface temperatures as winds cause mixing in the stable boundary layer and adiabatic warming (Vihma and others, 2011; Lenaerts and others, 2017). Lenaerts and others (2017) attributed a doubling in summer surface melt at the grounding line to the katabatic winds, and they also noted that the exposed blue ice surfaces will contribute to further melt, as they have a lower surface albedo than snow-covered surfaces. These processes help to explain the main patterns of ponded meltwater that we observe across the RBIS, as ponded meltwater is clustered near to the grounding line (Figs 6 and 7).

4.4 Errors arising from cloud and cloud shadows

In both the validation dataset and the larger Roi Baudouin dataset, errors of commission due to cloud and cloud shadows are evident (Figs 4–6), which highlights a limitation of our classifier. For example, from 31 December 2016 to 14 January 2017, and through to the end of the melt season, errors of commission are identified over the central and distal regions of the RBIS (e.g. see red panel in Fig. 6). Similar errors are identified within the maximum melt extent products (Fig. 7). This limitation has also been found in similar previous research (e.g. Halberstadt and others, 2020), with errors resulting from imperfect cloud-masking methods.

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

5. Conclusions

We have presented an 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. Although 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 research should look at improving cloud-masking approaches before applying the classifier, or developing 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 extents across all Antarctic ice shelves.

Finally, we applied the classifier to the RBIS for the 2013/14 to 2019/20 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. On average slush accounted for around two-thirds of the total meltwater extent. 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.

References

Alley KE, Scambos TA, Miller JZ, Long DG and MacFerrin M (2018) Quantifying vulnerability of Antarctic ice shelves to hydrofracture using microwave scattering properties. *Remote Sensing of Environment* **210**, 297–306. doi: 10.1016/j.rse.2018.03.025

Arthur JF, Stokes C, Jamieson SS, Carr JR and Leeson AA (2020a) Recent understanding of Antarctic supraglacial lakes using satellite remote sensing. *Progress in Physical Geography*, **44**(6), 837–869. doi: 10.1177/0309133320916114

Arthur J, Stokes C, Jamieson S, Carr JR and Leeson A (2020b) Distribution and seasonal evolution of supraglacial lakes on Shackleton Ice Shelf, East Antarctica. *The Cryosphere*, **14**(11) 4103–4120. doi: 10.5194/tc-14-4103-2020

Arthur D and Vassilvitskii S (2007) k-means++: the advantages of careful seeding. Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms, 07–09 January, pp. 1027–1035, Stanford.

Bamber JL and Aspinall WP (2013) An expert judgement assessment of future sea level rise from the ice sheets. *Nature Climate Change* **3**(4), 424–427. doi: 10.1038/nclimate1778

Banwell AF and 5 others (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. doi: 10.3189/2014AoG66A049

Banwell AF and 7 others (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. doi: 10.5194/tc-15-909-2021

Banwell AF and MacAyeal DR (2015) Ice-shelf fracture due to viscoelastic-flexure stress induced by fill/drain cycles of supraglacial lakes. *Antarctic Science* **27**(6), 587–597. doi: 10.1017/S0954102015000292

Banwell AF, MacAyeal DR and Sergienko OV (2013) Breakup of the Larsen B Ice Shelf triggered by chain reaction drainage of supraglacial lakes. *Geophysical Research Letters* **40**(22), 5872–5876. doi: 10.1002/2013GL057694

Banwell AF, Willis IC, Macdonald GJ, Goodsell B and MacAyeal DR (2019) Direct measurements of ice-shelf flexure caused by surface meltwater ponding and drainage. *Nature Communications* **10**(1), 1–10. doi: 10.1038/s41467-019-08522-5

Bell RE and 6 others (2017) Antarctic ice shelf potentially stabilized by export of meltwater in surface river. *Nature* **544**(7650), 344–348. doi: 10.1038/nature22048

Bell RE, Banwell AF, Trusel LD and Kingslake J (2018) Antarctic surface hydrology and impacts on ice-sheet mass balance. *Nature Climate Change* **8**(12), 1044–1052. doi: 10.1038/s41558-018-0326-3

Berthier E, Scambos TA and Shuman CA (2012) Mass loss of Larsen B tributary glaciers (Antarctic Peninsula) unabated since 2002. *Geophysical Research Letters* **39**, L13501. doi: 10.1029/2012GL051755

Bindschadler R and 8 others (2008) The Landsat image mosaic of Antarctica. *Remote Sensing of Environment* **112**(12), 4214–4226.

Breiman L (2001) Random forests. *Machine Learning* **45**, 5–32.

Dell R and 6 others (2020) Lateral meltwater transfer across an Antarctic ice shelf. *The Cryosphere* **14**(7), 2313–2330. doi: 10.5194/tc-14-2313-2020

Dirscherl M, Dietz AJ, Kneisel C and Kuenzer C (2020) Automated mapping of Antarctic supraglacial lakes using a machine learning approach. *Remote Sensing* **1**(7), 1203. doi: 10.3390/rs12071203

Dirscherl M, Dietz AJ, Kneisel C and Kuenzer C (2021) A novel method for automated supraglacial lake mapping in Antarctica using Sentinel-1 SAR imagery and deep learning. *Remote Sensing* **13**(2), 1–27. doi: 10.3390/rs13020197

Dunmire D and 6 others (2020) Observations of buried lake drainage on the Antarctic ice sheet. *Geophysical Research Letters* **47**(15), e2020GL087970. doi: 10.1029/2020gl087970

Frezzotti M (1993) Glaciological study in terra Nova Bay, Antarctica, inferred from remote sensing analysis. *Annals of Glaciology* **17**, 63–71. doi: 10.3189/s0260305500012623

Fricker HA and 10 others (2021) ICESat-2 Meltwater Depth Estimates: Application to Surface Melt on Amery Ice Shelf, East Antarctica. *Geophysical Research Letters*, **48**(8), e2020GL090550. doi: 10.1029/2020gl090550

Fürst JJ and 6 others (2016) The safety band of Antarctic ice shelves. *Nature Climate Change* **6**(5), 479–482. doi: 10.1038/nclimate2912

Halberstadt ARW and 5 others (2020) Antarctic supraglacial lake identification using Landsat-8 image classification. *Remote Sensing* **12**(8), 1–29. doi: 10.3390/RS12081327

Hubbard B and 6 others (2016) Massive subsurface ice formed by refreezing of ice-shelf melt ponds. *Nature communications* **7**(1), 1–6.

IPCC (2019) *Special Report on the Ocean and Cryosphere in a Changing Climate – Technical Summary* (Final Draft). Retrieved from https://www.ipcc.ch/srocc/.

Kingslake J, Ely JC, Das I and Bell RE (2017) Widespread movement of meltwater onto and across Antarctic ice shelves. *Nature* **544**(7650), 349–352. doi: 10.1038/nature22049

Kuipers Munneke P, Ligtenberg SRM, Van Den Broeke MR and Vaughan DG (2014) Firn air depletion as a precursor of Antarctic ice-shelf collapse. *Journal of Glaciology* **60**(220), 205–214. doi: 10.3189/2014JoG13J183

Lai CY and 6 others (2020) Vulnerability of Antarctica's ice shelves to meltwater-driven fracture. *Nature* **584**(7822), 574–578. doi: 10.1038/s41586-020-2627-8

Langley ES, Leeson AA, Stokes CR and Jamieson SSR (2016) Seasonal evolution of supraglacial lakes on an East Antarctic outlet glacier. *Geophysical Research Letters* **43**(16), 8563–8571. doi: 10.1002/2016GL069511

Lenaerts J and 6 others (2017) Meltwater produced by wind–albedo interaction stored in an East Antarctic ice shelf. *Nature Climate Change* **7**, 58–63. doi: 10.1038/NCLIMATE3180

Macdonald GJ and 5 others (2019) Formation of pedestalled, relict lakes on the McMurdo Ice Shelf, Antarctica. *Journal of Glaciology* **65**(250), 337–343. doi: 10.1017/jog.2019.17

Montgomery L and 6 others (2020) Hydrologic properties of a highly permeable firn aquifer in the Wilkins Ice Shelf, Antarctica. *Geophysical Research Letters* **47**(22), e2020GL089552. doi: 10.1029/2020GL089552

Moussavi M and 5 others (2020) Antarctic supraglacial lake detection using Landsat 8 and Sentinel-2 imagery: towards continental generation of lake volumes. *Remote Sensing* **12**(1), 134. doi: 10.3390/rs12010134

Reynolds JM (1981) Lakes on George VI Ice Shelf, Antarctica. *Polar Record* **20** (128), 425–432. doi: 10.1017/S0032247400003636

Rignot E and 5 others (2004) Accelerated ice discharge from the Antarctic Peninsula following the collapse of Larsen B Ice Shelf. *Geophysical Research Letters* **31**(18), L18401. doi: 10.1029/2004GL020697

Robel AA and Banwell AF (2019) A speed limit on ice shelf collapse through hydrofracture. *Geophysical Research Letters* **46**(21), 12092–12100. doi: 10.1029/2019GL084397

Scambos TA, Bohlander JA, Shuman CA and Skvarca P (2004) Glacier acceleration and thinning after ice shelf collapse in the Larsen B embayment, Antarctica. *Geophysical Research Letters* **31**(18), L18402. doi: 10.1029/2004GL020670

Scambos T, Hulbe C and Fahnestock M (2003) Climate-induced ice shelf disintegration in the Antarctic Peninsula. In *Antarctica Peninsula Climate Variability: A Historical and Paleo-Environmental Perspective*, pp. 79–92. doi: 10.1029/AR079p0079

Spergel JJ, Kingslake J, Creyts T, van Wessem M and Fricker HA (2021) Surface meltwater drainage and ponding on Amery Ice Shelf, East Antarctica, 1973–2019. *Journal of Glaciology*, 1–14.

Stehman SV (1997) Selecting and interpreting measures of thematic classification accuracy. *Remote sensing of Environment* **62**(1), 77–89.

Stokes CR, Sanderson JE, Miles BWJ, Jamieson SSR and Leeson AA (2019) Widespread distribution of supraglacial lakes around the margin of the East Antarctic ice sheet. *Scientific Reports* **9**(1), 1–14. doi: 10.1038/s41598-019-50343-5

Trusel LD and 6 others (2015) Divergent trajectories of Antarctic surface melt under two twenty-first-century climate scenarios. *Nature Geoscience* **8**, 927–932. doi: 10.1038/ngeo2563

Vihma T, Tuovinen E and Savijärvi H (2011) Interaction of katabatic winds and near-surface temperatures in the Antarctic. *Journal of Geophysical Research: Atmospheres* **116**. doi:10.1029/2010JD014917

Wagner AC (1972) Flooding of the ice shelf in George VI sound. *Antarctic Survey Bulletin* **28**, 71–74.

Williamson AG, Banwell AF, Willis IC and Arnold NS (2018) Dual-satellite (Sentinel-2 and Landsat 8) remote sensing of supraglacial lakes in Greenland. *The Cryosphere* **12**(9), 3045–3065. doi: 10.5194/tc-12-3045-2018

Work EA and Gilmer DS (1976) Utilization of satellite data for inventorying prairie ponds and lakes. *Photogrammetric Engineering and Remote Sensing* **42**(5), 685–694.

Yang K and Smith LC (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. doi: 10.1109/LGRS.2012.2224316

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