remote sensing

MDPI

Supplementary Materials

Table S1. Landsat product ID and sun elevation for each training and application scene. Training scenes are denoted by red and application scenes are shown in black. Date of image collection is included in the product ID naming convention. Scenes are separated by location; (a) from Amery Ice Shelf and (b) from Roi Baudouin Ice Shelf. (c) Scenes used for manual polygon and pixel validation datasets.

A. Amery Ice Shelf		**B. Roi Baudouin Ice Shelf**	
Sun Elevation (°)	Landsat Product ID	Sun Elevation (°)	Landsat Product ID
5.3	LC08_L1GT_128111_20150403	10.3	LC08_L1GT_155110_20180323
5.3	LC08_L1GT_128111_20150403	11.3	LC08_L1GT_155110_20170320
5.5	LC08_L1GT_126111_20140909	11.9	LC08_L1GT_153109_20170322
5.6	LC08_L1GT_126111_20140402	14.0	LC08_L1GT_154109_20180924
7.4	LC08_L1GT_129111_20140914	15.1	LC08_L1GT_154109_20170313
10.5	LC08_L1GT_126112_20140925	16.5	LC08_L1GT_153109_20170930
11.8	LC08_L1GT_126111_20140925	18.6	LC08_L1GT_153109_20160303
12.3	LC08_L1GT_129111_20170314	20.8	LC08_L1GT_154109_20170225
12.9	LC08_L1GT_126112_20170309	23.2	LC08_L1GT_154109_20151018
13.1	LC08_L1GT_126111_20180312	23.5	LC08_L1GT_155110_20160214
14.8	LC08_L1GT_129111_20151003	24.9	LC08_L1GT_153109_20150213
17.3	LC08_L1GT_127111_20170228	26.1	LC08_L1GT_154109_20181026
17.9	LC08_L1GT_128111_20140227	27.7	LC08_L1GT_154109_20150204
19.5	LC08_L1GT_128111_20180222	29.3	LC08_L1GT_154110_20170124
21.4	LC08_L1GT_128111_20160217	31.2	LC08_L1GT_154109_20160122
22.5	LC08_L1GT_129111_20180213	31.3	LC08_L1GT_154110_20140116
23.2	LC08_L1GT_128111_20140211	32.2	LC08_L1GT_154110_20180111
24.7	LC08_L1GT_128111_20180206	32.5	LC08_L1GT_154109_20140116
25.5	LC08_L1GT_128111_20170203	32.7	LC08_L1GT_154109_20151119
26.3	LC08_L1GT_128111_20160201	33.3	LC08_L1GT_154109_20180111
27.0	LC08_L1GT_128111_20150129	33.4	LC08_L1GT_154110_20150103
27.3	LC08_L1GT_126112_20180123	33.5	LC08_L1GT_155110_20180102
27.4	LC08_L1GT_128111_20131107	34.3	LC08_L1GT_155110_20141225

27.4	LC08_L1GT_127111_20170127
27.6	LC08_L1GT_127112_20150122
27.8	LC08_L1GT_128112_20180121
27.8	LC08_L1GT_128111_20140126
28.1	LC08_L1GT_129112_20150120
28.8	LC08_L1GT_128111_20151113
29.0	LC08_L1GT_128111_20180121
29.4	LC08_L1GT_128111_20161115
29.6	LC08_L1GT_128111_20170118
30.0	LC08_L1GT_128111_20171118
30.8	LC08_L1GT_128112_20180105
31.1	LC08_L1GT_128111_20131123
31.2	LC08_L1GT_126112_20160102
31.3	LC08_L1GT_128111_20140110
32.3	LC08_L1GT_128111_20161201
32.3	LC08_L1GT_128111_20170102
32.9	LC08_L1GT_128111_20141228
33.0	LC08_L1GT_127111_20161226
33.0	LC08_L1GT_126111_20131227
33.1	LC08_L1GT_129111_20161224
33.1	LC08_L1GT_128111_20131209
33.2	LC08_L1GT_128111_20151215
33.2	LC08_L1GT_128111_20171220
34.8	LC08_L1GT_153109_20170101
35.1	LC08_L1GT_153109_20151230
35.4	LC08_L1GT_154109_20171226
C. Manual polygon and pixel validation scenes (Amery and Roi Baudouin(*) ice shelves)	
20.8	LC08_L1GT_154109_20170225 *
25.4	LC08_L1GT_127111_20140204
32.4	LC08_L1GT_128111_20170102
32.5	LC08_L1GT_154109_20140116 *
33.0	LC08_L1GT_127111_20161226
33.2	LC08_L1GT_127111_20141221

Section SI.1. Supervised Classification Algorithms

We assess the sensitivity of classified lake areas to classification algorithm using resubstitution accuracy and our pixel-level validation dataset (Table S2 and Table S3, respectively). Resubstitution accuracy was computed by blindly applying the trained classifier to the original training dataset and calculating classification accuracy compared to a known input (i.e., the training data). This analysis allowed us to evaluate the sensitivity of our training data and ensure that classes are spectrally distinct (Table S2).

In Figure S1, we show the best-performing algorithms (those with validation accuracy greater than 91% for training dataset t11AB; Table S3), all generated from the same training dataset and applied across Amery Ice Shelf. The Random Forest (RF) and Classification and Regression Trees

(CART) classifiers generally produce similar lake areas and have the highest resubstitution accuracies (Table S2); the Minimum Distance (MDM) classifier has lower validation and resubstitution accuracies, and can produce significantly higher lake areas through misclassification of some application scenes. The Support Vector Machine (SVM) classifier has the same validation accuracy as RF; we note that SVM is slightly better than RF at reducing cloud shadow commission and correctly identifying lakes under cloud shadows and in low-sun-elevation scenes. However, SVM produces slightly lower resubstitution accuracy, and misses some shallow lake areas compared to RF.

We therefore selected the RF algorithm (with 50 Rifle decision trees) to generate our trained supervised classifiers shown in this work based on its high resubstitution accuracy (Table S2), high accuracy based on our pixel lake validation dataset (Table S3), and consistently accurate visual lake identification (for example, across the set of Amery Ice Shelf application scenes, Figure S1).

Table S2. Resubstitution accuracy for supervised classifiers generated from the 11-class training dataset (t11) using different classification algorithms. Default Google Earth Engine algorithm settings were used if not specified otherwise. Percentages are reported as total (11-class) resubstitution accuracy / lake versus non-lake resubstitution accuracy (in bold).

Supervised classification algorithm		Resubstitution accuracy (t11AB)
Classification and Regression Trees [1]		88.9% / **98.8%**
Random Forest [2]	50 Rifle decision trees per class	99.3% / **99.9%**
	20 Rifle decision trees per class	98.4% / **99.8%**
	1 Rifle decision trees per class	87.0% / **98.0%**
Minimum Distance	Cosine distance from the class mean	46.8% / **91.8%**
	Mahalanobis distance from class mean	74.4% / **96.7%**
	Euclidean distance from class mean	19.3% / **87.6%**
Naive Bayes [3]	default	61.5% / **94.5%**
	continuous	62.6% / **94.9%**
Maximum Entropy		53.2% / **93.8%**
Support Vector Machine [4]		70.1% / **96.3%**

Table S3. Validation accuracy assessed using the manual pixel lake dataset for every combination of training dataset and classification algorithm: Classification and Regression Trees (CART); Random Forest with 50, 20, and 1 Rifle decision trees ('RF 50', 'RF 20', 'RF 1'); Minimum Distance with cosine, mahalanobis, and Euclidean distance metrics ('MDC', 'MDM', 'MDE'); Naive Bayes ('Bayes' and continuous 'Bayes C); Maximum Entropy ('Max Ent'); and Support Vector Machine (SVM).

Amery Ice Shelf validation pixels											
	t6A	**t7A**	**t9A**	**t11A**	**tOB11A**	**t11AB**	**t6B**	**t7B**	**t9B**	**t11B**	**tOB11B**
CART	91.4	90.3	89.0	89.3	84.0	91.4	88.3	90.1	72.6	72.6	80.9
RF 50	91.4	91.6	93.3	92.0	89.4	94.9	88.6	92.4	78.4	79.3	78.4
RF 20	90.9	91.8	93.0	91.4	89.6	94.6	88.4	91.9	78.4	80.6	89.6
RF 1	88.5	90.1	87.0	87.9	80.0	88.1	85.5	92.3	76.9	82.1	80.0

MDC	90.3	89.6	90.1	90.3	57.4	90.4	89.4	92.4	91.8	91.8	58.3
MDM	90.8	88.9	93.8	93.9	80.0	93.8	80.9	87.3	94.0	87.9	77.4
MDE	66.1	68.1	63.1	63.4	57.4	59.1	57.8	67.0	67.0	67.0	57.4
Bayes	91.8	91.8	92.1	91.4	57.4	90.1	91.9	93.8	93.6	84.9	57.4
Bayes C	90.8	91.0	89.8	88.9	74.8	88.5	88.5	89.5	67.4	67.4	59.9
Max Ent	90.6	89.1	89.9	89.8	42.6	90.8	89.1	92.1	90.0	90.4	43.4
SVM	90.8	89.5	91.9	92.0	54.0	94.9	87.0	94.4	83.0	83.0	71.4
Roi Baudouin Ice Shelf validation pixels											
CART	94.3	77.8	74.5	76.0	65.3	85.3	86.8	91.0	78.5	90.3	74.5
RF 50	95.5	80.3	79.5	76.8	68.8	89.5	97.0	97.3	98.3	97.3	85.8
RF 20	95.5	80.5	79.0	78.5	69.5	88.5	96.8	96.8	97.5	95.8	69.5
RF 1	81.5	78.8	77.3	76.3	64.0	78.5	90.5	94.0	94.5	93.3	64.0
MDC	96.0	91.0	90.1	85.8	51.3	87.0	95.0	89.8	88.3	88.3	53.3
MDM	95.5	85.5	85.8	86.5	71.3	92.3	93.0	91.5	91.8	91.3	64.8
MDE	64.0	51.8	50.3	50.3	51.0	51.0	51.8	65.3	65.3	65.3	51.0
Bayes	96.0	82.0	82.0	81.0	51.3	86.3	97.0	91.5	92.0	90.5	53.0
Bayes C	91.3	89.0	82.3	80.8	73.3	88.0	96.0	91.3	75.8	75.8	63.8
Max Ent	95.0	88.0	81.0	81.3	47.3	88.0	95.3	90.3	87.0	91.3	53.5
SVM	96.5	72.8	75.0	74.8	49.0	87.5	94.5	92.3	91.0	91.0	74.5

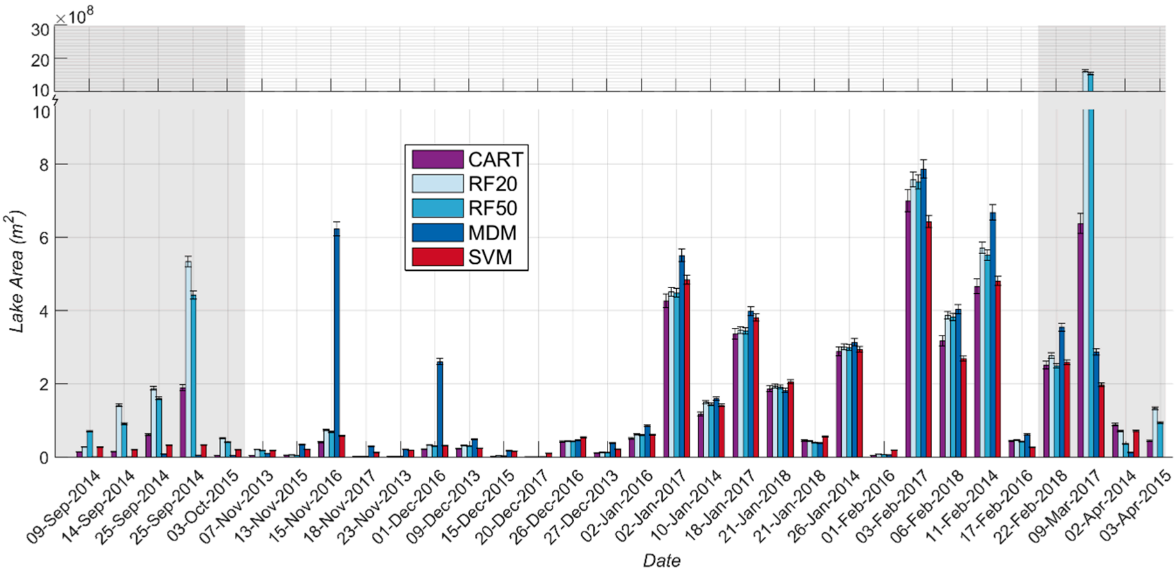

Figure S1. Lake areas identified across a set of application scenes across the Amery Ice Shelf, categorized by date of image acquisition. Classifiers were generated from the t11AB training dataset, using different high-performing classification algorithms. Manual cloud/cloud shadow polygons and the rock mask were applied prior to classification. Grey shading indicates scenes with sun elevation below 20°. Note the upper scale is compressed to show the exceedingly high (misclassified) lake areas for some low-sun-elevation scenes. Algorithm acronyms shown in legend: 'CART' is Classification and Regression Trees; 'RF50' and 'RF20' are Random Forest with 50 and 20 Rifle decision trees, respectively; 'MDM' is Minimum Distance with mahalanobis distance metric; and 'SVM' is Support Vector Machine. Error bars reflect the validation accuracy of each classifier (Table S3).

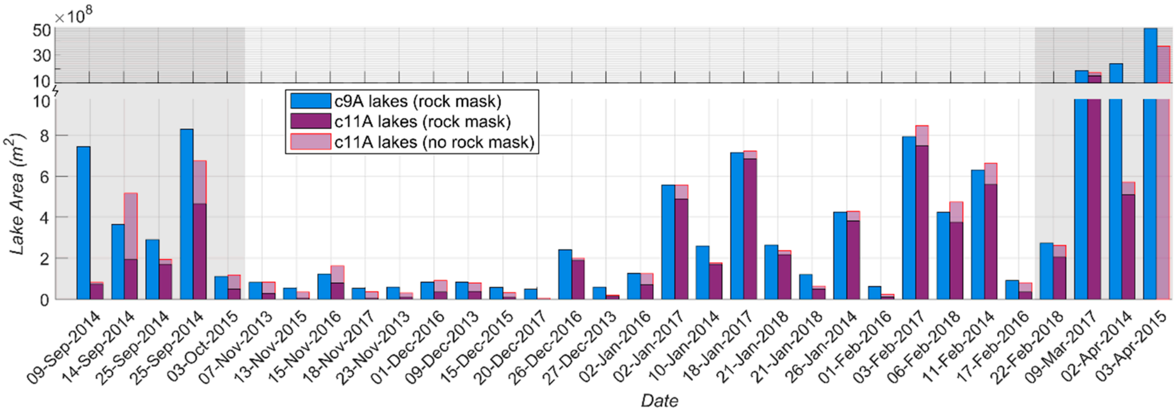

Figure S2: Supervised classifier results compared across the Amery Ice Shelf application scenes to investigate the effect of pre-classification rock masking versus including rock training classes. Lake areas were produced by c9A and c11A from rock-masked scenes with automated cloud removal; c11A was also applied to scenes without rock masking. Gray shading indicates scenes acquired under sun elevations < 20°. Note the upper scale is compressed to show the exceedingly high (misclassified) lake areas for some low-sun-elevation scenes.

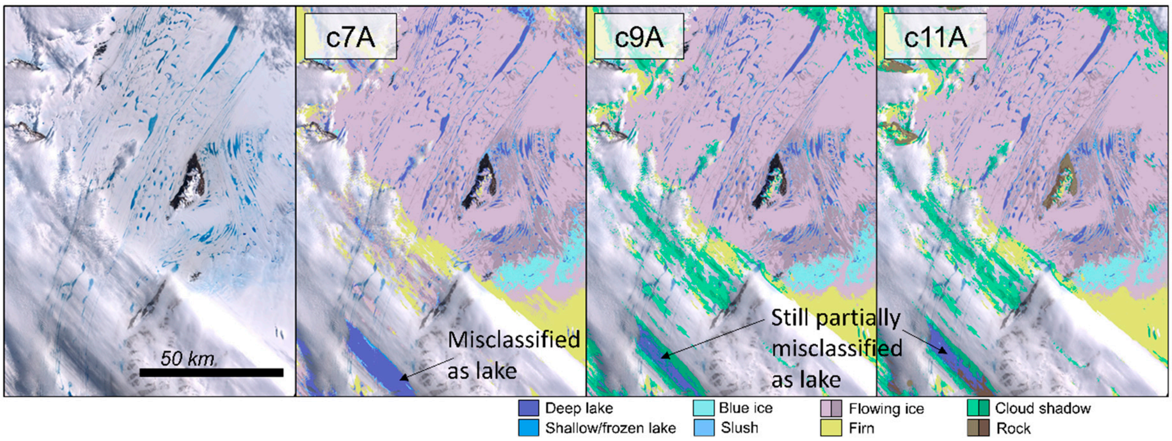

Figure S3. Supervised classifiers generated from different numbers of training classes are applied to Amery Ice Shelf scene LC08_L1GT_128111_20170118, following automated cloud removal.

References

1. Breiman, L.; Friedman, J.; Olshen, R.; Stone, C. Classification and Regression Trees: Wadsworth Int. Group: Wadsworth, OH, USA. **1984**, *37*, 237–251.
2. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32.
3. Huang, J.; Lu, J.; Ling, C.X. Comparing naive Bayes, decision trees, and SVM with AUC and accuracy. In Proceedings of the Third IEEE International Conference on Data Mining; IEEE, 2003; pp. 553–556.
4. Burges, C.J.C. A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.* **1998**, *2*, 121–167.