Automated mapping of Antarctic supraglacial lakes and streams using machine learning

Mariel Dirscherl[1,*], Andreas Dietz[1], Celia Baumhoer[1], Christof Kneisel[2], Claudia Kuenzer[1,2]

[1]German Aerospace Center (DLR), Earth Observation Center (EOC), German Remote Sensing Data Center (DFD)
[2]University Wuerzburg, Institute of Geography and Geology
[*]E-Mail: mariel.dirscherl@dlr.de

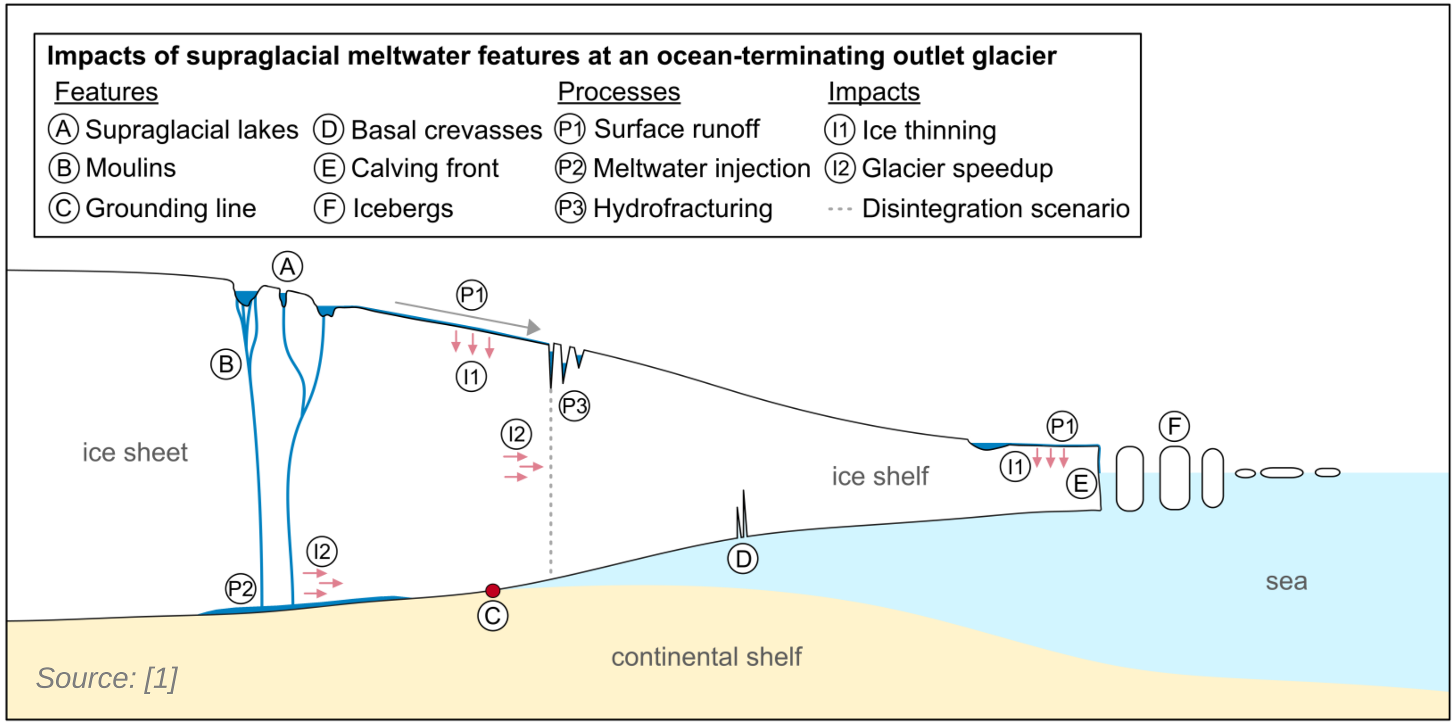
Link to Publication

DLR

EGU General Assembly 2020
Online | 4–8 May 2020

Why is the mapping of Antarctic supraglacial lakes important?

Supraglacial lakes may impact Antarctic ice dynamics through three main processes (P1-P3):



→ A circum-antarctic mapping of Antarctic supraglacial lakes is overdue and required to study these processes in more detail!

Study aim & overall workflow

Study aim:

Development of an **automated** supraglacial lake mapping method transferable in space and time using spaceborne Sentinel-2 data and state-of-the-art image processing.

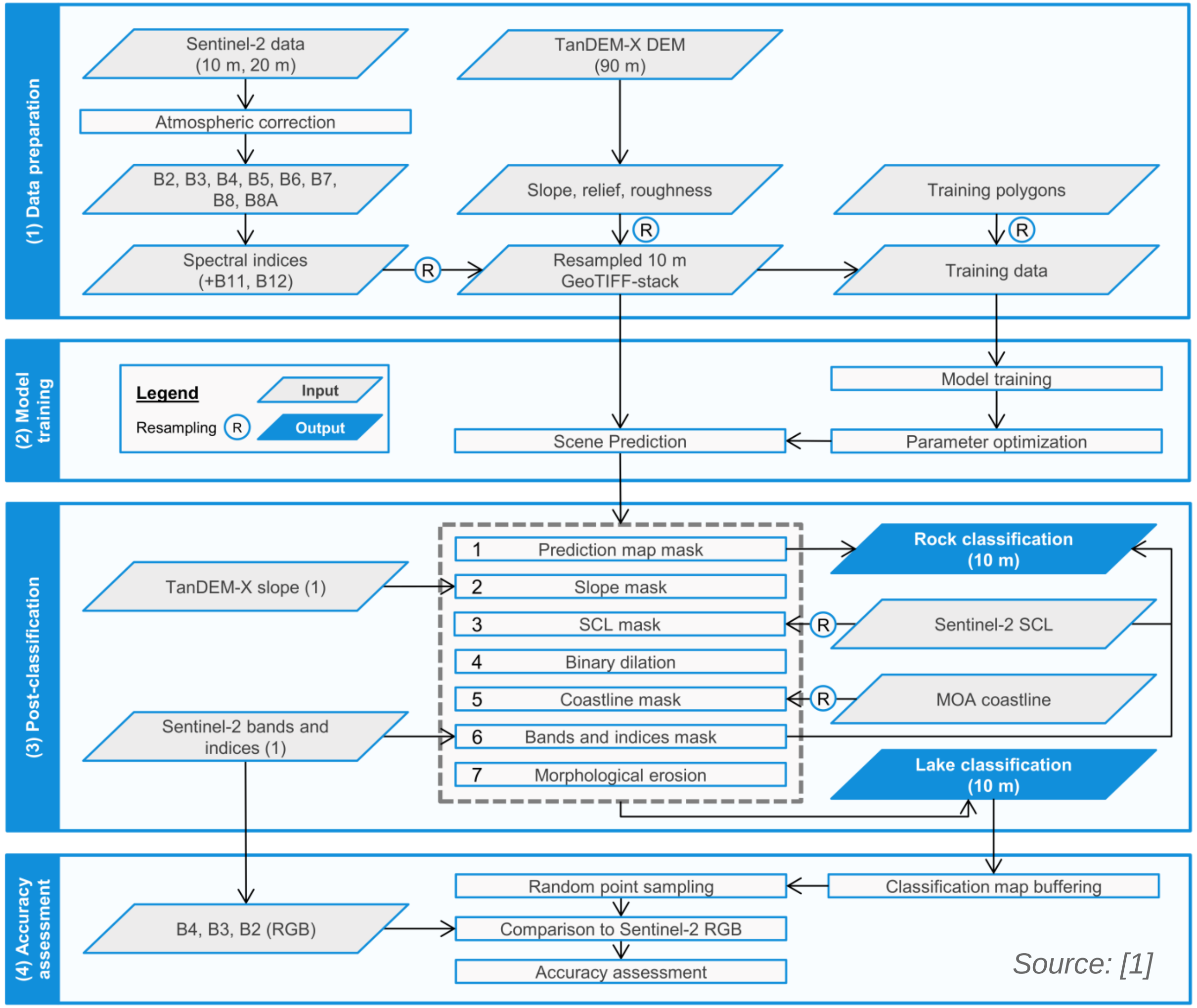
Study design:

Application of a supervised **Machine Learning** algorithm, namely Random Forest, trained on optical Sentinel-2 and auxiliary TanDEM-X topographic data.

Overall workflow:

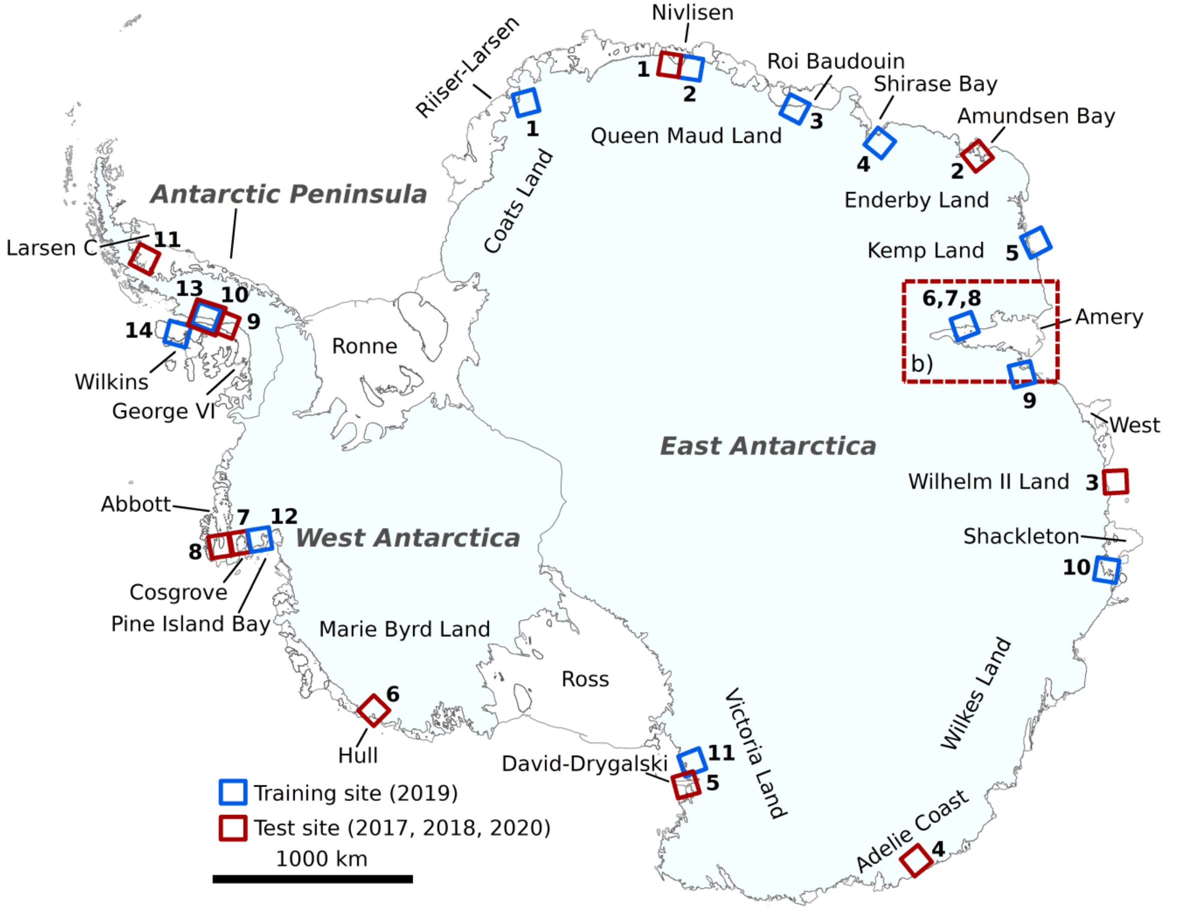
- Data preparation
- Model training and prediction
- Post-classification
- Accuracy assessment

Workflow for automated supraglacial lake mapping



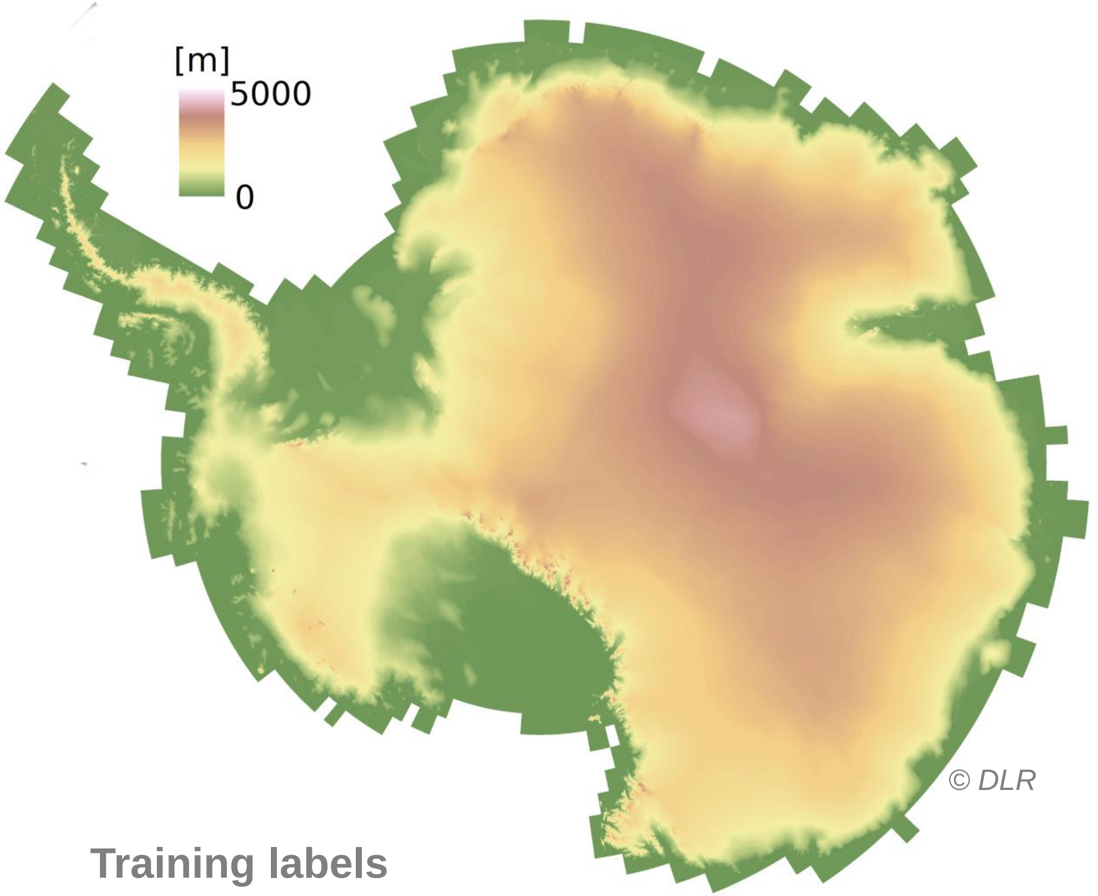
Data preparation: input data

Sentinel-2 training (2019) and test data (2017, 2018, 2020)



Modified after [1]

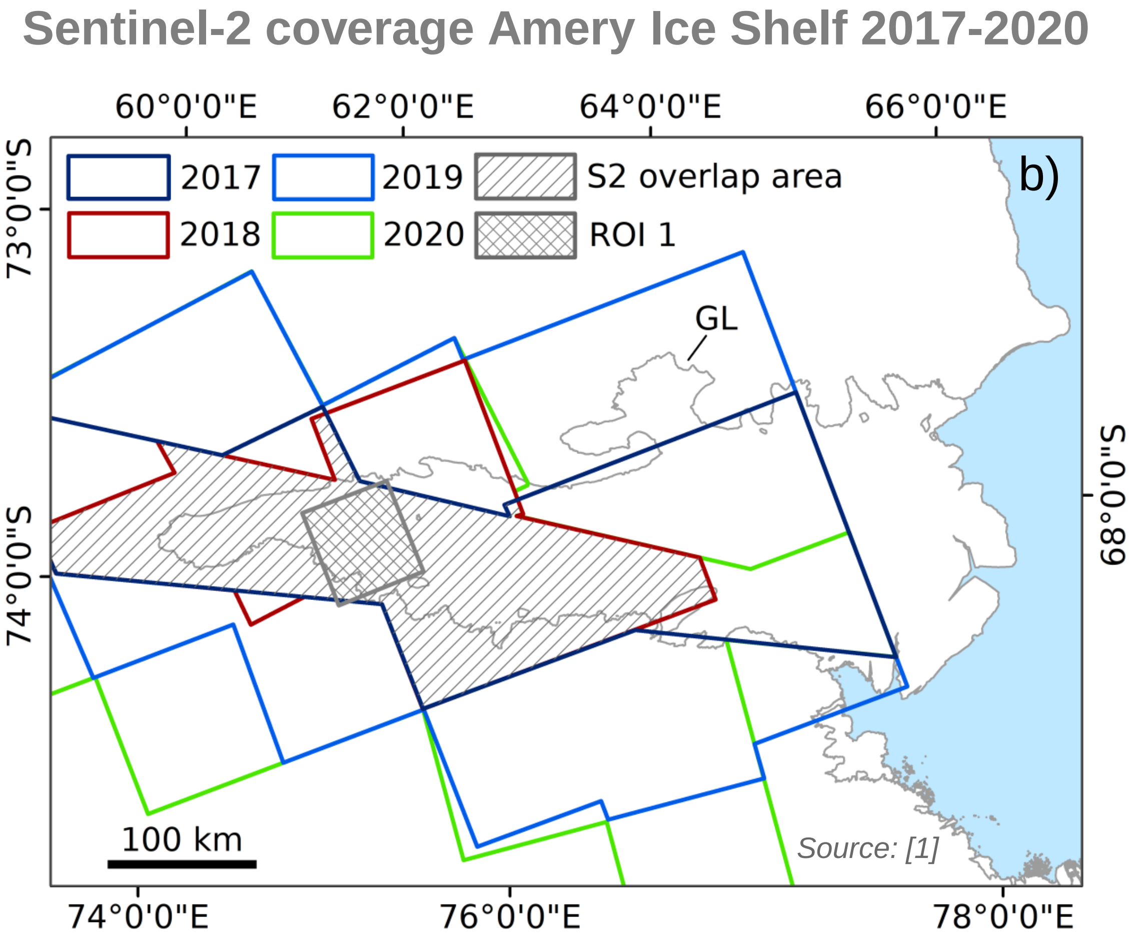
Edited 90-m Antarctic TanDEM-X DEM (2013-2014)



Training labels

- Created on basis of Sentinel-2 in GEE
- 4 main classes: “water”, “snow/ice”, “rock”, “shadow”

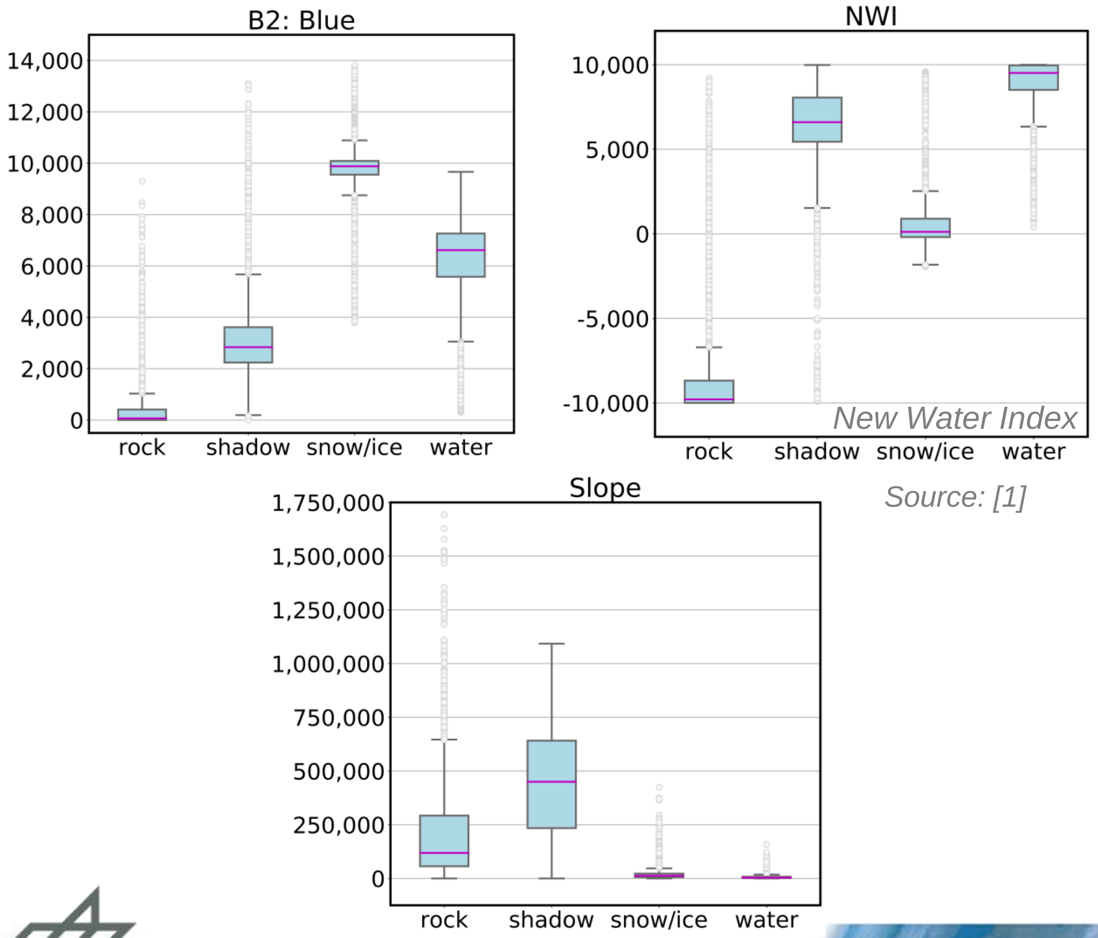
Data preparation: input data for application example



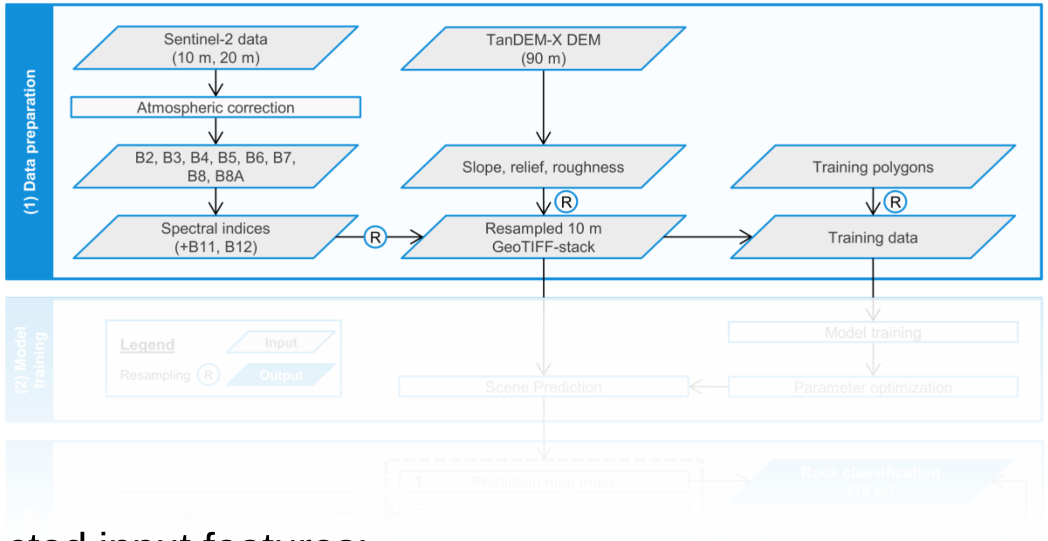
→ 84 additional acquisitions were selected for Amery Ice Shelf to test our algorithm for mapping of maximum lake extents over four consecutive melt seasons at full ice shelf coverage

Data preparation: variable selection

Input features were selected on basis of a discrimination analysis of the reflectance / topographic properties of water, snow/ice, rock and shadow on ice:



Source: [1]



Selected input features:

- **Sentinel-2**
  - **Bands**: 2-8A
  - **Indices**: e.g. $NDWI_1$, $NDWI_2$, SWI, SAVI, SI, NDGI, NDSI, NWI, $AWEI_{sh}$, $AWEI_{nsh}$, TC_{wet} [2-12]
- **TanDEM-X topographic variables**
  - DEM, slope, relief, roughness

Test data

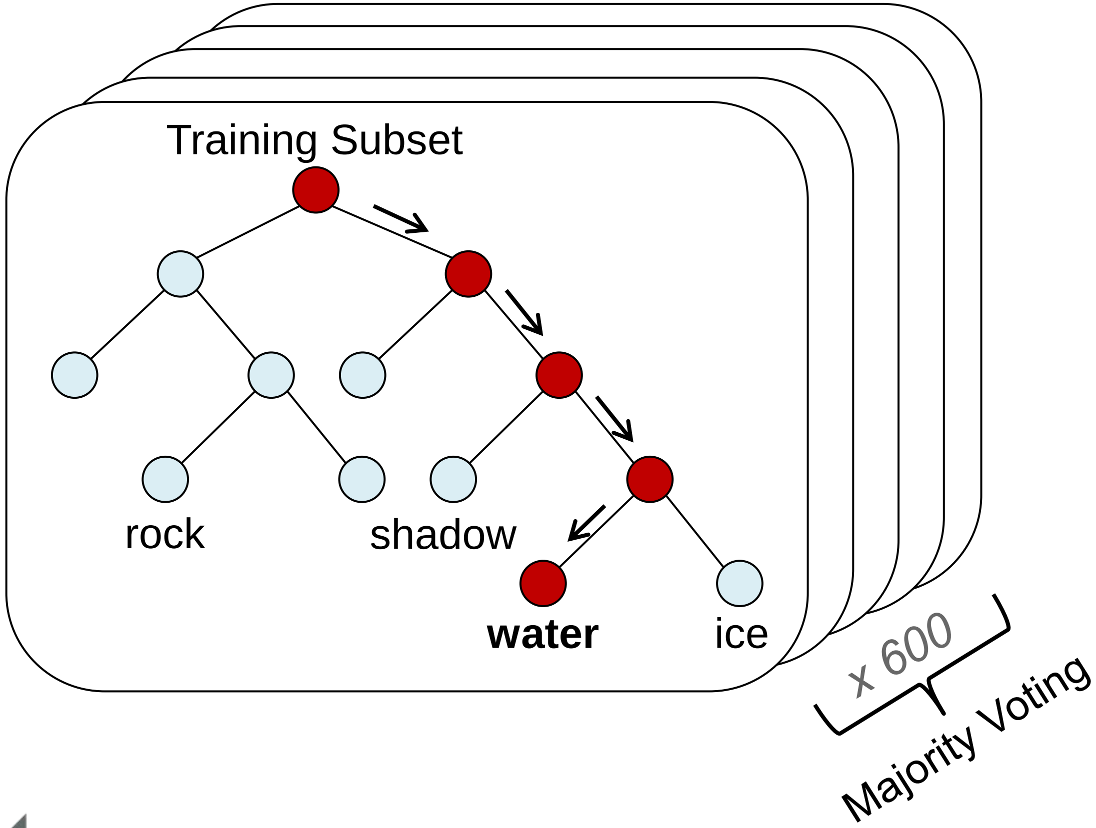
- **Training labels**

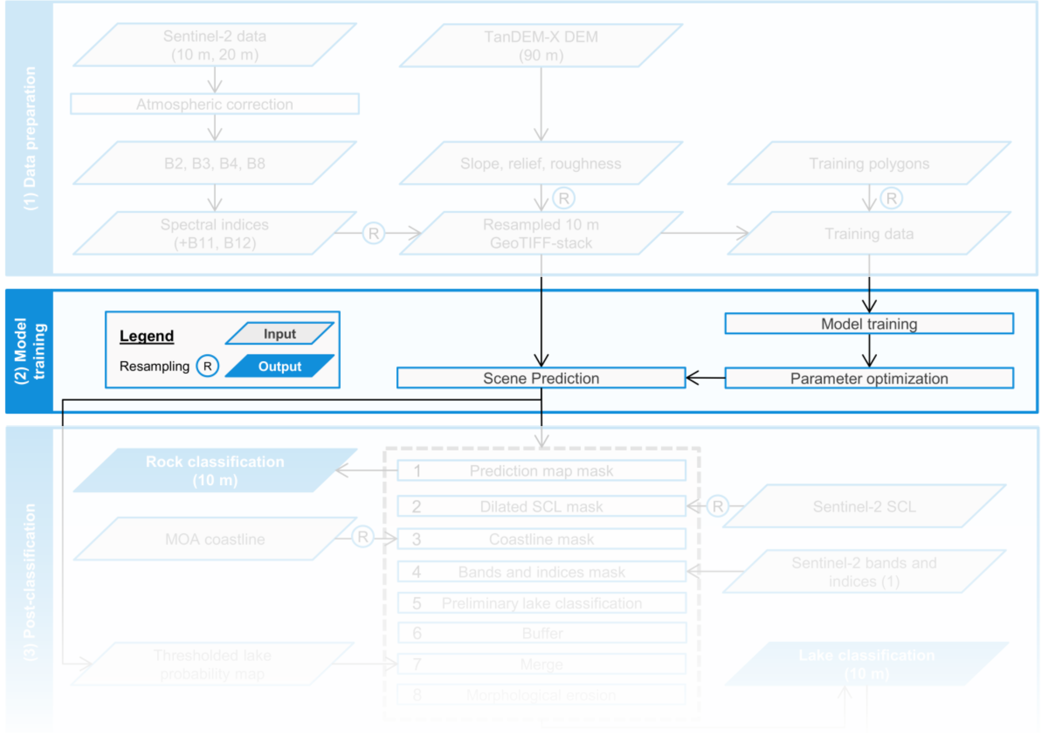
Training data

Model training and prediction

Random Forest was trained on subsets (~70%) of all 14 collocated training datasets using the Python programming language.

Background Random Forest:



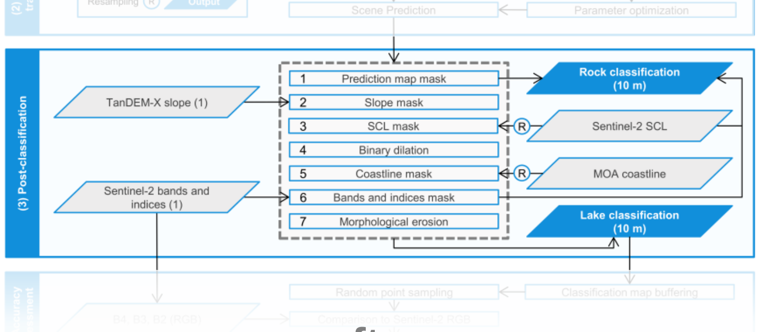


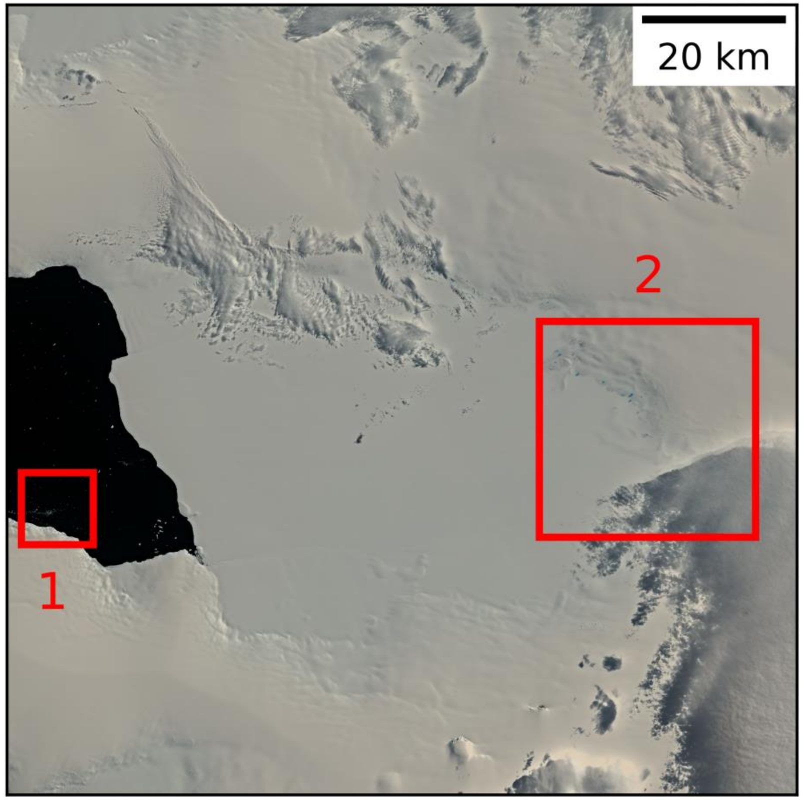
- Random Forest is characterized by an ensemble of uncorrelated decision trees, each built on the basis of a randomly sampled subset of training data (bagging)
- New unclassified data is predicted based upon the maximum votes of all independent decision trees

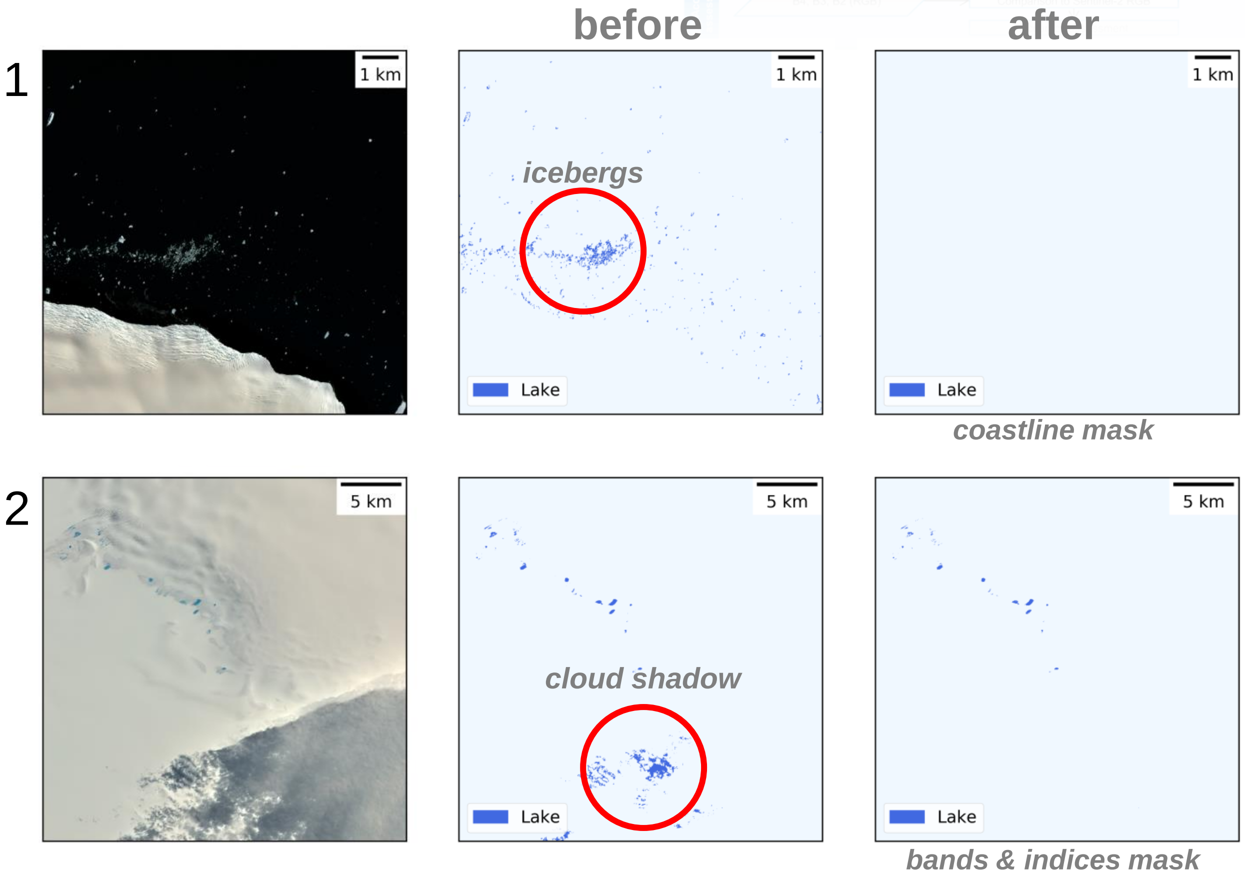
e.g. [13-15]

Post-classification: before vs. after

Cosgrove Ice Shelf, WAIS



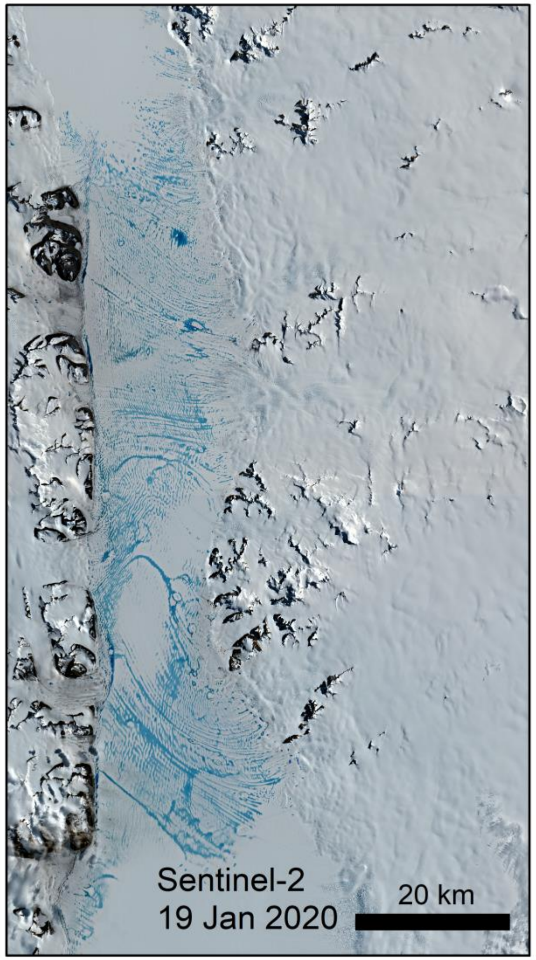


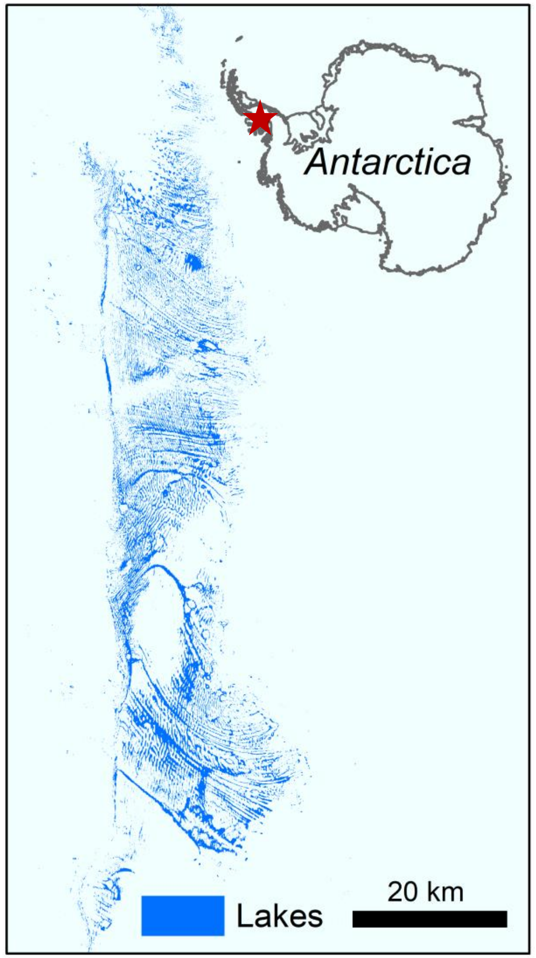


© Copernicus Sentinel-2 data, 12 January 2017
Modified after [1]

Automated mapping results

A) George VI Ice Shelf, Antarctic Peninsula

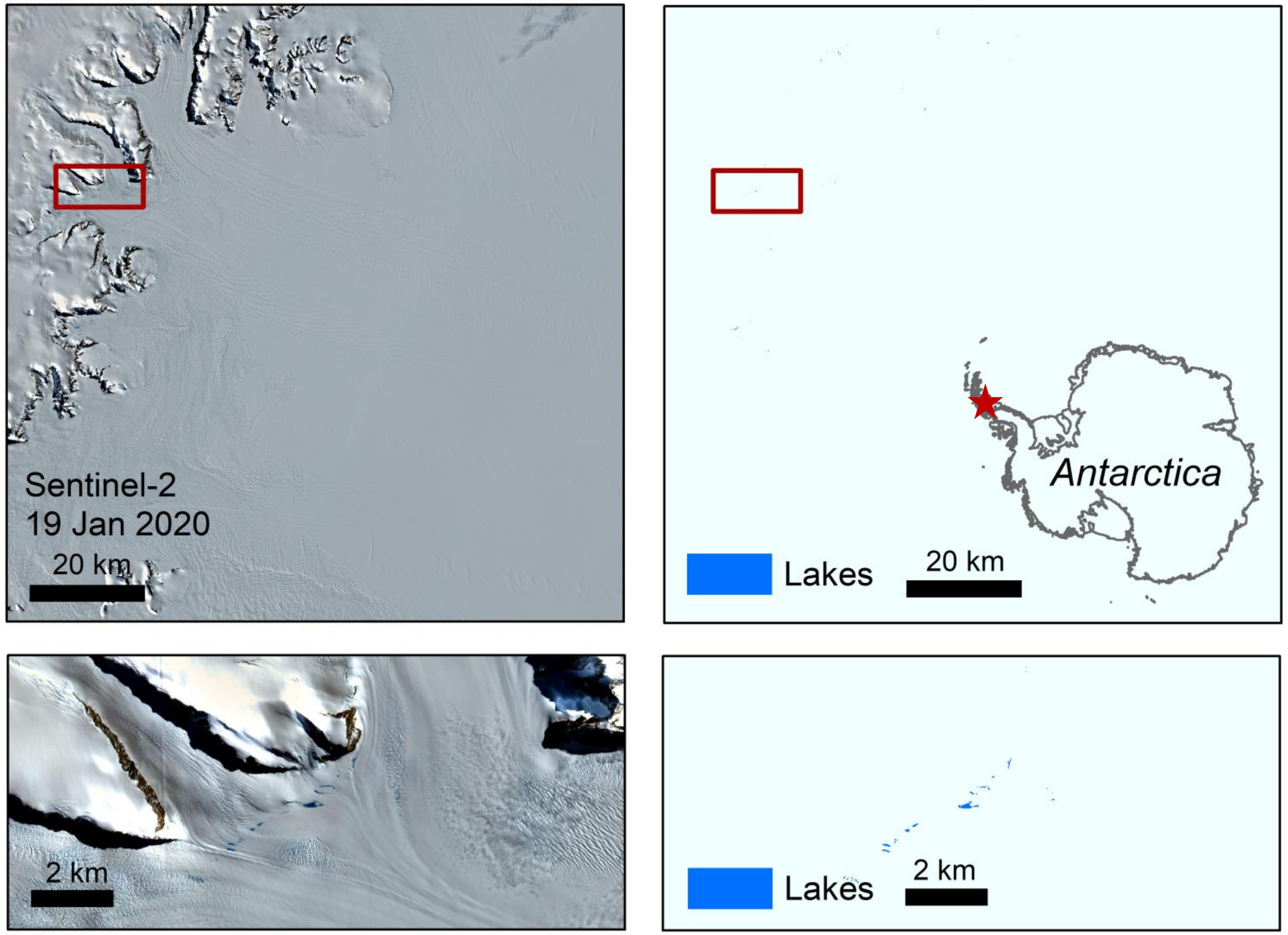




- Extensive supraglacial meltwater network visible on ice shelf on 19 January 2020
- **~831.7 km²** covered by supraglacial meltwater
- Very long meltwater channels and large surface ponds
- Artefacts (e.g. topographic shadow, shadow in crevasses) successfully masked

Automated mapping results

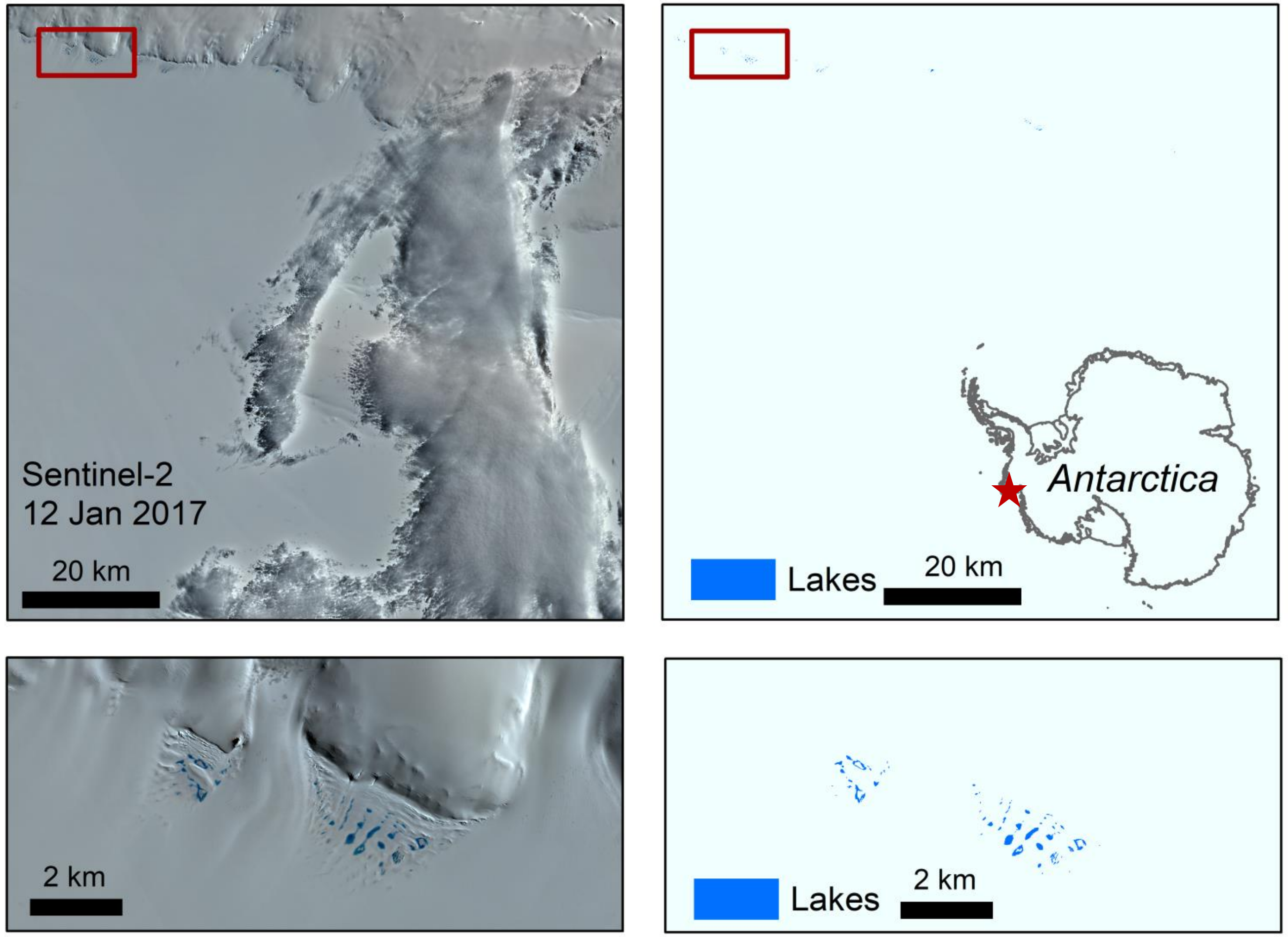
B) Larsen C tributaries, Antarctic Peninsula



- Widespread surface melt visible on 19 January 2020
- **~0.88 km²** covered by supraglacial lakes
- Mainly small melt ponds
- Artefacts (e.g. topographic shadow, shadow in crevasses) successfully masked

Automated mapping results

C) Abbott Ice Shelf, West Antarctica

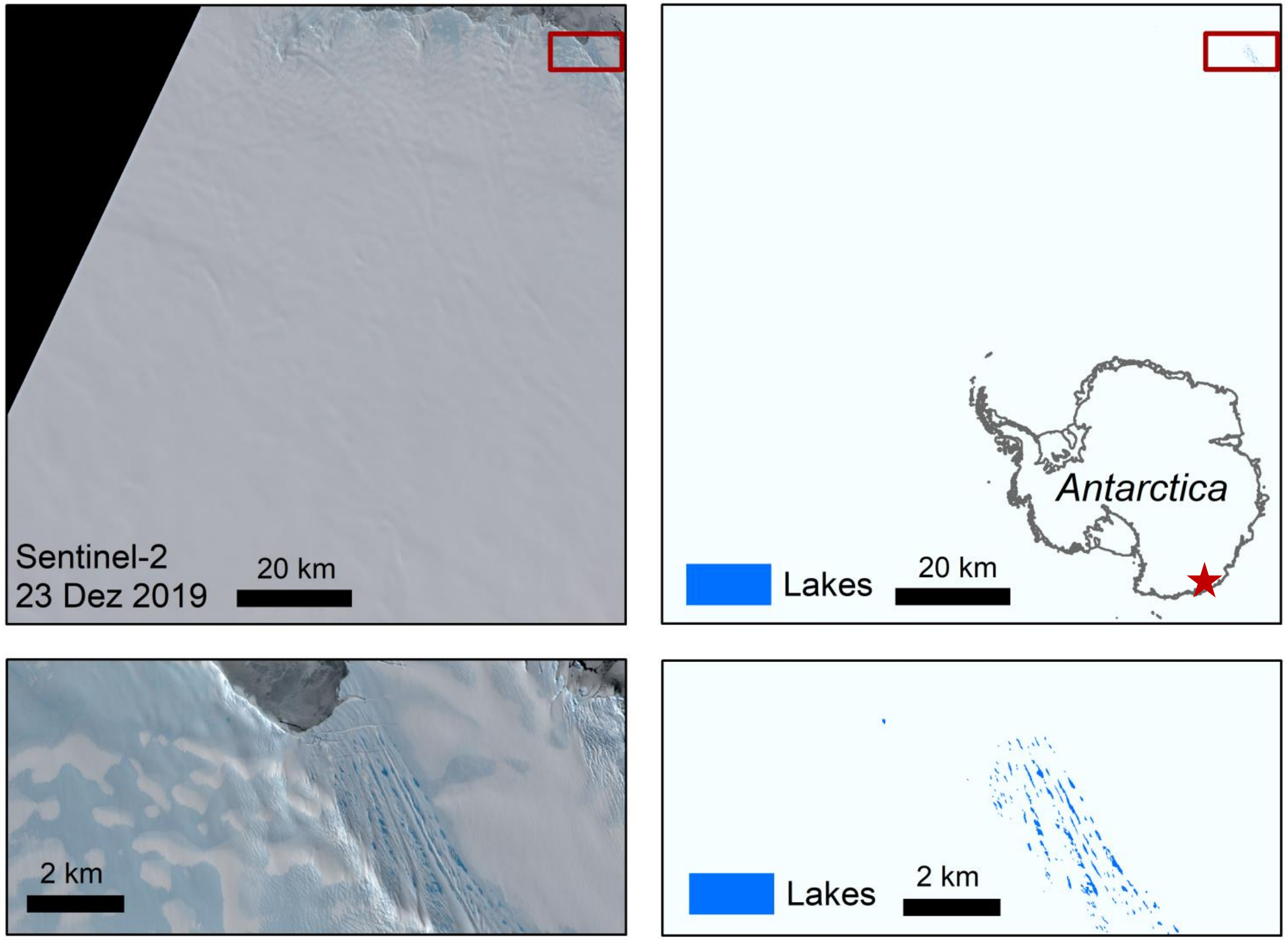


© Copernicus Sentinel-2 data

- Widespread surface melt visible near grounding line on 12 January 2017
- **~0.81 km²** covered by supraglacial lakes
- Mainly small melt ponds
- Artefacts (e.g. cloud shadow) successfully masked

Automated mapping results

D) Adelie Coast, East Antarctica

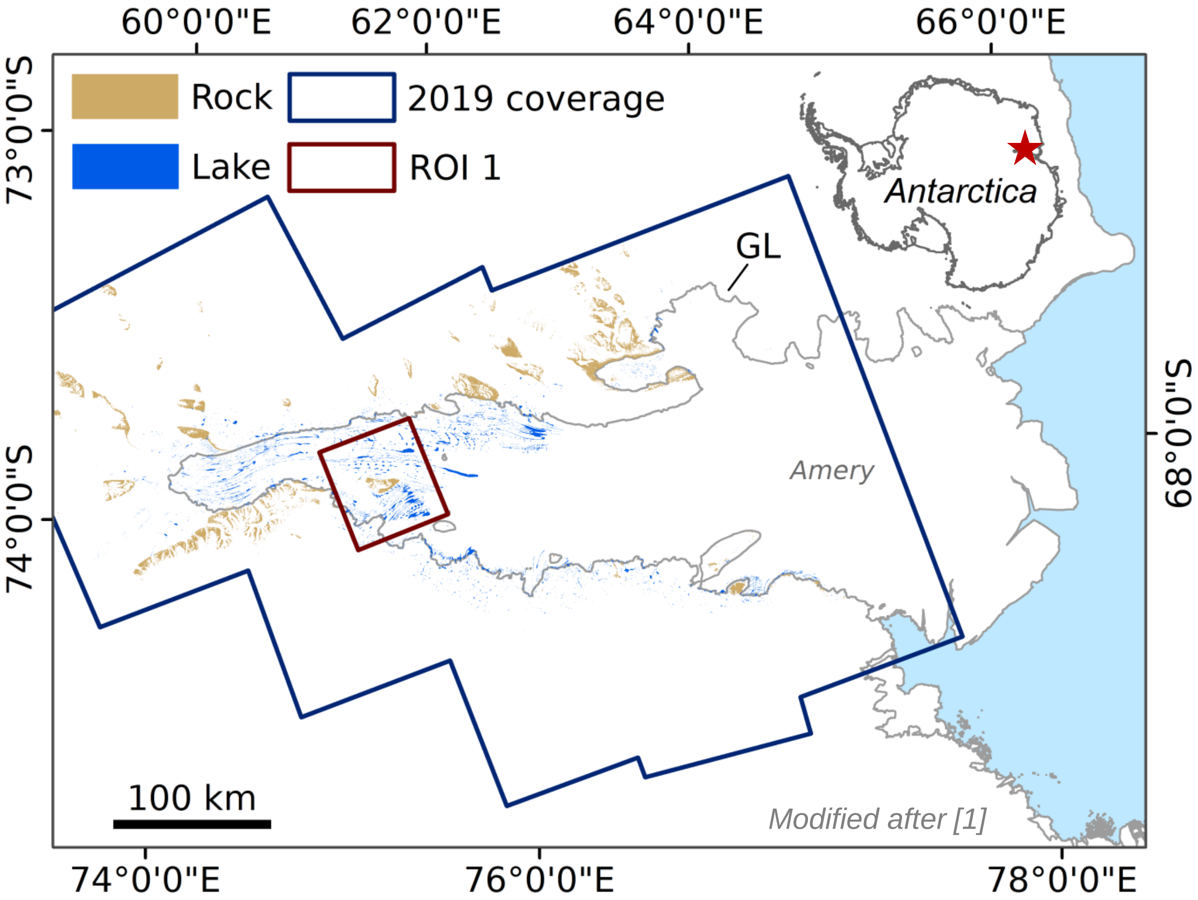


© Copernicus Sentinel-2 data

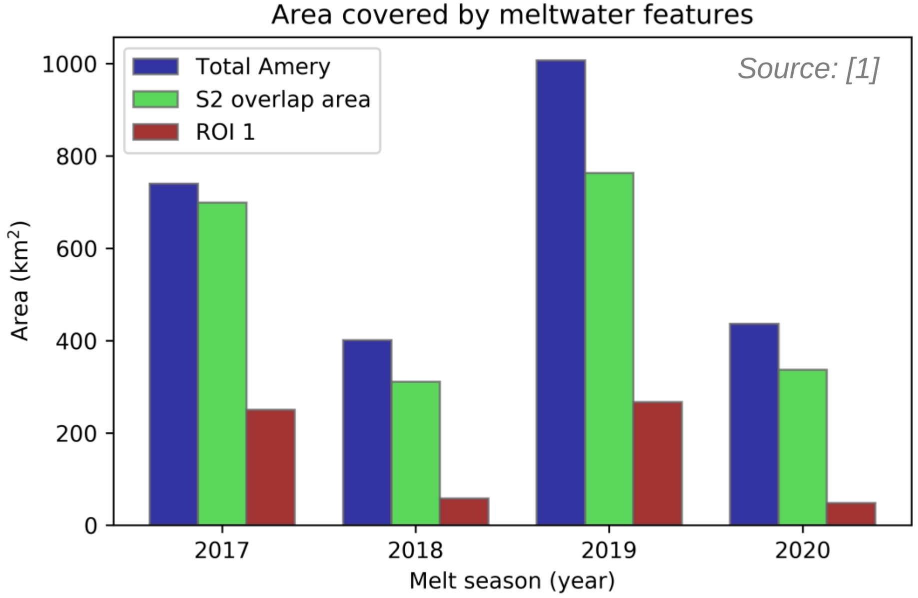
- Widespread surface melt visible on ice tongue on 23 December 2019
- **~0.58 km²** covered by supraglacial lakes
- Mainly small melt ponds
- Artefacts (e.g. blue ice, ocean, shadow in crevasses) successfully masked

Application example: spatio-temporal lake dynamics on Amery Ice Shelf

Maximum lake extent and rock outcrop map for 2019



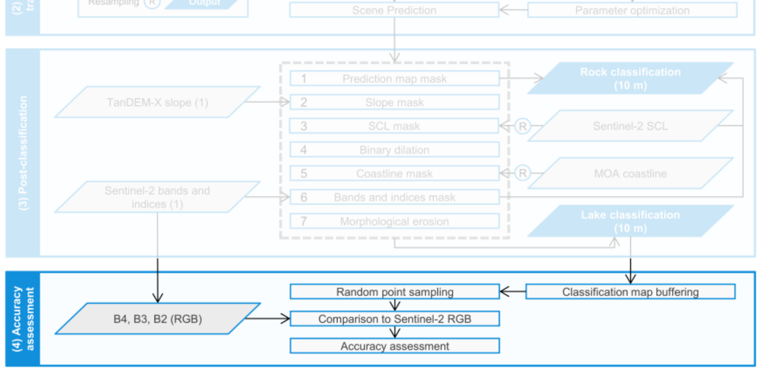
Maximum lake extent development over four years



Supraglacial lake occurrence within defined geographical units

Source: [1]

	Geographical Unit			
	Distance to Grounding Line ≤10 km	Distance to Coastline ≥300 km	Distance to Rock Outcrop ≤5 km	On Floating Ice Shelf
Year	Number of Supraglacial Lakes [%]			
2017	48	61	24	87
2018	74	59	58	72
2019	53	52	35	81
2020	74	67	56	71



Accuracy Assessment

The mapping results were evaluated by means of a confusion matrix which allowed deriving common statistical accuracy metrics including Recall (R) and Precision (P), F-score (F_1), Errors of Commission (EC) and Omission (EO) as well as Cohen's Kappa (κ) (e.g., [16-19]).

- Overall Kappa both classes: **0.883**
- Average F_1 water class: **88.62 %**
- Increased false positives (EC, P) mainly due to shadow on ice below clouds (e.g. Hull Glacier)
- Increased false negative pixels (EO, R) mainly at lake edges and due to the TanDEM DEM being from a different time step (e.g. Amundsen Bay, Wilhelm II Coast)

Classes	Water					Non-water					Both
Accuracy Metrics	EO [%]	EC [%]	R [%]	P [%]	F_1 [%]	EO [%]	EC [%]	R [%]	P [%]	F_1 [%]	K
Nivlisen Ice Shelf	3.16	8.00	96.84	92.00	94.36	0.88	0.33	99.12	99.67	99.39	0.938
Amundsen Bay	20.37	6.52	79.63	93.48	86.00	0.15	0.56	99.85	99.44	99.64	0.856
Wilhelm II Coast	17.14	9.38	82.86	90.63	86.57	0.15	0.30	99.85	99.70	99.77	0.863
Adelie Coast	0.00	5.56	100.00	94.44	97.14	0.10	0.00	99.90	100.00	99.95	0.971
Drygalski Ice Tongue	16.67	0.00	83.33	100.00	90.91	0.00	0.36	100.00	99.64	99.82	0.907
Hull Glacier	35.29	50.00	64.71	50.00	56.41	0.55	0.30	99.45	99.70	99.57	0.560
Abbott Ice Shelf	30.99	0.00	69.01	100.00	81.67	0.00	1.13	100.00	98.87	99.43	0.811
Cosgrove Ice Shelf	10.45	0.00	89.55	100.00	94.49	0.00	0.36	100.00	99.64	99.82	0.943
George VI Ice Shelf I	0.00	6.01	100.00	93.99	96.90	0.60	0.00	99.40	100.00	99.70	0.966
George VI Ice Shelf II	0.44	3.42	99.56	100.00	98.05	0.45	0.05	99.55	99.94	99.75	0.978
Larsen C tributaries	11.11	4.0	88.89	96.00	92.31	0.05	0.15	99.95	99.85	99.90	0.922
Average	13.24	8.44	86.76	91.87	88.62	0.27	0.32	99.73	99.68	99.70	0.883

Modified after [1]

Conclusion & Outlook

Conclusion

- Random Forest has proven its applicability for mapping of supraglacial lakes in Antarctica and enabled the development of the **first automated mapping method** applied to Sentinel-2 data distributed across all three Antarctic regions

- The average F_1 score for the classification of surface lakes across all test sites was computed at ~89 % and the overall Kappa reached 0.883 suggesting the good functionality and **spatio-temporal transferability** of our workflow

- The main **remaining limitations** of our workflow are associated with (1) the lack of up-to-date topographic (and coastline) data, (2) difficulties in classifying pixels at lake edges and (3) shadow on ice below thick clouds in Sentinel-2 imagery

Outlook

- Ongoing work involves the improvement of the Random Forest model with **more training data** e.g. on shadow on ice as well as the application of our workflow to the whole Antarctic continent

- Besides, the results of this study are used for further methodological developments using **Sentinel-1**

References

(1) Dirscherl, M.; Dietz, A.J.; Kneisel, C.; Kuenzer, C. Automated Mapping of Antarctic Supraglacial Lakes Using a Machine Learning Approach. Remote Sens. 2020, 12, 1203.

(2) Williamson, A.G.; Arnold, N.S.; Banwell, A.F.; Willis, I.C. A Fully Automated Supraglacial lake area and volume Tracking ("FAST") algorithm: Development and application using MODIS imagery of West Greenland. Remote Sensing of Environment 2017, 196, 113–133.

(3) Yang, K.; Smith, L.C. Supraglacial Streams on the Greenland Ice Sheet Delineated From Combined Spectral–Shape Information in High-Resolution Satellite Imagery. IEEE Geoscience and Remote Sensing Letters 2013, 10, 801–805.

(4) McFeeters, S.K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. International Journal of Remote Sensing 1996, 17, 1425–1432.

(5) Ding, F. Study on information extraction of water body with a new water index (NWI). Sci. Surv. Mapp. 2009, 155–157.

(6) Feyisa, G.L.; Meilby, H.; Fensholt, R.; Proud, S.R. Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. Remote Sensing of Environment 2014, 140, 23–35.

(7) Kauth, R.J.; Thomas, G.S. The tasselled cap - A graphic description of the spectral-temporal development of agricultural crops as seen by Landsat. In Proceedings of the Proceedings of the Symposium on Machine Processing of Remotely Sensed Data; West Lafayette, IN, USA, 1976; Vol. 4B, pp. 41–51.

(8) Schwatke, C.; Scherer, D.; Dettmering, D. Automated Extraction of Consistent Time-Variable Water Surfaces of Lakes and Reservoirs Based on Landsat and Sentinel-2. *Remote Sensing* **2019**, *11*, 1010.

(9) Huete, A.R. A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment* **1988**, *25*, 295–309.

(10) ESA Sentinel-2 MSI Level-2A Algorithm Overview Available online: https://earth.esa.int/web/sentinel/technical-guides/sentinel-2-msi/level-2a/algorithm.

(11) Keshri, A.K.; Shukla, A.; Gupta, R.P. ASTER ratio indices for supraglacial terrain mapping. *International Journal of Remote Sensing* **2009**, *30*, 519–524.

(12) Li, H.; Xu, L.; Shen, H.; Zhang, L. A general variational framework considering cast shadows for the topographic correction of remote sensing imagery. ISPRS Journal of Photogrammetry and Remote Sensing 2016, 117, 161–171.

(13) Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32.

(14) Belgiu, M.; Drăgu¸t, L. Random forest in remote sensing: A review of applications and future directions.*ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31.

(15) Pal, M. Random forest classifier for remote sensing classification. *Int. J. Remote Sens.* **2005**, *26*, 217–222.

(16) Jolly, K. Machine Learning with Scikit-Learn Quick Start Guide; Packt Publishing Ltd.: Birmingham, UK, 2018; ISBN 978-1-78934-370-0.

(17) Müller, C.; Guido, S. Introduction to Machine Learning with Python: A Guide for Data Scientists; O'Reilly Media Inc.: Sebastopol, CA, USA, 2016; Volume 1, ISBN 978-1-4493-6990-3.

(18) Cohen, J. A Coefficient of Agreement for Nominal Scales. Educ. Psychol. Meas. 1960, 20, 37–46. 83.

(19) Landis, J.R.; Koch, G.G. The Measurement of Observer Agreement for Categorical Data. Biometrics 1977, 33, 159–174.