Appendix A KNN-Subspace Prediction Model

In this supplement, we provide additional information on the process used to refine the KNN subspace model to create our preferred machine learning classifier.

A.1 Refining Subspace KNN model

A subspace KNN model is a form of ensemble model in which the KNN method of classifying a point is preformed over a number of ensembles consisting of a select number of dimensions (our environmental parameters). These scores from each ensemble are then weighted against each other with the most common classification across the ensembles being chosen as the final classification. The model may be refined by adjusting two hyperparameters:

1. the number of dimensions being used, which can range from 1, i.e., classification based on a single environmental parameters, to all 6 environmental parameters used in this study—equivalent to a single KNN model; and

2. the number of number of learners (trained classifiers) used to create the ensemble.

The number of learners is not bounded, though there is a practical limit determined by the number of dimensions used for each classifier and the total number of training parameters. Choosing a value higher than this will result in oversampling of combinations.

Choice of the learners and subspace dimensions was performed by varying each parameter over a range of values and selecting the point where the L-curve flattens. For the learners, a marked improvement was observed in the accuracy of our models in classifying the training data between 1-7 learners. However, past this point we see minimal to no improvement in our model accuracy (Supplementary Figure 1 A). Thus, to prevent oversampling of our learners, we will uses 7 learners in the final model. For dimensions a similar trend is observed as seen in the learners in which accuracy improves greatly between 1-4 dimensions but little to no improvement is seen beyond that point (Supplementary Figure 1 B) . We will uses 4 dimensions as 5 and 6 dimensions are somewhat limiting in ensemble arrangements in comparison.

A.2 Refining Lake and Antarctica data

Establishing null locations (regions devoid) of lakes is necessary to identify regions without lakes. Doing so is a tricky proposition as choosing regions of null from radar maps or active lake predictions may exclude regions where melting is occurring but lakes cannot form due to infiltration of water into the bedrock, or unfavourable topographic conditions to capture the water. Randomly assigning locations without lakes also has the potential to exclude regions where lakes are present, but no observations exist. However, a large fraction of Antarctica has been surveyed by radar and satellite observations that cover the entire continent. Therefore, we have opted for a strategy based on the latter approach.

To define null lake positions, we randomly assign grid points as null, restricted only by their proximity to known lakes. A distance too close and one risks excluding regions where lakes are possible, a distance too far and one risks spatially isolated and anomalous environmental regions where no lakes exist, but are not particularly helpful to determining null regions more broadly. To determine the optimal distance for training null cells to lakes, tested a range of minimum distances for selecting the randomized null cells. As the minimum distance to known lakes is increased there is a decrease in the classification error (Supplementary Figure 1C). Although there is a distinct change in the error at 300 km, we chose a value of 200 km because larger values tend to force most null points onto the Antarctic Peninsula and portions of the eastern coast (Supplementary Figure 2). Thus, we chose to set our minimum proximity as 200 km, which results in high accuracy while preserving a wide distribution of null cell properties that are generally representative of the distribution of Antarctic environments (Supplementary Figure 2B).

Because geothermal heat flux is a poorly known property in Antarctica, we test the individual accuracy of the 6 proxy models used along with the combined mean (1 D). Surprisingly, we see newer proxy models preforming more poorly in comparison to older models, which conflicts with our assumption that the newer proxy models Shen et al. (2020) and Stål et al. (2021) are more accurate representation of the true Antarctic geothermal heat flux. Regardless, the mean performs better than any individual proxy model on its own, and thus we will use the mean of all 6 proxy models as our value for heat flux.

A.3 Final Subspace KNN model's ROC curve

Receiver operating characteristic (ROC) curves show the result of our models in identifying each classification types (stable, active, and null) at different classification thresholds on the Subspace KNN output scores (Supplementary Figure 3). The better the subspace KNN model is at classifying the data, the more rapidly the true positive value (e.g. stable lake being identified as a stable lake) will approach 1. This accuracy result can be measured by the area under the curve (AUC) with a value as close to 1 being desirable. The dots in Supplementary Figure 3 represent the cut-off thresholds our final subspace KNN model uses when classifying the data.

For stable lakes and null cells, we see very strong ROC curves both with AUC values of 0.94 showing that our final Subspace KNN model identify most of the training data's stable lakes and null cells without many false identifications. As a result, our final Subspace KNN model is, on average, able to identify 85% of null cells and 90% of stable lakes while only incorrectly identifying 5% of cells as null cells and 11% as stable lakes. While the final Subspace KNN model is excellent at identifying stable lakes and null cells, it is not as accurate at identifying active lakes. Our final Subspace KNN model is only able to identify 54 % of active lakes before the false positive rate becomes too large. As addressed in the main text, poor the accuracy of our model in identify active lakes is likely a result of the limited number of active lakes in our training data combined with the wide distribution of active lake properties.

A.4 Separate Basal Heat Flux Models

To test the individual geothermal heat flux models on the classifier, we preform the same hyperparameters used to train the preferred classifier 4. The result is fairly similar for each of the heat flux models, though there are a few differences in the details. For example, there are some differences in Dronning Maud Land, Ellsworth Land (Pine Island Glacier), and west of the Transantarctic Mountains. There is only a minor influence on the resulting classification accuracy as a result of these minor differences (Supplmentary Figure 1D).

References

An, M., Wiens, D. A., Zhao, Y., Feng, M., Nyblade, A. A., Kanao, M., Li, Y., Maggi, A., and Lévêque, J.-J. (2015). S-wave velocity model and inferred Moho topography beneath the Antarctic Plate from Rayleigh waves. *Journal of Geophysical Research: Solid Earth*, 120(1):359–383.

Ashmore, D. W. and Bingham, R. G. (2014). Antarctic subglacial hydrology: current knowledge and future challenges. *Antarctic Science*, 26(6):758–773.

Baranov, A., Tenzer, R., and Bagherbandi, M. (2017). Combined gravimetric–seismic crustal model for antarctica. *Surveys in Geophysics*, 39(1):23–56.

Behrendt, J. C. (1999). Crustal and lithospheric structure of the West Antarctic Rift System from geophysical investigations—a review. *Global and Planetary Change*, 23(1):25–44.

Bell, R. E., Studinger, M., Shuman, C. A., Fahnestock, M. A., and Joughin, I. (2007). Large subglacial lakes in East Antarctica at the onset of fast-flowing ice streams. *Nature*, 445(7130):904–907.

Carter, S. P., Blankenship, D. D., Peters, M. E., Young, D. A., Holt, J. W., and Morse, D. L. (2007). Radar-based subglacial lake classification in antarctica. *Geochemistry, Geophysics, Geosystems*, 8(3):n/a–n/a.

Christner, B. C., Priscu, J. C., Achberger, A. M., Barbante, C., Carter, S. P., Christianson, K., Michaud, A. B., Mikucki, J. A., Mitchell, A. C., Skidmore, M. L., Vick-Majors, T. J., Adkins, W. P., Anandakrishnan, S., Barcheck, G., Beem, L., Behar, A., Beitch, M., Bolsey, R., Branecky, C., Edwards, R., Fisher, A., Fricker, H. A., Foley, N., Guthrie, B., Hodson, T., Horgan, H., Jacobel, R., Kelley, S., Mankoff, K. D., McBryan, E., Powell, R., Purcell, A., Sampson, D., Scherer, R., Sherve, J., Siegfried, M., and and, S. T. (2014). A microbial ecosystem beneath the West Antarctic ice sheet. *Nature*, 512(7514):310–313.

Couston, L.-A. and Siegert, M. (2021). Dynamic flows create potentially habitable conditions in antarctic subglacial lakes. *Science Advances*, 7(8):eabc3972.

Fretwell, P., Pritchard, H. D., Vaughan, D. G., Bamber, J. L., Barrand, N. E., Bell, R., Bianchi, C., Bingham, R. G., Blankenship, D. D., Casassa, G., Catania, G., Callens, D., Conway, H., Cook, A. J., Corr, H. F. J., Damaske, D., Damm, V., Ferraccioli, F., Forsberg, R., Fujita, S., Gim, Y., Gogineni, P., Griggs, J. A., Hindmarsh, R. C. A., Holmlund, P., Holt, J. W., Jacobel, R. W., Jenkins, A., Jokat, W., Jordan, T., King, E. C., Kohler, J., Krabill, W., Riger-Kusk, M., Langley, K. A., Leitchenkov, G., Leuschen, C., Luyendyk, B. P., Matsuoka, K., Mouginot, J., Nitsche, F. O., Nogi, Y., Nost, O. A., Popov, S. V., Rignot, E., Rippin, D. M., Rivera, A., Roberts, J., Ross, N., Siegert, M. J., Smith, A. M., Steinhage, D., Studinger, M., Sun, B., Tinto, B. K., Welch, B. C., Wilson, D., Young, D. A., Xiangbin, C., and Zirizzotti, A. (2013). Bedmap2: improved ice bed, surface and thickness datasets for antarctica. *The Cryosphere*, 7(1):375–393.

Fricker, H. A., Scambos, T., Bindschadler, R., and Padman, L. (2007). An active subglacial water system in west antarctica mapped from space. *Science*, 315(5818):1544–1548.

Fricker, H. A., Siegfried, M. R., Carter, S. P., and Scambos, T. A. (2016). A decade of progress in observing and modelling Antarctic subglacial water systems. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2059):20140294.

Gard, M. and Hasterok, D. (2021). A global Curie depth model utilising the equivalent source magnetic dipole method. *Physics of the Earth and Planetary Interiors*, 313:106672.

Goeller, S., Steinhage, D., Thoma, M., and Grosfeld, K. (2016). Assessing the subglacial lake coverage of Antarctica. *Annals of Glaciology*, 57(72):109–117.

Goes, S., Hasterok, D., Schutt, D. L., and Klöcking, M. (2020). Continental lithospheric temperatures: A review. *Physics of the Earth and Planetary Interiors*, 306:106509.

Gray, L. (2005). Evidence for subglacial water transport in the west antarctic ice sheet through three-dimensional satellite radar interferometry. *Geophysical Research Letters*, 32(3).

Gudlaugsson, E., Humbert, A., Kleiner, T., Kohler, J., and Andreassen, K. (2016). The influence of a model subglacial lake on ice dynamics and internal layering. *The Cryosphere*, 10(2):751–760.

Guimarães, S. N. P., Vieira, F. P., and Hamza, V. M. (2020). Heat flow variations in the antarctic continent. *International Journal of Terrestrial Heat Flow and Applications*, 3(1):1–10.

Hasterok, D. and Chapman, D. (2011). Heat production and geotherms for the continental lithosphere. *Earth and Planetary Science Letters*, 307(1-2):59–70.

Hasterok, D. and Gard, M. (2016). Utilizing thermal isostasy to estimate sub-lithospheric heat flow and anomalous crustal radioactivity. *Earth and Planetary Science Letters*, 450:197–207.

Hasterok, D., Gard, M., and Webb, J. (2018). On the radiogenic heat production of metamorphic, igneous, and sedimentary rocks. *Geoscience Frontiers*, 9(6):1777–1794.

Hasterok, D. and Webb, J. (2017). On the radiogenic heat production of igneous rocks. *Geoscience Frontiers*, 8(5):919–940.

Horgan, H. J., Anandakrishnan, S., Jacobel, R. W., Christianson, K., Alley, R. B., Heeszel, D. S., Picotti, S., and Walter, J. I. (2012). Subglacial Lake Whillans — Seismic observations of a shallow active reservoir beneath a West Antarctic ice stream. *Earth and Planetary Science Letters*, 331-332:201–209.

Humbert, A., Steinhage, D., Helm, V., Beyer, S., and Kleiner, T. (2018). Missing evidence of widespread subglacial lakes at Recovery Glacier, Antarctica. *Journal of Geophysical Research: Earth Surface*, 123(11):2802–2826.

Jennings, S., Hasterok, D., and Payne, J. (2019). A new compositionally-based thermal conductivity model for plutonic rocks. *Geophysical Journal International*, 219(2):1377–1394.

Jordan, T. A., Riley, T. R., and Siddoway, C. S. (2020). The geological history and evolution of west antarctica. *Nature Reviews Earth & Environment*, 1:1–17.

King, J. C. and Turner, J. (1997). *Antarctic Meteorology and Climatology*. Cambridge University Press.

Krynauw, J. (1996). A review of the geology of East Antarctica, with special Reference to the c. 1000 Ma and c. 500 Ma events. *Terra Antarctica*, 3:77–89.

Lai, C.-Y., Stevens, L. A., Chase, D. L., Creyts, T. T., Behn, M. D., Das, S. B., and Stone, H. A. (2021). Hydraulic transmissivity inferred from ice-sheet relaxation following Greenland supraglacial lake drainages. *Nature Communications*, 12(1):3955.

Livingstone, S. J., Clark, C. D., Woodward, J., and Kingslake, J. (2013). Potential subglacial lake locations and meltwater drainage pathways beneath the Antarctic and Greenland ice sheets. *The Cryosphere*, 7(6):1721–1740.

Livingstone, S. J., Li, Y., Rutishauser, A., Sanderson, R. J., Winter, K., Mikucki, J. A., Björnsson, H., Bowling, J. S., Chu, W., Dow, C. F., Fricker, H. A., McMillan, M., Ng, F. S. L., Ross, N., Siegert, M. J., Siegfried, M., and Sole, A. J. (2022). Subglacial lakes and their changing role in a warming climate. *Nature Reviews Earth & Environment*, 3(2):106–124.

Llubes, M., Lanseau, C., and Rémy, F. (2006). Relations between basal condition, subglacial hydrological networks and geothermal flux in Antarctica. *Earth and Planetary Science Letters*, 241(3-4):655–662.

MacKie, E. J., Schroeder, D. M., Caers, J., Siegfried, M. R., and Scheidt, C. (2020). Antarctic topographic realizations and geostatistical modeling used to map subglacial lakes. *Journal of Geophysical Research: Earth Surface*, 125(3).

Magnússon, E., Pálsson, F., Gudmundsson, M. T., Högnadóttir, T., Rossi, C., Thorsteinsson, T., Ófeigsson, B. G., Sturkell, E., and Jóhannesson, T. (2021). Development of a subglacial lake monitored with radio-echo sounding: case study from the eastern Skaftá cauldron in the Vatnajökull ice cap, Iceland. *The Cryosphere*, 15(8):3731–3749.

Maguire, R., Schmerr, N., Pettit, E., Riverman, K., Gardner, C., Della-Giustina, D., Avenson, B., Wagner, N., Marusiak, A. G., Habib, N., Broadbeck, J. I., Bray, V. J., and Bailey, H. (2021). Geophysical constraints on the properties of a subglacial lake in northwest Greenland. *The Cryosphere*, 15:3279–3291.

Mareschal, J.-C. and Jaupart, C. (2013). Radiogenic heat production, thermal regime and evolution of continental crust. *Tectonophysics*, 609:524–534.

Martos, Y. M., Catalán, M., Jordan, T. A., Golynsky, A., Golynsky, D., Eagles, G., and Vaughan, D. G. (2017). Heat flux distribution of Antarctica unveiled. *Geophysical Research Letters*, 44(22):11,417–11,426.

MATLAB (2020). *version 9.8.0.1359463 (R2020a)*. The MathWorks Inc., Natick, Massachusetts.

Maule, C. F. (2005). Heat flux anomalies in antarctica revealed by satellite magnetic data. *Science*, 309(5733):464–467.

Messager, M. L., Lehner, B., Grill, G., Nedeva, I., and Schmitt, O. (2016). Estimating the volume and age of water stored in global lakes using a geo-statistical approach. *Nature Communications*, 7(1).

Mony, L., Roberts, J. L., and Halpin, J. A. (2020). Inferring geothermal heat flux from an ice-borehole temperature profile at Law Dome, East Antarctica. *Journal of Glaciology*, 66(257):509–519.

Morlighem, M., Rignot, E., Binder, T., Blankenship, D., Drews, R., Eagles, G., Eisen, O., Ferraccioli, F., Forsberg, R., Fretwell, P., Goel, V., Greenbaum, J. S., Gudmundsson, H., Guo, J., Helm, V., Hofstede, C., Howat, I., Humbert, A., Jokat, W., Karlsson, N. B., Lee, W. S., Matsuoka, K., Millan, R., Mouginot, J., Paden, J., Pattyn, F., Roberts, J., Rosier, S., Ruppel, A., Seroussi, H., Smith, E. C., Steinhage, D., Sun, B., van den Broeke, M. R., van Ommen, T. D., van Wessem, M., and Young, D. A. (2019). Deep glacial troughs and stabilizing ridges unveiled beneath the margins of the Antarctic ice sheet. *Nature Geoscience*, 13(2):132–137.

Oswald, G. K. A. and Robin, G. D. Q. (1973). Lakes beneath the Antarctic ice sheet. *Nature*, 245(5423):251–254.

Palmer, S. J., Dowdeswell, J. A., Christoffersen, P., Young, D. A., Blankenship, D. D., Greenbaum, J. S., Benham, T., Bamber, J., and Siegert, M. J. (2013). Greenland subglacial lakes detected by radar. *Geophysical Research Letters*, 40(23):6154–6159.

Paterson, W. S. B. (1994). *The Physics of Glaciers*. Pergamon.

Pattyn, F., Carter, S. P., and Thoma, M. (2016). Advances in modelling subglacial lakes and their interaction with the Antarctic ice sheet. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2059):20140296.

Pollett, A., Hasterok, D., Raimondo, T., Halpin, J. A., Hand, M., Bendall, B., and McLaren, S. (2019). Heat Flow in Southern Australia and Connections With East Antarctica. *Geochemistry, Geophysics, Geosystems*, 20(11):5352–5370.

Rignot, E. (2019). MEaSUREs Phase Map of Antarctic Ice Velocity, Version 1.

Robin, G. d. Q., Swithinbank, C., and Smith, B. (1970). Radio echo exploration of the antarctic ice sheet. *International Symposium on Antarctic Glaciological Exploration (ISAGE)*, 3(7):97–115.

Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., and Chica-Rivas, M. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geology Reviews*, 71:804–818.

Schroeder, D. M., Blankenship, D. D., Young, D. A., and Quartini, E. (2014). Evidence for elevated and spatially variable geothermal flux beneath the West Antarctic Ice Sheet. *Proceedings of the National Academy of Sciences*, 111(25):9070–9072.

Shapiro, N. and Ritzwoller, M. (2004). Inferring surface heat flux distributions guided by a global seismic model: particular application to Antarctica. *Earth and Planetary Science Letters*, 223(1):213–224.

Shen, W., Wiens, D. A., Lloyd, A. J., and Nyblade, A. A. (2020). A Geothermal Heat Flux Map of Antarctica Empirically Constrained by Seismic Structure. *Geophysical Research Letters*, 47(14).

Siegert, M. J. (2000). Antarctic subglacial lakes. *Earth-Science Reviews*, 50(1-2):29–50.

Siegert, M. J., Ellis-Evans, J. C., Tranter, M., Mayer, C., Petit, J.-R., Salamatin, A., and Priscu, J. C. (2001). Physical, chemical and biological processes in Lake Vostok and other Antarctic subglacial lakes. *Nature*, 414(6864):603–609.

Siegert, M. J., Kulessa, B., Bougamont, M., Christoffersen, P., Key, K., Andersen, K. R., Booth, A. D., and Smith, A. M. (2017). Antarctic subglacial groundwater: a concept paper on its measurement and potential influence on ice flow. *Geological Society, London, Special Publications*, 461(1):197–213.

Stål, T., Reading, A. M., Halpin, J. A., and Whittaker, J. M. (2021). Antarctic geothermal heat flow model: Aq1. *Geochemistry, Geophysics, Geosystems*, 22(2).

Stearns, L. A., Smith, B. E., and Hamilton, G. S. (2008). Increased flow speed on a large east antarctic outlet glacier caused by subglacial floods. *Nature Geoscience*, 1(12):827–831.

Thatje, S., Brown, A., and Hillenbrand, C.-D. (2019). Prospects for metazoan life in sub-glacial Antarctic lakes: the most extreme life on Earth? *International Journal of Astrobiology*, 18(5):416–419.

van der Veen, C. J., Leftwich, T., R. von Frese, B. M. C., and Li, J. (2007). Subglacial topography and geothermal heat flux: Potential interactions with drainage of the Greenland ice sheet. *Geophysical Research Letters*.

van Wessem, J. M., Reijmer, C. H., Lenaerts, J. T. M., van de Berg, W. J., van den Broeke, M. R., and van Meijgaard, E. (2014). Updated cloud physics in a regional atmospheric climate model improves the modelled surface energy balance of Antarctica. *The Cryosphere*, 8(1):125–135.

Willcocks, S., Hasterok, D., and Jennings, S. (2021). Thermal refraction: implications for subglacial heat flux. *Journal of Glaciology*, pages 1–10.

Wolff, E. and Doake, C. (1986). Implications of the form of the flow law for vertical velocity and age–depth profiles in polar ice. *Journal of Glaciology*, 32(112):366–370.

Wright, A. and Siegert, M. (2012). A fourth inventory of antarctic subglacial lakes. *Antarctic Science*, 24(6):659–664.

Wright, A., Young, D., Roberts, J., Schroeder, D., Bamber, J., Dowdeswell, J., Young, N., Le Brocq, A., Warner, R., Payne, A., et al. (2012). Evidence of a hydrological connection between the ice divide and ice sheet margin in the Aurora Subglacial Basin, East Antarctica. *Journal of Geophysical Research: Earth Surface*, 117(F1).

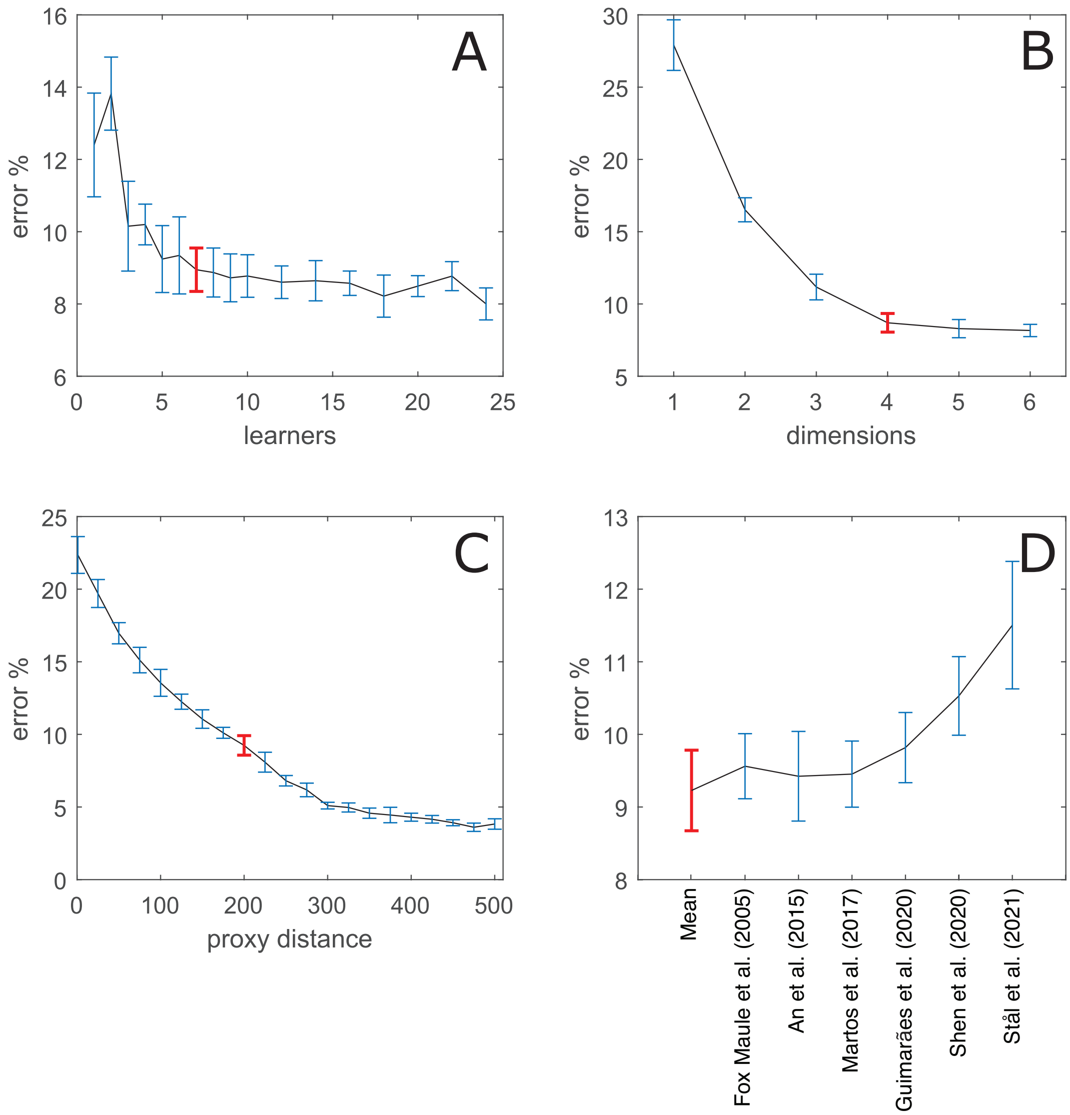

Figure 1: Tests used to develop the preferred subspace KNN classifier. Assessment is made using the error (incorrect classified/total × 100) shown with error bars produced by the standard deviation of 11 runs for each parameter set. Tests included (A) the number of learners, (B) the number of dimensions, (C) distance of null cells to known lakes, and (D) geothermal heat flow model. Heavy red bars indicate the preferred value.

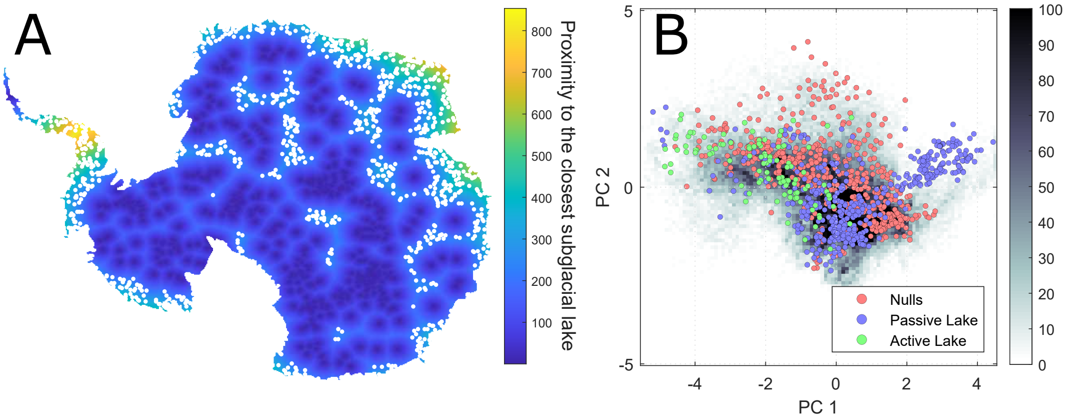

Figure 2: A) Map of showing proximity to subglacial lakes. A randomized set of null lakes with a minimum distance of 200 km is identified by the white points. B) Distribution of subglacial lakes and null cell location scores along the first and second principal axis (see main text).

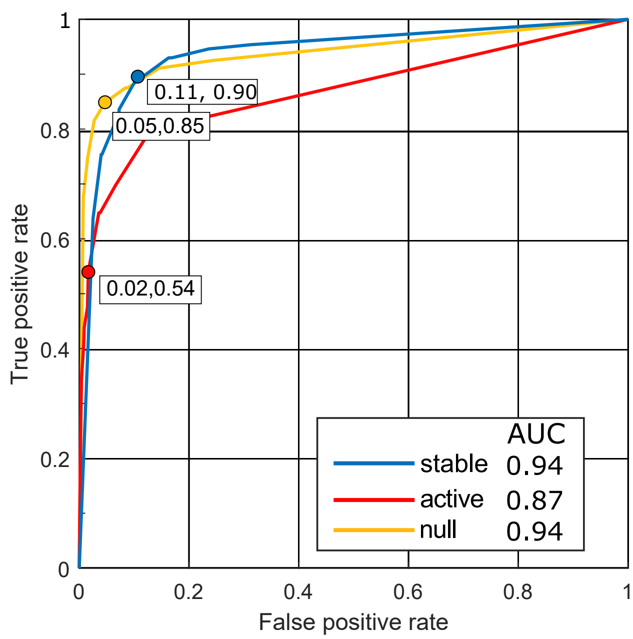

Figure 3: Receiver operating characteristic (ROC) curves for the active lake, stable lake and null cells along with the area under each curve (AUC). The point on each curve is the preferred trained classifier along with values for the false positive and true positive rates.

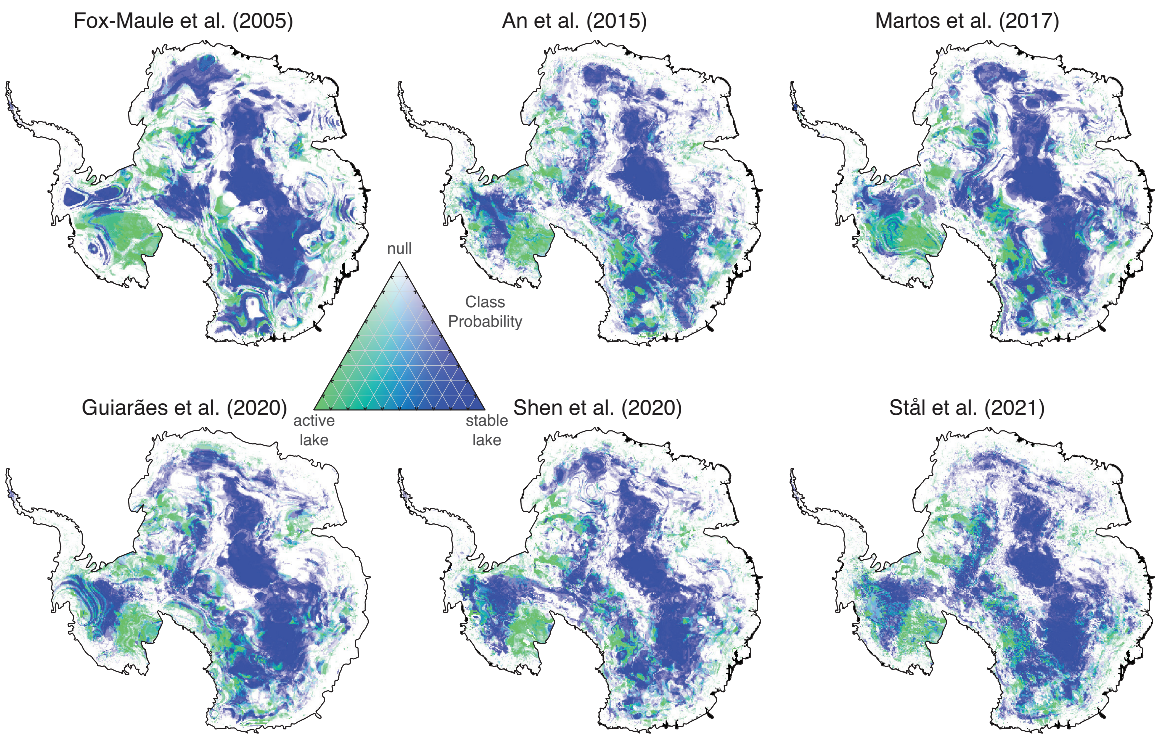

Figure 4: Machine learning Classifier of lake melt sources as preformed in Figure 6 of the main text. Each map uses a one of the 6 proxy models as its value for basal heat flux.