

Thank You for the interesting topic and manuscript. The manuscript will still require reasoning and clarification of the applied methodology and presentation of the results. Therefore, I propose a major revision.

*We would like to thank the referee for valuable comments which have substantially helped improve the clarity and quality of the manuscript and stimulated interesting and constructive discussion. \*Please, refer to the supplement for all tables and figures.*

Major comments:

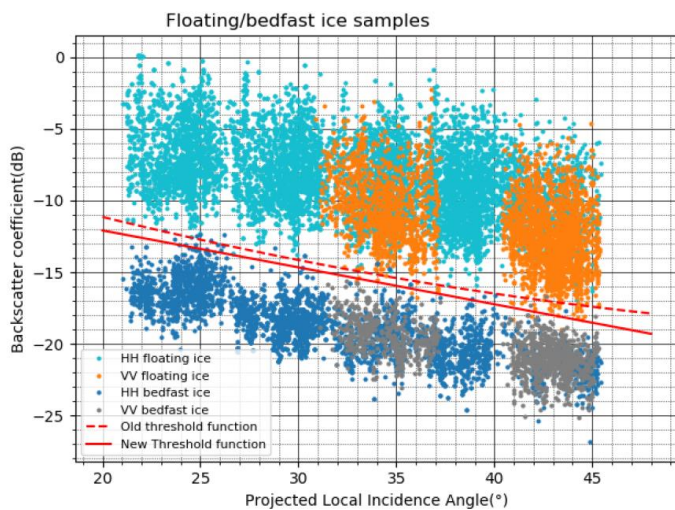
My major comments are related to the applied methodology and presentation.

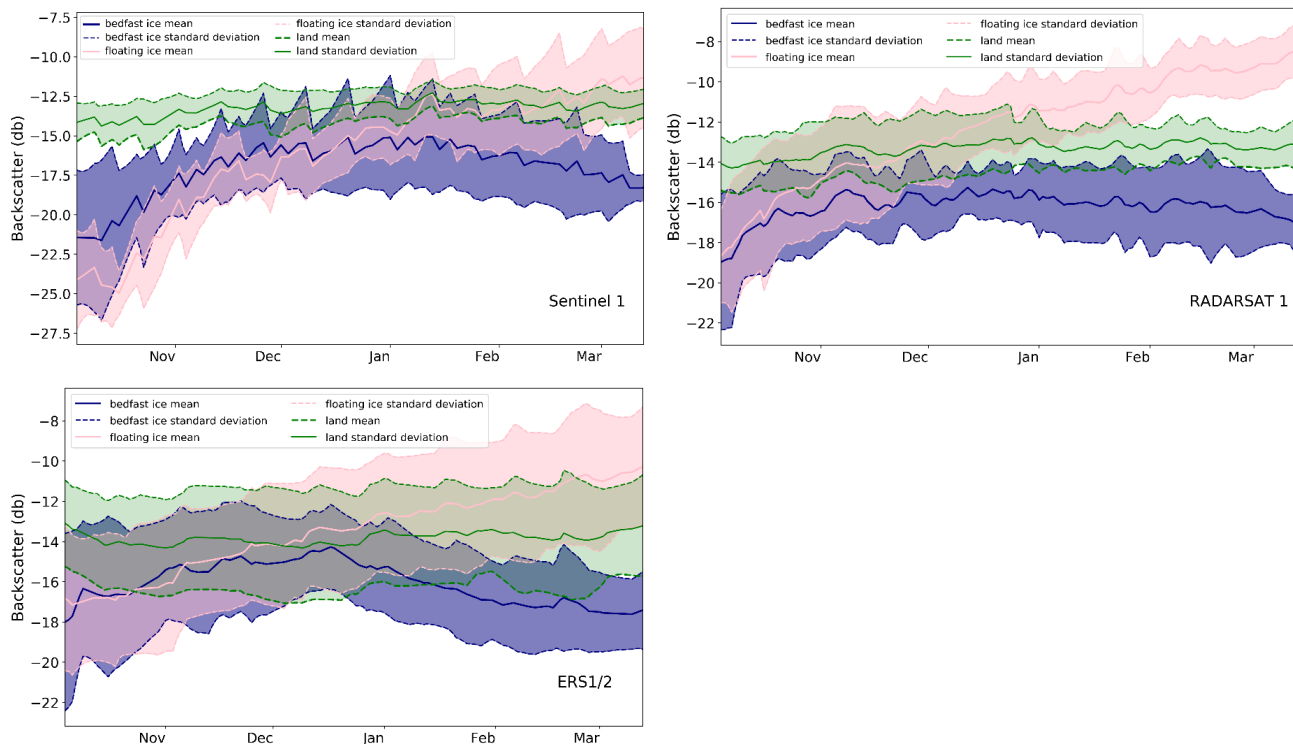
1) As pointed out by the other reviewers, the input data comes from different instruments and they even have different polarizations. The effect of this should be analyzed. Could e.g. the instrument type be fed into the NN as one additional input, or separate NN's used for different instruments? How much would the result improve by taking separate instrument into account (if any)? Also some kind of analysis of the backscattering of the separate instruments for the lake ice (and surrounding land) would be interesting (how do they differ or are they very similar statistically).

*Thank you for a valuable comment. The goal of this study was to present a method capable of working with different C-band SAR platforms and different polarizations. One of the strengths of a deep learning approach is the ability of the network to be trained on different types of data, learn all of its different aspects, and then be able to recognize and classify them correctly. It has been shown by Claude Duguay's research team (Duguay and Wang, 2019) that HH and VV backscatter of floating and bedfast ice is comparable and similar enough. Please, refer to the graph below.*

*In addition, as you can see in the three graphs below, although different instruments do have differences, the backscatter patterns for bedfast ice, floating ice, and land are quite comparable between the three platforms.*

*Duguay, C.R. and J. Wang, 2019. Arctic-wide ground-fast lake ice mapping with Sentinel-1. ESA Living Planet Symposium, Milan, Italy, 13-17 May.*





2) Selection of TempCNN as a method and parametrization TempCNN was selected to be used but it would be useful to compare the performance of TempCNN to some simple method (thresholding) to give evidence that it performs better (or much better?). Also more detailed reasoning of the selected structure would be useful to be included, e.g. why there are just three convolutional layers etc.? The selections could also be reasoned by referring to publications where the selections have been justified.

*Thank you for a valuable comment.*

- *The goal of this manuscript was not to compare methods, but rather present a new method not previously used for lake ice mapping describing its benefits for the Old Crow Flats. The method presented in our paper could, however, be compared to other previously published approaches in a follow-up study either in the same or other study area(s). Although, this method is not necessarily better than other methods it does present advantages, namely 1) it takes advantage of the temporal evolution of ice backscatter over the entire lake ice season and is not reliant on a single value, as is the case for other methods; as new SAR platforms are being launched with higher quality, spatial resolution, and denser temporal coverage it is only reasonable to look for ways to make use of the wealth of available data; 2) wetland landscapes are dynamic in nature and the presented TempCNN method does not require a lake mask, which is necessary for all the previously presented methods; 3) a well-trained TempCNN can be applied to various SAR sensors, both HH and VV polarizations, and different spatial resolutions. The accuracy of the method was assessed using a training set and the overall accuracy values fall in the range reported by other ice regime mapping solutions.*
- *This work closely follows the TempCNN approach presented for land cover mapping using optical imagery by Pelletier et al., 2019. In terms of model parameters, the number of convolutional units for each convolutional layer was selected to be 64 based on temporal cross validation using 4, 8, 16, 32, 64, and 128 units (Please, refer to Figure 5). Batch size was adjusted to achieve best results. In addition, the number of training epochs is controlled by an early stopping technique with a patience of 10 epochs and a validation set comprising 5% of the training set.*

*Pelletier, C., Webb, G. I., and Petitjean, F.: Temporal convolutional neural network for the classification of satellite image time series, Remote Sens., 11, 1–25, 2019.*

3) Use of data, division to training and test data sets: It should be better reasoned why the model training setup was such complicated. On what are the selections and divisions based? Cross validation is a good way to train and test if there is little data. If You want to continue the time series (with the same training) it would also be good to have a training with

good generalization property. In any case, give reasoning for the use of the data sets and division to training and test. Would a more simple approach be feasible or even better?

*Thank you for a valuable comment. The 15 experiments described in the subsection 3.3.4 TempCNN training and testing aim to test how sensitive the model is to inclusion or exclusion of certain years of data. We have used three different ways to split the data into training and testing sets: 1) a random split of all data points into 80% for training and 20% for testing; 2) three complete years (each from a different SAR platform) were left out for testing, 15 years of data were used for training; 3) training was done using 17 years of data and testing using data from the 2020/2021 ice season which was originally reserved for validation and not used when determining optimal neural network architecture.*

*To clarify this issues, caption of Table 2 has been updated as follows:*

*“Table 2. TempCNN overall classification accuracy for 15 experiments designed to test sensitivity of the network to removing certain years of data from the training set. Runs 1-5 correspond to the 20/80% split of the entire dataset, runs 6-10 were performed by training the network on 15 years of data and testing it on 3 each from a different sensor, runs 10-15 were carried out by training the network on 17 years of data and testing it on 1 year of data that was originally reserved and was not part of the cross validation procedure for determining the best architecture. Subsequently, mean accuracy for each set of 5 runs was calculated, and finally, mean of the three means is shown in the last row of the table.”*

4) The effect of speckle should be evaluated by comparing the results without and with speckle filtering. Could the filtering be included in the neural network model? Could e.g. a small neighborhood around each pixel be used instead of single pixel values (applying a 2-D convolution)?

*Thank you for a valuable comment.*

*- Speckle filtering is a standard processing step when working with C-band SAR (e.g., Surdu et al., 2014; Arp et al., 2012; Bartsch et al., 2017; Engram et al., 2018; Duguay and Wang, 2019) and is always performed prior to any environmental analysis. Without speckle filtering the salt and pepper effect is so strong that it would not be possible to extract any useful information.*

*- This work considered pixel-based classification. However, patch-based classification will be the next step in our research.*

5) The reference data are not very good. Is there any way to evaluate the accuracy of the reference data e.g. w.r.t. the existing field measurements?

*Thank you for a valuable comment. To the best of the authors' knowledge, lake (bedfast and floating) ice regimes have not been previously investigated in the Old Crow Flats. As such, field data is very scarce. All the available field measurements have been used in the study. If possible, more ice regime field work will be carried out in the future.*

6) Analysis results: There are a lot of details and figures of selected subregions. What I miss would be a clear conclusion of the analysis indicating by a few numbers of one figure the most essential results of the analysis for the tundra and taiga lakes (in general) and possibly estimated uncertainty estimates. These could be given in a separate shortish subsection

*Thank you for a valuable comment. We appreciate the suggestion. However, we have not done the analysis to compare tundra and taiga ice regimes. The two ecotones are likely responding differently, and a follow-up study would be necessary to analyse and compare lake ice regime dynamics in taiga and tundra areas of the Old Crow Flats.*

Some detailed comments:

L70-74: There are some studies using the separate between static and drifting sea ice based on ice drift or correlation estimated from (SAR) image pairs. Such method has been applied e.g. in Makynen, M.; Karvonen, J.; Cheng, B.; Hiltunen, M.; Eriksson, P.B. Operational Service for Mapping the Baltic Sea Landfast Ice Properties. Remote Sens. 2020, 12, 4032. <https://doi.org/10.3390/rs12244032> Also provide a reference to this kind of approach where backscattering is not directly used would complement the manuscript.

*Thank you for a valuable comment. The following sentence has been added to the Introduction section:*

“Nonetheless, not all bedfast mapping approaches rely directly on the SAR backscatter, for instance, some sea ice studies identified bedfast ice using SAR interferometry (Dammann et al., 2018) and landfast ice using SAR image pairs (Makynen et al., 2020).”

Dammann, D.O., Eriksson, L.E.B., Mahoney, A.R., Stevens, C.W., Van der Sanden, J., Eicken, H., Meyer, F.J., and Tweedie, C.E.: Mapping Arctic Bottomfast Sea Ice Using SAR Interferometry, *Remote Sens.*, 10, 720, <https://doi.org/10.3390/rs10050720>, 2018.

Makynen, M., Karvonen, J., Cheng, B.; Hiltunen, M., and Eriksson, P.B.: Operational Service for Mapping the Baltic Sea Landfast Ice Properties, *Remote Sens.*, 12, 4032. <https://doi.org/10.3390/rs12244032>, 2020.

Table 1: This indicates that VV mode has been used, except for RS-1. Were there not HH mode data available (e.g. S-1 EW mode data in HH/HV)? Would including cross-polarized channel improve the detection (or has this been studied by anyone)? This is interesting because there exist a lot of HH/HV or VV/VH data acquired by RS-2 and S-1.

*Sentinel-1 EW mode HH/HV imagery was available only for the season of 2016/2017. However, the earliest scene for that season was October 14. Starting the time-series in mid-October could have caused missing the major component of the lake ice lifecycle, namely the initial drop of backscatter as thin ice forms over the water surface. For the remaining years, there was no HH/HV coverage (all the available scenes were located slightly to the north and north-east of the study area).*

*HH and VV polarizations are commonly used for bedfast lake ice mapping (e.g., Engram et al., 2018). To the best of the author’s knowledge, cross-polarized SAR imagery is not used for this task. There is a study that performs a classification of decaying ice and open water using HH and HV channels (Geldsetzer et al., 2010).*

*In addition, early C-band SAR data is available only as co-polarized (e.g., ERS1/2). As such, the authors were interested in developing a method that would be able to extend back and use the early platforms as well as take advantage of the new C-band SAR platforms.*

*Engram, M., Arp, C. D., Jones, B. M., Ajadi, O. A., and Meyer, F. J.: Analyzing floating and bedfast lake ice regimes across 580 Arctic Alaska using 25 years of space-borne SAR imagery, *Remote Sens. Environ.*, 209, 660–676, 2018.*

*Geldsetzer, T., Sanden, J., and Brisco, B.: Monitoring lake ice during spring melt using RADARSAT-2 SAR, *Can. J. Remote Sens.*, 36(sup2):S391-S400, 2010.*

L155: Give reference to Lee filter used (there exist variants of Lee filter). On what is the 7x7 window size based?

*Thank you for a valuable comment.*

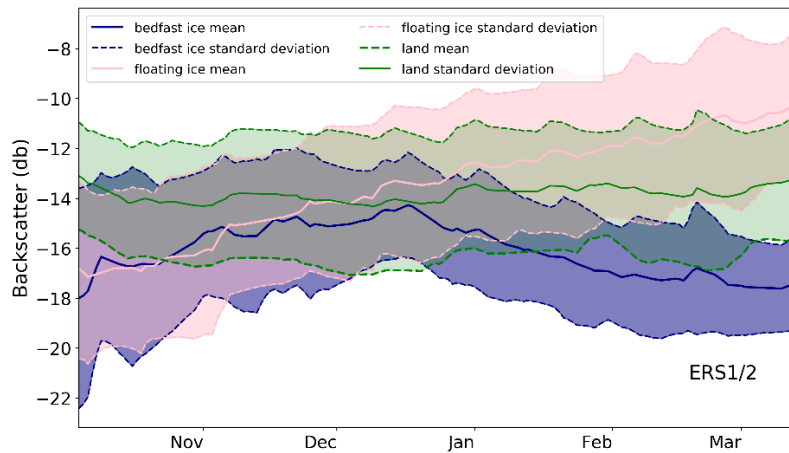
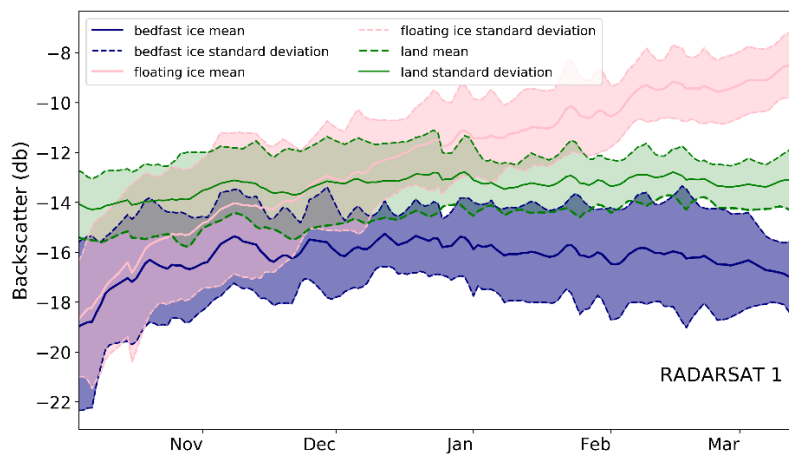
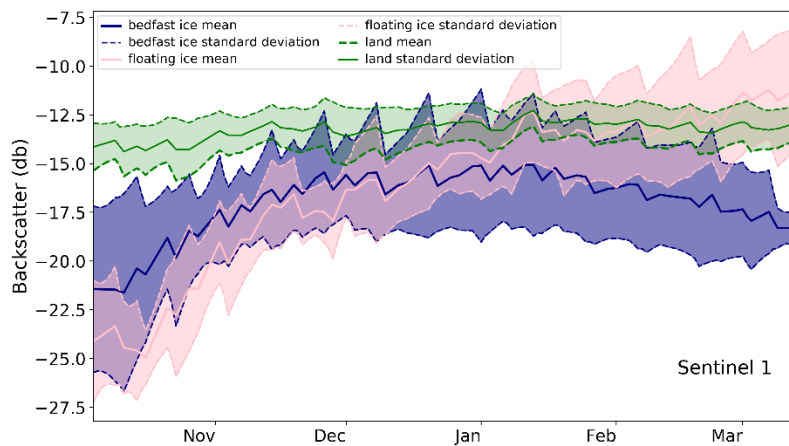
*Speckle filtering is part of the Radiometric Terrain Correction process carried out by the Alaska Satellite Facility for Sentinel-1. They use Enhanced Lee filter (reference has been added: Banerjee et al., 2021) to remove speckle while preserving edges. When applied, the filter is set to a dampening factor of 1, with a box size of 7x7 pixels and 180 looks. The speckle filter size for the other two sensors was adjusted based on the spatial resolution to cover a similar area as the 7x7 window used for Sentinel-1. Please, refer to the table below.*

Platform	Pixel size (m)	Filter size (pixels)	Area covered by each filter (m <sup>2</sup> )
Sentinel 1	30 x 30	7 x 7	44,100
ERS1/2	12.5 x 12.5	16 x 16	40,000
RADARSAT 1	50 x 50	4 x 4	40,000

*Banerjee, S., Sinha Chaudhuri, S., Mehra, R., Misra, A.: A Survey on Lee Filter and Its Improved Variants, in: *Advances in Smart Communication Technology and Information Processing*, edited by Banerjee, S., Mandal, J.K., *Lecture Notes in Networks and Systems*, vol 165. Springer, Singapore, 2021 [https://doi.org/10.1007/978-981-15-9433-5\\_36](https://doi.org/10.1007/978-981-15-9433-5_36)*

Figure 2: It would be more clear to show e.g. average and deviation for the classes, now the many curves are shadowing each other.

*Thank you for a valuable comment. Following the referee's suggestion, the means and standard deviation corridors have been plotted instead. Please, see the modified graphs below:*



L254: 330m, how was this elevation threshold selected. Give some reasoning for this selection.

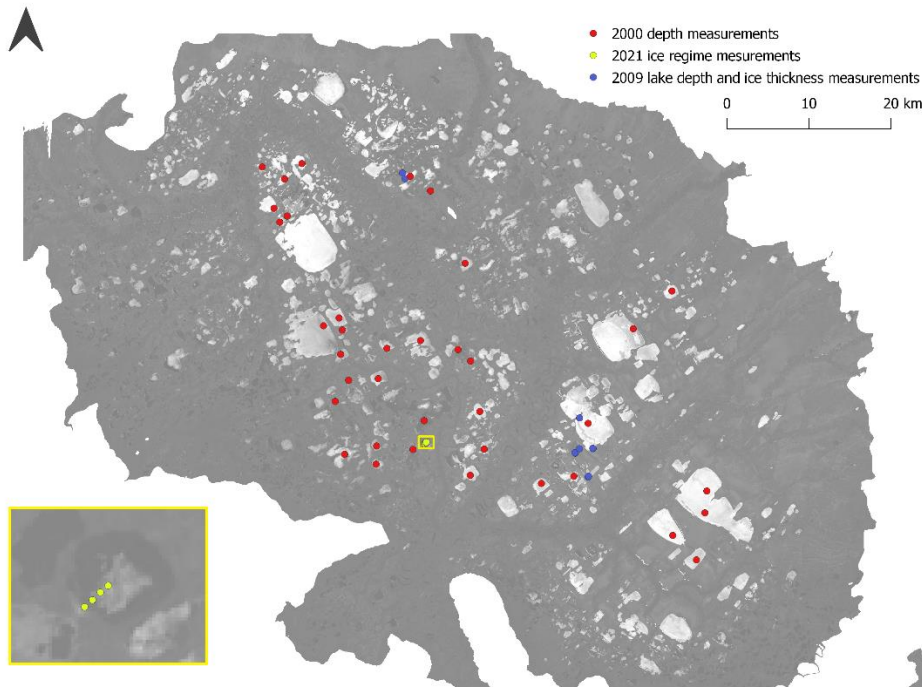
*The threshold was selected by attempting to define an area that would encompass all of the wetland with the majority of lakes, but exclude the surrounding mountainous areas.*

L257: Give the number(s) of field measurements here.

*Thank you for a valuable comment. The manuscript has been modified as follows:*

*“Apart from statistical classification based on the test set, accuracy of the ice regime maps was assessed using a set of 51 field observations.”*

*In addition, we are providing a map that demonstrates the distribution of field measurements for the referee’s reference.*



L282: (Date1\*10)+Date2, You probably mean class(date1)\*10 + class(date2) or something similar?

*Thank you for a valuable comment. The formula has been replaced with:*

*“class on date 1” \*10) + “class on date 2”*

L285-286: experimentally defined threshold, be more specific with this. A simple way to define a threshold statistically (experimentally) would be the Bayesian approach based on class distributions. Was this approach used or how was the threshold defined experimentally?

*Thank you for a valuable comment. The threshold was identified by trial and error. We have tried multiple backscatter values moving by 0.5 dB. The selected threshold (-16.5 dB) captured the lake boundaries the best. The manuscript has been modified as follows:*

*“The lake mask was created using an early October 2020 scene and a threshold of -16.5 dB identified through the process of trial and error.”*

L291: pyManKendall, give a reference to this python package.

*Thank you for a valuable comment. A reference for the python package has been added:*

*“Hussain et al., (2019). pyMannKendall: a python package for non parametric Mann Kendall family of trend tests. Journal of Open Source Software, 4(39), 1556, <https://doi.org/10.21105/joss.01556>”*

L305: Is linear interpolation really feasible. To get evidence, You could compare linear interpolation of the periods with data to get error estimates of the linear interpolation.

*Thank you for a valuable comment. The authors are applying linear interpolation following the lead of Pelletier et al., 2019 who propose the TempCNN method for land cover classification using optical imagery. The above-mentioned study*

indicates that upon investigation they came to a conclusion that more complex interpolation methods have little influence on classification accuracy. In our work linear interpolation does work as is shown by the accuracy assessment. However, following referee's suggestion we will definitely investigate more complex interpolation techniques in the future research.

Table 2: Please, give formulas /explanations of the accuracy measures (in the text).

Thank you for a valuable comment. The text of the manuscript has been modified as follows:

“As has been mentioned in Sect. 3, the TempCNN model classification accuracy was evaluated through 15 experiments. Table 2 presents the overall accuracy of the 15 experiments calculated as total number of correctly classified time-series in the training set divided by the total size of the training set and multiplied by 100%.”

Fig 7: Better distinguishable colors could be used for different classes. Now the blue tones are difficult to distinguish in a printed version.

Thank you for a valuable comment. Colours have been updated, please see below:

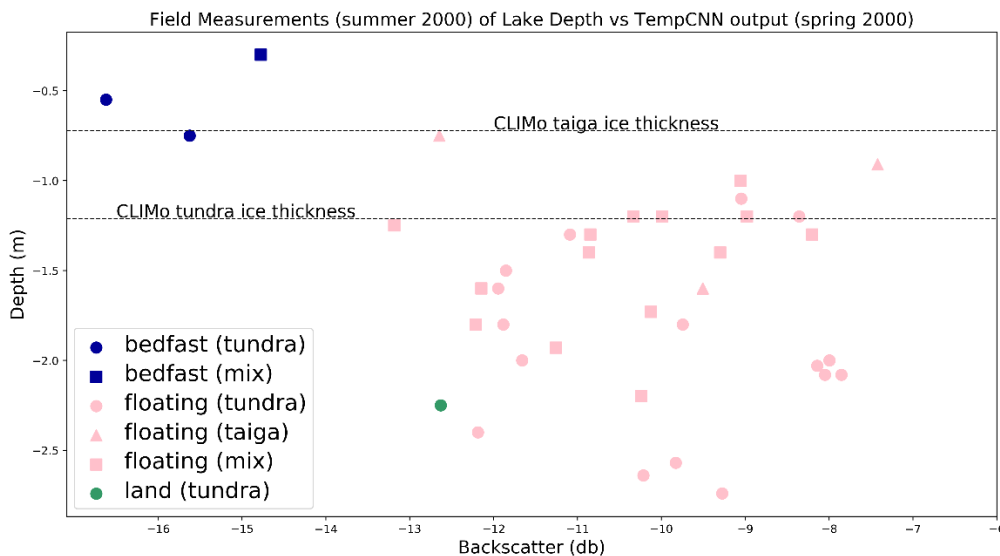


Fig.8: Are these time series really informative? Maybe only ice thickness and snow depth for certain day(s) could be shown, now it seems these figures include too much information.

Thank you for a valuable comment. However, authors would like to keep both ice thickness and snow depth for the full year. We believe that having the maximum value, as well as the seasonal trajectory along with the value on March 13 creates helpful context for the ice regime analysis, comparison of different years, and might be useful for the reader.

We wanted to illustrate to the reader the interannual variability and evolution of the snow and ice thickness during the ice growing season relative to the last date of the time-series for which the classification was performed.

L534: SWOT, can You give any reference to this?

Thank you for a valuable comment. SWOT (Surface Water and Ocean Topography) mission is due for launch in 1 month and 20 days. The manuscript has been modified as follows to include the link of the mission webpage that contains all the details:

“Future analyses of the bedfast/floating ice regime of the OCF will also benefit from data of the upcoming Surface Water and Ocean Topography (SWOT) mission (due for launch in November 2022; <https://swot.jpl.nasa.gov/>), which will allow for higher accuracy mapping of water level of lakes than is currently possible from current radar altimetry missions.”