Response to Michael Wolovick (RC1)

Overview

In this manuscript, the authors apply Deep Learning (DL) techniques to the problem of identifying subglacial lakes using ice-penetrating radar data. They use a multi-step method consisting of first encoding the information contained in the shape of each vertical reflector trace into a lower-dimensional latent space, and then applying a clustering algorithm to that space in order to identify populations with similar trace shapes. Of these clusters, they identify the one with a narrow symmetric peak in reflection power as representing subglacial lakes, and they further refine the population of subglacial lakes by performing a simple linear attenuation correction to the bed reflection power. Thus, their final identification procedure can be viewed as containing two parts: one part, the encoding-clustering analysis, focuses on the shape of the reflector trace, looking for reflections that are narrow and symmetric, as would be expected for a specular interface. The second, the attenuation analysis, focuses on the strength of the reflector rather than the shape. The authors apply their method to the AGAP-S dataset from East Antarctica and compare their identified lakes with previously published lake compilations.

This manuscript is appropriate for publication in The Cryosphere. It represents a new method in the analysis of ice-penetrating radar data with machine learning techniques and a new method for the identification of subglacial water. However, before it can be accepted in final form, I think that the authors need to provide more justification and explanation around their choice to use a clustering algorithm and on their choice of a particular number of clusters to use in that algorithm. In the remainder of my review, I first explain my major concern about the clustering algorithm, and then I give detailed comments on the rest of the paper.

Reply: We greatly appreciate your helpful comments and detailed advice regarding this paper.

Major Concern

My biggest concern with the analysis in this paper is the decision to use a clustering algorithm, which splits the data into discrete non-overlapping clusters, and the arbitrary decision to use 15 clusters. The data presented in Figure 3a do not appear to display any inherent clustering on their own. Rather, the data points appear to vary continuously across the Z1,Z2 plane. This point is confirmed by the author's own elbow curve (supplemental figure 1), which does not display a clear cutoff, and this point is acknowledged by the authors themselves on lines 166-168. Thus, a mode of analysis that breaks the data up into discrete categories may not be appropriate, and the authors' decision to use 15 categories is arbitrary and unsupported. Nonetheless, subsequent steps in the analysis protocol are dependent on the use of a clustering analysis earlier on.

The ultimate end state of the analysis- a list of positively identified water bodies- requires that the data be split into discrete categories at some point. At some point a threshold must be applied to distinguish “water” from “not water”, and if the clustering analysis is not used for this purpose, then some other means of setting a threshold must be used. Additionally, if I have interpreted the authors’ reflection power analysis correctly, then the clustering analysis may also be needed at this stage as well, since I think that they are averaging reflection power values within contiguous reflectors identified through cluster analysis before analyzing reflection power (but note that their explanation of this part of the method was somewhat unclear, so I am not 100% confident that I have interpreted their procedure correctly; I discuss the need for more clarity around this method in the Detailed Comments section below). Averaging reflection power within contiguous similar reflectors is a reasonable methodological choice, since reflection power can be quite variable along-track. Basically, the authors are using the clustering analysis to identify contiguous regions along the bed that have basically the same reflection trace, and they are defining each of those regions as “one reflector” with a single average reflection power for the purpose of reflection power analysis.

Thus, I face a dilemma: on the one hand, I do not want to recommend that the authors remove the cluster analysis entirely, since doing so may necessitate downstream changes throughout their method, including changes to parts of the method that seem sensible; but on the other hand, the data do not seem to support the use of discrete clusters and the particular choice of 15 clusters is unsupported. The arbitrary decision to use 15 total clusters can be regarded as an indirect means of setting the threshold separating “water” from “not water”, since the size of each cluster varies inversely with the total number of clusters. A low total number of clusters will increase the size of each individual cluster, thus increasing the diversity of reflectors in the “water” cluster, while a high total number of clusters will make the “water” cluster smaller. The authors state on line 168 that they have tested different values of K (the total number of clusters), but they do not show the results of those tests in the manuscript.

Therefore, as a minimum condition for publication, I think that the authors should show the results of some of these sensitivity tests. How does the final population of water bodies depend on the choice of K? What percentage of the identified water bodies are robust to the choice of K? Perhaps the subpopulation of water bodies that emerge for multiple values of K could be considered a more robust identification of subglacial lakes. (As an aside, if some values of K result in the clustering algorithm splitting the upper right “water corner” of Z1,Z2 space into two clusters, then that would be a valid argument for omitting those values of K; the K sensitivity test should only include values of K for which there is one unambiguous “water cluster”). Alternatively, if the authors can provide argumentation to justify their particular choice of K=15, then that could also satisfy my concerns, although the authors’ own statement on lines 166-168 seems to indicate that they do not believe a particular value of K is supported by the data. It may also be worth plotting the elbow diagram (supplemental figure 1) on log/log axes to see if a corner emerges when the data are plotted in that fashion. Once the authors have either justified their particular choice of K, explored the sensitivity of their results to K, or both, then I think that this manuscript will make an excellent addition to The Cryosphere.

Reply: Much appreciate your thoughtful and detailed feedback on our method. We agree that the application of 15 clusters on latent space analysis is arbitrary and unsupported. This number of 15 is chosen according to the final results of subglacial detection, which displays a better matching between the visual discriminations and the encode-cluster method. Indeed, we were still seeking a reasonable method to validate and evaluate the K number when we submitted the primary manuscript. Thus, we would like to acknowledge your helpful advice.

According to your concerns and advice, we have modified the manuscript by the following points:

(1) We have appended more content about both the cluster ranges in latent space and subglacial lake detections when different K values are applied. The result has shown the sensitivity between detections and K values. We also appended more discussion on the results from different K-values applied.

(2) We have appended more discussion about the potential studies by introducing reflection power into clustering. The contiguous reflectors identified as the same cluster, which applied to obtain the averaging reflection power, rely on a step of clustering on latent space to separate the different clusters of reflectors. So if we apply clustering on both the average reflection power and latent space, there will be a dilemma on the method's step, as well as you mentioned.

(3) We have plotted the elbow curve on log/log axes, and have updated this diagram on Figure S1(S2 in the revised version). We agree that although the log/log elbow curve does not have a significant cut-off point, this external plotting could provide an additional reference.

Detailed Comments

L14 Some of these references aren't really appropriate to use as a general background on subglacial water. Robin 1970 is about ice-penetrating radar, Siegert 2000 is about subglacial lakes, and Pattyn 2010 is about the results of a specific model. Pattyn and Siegert could work, although they aren't necessarily the best citations for this purpose, but Robin 1970 is definitely the wrong reference to use here. Chapters 6 and 9 of (Cuffey and Paterson, 2010) could be cited here, although I understand that citing a textbook is a bit unsatisfying. Another important reference might be (Robin, 1955).

Reply: We have modified and updated these wrong references. Thank you for your indications.

L21 "...in recent years..."

I don't know if it is fair to describe the use of ice-penetrating radar for detecting the subsurface features of ice sheets as “recent”. Maybe this sentence would be better as, “Ice-penetrating radar can be used to detect the subsurface features of ice sheets”. Also, this would be a good place for the Robin (1970) reference, not L14. Another good reference might be (Robin et al., 1969) or (Bailey et al., 1964).

Reply: We have modified this sentence and added more citations. Thanks for your advice on enriching our introduction.

L50: “from the CReSIS”

Should be “from CReSIS”.

Reply: Done.

L52: “We then apply K-means clustering method”

This sentence should be reworded in one of the following 3 ways: 1) “We then apply theK-means clustering method”, 2) “We then apply a K-means clustering method”, or 3) “We then apply K-means clustering methods”. Wording (1) applies if there is only 1 version of the K-means clustering method, wording (2) applies if there are multiple versions of the method but you only use 1 of them, and wording 3) applies if you use multiple versions of the method.

Reply: Much appreciate your detailed advice on improving our expression. According to our updated version, we have applied multiple K-means by using different K values. Thus, we reworded this sentence according to (3).

L54: “We notice a cluster”

“We identify a cluster” sounds better.

Reply: Thanks for your indication, we have updated that.

L61: “...to detect and label the other clusters...” “The” is unnecessary here.

Reply: Done, Thanks.

L72: “The radar images also contain the positions of ice bottom reflectors, which were extracted by hybrid manual-automatic method (Wolovick et al., 2013).”

Note that the bed picks produced by (Wolovick et al., 2013) are not the same bed picks included in the CReSIS data release. The AGAP-S data was processed in parallel at both the Lamont-Doherty Earth Observatory (LDEO) and CReSIS. The results of the LDEO processing are available at: https://pgg.ldeo.columbia.edu/data/agap-gambit. Though the original raw data is the same for both institutions, the code for SAR migration and bed picking was different, and of course different human operators provided the “manual” part of the manual-automatic bed picking. Both institutions used “hybrid manual-automatic” bed picking, but it might be better to cite a CReSIS source here if you are using the CReSIS version of the data.

Reply: Thanks a lot for the detailed interpretation of the data. We received the difference in data, and have updated the citation.

L81: “Second, we apply the reflector position markers in the dataset to truncate the 1-D data within the ±200 sampling points near the reflector position for every single trace along Z-axis.”

It might clarify things a bit to say that the reflector in this case is the bed. So maybe, “Second, we use the bed picks in the dataset to truncate the 1-D data within ±200 sampling points near the bed reflector position for every single vertical trace.”

Reply: Much appreciate your advice. It looks much better after modification.

L88: Gaussian filter, normalization.

What is the filter width? How are the data normalized?

Reply: Thanks for your indications, we have added the details about these operations.

L91: “1488600 1-D Z-axis (A-Scope) radar echo traces.”

Were you using the original 1.3 m along-track spacing of the data, or are you working with data that have been downsampled? I did not work with the CReSIS version of the AGAP-S data, but I know that other CReSIS data products are generally released at coarser horizontal resolution than this, and when processing the AGAP-S at LDEO, I downsampled the data by a factor of 10 in the along-track dimension. The original 1.3 m data should have a lot more than 1.4 million traces. If you are using downsampled data, then you should mention that.

Reply: Much appreciate your indication, we have amended and updated the introduction of the data product according to the database's manual.

L105-110: MSE

What does MSE stand for?

Reply: MSE represents "mean squared error". We have fixed that. Thanks for the indication.

L132: “...are challenging to be reconstructed...” Change to: “...are challenging to reconstruct...”

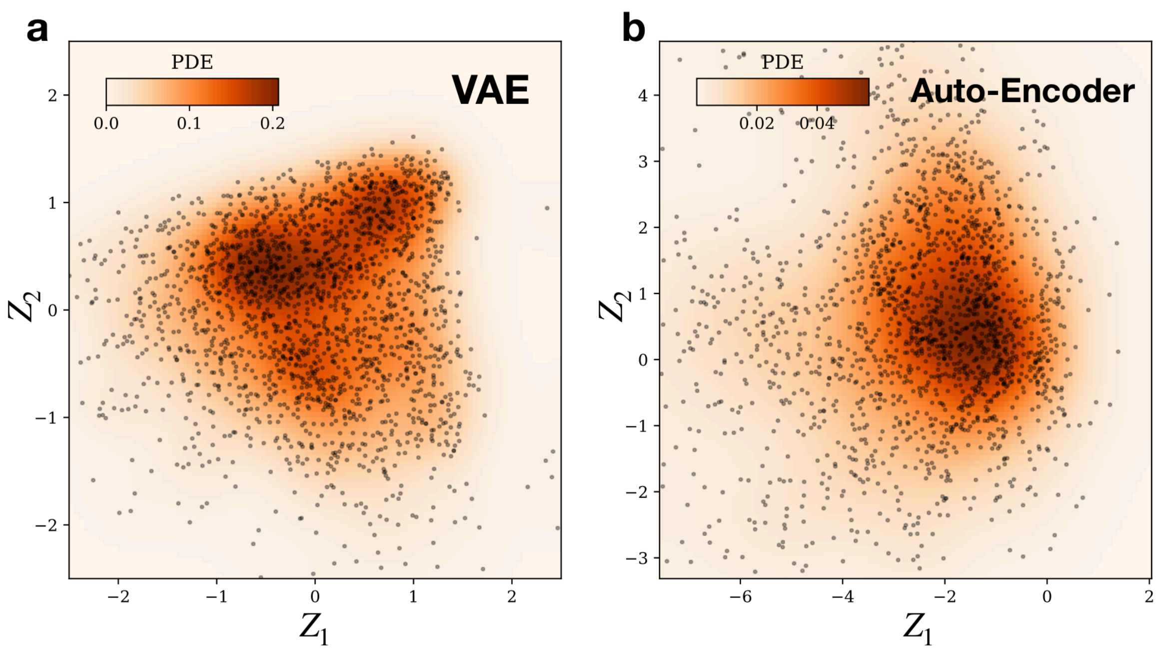
Reply: Done.

L166-168: “However, the elbow curve does not show a clear cutoff point, possibly due to the distribution of vectors in the latent space (Figure 3a) not displaying a distinct trend of multiple classes.”

The “elbow curve” for this method seems to be analogous to an L-curve for inverse problems. Could you please present the elbow curve on log-log axes in addition to the linear axes you used in your supplemental figure? The problem with linear axes for this purpose is that inverse power laws always appear L-shaped on linear axes, despite having no intrinsically preferable value.

Additionally, when I look at Figure 3a, it appears to my eye that the data do not really have any clusters at all. Is that what you meant by “not displaying a distinct trend of multiple classes”? To my eye, it looks like the data are smoothly distributed within the central part of the latent space, with perhaps a greater number of outliers in the negative direction for both Z1 and Z2 than in the positive direction. It sounds as though the lack of visual clustering in Fig 3a is confirmed by the lack of a clear cutoff in the elbow curve. I discuss this issue at greater length in my “Major Concern” section above.

Reply: Thanks for your suggestions. We have updated the Elbow curves and contain a log/log axes version. The distribution in Figure 3a indeed shows no clustering trend according to the points scattered. We appended an additional test using auto-encoder without a variational module but however gained a similarly smooth and flat distribution in latent space as shown below. (We have also appended the result to supplement Figure).



These two distributions indicate there are no significant trends of multiple feature clusters in waveform shape features. Thus, we consider modifying the goal of clustering only to separate the continuous latent space into different parts, and later utilize the separated cluster to isolate similar reflector waveforms for detection.

Figure 3b

It might be easier to read and interpret this figure if you used a 10x10 grid of virtual waveforms instead of a 20x20 grid, and then made the individual waveforms twice as large. Additionally, it might be better to make the individual waveforms all black, and then overlay cluster boundaries as lines.

Reply: We have modified the figure according to your advice. Considering the black color corresponding to the subglacial lake in the clustering colormap, which may fuse the waveform with the stacked cluster boundaries, we plotted the waveforms in white instead.

L185-192: Identifying subglacial water using clustering analysis

It seems as though your major reason for using clustering analysis was to get to this step, where you use your method to automatically identify water. The basic argument you are making here seems to be that the upper right quadrant of Z1,Z2 space contains symmetrical sharp reflectors, and these reflectors are more likely to be water. This argument is simple, robust, and I believe it. But in addition to identifying points in this quadrant using a clustering analysis with an arbitrary number of clusters, it may help to make your analysis more robust if you also constructed alternate metrics to identify points in this quadrant. For instance, you could select traces for which Z1 and Z2 are both more than 1sigma above the mean. Or you could make a combined water index, I=Z1+Z2, and then select points with a high value of this index. Alternatively, you could construct a not-water index by taking the euclidean distance from each point to the upper right corner (+2sigma,+2sigma). These sorts of continuous water indices may help reduce the dependence on an arbitrary choice of K.

Reply: We consider your suggestions highly reasonable, as this approach provides a quantitative assessment of feature similarity. However, certain issues may exist when applying based on the current method, primarily related to the order of steps. In my opinion, the new strategy potentially consists of the following steps:

(1) Determine the centroid in the latent space corresponding to subglacial lakes, based on the detected subglacial lake list.

(2) Assess the certainty of subglacial lakes by calculating the distance/index from the centroid, and apply the judgment to all reflectors.

(3) Detect subglacial lake and record the subglacial lake boundaries based on the extracted reflector locations.

Notably, Step (1) requires the existing distributions of subglacial lakes (requires start/end point to utilize all reflectors for encoding), which can be obtained by isolating corresponding subglacial basal reflections using known subglacial lake ranges (if the ranges of the lake given). However, existing subglacial lake catalogs typically provide only the location and size of the lakes, resulting in only one data point available for each regional lake. Conversely, Step (3) can conveniently provide a reasonable subglacial lake distribution range, allowing us to determine the average centroid and boundaries for Step (1).

This new strategy can be implemented based on the distributions and ranges of subglacial lakes provided by the current clustering method. Thus, we consider that this new strategy will be a potential improvement to our current method. Especially, when widely applying the encoding method in updating the known lake inventory, the new strategy could provide valuable estimations of subglacial reflectors from different regions.

We greatly appreciate your suggestions, we will make an attempt for this potential improvement as the next step for our current method.

L199: “Detected subglacial water bodies should contain a continuous ice bottom segmentation in subglacial water type with a width greater than 8 traces (corresponding to an average spatial distance of 10.4 m).”

You should double-check the along-track spacing of the data product you use. If the data have been downsampled from the original 1.3 m spacing, then your 8 trace threshold will correspond to a longer distance.

Reply: Thanks for your indications. We have modified the spatial distance to 112 m.

Figure 4c:

This plot would definitely benefit from a continuous approach to reflector categorization.

The different colors here represent different categories, but it is hard to tell how close each category is to the water category. By contrast, a “water index” would provide a continuous metric that could be displayed here.

Reply: Thanks for the indication. We have updated Figure 3b to demonstrate the adjacent relation of different clusters, which provides references for this Figure. The "water index" indeed could provide a direct estimation for the similarity of waveforms. We will consider this new strategy in further studies.

L204-218, Figure 5: Reflection power analysis

Did you correct the bed returned power for geometric spreading before doing this analysis? Signal loss with depth comes from both attenuation within the ice and from simple geometric spreading with range. The effect of geometric spreading can be calculated and removed.

Additionally, I am curious whether Figure 5 shows the entire dataset, or only a subsample of the dataset? When I did a similar analysis for the 2013 paper, I found many data points that were 3sigma or even 4sigma above the linear best-fit. However, in this figure it looks like you would have perhaps 2 or 3 data points at a 3sigma level, and no data points at a 4sigma level. Is the total sample size smaller here? What exactly is being plotted in Figure 5? Does each point represent a single trace, or does each point represent an along-track average of candidate water bodies? I feel like this method needs a better explanation.

After thinking about it for a bit, my guess is that you have done something like this:

1) Apply clustering analysis to the traces

2) Apply the 8-trace rule to generate contiguous reflectors that all belong to a particular cluster.

3) Compute average peak power for each contiguous reflector

4) Perform the attenuation analysis using this smaller dataset of horizontally averaged power data

Am I correct? Is that the procedure that you followed? If yes, then this should be explained in more clarity. In particular, it should be clear that you used grouped adjacent reflectors according to their cluster, and that you have done this for all clusters, not just the water cluster. If this is what you have done, then that also explains why you find far fewer high-reflectivity outliers than I did in the 2013 paper, since peak reflection power can be highly variable along-track and the averaging process will tend to reduce the amplitude of individual bright spots.

It also seems to me that the arbitrary choice to use 15 clusters will have a big impact at this stage, since it will determine the along-track length of contiguous regions that you average together into the analysis. It would be interesting to see in the sensitivity analysis how changing the value of K affects this part of the analysis.

Reply: The plotted data points were randomly chosen from all the AGAP-S radar images because the full plotting of the sample points will fill the figure into black or other pure color selected. Thus, we reduce the number of data points used in this figure, which perhaps has also reduced the number of sparse points above +3/4 sigma above the linear best fit. We were also puzzled by the difference between this figure and your 2013 paper, but now we guess that maybe the difference in these figures is caused by the different preprocessing of different raw data sources. In this figure, we did not correct the bed returned to power for geometric spreading, because the data download from CReSIS seems to contain no enough information and parameters for correction (actually we are not sure if they were already corrected in preprocessing). Thus we applied +1 sigma, which lowers the threshold in your 2013 paper to ensure a larger tolerance to the uncorrected signal.

The reflection power is directly extracted from the radar image data after 10*log_10(X) processing. So the sparse sample points are applied without averaging in clusters or windows. We apply the average return power as follows: When this threshold line was applied in the subglacial lake detection, the average return power of each detected lake range is calculated. If the average return power is higher than the depth-corresponding threshold, the lake will be collected and recorded in the list.

Figures 4, 6, 7, 9, 10: Radar results figures

These figures would all benefit from being zoomed in on the bed. The vertical scale

could be cropped between 2 km and 4km (or perhaps slightly below 4km, to accommodate the deep lake in Fig 9c). In addition, the color scale of the echograms should be adjusted so that the lower limit is just a bit below the noise floor and the upper limit is closer to the brightest bed. These changes would make it easier to follow along when the text goes into detail about specific features in these figures.

Reply: Much appreciate your suggestion for enhancing the display. We have updated the color scales and vertical scales in these figures. The modified version indeed highlights the feature of subglacial lakes in radar images.

L260 "...but some are also sparsely detected."

This wording is awkward. Perhaps, "...but some isolated points are also detected."

Reply: Done, Thanks.

Figure 8: map figure

It is hard to tell what the text labels (L#, E#, N#) refer to. Maybe you could move the text labels further away from their targets, and then add annotation arrows pointing from the text to the target?

Additionally, the main map should be bigger and the other elements of this figure should be smaller. There is way too much empty white space in this figure. The main map containing the central AGAP survey contains all of the important information in this figure. Therefore, that main map should be as big as possible. The other two items in the figure, the inset location map and the legend, can be placed in unused corners of the main map.

Reply: Thanks a lot for your advice. We have modified the map according to your advice.

L255-268: L#, E#, N#

What do L, E, and N stand for? Are N1-N4 new subglacial lakes?

Reply: Yes, we have added more details in the context about the meaning of L#, E#, N#.

L 278: "Considering the dense distribution of subglacial water bodies nearby, these thicker reflection features are possibly formed by frozen-on ice due to ice flow."

Freeze-on isn't caused by ice flow. Freeze-on is caused by either conductive cooling or supercooling. Perhaps a better way to phrase this sentence would be, "Considering the dense distribution of subglacial water bodies nearby, these thicker reflection features are possibly formed by frozen-on ice that complicates the shape of the near-basal reflection trace."

Reply: Thanks for your indication, we have amended this sentence.

L281: "...the sparse but regionally dense distribution..."

What exactly does "sparse but regionally dense" mean?

Reply: Here we would like to indicate the regionally dense distributed subglacial lakes. We have removed "sparse" and modified the sentence.

L309: "The unsupervised clustering analysis applied in the latent vectors relies on the implied feature difference of the reflection waveform, effectively excluding subjective and external factors in finding potential classifications of subglacial conditions, and reducing the dependence on model assumptions."

Except for the subjective choice to use 15 clusters. This choice has downstream effects in terms of determining the size of the “water” cluster (because average cluster size should vary inversely with the number of clusters), so this arbitrary choice indirectly determines how much variability in reflector shape you are willing to tolerate while still calling something “water”. Additionally, the choice to use a 2D latent space instead of a higher dimensional space was also arbitrary. All methods require some degree of human choice on the part of the scientists employing the method.

It seems to me that the big advances achieved here are in 1) having a new method to quantify and classify the shape of the reflection waveform, and 2) using that method to help classify the physical setting of the ice sheet bed, particularly by helping to identify subglacial lakes. It is not really fair to say that you have excluded subjective and external factors, those factors simply enter into your analysis in a different way than they do in other analyses.

Reply: Much appreciate your indications and suggestions. We have modified the discussions here according to your suggestions.