Response to Veronica Tollenaar (RC2)

General comments

The paper discusses a subglacial lake detection method applied to a region near the center of the continent of Antarctica. With the available data, the problem can be seen as a positive and unlabeled problem, where some subglacial lakes have been outlined in earlier studies (positive labeled examples), while for the remaining area the presence or absence of subglacial lakes is unknown (unlabeled examples). The authors take an unsupervised learning approach to this problem, which is a valid choice.

The unsupervised learning consists of an auto-encoder, which basically reduces the dimensionality of the data, and a clustering, where one of the clusters is assumed to correspond to the presence of a subglacial lake. Although this approach is smart, novel, and has a high potential in delineating subglacial lakes, I see several weakly motivated choices in the methodology that I will also try to outline further through the specific comments per section.

My main issue is that the authors perform a clustering analysis on a (2-dimensionally) normally distributed set of samples. These samples are normally distributed through the applied loss function in the encoder. However, per definition, in this set of samples there is only a single cluster, otherwise the loss function should have allowed a certain number of gaussian distributions in the latent space. This caveat is also confirmed by the fact that there is no clear cutoff point in the elbow function to determine the number of clusters present in the data. In my view there are three potential approaches to adjust the manuscript to overcome these caveats in the methodology.

(i) The authors can illustrate quantitatively that the results are convincing, despite the conceptual problem with the methodology, making the study a pragmatic approach toward subglacial lake detection. With the absence of correctly labeled negative examples (i.e., the absence of subglacial lakes), traditional performance metrics such as precision and accuracy cannot be estimated. Nevertheless, a sensitivity estimate of the results, which is currently not part of the manuscript, can be included.

(ii) Instead of the clustering, the authors can identify where the currently known subglacial lakes are located in the latent space (i.e., plot these samples in Figure 3a). As “the distance between vectors in the latent space can serve as a statistical similarity indicator for reflector features” (Line 308-309), samples within a certain distance from the located latent-space vector of known subglacial lakes can be identified as subglacial lakes.

(iii) The authors could use another approach to deep clustering as discussed in various deep learning literature. The simplest solution would be to use an auto-encoder instead of a variational auto-encoder, despite obtaining a less meaningful latent space in the sense that the distance between latent vectors does not reflect a similarity. Nevertheless, it might appear that there are distinct clusters in the latent space.

I think that through adopting (a combination of) the above approaches, or by taking another approach that overcomes the illustrated problem, the study can significantly contribute to the development of an automated approach for the detection of subglacial lakes. This method will be essential to process the ever-growing amounts of data across the continent (and beyond) efficiently, and the authors already convey this message clearly through an elaborate discussion of their results and informative figures.

Reply: Much appreciate your encouraging comments and valuable suggestions. We have updated the manuscript according to your concerns, as the following points:

(i) We have appended the clustered areas in latent space corresponding to subglacial lakes when different K value is applied in clustering analysis. We also traced the detected ranges of subglacial lakes in different K values applied.

(ii)We trained another auto-encoder which contains no variational module, and used the same reflector samples as Figure 3 to exhibit the latent space distribution. We have appended an additional comparison between VAE and Auto-Encoder on the same samples' latent space distributions and their probability density estimations.

For advice (ii), we agree that locating (mapping) the known subglacial lakes from the latest lakes list in latent space could provide a more reasonable indicator for the newly detected lakes. However, the known subglacial lakes list only implied the location of each lake. The absence of widths/ranges of known subglacial lakes makes only one trace of the reflector that can be extracted based on the longitude and latitude, which may also induce few known indicators that can be utilized in latent space and influence the detection. So, we consider this known lakes-supervised method to be a potential application for future study.

Thanks again for your detailed suggestions to benefit our work.

Specific comments per section

Title and abstract

Title: I think “Subglacial Radar Reflectance” sounds better than “Radar Subglacial Reflector”. Also, apart from a very elaborate qualitative analysis of the results, there is no hard or independent evidence that the detected lakes are really lakes, let alone that they are “new”, which implies that they were not there before (in time). Leaving the word “new” out of the title solves this issue. Otherwise, rephrasing toward something like “An automated method for subglacial lake detection based on deep clustering” could be nice, but it depends on the intention of the authors.

Reply: Thanks for your advice on the title. We have modified the title according to your advice.

Line 3: It is confusing to read that you generate a dataset. Maybe better to rephrase as “In this study, we use available IPR images in the Gamburtsev Subglacial Mountains to extract one- dimensional reflector waveform features of the ice-bedrock interface.”

Line 4: The method remains very mystical, maybe good to clarify that you apply a deep learning method to reduce the dimension of the data so that you can perform a cluster analysis.

Reply: Thanks for the indications. According to your advice, we have modified and simplified these sentences.

1 Introduction

Line 13: The sentence does not read well. I would suggest: “Subglacial water, i.e., water between bedrock and ice sheet, is formed through a complex interplay..”

Reply: Done. Thanks.

Line 15: Potentially also include the recent publication of Kazmierczak et al. in The Cryosphere:

E. Kazmierczak, S. Sun, V. Coulon, F. Pattyn, Subglacial hydrology modulates basal sliding response of the Antarctic ice sheet to climate forcing. The Cryosphere, 16, 4537–4552 (2022).

Reply: We have added this citation.

Line 16-20: The importance of research in subglacial lakes is well outlined, but the order is a bit confusing. I would start with the ice sheet meltwater (following the previous sentence about ice flow and dynamics), then the history of climate change and ice sheet evolution, then the subglacial lake sediments, then the unique lacustrine ecosystems.

Reply: Thank you for your helpful advice, we have modified the order.

Line 21: Potentially write out the acronym of radar (radio detection and ranging).

Line 21: Potentially remove “in recent years”, the next sentence refers to a publication of 1973.

Reply: Done. Thanks.

Line 22: The sentence starting with “Subglacial water bodies” could fit better in the next paragraph, where these visual features are discussed again.

Reply: We have moved this sentence to the next paragraph, thanks for your advice.

Line 23: I would swap around the subject and the object of these sentences so that it is easier for the reader to understand that here the authors are going to refer to other measurement techniques: “The thickness of the subglacial water layer and sediment characteristics at the bottom of lakes are also investigated with active seismic surveys

(Paden et al., 2010; Arnold et al., 2020) and gravimetry and electromagnetic methods (Studinger et al, 2004, Key and Siegfried, 2017).”

Reply: We have swapped these sentences to the improved version. Thanks a lot for advising.

Line 35: the “subjective factors” are not ruled out in this study: heavy postprocessing is applied and the results are discussed mainly in a qualitative way.

Reply: We have removed "subjective factors" in this sentence.

Line 36: the “absence of a complete interpretation of basal radar reflectance features” is also the case for the study: only a narrow window including the reflectance near the bedrock is considered, and the spatial context, i.e., along the bedrock, is only considered through a rather pragmatic postprocessing step that filters the results spatially. Deep learning is a powerful tool to consider these spatial relationships directly. If not adapting the methodology to actually rule out “subjective factors” and have a “complete interpretation of basal radar reflectance features”, I would suggest a more elaborate and precise discussion of other methods, to illustrate more in detail in which aspects the proposed methodology is better.

Reply: Thanks for your suggestion. We have appended precise discussions of each conventional method, and have modified the context about the advantages.

Line 37: I would suggest an easier rephrasing: “In recent years, deep learning has been applied as a powerful tool to detect different features in IPR images, including bedrock interfaces, internal ice layers, snow accumulation layers”. For the “radar semantic segmentation”, that is an automated feature extraction in se, so I’d suggest to either refer to what is semantically segmented or remove.

Reply: We have updated these sentences to according to your suggestion.

Line 40: I am not sure if I understand the difference between this sentence and the previous: is the previous specifically about the detection of layers? If not, I would try to combine this sentence with the previous one and specify the subglacial features. For me it is not clear whether the subglacial features refer to anything under the surface or just features at the ice- bedrock interface.

Line 42: I would rephrase this sentence with: “Moreover, deep learning applied to IPR has also contributed to estimates of ice thickness (to enable data application in ice sheet studies.)”, with the part in brackets potentially removed.

Reply: We have combined and simplified these sentences according to your helpful advice.

Line 46: Potentially include a reference to the dataset directly (see: https://data.cresis.ku.edu/#ACRDU)

Reply: Done. Thanks.

Line 50: I think it is a bit confusing to use the wording “construct a dataset”, it suggests that you collected the data in the field. I suggest the rephrasing: “In this study, we select IPR images in the region of the Gamburtsev Sublgacial Mountains from the CReSIS database. We crop these images around the ice bottom, to obtain a set of one-dimensional waveforms that capture the ice bottom reflectance characteristics. Using this data, we train ...”

Reply: Thank you for your suggestion, these modified sentences read much better.

Line 52: The “time-domain waveform features” are confusing. Either introduce the time-domain aspect in an additional sentence (something like: “The radar is reflected most strongly by the bedrock beneath the ice sheet, resulting in a peak in the return signal received by the radar over time. Moreover, bedrock characteristics, such as roughness or the presence of water, influence the intensity and shape of the peak signal, to which we refer to as the waveform features of basal reflectors.”)

Reply: Thanks for your indication. Here we modified the "time-domain waveform features" to "one-dimensional waveform features" to contain the continuous logic with the previous sentence.

Line 55: Do you mean the features that correspond to subglacial lakes? Line 55: I would specify that this is a kind of post-processing step.

Reply: We have added the specific subglacial lake feature in this sentence. Thanks.

Line 58-60: What is the benefit of extracting reflectors with similar waveform characteristics as water bodies? How does that improve the efficiency and accuracy of the detection of subglacial lakes?

Reply: We have separated this sentence into two parts and introduced the benefits of efficiency and accuracy separately. Thanks for the indication.

Line 61: Indeed, it is nice that you can characterize/cluster the subglacial features through this method.

Reply: Thanks.

2 Data and Methods

Figure 1: The Figure looks nice, and summarizes the workflow well, but there are several details that need to be adjusted: What is “Z-Scope”? What is “A-Scope”? “Ice Buttom” should be “Ice Bottom”, “Reconstructed Reflector Feature” should be “Reconstructed Reflector” (as in “Ice Bottom Radar Reflector”). Both waveforms need axes with labels (time and power I guess). For the caption “(b) VAE reconstructs and encoding of the sampled ice bottom reflector features.” should be changed to “(b) The VAE encodes and reconstructs the sampled ice bottom reflector.” For the subpanel (c), the caption says “Supervised”, while I think the authors mean “Unsupervised”.

Reply: We have modified both the figure and caption according to your detailed indication and helpful suggestions. Thanks a lot.

Line 69: This sentence about the lake inventories seems out of place. I think, together with the sentence “According to the lakes inventory...” on line 71, these sentences should be moved to the introduction in the paragraph that starts on line 50, so that paragraph 2.1 really focusses on the radar data.

Reply: We have moved this sentence to the introduction (line 50). This modified version is indeed better. Thank you for the suggestion.

Line 70: I miss a reference here: is it this dataset that’s been used? https://data.bas.ac.uk/full-record.php?id=GB/NERC/BAS/PDC/01544

Reply: We have added this link in this sentence, thanks for indication.

Line 74: “The radar data were acquired from L1B..” can be rephrased to “We use the L1B data product” to avoid confusion whether the data has been acquired by the authors.

Reply: We have modified this sentence.

Line 81: Is there a physical motivation for truncating the signal to this narrow range around the bedrock? When I see the radar images shown in the different Figures (e.g., Figure 4), I find it remarkable to see a distinct reflectance below the bedrock for each of the subglacial lakes that seems to be not captured anymore by choosing the narrow window.

Reply: The selection of time window width for truncating the signal is applied based on the experience. We did notice there are some distinct reflectances below the subglacial lake interface reflections, but some subglacial lakes from the known inventory (e.g., the left lake in Figure 9a) do not contain this specific feature. Therefore, we apply a narrower time window to reduce the sensitivity of this additional reflectance. We have appended more description of the motivation for the window width chosen here.

Line 85: Assuming that the peak signal corresponds to a single point, I would guess the length of the truncated signal would be 64 + 1 + 64 = 129, but it reads 128.

Reply: Thanks for the indication. We have modified the range to "-64 to +63". The length of 64 is utilized in the raw programming code, in which the index starts from zeros.

Line 88: Could you provide the bandwith/sigma of the gaussian kernel?

Reply: We have appended more details about the gaussian kernel. Thanks for the indication.

Line 89: How do you perform this normalization? Somehow I get the impression that all of the nearly 1,5 million (incredible number, congrats!) reflectance traces are normalized individually: or do you calculate a global mean and standard deviation and set these to 0 and 1? If normalized individually, I think this might be the cause of why you need to use the post- processing step where you use the peak power reflectance. I would advise to

either (i) normalize all data with the statistics of the entire dataset as otherwise you're comparing different units to each other, or (ii) already implement the depth/power relationship while normalizing, or (iii), more experimental, normalize each individual waveform, but provide the peak power and the ice thickness as additional input to the VAE.

Reply: Yes, the normalization is applied in every single waveform trace. In the early phase of our method concept design, we considered the strategy of all data normalization as you mentioned. However, the VAE failed to learn the waveform in this situation. The potential reason is 2*1 bottleneck was too small to reconstruct the waveform feature consisting of both the waveform shape and dynamic ranges. Thus, we applied single-trace normalization here to simply feature by excluding the dynamic ranges of echo power. According to your indication, we have appended more details about the normalization and its corresponding function in reducing features.

Line 97: What do you mean with the sentence starting with "And the.."? I think it deviates the attention from why you use the VAE: to reduce the dimension of your data.

Reply: We agree that this sentence is redundant here and have removed it. We used this sentence to explain the specific feature of VAE, but it seems useless in this paragraph.

Line 102: I think you use it to reduce the dimension of the reflector waveform features from the ice bottom, right? It is confusing to think that the goal is to reconstruct something that you already know.

Reply: We have modified this sentence to match the final goal of our VAE application. Thanks a lot for your indication and advice.

Line 104: Your bottleneck consists of a two-dimensional latent distribution, enforced to follow a normal distribution through using the KL divergence in your loss function. I find the motivation for choosing to sample only two samples from your latent distribution just for visual representation weak. Another motivation can be that it is easier to perform the clustering in two dimensions, or that in other work it has been proven sufficient (for example in the referenced work of Li 2022).

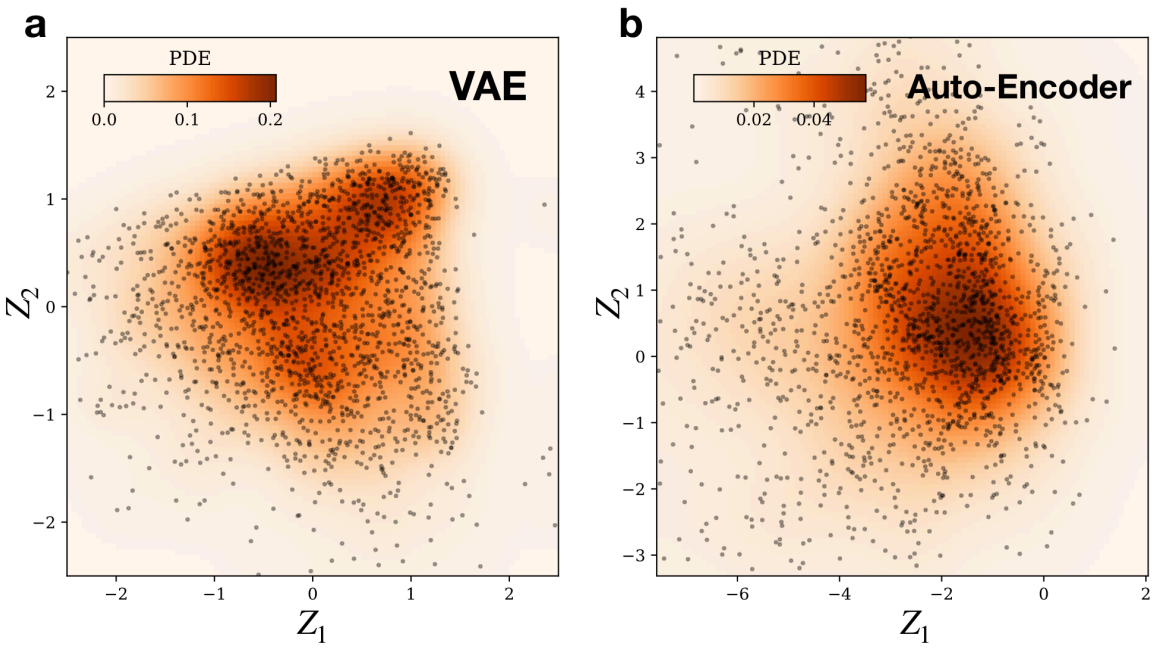
Reply: Thanks for your advice for updating the motivation of the 2-D latent space application, which indeed we think was weak before. We have appended more descriptions here according to your suggestion.

Line 106: Conceptually I don't understand why the KL is used in the loss function: it forces the latent space to be normally distributed, which is essential when using VAE for generative purposes. However, as the authors want to perform a cluster analysis, I think there is a fundamental conflict. Clustering data that is normally distributed will not yield in clearly separable clusters. Or, differently put: the underlying assumption for clustering should be that there a different clusters, which, of course, can be each normally distributed, but through VAE the latent space is constructed as one single big cluster. The fact that there is no clear cutoff point of the elbow curve that the authors want to use to

determine the number of clusters confirms that there are no separable clusters in the latent space. I have not read enough into the literature to know whether there are other examples of the approach that the authors take that still yield useful results – but a quick search indicated that there are fancy solutions for this mismatching of concepts, e.g., Lim et al., 2020. A simple solution would be to just use an Auto Encoder and perform the clustering on those results.

Lim, Kart-Leong, Xudong Jiang, and Chenyu Yi, Deep clustering with variational autoencoder. IEEE Signal Processing Letters, 27, 231-235 (2020).

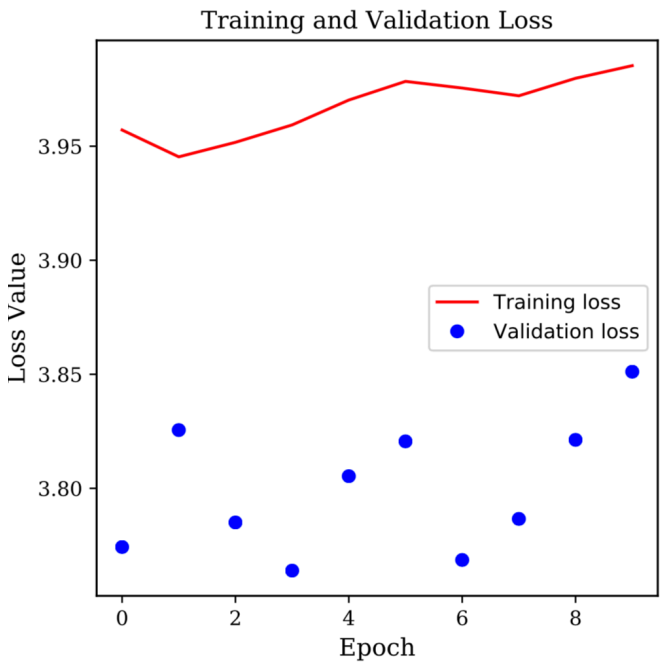
Reply: We agree that the context here indeed confusing due to the conflicted motivation when using KL in loss function but applied clustering later in the dimension-reduced latent space. The goal of the VAE application in this study is to obtain a continuous-presenting latent space so that we can generate synthetic reflector waveforms (as shown in Figure 3b). After clustering, we can directly choose the cluster that covers the latent space corresponding to the subglacial lakes' feature. This goal is also the motivation we would like to exhibit Figure 3b. Thanks for your suggestion. We tested the auto-encoder without variational models and KL in the loss function. The distributions of the same data samples in different auto-encoders are shown below:



Similar to VAE's distribution(a), the result of auto-encoder(b) does not show a distinct trend of the cluster in the latent space distribution. These two distributions indicate that the waveforms may contain no potential clusters by the feature presenting. We have appended this comparison to supplyment Figures, and have modified the confusing description which may cause conflicted goals in the VAE application.

Line 122: Why do you stop training at epoch 10 if the training loss does not descend more after epoch 4? Can you report the generalization error? If the training loss does not decrease, but you continue training (epoch 5-10), you start to overfit to your training data.

Reply: Yes, because we noticed the potential overfitting, the final model we applied in the encode and generalization is from epoch 4. We have appended the training loss curve until epoch 4 of both the training and validating datasets in Supplemental Figure S1(unfortunately, we lost the loss information between epochs 5-10). Thus, we additionally repeat the training of VAE using the same dataset until epoch 10 to demonstrate the potential overfitting as shown right-side. We have modified the descriptions of the epochs in training. Thanks for indication



Line 123: The word “evaluate” suggests a quantitative estimation, for example based on independent test data. Could you either provide this, or change to “illustrate”?

Reply: We have changed the word "evaluate" to "illustrate" here. Thanks for the indication.

Figure 2: Could you provide axes and labels for all subpanels? Could you provide the MSE for all examples? Potentially the learning curve, and the generalization error could be included in this Figure.

Reply: Thanks for your suggestion. Because of the dynamic normalization, in all the subpanels' vertical axis is unified to 0-1. Here, we would like to better exhibit the waveform difference between the raw and reconstructor, so we simplified the axes and label. According to your suggestion, we have appended more description to the caption of Figure 2 and detailed the MSE value for all examples (Due to the missing label of raw data, we have replaced the examples in the raw version with similar waveforms).

Line 139: These vectors consist of two samples from the latent distribution, right?

Reply: Yes. We have appended 'from two reflector samples' in this sentence.

Line 143: How does this subset vary from the validation subset mentioned in line 119? It seems like you are going to use these samples for clustering and not for “validate the encoder”?

Reply: Yes, we use these samples for clustering, instead of validating the encoder. We are sorry for the mistake and have amended this sentence.

Line 147: That gaussian distribution poses problems for the clustering (see earlier remark about line 106).

Reply: We have appended more descriptions about the gaussian distribution and the motivation of the clustering applied.

Line 148-153: This is almost philosophical, could you rephrase it with more direct wording?

Reply: We have modified this sentence. Thanks for indicating.

Line 156: I do not directly see that 2000 reflectors are sufficient for clustering. From Figure 3, to me, the clusters seem rather arbitrary. Also, given that you have 1.5 million reflectors and you perform the dimension reduction to enable efficient clustering, I think the sample of 2000 is rather small (~0.1 % of all data). How long does it take to perform the clustering analysis?

Reply: Thanks for your indication. The clustering analysis takes about 20s. We have used a larger dataset of 0.1 % of all data for clustering according to your advice. However, because of the difference between the data applied, the region of each cluster show a slight difference in latent space, which could impact all the following results and requires huge works on replotting figures and map. Therefore, we consider appending an additional comparison of the detected range of subglacial lakes in same radar image in Figure 11. We agree that using a larger amount of samples can provide more reasonable clustering results. We have appended more description about the comparison in detection ranges and potential improvement in the discussion.

Figure 3: The generative capacity of VAE is nice, and Figure 3b is a pretty visualization of this capacity. However, I do miss a link to the physical phenomenon, and therefore I would suggest to remove the subfigure or move it to Supplementary Materials.

Reply: Thanks for your advice. The purpose of this subfigure is to demonstrate the shapes of waveforms corresponding to the vectors in different clusters of latent space. Based on these reconstructed waveforms, we can select the cluster of reflectors which visually similar to the ice-water interface. Thus, we think this subfigure is relatively necessary here. This subfigure also provides a potential reference for the reflector waveform in the subglacial lake cluster (in black color). We have appended more descriptions about the purpose of this subfigure. We have also appended additional color blocks in the background to demonstrate the boundaries between different clusters.

Line 169: Here I miss evidence for the statement: what motivates the authors to conclude that there is an effective separation of bottom reflector features? And how do they correspond to different conditions?

Reply: We agree that the statement is missing here. We have modified the conclusion of the effective separation of bottom reflector features. We have removed this arbitrary conclusion and modified the sentence.

Line 171-183: Similar to Figure 3b: a physical interpretation is lacking, and I would move this to Supplementary Materials.

Reply: Thanks for the advice. We have modified Figure 3b, and have appended an additional background color block to indicate the boundary between different clusters. We have also modified the descriptions here.

Line 184: In this section the authors discuss how to detect subglacial lakes using the results of the clustering analysis. The main points discussed are related to post-processing steps, and I think this is not clearly reflected in the section title. Potential other titles could be “Subglacial lake detection” or “Post processing to detect subglacial lakes”.

Reply: Thanks for your advice, we have modified the title of this section to "Subglacial lake detection".

Line 192: I do not understand the conclusion here. I guess you want to say that one of the clusters seems to correspond to subglacial lakes, right? Another way to confirm this is to give statistics of to what clusters the waveforms at earlier detected subglacial lakes belong, e.g., 80% of known subglacial lakes have a bottom reflector that falls into cluster x.

Reply: Sorry, maybe we missed the position of these sentences due to the mismatching line numbers. Maybe the paragraph that causes your confusion is the first paragraph in this section. We agree with your opinion here. However, as we mentioned in the reply of major concern, only points of location subglacial lakes are provided by the inventory, which causes difficulties in tracing the reflectors corresponding to the known lakes. According to your advice, we have modified these primary conclusions.

Line 198: What do you mean by “based on experimental experience”? Is there a reference? A solution could be to remove that specification.

Reply: Thanks for your advice, we have removed this sentence. The experimental experience was from the final result analysis after this step. We filtered the small subglacial lakes with a threshold on the lake range and compared the result with the known lake inventory. After multiple attempts, we finally chose this value.

Line 203: What do you mean by “interpolation artifacts due to specific noise?”

Reply: Thanks for your indication of this redundant description. We have modified and simplified this sentence to "mistaken detection caused by abundant interpolation"

Line 209: If I understood it well, before you used this peak echo power to normalize the data for the encoder. I wonder if this postprocessing step would still be necessary if don’t apply this normalization earlier. That would potentially be something to investigate and report on.

Reply: We agree that there was potential content that needed to be reported. We did apply the raw signals without normalizations in the VAE training. However, the VAE failed to reconstruct the input signals. The potential reason is the raw signals before normalizations contain more features (especially for the peak echo power), so it is relatively more difficult to reconstruct by an auto-encoder with a smaller size of latent space (bottleneck). Therefore, we applied power normalization for all the reflector waveforms before VAE training. We have applied more descriptions about the motivation of normalization in the VAE section, as well as the peak echo power postprocessing. Thanks for your advice.

Figure 4: For panel d, would it be possible to have the same colors as panel c? So black for the lake, and other colors corresponding to the different clusters that have been filtered out during the post processing?

Reply: Thanks for the advice. We have modified these figures by changing the colormap on panel d.

Line 211: How did you calculate the best linear fit? Somehow, I get the impression that the orange dashed line should be steeper in Figure 5, but this might be an optical illusion.

Reply: We used LinearRegression module from scikit-learn toolkit in Python. We did notice the mismatch of the steep on fitting, which we considered as the algorithm difference between linear fitting and probability density estimation.

Figure 5: Potentially only show the +1 sigma as that's the threshold you use, to avoid confusion.

Reply: Thanks for your advice, We have modified this figure and removed the dashed line of +2 sigma.

3 Results

Line 229-230: If I understand it correct you are claiming that the results are reliable because the subglacial water bodies look like known subglacial waterbodies, right? Out of interest, what do you mean by the geothermal environment in adjacent areas?

Reply: Thanks for your indications. In this sentence, we would like to describe that the geothermal and subglacial environments should be similar in the same radar image, which was continuously recorded in adjacent areas. We have modified this sentence to be more readable.

Line 237-240: This statement is very similar to the statement in the previous paragraph. I think you do not need to convince the reader of the value of an automated method for detection, it is already clear that this is very valuable.

Reply: We have removed this redundant description about the automated advantages. Thanks for your advice.

Line 241: I think it should be “(at about 40 km along the transect)” or so, it looks like the lake is ~3 km wide.

Reply: Thanks for the suggestion. It did look better after modification.

Line 241-253: Nice discussion of results.

Reply: Thanks.

Line 255-260: Somehow this paragraph makes me doubt that for the results in Figures 4, 6, and 7, the peak power post-processing step is not applied? Could you clarify that in the text?

Reply: The post-processing of the peak power threshold helps to obtain the weak reflections. Results in Figures 4, 6, and 7 show strong reflections and are therefore validated in this post-processing. We have appended more context about the post-processing step. Thanks for your advice.

Line 260: By “sparsely detected”, do you mean that these are isolated lakes? Or just along a single IPR line?

Reply: Yes, we have modified this description.

Line 261: Normally it should be “compare to something”: rephrase as “We compare the subglacial lakes detected in this study to the previously identified ...”

Reply: Thanks for the indication. We have amended that.

Line 265: remove “which is newly detected”, that is already clear from the first part of the sentence.

Reply: Done.

Line 277: Do you mean that the red arrows show lakes that have not been detected?

Reply: In this context, the red arrows indicate other continuous reflector features within the same cluster, though they do not correspond to the subglacial lakes cluster. We noticed the context near this sentence may cause confusion. Thus, we have modified this part, separated this sentence into a new paragraph, and added more description about this radar image.

Line 278: In Figure 7c you associate the yellow cluster with frozen-on ice and ice flow dynamics. But in Figure 9 it looks like different shades of purple. Do you think multiple clusters do show this frozen-on ice? And are these clusters next to each other (it’s hard to link the shades of purple in the Figures with the shades of purple in Figure 3a).

Reply: Yes. We consider that different clusters (which appears continuously in radar reflectors) may correspond to different phases or situation of frozen-on ice. However, the relations still need further studies and field observations. We have appended a color block in Figure 3b to demonstrate the adjacent relation of different clusters. From the Figure 3b, we notice these clusters are next to each other.

Line 280: I think the origin of the water bodies is very suggestive. What do you mean by the sparse but regionally dense distribution of subglacial water bodies?

Reply: We have removed this confusing description of the "sparse but" and modified that to "the regionally dense distribution of subglacial water bodies". Thanks for the indication.

Figure 8: I think the Figure is very essential for the study. It took a long time to understand the link between the regions and the labels, but I understand now that it is related to the thin black arrows. Potentially it would be nice to clarify that in the main text, as well as in the caption. Moreover, the two blue colors (blue and cyan), might be confusing, and the labelling can be “detected lakes (no post-processing)” and “lakes (post-processed)” or so, now it is not clear what is what exactly. Other questions that pop up when seeing the figure are: (i) in the region near “N3”, going perpendicular to the radar lines, there is a clear line of lakes, does that correspond to a kind of channel in the subsurface topography? It could be interesting to overlay the detections on bed topography data, but that is probably out of scope for this study. (ii) There are a lot of “candidate lakes” on the southern part of the survey, it almost looks like an artifact, is that the case?

Reply: We have appended more descriptions to the main text about the markers used in the map. According to your advice, we have modified this map in both color usage and caption. For question (i), we agree that will be an interesting illustration by tracing the nearby subglacial lakes in radar images and comparing them with bed topography data. There will be the next studies after our arranging of the new subglacial lake list. For question (ii), the candidate lakes on the southern part are invalidated by the echo power filtering. Flatten topography with weak reflections is exhibited in the radar image in this region, which mismatches the features of subglacial lakes. We have appended more descriptions of this abnormality.

Line 287: What do you mean by “differ visually”?

Reply: It should be "visually different from...". We are sorry for this confusing description.

Line 298-304: I think the conclusion is very bold, basically saying that the previous inventories are wrong in places where the authors do not detect lakes. I would be a bit more reserved and steer in the direction that this automated method is promising, and that further investigation is needed (as already suggested). Moreover, there is the remark about “multi-trace detection methods”, but in some sense the applied post-processing of grouping 8 neighboring traces makes this method also a “multi-trace detection method”, right? Or is this not applied for obtaining the map?

Reply: Thanks for the indication. We agree that this conclusion is too bold. We have modified these sentence, and appended more context about the automated method application in updating the lake inventory. According to your advice about “multi-trace detection methods”, we have modified the sentence and appended more details about the "multi-trace method".

4 Discussion

Line 307: I understand what you mean by “all reflection information”, but actually you crop the reflectance to contain only the signal of the bottom.

Reply: Thanks for your indication. We agree that the reflectance was cropped. Therefore, we have modified "all reflection information" to "ice bottom echo waveform information" according to your advice.

Line 308: I miss a sentence that states what has been done, something like “We encoded the waveforms to obtain two-dimensional vectors that conceptually summarize the waveform in the so-called latent space of an auto-encoder. The distance between vectors in the latent space...”

Reply: Much appreciate your helpful supplement. We have added this sentence here, which greatly improves the context.

Line 328: What do you mean by this sentence? The clustering analysis can be used as input for other models?

Reply: We agree that the usage of "data" may cause confusion. Similar to the subglacial lakes, reflectors classified in the same clusters by the analysis may correspond to different subglacial environments. Although the relations between cluster and environment are still waiting for further study. We think this primary clustering analysis can reduce the data complexity for other models.

Line 330: What do you mean by “an automated analysis data”? “automated analysis of the data”?

Reply: We have modified this context. Thank you for the indication.

Line 336: “As such, the method has potential..”

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

Line 337: What do you mean by classifications for single-track radar data?

Reply: The "classifications" here means "analysis", and "single-track radar data" means the reflection waveform from single-trace radar observations. We have modified the description and appended a citation here. Thanks for the indication.

Line 339: Sorry for the noob question: does ice penetrating radar on Mars exist? Can you obtain those kinds of observations from space? And in general, DL methods are known to perform badly on out-of-distribution examples, so is it realistic to apply the method to data that is very dissimilar from airborne observations?

Reply: Thanks for the indication. We have modified this sentence to "provide a potential reference for analyzing ...", and have appended more missing citations here. There are public data on radar-sounding observation from Mars, such as the SHARAD[1] and MARSIS[2]. Some observation tracks from orbit have covered Mars' southern ice cap[3]. Studies (e.g.,[4]) have discussed the detection of candidate martian subglacial water

bodies. We agree with the potential challenges in transferring the model on out-of-distribution examples, thus we have modified the discussion here.

[1] Seu, R., Phillips, R. J., Biccari, D., Orosei, R., Masdea, A., Picardi, G., ... & Nunes, D. C. (2007). SHARAD sounding radar on the Mars Reconnaissance Orbiter. Journal of Geophysical Research: Planets, 112(E5).

[2] Picardi, G., Biccari, D., Seu, R., Plaut, J., Johnson, W. T. K., Jordan, R. L., ... & Zampolini, E. (2004, August). MARSIS: Mars advanced radar for subsurface and ionosphere sounding. In Mars express: The scientific payload (Vol. 1240, pp. 51-69). (https://pds-geosciences.wustl.edu/missions/mars_express/marsis.htm)

[3] Orosei, R., Lauro, S. E., Pettinelli, E., Cicchetti, A. N. D. R. E. A., Coradini, M., Cosciotti, B., ... & Seu, R. (2018). Radar evidence of subglacial liquid water on Mars. Science, 361(6401), 490-493.

[4] Carrer, L., & Bruzzone, L. (2021). A novel approach to the detection and imaging of candidate martian subglacial water bodies by radar sounder data. IEEE Transactions on Geoscience and Remote Sensing, 60, 1-15.

5 Conclusions

Concise, clear

Reply: Thanks.

Data availability

Will you share your clustered data, i.e., the data in Figure 4, 6-10? Will you share your code in a repository?

Reply: Thanks for the suggestion. We are still arranging and packing the code and results. We will update the open-source information about both the data and code in this section before the final publication.

Technical comments

Line 21: introduce the acronym IPR here

Reply: Done.

Line 21: “subsurface features”

Reply: Done.

Line 37: remove (DL), acronym is not used often in the paper, and it complicates reading.
Line 42: “These deep learning-based approaches”

Reply: Modified.

Line 66: remove “reduction”, add “the” before variational auto=encoder

Reply: Done.

Line 116: “n” instead of “N”

Reply: Done.

Line 149: Brackets around Kingma and Welling 2013

Reply: Done.

Line 185: “different type’s ice bottom” should be “different types of ice bottom”

Reply: Done.

Caption Figure 4: “Fist example” instead of “Example 1”

Reply: Modified

Caption Figure 4: “Results of the unsupervised clustering of the latent vectors”

Reply: Done.

Line 220: “dataset” instead of “database”

Reply: Done.

Line 243: Remove “This subglacial ... return power”, it repeats the previous sentence

Reply: Done.

Caption Figure 7: “continuous” instead of “continus”

Reply: Done.

Line 270: “Figure” instead of “Figures”

Reply: Done.

Line 335: “covering the Arctic” can be “covering, e.g., the Arctic”

Reply: Done.

Line 356: remove “A.”

Reply: Done. Thanks a lot for indicating the technical issues above.