Title: Towards the systematic reconnaissance of seismic signals from glaciers and ice sheets - Part B: Unsupervised learning for source process characterisation
Author(s): Rebecca B. Latto et al.
MS No.: egusphere-2023-1341
MS type: Research article

Please access the discussion at:
https://egusphere.copernicus.org/preprints/2023/egusphere-2023-1341/#discussion

---

Below, the R1 comments are copied in grey. Author Comments continue in blue.

I have read with great interest the scientific article titled "Towards the systematic reconnaissance of seismic signals from glaciers and ice sheets - Part B: Unsupervised learning for source process characterisation" submitted by Latto et al., for publication in *The Cryosphere*. In this article, the authors propose to systematically explore seismic data acquired by a network of stations deployed on the Ross Ice Shelf during the austral summer of 2010-2011. The data processing pipeline, based on a priori detection (detailed in a companion paper), relies on the clustering of seismic events through the deployment of the K-means clustering method on a feature space computed from curated seismic signals features. The authors discuss the influence of feature selection and the number of clusters (one of the hyperparameters of the K-means method) on their results. They demonstrate that this approach is at least capable of revealing relatively pure clusters containing microseisms generated by stick-slip phenomena associated with the dynamics of the ice shelf. This clustering also allows for the identification of new microseismic events associated with tidal forcings.

The paper proposed by Latto et al. is remarkable for several reasons. Firstly, it is very well written and easy to follow. The literature review is particularly relevant while remaining concise. All critical information is contained within the paper, but the authors also provide a significant amount of supplementary results that address questions the reader may have. The division between the main content and supplementary content is particularly relevant. The results are convincing, and the discussion on clusters, methodology, and especially the choices of hyperparameters and features is comprehensive and very interesting. The figures are of good quality, although a few minor improvements could be made (see my comments below). Overall, this is an excellent contribution that will undoubtedly have a significant impact on communities interested in cryo-seismicity and other applications in environmental seismology. Therefore, I strongly recommend the publication of this article. However, I do have some minor comments that I will detail below.
Many thanks for this warm appraisal. We're happy to address the general and minor comments as given.

The choice of the K-means clustering method appears appropriate for the present study, and the following suggestions are by no means an invitation to completely revise this paper. I believe it is already comprehensive and insightful enough for publication as it is. However, I would like to draw the authors' attention to the fact that this clustering method may not be the most relevant for working with the features proposed by Provost et al. (2017). These features often have values distribution overlapping between each class of events, but K-means is not able to consider the "fuzzy" boundaries between different cluster. Methods

such as Gaussian Mixture Models seem to be more suitable, and I suggest that the authors at least explore this family of methods in future work.
We will add a note that we chose k-means++ due to its transparency of usage. We agree that GMM is a sensible suggestion and will incorporate this suggestion into our response to Part B R1.

The choice of the clustering method used, however, is secondary compared to the more important question of feature selection. In this article, the authors propose to reduce the number of features by calculating pairwise correlation coefficients between features. This step is absolutely necessary for K-means to work well (as demonstrated in this study). However, this requirement limits the robustness and versatility of the approach proposed. The chosen set of features may work well for this dataset, but the correlations may be different for another dataset. Furthermore, even correlated features can carry complementary and, more importantly, relevant information for discriminating certain events (in supervised approaches like Random Forest or Gradient Boosting, removing correlated features usually reduces precision scores). An alternative approach to reducing the number of features would be to reduce the dimensionality of the feature space using dimensionality reduction methods, which would eliminate the cross-correlation selection step and retain some of the information carried by each feature. I encourage the authors to consider the possibility of using methods like PCA (although it rarely works on seismic data), or even better, t-SNE (Van der Maaten and Hinton, 2008) or UMAP (McInnes, Healy, & Melville, 2018) for future work.
Many thanks for this suggestion. We'll include the information given in the discussion section.

The question of feature normalization is also critical. The authors chose to normalize by the standard deviation. This normalization helps to homogenize the overall distribution in the feature values space, but it sacrifices some level of information about each feature. The absolute value of the feature is probably as important as the position of the feature value in the overall distribution. Figure 4 presented in this paper is a clear example of this. What would your clustering result look like if you applied K-means (or another method) with non-normalized values in a two-dimensional space (Characteristic frequency - Duration) or three-dimensional space (Characteristic frequency - Duration - Peak Amplitude)? Have the authors tested their approach without normalization? Finally, why was standard deviation chosen for normalization? If the goal is to preserve the properties of the distributions of each feature, there are other normalization approaches that may be more relevant (Yeo-Johnson transform, Quantile Transform, Unit Vector Scaling, Sigmoid scaling?). I suggest testing these in future work.
We are happy to add a short note of clarification on this point.

We agree that k-means without normalisation can homogenise feature distribution and is certainly a factor to consider in terms of the high-dimensionality curse (Aggarwal et al.,. 2001). We, however, scaled the data such that each feature was not necessarily transformed to a normal distribution (Fig. 2), thereby providing critical diversity to the inputs to our k-means++ unsupervised clustering. We agree that in some cases, clustering algorithms that are more predictive (e.g. that in Provost et al., 2017, a key reference in the m/s) can benefit from a different treatment of features.

Aggarwal, C.C., Hinneburg, A., Keim, D.A. (2001). On the Surprising Behavior of Distance Metrics in High Dimensional Space. In: Van den Bussche, J., Vianu, V. (eds) Database Theory — ICDT 2001. ICDT 2001. Lecture Notes in Computer Science, vol 1973. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-44503-X_27

In the first stages of the m/s development, we did try k-means without normalization. But, because of the large magnitude differences found that results were skewed such that limited information about glacier processes were discernible (e.g. in one case clusters were only bounded by energy levels, which were features of the highest magnitudes between 10^4 to 10^10). We then chose standard deviation as a transparent approach to scaling features, but are happy to note an example of the other suggestions in the discussion. Particularly, the quantile transform would be an interesting method to consider.

Minor Comments :

L.102 : A missing space “fracture(Hammer) [...]” Will be corrected.

L.163-164 : This needs some clarification : you computed the median, the mean or something else of the values of the features of the seismic signals recorded at each station for a given event ?
We agree that this line requires clarification and citation. Feature calculations are more explicitly expounded upon in the companion paper (Part A: Latto et al., 2023a) and in our reference work (Turner et al., 2021). First, for a given event, we compute and sort trace catalogue seismometer records in order from most energetic (i.e. largest maximum peak amplitude of a record) to least. For that event, feature values are chosen as the median — or second highest — from the top 3 most energetic records. Please refer to the supplementary Table S1 for feature-specific information where this general formula can vary per feature.

Latto, R. B., Turner, R. J., Reading, A. M., and Winberry, J. P.: Towards the systematic reconnaissance of seismic signals from glaciers and ice sheets – Part A: Event detection for cryoseismology, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2023-1340, 2023.

Turner, R.J., Latto, R.B. and Reading, A.M., 2021. An ObsPy Library for Event Detection and Seismic Attribute Calculation: Preparing Waveforms for Automated Analysis. *Journal of Open Research Software*, 9(1), p.29.DOI: https://doi.org/10.5334/jors.365

Have you also considered adding in your feature arrays the standard deviation of the values of the features of the seismic signals recorded at each station ? This can provide some valuable information on the location of the source and on other geometrical properties of the event (e.g. near of far field origin, dip, etc.).
We are happy to add to the discussion regarding use of the feature information as suggested.

L.175 : If for each feature you have the same distribution then those features become useless for any identification methods. Please rephrase and clarify.
We agree that the m/s insinuation that the features have the same distributions is misleading. As noted above, the features are scaled (not normalized– as in not transformed into normal distributions necessarily). The features are transformed by standard deviation to be between -4 and 4, but otherwise have diverse distribution shapes (Fig. 2), lending to interesting clustering results.

L.207 : There are other estimators used to try to determine the relevance of a clustering result besides the Silhouette test (e.g. Davies-Bouldin score, Elbow score, Calinski-Harabasz index, Dunn index). It would have been interesting to see how these factors evolve in relation to the number of clusters chosen and in comparison to the ideal number decided by you.
We agree that we can note in the paper that there were a variety of estimators to choose from. In our preliminary work for this m/s, we considered each of these other estimators along with the Silhouette test for their individual definitions of *similarity* to help guide the choice of k. However, ultimately, we decided to construct our own test for measuring similarity because these typical scoring methods were difficult to interpret in our high-dimensional feature space where *separation* vs *similarity* among 30+ features is convoluted to quantify. We chose to reference and show the results of the Silhouette test as a widely-used assessment that could provide an established point of comparison for our new test but agree that the evolution of cluster numbers as determined by other estimators could be a future avenue of investigation.

L.231-232 : You find 136 stick-slip events identified as such by Pratt et al. (2014), 4 new ones, but how many did you miss from the Pratt et al. (2014) catalog ?
We are happy to clarify as this is not stated in this m/s (but covered in Part A). We did not miss any of the stick-slips in the Pratt et al., 2014 catalogue.

L.241-243 : Indeed, but don't you lose this information by your normalization of the features (see general comment)? The duration/peak amplitude correlation also seems quite strong (as is often observed for microseismic sources, which are sometimes assigned a "duration magnitude"). Is one of these two features therefore excluded from your analysis?
We are happy to add a note in the earlier section that subsequent comparisons refer to normalised values where appropriate. And also a sentence or two in the discussion about possible information loss at various stages of the workflow. We agree that this would be helpful to others using the workflow.

Figure 4 and Figure 7 miss subplot labels “a”, ”b”, ”c”, etc.
Thank you for this careful read of the figures. In Fig. 4, our intention was to leave off subplot labels and they are not referred to in the text. However, we can follow editor guidance on this point. We intended the same for Fig. 7, though see now that subplot labels are referenced on lines 315–316. Based on editor guidance, we can either clean up the text such that (a)--(c) are not referenced as such, or we can add subplot labels to the figure and caption.

Figure 6 (b) : The X-axis label “Days after December 14, 2010” is not convenient for interpretation I think. I suggest giving the real dates.
Our intention with the x-axis label here was to be able to assess periodicity over the two months, and use the # of days as a value to refer back to in the Discussion. However, we also agree with this suggestion, and can add annotation to certain dates (e.g. Jan 1, Jan 31), to provide more information such that the reader does not have to add from Dec 14 to investigate a date of interest noted in our interpretation (e.g. Jan 15).

L.317 : Define “less-defined”. Maybe quantify this (percentage of events from a given class?)
As this is not central to the presented work, we prefer to retain the qualitative comparison, accessible to the reader in the figure.