Review of
Automatic detection of avalanches using a combined array classification and localization
by Heck et al.

1st Revision

The two reviewers of the original submission pointed out several specific issues within the manuscript. The authors responded to all of these comments, yet sometimes it is hard for me to check if or how those replies made it into the revised version of the manuscript. Additionally I think the manuscript still needs work before it can be accepted for publication. Please find my detailed comments below.

Best regards,
Florian Fuchs

Handling of reviewer comments:

The authors reply properly to all reviewer comments. However, especially when answering to Reviewer #1 comments on pages 12 and later it's not clear anymore if those replies were integrated into the manuscript. I do support all of the reviewers comments and questions and I do see that the authors know how to respond to those. But I strongly suggest to implement all of them – at least briefly – into the manuscript. The same holds for almost all comments by Reviewer #2.

Please insert all of those replies to the text (at least briefly) and indicate all the changes in the rebuttal letter. Otherwise, it's hard to follow, not having done the 1st round of reviews.

Additional comments:

Although all/many comments from the first round of reviews were already taken care of, I must unfortunately admit that I still had a hard time reading the manuscript. I do believe that work itself is interesting and the findings are worth reporting. Yet, the manuscript is not easily comprehensible in the current shape. Mainly, I am missing a clear and concise structure and more precision in the wording and figures. I also suggest to make use of the Copernicus English grammar and spellchecking service.

General structure and ease of reading:

- Please be more precise throughout the entire manuscript. When you say "high", "low", "good", "poor", "better", "most", "large" please try to give values, if possible. E.g., what number of percentage can be considered a "good" classification result? When you speak of "features" that "change" and are "common" please describe specifically which features you mean and how they change.

- Your chain of processing kind of gets obscured throughout the manuscript. I'd suggest that somewhere you briefly list your work flow. Figure 5 somewhat tries to summarize this, but I think text would help here. Additionally, Figure 5 could use some instructive labels, e.g. you

could indicated the length of the data windows. The panels “pre-processing” and especially “post-processing” could indicate what’s actually done. E.g. that post-processing is the MUSIC beamforming.

- I am missing a short subsection on “post-processing” in section 3, “methods”. You repeatedly emphasize the need for “post-processing” but it’s not clear what this is.

- In principle it is a good idea to have dedicated sessions on methods and results (sections 3 and 4 in this manuscript). Yet, you mix methods, observations, interpretation and references repeatedly. In the “Methods” section you should be as brief and precise and necessary. You should not evaluate the results of other work here, but only briefly repeat the main points you make use of. All the rest is better placed in the discussion section. Likewise, in the results section (4) you repeatedly evaluate the quality of the results (this should be done in the discussion section) or introduce new steps in the processing. Please double-check to clarify.

- It’s difficult to track how many events you remove during the different processing steps and how many events actually remain as final detections. Maybe a table listing the number of events and how many get discarded by each processing step would help.

Discussion section:

- I agree with Reviewer #1 that the “Discussion” section in the current state is rather a repetition and summary of the previous chapters. This needs to be changed. Here I’d like to see you discuss the benefits and limitations of your methods. E.g. it is very interesting to read that the sensor installation itself already has a huge impact on the classification results. Why? What else can influence the classification that strongly? The airplane signal could also be discussed here (or in the supplemental, see below), as reviewer #2 points out the strong difference to other observed airplane signals. What about anthropogenic signals? Are there roads/cars nearby?

- Most importantly, the choice of the training event should be discussed, as it surely has a huge impact. For example why did you only choose a part of the avalanche signal in Figure 7 as the training event? Half of the signal seems to be missing … You may not have the time and patience now to carefully double-check the performance of your routine based on different training events, but this would of course be desirable. Do you maybe at least have some experience from other datasets that you can report on? Why can’t you simply use more than one training event?

- Could you think of other “features” that could help to distinguish avalanches from airplanes and earthquakes? After all, the ones you use don’t seem to do the job. I’d personally like to see you speculating here …

- Obviously, broadband sensors will not necessarily improve your data quality, neither will they automatically detect more distant avalanches. This needs to be rephrased. Only in the rare case of huge, catastrophic events – which generate long period seismic radiation, in contrast to the small local ones – they might be an advantage over the short-period geophones. The fact that “common” avalanches can only be detected within few km distance is probably due to the weak seismic signal they generate, and the only chance to improve the data quality is to have more sensors (signal-to-noise ratio) closer to the events (less attenuation). Of course, this is not always possible.

Efficiency of computations:

- When discussing the “speed of processing” you refer to a “standard 8 core processor with 16GB RAM”. It may seem picky now, but do you actually make use of all the 8 cores? Is there some kind of parallelization involved in your processing? If yes, please comment on this, if not I think the community usually refers to “a standard personal desktop/laptop computer” to indicate that no supercomputing powers or high-level workstations are required. Similarly, do you actually need the 16GB RAM? If yes, what for?

  Reviewer #2 also pointed this out and it’s actually an interesting point. In fact, probably the computing power wouldn’t really matter and you would not have to comment on it, if data were only processed “off-field” in some data center. However, In your reply, you indicate that some of the processing is done on-site in the field – this of course strongly limits computational power and is a very interesting and crucial point that is not mentioned at all in the manuscript. Please include this in the Instrumentation/Methods section! This will also clarify why you perform some of the processing steps and why computation time is crucial.

Figures:

- There are a lot of Figures, which complicates the reading. I suggest to e.g. somehow merge Figures 6, 8, 12 and 14 as they all show the same information. If all the panels were shown below each other, a comparison of the observations would be easier.

- Similarly, maybe Figures 2 and 3 could be merged.

- Please highlight the avalanches in Figure 4.

- Figures 9 + 11 are not relevant for the understanding of the text and I suggest to move those to the supplemental material. Reviewer #2 raised doubts about the origin of the airplane signals, since they look different in other studies. The authors claim to be certain about their interpretation. This point might also be discussed in the supplemental material, as it’s not crucial for the understanding of the main text.