Repsonse to reviewer #1

This study focuses on the seismic detection and identification of the signals generated by snow avalanches in the Davos area in the Swiss Alps during the winter 2010. The authors tested the capability of a machine learning algorithm (hidden Markov models - HMM) to perform this detection and identification from continuous seismic data. They used a reference catalog to evaluate the performance of the algorithm. The first results showed that the algorithm is capable to achieve relatively high positive identification rates of the avalanches in the catalog (70-95% depending on the station that recorded the signals), but also with a high rate of supposedly false detections. This led the authors to propose a post-processing strategy. Three post-processing steps were investigated: (i) analysis of the duration of the signals; (ii) computation of a correlation factor to evaluate the coherence of the signal between each sensor; and (iii) a voting system based on the classification returned by each station for a given event. Using one, or a combination of those proposed post-processing steps, led to a decrease of false alarm rates, but also in most cases to a decrease of the rate of good identification.

The use of seismology to study environmental processes is of growing interest as it allows producing observations with a unique spatio-temporal resolution. This new approach can help to better understand the triggering factors of natural hazards and to mitigate their consequences on our societies. In this context, this study contributes to the continuing effort to develop robust and versatile methods to explore years of continuous data and for the implementation of real-time seismology-based warning systems. Overall, I think the paper is clearly written, and that the Authors have made a good effort to carefully explore the data, explain their approach and discuss their results. Nevertheless I listed below several comments and suggestions that might help to improve this paper.

General comment:

The only major concern I have regarding this work is that the downsides of using the HMM algorithm are not discussed while most of the results presented in this paper suggests that HMM alone, without pre- and post-processing, cannot perform identification of seismic sources with a high success rate. The strengths of the HMM are usually stated to be:

- i) it does not need any pre-detection or picking (STA/LTA, etc.), which should ensure that no event is missed;
- ii) it does not require any inputs from experts.

Yet this paper demonstrates that

- i) a pre-detection can be suitable to remove low-amplitude/noise signals (figure 2);
- ii) post-processing steps with thresholds set by experts (duration, etc.) is needed to achieve a high accuracy.

Moreover, the Authors are building their post-processing strategy based on features that can be incorporated in the identification models constructed with other machine learning algorithms. The post-processing steps the Authors propose seem necessary because the HMM cannot include these features (durations of the signals, coherence between signals recorded at different stations, vote among stations) in the model due to its core design, which is to consider chunk of continuous data and not the entire signal generated by the event. This forces the Authors to manually set thresholds on those features, while with some other algorithms those thresholds are determined through a statistical analysis of the reference data. I think the authors must include a more thorough and objective discussion on the pros (which definitely exist) and the cons of HMM compared to other algorithms/studies in the light of the results of this work.

We agree with the reviewer that we did not sufficiently discuss the shortcomings of HMM models. We now address this in more detail in the Discussion.
We would like to clarify that the signal amplitude threshold value we used should not be considered a pre-detection. We merely applied this threshold to reduce the data volume for the feature computation. Since the goal is to develop this method for operational application, reducing the computational time is of crucial importance. We would also like to point out that the HMM identifies events and returns a duration for the events. Event duration is thus a feature obtained by the model. Furthermore, the coherence between the sensors could also be used as a feature in the HMM. However, due to the high computational time required to calculate the coherence between 21 receiver pairs, we decided to use this feature only during post-processing. Calculating the coherence for a small number of detections is faster than calculating the coherence for the continuous data set. Finally, the reviewer is correct to state that the voting based classification cannot be included in the HMM.

Specific comments:
P3 L6-8: Are those false alarm rates related to the choice of the algorithms or to the choice of the features used to parametrized the signals? The latter might be more important and should be mentioned.

These false alarms rates are related to the choice of the algorithm.

Figure 1: I think a colorscale with more colors would allow to better observe the features of the signals generated by the different sources, especially at frequencies below 50 Hz. This is important as the readers might want to understand what guided your choice of features. Also this figure can be larger.

We chose this colour scale since it is perceptually uniform. We adapted the colour scale and enlarged the figure to more clearly highlight the features of the signals. However, the main goal of this figure was to highlight the ubiquitous environmental noise and not necessarily those of avalanche signals. The features of a typical avalanche signal are more clearly visible in Figure 7.

P8 L25-27: How do you compute the duration?

The duration of the avalanches in the reference catalogue was determined by visual inspection of the seismic data.
The onset was defined as the first appearance of energetic low frequency signals (i.e. between 15 and 25 Hz), while the end of the signal was defined as the time when low frequency signals reverted back to background levels (P5 L4-6).
The duration of any automatically detected event was determined by the HMM (P10 L9).
The minimal duration T_{min} was determined manually and is described in the results section (P14 L12 – P16 L8).

P9 L1-3: How does the voting step in the post-processing would impact the detection of “small” events (especially with a threshold set at 5 stations for a network with 7 sensors)? Are “small” events detected by the whole network? A figure showing the locations of the avalanche corridors and the seismic network would be interesting.

We have added more details on the spacing between the geophones (see P 4 L5-9). A figure of the array can be found in van Herwijnen and Schweizer (2011a) and we refer the reader specifically to their figures (see Figure 3 and 4 in van Herwijnen and Schweizer, 2011a and Figure 2 in van Herwijnen and Schweizer, 2011b).
Due to the short distance between the sensors, the voting step does not neglect small avalanches. Signals of small events are recorded at all stations. The deployment of the sensors resulting in a less desirable SNR impacts more on the detection of small avalanches. Furthermore, most of the confirmed events can be regarded as small avalanches regarding international avalanche classifications.

P11 L4-5: Indeed. How would it have impacted your results if you had chosen another master event? Is this has been investigated in Hammer et al. (2017)? If yes it should be mentioned and referenced.

We used an avalanche event from the 100% group as master event. We compared it with other avalanche events and these all had similar features (Fig 5). We therefore can expect, that the classification results are similar.
We investigated the effect of using another training event and the overall results were very similar. However, since this is a very time consuming endeavour, we did not perform a more in-depth analysis of the influence of the training event.
Hammer et al (2012) investigated the dependence of classification performance on the reference event in detail. They show that the proposed approach is very robust in face of various master events.

P13 L11-12: So the selection of the threshold on the duration is not done by considering the physics of the sources or the distribution of the durations in the reference catalog, but to optimize the POD-FAR ratio? By applying this threshold to the reference data set you lose 40% of the events (P14, last line). How is this threshold choice impacting the detection of "small" avalanches? Can you show a histogram of the duration of the events in the reference catalog? You state on P3 L24 that "For avalanche forecasting information on smaller avalanches is also required". Hence is the approach you propose suitable to detect the smaller events?

Indeed, we optimized the duration threshold with the POD-FAR ratio. Since signal duration likely relates to avalanche size (van Herwijnen et al., 2013), it is clear that by increasing the signal duration threshold the detection rate of the smallest avalanches decreases. However, overall most of the avalanches in the reference catalogue can be considered to be 'small' for the purpose of avalanche forecasting. As such, we believe our approach is still suitable to detect smaller events.

P13 L14-16, P20 L6-7, Table 3: Again, it would be great to have a map of the seismic network to discuss the discrepancies observed at the different sensors. Distance to the sources, traveled paths, attenuation, dispersion, etc., can also be factors impacting the amplitude, the duration and more generally the features of the signals that might in return change the POD at different stations. This could be discussed.

Since the distance between the sensors was rather small (<12 m for the longest distance), we do not believe that local site effects substantially influence the signals. However, since the sensors were inserted in the snow, we are convinced that the main reason for the discrepancies in model performance between the different sensors is due to differences in snow cover properties, as discussed in P22 L5-10. To clarify this point in more detail, we now also mention in Section 2 that sensors 1 and 4 were most deeply covered by snow (P4 L5-8).

P20 L17-18: So in the best case what is your overall accuracy? Considering which range of avalanche sizes? I think it is this information that the readers will seek.

Overall accuracy is difficult to estimate. Although we have a reference catalogue, the actual number of avalanches in the continuous seismic data remains unknown. This is why we implemented the subjective probability classes in the reference data set. While such an approach does not provide a single performance value for the HMM, it shows that for clear avalanche signals (100% class) model performance is reasonable. Furthermore, it highlights the difficulties in obtaining a reliable and independent avalanche catalogue. However, in response to reviewer 2, we now also added information on the detection rate of the avalanches that were confirmed by images from the automatic cameras. These results show that at least 80% of the confirmed events were detected by all sensors.

P20 28-30: Are the results presented in this study supporting this statement or is it based solely on the study by Hammer et al. (2017)?

Our results support this statement since we also only used one training event for the HMM.

P21 L4-5: How can you incorporate localization parameters in the HMM? Is this done directly in the model or during the post-processing?

Both methods are possible. Localization metrics, in particular back azimuth and apparent velocity, of signals above the amplitude threshold could be used as feature for the HMM. On the other hand, it would also be possible to use the localization metrics during post processing by rejecting all detections having clear path.

Repsonse to reviewer #2

General comments:

The paper discusses the performance of an automatic seismic detection system of snow avalanches using hidden Markov models (HMMs). The study is based on a 107- day continuous dataset acquired during the winter 2010 by a small seismic array consisting of 7 vertical geophones deployed above Davos, Switzerland, and surrounded by several avalanches starting zones. The HMMs system is tested against a reference avalanche catalogue that is based on the work by van Herwijnen and Schweizer (2011a). The reference avalanche catalogue gathers 283 events distributed into 7 probability classes following an independent re-evaluation of the avalanche signals of van Herwijnen and Schweizer (2011a) by three of the authors. Among the 283 events, only 20 (7%) are finally classified by the aforementioned three authors as being with 100% probability a snow avalanche. For the HMMs detector, the authors apply the approach developed by Hammer et al. (2012) that requires a background noise model and only one training event to learn the system. Because important diurnal and seasonal variations are observed in the seismic feature of the background noise over the season, a background noise model is recalculated for each day. Since the authors observe that the change from dry to wet avalanches through the winter does not have a strong impact on the selected seismic features used for the HMMs, only one avalanche signal is selected as a training event for a single avalanche class HMMs detector. The first HMMs avalanche detector results in high probability of detection (POD 70-95%) in the 100% probability class of the reference avalanche catalogue for all 7 stations. However, a large number (up to 2091) of additional events (unassigned detections) are detected by the HMMs for each sensors. In order to reduce the number of unassigned detections, the authors decide to introduce post-processing steps to the HMMs: (i) a duration based classification and (ii) an array based (multiple-station detection/coherence of signal among station pairs) classification. The postprocessing steps reduce the number of unassigned detections by the HMMs; however, the POD in the highest probability class of the reference avalanche catalogue is also reduced. A manual re-evaluation of the unassigned detections after post-processing shows that part of these detections correspond to avalanche missed during the manual detection procedure.

The paper is well structured and clearly written. Despite the noisy character of the data, the authors show HMMs as developed by Hammer et al. (2012) to be a promising approach for operational avalanche forecasting; however, with fine tuning. Strategies will be needed for a regular update of the background noise model and post-processing must be implemented. Much attention should be paid to the stations deployment. The author recommend to install the sensors 30-50 cm in the ground below a homogenous snow cover in order to reduce environmental noise and increase coherence of the data among the network and thus, stabilize the array based post-processing. In addition, a larger sensor inter-spacing would allow the system to apply array-processing techniques and enable to incorporate localization
parameters of avalanche source area in the automatic system. In my opinion, the implementation of the post-processing step is of great interest, as it helps to deal with the noise contamination that is found at any surface site data. The main concern I have is that the authors didn’t make use of the 33 confirmed avalanches of van Herwijnen and Schweizer (2011a) and used instead a re-processed

probabilistic reference catalogue where only 20 avalanches remain considered as certain. In the following, a number of specific comments are related to this concern. I leave it to the authors to decide to which point they want to integrate these suggestions. It would be interesting to know how the 33 confirmed events match the post-processing steps: do they have durations above 12 seconds, are they detected by 5 sensors or more, what are the inter-station correlation indices for these events. In addition, if the location of the 33 confirmed events is known, how important are the variations in the spectral feature of the signals as a function of station-source distance? Finally, what is the performance of the HMMs system including the three post-processing steps against this original 385 events/33 confirmed events reference catalogue?

We agree with the reviewer that adding the model performance for the confirmed avalanches would be insightful. We therefore updated figures (Figures 3 and 9) and tables (Table 1,2 and 3) to include these results and discuss them throughout the paper (e.g. P10 L10-11). Note that the number of confirmed avalanches originally mentioned in the manuscript was wrong and should have been 25. We believe that these results better highlight the model performance and we thank the reviewer for this valuable suggestion.

Specific comments:

P1 L9-12. Please rephrase and clarify this part of the abstract (according to remarks about P4-7 Section 3).

We rewrote this sentence (P1 L10).

P4 L5-7. Although the station inter-spacing is specified in van Herwijnen and Schweizer (2011b), you should mention it here, so the reader has an idea of the size of the network.

We now added this information in the text (P4 L6).

P4-7. I think the section 3 needs some clarification so the reader can get a better idea of the reference data:

1) What are the exact data that were analyzed visually by van Herwijnen and Schweizer (2011a) (385 detection, 33 confirmed events): (1) 107-day of continuous data without pre-processing of section 3.1 at sensor number X or (2) XX% of pre-processed data (section 3.1) at sensor number X? In case (1), I would place P4 L28-30 before section 3.1 for clarity.

We now clarify these points (P5 L1) but prefer to keep the sentence where it is.

2) Section 3.1: The pre-processing of the data is obscure to me. Was it apply only to reduce the amount of data to (re-)process manually? Did van Herwijnen and Schweizer (2011a)/Heck et al. AND the HMMs perform only on the 20% of high

amplitude data (which would impact on the paper title as it would be no more continuous data)? If a local network is intentionally deployed to detect small avalanches, why place an energy threshold in the investigated data series? See for example avalanche number 5 (Av 5) in Figure 7 of van Herwijnen and Schweizer (2011b) which has poor SNR. Furthermore, HMMs are not dependent on energy thresholds, please clarify.

The manual detection of avalanches was performed on the complete data set. However, calculating the features for all seven sensors and the entire data set is extremely time consuming. We therefore applied an amplitude threshold to reduce the amount of data by only taking those parts of the time series when some energy is arriving at the sensors. The amplitude threshold we used was very conservative and all the confirmed avalanches were still in the pre-processed data (Table 1 or Figure 3). Since the goal is to develop this method for operational application, reducing the computational cost is of crucial importance. As such, we do not believe that the pre-processing step we applied warrants a change of the title.

3) Section 3.2: 20 merged avalanches represent 7% of the 283 reference events. In these data, where are the 33 confirmed avalanches of P4 L30? You could have used 33 confirmed events (9%) against 352 uncertain events of the original 385 events of van Herwijnen and Schweizer?

We now include these confirmed events in Figure 3 and Table 1,2 and 3. Most of these events belong to the 100% or 83% class of the re-evaluated reference data set. For sensor 1, 20 of these events were detected and 5 missed.
For the voting based detection without using a minimal duration, also 20 were detected and 5 missed. By applying a minimal duration, the number of detections reduced to 18 hits and 7 missed.

4) P6 L11-15: I think it would be interesting to discuss here in a few words the 283 events reference avalanche catalogue in more details. For example, do you find trends in the probability classes, are the 20 certain events confirmed detections, long duration signals, high energy events detected by all stations? Are the lower confidence event simply lower quality events (low SNR, only recorded at a few stations and not consistent among the stations)? These low probability events represent in my opinion the real challenge of environmental seismology case studies. And I expect the subjectivity of the analyst to have more influence on low quality data (how would/did you rate avalanche Av 5 in Figure 7 of van Herwijnen and Schweizer (2011b)) than on high SNR data? This is where automatic, quantitative systems should help :)

We clarified these points by adding model performance statistics for the confirmed avalanche events throughout the text (Figures 9 and Tables 2 and 3). Furthermore, we added a figure (Figure 4 in the new manuscript) which shows the distribution of signal duration for each probability class. This figure clearly shows that the number of short duration events decreases with increasing probability class.

5) P7 Figure 3. How does the original catalogue with 385 events, 33 confirmed avalanches plot? Are the two periods of higher activity well represented? I personally would insert a plot of the 33 confirmed events at the bottom of Figure 3.

We now show the confirmed avalanches in Figure 4. Indeed, most of these avalanches also released during the high activity periods in March and April.

P10 L8-20. Does this very sophisticated approach to find a threshold value bring more than an approach using common sense (for which arguments are well described in P8 L24-29 – P9 L1-20)? The minimum duration can be evaluated (or decided) as a function of the expected distance/size of the target events; the number of stations that must detect the signal can be selected as a reasonable value knowing the network/stations specific performances; the inter-station correlation threshold can be evaluated by investigating the 33 confirmed events. Please comment.

We used a systematic approach to find threshold values by optimizing model performance. Since we had a reference avalanche catalogue this was possible. However, in the absence of such a reference data set, for instance when installing a system at a new location, we agree that to some extent it should be possible to select these thresholds based on some a priori available knowledge on the local topography and the array geometry. However, this requires some assumptions on the minimum avalanche size that can be detected as well as sensor performance.

P11 L11-13. I agree that in Figure 5a the feature distribution is very similar for the 4 selected events. However, in Figure 5b, I find events of 21 Jan/21Feb dissimilar to events of 22Mar/24Apr, especially at the events onset where one group goes up while the other one goes down. Please comment on that and see remark on P15 Figure 8.

Due to the different length of the signals, we decided to use a normalized time. As the events in March and April were quite long, the decrease for cepstral coefficient is longer. In January and March the decrease is shorter, thats why it seems as it only increases. This is also visible for the central frequency, where the frequency decreases slower for the March and April events.
Furthermore, for the HMM all 4 events are similar, since the feature behaviour (the decrease) is important. The absolute appearance of the features does not matter. This explains why the detection of events with different durations is possible.

P12 Figure 5. In (a) and (b) vertical dotted lines at normalized time 0 and 1 would help to visualize start and end time of event.

Figure 5 changed as suggested.

P15 Figure 8. For both sensor 1 and 7, the HMMs missed 4 events at the end of January:
1) I don't find this 4-event spike in Figure 3?
2) Are any of the 33 confirmed avalanches found at this date? I see there is one 100% probability class event around this date in Figure 3.
3) Could this speak for an influence of the HMMs system by the avalanche type (dry-
wet) selected to train the system?

The small spike in avalanche activity at the end of January was also present in Figure 3, marked as 33% avalanches. However, the spike is more obvious in Figure 8 since there we also applied the signal duration threshold to the reference data set. None of the confirmed avalanche events occurred on this day, as can now be seen in Figure 3 where we added the confirmed events.

P18 L1-6. Clarification for the single sensor results:
1) Did you investigate the signals of the unassigned detections individually at station 1 and 6? Or, 2) Did you look at the signals recorded by all stations at the time of the unassigned detections of sensor 1 and 6?
I think these results should be commented. What could explain the higher false alarm rate of the array based approach against the single sensor approach?

We investigated the results for each sensor individually. We clarified the sentence (P20 L1-2).
The higher false alarm rate is due to the higher signal-to-noise ratio for the other 5 sensors, as stated in the Discussion section (P22 L5-10).

P19 L8-10. Related to comment on section 3. Please clarify. At P4 L28-30, the 33 events are part of the 385 van Herwijnen and Schweizer (2011a) events. (sorry to insist ;))

The 25 confirmed events were part of the 385 original events and also remained in the 283 events after applying the amplitude threshold during pre-processing.

Automatic detection of snow avalanches in continuous seismic data using hidden Markov models

Matthias Heck[1], Conny Hammer[2], Alec van Herwijnen[1], Jürg Schweizer[1], and Donat Fäh[2]

[1]WSL Institute for Snow and Avalanche Research SLF, Davos

[2]Swiss Seismological Service SED, ETH Zurich, Zurich

Correspondence to: Matthias Heck (matthias.heck@slf.ch)

Abstract.

Snow avalanches generate seismic signals as many other mass movements. Detection of avalanches by seismic monitoring is highly relevant to assess avalanche danger. In contrast to other seismic events, signals generated by avalanches do not have a characteristic first arrival nor is it possible to detect different wave phases. In addition, the moving source character of avalanches increases the intricacy of the signals. Although it is possible to visually detect seismic signals produced by avalanches, reliable automatic detection methods for all types of avalanches do not exist yet. We therefore evaluate whether hidden Markov models (HMMs) are suitable for the automatic detection of avalanches in continuous seismic data. We analyzed data recorded during the winter season 2010 by a seismic array deployed in an avalanche starting zone above Davos, Switzerland. We ~~first visually inspected the data and identified more than 200 events we assume to be generated by avalanches.~~ re-evaluated a reference catalogue containing 385 events by grouping the events in seven probability classes. Since most of the data consists of noise we first applied a simple amplitude threshold to reduce the amount of data. As first classification results were unsatisfying, we analyzed the temporal behaviour of the seismic signals for the whole data set and found that there is a high variability in the seismic signals. We therefore applied further post-processing steps to reduce the number of false alarms by defining a minimal duration for the detected event, implementing a voting based approach and analyzing the coherence of the detected events. We obtained the best classification results for events detected by at least 5 sensors and with a minimal duration of 12 s. These processing steps allowed identifying two ~~known~~ periods of high avalanche activity, suggesting that HMMs are suitable for the automatic detection of avalanches in seismic data. However our results also showed that more sensitive sensors and more appropriate sensor locations are needed to improve the signal-to-noise ratio of the signals and therefore the classification.

1 Introduction

During the winter season, snow avalanches may threaten people and infrastructure in mountainous regions throughout the world. Avalanche forecasting services therefore regularly issue avalanche bulletins to inform the public about the avalanche conditions. Such an avalanche forecast requires meteorological data, information about the snowpack and avalanche activity data. The latter are mostly obtained through visual observations requiring good visibility. Avalanche activity data are therefore often lacking during periods of intense snowfall, which are typically the periods when they are most important for forecasting. A possible alternative approach to determine the avalanche activity is to use a seismic monitoring system (e.g. van Herwijnen and Schweizer, 2011a).

Seismic monitoring systems are well suited to detect mass movements such as rockfalls, pyroclastic flows and snow and ice avalanches (Faillettaz et al., 2015; Podolskiy and Walter, 2016; Caplan-Auerbach and Huggel, 2007; Suriñach et al., 2005; Zobin et al., 2009). The ability to detect snow avalanches through seismic methods was first demonstrated in the 1970s. St. Lawrence and Williams (1976) and Harrison (1976) deployed geophones near avalanche paths and manually identified signals generated by avalanches in the seismogram. They showed that the seismic signature of avalanches differs from other seismic events such as earthquakes or nearby blasts. A more in-depth analysis of seismic signals generated by avalanches was performed 20 years later, identifying typical characteristics in both the time and time-frequency domain (Kishimura and Izumi, 1997; Sabot et al., 1998). Using automatic cameras to film avalanches, Sabot et al. (1998) showed that specific features in the seismic signals were related directly to changes in the flow of the avalanche; these findings were confirmed by Suriñach et al. (2000) and Suriñach et al. (2001). Since then, seismic signal characteristics were used to estimate specific properties of single avalanches such as the flow velocity (Vilajosana et al., 2007a), the total energy of the avalanche (Vilajosana et al., 2007b) or the runout distance (Pérez-Guillén et al., 2016; van Herwijnen et al., 2013).

While many studies focused on using seismic signals to better understand the properties of single avalanches, continuous monitoring of avalanche starting zones to obtain more accurate avalanche activity data is of particular interest for avalanche forecasting (e.g. van Herwijnen et al., 2016). Leprettre et al. (1996) deployed three-component seismic sensors at two different field sites and compared seismic signal characteristics to a database including avalanches, helicopters, thunder rolls and earthquakes. Lacroix et al. (2012) improved the seismic monitoring system used by Leprettre et al. (1996) by deploying a seismic array between two known avalanche paths. The array consisted of six vertical component geophones arranged in a circle around a three component geophone in the centre. Using array techniques, Lacroix et al. (2012) determined the release area and the path of manually identified avalanches and estimated their speed. These studies mainly monitored medium and large avalanches. van Herwijnen and Schweizer (2011a), however, deployed seismic sensors near an avalanche starting zone above Davos in the eastern Swiss Alps to also detect small avalanches. They manually identified several hundred avalanche events in the continuous seismic data during four winter months.

While these studies have highlighted the usefulness of seismic monitoring to obtain more accurate and complete avalanche activity data, using machine learning algorithms to automatically detect snow avalanches has thus far remained relatively unsuccessful. Nevertheless the interest in these techniques has been evident for several decades (Leprettre et al., 1998; Bessason et al., 2007; Rubin et al., 2012). The first attempt to automatically detect avalanches focused on using fuzzy logic rules and credibility factors derived from features of the seismic signal in the time and time-frequency domain (Leprettre et al., 1996, 1998). In a first step, the features of unambiguously identified seismic events were analysed including avalanches, blast and teleseismic events. They then formulated several fuzzy logic rules for each type of event to train a classifier used to identify the type of a new unknown seismic event. While the probability of detection (POD), i.e. the number of detected avalanches divided by the total number of observed avalanches, was high ($\approx 90\%$), one of the main drawbacks of this method is the subjective expert knowledge used to derive the fuzzy logic rules and the need to adapt these rules to each individual field site.

Bessason et al. (2007) deployed seismic sensors in several known avalanche paths along an exposed road in Iceland. They used a nearest-neighbour method to automatically identify avalanche events. The method consists of comparing new events with those in a database. Although a 10-year database was used, the identification performed rather poorly. Seismic signals generated by rockfalls and debris flows were wrongly classified as avalanches and vice versa, resulting in a POD of about 65%.

In an attempt to improve the automatic detection of avalanches, Rubin et al. (2012) used a seismic avalanche catalogue presented by van Herwijnen and Schweizer (2011b) and compared the performance of 12 different machine learning algorithms. The PODs of all classifiers were high (between 84% - 93%). However, the main drawback were the high false alarm rates, much too high for operational tasks.

The methods described above are generally difficult to apply at new sites since they require time to build a training data set and/or expert knowledge to define thresholds and rules. To overcome these drawbacks, we investigate using a hidden Markov model (HMM). A HMM is a statistical pattern recognition tool commonly used for speech recognition (Rabiner, 1989) and was first introduced for the classification of seismic traces by Ohrnberger (2001). The advantage of HMMs compared to other classification algorithms is that the time dependency of the data is explicitly taken into account. First studies using HMMs for the classification of seismic data relied on large training data sets (Ohrnberger, 2001). Using this approach Beyreuther et al. (2012) created an earthquake detector.

More recently, Hammer et al. (2012) developed a new approach which only requires one training event. This approach was applied for a volcano fast response system (Hammer et al., 2012) and the detection of rockfalls, earthquakes and quarry blasts on seismic broad-band stations of the Swiss seismological service (SED) (Hammer et al., 2013; Dammeier et al., 2016). Furthermore, Hammer et al. (2017) also detected snow avalanches using data from a seismic broadband station of the Swiss Seismological Service. During a period with high avalanche activity in February 1999, 43 very large confirmed avalanches were detected over a 5 day period with only 4 presumable false alarms. While these detection rates are very encouraging, the investigated avalanche period was exceptional. Furthermore, due to the location of the broadband station at valley bottom, they could not detect small or medium sized avalanches.

For avalanche forecasting information on smaller avalanches is also required. To resolve this issue, we investigate a method to obtain local (i.e. scale of a small valley) avalanche activity data using a seismic monitoring system (van Herwijnen and Schweizer, 2011a). We therefore implement the approach outlined by Hammer et al. (2017) to detect avalanches in the continuous seismic data obtained with the sensors deployed near avalanche starting zones (van Herwijnen and Schweizer, 2011b). Based on the duration of the seismic signals, the majority of these avalanches were likely rather small (van Herwijnen et al., 2016). We therefore used this avalanche catalogue to train and evaluate the performance of a HMM model and discuss limitations and possible improvements.

2 Field site and instrumentation

We analysed data obtained from a seismic array deployed above Davos, Switzerland. The field site is located at 2500 m a.s.l. and is surrounded by several avalanche starting zones. The site is easily accessible during the entire year and is also equipped with various automatic weather stations and automatic cameras observing the adjacent slopes.

The array consisted of seven vertical geophones with an eigenfrequency of 14 Hz. The maximum distance between the sensors was $12\,\mathrm{m}$ Six of the geophones were inserted in a styrofoam housing and placed within the snow, whereas the seventh geophone (Sensor 7) was inserted in the ground with a spike (see Figure 3 and 4 in van Herwijnen and Schweizer (2011a)). Sensors 1 and 4 are the ones nearest to the ridge and most deeply covered (see Figure 2 in van Herwijnen and Schweizer (2011b)). The seismic sensors were deployed at the field site from early December until the snow had melted.

The instrumentation was originally designed to record higher frequency signals in order to detect precursor signals of avalanche release (van Herwijnen and Schweizer, 2011b). A 24-bit data acquisition system (Seismic Instruments) was used to continuously acquire data from the sensors at a sampling rate of 500 Hz. The data were stored locally on a low power computer and manually retrieved approximately every 10 days. A more detailed description of the field site and the instrumentation can be found in van Herwijnen and Schweizer (2011a) and van Herwijnen and Schweizer (2011b).

3 Data

Continuous seismic data were recorded from 12 Jan 2010 to 30 April 2010. These data were previously used by van Herwijnen and Schweizer (2011b), Rubin et al. (2012) and van Herwijnen et al. (2016). The recorded seismic data contains various types of events (van Herwijnen and Schweizer, 2011a), including aeroplanes, helicopters and of course avalanches (Figure 1).

3.1 Pre-processing of the seismic data

While during the 107 day period several hundred avalanches were identified, the vast majority of the data consist of noise or seismic events produced by other sources. To reduce the amount of data to process, we applied a simple threshold based event detection method. It consisted of dividing the continuous seismic data stream in non-overlapping windows of 1024 samples. For each window, a mean absolute amplitude A_i was determined. When $A_i \geq 5\overline{A}$, with $\overline{A}$ the daily mean amplitude, the data were used. Furthermore, a data section of $\Delta t = 60\,\mathrm{s}$ before and after each event was also included to ensure that the onset and coda of each event were incorporated. The amount of data to process was thus reduced by 80%(Figure 2).

Finally, before we used the data for the detection, we applied a bandpass filter. Previous studies showed that seismic signals generated by avalanches typically have a frequency below 50 Hz (e.g. Harrison, 1976; Schaerer and Salway, 1980). We therefore applied a 4th order Butterworth bandpass filter to our data between 1 and 50 Hz.

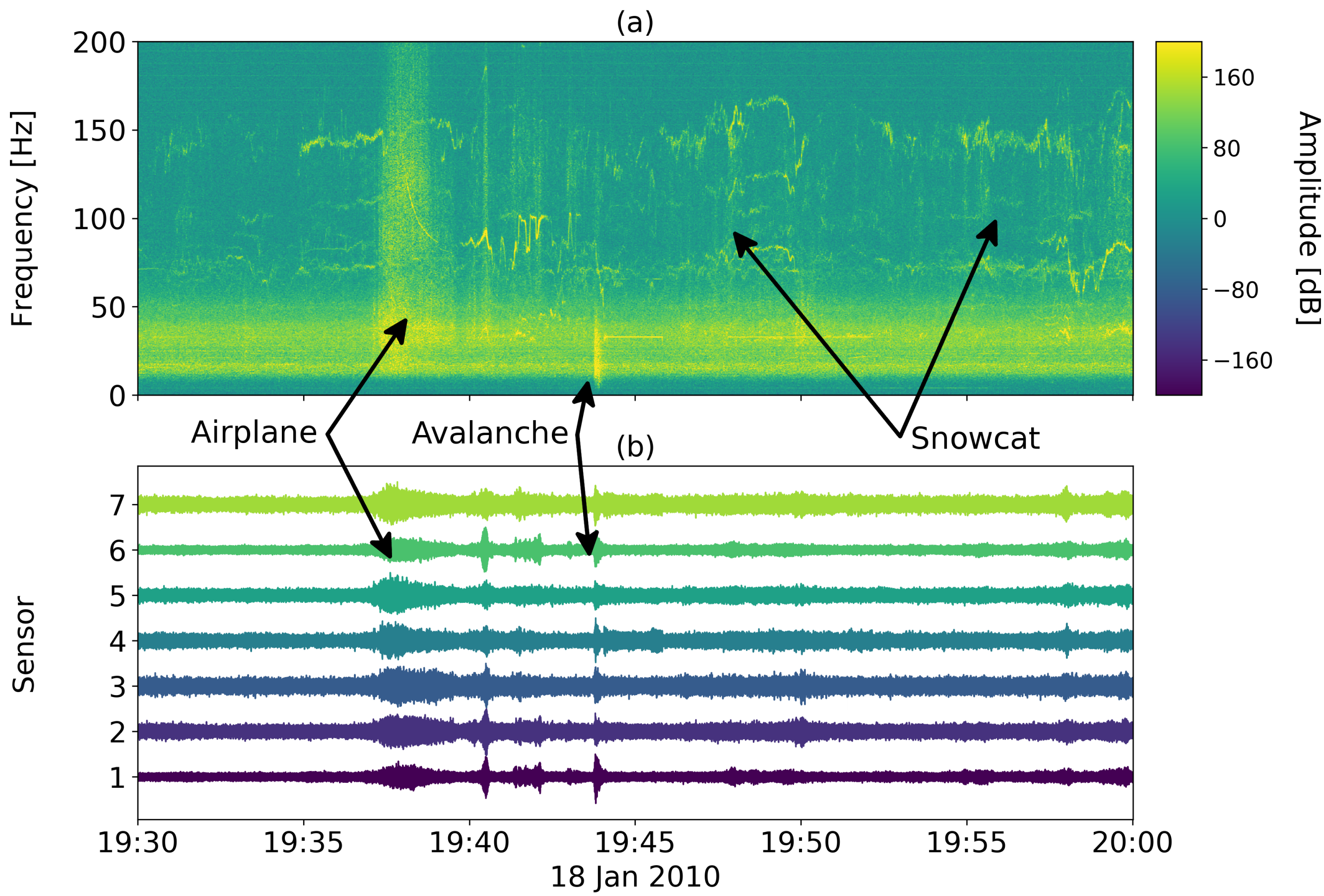

Figure 1. (a) Spectrogram of an unfiltered 30 minute time series. (b) Corresponding time series. An airplane as well as an avalanche are visible. Furthermore, noise produced by a snowcat ~~is visible~~ can be seen in the second half of the time series.

3.2 Reference avalanche catalogue

van Herwijnen and Schweizer (2011a) visually analysed the unprocessed seismic time series and the corresponding spectrogram of one sensor and identified $N = 385$ avalanches between 12 January and 30 April 2010. They thus obtained an avalanche catalogue consisting of the release time t_i and duration T_i for each avalanche. The onset was defined as the first appearance of energetic low frequency signals (i.e. between 15 and 25 Hz), while the end of the signal was defined as the time when low frequency signals reverted back to background levels. However, only ~~33~~ 25 of these avalanches were confirmed by visual observations, i.e. ~~by field observations or~~ on images obtained from automatic cameras. Hence there remains substantial uncertainty about the nature of the identified events.

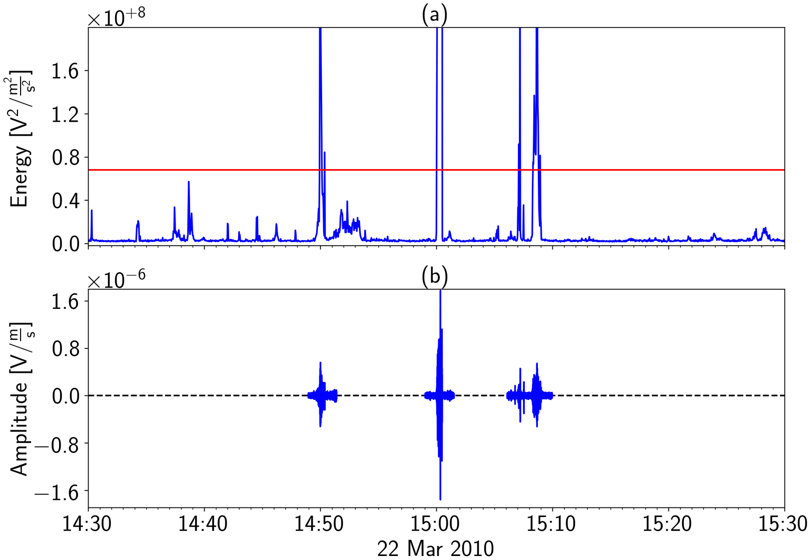

Figure 2. Example of the pre-processing. (a) The mean energy values of each window shown as the blue line and the threshold value indicated by the red line. (b) The remaining data cut by the pre-processing step.

To reduce the uncertainty, three of the authors therefore independently re-evaluated this avalanche catalogue. From the 385 avalanches in the original avalanche catalogue only $N_{\text{pre}} = 283$ remained after pre-processing (see section 3.1); none of the 25 confirmed avalanches were dismissed. By visually inspecting the seismic time series of the seven sensors and the stacked spectrogram for each event, they then assigned a subjective probability p_e to each of these possible avalanche events. Three probabilities were assigned: 1 when it was certain that the observed event was an avalanche, 0 when it was certain that the observed event was not an avalanche and 0.5 when it was uncertain whether the event was an avalanche or not. The probabilities were then combined into seven probability classes depending on the mean probability of each event:

$$P_{\text{ava}} = \frac{1}{3}\sum_{e=1}^{3} p_e \tag{1}$$

with p_e the subjective probability that each assigned to a specific event. In Table 1 the number of events in each probability class is listed. In the reclassified data set, only 20 avalanches ~~merged~~ were considered as certain avalanche by all three evaluators and 58 events were marked as certainly not an avalanche. Furthermore 18 of the 25 visually confirmed avalanches were within the two highest probability classes.

The avalanche activity of the season 2010 is shown in Figure 3. Overall, most avalanches were detected in the second half of the investigated period, with two distinct peaks in the avalanche activity around 22 March and 24 April.

Table 1. Number of events per probability class P_{ava} after re-evaluation and the corresponding number of confirmed events for each class. The first row shows the possible combinations of subjective probability (p_e): $p_e = 1$ it was certain that the event was an avalanche, $p_e = 0.5$ it was uncertain and $p_e = 0$ it was certain that the event was not an avalanche.

p_e combinations	0,0,0	0.5,0,0	1,0,0 0.5,0.5,0	0.5,0.5,0.5 1,0.5,0	1,1,0 1,0.5,0.5	1,1,0.5	1,1,1
P_{ava}	0%	16.6%	33.3%	50%	66.6%	83.3%	100%
Number of events	58	66	48	34	30	27	20
Number of visually confirmed events	0	0	1	3	3	9	9

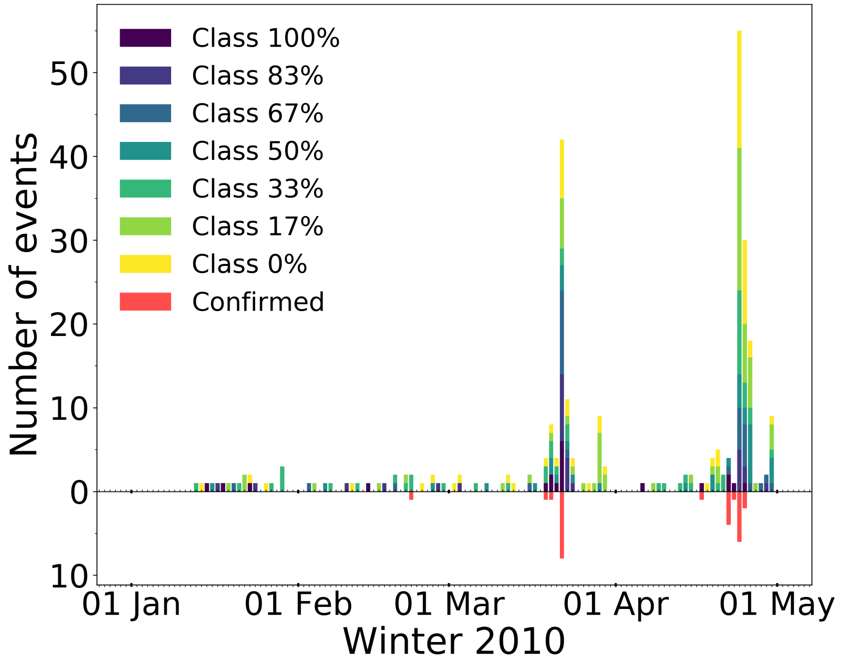

Figure 3. Avalanche activity for the winter season of 2010. The different colours above the zero line indicate the different probability classes derived from the manual detection. The red bars below the zero line are indicates the visually confirmed events.

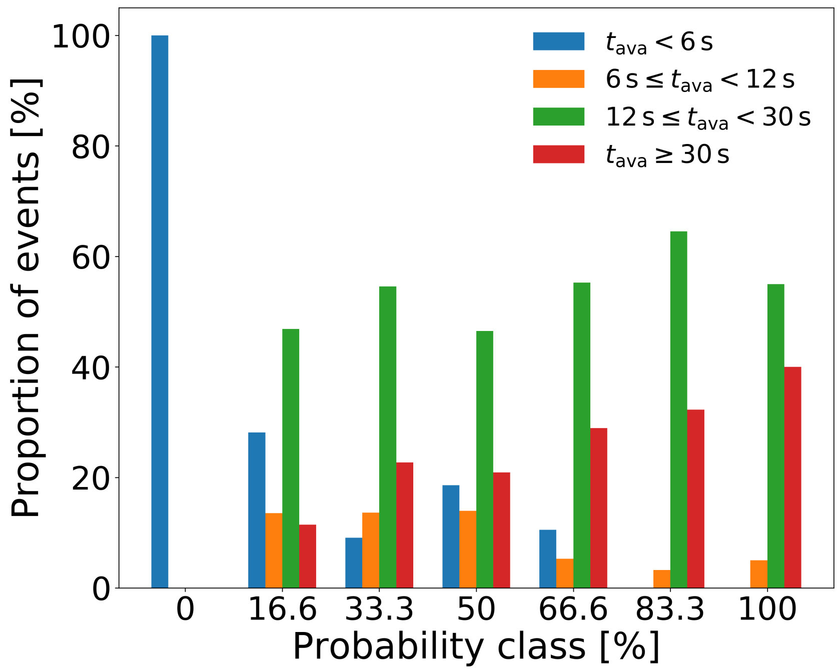

Figure 4. Event duration distribution per probability class.

The distribution of the event duration changed as the probability class increased (Figure 4). All events in the lowest probability class had a duration $< 6\,s$. More than $90\,\%$ of the events had durations $\geq 12\,s$ for the three highest probability classes.

4 Methods

4.1 Hidden Markov model

~~In this study~~ To automatically detect avalanches in our continuous seismic data we used hidden Markov models (HMMs)~~to detect avalanches in a continuous seismic data set~~. This statistical classifier models observations (i.e. the seismic time series or its features) by a sequence of multivariate Gaussian probability ~~distribution~~distributions. The characteristics ~~for~~ of the distributions (i.e. mean and covariance) are derived from training sets of known events, so called pre-labelled training sets. Several classes describing different types of events can be implemented in the classifier but each class needs its own training set to determine its unique distribution characteristics. Therefore, the actual classifier consists of several HMMs, one for each class. This classical approach, as used for the classification of seismic time series by e.g. Ohrnberger (2001) and Beyreuther et al. (2012), relies on well known pre-labelled training sets. In our case, however, avalanches are rare events and it is nearly impossible or too time consuming to obtain an adequate training set. To circumvent the problem of obtaining sufficiently large training sets, we used a new approach developed by Hammer et al. (2012). This classification approach exploits the abundance of data containing mainly background signals to obtain general wave-field properties. Using these properties, a widespread background model can be learned from the general properties. A new event class is then implemented by using the background model to adjust the event model description by using only one training event. The so obtained classifier therefore consists of the background model and one model for each implemented event class. The classification process itself calculates the likelihood that an unknown data stream has been generated by a specific class for each individual class HMM. More detailed information can be found in Hammer et al. (2012, 2013).

4.2 Feature calculation

Although it is possible to use the raw seismic data as input for the hidden Markov model, we used a compressed form of it, so called features. Several features can be calculated representing different aspects of the time series such as spectral, temporal or polarization characteristics. The representation of the seismic signals by features is more adequate to highlight differences between diverse event types. Since we used single component geophones, we only used spectral and temporal features. Based on preliminary analysis, we used the features listed below.

- Central frequency
- Dominant frequency
- Instantaneous bandwidth
- Instantaneous frequency
- Cepstral coefficients
- Half-octave bands

A detailed list of the functions used to calculate these features can be found in Hammer et al. (2012).

To calculate the features, we used a sliding window with width $w = 1024$ samples. The sliding window is then moved forward with a step of $0.05\,\mathrm{s}$ or 25 samples resulting in an overlap of 97%.

For the half-octave bands we used a central frequency of $f_c = 1.3\,\mathrm{Hz}$ for the first band and a total number of 9 bands. Since the geophones have an eigenfrequency of 14 Hz, only signals with a higher frequency are recorded without any loss of information. However, preliminary results showed that half-octave bands with a central frequency higher than $f_{\min} = 5\,\mathrm{Hz}$ are adequate.

4.3 Post-processing

The HMM classification resulted in several hundred events in the avalanche class. Many of these events were however of very short duration or only identified at one sensor and did not necessarily coincide with avalanches in the reference catalogue, i.e. these events were likely false alarms. We therefore investigated three post-processing methods to reduce the number of false alarms, namely:

1. Applying a duration threshold for the detected events
2. Analyzing the results of all sensors by introducing a voting based classification
3. Analyzing the coherence between all sensors for each detected event

First, we used an event duration threshold. The duration T_j of any automatically detected event j was determined by the HMM. Based on the analysis of van Herwijnen et al. (2013) and van Herwijnen et al. (2016), we can assume that event duration correlates with avalanche size. The first post-processing method therefore consisted of using a minimal duration $T_{\min}$ for the events as described below in 5.3.1. Any automatically detected event j with $T_j \leq T_{\min}$ was thus removed. Similarly, any avalanche i in the reference catalogue with a duration $T_i \leq T_{\min}$ was also removed.

Second, we used a voting based threshold by tallying the classified events of each sensor. Rubin et al. (2012) used a similar approach and found that with increasing votes the false alarm rate decreased. The overall idea is that although an avalanche event might not be recorded by one sensor, for instance due to poor coupling of the sensor, it is unlikely that an avalanche is missed by all sensors, especially larger avalanches (Faillettaz et al., 2016). Any automatically detected event with $V_j \leq V_{\min}$ with V the number of votes was removed.

Third, we used a threshold based on the cross correlation coefficient between the seven sensors. Wave-fields generated by avalanches should be relatively coherent, while wave fields generated by noise (e.g. wind) are expected to be incoherent. We therefore divided the seismic data in non-overlapping windows of 1024 samples and for each window we defined a mean normalized correlation coefficient

$$R(t_{\text{win}}) = \frac{1}{N_{pairs}} \sum_{k=1}^{N_{pairs}} r_{kl}(t_{\text{win}}) \tag{2}$$

with $N_{\text{pairs}} = 21$ the number of sensor pairs, $r_{kl}(t_{\text{win}})$ the maximum in the normalized cross correlation between sensor k and l and t_{win} the time of the sliding window. The normalized cross correlation is defined as:

$$\overline{\phi}_{kl}(t) = \frac{\phi_{kl}(t)}{\sqrt{\phi_{kk}(0)\phi_{ll}(0)}} \tag{3}$$

with $\phi_{kk}(0)$ and $\phi_{ll}(0)$ the zero lag autocorrelation of each sensor which is equal to the energy of each single time window. The maximum of the normalized correlation is picked for a maximum lag of $t_{\max} = 0.05\,\mathrm{s}$, which is the time a sonic wave-field at a speed of $330\,\mathrm{m/s}$ needs to travel the maximum distance between the most distant receiver pair ($\approx 15\,\mathrm{m}$):

$$r_{kl}(t_{\text{win}}) = \max \overline{\phi}_{kl}(t), |t| \leq 0.05\,\mathrm{s} \tag{4}$$

The normalized cross correlation only yields values between -1 (perfectly anti-correlated) and 1 (perfectly correlated). A value of 0 means that the signals are completely uncorrelated. Finally, the coherence of each automatically detected event was defined as

$$C_j = \max R(t_{\mathrm{win}}), 0 \leq t_{\mathrm{win}} \leq T_j \tag{5}$$

Any automatically detected event with $C_j \leq C_{\mathrm{min}}$ was removed.

4.4 Model performance evaluation

To evaluate the performance of the HMM classification, we compared the automatic picks with the reference data set described in section 3.2. To assign an event classified by the HMM as a positive detection we defined a tolerance interval d: the time t_j^{HMM} of an event had to be within the interval $t_i - d \leq t_j^{HMM} \leq t_i + d$, with t_i the release time of the i-th avalanche in the reference data set and $d = 60\,\mathrm{s}$. The tolerance interval d was necessary, since the release times t_i of the avalanches were picked manually and may contain some uncertainties. In addition, the releases times t_j^{HMM} do not necessarily coincide with the reference data since the classifier is not an onset picker.

To describe the performance of the classifier we used three values N_{hit}, $N_{\mathrm{unassigned}}$ and N_{miss}. The first value N_{hit} describes the total number of avalanches which were correctly detected by the classifier, i.e. events identified by the classifier which corresponded to avalanches in the reference data set. The second value $N_{\mathrm{unassigned}}$ describes the number of events identified by the classifier which did not correspond to an avalanche in the reference data set. We do not call these events false alarms as during the manual detection some avalanche events might have been missed that are therefore not present in the reference data set. Avalanche events may still be found in the unassigned detections. Finally, the third value N_{miss} describes the number of avalanches in the reference data that were not identified by the HMM classifier. The three values were used to evaluate the overall model performance in terms of probability of detection (POD) and false alarm ratio (FAR), defined as $POD = \frac{N_{\mathrm{hit}}}{N_{\mathrm{hit}} + N_{\mathrm{miss}}}$ and $FAR = \frac{N_{\mathrm{unassigned}}}{N_{\mathrm{hit}} + N_{\mathrm{unassigned}}}$ (Wilks, 2011).

To determine the best threshold values for the post-processing steps (section 4.3), we plotted POD against FAR values for all probability classes and for different values of T_{min}, v_{min} and C_{min}. One curve therefore illustrates the POD and FAR with respect to the probability classes for a fixed threshold value. Ideally, when only considering the 100% probability class, the POD value should be 1 while the FAR value should also be relatively high since all detections of the lower classes are counted as false alarms. By taking more probability classes into account, the FAR will decrease while the POD value should stay close to 1. A perfect model should therefore result in constant POD values of 1 whereas the FAR values should decrease to 0. For realistic models, however, POD and FAR values are expected to decrease when accounting for more probability classes. By plotting POD against FAR values for different threshold values, the optimal threshold value can be found by searching for the largest area under the curve (AUC). A perfect model would result in an AUC value of 1, while an AUC value of 0.5 corresponds to a model with a 50-50 chance. Threshold values obtained for models with an AUC value lower than 0.5 are therefore not reliable and any random threshold value can be chosen. In our case, the POD-FAR curves did not span from 0 to 1 on the x-axis. To determine the AUC value we therefore calculated the ratio between the area under the POD-FAR curve and the area

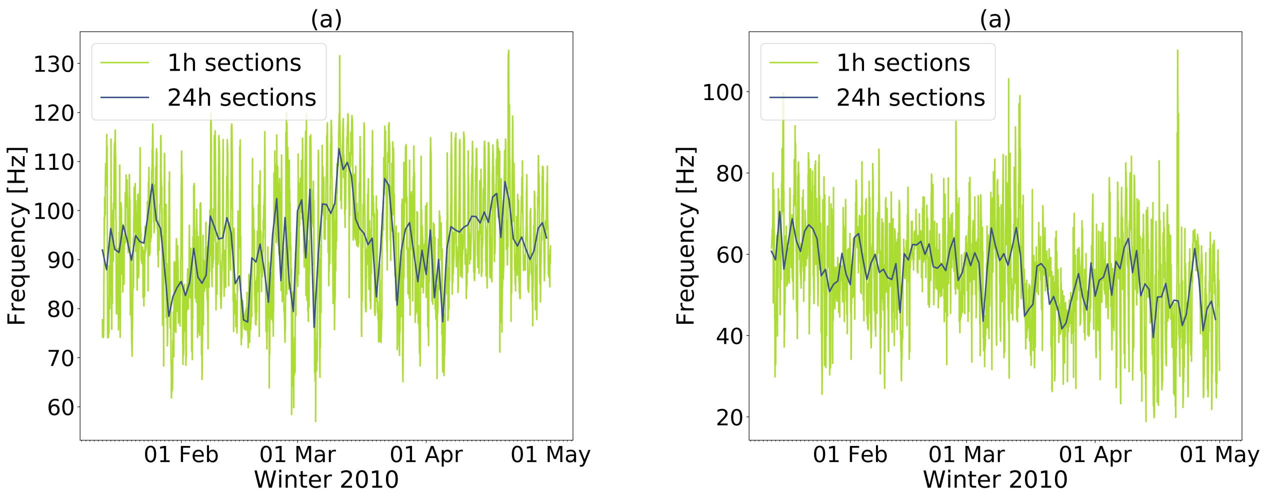

Figure 5. (a) Temporal variations in the central frequency hourly average (yellow) and daily average (blue). (b) Temporal variations in the dominant frequency.

under the bisector with the same x-axis limits

$$AUC = \frac{AUC_{\text{hmm}}}{AUC_{\text{bisector}}} \frac{AUC_{\text{HMM}}}{AUC_{\text{bisector}}} \cdot 0.5. \tag{6}$$

5 Results

5.1 Temporal feature distribution

To investigate changes in feature distribution over time, for instance, due to diurnal changes in environmental noise levels or seasonal changes in snow cover properties, we calculated hourly and daily mean values for all features (see Figure 5 for the central frequency and dominant frequency). Throughout the season there were large variations in the feature distribution at various time scales. First, there were strong diurnal variations (yellow lines in Figure 5). These were observed in all features (not shown). Second, there were also large variations at longer time scales (blue lines in Figure 5). While for some features there were significant seasonal trends, for instance, for the dominant frequency (Pearson $r = -0.56$, $p < 0.001$; Figure 5b), for others there were no clear trends, for instance, for the central frequency (Pearson $r = 0.23$, $p = 0.02$; Figure 5a). Building a single background model to classify the entire season would therefore likely not result in a reliable classification. The background model thus has to be regularly recalculated. We therefore decided to recalculate the background model each day to classify the events within the same day.

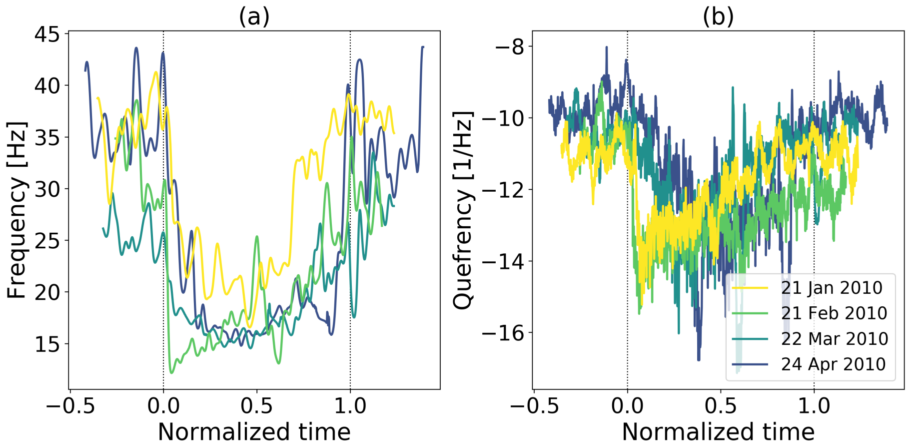

Figure 6. (a) Central frequency with normalized time for 4 different avalanche events (colours) from 4 different months. For comparison, the time was normalized by the event duration with 0 indicating the start of the avalanche and 1 the end of the event. (b) 2^{nd} cepstral coefficient.

5.2 Training event

While building a representative background model is important, choosing an appropriate training event is of utmost importance for the classifier. As outlined in van Herwijnen et al. (2016), the avalanche catalogue consists of various avalanches of different size and type. At the beginning of the season the avalanches are most likely dry-snow avalanches, while at the end of the season there a mostly wet-snow avalanches. Our avalanche catalogue only consists of the release time and little information is available on the type of the avalanches. We therefore compared the feature distribution of four different avalanche events, of 21 January, 27 February, 22 March and 24 April 2010 to investigate whether substantial differences related to avalanche type existed (Figure 6).

While there were some subtle differences in the feature distribution for the four avalanches (e.g. between the avalanches on 21 January 2010 and 22 March 2010 in Figure 6 (a)), overall the four avalanches exhibited very similar behaviour. Thus, when viewing avalanches in the feature space, wet- and dry-snow avalanches appear to be very similar. We therefore used one single avalanche class for the HMM classifier and used one training event to learn the model. Specifically, we used an avalanche with $P_{\text{ava}} = 100\%$ recorded on 22 March 2010 (Figure 7) with a duration of $30\,\text{s}$. However, as seen in Figure 6, the most rapid changes in feature values occurred at the beginning of the events. In the coda changes in feature values were rather slow, providing limited relevant information for the classifier. We therefore only used the first 8 s of the training event (marked by the red rectangle in Figure 7).

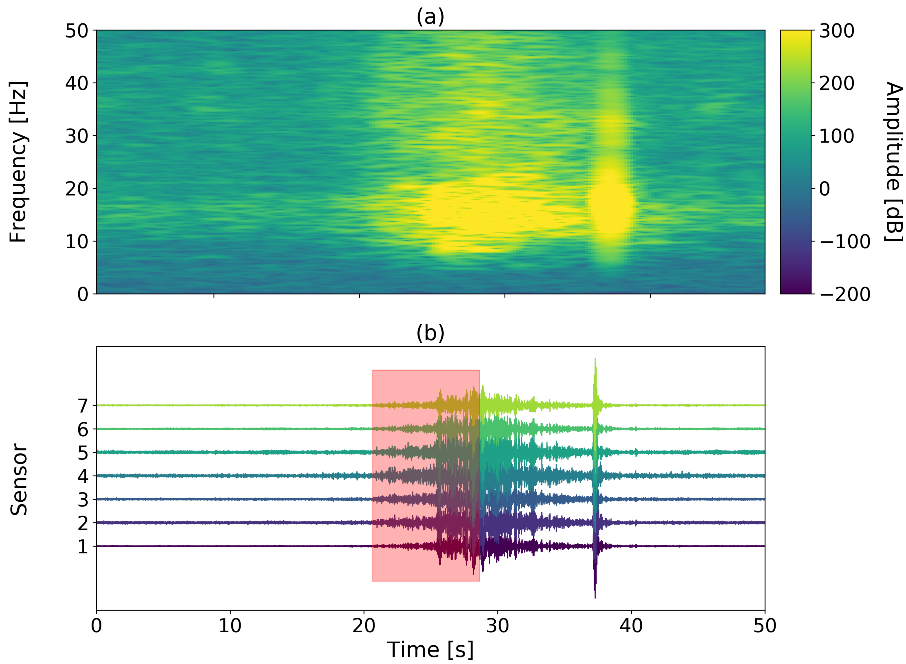

Figure 7. The event used for training the HMM. (a) Stacked spectrogram of all 7 sensors. (b) Seismic waveform for each individual sensor (colours). The red area highlights the part of the signal used as the training event for the HMM.

5.3 Automatic avalanche classification

5.3.1 Single sensor classification

We used the entire reference data set containing 283 avalanche events to evaluate model performance as function of probability class.

The first model was built for each individual sensor and without any post-processing. For each sensor a separate model was built containing only the data of the specific sensor. In Table 2 the number of detections for each probability class is listed. For the highest probability class the POD values were relatively high, ranging from 70% to 95% and the values generally decreased with decreasing probability class. For the confirmed avalanches, the probability of detection ranged between 80% and 92%. Nevertheless, even for the lowest probability class events were still detected. Furthermore, numerous events, between 124 and 2091 events, were detected for each sensor which were not listed in the reference data set. Clearly without any post-processing the number of unassigned events ~~is~~ was high.

To reduce the number of unassigned events we applied a minimum duration $T_{\min}$ to remove events. To obtain a reasonable threshold value, we determined the area under the POD-FAR curve for different $T_{\min}$ values (Figure 8).

Table 2. Number of detections (N_{hit}) and probability of detection (POD) for each sensor and each probability class as well as the POD for confirmed events Evt_{conf} and number of events that were not in the reference data set ($N_{\mathrm{unassigned}}$).

Probability Class		0%	16.6%	33.3%	50%	66.6%	83.3%	100%	Evt_{conf}	$N_{\mathrm{unassigned}}$
Sensor 1	POD	$\frac{12}{58}$ 21%	$\frac{17}{66}$ 26%	$\frac{21}{48}$ 44%	$\frac{17}{34}$ 50%	$\frac{14}{30}$ 47%	$\frac{21}{27}$ 78%	$\frac{19}{20}$ 95%	$\frac{23}{25}$ 92%	149
Sensor 2	POD	$\frac{28}{58}$ 48%	$\frac{24}{66}$ 36%	$\frac{19}{48}$ 40%	$\frac{11}{34}$ 32%	$\frac{17}{30}$ 57%	$\frac{17}{27}$ 63%	$\frac{14}{20}$ 70%	$\frac{21}{25}$ 84%	1432
Sensor 3	POD	$\frac{21}{58}$ 36%	$\frac{25}{66}$ 38%	$\frac{21}{48}$ 44%	$\frac{12}{34}$ 35%	$\frac{16}{30}$ 53%	$\frac{17}{27}$ 63%	$\frac{15}{20}$ 75%	$\frac{21}{25}$ 84%	1347
Sensor 4	POD	~~$\frac{20}{58}$~~ $\frac{11}{58}$ ~~35~~19%	~~$\frac{19}{66}$~~ $\frac{11}{66}$ ~~29~~17%	~~$\frac{18}{48}$~~ $\frac{16}{48}$ ~~38~~33%	~~$\frac{9}{34}$~~ $\frac{14}{34}$ ~~27~~41%	$\frac{15}{30}$ 50%	~~$\frac{15}{27}$~~ $\frac{20}{27}$ ~~56~~74%	~~$\frac{14}{20}$~~ $\frac{18}{20}$ ~~70~~90%	$\frac{21}{25}$ 84%	124
Sensor 5	POD	$\frac{22}{58}$ 38%	$\frac{27}{66}$ 41%	$\frac{17}{48}$ 35%	$\frac{13}{34}$ 38%	$\frac{19}{30}$ 63%	$\frac{17}{27}$ 63%	$\frac{15}{20}$ 75%	$\frac{21}{25}$ 84%	786
Sensor 6	POD	~~$\frac{11}{58}$~~ $\frac{20}{58}$ ~~19~~35%	~~$\frac{11}{66}$~~ $\frac{19}{66}$ ~~17~~29%	~~$\frac{16}{48}$~~ $\frac{18}{48}$ ~~33~~38%	~~$\frac{14}{34}$~~ $\frac{9}{34}$ ~~41~~27%	$\frac{15}{30}$ 50%	~~$\frac{20}{27}$~~ $\frac{15}{27}$ ~~74~~56%	~~$\frac{18}{20}$~~ $\frac{14}{20}$ ~~90~~70%	$\frac{20}{25}$ 80%	2091
Sensor 7	POD	$\frac{29}{58}$ 50%	$\frac{26}{66}$ 39%	$\frac{23}{48}$ 48%	$\frac{16}{34}$ 47%	$\frac{19}{30}$ 63%	$\frac{17}{27}$ 63%	$\frac{14}{20}$ 70%	$\frac{21}{25}$ 84%	2094

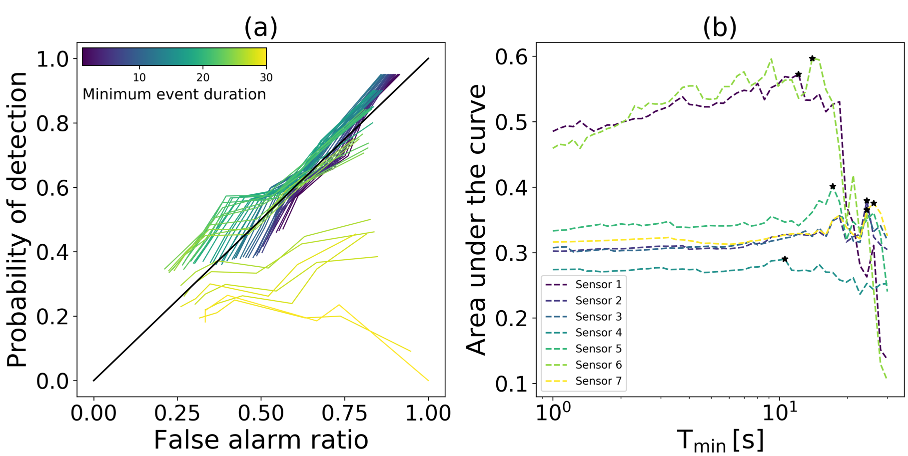

Figure 8. (a) POD-FAR curves for sensor 1 for different minimum event durations T_{min} (colours). (b) Area under the POD-FAR curve with T_{min} for all sensors (colours). The stars show the T_{min} value with the largest area under the curve for each sensor.

Table 3. Number of detections (N_{hit}) and probability of detection (POD) for each sensor and each probability class as well as the POD for confirmed events Evt_{conf} and number of events that were not in the reference data set ($N_{\mathrm{unassigned}}$) using a minimal event duration of $T_{\mathrm{min}} = 12\,\mathrm{s}$.

Probability classes		0%	16.6%	33.3%	50%	66.6%	83.3%	100%	Evt_{conf}	$N_{\mathrm{unassigned}}$
Sensor 1	POD	$\frac{0}{0}$ 0%	$\frac{4}{39}$ 10%	$\frac{6}{36}$ 17%	$\frac{11}{23}$ 48%	$\frac{10}{26}$ 39%	$\frac{15}{27}$ 56%	$\frac{16}{19}$ 84%	$\frac{20}{25}$ 80%	22
Sensor 2	POD	$\frac{0}{0}$ 0%	$\frac{17}{39}$ 44%	$\frac{12}{36}$ 33%	$\frac{7}{23}$ 30%	$\frac{14}{26}$ 54%	$\frac{17}{27}$ 63%	$\frac{14}{19}$ 74%	$\frac{21}{25}$ 84%	408
Sensor 3	POD	$\frac{0}{0}$ 0%	$\frac{14}{39}$ 36%	$\frac{13}{36}$ 36%	$\frac{9}{23}$ 39%	$\frac{13}{26}$ 50%	$\frac{16}{27}$ 59%	$\frac{14}{19}$ 74%	$\frac{20}{25}$ 80%	340
Sensor 4	POD	$\frac{0}{0}$ 0%	$\cancel{\frac{7}{39}}$ $\frac{3}{38}$ ~~18~~8%	$\cancel{\frac{13}{36}}$ $\frac{5}{35}$ ~~36~~14%	$\cancel{\frac{5}{23}}$ $\frac{6}{22}$ ~~22~~27%	$\cancel{\frac{10}{26}}$ $\frac{8}{26}$ ~~39~~31%	$\cancel{\frac{14}{27}}$ $\frac{12}{26}$ ~~52~~46%	$\cancel{\frac{13}{19}}$ $\frac{15}{19}$ ~~69~~79%	$\frac{15}{25}$ 60%	15
Sensor 5	POD	$\frac{0}{0}$ 0%	$\frac{16}{39}$ 41%	$\frac{11}{36}$ 31%	$\frac{9}{23}$ 39%	$\frac{15}{26}$ 58%	$\frac{14}{27}$ 52%	$\frac{13}{19}$ 68%	$\frac{21}{25}$ 84%	157
Sensor 6	POD	$\frac{0}{0}$ 0%	$\cancel{\frac{3}{38}}$ $\frac{7}{39}$ ~~8~~18%	$\cancel{\frac{5}{35}}$ $\frac{13}{36}$ ~~14~~36%	$\cancel{\frac{6}{22}}$ $\frac{5}{23}$ ~~27~~22%	$\cancel{\frac{8}{26}}$ $\frac{10}{26}$ ~~31~~39%	$\cancel{\frac{12}{26}}$ $\frac{14}{27}$ ~~46~~52%	$\cancel{\frac{15}{19}}$ $\frac{13}{19}$ ~~79~~69%	$\frac{18}{25}$ 72%	597
Sensor 7	POD	$\frac{0}{0}$ 0%	$\frac{17}{39}$ 43%	$\frac{14}{36}$ 39%	$\frac{10}{23}$ 44%	$\frac{15}{26}$ 58%	$\frac{17}{27}$ 63%	$\frac{14}{19}$ 74%	$\frac{21}{25}$ 84%	653

Due to the large number of unassigned events (see Table 2) AUC values were generally below 0.5 and using a minimum duration threshold did not result in much improvement. However, for sensors 1 and 6 the number of unassigned events was much lower and there was an optimum T_{min} threshold value around $12\,\mathrm{s}$ (Figure 8 ~~(b)~~b). In the following, we therefore used the same minimal duration $T_{\mathrm{min}} = 12\,\mathrm{s}$ for all sensors. While with this T_{min} threshold the POD values for the highest probability class somewhat decreased, they still remained high (between 68% and 84%; Table 3). Furthermore, the probability of detection for the confirmed avalanches also remained relatively high, ranging from 60% to 84%. Note that the duration threshold was also applied to remove events from the reference data set. For a threshold value of $T_{\mathrm{min}} = 12\,\mathrm{s}$, the number of events in the reference data set reduced from 283 to ~~170.~~ 170 while the number of confirmed avalanches was not affected.

Overall the number of unassigned events substantially decreased (compare Tables 2 and 3), especially for sensors 1 and 6. For these two sensors the POD values also substantially decreased for the lower probability classes but the two main avalanche periods in March and April are clearly visible in the detections (see Figure 9a for sensor 1). However, for the other sensors, the POD values remained relatively high for the lower probability classes and there were still many unassigned events and the two main avalanche periods were less evident (see Figure 9b for sensor 7). Clearly, for some of the sensors the number of unassigned events remained very high.

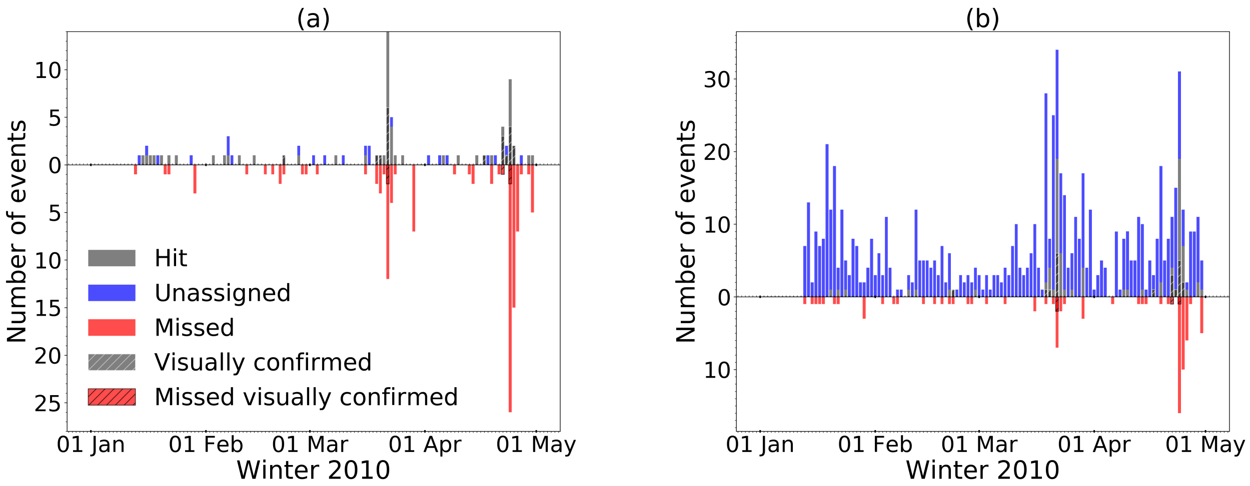

Figure 9. Classification results using a minimal event duration length of $12\,\mathrm{s}$ for (a) sensor 1 and (b) sensor 7. Gray bars show the number of matching detections (hit), blue bars show the unassigned events and the red bars show the events that were not detected as an avalanche (missed). The hatched bars indicate the number of hits (grey) or misses (red) for the confirmed events.

5.3.2 Array based classification

To further reduce the number of unassigned events (potential false alarms), we applied two array based post-processing methods to eliminate events, namely a minimum number of votes and coherence threshold (see Section 4.3). To define an optimal number of votes or coherence threshold we used the same procedure as for the determination of a minimal duration threshold (Figure 8). Since for these array based methods the detections from all sensors were pooled, the number of unassigned events was very high. This resulted in much higher FAR values and thus poor model performance with low AUC values, all below 0.5. Arbitrary values for $v_{\min}$ or $C_{\min}$ could therefore be chosen. However, the overall goal of these post-processing steps was to reduce the number of unassigned events, while still retaining a reasonable probability of detection. We therefore analysed the effect of different threshold values on FAR values as well as on POD values for the probability classes.

Overall, for both processing steps the number of unassigned events decreased with increasing threshold values (Figures 10 and 11). By using a minimal number of 5 votes the POD stayed relatively high with a low number of unassigned events. Following the same procedure we selected a coherence value of 0.6. Using this value, the POD was relatively high and the number of unassigned events could be reduced to less than 100 events.

In total, we thus have three different post-processing steps which can be applied to the data: a minimal event duration $T_{\min} = 12\,\mathrm{s}$, a minimum number of votes $v_{\min} = 5$ and a coherence threshold $C_{\min} = 0.6$. By combining these three steps, six array based post-processing work flows were implemented:

– $v : v_j \geq v_{\min}$

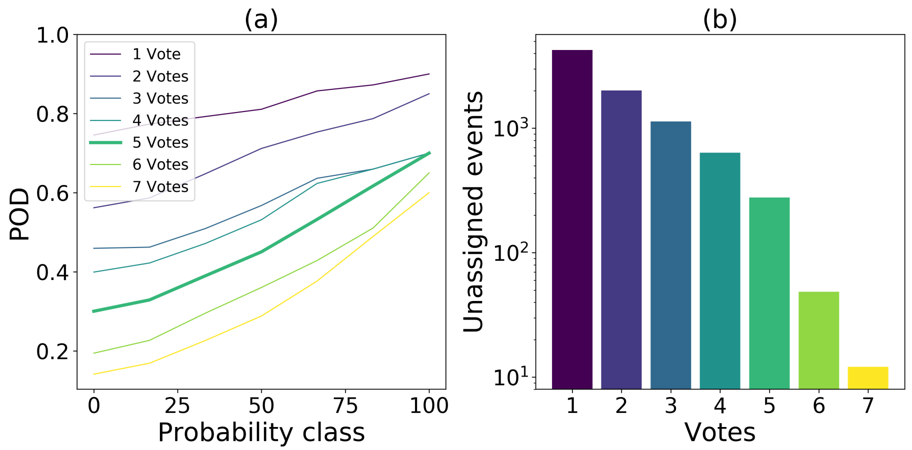

Figure 10. (a) POD for each probability class depending on the minimum number of votes (colours). (b) Number of unassigned events for different number of votes (colours).

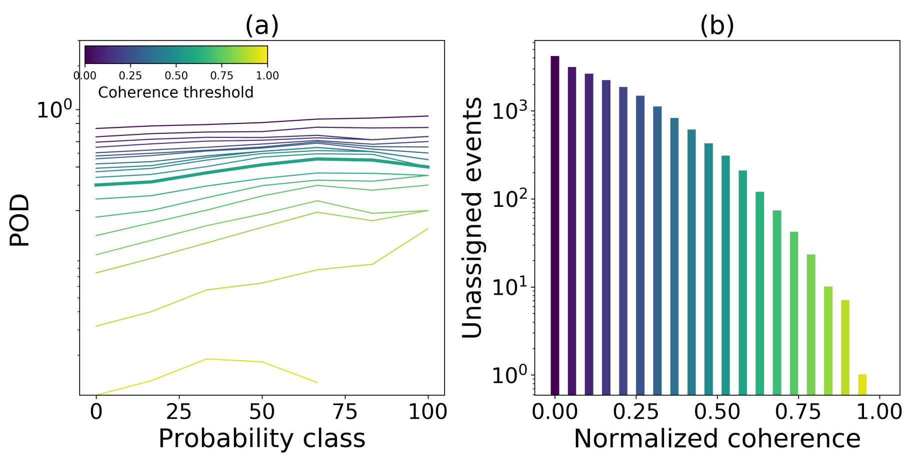

Figure 11. (a) POD for each probability class depending on the minimal coherence (colours). (b) Number of unassigned events for different coherence threshold values (colours).

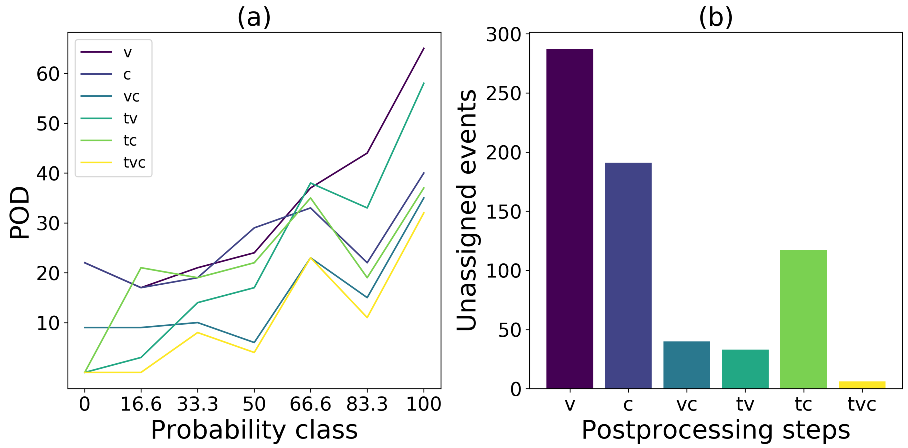

Figure 12. (a) POD for each probability class for different post-processing work flows (colours). v = minimal number of votes used, t = minimal event duration and c = minimal coherence. (b) Number of unassigned events remaining after the post processing for each work flow.

– $c : C_j \geq C_{\min}$

– $vc : v_j \geq v_{\min}$ and $C_j \geq C_{\min}$

– $tv : T_j \geq T_{\min}$ and $v_j \geq v_{\min}$

– $tc : T_j \geq T_{\min}$ and $C_j \geq C_{\min}$

– $tvc : T_j \geq T_{\min}, v_j \geq v_{\min}$ and $C_j \geq C_{\min}$

The number of unassigned events decreased most with a combined approach always including the number of votes (vc, tv, or tvc in Figure 12). When using only one array based post-processing step, the number of unassigned events remained high (v and c in Figure 12b). While the lowest number of unassigned events was achieved when combining all three post-processing steps, this model also resulted in low POD values for all probability classes (tvc Figure 12a). Overall, the highest POD values and the steepest decrease for the lowest probability classes were obtained for the voting based processing with and without a minimal duration of the events (v and tv in Figure 12a). For both these post-processing work flows the two periods of high avalanche activity are visible (Figure 13). However, by also applying the duration threshold, the total number of detections decreased and the two periods in March and April became more clearly visible (Figure 13b).

5.3.3 Unassigned events

We compared the automatic detections with the reference avalanche catalogue and obtained a large number of unassigned events. It remains unclear whether the unassigned events are all false alarms or partly correspond to avalanches that were not identified in the reference data. We therefore visually inspected the unassigned events. In order to keep this reanalysis

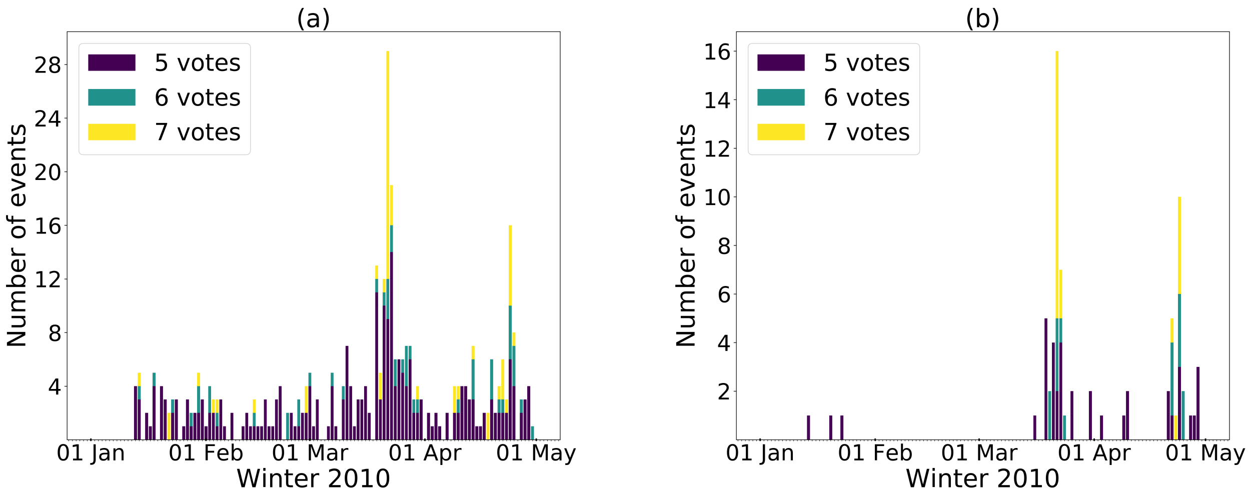

Figure 13. All events detected by at least 5 sensors. The different colours indicate the number of votes. (a) the results without any limitation of the duration of the events are plotted, (b) only events with a minimum duration as mentioned before are taken into account. On the right the two main avalanche periods are more clearly visible.

Table 4. Results of the reanalysis of the detections not covered by the test data set. Left side of the table shows the results of the reanalysis of two single sensors, the right side shows the results of the array based classification. For the single sensors a minimum duration for the events of $t_{\min} = 12s$ was taken into account. The voting based processing steps analysed are minimum number of votes (v), minimum duration (t) and coherence (c).

	Single sensors		Array based		
	Sensor 1	Sensor ~~6~~ 4	tv	vc	tvc
Unassigned	22	15	33	40	6
Confirmed as avalanche	36% (8)	47% (7)	12% (4)	28% (11)	33% (2)
Confirmed as false alarm	27% (6)	27% (4)	79% (26)	45% (18)	50% (3)
Uncertain	36% (8)	27% (4)	9% (3)	28% (11)	17% (1)

manageable, we only focused on the post-processing steps which resulted in less than 50 unassigned events. Thus, we ~~only~~ individually investigated single sensor results from sensor 1 and ~~6~~ 4 with $T_{\min} = 12\,\mathrm{s}$ and array based results for tv, vc and tvc. The visual inspection showed that for the single sensor results between 36% and 47% of the unassigned events were likely unidentified avalanches and less than one third were false alarms (Table 4). For the array based results, however, most of the unassigned events (between 45% and 79%) were false alarms while fewer events were likely associated with avalanches that were missed by van Herwijnen and Schweizer (2011a).

6 Discussion and Summary

We trained hidden Markov models (HMMs), a machine learning algorithm, to automatically detect avalanches in continuous seismic data recorded near an avalanche starting zone above Davos, Switzerland for the winter season of 2010. To reduce the amount of data to process, we pre-processed the continuous data using an amplitude threshold (Figure 2). We then implemented single sensor and array based post-processing steps and the performance of the models was evaluated using a previously published reference avalanche catalogue obtained from the same seismic data (van Herwijnen and Schweizer, 2011a; van Herwijnen et al., 2016).

After pre-processing the data, the reference avalanche catalogue contained 283 avalanches between 12 January and 30 April 2010, events that were identified by visual inspection of the waveform and spectrogram of a single sensor (van Herwijnen and Schweizer, 2011a). Since only ~~33~~ 25 of these events were independently confirmed avalanches, considerable uncertainty remained about the identified events. To reduce the uncertainty in the reference catalogue, three of the authors therefore re-evaluated the data. This allowed us to assign 7 subjective probability classes between 0% and 100% to each event. Overall, only 20 events were marked as certain avalanche (Table 1) and hence the performance ~~,~~ of the classifiers can only be evaluated for these particular events. For the remaining events, there are still uncertainties and hence the performance of the classifier can only be estimated. Furthermore, this reanalysis highlighted the difficulty in obtaining an objective and reliable reference avalanche catalogue. It also showed that expert decisions are biased and there is a need for a reliable automatic classifier to identify avalanches in continuous seismic data.

Recent work by Hammer et al. (2017) showed very promising results for applying a HMM to automatically detect avalanches in continuous seismic data. While they only focussed on a five day period during an exceptional avalanche cycle in 1999, our goal was to classify continuous seismic data spanning more than 100 days. This prevented us from building a single background model to classify the entire season since temporal variations in feature distributions at various time scales were present (Figure 5). Indeed, when using a single background model to classify the entire season for sensor 1 only two thirds of the events were detected by having almost 6 times the number of unassigned events. One possible reason for these variations in feature distribution was likely the setup of the sensor array. The geophones were packed in a Styrofoam housing and inserted within the snowpack. As such, less snow covered the sensors than if they has been inserted in the ground, making them more susceptible to environmental noise. Furthermore, it is also likely that the snow cover introduced additional noise in spring due to the rapid settlement and water infiltration. We therefore recalculated the background model for each day and for each sensor to classify the data from the same day. However, for the operational implementation this would be impractical, since there would always be a 24 hour delay in the detections. Other strategies for regularly updating the background model should therefore be investigated (e.g. Riggelsen and Ohrnberger, 2014).

We performed the automatic classification over the entire season by recalculating the classifier for each day and for each sensor. Overall, probability of detection (POD) values decreased with decreasing probability class and the highest POD values were associated with the highest probability class for all sensors (Table 2). Indeed, between 70% and 95% of all avalanches in the highest probability class were detected, which is comparable to the results presented by Bessason et al. (2007) and

Leprettre et al. (1996), who reported POD values of approximately 65% and 90%. Nevertheless, without any post-processing, the number of unassigned events was high, questioning the reliability of the models as many of these events were likely false alarms. Post-processing of the results was therefore required. Applying a minimal signal duration drastically reduced the number of unassigned events while still retaining reasonable POD values, in particular for sensor 1 and sensor ~~6~~ 4 (Table 3). However, there were large differences in model performance between the sensors (Figures 8 and 9). The reason for these performance differences is very likely the deployment of the sensors. Indeed, sensor 1 and ~~6~~ 4 were deployed at the top of the slope closest to a cornice where the snow was the deepest (van Herwijnen and Schweizer, 2011a). The other 5 sensors were covered by less snow due to local inhomogeneities, leaving these sensors more sensitive to environmental noise. For future deployments it will thus be important to deploy the sensors below a homogeneous snow cover and not within the snow cover. This should reduce the amount of environmental noise and consequently the number of false alarms.

To further reduce the number of false alarms, we implemented two array based post-processing steps, namely a voting based approach and a signal coherence threshold. In combination with the minimal event duration, we thus investigated 6 array based post-processing work flows. Results showed that these array based methods were effective in reducing the number of unassigned events (Figure 12). However, the POD values generally also decreased, resulting in overall fewer detections. Combined post-processing methods which included the voting based approach resulted in better model performance, in line with results presented by Rubin et al. (2012). The best model performance was obtained by combining the event duration threshold for events with at least 5 votes. The number of unassigned events reduced to about 30 and POD values were highest ($\sim$ 55%) for the highest probability class and decreased for the lower classes. Despite the large differences in model performance for the individual sensors, the model still performed marginally better when pooling the data from the entire array. These results are promising as with an improved sensor deployment strategy array based post-processing is likely to further improve.

Comparing our model performance to previously published studies is not straightforward. We assigned subjective probability classes to our reference avalanche catalogue rather than using a yes or no approach. Furthermore, we used geophones deployed in an avalanche starting zone, while Bessason et al. (2007), Leprettre et al. (1996) and Hammer et al. (2017) used sensitive broadband seismometers deployed at valley bottom. Therefore, it is very likely that there was more environmental noise in our data and many of the detected avalanches in our reference data set were rather small (van Herwijnen et al., 2016). Given these differences in instrumentation and deployment, our detection results are encouraging and highlight the advantage of using HMMs for the automatic identification of avalanches in continuous seismic data. ~~Indeed, our model only required~~

The main advantages of the proposed approach is that only one training event (Figure 7) is needed to classify the entire season. As shown by Hammer et al. (2017), for large avalanches it is possible to build a HMM with a high POD and very low FAR with one training event. ~~As such, HMMs are well suited to detect avalanchesas they~~ Even though we used less sensitive sensors in this work, we were also able to identify periods of high avalanche activity (compare Figure 3 with Figure 9 and 13). Furthermore, when only considering the visually confirmed avalanches, the probability of detection was typically around 80% (see Tables 2 and 3). This suggests that HMMs can easily be implemented at new sites. In contrast, the model used by Bessason et al. (2007) relied on a 10-year data base, and Leprettre et al. (1996) used a set of fuzzy logic rules derived by the experts. Note that the post-processing steps we investigated are likely site-dependent, in particular the event duration threshold. However, such a threshold value is intuitive, has a linear influence on model outcome and is thus easily tunable.

For operational use, the model should be able to automatically detect avalanches in near real-time. The main disadvantages of the proposed approach is its computational cost. The feature calculation for one day takes $\approx 1\,\mathrm{h}$ for the pre-processed data and $\approx 7\,\mathrm{h}$ for the unprocessed data. Replacing the used features with computationally less expensive attributes would decrease the processing time drastically and encourage real-time applications. In this work, we classified the data from each sensor individually, requiring a separate background model for each sensor. The results from the different sensors were then combined using post-processing rules either on a voting based approach or taking the coherence into account. Strictly speaking, the coherence could have been added to the model as an additional feature. However, calculating the coherence for 21 receiver pairs, even after applying the amplitude threshold during pre-processing, was still very time consuming ($\approx 200\%$ real time).

Overall, our results suggest that HMMs may be well suited for the automatic detection of avalanches in ~~a~~ continuous seismic data~~for operational avalanche forecasting~~. The variable model performance between the different sensors highlighted ~~some~~ problems which can likely be overcome by improving the sensor deployment strategy. Specifically, we suggest that the sensors should be deployed 30 to 50 cm underground at a site with a homogeneous and preferably thick snow cover ~~. Furthermore,~~ and to increase the distance between the sensors ~~should be increased~~ to apply array processing techniques for source localization (Lacroix et al., 2012). In addition, further avalanche events may be used for training to improve model performance. Finally, incorporating localization parameters as new features in the HMM could open the door for further model improvement, as is done for the automatic detection of avalanches in continuous infrasound data (Marchetti et al., 2015; Thüring et al., 2015). These features can then either be implemented directly into the HMM or be used in additional post-processing steps.

Acknowledgements. M.H. was supported by a grant of the Swiss National Science Foundation (200021_149329). We thank numerous colleagues from SLF for help with field work and maintaining the instrumentation.

Competing interests. The authors declare that they have no conflict of interest.

References

Bessason, B., Eiriksson, G., Thorarinsson, O., Thorarinsson, A., and Einarsson, S.: Automatic detection of avalanches and debris flows by seismic methods, J. Glaciol., 53, 461–472, 2007.

Beyreuther, M., Hammer, C., Wassermann, J., and Ohrnberger, M.: Constructing a Hidden Markov Model based earthquake detector: application to induced seismicity, Geophys. J. Int., 189, 602–610, 2012.

Caplan-Auerbach, J. and Huggel, C.: Precursory seismicity associated with frequent, large ice avalanches on Iliamna volcano, Alaska, USA, J. Glaciol., 53, 128–140, doi:10.3189/172756507781833866, 2007.

Dammeier, F., Moore, J. R., Hammer, C., Haslinger, F., and Loew, S.: Automatic detection of alpine rockslides in continuous seismic data using hidden Markov models, J. Geophys. Res. Earth Surf, 121, 351–371, 2016.

Faillettaz, J., Funk, M., and Vincent, C.: Avalanching glacier instabilities: Review on processes and early warning perspectives, Rev. Geophys., 53, 203–224, doi:10.1002/2014RG000466, http://dx.doi.org/10.1002/2014RG000466, 2014RG000466, 2015.

Faillettaz, J., Or, D., and Reiweger, I.: Codetection of acoustic emissions during failure of heterogeneous media: New perspectives for natural hazard early warning, Rev. Geophys., 43, 1075–1083, doi:10.1002/2015GL067435, http://dx.doi.org/10.1002/2015GL067435, 2015GL067435, 2016.

Hammer, C., Beyreuther, M., and Ohrnberger, M.: A seismic-event spotting system for Volcano fast-response systems, B. Seismol. Soc. Am., 102, 948–960, 2012.

Hammer, C., Ohrnberger, M., and Fäh, D.: Classifying seismic waveforms from scratch: a case study in the alpine environment, Geophys. J. Int., 192, 425–439, 2013.

Hammer, C., Fäh, D., and Ohrnberger, M.: Automatic detection of wet-snow avalanche seismic signals, Nat. Hazards, 86, 601–618, doi:10.1007/s11069-016-2707-0, http://dx.doi.org/10.1007/s11069-016-2707-0, 2017.

Harrison, J.: Seismic signals from avalanches, Armstrong and Ives (Eds.), Avalanche release and snow characteristics. Institute of Arctic and Alpine Research, University of Colorado. Occasional Paper No. 19, pp. 145–150, 1976.

Kishimura, K. and Izumi, K.: Seismic signals induced by snow avalanche flow, Nat. Hazards, 15, 89–100, 1997.

Lacroix, P., Grasso, J.-R., Roulle, J., Giraud, G., Goetz, D., Morin, S., and Helmstetter, A.: Monitoring of snow avalanches using a seismic array: Location, speed estimation, and relationships to meteorological variables, J. Geophys. Res. Earth Surf, 117, F01 034, doi:10.1029/2011JF002106, 2012.

Leprettre, B., Martin, N., Glangeaud, F., and Navarre, J.: Three-Component Signal Recognition Using Time, Time–Frequency, and Polarization Information - Application to Seismic Detection of Avalanches, IEEE Trans. Signal Process., 46, 83–102, 1998.

Leprettre, B., Navarre, J., and Taillefer, A.: First results from a pre-operational system for automatic detection and recognition of seismic signals associated with avalanches, J. Glaciol., 42, 352–363, 1996.

Marchetti, E., Ripepe, M., Ulivieri, G., and Kogelnig, A.: Infrasound array criteria for automatic detection and front velocity estimation of snow avalanches: towards a real-time early-warning system, Nat. Hazards, 15, 2545 – 2555, 2015.

Ohrnberger, M.: Continuous automatic classification of seismic signals of volcanic origin at Mt. Merapi, Java, Indonesia, PhD. thesis, 2001.

Pérez-Guillén, C., Sovilla, B., E. Suriñach, E., Tapia, M., and Köhler, A.: Deducing avalanche size and flow regimes from seismic measurements, Cold Reg. Sci. Tech., 121, 25–41, 2016.

Podolskiy, E. A. and Walter, F.: Cryoseismology, Rev. Geophys., 54, 708–758, doi:10.1002/2016RG000526, http://dx.doi.org/10.1002/2016RG000526, 2016RG000526, 2016.

Rabiner, L.: A tutorial on Hidden Markov Models and selected application in speech recognition, Proceedings of the IEEE, 77, 257–286, 1989.

Riggelsen, C. and Ohrnberger, M.: A Machine Learning Approach for Improving the Detection Capabilities at 3C Seismic Stations, Pure Appl. Geophys., 171, 395–411, doi:10.1007/s00024-012-0592-3, 2014.

Rubin, M., Camp, T., van Herwijnen, A., and Schweizer, J.: Automatically detecting avalanche events in passive seismic data, IEEE International Conference on Machine Learning and Applications, 1, 13–20, 2012.

Sabot, F., Naaim, M., Granada, F., Suriñach, E., Planet, P., and Furada, G.: Study of avalanche dynamics by seismic methods, image-processing techniques and numerical models, Ann. Glaciol., 26, 319–323, 1998.

Schaerer, P. A. and Salway, A. A.: Seismic and impact-pressure monitoring of flowing avalanches, J. Glaciol., 26, 179–187, 1980.

St. Lawrence, W. and Williams, T.: Seismic signals associated with avalanches, J. Glaciol., 17, 521–526, 1976.

Suriñach, E., Furdada, G., Sabot, F., Biescas, B., and Vilaplana, J.: On the characterization of seismic signals generated by snow avalanches for monitoring purposes, Ann. Glaciol., 32, 268–274, doi:10.3189/172756401781819634, 2001.

Suriñach, E., Sabot, F., Furdada, G., and Vilaplana, J.: Study of seismic signals of artificially released snow avalanches for monitoring purposes, Phys. Chem. Earth Pt. B, 25, 721–727, 2000.

Suriñach, E., Vilajosana, I., Khazaradze, G., Biescas, B., Furdada, G., and Vilaplana, J.: Seismic detection and characterization of landslides and other mass movements, Nat. Hazards Earth Syst. Sci., 5, 791–798, 2005.

Thüring, T., Schoch, M., van Herwijnen, A., and Schweizer, J.: Robust snow avalanche detection using supervised machine learning with infrasonic sensor arrays, Cold Reg. Sci. Tech., 111, 60 – 66, 2015.

van Herwijnen, A. and Schweizer, J.: Seismic sensor array for monitoring an avalanche start zone: design, deployment and preliminary results, J. Glaciol., 57, 257–264, 2011a.

van Herwijnen, A. and Schweizer, J.: Monitoring avalanche activity using a seismic sensor, Cold Reg. Sci. Tech., 69, 165–176, 2011b.

van Herwijnen, A., Dreier, L., and Bartelt, P.: Towards a basic avalanche characterization based on the generated seismic signal, Proceedings 2013 International Snow Science Workshop, Grenoble, France, pp. 1033–1037, 2013.

van Herwijnen, A., Heck, M., and Schweizer, J.: Forecasting snow avalanches by using highly resolved avalanche activity data obtained through seismic monitoring, Cold Reg. Sci. Tech., 132, 68–80, 2016.

Vilajosana, I., Khazaradze, G., Surinach, E., Lied, E., and Kristensen, K.: Snow avalanche speed determination using seismic methods, Cold Reg. Sci. Tech., 49, 2–10, doi:10.1016/j.coldregions.2006.09.007, 2007a.

Vilajosana, I., Suriñach, E., Khazaradze, G., and Gauer, P.: Snow avalanche energy estimation from seismic signal analysis, Cold Reg. Sci. Tech., 50, 72–85, doi:10.1016/j.coldregions.2007.03.007, 2007b.

Wilks: Statistical methods in the atmospheric sciences, vol. 100, Academic press, 2011.

Zobin, V. M., Plascencia, I., Reyes, G., and Navarro, C.: The characteristics of seismic signals produced by lahars and pyroclastic flows: Volcán de Colima, México, Journal of Volcanology and Geothermal Research, 179, 157 – 167, doi:10.1016/j.jvolgeores.2008.11.001, 2009.