EC: Editor Comment
AC: Author Comment

1 Response to EC1

AC: Dear Jens Turowski
thank you very much for your feedback. We addressed your comments in the following and responded to each referee individually. Attached to this response you find a preliminary version of the revised manuscript and a version highlighting all modifications.

EC: Dear authors,
I agree with the reviewers that the papers presents a potentially very useful contribution, however, currently there are weaknesses and the presentation (language and structure).
EC: When revising the paper, I ask you specifically to think from the perspective of a potential user. Is all the information available to reproduce the analysis? Can the reasoning be followed easily? Can the necessary information be clearly accessed in the manuscript? Is the approach described separately from specific features of the case study? Is the evaluation of the method objective?

AC: We provided additional information by clarifying parameters, such as the impact of STA/LTA settings, learning rate, training iterations. We updated the structure of the manuscript, especially in the introduction and concept sections, to make it simpler to follow and access the presented information. To take especially the concerns of reviewer one into account we defined precisely what belongs to our method and what to the case study. In addition, we are more precise in the terminology we use and the geoscientific field we consider. In the evaluation we objectively demonstrate the benefits our method and detach evaluation of the case study from the evaluation of our method. We purposely leave out any geophysical interpretation of the results we obtain after applying our method to the case study.

EC: The comments of the reviewers largely speak for themselves; although they assess the paper from different angles, the common feature of their assessment is that there is a lack of clarity.

AC: We addressed this issue by defining precise terminology at the beginning of the revised manuscript which is used persistently throughout the manuscript. Moreover we avoid misleading formulations. Moreover, we condensed significant parts of the introduction and reorganized the structure of the paper.

EC: Reviewer #1 focuses more on the language and structure problems. In addition, he asks to evaluate your work in a broader context in the discussion/conclusion.

AC: In addition to restructuring the first part of the manuscript, including updates to figures, we changed the discussion to include a broader . Specifically we introduced two new subsections (”Feature Extraction” and ”Overfitting”) and extended the existing subsections. Additionally, we added more literature to the discussion. Moreover, we gave hints to possible future works in the evaluation, discussion and outlook.

EC: Similarly, reviewer #2 asks for a more detailed description of the method and a better acknowledgment of previous work. I largely agree with that.

AC: We added a more concise method description and introduced a new subsection ”Convolutional neural networks” to explain the concepts of convolutional neural networks. Moreover, in addition to new geoscientific literature we acknowledge additional literature from the machine learning for seismology field.

EC: In addition, I want to highlight a few specific stylistic points, which may help you to revise the paper and achieve the aims outlined above. I will do this quoting specific example (page.line). 2.11 ‘ ... very unattractive overall solution ... ’: please avoid subjective judgements such at this. It is better to list the pros and cons, and then explain your priorities and your reasoning. 5.3 ‘Care is taken ... ’; 6.12 ‘But the meticulous care does pay off.’; 7.8 ‘Care has been taken to prevent significant data gaps ... ’: Such statements are not helpful to the reader. The phrasing is meant to convey some particularly high standards of scientific rigour. However, it is unclear what you have actually done, and thus your ’care’ is not reproducible. It would be better if you describe your actual measures (e.g., for preventing significant data gaps) and then describe how well they worked, and if they failed, why they failed. Again, here it is important to keep the reproducability of the work in mind.

AC: We have worked through the paper to remove any subjective judgment, imprecise wording or not well-documented phrases. For example we have updated the example you mention to ”Significant data gaps are prevented by using solar panels, durable batteries and field-tested sensors..."

EC: 8.6 'It becomes apparent in Fig. 5 (b)-(c) that anthropogenic noise, such as mountaineers walking by or helicopters, can have a strong influence on seismic recordings.': This sentence is an example of how results are mixed into the method description. There are multiple other instances. I ask you to separate this and present the methodology in the methods section and the results in the results section.

AC: We corrected your example and reformulated the sentence. In addition, we have restructured certain parts of the manuscript to avoid a mix-up of methods and results. For example we have placed all of our findings of the statistical evaluation into the results section.

EC: 18.2 'The results of the classifier experiments from Sect. 3.2 are listed in Table 3.': Such sentences contain little information. It would be better to state the main result or feature (that is important in the current context) and then cite the table in parentheses.
Looking forward to seeing your revised paper, best wishes, Jens Turowski

AC: We have reformulated the presented example and have worked through the paper to address similar issues. For example we updated the example you mention to "The results of the classifier experiments (Table 2) show that ..."

RC: Referee Comment
AC: Author Comment

1 Response to RC1

AC: We would like to thank the anonymous reviewer for the extensive review and the valuable feedback. We will incorporate the feedback and address all comments in the following. For a preliminary revised manuscript highlighting all the modifications please refer to the response to the editor.

RC: This paper addresses the issue of accurate attribution of seismic events to the correct source in long-term/large (micro-)seismic datasets. This paper has the potential to form a helpful methodological contribution to the geomorphic literature, and the overall result is promising. However, I do not believe the paper is ready for publication in its current format. Whilst there is some interesting information presented here, the focus, clarity and structure of the paper require further work.

AC: We restructured the manuscript to make it more precise in terms of terminology and the geoscientific field we consider. Moreover, we have condensed the information to make it more accessible.

RC: General points The language is often vague, with loose use of specific terminology. For example, in the abstract, the authors mention that 'Successful analysis depends strongly on the capability to cope with such external influences'. What do they mean by 'successful analysis' and 'coping' with these influences?

AC: We acknowledge the loose use of specific terminology in the initial submission. In the revised manuscript we use a more precise terminology and avoided misleading formulations, such as 'successful analysis' mentioned in your comment.

RC: Similarly, the authors mention 'correct slope characterisation' in the next sentence. What does this mean? It suggests consideration of the structural/strength/geometric properties and/or damage condition of the slopes. It is not clear which the authors are addressing, and why. Linked to this, Fig. 5 suggests that the focus

of the paper is on rockfalls, which again is different to 'slope characterisation'. In short, what is the geomorphic nature of the seismic activity the authors are considering?
RC: Links to geomorphic processes are implicit at best, and largely absent. For example, what exactly are you trying to monitor? Rockfall occurrence? Ground cracking and associated micro-seismic signal? This isn't clear. There is also a stark lack of reference to appropriate literature (e.g. page 2, lines 10 – 16).

AC: In contrast to the submitted manuscript, we have precisely defined the application context in the revised manuscript. The context is geophysical analysis on micro-seismic signals in general and event-based analysis in particular. The specific analysis performed using event-based methods depends on the use case and is not the focus of our study. For example the case study we use to demonstrate our method focuses on rock fracturing. We have also added additional literature to define the different geophysical applications and state more precisely which application we focus on: (Hardy, 2003), (Michlmayr et al., 2012), (Gischig et al., 2015), (Burjánek et al., 2012), (Weber et al., 2018).

RC: The final sentence in the abstract is also rather obvious and can be made without the detailed assessment presented in the paper. Indeed, this type of source characterization is commonly done (and done well) by geomorphologists (see e.g. the work of Adam Young on coastal microseismic monitoring). The most interesting part here is the ability to distinguish between sources of microseismic activity in large/long-term monitoring datasets, and this needs to be more clearly presented.

AC: You are correct that the final sentence "Due to these findings we argue that a systematic identification of external influences, like presented in this paper, is a prerequisite for a qualitative analysis." is rather obvious. Indeed, we want to show in this paper how the source characterization can be done systematically for a large and long-term monitoring experiment. Consequently, we have updated the statement to be more precise: "Due to these findings we argue that a systematic identification of external influences using a semi-automated approach and machine learning techniques as presented in this paper is a prerequisite for the qualitative and quantitative analysis of long-term monitoring experiments."

RC: The Introduction repeats the same points multiple times in subtly different ways – this section could be condensed considerably.
RC: Section 2 again repeats much of what we have already been told in the introduction.
RC: Sections 2.4.2 –3.4 contain some ostensibly important methodological steps, but again much of these sections feels descriptive, lacks an appropriate justifi-

cation and a logical structure to follow the workflow and the choices made.
RC: The aim of the paper is not clear and the authors present instead a bullet-point list of study conclusions. What is the focus here and what is novel?

AC: The introduction has been condensed significantly. Moreover, due to the reorganization of the first sections the reading flow has been improved. Whereas the initial manuscript had a potentially confusing structure, the updated manuscript follows a clear structure of

- Introduction
- Concept of the classification method
- Case Study
- Manual Data Assessment
- Classifier Selection and Training
- Automatic Classification
- Evaluation
- Discussion
- Conclusions

The aims of the paper are now highlighted at the end of the introduction with precise statements about the contributions and novelty.

RC: The methods section is again repetitive, justifying the need for, and broad benefits of, the approach, rather than stating concisely how it works. Much of the information here is not clear. For example, Page 6, Lines 16 -21 - there is no specific detail about how tasks are undertaken and how 'a good set of classifiers' is objectively specified. Much of the methods section lacks detail and feels very descriptive and subjective; many of the choices made are not fully/objectively demonstrated.

AC: We have condensed the methods section (now called "Manual Data Assessment") significantly and concisely described the steps required for manual data assessment. A detailed description to objectively demonstrate our choices is given in each subsection. Additionally, to avoid the "descriptive feel" we have added specific examples to the classifier selection description (see p. 6 l. 24 - 31). In this way it remains clear which steps are to be taken on a high, methodology level while having short specific details on e.g. 'how a good set of classifiers' is specified (which is later defined in detail in the respective subsection).

RC: For example, Fig. 6 does not clearly demonstrate the wind speed threshold required for a 'visible influence' on tremor amplitude. Important definitions do not appear in logical places (e.g. tremor amplitude is defined after it has been used in the text).

AC: For better understanding we have indicated the wind speed threshold in the respective figure. Moreover, we have rephrased the section about the tremor amplitude such that the explanation is directly available to the reader (see Page 9, Lines 1 - 5).

RC: The key aspect of the event trigger threshold by STA/LTA is not appropriately addressed; I would like to see more critique of the application of the method in this setting. Is it too sensitive and/or appropriate given the plots in Fig. 5?

AC: We discussed the STA/LTA characteristics throughout the submitted manuscript. In the revised manuscript we added an additional paragraph to discuss the STA/LTA settings in the context of Fig.5, including the effect of the threshold (see Page 9, Lines 12 - 17). We hope that thereby the application of the method in this setting are described in greater detail.

RC: How is the accuracy of event attribution assessed, other than by ruling out mountaineers etc. and process of elimination (page 9, lines 8 – 10 suggests this is the case)?

AC: We assume that the term event in this question relates to geophysical events for example the rockfalls discussed on page 9, lines 8 – 10 in the initial submission. In this case the accuracy of the attribution is related to the accuracy of our sources, which are incidents reported and logged by local observers, for example during maintenance of the monitoring setup. In addition, some rockfalls can also be seen by analyzing image sequences. We relate the characteristics of the micro-seismic signal to the timestamps of a rockfall report containing beginning and end timestamp. We do not use additional information/knowledge about a characteristic signal pattern which makes a rockfall identifiable only with the micro-seismic signal. Since we take only verified events into account the accuracy of these events is rather high but it also means that we probably missed to annotate rockfalls occurrences during the two years. As a result, we did not use our classifier to automatically annotate rockfall occurrences since the dataset is not accurate enough to train a rockfall classifier. Additionally, to evaluate the accuracy of event attribution even more we introduce a new evaluation which investigates how false labels affect the classifier performance. This evaluation is presented and discussed in the section "Classifier Evaluation".

RC: The level of assumed knowledge about neural network is also rather high.

AC: We have added a new subsection "Convolutional Neural Networks" (starting Page 9, Line 7) to explain the concept of convolutional neural networks and we recommend additional literature for the interested reader

RC: I am not convinced by the 'statistical analysis' presented in Table 1 – this seems rather weak and limited in terms of the depth of data analysis.

AC: We have restructured the statistical evaluation and improved the analysis. First, methods and evaluation are strictly separated into their individual section. Second, the "Statistical Evaluation" section now comprises all results from the initial submission and new results we have added as a consequence of your comment. The results presented and discussed in depth are (i) statistics for the manually annotated test set, (ii) statistics for the automatically annotated set for the year 2017, (iii) a plot which illustrates the distribution of STA/LTA events over time.

RC: The results section draws out the key argument that the authors wish to make, but I would like to see more assessment of the data presented in Fig. 5, even at the basic level, including the duration and frequency range/spectral density of different seismic sources. Can this information be used in a simpler manner to draw the same conclusions?

AC: In the new section "Feature Extraction" we have addressed the before mentioned comment. We make an assessment of the different source characteristics and how these can be used to classify/distinguish event sources. Moreover, we discuss the pro/cons of classifying based on manually extracted features versus classifying on learned features.

RC: How sensitive are the patterns shown by the graphs to the colour scale of the spectral density information?

AC: The color scale is only a visualization guideline and has been set to the same range for all subplots to maintain comparability between the subplots.

The visibility of the patterns would change with a different scale but please note that this visualization is only used for illustration in the paper. The input to the convolutional neural network is not using a color-coded representation but uses the raw spectrogram matrix.

RC: The discussion section is underdeveloped, lacks grounding and critique in the context of related literature and does not address the geomorphic significance of the approach addressed.

AC: We have addressed the fact that the discussion section is underdeveloped and expanded by adding two new subsections ("Feature Extraction" and "Overfitting") and extending the existing subsections. Moreover, we added more literature, such as (Walter et al., 2008), (Kuyuk et al., 2011), (Eibl et al., 2017), (Fei-Fei et al., 2006). Additionally, we have included more information in the evaluation section to support discussion about the possible advantages/disadvantages of our approach in regards to the geophysical application of event-based analysis. We show how our method can be used in our case study to extend event-based geophysical analysis. However, we find that a more detailed assessment and discussion about the results of our case study (in regards to geophysical significance) is out of scope of the submitted manuscript.

RC: How does the constrained uncertainty of the approach considered compare to other sources of uncertainty, such as seismometer tilt (indeed, which component of the seismometer is being used, and why? – again see the work of Adam Young)

AC: We have updated the manuscript to include which components of the seismometer are used (all three components, see Page 9, Lines 1-2). However, since we are not performing a device characterization study the analysis of tilt impact is not in the scope of our work.

RC: [How does the constrained uncertainty of the approach considered compare to other sources of uncertainty, such as] rock slope resonance and site effects (see e.g. Burjanek et al, 2012, 2017 GJI)? Some of the claims made about trade-off between time and accuracy feel poorly considered and require a more robust demonstration. There is also no discussion of the representativity of the case studies provided and how changes in the nature of the rock mass may affect the accuracy/source attribution of the seismic readings through time (e.g. resonance effects on duration and frequency as a rock mass degrades).

AC: We accounted for changes in the nature of the rock by using a test set from a different year. We assume that the test set is representative for upcoming, never-seen-before data (see Discussion section "Overfitting" of the revised manuscript). Since the classifier shows good performance on the test set, the changes in nature of the rock are not assumed to have a significant impact on the performance of our classifier. This assessment is of course only valid in the scope of our dataset and an interesting future work would be to investigate how specific resonance or side effects have an impact on the accuracy of the classifier.

RC: Specific comments (not exhaustive)

blackAC: The following reviewer comments have been acknowledged and corrected but do not need a dedicated answer in our opinion.

RC: There are many uses of e.g. in the manuscript – remove these and replace with 'such as' or 'for example' as appropriate.
RC: Brackets for citations are not always used correctly. Please check and amend.
RC: Page 6 Line 28 - has monitored? Tense is not correct.
RC: [Page 7] Line 32 – sentence beginning 'Additionally' is not clear.
RC: Page 10 Line 5 – do not use comma splices (re: therefore)

blackAC: The passages the following comments refer to have been removed in the revised manuscript.

RC: [Page 2] Line 3 – HAS been demonstrated [Page 2] Line 5 – micro-seismic RECORDS?
RC: [Page 2] Line 6 – biased assessments of what? [Page 2] Lines 3 – 4 and 5 – 9 – very repetitive.
RC: [Page 2] Line 22 – its (not it's)
RC: [Page 2] Lines 18 – 19 – Do you mean the accurate attribution of seismic events?

blackAC: The following specific comments are replied to individually.

RC: Fig. 5 - what are the red/purple circles? Are these triggered microseismic events? This isn't clear in the figure or the caption.

AC: The red circles indicate the timestamps of the STA/LTA triggers we use for the paper. We have updated the caption accordingly.

RC: Page 1 Line 19 – check terminology. Rockfalls are a type of landslides (see e.g. Varnes, 1978, and subsequent iterations of this work).

AC: We acknowledge a loose usage of terminology and a possible sources of misunderstanding given our formulation and thus we have rewritten the passage mentioned. However, we would like to highlight that in Varnes, 1978 and in the subsequent work it is recommended that the term slope movement is used instead of landslides to avoid confusion. Consequently, you are right that rockfalls are a type of slope movement.

RC: Page 2 Line 1 – what is the difference between acoustic emission and micro-seismic emission? Clarify.

AC: The difference is the frequency range in which the emission is detected. The particular sentence related to this question was rewritten in the revised manuscript. Now, acoustic emission is not mentioned anymore to avoid confusion since the focus of the study is only on micro-seismic emission.

RC: [Page 2] Line 9 – expand on scaling issues.
RC: [Page 7] Line 8 – expand on 'scaling issues' – this is unclear.

AC: By scaling issues we mean that it is for example unfeasible to manually analyze and annotate continuous micro-seismic recordings of many years. We have reformulate a similar statement to "... manual methods suffer from their inability to scale to increasing data volumes ..."

RC: [Page 2] Line 21 – what is the significance of footsteps?

AC: The paragraph has been changed in the revised manuscript. It is now made clear that the STA/LTA event detectors can be used to register external influences, such as footsteps, but "cannot reliably discriminate geophysical seismic activity from external (unwanted) influence factors"

RC: Page 3 Line 16 – F1 is not defined at this point. Indeed, much of the terminology in this section (e.g. ensemble classifier) is not clearly defined.

AC: We have addressed this point by defining F1 before it is used in the text (except for the abstract where the value is required to make a statement about the performance of our method). Additionally, it is made clear than an ensemble classifier consists of multiple classifiers.

RC: Page 7 Lines 6/7 – clarify 'sampling rate' – how was the sampling done? Or are you referring to the data transmission interval?

AC: The sampling is done by performing a measurement every two minutes with the respective sensor and then transmitting that measurement via the wireless sensor network to the server. We have adjusted the phrasing to make this aspect clear. To answer your question: In our case the sampling rate equals the inverse data transmission interval.

RC: Lines 26 – 29 – the distinction between acoustic events and seismic events is confusing, seems a little arbitrary and lacks reference to the literature; some of the terms do not follow some conventions in e.g. laboratory monitoring of acoustic emissions; this is important for a contribution to the geophysical literature. These definitions and distinctions also come too late in the manuscript, since these terms are used earlier.

AC: We acknowledge that given the geophysical background a consistent usage of terms is required. We have updated our definitions such that an event is defined as a trigger from a STA/LTA event detector. An event can have sources of geophysical or non-geophysical nature. In our study we apply a systematic method to identify non-geophysical sources in order to take them into account when analyzing geophysical sources. The definitions are now defined at the beginning of the paper and are consistently used throughout the manuscript

RC: Table 2 – this needs a lot more detail – what is this showing?

AC: The table caption has been updated to explain the table content, the structure of the neural network, its layers, strides and output channels.

RC: [Page 6] Line 19 – triaxial or three-axis?

AC: We use the word three-component in the revised manuscript

RC: [Page 7] Line 11 – The microseismic records considered in the case study were affected?

AC: It has been updated to "The recordings of the case study were affected..."

RC: [Page 7] Line 24 – is 'sounds' the correct word here?

AC: The passage the comment refers to has been removed in the revised manuscript.

RC: Table 1 – Reword the caption. It is not the case that none of the other categories 'apply'. Rather, it is where you have not been able to classify the signal as one of the three categories discussed.

AC: It has been updated to "when none of the other categories could have been identified".

RC: Page 9 Line 9 – Figure 5(e) does not show an example of a rockfall. It shows an example of the seismic signature of a rockfall event.

AC: It has been updated to "Shown are seismic signatures of ..."

RC: Referee Comment
AC: Author Comment

1 Response to RC2

AC: We thank the anonymous reviewer for the extensive review and the hint to additional literature. In the following we will address the comments and hint to the improvements in our manuscript. For a preliminary revised manuscript highlighting all the modifications please refer to the response to the editor.

RC: This paper proposed a new method for identifying external influences such as winds or mountaineers in micro-seismic recordings. Because the external influences may cause bias interpretations, its identification is very important for understanding micro-seismic recordings. In addition, the method may help to interpret the external influences which are keys to improve our understanding of rock-slope failure processes. The similar idea using machine learning is already applied to seismic wave discrimination such as Li et al., 2018. This study is interesting and suitable for the publication after moderate revisions. I suggest the authors revise this manuscript and pay attention to the following list as general suggestions:
RC: 1. Acknowledge previous studies on this topic or related topics and make sure the readers understand your contribution;

AC: We have acknowledged them. Moreover, we have rewritten the introduction and now the contributions are clearly demonstrated.

RC: 2. Introduce more about methods especially for Convolution Neural Network since readers may not be familiar with this method at all;

AC: We have added a new subsection "Convolutional Neural Networks" (starting Page 9, Line 7) which introduces convolutional neural networks and we have referenced additional literature to provide the reader with more background information on neural networks.

RC: 3. Discuss more future works such as how to automatically learn signal pattern in the external influences to improve the classification and interpreta-

tion of rock-slope failure processes;
RC: Page 20 line 14-22: The method is trained based on negative examples. But in most conditions, we should pay more attention to the phenomena of interest. In the discussion part, the author should discuss how this method can improve our interpretation of the processes of interest.

AC: We have added another paragraph to the discussion subsection "Classification of Negative Examples" (starting Page 23, Line 20) in which we explain how the method can improve the interpretation of the phenomena of interest. Moreover, we have extended our outlook (Page 26, Line 9) to include other methods which can improve the classification of external influences for example using semi-supervised and one-shot classification.

RC: Here are more specific comments:

RC: Page 9, line 1-4: I am confused about this part. Does this part mean that the dataset may be mislabeled due to fog, lens flares or other reasons? Have this data been included in the training dataset? Similar problem for the rockfalls in line 8-10?

AC: In case of limited visibility the images are not mislabeled, since the label represents what the annotator sees. However, since the seismic data is labeled with the help of images a certain probability of mislabeled samples exist if only images are used for annotation. In our case we reduce this probability by using an experienced annotator who can identify mountaineers on spectrograms and by using image sequences for annotation (before/after) when applicable. In the case of rockfalls we can only annotate time periods where we have additional information. Therefore it is most likely that we were unable to annotate all rockfall occurrences. As a consequence we did not consider a rockfall classifier. We have added this information on (see Page 9, Lines 6 - 9; Page 10, Lines 9 - 10).

RC: Page 11, line 12-15, the dataset including training dataset and test dataset seems to be small and may have serious overfitting problem. The authors need to address this issue during the discussion part and prove the trained model can handle it well.

AC: In the initial manuscript we have addressed the problem of overfitting in several paragraphs. In the revised manuscript we added another subsection "Overfitting" (starting Page 25, Line 31) to the discussion section explaining the impact of overfitting in our study.

RC: Page 14, line 12-14 The results with ten iterations are presented in this paper, but it will be better to show how the results change for a different number of iterations (such as 1, 5, 10, 20 iterations).

AC: We extended the evaluation section and have evaluated the impact of different training/test iterations in the "Classifier Evaluation" section (Page 20, Lines 9 - 19) and in Figure 11. We can confirm our choice of ten iterations.

RC: Page 15, Line 10: The learning rate is very small, which may make the code very slow. Is there any specific reason to set this small value?

AC: The value is the outcome of a preliminary hyper-parameter search and has been fixed for classifier training. Since the number of required iterations until convergence is rather small in comparison to other datasets/networks we found it reasonable to use such a small learning rate without negative impact regarding the total training duration. We have added the information explaining the learning rate to the revised manuscript (Page 16, Line 23 - 24).

RC: Page 16, line 10-16: Since it needs to manually relabel for the dataset in some cases, it will be worth to discuss how the potential human errors during data labeling will influence the classifier performance.

AC: We introduce a new evaluation which investigates how false labels affect the classifier performance. This evaluation is presented and discussed in the section "Classifier Evaluation" (Page 20, Lines 20 - 22) and in Table 3.

1 List of Changes

- We addressed all comments by the reviewers
- We provided additional information by clarifying parameters, such as the impact of STA/LTA settings, learning rate, training iterations.
- We updated the structure of the manuscript, especially in the introduction and concept sections, to make it simpler to follow and access the presented information.
- We condensed significant parts of the introduction and reorganized the structure of the paper.
- To take especially the concerns of reviewer one into account we defined precisely what belongs to our method and what to the case study. In addition, we are more precise in the terminology we use and the geoscientific field we consider.
- In the evaluation we objectively demonstrate the benefits our method and detach evaluation of the case study from the evaluation of our method. We purposely leave out any geophysical interpretation of the results we obtain after applying our method to the case study.
- We define precise terminology at the beginning of the revised manuscript which is used persistently throughout the manuscript.
- We introduced two new subsections (”Feature Extraction” and ”Overfitting”) and extended the existing subsections.
- We added a more concise method description and introduced a new subsection ”Convolutional neural networks” to explain the concepts of convolutional neural networks.
- We gave hints to possible future works in the evaluation, discussion and outlook
- We reference more geoscientific and machine learning literature

Systematic Identification of External Influences in Multi-Year Micro-Seismic Recordings Using Convolutional Neural Networks

Matthias Meyer[1], Samuel Weber[2,1], Jan Beutel[1], and Lothar Thiele[1]

[1]Computer Engineering and Networks Laboratory, ETH Zurich, Zurich, Switzerland
[2]Department of Geography, University of Zurich, Zurich, Switzerland

Correspondence: Matthias Meyer (matthias.meyer@tik.ee.ethz.ch)

Abstract. ~~Natural hazards, e.g. due to slope instabilities, are a significant risk for the population of mountainous regions. Monitoring of micro-seismic signals~~ Passive monitoring of ground motion can be used for geophysical process analysis and ~~risk assessment. However, these signals are subject to~~ natural hazard assessment. Detecting events in micro-seismic signals can provide responsive insights into active geophysical processes. However, in the raw signals micro-seismic events are superimposed by external influences, ~~e.g~~ for example anthropogenic or natural noise ~~. Successful analysis depends strongly on the capability to cope with such external influences. For correct slope characterization it is thus important to be able to identify, quantify and take these influences~~ sources that distort analysis results. In order to be able to perform event-based, geophysical analysis with such micro-seismic data records it is imperative that negative influence factors can be systematically and efficiently identified, quantified and taken into account.

~~In~~ Current identification methods (manual and automatic) are subject to variable quality, inconsistencies or human errors. Moreover, manual methods suffer from their inability to scale to increasing data volumes, an important property when dealing with very large data volumes as in the case of long-term ~~monitoring scenarios manual identification is infeasible due to large data quantities demanding accurate automated analysis methods.~~ monitoring.

In this work we present a systematic strategy to identify ~~multiple external influences, characterize their impact on micro-seismic analysis~~ a multitude of external influence sources, characterize and quantify their impact and develop methods for automated identification in micro-seismic signals. We apply the ~~developed strategy~~ strategy developed to a real-word, multi-sensor, multi-year micro-seismic monitoring experiment ~~on the Matterhorn Hörnliridge~~ performed at the Matterhorn Hörnligrat (CH). We ~~present a convolutional neural network~~ develop and present an approach based on convolutional neural networks for micro-seismic data to detect external influences originating in mountaineers, a major unwanted influence, with an error rate of less than 1 %, ~~which is~~ 3x lower than comparable algorithms. Moreover, we present an ensemble classifier for the same task obtaining an error rate of 0.79 % and an F1 score of 0.9383 by ~~using images~~ jointly using time-lapse image and micro-seismic data ~~. Applying the~~ on a annotated subset of the monitoring data. Applying these classifiers to the ~~experiment data~~ whole experimental dataset reveals that approximately 1/4 of events detected ~~with~~ by an event detector without such a pre-processing step are not due to seismic activity but due to anthropogenic ~~mountaineering~~ influences and that time periods with mountaineer activity have a 9x higher event rate. Due to these findings we argue that a systematic identification of external influences ~~, like~~

using a semi-automated approach and machine learning techniques as presented in this paper ~~,~~ is a prerequisite for ~~a qualitative analysis~~the qualitative and quantitative analysis of long-term monitoring experiments.

1 Introduction

~~Rock-slope failures and landslides pose a significant risk for humans and infrastructure (Glade et al., 2005). Identifying and monitoring potential slope instabilities is therefore of great importance. Micro-seismic analysis can be used to assess the stability and characteristics of slopes in various environments (Spillmann et al., 2007) and as a result is beneficial for two main applications: (i) process analysis and scientific understanding and (ii)natural hazard warning systems. With the advent of low-power wireless sensor networks, long-term monitoring with continuous in-situ monitoring of acoustic emission and~~ Passive monitoring of elastic waves, which are generated by the rapid release of energy within a material (Hardy, 2003), is a non-destructive analysis technique and a popular approach to investigate various processes in rock slopes (Amitrano et al., 2010; Occhiena e allowing for a wide range of possible applications (Michlmayr et al., 2012). Passive monitoring techniques may be broadly divided into three categories, characterized by the number of stations (single vs. array), the duration of recording (snapshot vs. monitoring) and the type of analysis (continuous vs. event-based). On the one hand, continuous methods such as the analysis of ambient seismic vibrations can provide information on internal structure of a rock slope (Burjánek et al., 2012; Gischig et al., 2015; Weber e . On the other hand, event-based methods such as the detection of micro-seismic ~~signals has become possible even in remote areas (Girard et al., 2012) and the feasibility of the two aforementioned application areas have been demonstrated. Especially if deployed in larger quantities, the data volumes originating from such high-rate sensors necessitate an automated analysis workflow since manual treatment of the sensor data is infeasible~~events (which are focus of this study) can give immediate insight into active processes, such as local irreversible (non-elastic) deformation occurring due to the mechanical loading of rocks (Grosse and Ohtsu, 2008). However, for the reliable detection of events irrespective of the detection method the signal source of concern has to be distinguishable from noise, for example background seismicity or other source types. This discrimination is a common and major problem for analyzing micro-seismic data.

In ~~practice, micro-seismic signals are influenced by anthropogenic or extrinsic ambient noise (Eibl et al., 2017; van Herwijnen and Schwe , leading to biased assessments~~general, event-based geoscientific investigations focus on events originating from geophysical sources such as mechanical damage, rupture or fracture in soil, rock and/or ice. These sources originate for example in thermal stresses, pressure variations or earthquakes (Amitrano et al., 2012). However, non-geophysical sources can trigger events as well: (i) anthropogenic influences such as helicopter or mountaineers (Eibl et al., 2017; van Herwijnen and Schweizer, 2011; Weber et al., and (ii) environmental influences / disturbances, such as wind or rain (Amitrano et al., 2010). One way to account for such external influences is to manually identify their sources in the recordings (van Herwijnen and Schweizer, 2011). This procedure, however, is not feasible for ~~continuous, autonomous monitoring due to scaling issues~~autonomous monitoring because manual identification does not scale well for increasing amounts of data. Another approach is to limit to field sites far away from possible sources of uncontrolled (man-made) interference or to focus and limit analysis to decisively chosen time periods known not to be influenced by ~~e.g.~~ for example anthropogenic noise (Occhiena et al., 2012)~~or limit to field sites far away from possible sources of uncontrolled (man-made) interference~~. In practice both the temporal limitation as well as the spatial limitation pose severe restrictions~~resulting in a very unattractive overall solution. Research~~ . First, research applications can benefit from close proximity to man-made infrastructure since set up and maintenance of monitoring infrastructure is facilitated~~. Applications~~ (Werner-Allen et al., 2006). Second, applications in natural hazard early warning must not be restricted to special time-periods only. Moreover, they are specifically required to be usable close to inhabited areas with an increasing likelihood for human interference on the signals recorded. As a conclusion it is a requirement that external influences can be taken into account with an automated workflow, including pre-processing, cleaning and analysis of micro-seismic data.

~~A method~~

A frequently used example of an event detection mechanism is an event detector based on the ratio of short-term average to long-term average (STA/LTA). Due to its simplicity, this event detector is commonly used to assess seismic activity ~~is to calculate~~ by calculating the number of ~~seismic~~ triggering events per time interval for ~~the~~ a time period of interest ~~(Amitrano et al., 2005; Sen~~ ~~. The correct detection of seismic events is thus of importance for a good analysis. Due to its simplicity, a popular filtering technique for event detection is to use short-term/long-term average triggering (STA/LTA) (Withers et al., 1998). This~~ (Withers et al., 1998; . It is often used in the analysis of unstable slopes (Colombero et al., 2018; Levy et al., 2011) ~~, is available in commercial~~ and is available integrated into many commercially available digitizers and data loggers (Geometrics, 2018)~~and can be~~ . With respect to unwanted signal components, STA/LTA has also been used to detect external influence factors such as footsteps (Anchal et al., 2018) ~~. Due to it's~~ but due to its inherent simplicity, ~~STA/LTA~~ it cannot reliably discriminate geophysical seismic activity from external (unwanted) influence factors such as noise from ~~human beings,~~ humans, natural sources like wind, rain or hail without manually supervising and intervening with the detection process on a case by case basis. As a result the blind application of STA/LTA will inevitably lead to the false estimation of relevant ~~intrinsic seismic events~~ geophysical processes if significant external influences, ~~e.g.~~ such as wind, are present (Allen, 1978).

There exist several algorithmic approaches to mitigate the problem of external influences by increasing the ~~sensitivity of seismic~~ selectivity of event detection. Auto-correlation and cross-correlation methods (Brown et al., 2008; Gibbons and Ringdal, 2006) use ~~seismic~~ event examples to find similar events, failing if events differ significantly in "shape" or if the transmission medium is very inhomogeneous (Weber et al., 2018b). The most recent advanced methods are based on machine learning techniques ~~(Reynen and Audet, 2017)~~(Reynen and Audet, 2017; Olivier et al., 2018). The use of neural networks ~~(Kislov and Gravirov, 2017; Perol et al., 2018)~~ (Kislov and Gravirov, 2017; Perol et al., 2018; Li et al., 2018; Ross et al., 2018) shows promising results with the drawback that large datasets containing ground truth (verified ~~seismic~~ events) are required to train these ~~networks~~automated classifiers. In earthquake research ~~these large database of known seismic~~ large databases of known events exist, but ~~are difficult to obtain in slope stability~~ in scenarios like slope instability analysis where effects are ~~strictly local~~ on a local scale and specific to a given field site such data are inexistent. Here, inhomogeneities are ~~commonly found on a~~ present on a very small scale and ~~each field site differs in its~~ field sites differ in their specific characteristics with respect to signal attenuation and impulse response. ~~Thus, obtaining a database~~ In order to apply such automated learning methods to these scenarios obtaining a dataset of known events is required for each new field site ~~which requires expert knowledge and is generally known to be an arduoustime consuming task.~~

~~By using additional sensors like weather stations, cameras or microphones and external knowledge such as helicopter flight plans or mountain hut occupancy it is possible to semi-automatically label non-seismic events, e.g. helicopters, footsteps or wind without the need for expert knowledge. Such "external" information sources can be used to establish an annotated data subset that allows to train an algorithm that is able to identify unwanted external influence factors in the complete dataset. Following this pre-processing step, seismic activity can be assessed by using the simple STA/LTA method mentioned earlier or more complex approaches taking into account the external influences identified in the earlier step. The efficient labeling of multiple external influence sources in large seismic datasets enables a broad set of algorithms to be applied subsequently. The concept of this method is generic and can be applied to many signal source/influence pairs.~~

~~We focus on an application to analyze micro-seismic signals originating from steep, fractured bedrock permafrost that is used as a driving example to illustrate the method and quantify its benefit on a real world example. Our case study is based on a multi-sensor~~requiring substantial expert knowledge for a very arduous, ~~multi-year~~ time-consuming task. The aim of this study is to use a semi-automatic workflow to train a classifier which enables the automatic identification of unwanted external influences in real-world micro-seismic ~~monitoring experiment on the Matterhorn Hörnligrat (CH) which is affected by mountaineer and wind~~ data. By this means, the geophysical phenomena of interest can be analyzed without the distortions of external influences.

~~In this paper we~~ To address these problems, this paper contains the following contributions. We propose a strategy to identify and deal with unwanted external influences in multi-sensor, multi-year experiments. ~~propose a convolutional neural network for acoustic event detection using micro-seismic data which detects mountaineers with an error rate of 0.96 % and a F1 score of 0.9167. present an ensemble classifier which detects mountaineers on images and micro-seismic data with an error rate of 0.79 % and a F1 score of 0.9383.~~ We compare the suitability of multiple algorithms for mountaineer detection using a combination of micro-seismic ~~data and images and show that our approach shows a 3x lower error rate than other algorithms.~~ signals and time-lapse images. We propose a convolutional neural network (CNN) for source identification. We exemplify our strategy for the case of ~~micro-seismic event detection~~ source identification on real-world ~~data from measurements~~ micro-seismic data using monitoring data in steep, fractured bedrock permafrost~~and show that time periods with mountaineer activity have a approximately 9x higher event rate and that 25% of all detected events are of non-seismic nature due to mountaineer interference argue that due these findings extensive identification of external influences is a prerequisite for qualitative analysis~~. We further provide the real-world micro-seismic and image data as an annotated dataset containing data from a period of two years as well as an open source implementation of the algorithms presented.

~~In the following, the concept of identifying external influences will be presented in Sect. 2 using a case study. The strategy to select and train a set of classifiers will be described in Sect. 5. Finally, the results are presented in Sect. **??** and the advantages and disadvantages of the presented method are discussed in Sect. 8.~~

2 ~~Identification of External Influences~~

~~As noted earlier, measurements in a~~

2 Concept of the classification method

In this work we present a systematic and automated approach to identify unwanted external influences in long-term, real-world ~~scenario do not only contain signals~~ micro-seismic datasets and preparing this data for subsequent analysis using a domain-specific analysis method, as illustrated in Fig.1. Traditionally, the signal, consisting of the phenomena of interest ~~but are also affected by external influencesas depicted in Fig. 1. For quantitative analysis it is crucial to account for such external influences, possibly also differentiating between different types of external influences present in a given signal. It is important to point out that externalinfluences are not to be seen as something strictly negative that ought to be mitigated and removed for as much as possible. External influences can serve as context further describing the circumstances at a given point in time and space. Therefore a simple removal of all external influences is not desirable and we advocate that~~ and superimposed external influences, is analyzed directly as described earlier. However, this analysis might suffer from distortion through the external influences. By using additional sensors like weather stations, cameras or microphones and external knowledge such as helicopter flight plans or mountain hut occupancy it is possible to semi-automatically label events originating from non-geophysical sources, such as helicopters, footsteps or wind without the need of expert knowledge. Such "external" information sources can be used to train an algorithm that is then able to identify unwanted external influence. Using this approach multiple external influences are ~~classified in a~~ first classified and labeled in an automated pre-processing step ~~as presented in this paper. Using this extra information, tailored, domain specific analysis methods~~ with the help of state-of-the-art machine learning methods. Subsequent to this classification, the additional information can be used for domain-specific analysis for example to separate geophysical and unwanted events triggered by a simple event detector such as an STA/LTA event detector. Alternatively, more complex approaches can be used ~~on the signal benefiting from the additional information where applicable. An example for domain specific analysis~~ taking into account signal content, event-detections and classifier labels of the external influences. However the specifics of such advanced domain-specific analysis methods is beyond the scope of this paper and subject to future work. A basic example of a custom domain-specific analysis method is the estimation of ~~the~~ separate STA/LTA event ~~rate~~ rates for time periods when mountaineers are present and when they are not ~~, which will be used in~~ which we use as a case study in the evaluation section of this paper to exemplify our method. ~~The approach described is flexible, modular and extensible and does not constitute an information loss but rather an information gain constituting a significant advantage over competing approaches.~~

~~Illustration of the presented concept to enable domain-specific analysis on an existing measurement setup. Primary signal (light green) and secondary data (dark blue) are combined to form a dataset on which automatic classification can be performed. Secondary data is a combination of sensor and auxiliary data. The resulting annotated dataset can be used for domain-specific analysis. In a prior step, the correct settings and tools for automatic classification are determined in a manual preparation phase based on a subset of the dataset.~~

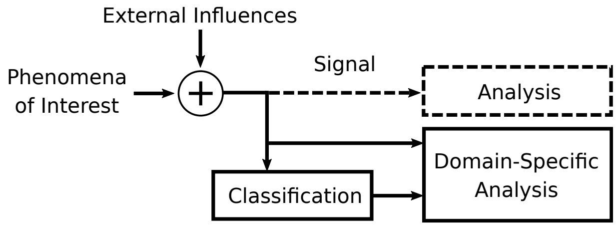

Figure 1. ~~Model of the sensing method. In a real~~ Real world measurement signals contain the phenomena of interest ~~is~~ superimposed ~~by~~ with external influences. ~~In a first approach the resulting signal can be~~ If directly analyzed ~~by appropriate methods, which could be constrained~~ the results are perturbed by the external influences. In ~~a second~~ contrast to this approach (dashed lines), ~~which we suggest~~ in this paper ~~, classification~~ we suggest a systematic and automated approach to first identify a multitude of external ~~influences can provide further information which~~ influence sources in micro-seismic signals using a classifier. The classifier result data can then be ~~utilized by domain-specific analysis~~ used to quantify unwanted signal components as well as drive more extensive and powerful event detection and characterization methods leveraging combinations of both the signals as well labelled and classified noise data (solid lines).

~~The manual preparation phase is subdivided into data evaluation and classifier selection and training. First, the data subset is characterized and annotated. This information can be used to do a statistical evaluation and select data types which are useful for classification. Domain experts are not required for the labor intensive task of annotation. The classifiers are selected, trained and optimized in a feedback loop until the best set of classifiers is found.~~

2.1 ~~Method~~

~~In the following we provide an overview how to identify and quantify external influence sources using a systematic combination of manual, semi-automatic and automated steps. Care is taken to reduce work requiring a human in the loop and to use automated methods as much as possible.~~ Figure 2 illustrates the ~~concept~~overall concept in detail. In a first step the available data sources ~~are assessed~~of a case study are assessed and cataloged. Given a ~~measurement setup~~ case study (Sect. 3) consisting of multiple sensors, one or more sensor signals are specified as primary signals (~~e.g.~~ for example the micro-seismic signals, highlighted with a light green arrow in the figure) targeted by a subsequent domain-specific analysis method. Additionally, secondary data ~~are chosen which can~~ (highlighted with blue arrows) are chosen to support the classification of external influences ~~. These~~ contained in the primary signal. Conceptually these secondary data can be ~~either other~~ of different nature, either different sensor signals, e.g. time-lapse images or weather ~~information,~~ data or auxiliary data ~~, e.g.~~ such as local observations or helicopter flight data. All data sources are combined into a dataset. However, ~~the~~ this resulting dataset is not yet annotated ~~which is~~ as required to perform domain-specific analysis ~~.~~ leveraging the identified and quantified external influences.

Two key challenges need to be addressed ~~before~~ in order to establish such an annotated dataset ~~can be established~~ by automatic classification: (i) suitable data types need to be selected for classification since not every data type can be used to continuously classify every external influence (~~e.g.~~ for example wind sensors are not designed to capture the sound of footsteps; flight data ~~cannot be obtained~~ may note be available for each time step) and (ii) ~~one or~~ a single (preferred) or at least a set of suitable, ~~good-performing~~ well-performing classifiers have to be found for each ~~external influence. These challenges can be addressed in a manual preparation phase, which includes dataset evaluation as well as classifier training, evaluation and selection.~~

2.1 ~~Manual preparation~~

~~A ground truth is often needed for state-of-the-art classifiers (e.g. neural networks), which requires manual annotation of data points. In a realistic setting the dataset comprises large data quantities which is impractical for manual annotation. To reduce the amount of manual labor only~~ type of external influence source. Once these challenges have been solved a subset of the dataset is ~~selected and used in a manual preparation phase , which consists of data evaluation and classifier training and selection as depicted in Fig. 4. Data evaluation can be subdivided into four parts: (i)characterization of external influences in the primary signal (i.e.the relation between primary and secondary signals), (ii) annotating the subset based on the primary and secondary signals, (iii) selecting the data types suitable for classification and (iv)performing a first statistical evaluation with the annotated dataset, which facilitates the selection of a classifier. Much of the preparatory work in creating this ground truth dataset is of manual nature and varies for different secondary data types depending on their specifics. But the meticulous care does pay off. By separately treating different categories of external influences and secondary data it is possible to evaluate in detail the impact of every factor in solitude or in combination, including detailed statistics. Moreover, the presented strategy reduces the work requiring an expert. Characterization and statistical evaluation are the only steps where domain expertise is required while it is not required for the time and labor intensive annotation process.~~

~~The classifier selection and training phase presumes the availability of a variety of classifiers for different input data types. The classifiers do not perform equally well on the given task with the given subset. Therefore classifiers have to be selected based on their suitability for classification given the task and the data. A selection of classifiers is therefore trained and tested with the annotated subset and optimized for best performance. The classifier selection, training and optimization is repeated until a sufficiently good set of classifiers has been found. These classifiers can then be used for application in the automatic classification process.~~

~~In the following, the previously explained methodwill be exemplified for wind and mountaineer detection using micro-seismic, wind and image data from a real-world experiment~~manually annotated in order to select and train the classifier(s) in a "preparatory" phase required to be performed only once, which includes manual data assessment (Sect. 4) as well as classifier selection and training (Sect. 5). The trained classifier is then used in an automated setup to annotate the whole dataset (Sect. 6). This "execution" phase can be performed in a one-shot fashion (post-processing all data in one effort) or executed regularly, for example on a daily or weekly basis if applied to continuously retrieved real-time monitoring data. These additional information can be used to perform a subsequent domain-specific analysis. This study concludes with an evaluation (Sect. 7) and discussion (Sect. 8) of the presented method.

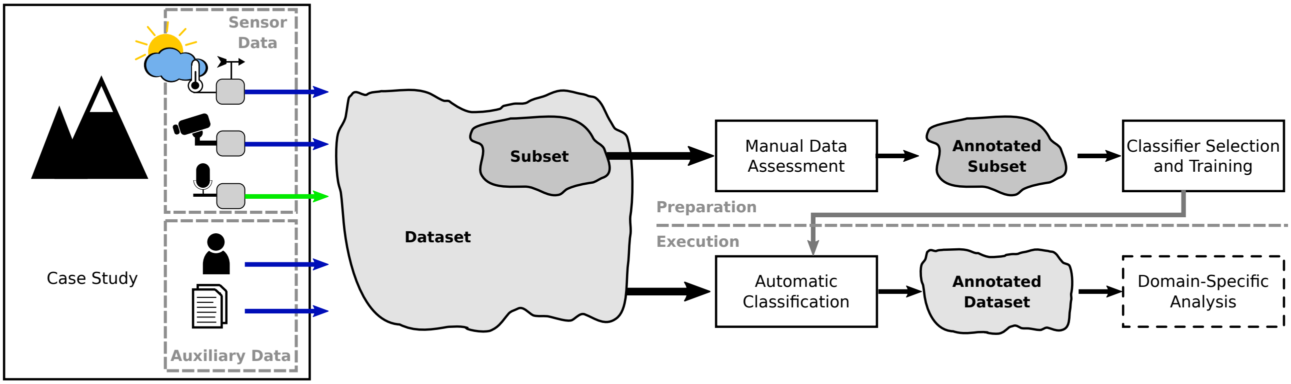

Figure 2. Conceptual illustration of the classification method to enable domain-specific analysis of a primary sensor signal (in our case micro-seismic signals denoted by the light green arrow) based on annotated datasets: A subset of the dataset containing both sensor and auxiliary data, is used to select and train a classifier that is subsequently applied to the whole dataset. By automatically and systematically annotating the whole dataset of the primary signal of concern, advanced methods can be applied that are able to leverage both multi sensor data as well as annotation information.

2.1 ~~Case Study~~

3 Case Study

The data used in this paper originate from a multi-sensor and multi-year experiment (Weber et al., 2018b) focusing on slope stability in high-alpine permafrost rock walls and understanding the underlying processes. Specifically, the sensor data is collected at the site of the 2003 rockfall event on the Matterhorn Hörnligrat, (Zermatt, Switzerland) at 3500 m a.s.l. where an ensemble of different sensors ~~monitors~~ has monitored the rockfall scar and surrounding environment over the past ten years. Relevant for this work are data from a ~~three-axial~~ three-component seismometer (Lennartz LE-3Dlite MkIII), images from a remote controlled high-resolution camera (Nikon D300, 24 mm fixed focus), rock surface temperature measurements, net radiation measurements and ambient weather conditions, specifically wind speed from a co-located local weather station (Vaisala WXT520).

~~In (Weber et al., 2018b) a seismometer~~ The seismometer used in the case study presented is used to assess the seismic activity by using an STA/LTA event detector~~. The~~ , which means for our application that the seismometer is chosen as the primary sensor and STA/LTA triggering is used as the reference method~~to assess seismic activity~~. Seismic data is recorded locally using a ~~nanometrics Centaur data logger~~ Nanometrics Centaur digitizer and transferred daily by means of directional WLAN~~connectivity~~. The data is processed on-demand ~~by~~ using STA/LTA triggering. The high-resolution camera's (Keller et al., 2009) field of view covers the immediate surroundings of the seismic sensor location as well as some backdrop areas further away on the mountain ridge. Figure 3 shows an overview of the field site including the location of the seismometer and an example image acquired with the camera. The standard image size is 1424x2144 pixels captured every 4 minutes. The Vaisala WXT520 weather data as well as the rock surface temperature are transmitted using a custom wireless sensor network infrastructure~~with a sampling rate once~~. A new measurement is performed on the sensors every 2 minutes and transmitted to the base station, resulting in a sampling rate of 30 samples per hour.

~~Care has been taken to prevent significant data gaps~~ Significant data gaps are prevented by using solar panels, durable batteries and field-tested sensors but given the circumstances on such a demanding high-alpine field site certain outages of single sensors, ~~e.g.~~ for example due to power failures or ~~also~~ during maintenance could not be prevented. Nevertheless this dataset constitutes an extensive and close-to-complete dataset.

The ~~case study was~~ recordings of the case study were affected by external influences, especially mountaineers and wind. This reduced the set of possible analysis tools. Auxiliary data which helps to characterize the external influences is collected in addition to the continuous data from the sensors. In the presented case the auxiliary data is non-continuous and consist of local observations, pre-processed STA/LTA triggers from (Weber et al., 2018b), accommodation occupancy of a nearby hut and a non-exhaustive list of helicopter flight data from a duration of approximately 7 weeks provided by a local helicopter company.

In following, we use this case study to exemplify ~~the~~ our method presented in the previous sections.

3.1 ~~Data Evaluation~~

~~As explained in Sect. ??, a prerequisite for automatic classification is the analysis of the given datasetto specify the requirements of the classifier. This sections explains the required steps of subset creation, characterization, annotation , statistical evaluation and the selection of the data type for classification. The subset is created by collecting continuous data from the sensors and additional data, which helps to characterize the external influences. In~~

4 Manual Data Assessment

A ground truth is often needed for state-of-the-art classifiers (such as artificial neural networks). To establish this ground truth while reducing the amount of manual labor only a subset of the dataset is selected and used in a manual data assessment phase, which consists of data evaluation, classifier training and classifier selection as depicted in Fig. 4. Data evaluation can be subdivided into four parts: (i) characterization of external influences in the primary signal (that is the relation between primary and secondary signals), (ii) annotating the subset based on the primary and secondary signals, (iii) selecting the data types suitable for classification and (iv) performing a first statistical evaluation with the annotated dataset, which facilitates the selection of a classifier. Characterization and statistical evaluation are the only steps where domain expertise is required while it is not required for the time and labor intensive annotation process.

The classifier selection and training phase presumes the availability of a variety of classifiers for different input data types, for example the broad range of available image classifiers (Russakovsky et al., 2015). The classifiers do not perform equally well on the given task with the given subset. Therefore classifiers have to be selected based on their suitability for classification given the task and the data. A selection of classifiers is therefore trained and tested with the annotated subset and optimized for best performance which can for example be done by selecting the classifier with the lowest error rate on a defined test set. The classifier selection, training and optimization is repeated until a sufficiently good set of classifiers has been found. This

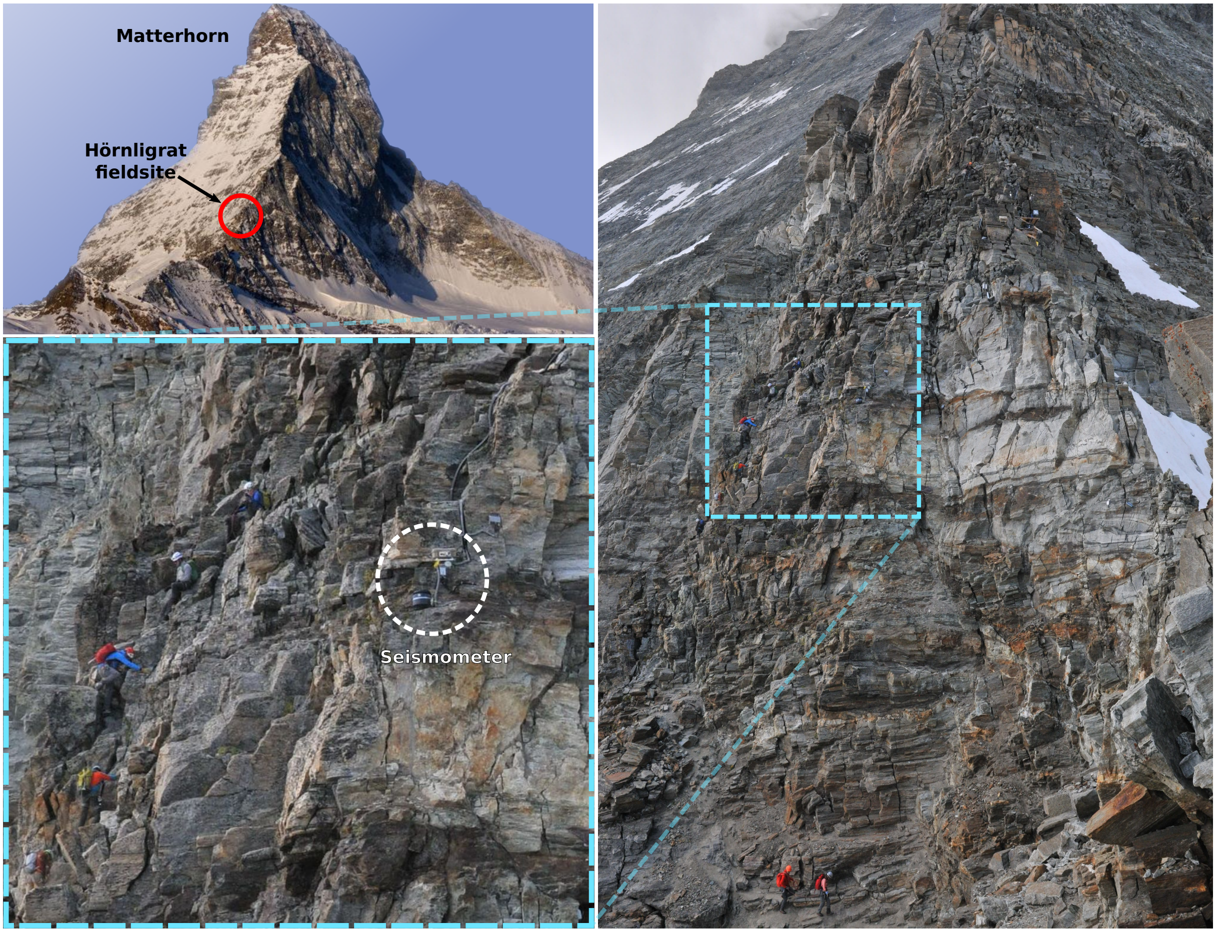

Figure 3. The field site is located on the Matterhorn Hörnligrat at 3500 m a.s.l. which is denoted with a red circle. The photograph on the right is taken by a remotely controlled DSLR camera on the field site at 2016-08-04T12:00:11. The seismometer of interest (white circle) is located on a rock instability which is close to a frequently used climbing route.

suitability is defined by the ~~presented case the additional data is non-continuous and consist of local observations, pre-processed STA/LTA triggers from (Weber et al., 2018b), accommodation occupancy of a nearby hut and a non-exhaustive list of helicopter flight~~ user and can for example mean that the classifier is better than a critical error rate. These classifiers can then be used for application in the automatic classification process.

In the following, the previously explained method will be exemplified for wind and mountaineer detection using micro-seismic, wind and image data from a ~~duration of approximately 7 weeks provided by a local helicopter company~~real-world experiment. The required steps of subset creation, characterization, annotation, statistical evaluation and the selection of the data type for classification are explained. Before an annotated subset can be created the collected data must be characterized for its useful-

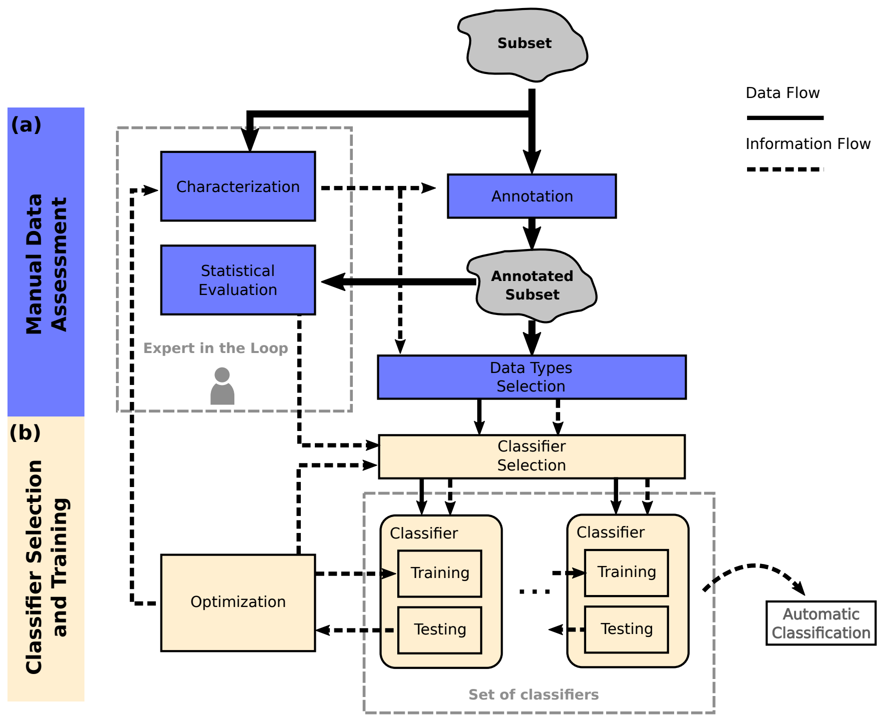

Figure 4. The manual preparation phase is subdivided into data evaluation (a) and classifier selection and training (b). First, the data subset is characterized and annotated. This information can be used to do a statistical evaluation and select data types which are useful for classification. Domain experts are not required for the labor intensive task of annotation. The classifiers are selected, trained and optimized in a feedback loop until the best set of classifiers is found.

ness in the annotations process, i.e. which data type can be used to annotate which external influence. ~~In the following the steps for creation of an annotated subset, as illustrated in Fig. 4 (a) are explained.~~

4.0.1 ~~Characterization~~

4.1 Characterization

The seismometer ~~data consists of sounds~~ captures elastic waves originating from different sources. ~~This section will discuss commonly captured sources in a real world measurement setup. The definition of an event can be ambiguous. Geoscientific studies often aim at identifying events related to rupture or fracture in rock and/or ice originating for example from thermal stresses, pressure variations or earthquakes (Amitrano et al., 2012). We will refer to them as events or seismic events. In audio classification literature an event is a specific sound which identifies a certain source, e.g. footsteps identifying a human. We will refer to them as acoustic events. While a seismic event is mostly regarded as a short-duration impulse, acoustic events can be a combination of different sounds of varying duration and spectral properties.~~

In this study we will consider multiple ~~acoustic events~~non-geophysical sources, which are ~~the influences of~~ mountaineers, helicopters, wind and rockfalls. ~~Additionally, we will analyze the time frames where none of~~ Time periods where the before mentioned ~~acoustic events happen~~sources can not be identified are considered as relevant and thus we will include them in our analysis as a fifth source, the "unknown" source. The mountaineer impact will be characterized on a long-term scale by correlating with hut accommodation occupancy (see Fig. 10) and on a short-scale by person identification on images. Helicopter examples are identified by using flight data and local observations. By analyzing spectrograms one can get an intuition what mountaineers or helicopters "look" like, which facilitates the manual annotation process. In Fig. 5 different recordings from the field site are illustrated, which have been picked by using the additional information described at the beginning of this section. For six different examples the time domain signal, ~~it's~~ its corresponding spectrogram and STA/LTA triggers are depicted. The settings for the detector are the same for all the subsequent plots. It becomes apparent in Fig. 5 (b)-(c) that anthropogenic noise, such as mountaineers walking by or helicopters, ~~can have a strong influence on seismic recordings~~are recorded by seismometers. Moreover, it becomes apparent that it might be feasible to assess ~~acoustic events~~ non-geophysical sources on a larger time frame. Mountaineers for example show characteristic patterns of increasing or decreasing loudness and helicopters have distinct spectral patterns, which could be beneficial to classify these ~~events~~sources. Additionally, the images captured on site show when a mountaineer is present (see Fig. 3), but due to fog, lens flares or snow on the lens the visibility can be limited. The limited visibility needs to be taken into account for when seismic data is to be annotated with the help of images. Another limiting factor is the temporal resolution of one image every 4 minutes. Mountaineers could move through the visible are in between two images.

The wind sensor can directly be used to identify the impact on the seismic sensor. ~~By manually analyzing~~ Figure 6 illustrates the correlation between tremor amplitude and wind speed~~(e.g. in Fig. 6) it can be deduced that wind speeds above approximately 30 km/h have a visible influence on the tremor amplitude~~. Tremor amplitude is the frequency-selective, median, absolute ground velocity and has been calculated for the frequency range of 1-60 Hz according to (Bartholomaus et al., 2015). By manually analyzing the correlation between tremor amplitude and wind speed it can be deduced that wind speeds above approximately 30 km/h have a visible influence on the tremor amplitude.

Rockfalls can best be identified by local observations since the camera captures only a small fraction of the receptive range of the seismometers. Figure 5 (e) ~~shows an example~~ illustrates the seismic signature of a rockfall. The number of rockfall observations and rockfalls caught on camera are however very limited. Therefore it is most likely that we were unable to annotate all rockfall occurrences. As a consequence we will not consider a rockfall classifier in this study.

~~The subplot~~ It can be seen in Fig. 5 (a) ~~shows a time frame from~~ that during a period which is not strongly influenced by external influences ~~. Repetitive seismic events are visible and trigger the event detector~~the spectrogram shows mainly energy in the lower frequencies with occasional broadband impulses.

The red circles in the subplots in Fig. 5 indicate the timestamps of the STA/LTA events for a specific geophysical analysis (Weber et al., 2018b). By varying the threshold of the STA/LTA event trigger the number of events triggered by mountaineers can be reduced. However, since the STA/LTA event detector cannot discriminate between events from geophysical sources and events from mountaineers, changing the threshold would also influence the detection of events from geophysical sources.

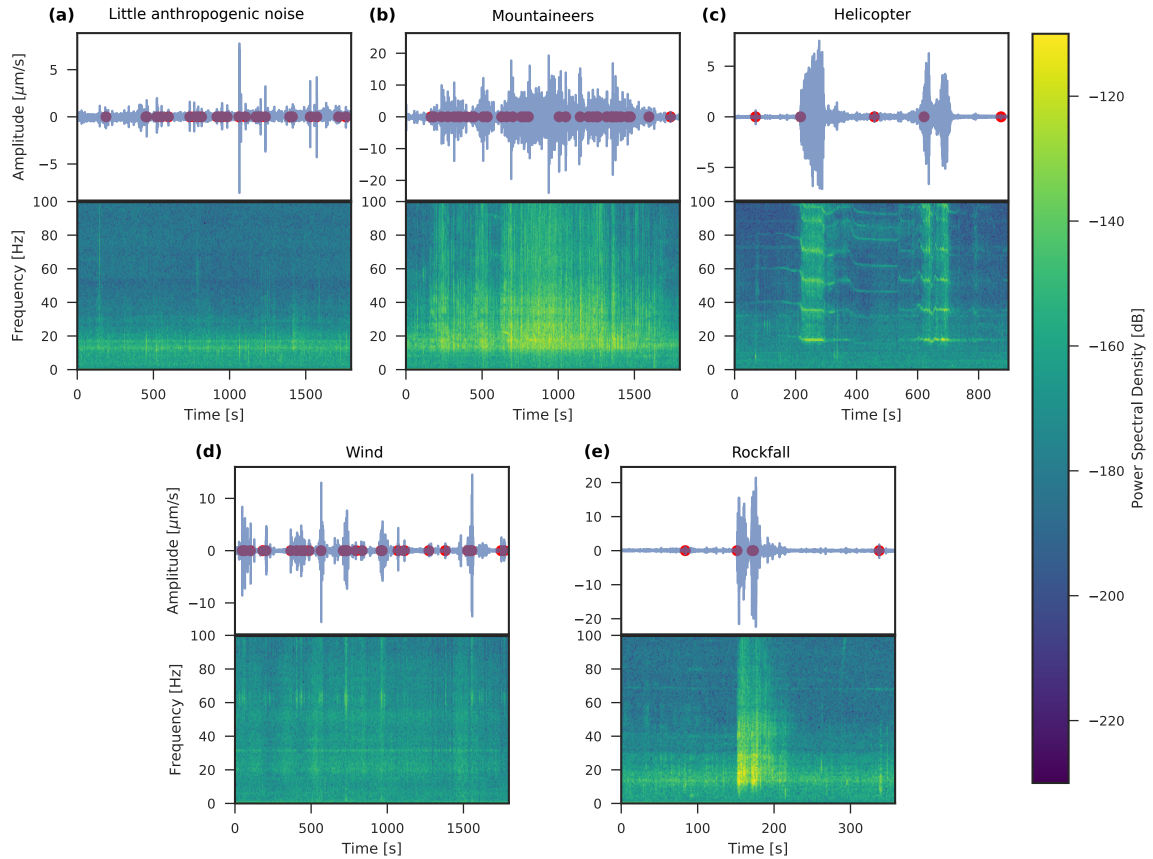

Figure 5. Micro-seismic signals and the impact of external influences: (a) During a period of little anthropogenic noise the seismic activity is dominant. (b) In the spectrogram the influence of mountaineers become apparent. Shown are seismic signatures of (c) a helicopter in close spatial proximity to the seismometer (d) wind ~~influence~~ influences on the signal (e) a rockfall in close proximity to the seismometer. The red dots in the signal plots indicate the timestamps of the STA/LTA triggers from (Weber et al., 2018b).

This fact would affect the quality of the analysis since the STA/LTA settings are determined by the geophysical application (Colombero et al., 2018; Weber et al., 2018b).

4.1.1 ~~Annotation~~

4.2 Annotation

~~For annotation and evaluation the~~ The continuous micro-seismic signals are segmented for annotation and evaluation. Figure 7 provides an overview of the three segmentation types, event-linked segments, image-linked segments and consecutive segments. Image-linked segments are extracted due the fact that a meaningful relation between seismic information and photos is only given in close temporal proximity~~, therefore~~. Therefore images and micro-seismic data are linked in the following way:

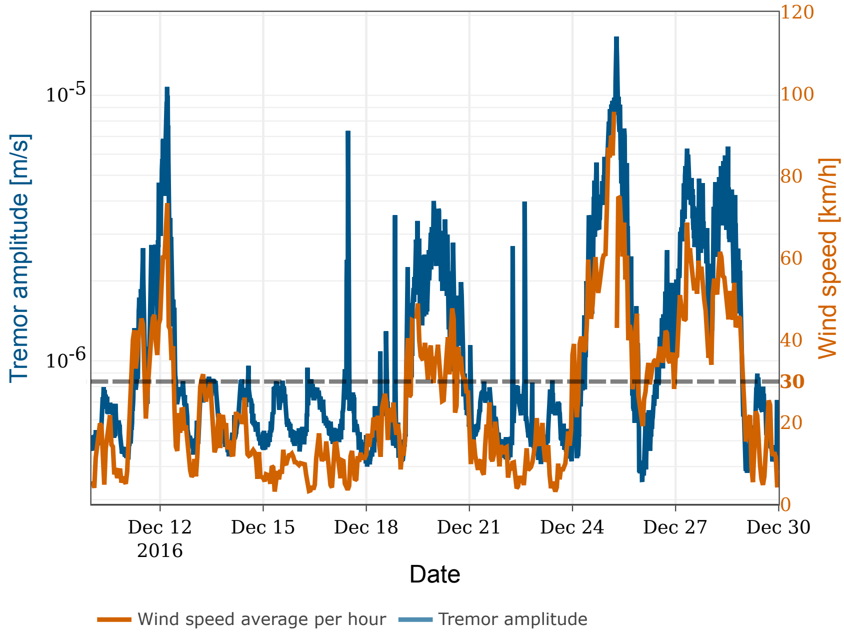

Figure 6. Impact of wind (light orange) on the seismic signal. The tremor amplitude (dark blue) is calculated according to (Bartholomaus et al., 2015). The effect of wind wind speed on tremor amplitude becomes apparent for wind speeds above approximately 30 km/h. Note the different scales on the y-axes.

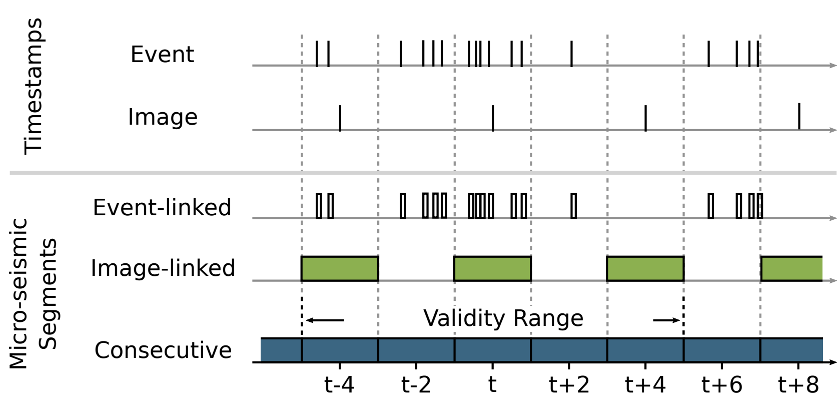

Figure 7. Illustration of micro-seismic segmentation. Event-linked segments are 10 second segments starting on event timestamps. Image-linked segments are two minute segments centered around an image timestamp. Consecutive segments are 2 minute segments sequentially extracted from the continuous micro-seismic signal.

For each image a 2 minute micro-seismic segment is extracted from the continuous micro-seismic signal. The micro-seismic segment's start time is set to 1 minute before the image timestamp. Event-linked segments are extracted based on the STA/LTA triggers from (Weber et al., 2018b). For each trigger 10 seconds following the timestamp of the trigger are extracted from the micro-seismic signal. Consecutive segments are 2 minute segments sequentially extracted from the continuous micro-seismic signal.

Only the image-linked segments are used during annotation, their label can however be transferred to other segmentation types by assigning the image-linked label to overlapping event-linked or consecutive segments. Image-linked and event-linked segments are used during data evaluation and classifier training. Consecutive segments are used for automatic classification on the complete dataset. Here, falsely classified segments are reduced by assigning each segment a validity range. A segment classified as mountaineer is only considered correct if the distance to the next (or previous) mountaineer is less than 5 minutes. This is based on an estimation of how long the mountaineers are typically in the audible range of the seismometer.

For mountaineer classification the required label is *mountaineer* but additional labels will be annotated which could be beneficial for classifier training and statistical analysis. These labels are *helicopters*, *rockfalls*, *wind*, *low visibility* (if the lens is partially obscured), and *lens flares*. The *wind* label applies to segments where the wind speed is higher than 30 km/h, which is the lower bound for noticeable wind impact as resulted from Sect. 4.1.

Figure 8 depicts the availability of image-linked segments per week during the relevant time frame. A fraction of the data is manually labeled by the authors, which is illustrated in Fig. 8. Two sets are created, a training set containing 5579 samples from the year 2016, and a test set containing 1260 data samples from 2017. The test set has been sampled randomly to avoid any human prejudgment. For each day in 2017 four samples have been chosen randomly, which are then labeled and added to the test set. The training set has been specifically sampled to include enough training data for each category. This means for example that more mountaineers samples come from the summer period where the climbing route is most frequently used. The number of verified rockfalls and helicopters is non-representative and although helicopters can be manually identified from spectrograms the significance of these annotations is not given due the limited ground truth from the secondary source. Therefore, for the rest of this study we will focus on mountaineers for qualitative evaluation. For statistical evaluation we will however use the manually annotated helicopter and rockfall samples to exclude them from the analysis. The labels for all categories slightly differ for micro-seismic data and images since the type of ~~events~~ sources which can be registered by each sensor differ. This means for instance that not every classifier uses all labels for training (~~e.g.~~ for example a micro-seismic classifier cannot detect a lens flare). It also means that for the same time instance one label might apply to the image but not to the image-linked micro-seismic segment (~~e.g.~~ for example mountaineers are audible but the image is partially obscured and the mountaineer is not visible). This becomes relevant in Sect. 5.3.4 when multiple classifiers are used for ensemble classification.

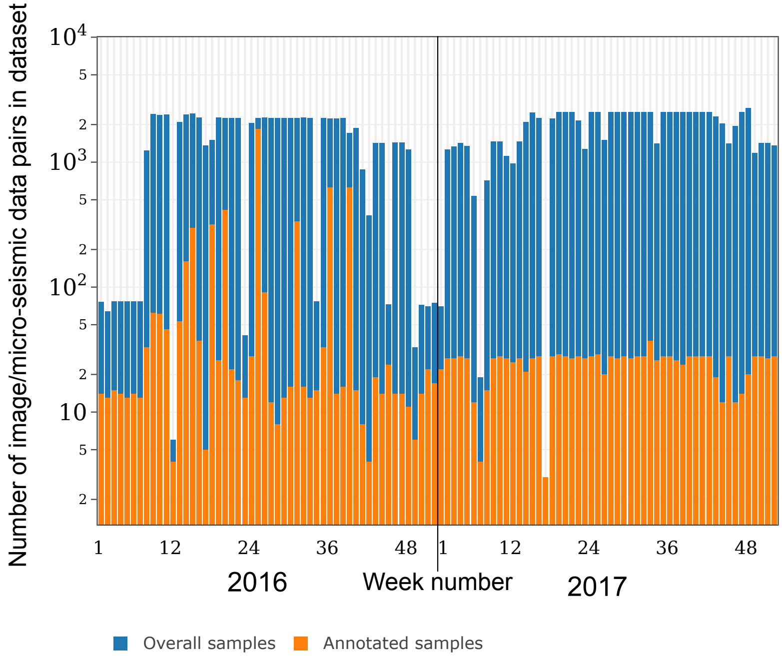

Figure 8. Number of image/micro-seismic data pairs in the dataset (dark blue) and in the annotated subset (light orange) displayed over the week number of the year 2016 and 2017. Note the logarithmic scale on the y-axis

4.2.1 ~~Data Types Selection~~

4.3 Data Types Selection

After the influences have been characterized the data type need to be selected which best describe each influence. The wind sensor delivers a continuous data stream and a direct measure of the external influence. In contrast, mountaineers, helicopters and rockfalls cannot directly be identified. A data type including information about these external influences needs to be selected. Local observations, accommodation occupancy and flight data can be discarded for the use as classifier input since the data cannot be continuously collected. According to Sect. 4.1 it seems possible to identify mountaineers, helicopters and rockfalls from micro-seismic data. Moreover, mountaineers can also be identified from images. As a consequence, the data types selected to perform classification are micro-seismic data, images and wind data.

4.3.1 ~~Statistical Evaluation~~

~~The annotated test set from Sect. 4.2 is used for a statistical evaluation involving the impact of annotated external influences on~~ The micro-seismic ~~analysis. The test set (2017) is chosen since wind data is not available for the whole training set due to a malfunction of the weather station in 2016. The experiment from (Weber et al., 2018b) provides STA/LTA event triggers for~~

~~None Mountaineer Wind Helicopter duration (hours) 28.87 1.9 6.6 3.73 mean number of events per hour 10.6 95.26 11.21 13.12 Statistics of the annotated test set (2017) per annotation category. "None" represents the category when none of the other categories apply. Given are the total duration of annotated segments per category and the mean number of STA/LTA events per category.~~

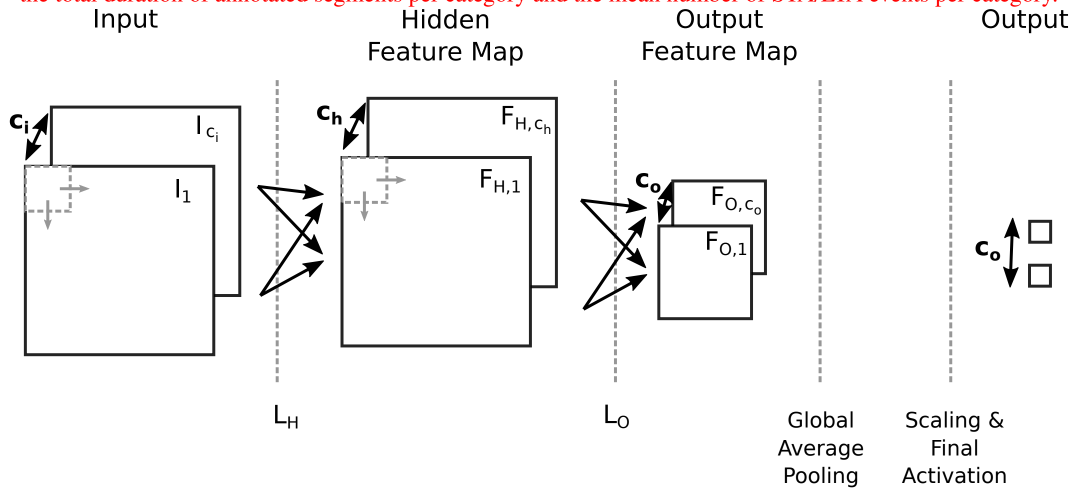

Figure 9. Simplified illustration of a convolutional neural network. An input signal, for example an image or spectrogram, with a given number of channels c_i is processed by a convolutional layer L_H. The output of the layer is a feature map with c_h channels. Layer L_O takes the hidden feature map as input an performs a strided convolution which results in the output feature map with reduced dimensions and number of channels c_o. Global average pooling is performed per channel and additional scaling and a final activation are applied.

~~2016 and 2017. Table **??** shows statistics for several categories, which are 3 external influences and one category where none of the 3 external influences are annotated (declared as "None"). For each category, the total duration of all annotated segments is given and how many events per hour are triggered. It becomes apparent that mountaineers have the biggest impact on the event analysis. 95.26 events per hour are detected on average during time periods with mountaineer activity, while during all other time periods the average ranges from 10.6 to 13.12 events per hour. This finding supports our choice to mainly focus on mountaineers in this paper~~data used are the signals from the three components of the seismometer.

5 Classifier Selection and Training

The following section describes the classifier pre-selection, training, testing and how the classifiers are used to annotate the whole data stream as illustrated in Fig. 4 (b). First, a brief introduction to convolutional neural networks is given. If the reader is unfamiliar with neural networks we recommend to read additional literature (Goodfellow et al., 2016).

5.1 Convolutional Neural Networks

Convolutional neural networks have gained a lot of interest due to their advanced feature extraction and classification capabilities. A convolutional neural network contains multiple adoptable parameters which can be updated in an iterative optimization procedure. This fact makes them generically applicable to a large range of datasets and a large range of different tasks. The convolutional neural network consists of multiple so-called convolutional layers. A convolutional layer transforms its input signal with c_i channels into c_h feature maps as illustrated in Fig. 9. A hidden feature map $F_{H,k}$ is calculated according to

$$F_{H,k} = g\left(\sum_{j=1}^{c_i} I_j * w_{k,j} + b_k\right)$$

where $*$ denotes the convolution operator, $g(\cdot)$ is a nonlinear function, I_j is an input channel, b_k is the bias related to the feature map $F_{H,k}$ and $w_{k,j}$ is the kernel related to the input image I_j and feature map $F_{H,k}$. Kernel and bias are trainable parameters of each layer. This principle can be applied to subsequent convolutional layers. Additionally, a strided convolution can be used which effectively reduces the dimension of a feature map as illustrated by L_1 in Fig. 9. In an all convolutional neural network (Springenberg et al., 2014) the feature maps of the output convolutional layer are averaged per channel. In our case, the number of output channels is chosen to be the number of event sources to be detected. Subsequent scaling and a final (non-linear) activation function are applied. If trained correctly each output represents the probability that the input contains the respective event source. In our case, this training is performed by calculating the the binary cross-entropy between the network output and the ground truth. The error is backpropagated through the neural network and the parameters are updated. The training procedure is performed for all samples in the dataset and is repeated multiple epochs.

5.2 Classifier Selection

Multiple classifiers are available for the previously selected data types: wind data, images and micro-seismic data.

For wind data a simple threshold classifier can be used, which indicates wind influences based on the wind speed. For simplicity the classifier labels time periods with wind speed above 30 km/h as *wind*. For images a convolutional neural network is selected to classify the presence of mountaineers in the image. The image classifier architecture is selected from the large pool of available image classifiers (Russakovsky et al., 2015). For micro-seismic data, three different classifiers will be pre-selected: (i) a footstep detector based on manually selected features (standard deviation, kurtosis and frequency band energies) using a linear support vector machine (LSVM) similar to the detector used in (Anchal et al., 2018), (ii) a seismic event classifier adopted from (Perol et al., 2018) and (iii) ~~an acoustic event classifier~~ a non-geophysical event classifier which we call MicroseismicCNN. We reimplemented the first two algorithms based on the information from the respective papers. The third is a major contribution in this paper and has been specifically designed ~~for acoustic event classification on~~ to identify non-geophysical sources in micro-seismic data.

The proposed convolutional neural network (CNN) ~~for acoustic event classification on~~ to identify non-geophysical sources in micro-seismic signals uses a time-frequency signal representation as input and consists of 2D convolutional layers. The time-domain signal, sampled at 1 kHz, is first offset-compensated and then transformed with a Short-Time Fourier Transformation (STFT). Subsequently, the STFT output is further processed by selecting the frequency range from 2 to 250 Hz and subdividing it into 64 linearly-spaced bands. The network consists of multiple convolutional, batch normalization and dropout layers, as

Layer	stride	output channels
Conv2D + BatchNorm + Linear	1	32
Conv2D + BatchNorm + ReLU	2	32
Dropout	-	32
Conv2D + BatchNorm + ReLU	2	32
Dropout	-	32
Conv2D + BatchNorm + ReLU	1	32
Conv2D + ~~BatchNorm +~~ ReLU	1	32
Dropout	-	32
Conv2D + ~~BatchNorm +~~ ReLU	1	1
Global Average Pooling	1	1
Dropout	-	1
Conv2D	1	1
Sigmoid Activation	-	1

Table 1. ~~Structure~~ Layout of the proposed ~~acoustic~~ non-geophysical event classifier, consisting of multiple layers which are executed in sequential order. The convolutional neural network consists of multiple 2D convolutional layers (Conv2D) with batch normalization (BatchNorm) and rectified linear units (ReLU). Dropout layers are used to minimize overfitting. The sequence of global average pooling layer, a scaling layer and the sigmoid activation compute one value between 0 and 1 resembling the probability of a detected mountaineer.

depicted in Table 1. Except for the first convolutional layer, all convolutional layers are followed by batch normalization and Rectified Linear Units (ReLU) activation. Finally, a set of global average pooling layer, dropout, trainable scaling (in form of a convolutional layer with kernel size 1) and sigmoid activation reduces the features to one value representing the probability ~~of a mountaineer~~ that a mountaineer is in the micro-seismic signal. In total the network has 30,243 parameters. In this architecture multiple measures have been taken to minimize overfitting: the network is all-convolutional (Springenberg et al., 2014), batch normalization (Ioffe and Szegedy, 2015) and dropout (Srivastava et al., 2014) are used and the size of the network is small compared to recent audio classification networks (Hershey et al., 2016).

5.3 Training and Testing

We will evaluate the micro-seismic algorithms in two scenarios in Sect. 7.1. In this section, we describe the training and test setup for the two scenarios as well as for image and ensemble classification. ~~The first scenario classifies~~ In the first scenario event-linked segments ~~. The second scenario compares~~ are classified. In the second scenario the classifiers on image-linked segments ~~, since~~ are compared. The second scenario stems from the fact that the characterization from Sect. 4.1 suggested that using a longer temporal input window could lead to a better classification because it can capture more characteristics of a mountaineer~~sound~~. Training is performed with the annotated subset from Sect. ~~??~~ 4 and a random 10 % of the training set are used as validation set, which is never used during training. For the ~~acoustic~~ non-geophysical and seismic event classifiers the number of epochs has been fixed to 100 and for the image classifier to 20. After each epoch the F1 score of the validation set is calculated and based on it the best performing network version is selected. The F1 score is defined as

$$\text{F1 score} = \frac{2 \cdot \text{true positive}}{2 \cdot \text{true positive} + \text{false negative} + \text{false positive}}$$

$$\text{F1 score} = \frac{2 \cdot \text{true positive}}{2 \cdot \text{true positive} + \text{false negative} + \text{false positive}}$$

The threshold for the network's output is determined by running a parameter search with the validation set's F1 score as metric. Training was performed in batches of 32 samples with the ADAM (Kingma and Ba, 2014) optimizer and cross-entropy loss. The Keras (Chollet and others, 2015) framework with a TensorFlow backend (Abadi et al., 2015) was used to implement and train the network. The authors of the seismic event classifier (Perol et al., 2018) provide TensorFlow source code, but to keep the training procedure the same it was re-implemented with the Keras framework. The footstep detector is trained with scikit-learn (Pedregosa et al., 2011). ~~Out of 10 runs the footstep detector which performed best on the validation set was selected.~~ Testing is performed on the test set which is independent of the training set and has not been used during training. The metrics error rate and F1 score are calculated.

~~Since neural networks are initialized with random values, and thus the classifier performance can vary, it~~ It is common to do multiple iterations of training and testing to get the best performing classifier instance. We perform a preliminary parameter search to estimate the number of iterations. The estimation takes into account the number of training types (10 different classifiers need to be trained) given the limited processing capabilities. As a result of the search, we train and test 10 iterations and select the best classifier instance of each classifier type to evaluate and compare their performances in Sect. ~~??~~7.

The input of the micro-seismic classifiers must be variable to be able to perform classification on event-linked segments and image-linked segments. Due to the principle of convolutional layers, the CNN architectures are independent of the input size and therefore no architectural changes have to be performed. The footstep detector's input features are averaged over time by design and are thus also time-invariant.

5.3.1 Event-linked Segments Experiment

Literature suggests that STA/LTA cannot distinguish ~~seismic events from other noise events~~ geophysical sources from non-geophysical sources (Allen, 1978). Therefore the first micro-seismic experiment investigates if the presented algorithms can distinguish events induced by mountaineers from other events in the signal. The event-linked segments are used for training and evaluation. The results will be discussed in Sect. 7.1.

5.3.2 Image-linked Segments Experiment

In the second micro-seismic experiment the image-linked segments will be used. Each classifier is trained and evaluated on the image-linked segments. The training parameters for training the classifiers on image-linked segments are as before but additionally data augmentation is used to minimize overfitting. Data augmentation includes random circular shift and random cropping on the time axis. Moreover, to account for the uneven distribution in the dataset, it is made sure that during training the convolutional neural networks see one example of a mountaineer every batch. The learning rate is set to 0.0001, which was determined with a preliminary parameter search. The classifiers are then evaluated on the image-linked segments.

To be able to compare the results of the classifiers trained on image-linked segments to the classifiers trained on event-linked segments (Sect. 5.3.1), the classifiers from Sect. 5.3.1 will be evaluated on the image-linked segments as well. The metrics can be calculated with the following assumption: If any of the event-linked segments which are overlapping with an image-linked segment are classified as mountaineer, the image-linked segment is considered to be classified as mountaineer as well.

The results will be discussed in Sect. 7.1.

5.3.3 Image Classification

Since convolutional neural networks are a predominant technique for image classification, a variety of network architectures have been developed. For this study, the MobileNet (Howard et al., 2017) architecture is used. The number of labeled images is small in comparison to the network size (approx. 3.2M parameters) and training the network on the Matterhorn images will lead to overfitting on the small dataset. To reduce overfitting a MobileNet implementation which has been pre-trained on ImageNet (Deng et al., 2009), a large-scale image dataset, will be used. Retraining is required since ImageNet has a different application focus than this study. The climbing route, containing the subject of interest, only covers a tiny fraction of the image and rescaling the image to 224x224, the input size of the MobileNet, would lead to ~~vanishingly~~ vanishing mountaineers (compare Fig. 3). However, the image size cannot be chosen arbitrarily large since a larger input requires more memory and results in a larger runtime. To overcome this problem the image has been ~~rescaled~~ scaled to 448x672 pixels and although the input size differs from the pretrained version network retraining still benefits from pre-trained weights. Data augmentation is used to minimize overfitting. For data augmentation each image is slightly zoomed in and shifted in width and height. The network has been trained to detect 5 different categories which are relevant for a good accuracy of the mountaineer classifier. These categories consist of mountaineer, low visibility (if the lens is partially obscured), lens flare, snowy (if the seismometer is covered in snow) and bad weather (as far as it can be deduced from the image). In this paper only the metrics for mountaineers are of interest for the evaluation and the metrics for the other labels are discarded in the following.

5.3.4 Ensemble

In certain cases, a sensor cannot identify a mountaineer although there is one, for example the seismometers cannot detect when the mountaineer is not moving or the camera does not capture the mountaineer if the visibility is low. The usage of multiple classifiers can be beneficial in these cases. In our case micro-seismic and image classifier will be jointly used for mountaineer prediction. Since micro-seismic labels and image labels are slightly different, as discussed in Sect. 4.2, the ground truth labels must be combined. For a given category, a sample is labeled true if any of micro-seismic or image labels are true (logical disjunction). After individual prediction by each classifier the outputs of the classifiers are combined similarly and can be compared to the ground truth.

5.4 Optimization

Due to potential human errors during data labeling, the training set has to be regarded as a weakly-labeled dataset. Such datasets can lead to a worse classifier performance. To overcome this issue a human-in-the-loop approach is followed where a preliminary set of classifier is trained on the training set. In the next step, each sample of the dataset is automatically classified. This procedure produces a number of true positives, false positives and true negatives. These samples are then manually relabeled and the labels for the dataset are updated based on human review. The procedure is repeated multiple times. This does however not completely avoid the possibility of falsely labeled samples in the dataset, since the algorithm might not find all human-labeled false negatives, but it increases the accuracy significantly. The impact of false labels on classifier performance will be evaluated in Sect 7.1.

5.5 ~~Automatic Annotation~~

6 Automatic Classification

In Sect. 7.1 it will be shown that the best set of classifiers are the ensemble of image classifier and MicroseismicCNN. Therefore, the trained image classifiers and MicroseismicCNN classifier are used to annotate the whole time period of collected data to quantitatively assess the impact of mountaineers. The image classifier and the MicroseismicCNN will be used to classify all the images and micro-seismic data, respectively. The consecutive segments and images are used for prediction. To avoid false positives we assume that a mountaineer requires a certain amount of time to pass by the seismometer as illustrated in Fig. 7, therefore a mountaineer annotation is only considered valid if its minimum distance to the next (or previous) mountaineer annotation is less than 5 minutes. Subsequently, events within time periods classified as mountaineer are removed and the event count per hour is calculated.

~~None Mountaineer Wind duration (hours) 6832.3 296.53 1364.2 mean number of events per hour 11.76 105.9 9.09 Statistics of the automatically annotated dataset (2017 only) per annotation category. "None" represents the category when none of the other categories apply. Given are the total duration of annotated segments per category and the mean number of STA/LTA events per category.~~

7 ~~Results~~Evaluation

In the following the results of the different classifiers experiments described in Sect. 5.3 will be presented to determine the best set of classifiers. Furthermore, in Sect. ~~??~~ 7.2 and Sect. 7.3 results of the automatic annotation process (Sect. ~~??~~6) will used to evaluate the impact of external influences on the whole dataset.

7.1 Classifier Evaluation

The results of the classifier experiments ~~from Sect. 5.3 are listed in Table2. The table shows~~ (Table 2) show that the footstep detector is the worst at ~~differentiating events~~classifying mountaineers with an error rate of 0.1702 on event-linked segments

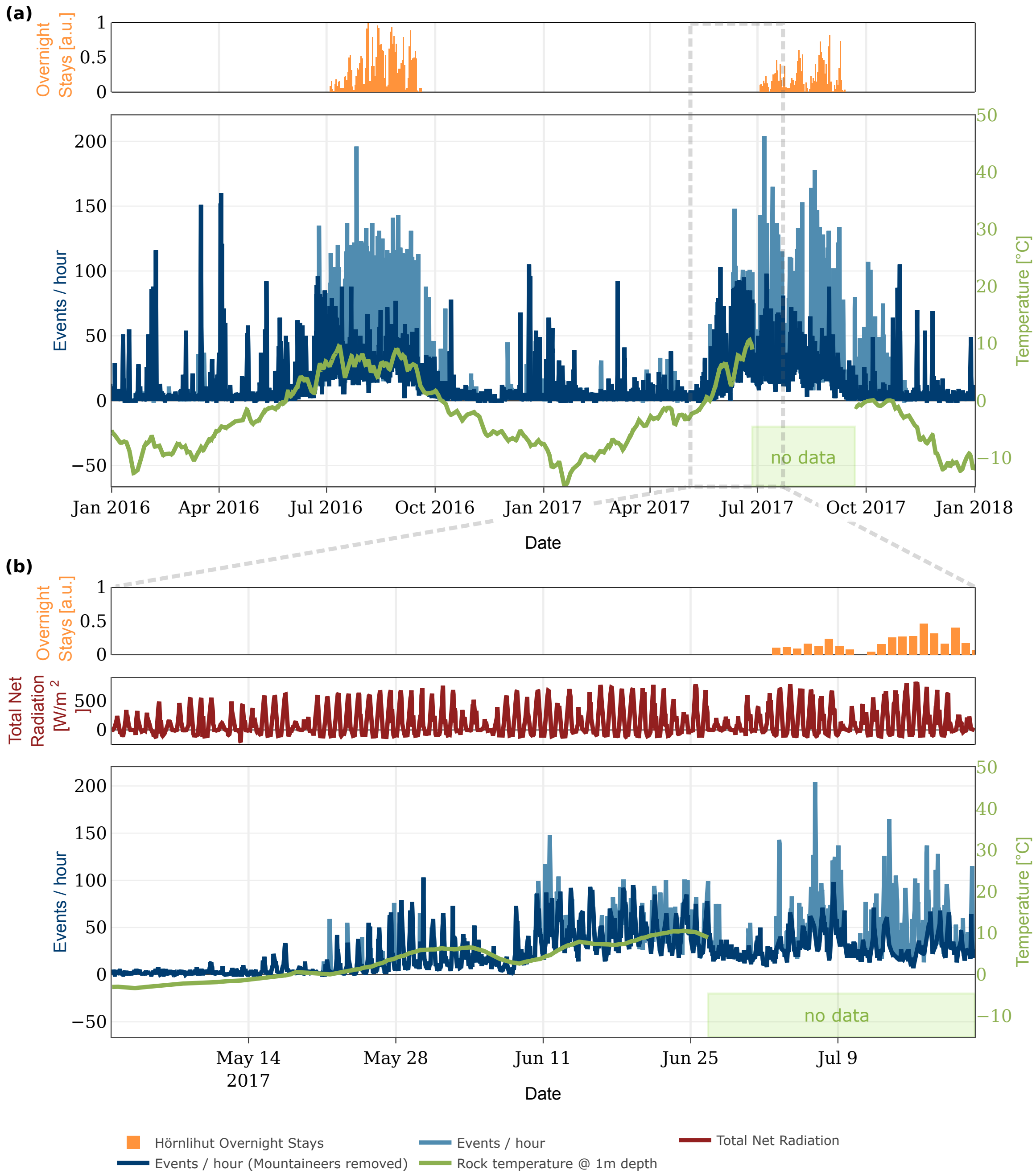

Figure 10. Event count, hut occupancy and rock temperature over time. (a) For the years 2016/2017 and (b) for a selected period during defreezing of the rock. The event rate from (Weber et al., 2018b) is illustrated in light blue and the rate after removal of mountaineer induced events in dark blue. The strong variations in event ~~count~~ rate correlate with the presence of mountaineer, hut occupancy and in (b) with the total net radiation. The impact of mountaineers is significant after July 9th and event detection analysis becomes unreliable.

	Error Rate	F1 Score
Event-linked Segments		
Footstep Detector (Events)	0.1702	0.7692
Seismic Event Classifier (Events)	0.1250	0.8291
MicroseismicCNN (Events)	**0.0641**	**0.9062**
Image-linked Segments		
Footstep Detector (Events)	0.0706	0.5389
Seismic Event Classifier (Events)	0.0540	0.6047
MicroseismicCNN (Events)	0.0309	0.731
Footstep Detector	0.0952	0.52
Seismic Event Classifier	0.0313	0.7383
MicroseismicCNN	~~0.0096~~**0.0096**	~~0.9167~~**0.9167**
Image Classifier	0.0088	0.9134
Ensemble	**0.0079**	**0.9383**

Table 2. Results of the different classifiers. The addition "(Events)" labels the classifier versions trained on event-linked segments

and 0.0952 on image-linked segments. Both convolutional neural networks score a lower error rate ~~of 0.0096~~ on image-linked segments of 0.0096 (MicroseismicCNN) and 0.0313 (Seismic Event Classifier). For the given dataset our proposed MicroseismicCNN network outperforms the seismic event classifier, in both the event-linked segment experiment as well as the image-linked segment experiment. The MicroseismicCNN using a longer input window (trained on image-linked segments) is comparable to classification on images and outperforms the classifier trained on event-linked segments. When combining image and micro-seismic classifiers the best results can be achieved.

The number of training/test iterations that were run for each classifier has been set to 10 through a preliminary parameter estimation. To validate our choice we have evaluated the influences of the number of experiments for only one classifier. The performance of the classifier is expected to depend on the number of training/test iterations (more iterations means a better chance of selecting the best classifier). However, the computing time is increasing linearly with increasing number of iterations. Hence, a reasonable trade off between the performance of the classifier and the computing time is desired to identify the ideal number of iterations. Figure 11 represents the statistical distribution of the classifier's performance for different number of training/test iterations. Each boxplot is based on ten independent sets of training/test iterations. While the box indicates the interquartile range (IQR) with the median value in orange, the whisker on the appropriate side is taken to $1.5 x IQR$ from the quartile instead of of the maximum or minimum if either type of outlier is present. Beyond the whiskers, data are considered outliers and are plotted as individual points. As can be seen in Fig. 11, the F1 score saturates at 9 iterations. Therefore our choice of 10 iterations is a reasonable choice.

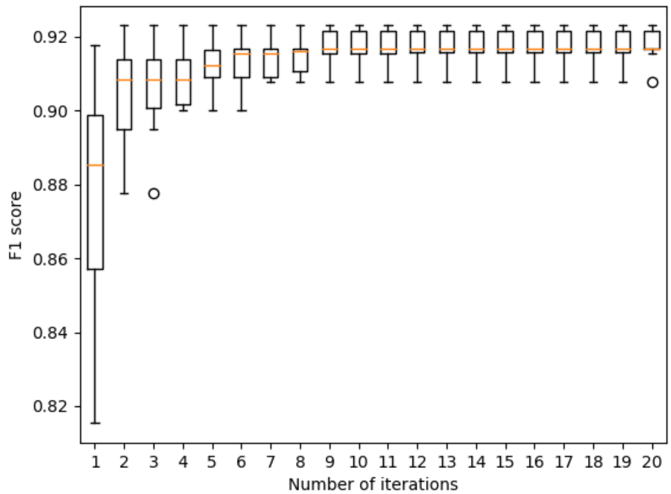

Figure 11. ~~Correlation of mountaineer activity and hut occupancy.~~ The ~~normalized number~~ statistical distribution of ~~mountaineer segments per week and~~ the ~~normalized~~ classifier's performance for different number of ~~overnight stays at the Hörnlihut per week plotted over time~~training/test iterations is illustrated. Each boxplot is based on ten independent sets of training/test iterations. The F1 score saturates after 9 iterations and validates our choice of 10 iterations.

False Labels (%)	25	12.5	6.25	3.125	0
F1 score (mean)	0.7953	0.8633	0.8761	0.8835	0.8911
Error rate (mean)	0.0208	0.0149	0.0139	0.0129	0.0122

Table 3. Influence of falsely labeled data points on the test performance. Shown are the mean values over 10 training/test iterations.

7.2 ~~External Influences~~

In Sect. ~~4.1 and 7.2 we have shown that external influences can have a strong impact on the quality of analysis.From Table ?? it can be deduced that the average number of triggered eventsper minute is high for times when the signal is influenced by mountaineers (~~5.4 the possibility of falsely labeled training samples has been discussed. As expected, our evaluation in Table 3 indicates that falsely labeled samples have an influence on the classification performance since the mean performances are worse for a high percentage of false labels.

7.2 Statistical Evaluation

The annotated test set from Sect. 4.2 and the automatically annotated set from Sect. 6 are used for a statistical evaluation involving the impact of external influences on micro-seismic events. Only data from 2017 is chosen since wind data is not available for the whole training set due to a malfunction of the weather station in 2016. The experiment from (Weber et al., 2018b)

		Unknown	Mountaineer	Wind	Helicopter
Manual	duration (hours)	28.87	1.9	6.6	3.73
	mean number of events per hour	10.6	95.26	11.21	13.12
Automatic	duration (hours)	6832.3	296.53	1364.2	-
	mean number of events per hour	11.76	105.9	9.09	-

Table 4. Statistics of the manually and automatically annotated set of 2017 per annotation category. "Unknown" represents the category when none of the other categories could have been identified. Given are the total duration of annotated segments per category and the mean number of STA/LTA events per category.

provides STA/LTA event triggers 2017. Table 4 shows statistics for several categories, which are 3 external influences and one category where none of the 3 external influences are annotated (declared as "Unknown"). For each category, the total duration of all annotated segments is given and how many events per hour are triggered. It becomes apparent that mountaineers have the biggest impact on the event analysis. Up to 105.9 events ~~/hour), which is an approximately 9x increase in comparison to periods without annotated external influences. This~~ per hour are detected on average during time periods with mountaineer activity, while during all other time periods the average ranges from 9.09 to 13.12 events per hour. This finding supports our choice to mainly focus on mountaineers in this paper and shows that mountaineers have a strong impact on the analysis~~and~~. As a consequence, a high activity detected by the event trigger does not correspond to a high seismic activity, thus relying only on this kind of event detection may lead to a false interpretation. From the automatic section in Table 4 it can be deduced that the average number of triggered events per hour for times when the signal is influenced by mountaineers is an approximately 9x increase in comparison to periods without annotated external influences. The effect of wind influences on event rate is not as clear as the influence of mountaineers. The values in Table ~~??~~ 4 indicate a decrease of events per hour during wind periods, which will be briefly discussed in Sect. 8.2.

As can be seen in Fig. 12, events are triggered over the course of the whole year whereas events that are annotated as coming from mountaineers occur mainly during the summer period. The main increase in event count occurs during the period when the rock is unfrozen which unfortunately coincides with the period of mountaineer activity. Therefore it is important to account for the mountaineers. However, even if the mountaineers are not considered the event count increases significantly during the unfrozen period. The interpretation of these results will not be part of this study but they are an interesting topic for further research.

7.3 Automatic Annotation in a Real-World Scenario

The results of applying the ensemble classifier to the whole dataset is visualized for two time periods in Fig. 10. The figure depicts the event count per hour before and after removing periods of mountaineer activity, as well as the rock temperature, the overnight stays at the Hörnlihut and the total net radiation. From Fig. 10 (a) ~~is~~ it becomes apparent that the mountaineer activity is mainly present during summer and autumn. An increase is also visible during increasing hut overnight stays. During winter

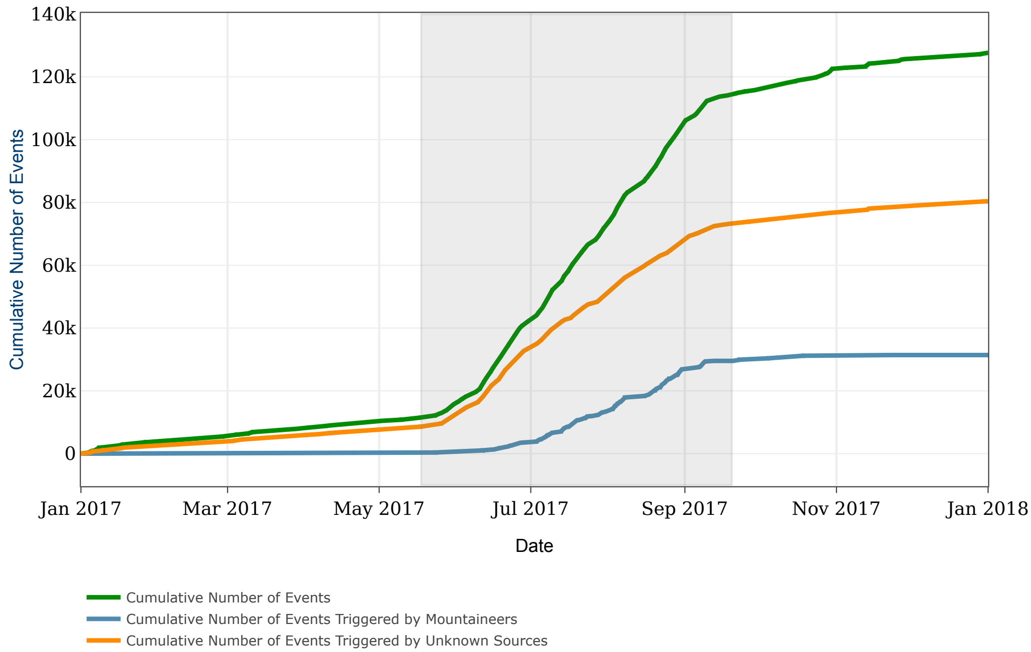

Figure 12. Illustration of the cumulative number of events triggered by the STA/LTA event detector for all events, for events triggered by mountaineers and for events triggered by unknown sources. The results presented in this paper were used to annotate the events. The time period during which the rock temperature in 1 m depth is above 0° C is shaded in gray.

and spring only few mountaineers are detected but some activity peaks remain. By manually review we were able to discard mountaineers as cause for most of these peaks, however further investigation is needed to explain their ~~occurence~~occurrence.

Figure 10 (b) focuses on the defreezing period. The zero-crossing of the rock temperature has a significant impact on the event count variability. A daily pattern becomes visible starting around the zero-crossing. Since few mountaineers are detected in May these can be discarded as the main influence for these patterns. The total net radiation however indicates an influences of solar radiation on the event count. Further in-depth analysis is needed but this examples shows the benefits of a domain-specific analysis, since the additional information gives an intuition of relevant processes and their ~~interdependences~~interdependencies. After July 9th, the impact of mountaineers is significant and the event detection analysis becomes unreliable. Different evaluation methods are required to mitigate the influence of mountaineers during these periods.

Figure 13 depicts that mountaineer predictions and hut occupancy correlate well, which indicates that the classifiers work well. The discrepancy in the first period of each summer needs further investigation. With the annotations for the whole timespan it can be estimated that from all events detected in (Weber et al., 2018b), approximately 25% originate in time periods with mountaineer activity and should therefore not be regarded as ~~seismic events~~originating from geophysical sources.

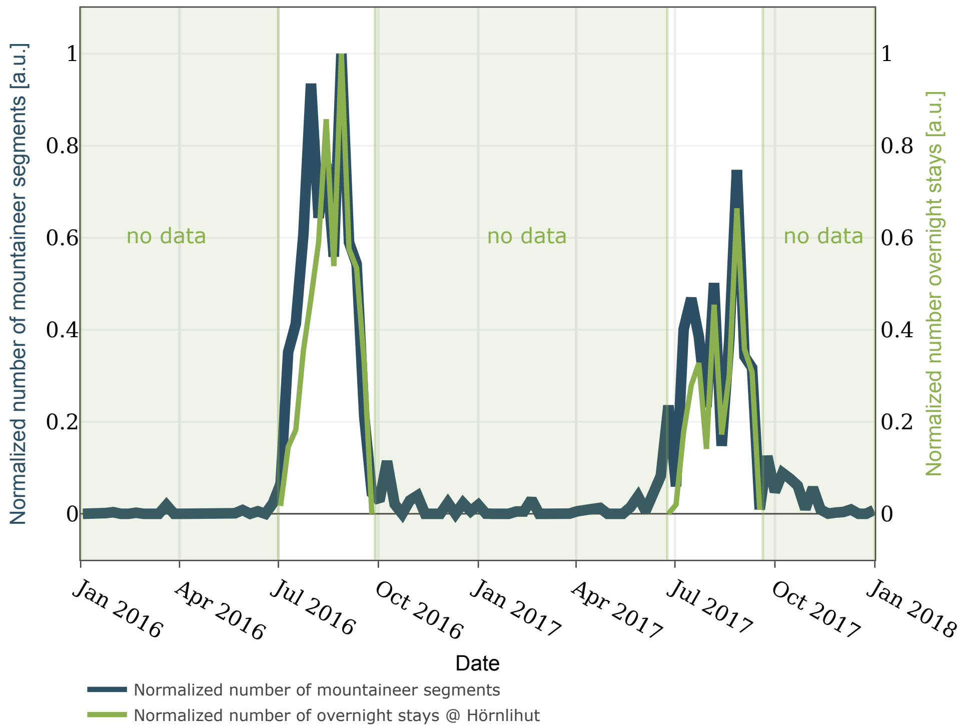

Figure 13. Correlation of mountaineer activity and hut occupancy. The normalized number of mountaineer segments per week and the normalized number of overnight stays at the Hörnlihut per week plotted over time.

8 Discussion

8.1 ~~Labeling~~Classification of Negative Examples

The previous section has shown that ~~an~~ a certain degree of understanding of the ~~collected data is necessary for~~ scenario and data collected is nevertheless necessary in order to achieve a significant analysis. The ~~creation of such an annotated~~ effort in creating an annotated data subset, despite being time and labor consuming, is ~~therefore an overhead which is outweighted~~ an overhead but as we show can be outweighed by the benefits of ~~a better analysis~~ better analysis results. For data annotation two ~~main~~ distinct approaches can be followed~~, annotating~~ : Annotating the phenomena of interest (positive examples) or annotating the external influences (negative examples). ~~For positive examples, which are~~ Positive examples, used in (Yuan et al., 2018; Ruano et al., 2014; Kislov and Gravirov, 2017), ~~we would be restricted to find seismic events in the sensor being influenced by compounding factors and thus~~ inherently contain a limitation as this approach requires that events as well as influencing factors must be identified and identifiable in the signal of concern. This is especially hard where no ground truth information except ~~for~~ (limited) experience by professionals ~~can be used~~is available. Therefore, the strategy presented in this work ~~creates~~ to create an annotated dataset using negative examples ~~, because they can be more ascertainable for human annotators allowing~~ is advisable to be used. It offers to perform cross checks if certain patterns can be found in different sensor/data types and in many cases the annotation process ~~to be outsourced~~can be performed by non-experts. Also, additional sensors ~~can provide a direct measure of possible influences~~allow to directly quantify possible influence factors. The detour required by first classifying negative examples and then analyzing the phenomena of interest offers further benefits: In cases where the characteristics of the phenomena of interest are not known in advance (no ground truth available) and in cases where a novel analysis method is to be applied or when treating very-long-term monitoring datasets working only on the primary signal of concern is hard and error margins are likely to be large. In these cases it is important to take into account all knowledge available including possible negative examples and it is significant to automate as much as possible using automatic classification methods.

8.2 Multi-Sensor Classification

In Sect. 5 multiple classifiers for different sensors have been presented. The advantages of classifying ~~on~~ micro-seismic signals are that continuous detection is possible and that no additional ~~sensor is~~ sensors are required. The classification accuracy of ~~our presented~~ the convolutional neural network and the image classifier presented are comparable. ~~Image classification has however~~ Classification of time-lapse images however has the disadvantage of a low time resolution ~~with a capture frequencyof at maximum~~ proportional to the capture frequency, for example a maximum of 15 images per hour in our example. Continuous video recording ~~would close the gap , but the complexity of the image classifierand~~ could close this gap at the cost of requiring a more complex image classifier, the size of ~~it's input results in a high processing time, which makes it unfeasible for video~~the data and more higher processing times, which are likely infeasible. The main advantage of images is that they can be used as additional independent sensor to augment and verify micro-seismic recordings. First, images can be used for annotations and second they can be used in an ensemble classifier to increase the overall accuracy. The different modalities strengthen the overall meaningfulness and make the classifier more robust. Table ~~??~~ 4 shows that during ~~wind~~ windy segments less events are triggered than ~~for periods without a category~~in periods that cannot be categorized ("unknown" category). A possible explanation ~~could be~~ is that the micro-seismic activity is superimposed by broadband noise ~~coming from~~ originating in the wind. For these time segments a ~~different parameter set or~~ variable trigger sensitivity (Walter et al., 2008) or a different event detection algorithm ~~could~~ can improve the analysis. ~~Shielding~~ Better shielding the seismometer from the wind would ~~probably reduce the wind influences significantly , but the common approach~~ reduce these influences significantly but the typical approach in seismology to embed it into the ground ~~is difficult~~ under a substantial soil column is next to impossible to implement in steep ~~rock and hard to reach regions. Nonethelesswe have shown that with the presented method we are able~~ bedrock and perennially frozen ground as found on our case-study field site. Nonetheless, Table 4 gives an intuition that our method performs well since the statistical distribution of manually and automatically annotated influences sources is similar. We therefore conclude that with our method presented it is possible to quantify the impact of external influences on a long-term scale and across variable conditions.

8.3 Feature Extraction

In Sect. 4.1 the different characteristics of event sources have been discussed. The characteristic features can be used to identify and classify each source type. The convolutional neural network accomplishes the task of feature extraction and classification simultaneously by training on an extensive annotated dataset. An approach without the requirement of an annotated dataset would be to manually identify the characteristics and then design a suitable algorithm to extract the features. For example the helicopter pattern in 5 (c) shows distinct energy bands indicating the presence of a fundamental frequency plus harmonics. These features could be traced to identify, model and and possibly localize a helicopter (Eibl et al., 2017) with the advantage of a relaxed dataset requirement. The disadvantage would be the requirement of further expertise in the broad field of digital signal processing and modeling as well as more detailed knowledge on each such phenomena class of interest. Also, it is likely that such an approach would require extensive sensitivity analysis to be performed alongside with modeling. Moreover, if the algorithm is handcrafted by using few examples it is prone to overfitting based on these examples (see also the next subsection). This problem of overfitting exist as well for algorithm training and can be solved by using more examples, however, it is easier to annotate a given pattern (with the help of additional information) than understanding its characteristics and thus the time- and labor-consuming task of annotation can be outsourced in the case of machine learning. Fig. 5 indicates that little anthropogenic noise (a) has less broadband background noise than wind (d) and the impulses occur in a different frequency band. However, the signal plots show a similar pattern. To identify wind from micro-seismic data manually one could utilize a frequency-selective event detector although it is not clear if this pattern and frequency range is representative for every occurrence of wind and if all non-wind events could be excluded with such a detector. Using a dedicated wind sensor for identification of wind periods as presented in this study overcomes these issues with the drawback of an additional sensor which needs to be installed and maintained and that during failure of the additional sensor no annotation can be performed.

8.4 Overfitting

A big problem with machine learning methods is overfitting due to too few data examples. Instead of learning representative characteristics the algorithm memorizes the examples. In our work overfitting is an apparent issue since the reference dataset is small as described in Sect. 4.2. As explained in the previous sections multiple measures have been introduced to reduce overfitting (data augmentation, few parameters, all convolutional neural network, dropout). The test set has been specifically selected to be from a different year to exclude that severe overfitting affects the classifier performance. The test set includes examples from all seasons, day and night time and is thus assumed representative for upcoming, never-seen-before data. However, overfitting might still exist in the sense that the classifier is optimized for one specific seismometer. Generalization to multiple seismometers still needs to be proven since we did not test the same classifier for multiple seismometers, which might differ in their specific location, type or frequency response. This will be an important study for the future since it will reduce the dataset collection and training time significantly if a new seismometer is deployed.

8.5 Outlook

This work has only focused on identifying external influences, what we have shown to be a prerequisite for micro-seismic analysis. Future work lies in finding and applying specific ~~analytical~~ analytic methods, especially finding good parameter sets and algorithms for each context. Additionally, the classifier could be extended to include helicopters ~~and rockfalls.~~ as well as geophysical sources such as rockfalls. A disadvantage of the present method is the requirement of a labeled dataset. Semi-supervised or unsupervised methods (Kuyuk et al., 2011) as well as one- or few-shot classification methods (Fei-Fei et al., 2006) could provide an alternative to the presented training concept without the requirement of a large annotated dataset.

9 Conclusions

In this paper we have presented a strategy to evaluate the impact of external influences on a micro-seismic measurement by categorizing the data with the help of additional sensors and information. With this knowledge a method to classify mountaineers has been presented. We have shown how additional sensors can be beneficial to isolate the information of interest from unwanted external influences and provide a ground truth in a long-term monitoring setup. Moreover, we have presented a mountaineer detector, implemented with a convolutional neural network, which scores an error rate of only 0.96 % (F1 score: ~~0.91670~~0.9167) on micro-seismic signals and a mountaineer detector ensemble which scores an error rate of 0.79 % (F1 score: 0.9383) on images and micro-seismic data. The classifiers outperform comparable algorithms. Their application to a real-word, multi-sensor, multi-year micro-seismic monitoring experiment showed that time periods with mountaineer activity have a approximately 9x higher event rate and that approximately 25% of all detected events are due to mountaineer interference. Finally, the findings of this paper show that an extensive, systematic identification of external influences is required for a ~~qualitative analysis~~quantitative and qualitative analysis on long-term monitoring experiments.

Code and data availability. The dataset is available under the DOI 10.5281/zenodo.1320835 and the accompanying code under the DOI 10.5281/zenodo.1321176.

Author contributions. Matthias Meyer, Samuel Weber, Jan Beutel and Lothar Thiele developed the concept. Matthias Meyer and Samuel Weber developed the code and maintained field site and data together with Jan Beutel. Matthias Meyer prepared and performed the experiments and evaluated the results with Samuel Weber. Matthias Meyer prepared the manuscript as well as the visualizations with contributions from all co-authors.

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

Acknowledgements. The work presented in this paper is part of the X-Sense 2 project. It was financed by http://www.nano-tera.ch (ref. no. 530659). We would like to thank Tonio Gsell and the rest of the PermaSense team for technical support. We acknowledgement Kurt Lauber for providing us with hut occupancy data and the Air Zermatt helicopter company for providing us with helicopter flight data. We thank Lukas Cavigelli for insightful discussions.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X.: TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, 2015.

Allen, R. V.: Automatic Earthquake Recognition and Timing from Single Traces, Bulletin of the Seismological Society of America, 68, 1521–1532, 1978.

Amitrano, D., Grasso, J. R., and Senfaute, G.: Seismic Precursory Patterns before a Cliff Collapse and Critical Point Phenomena, Geophysical Research Letters, 32, L08 314, https://doi.org/10.1029/2004GL022270, 2005.

Amitrano, D., Arattano, M., Chiarle, M., Mortara, G., Occhiena, C., Pirulli, M., and Scavia, C.: Microseismic Activity Analysis for the Study of the Rupture Mechanisms in Unstable Rock Masses, Nat. Hazards Earth Syst. Sci., 10, 831–841, https://doi.org/10.5194/nhess-10-831-2010, 2010.

Amitrano, D., Gruber, S., and Girard, L.: Evidence of Frost-Cracking Inferred from Acoustic Emissions in a High-Alpine Rock-Wall, Earth and Planetary Science Letters, 341-344, 86–93, https://doi.org/10.1016/j.epsl.2012.06.014, 2012.

Anchal, S., Mukhopadhyay, B., and Kar, S.: UREDT: Unsupervised Learning Based Real-Time Footfall Event Detection Technique in Seismic Signal, IEEE Sensors Letters, 2, 1–4, https://doi.org/10.1109/LSENS.2017.2787611, 2018.

Bartholomaus, T. C., Amundson, J. M., Walter, J. I., O'Neel, S., West, M. E., and Larsen, C. F.: Subglacial Discharge at Tidewater Glaciers Revealed by Seismic Tremor, Geophysical Research Letters, 42, 6391–6398, https://doi.org/10.1002/2015GL064590, 2015.

Brown, J. R., Beroza, G. C., and Shelly, D. R.: An Autocorrelation Method to Detect Low Frequency Earthquakes within Tremor, Geophysical Research Letters, 35, https://doi.org/10.1029/2008GL034560, 2008.

Burjánek, J., Moore, J. R., Molina, F. X. Y., and Fäh, D.: Instrumental Evidence of Normal Mode Rock Slope Vibration, Geophysical Journal International, 188, 559–569, https://doi.org/10.1111/j.1365-246X.2011.05272.x, 2012.

Chollet, F. and others: Keras, Python Framework, 2015.

Colombero, C., Comina, C., Vinciguerra, S., and Benson, P. M.: Microseismicity of an Unstable Rock Mass: From Field Monitoring to Laboratory Testing, Journal of Geophysical Research: Solid Earth, https://doi.org/10.1002/2017JB014612, 2018.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L.: ImageNet: A Large-Scale Hierarchical Image Database, in: CVPR09, 2009.

Eibl, E. P. S., Lokmer, I., Bean, C. J., and Akerlie, E.: Helicopter Location and Tracking Using Seismometer Recordings, Geophysical Journal International, 209, 901–908, https://doi.org/10.1093/gji/ggx048, 2017.

Fei-Fei, L., Fergus, R., and Perona, P.: One-Shot Learning of Object Categories, IEEE Transactions on Pattern Analysis and Machine Intelligence, 28, 594–611, https://doi.org/10.1109/TPAMI.2006.79, 2006.

Geometrics: Geode Exploration Seismograph Specification Sheet, version GeodeDS_v1 (0518), 2018.

Gibbons, S. J. and Ringdal, F.: The Detection of Low Magnitude Seismic Events Using Array-Based Waveform Correlation, Geophysical Journal International, 165, 149–166, https://doi.org/10.1111/j.1365-246X.2006.02865.x, 2006.

Girard, L., Beutel, J., Gruber, S., Hunziker, J., Lim, R., and Weber, S.: A Custom Acoustic Emission Monitoring System for Harsh Environments: Application to Freezing-Induced Damage in Alpine Rock Walls, Geoscientific Instrumentation, Methods and Data Systems, 1, 155–167, 2012.

Gischig, V. S., Eberhardt, E., Moore, J. R., and Hungr, O.: On the Seismic Response of Deep-Seated Rock Slope Instabilities — Insights from Numerical Modeling, Engineering Geology, 193, 1–18, https://doi.org/10.1016/j.enggeo.2015.04.003, 2015.

Glade, T., Anderson, M., and Crozier, M. J.: Landslide Hazard and Risk, John Wiley & Sons, Ltd, Chichester, West Sussex, England, https://doi.org/10.1002/9780470012659, 2005.

Goodfellow, I., Bengio, Y., and Courville, A.: Deep Learning, Adaptive computation and machine learning, The MIT Press, Cambridge, Massachusetts, 2016.

Grosse, C. U. and Ohtsu, M., eds.: Acoustic Emission Testing, Springer-Verlag, Berlin Heidelberg, 2008.

Hardy, H. R.: Acoustic Emission, Microseismic Activity, Balkema, Lisse, oCLC: ocm51911888, 2003.

Hershey, S., Chaudhuri, S., Ellis, D. P. W., Gemmeke, J. F., Jansen, A., Moore, R. C., Plakal, M., Platt, D., Saurous, R. A., Seybold, B., Slaney, M., Weiss, R. J., and Wilson, K.: CNN Architectures for Large-Scale Audio Classification, arXiv: 1609.09430, 2016.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H.: MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, arXiv:1704.04861 [cs], 2017.

Ioffe, S. and Szegedy, C.: Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, arXiv:1502.03167 [cs], 2015.

Keller, M., Yuecel, M., and Beutel, J.: High Resolution Imaging for Environmental Monitoring Applications, in: International Snow Science Workshop 2009: Programme and Abstracts, pp. 197–201, Davos, Switzerland, 2009.

Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, arXiv:1412.6980 [cs], 2014.

Kislov, K. V. and Gravirov, V. V.: Use of Artificial Neural Networks for Classification of Noisy Seismic Signals, Seismic Instruments, 53, 87–101, https://doi.org/10.3103/S0747923917010054, 2017.

Kuyuk, H. S., Yildirim, E., Dogan, E., and Horasan, G.: An Unsupervised Learning Algorithm: Application to the Discrimination of Seismic Events and Quarry Blasts in the Vicinity of Istanbul, Natural Hazards and Earth System Sciences, 11, 93–100, https://doi.org/https://doi.org/10.5194/nhess-11-93-2011, 2011.

Levy, C., Jongmans, D., and Baillet, L.: Analysis of Seismic Signals Recorded on a Prone-to-Fall Rock Column (Vercors Massif, French Alps), Geophysical Journal International, 186, 296–310, https://doi.org/10.1111/j.1365-246X.2011.05046.x, 2011.

Li, Z., Meier, M.-A., Hauksson, E., Zhan, Z., and Andrews, J.: Machine Learning Seismic Wave Discrimination: Application to Earthquake Early Warning, Geophysical Research Letters, 45, 4773–4779, https://doi.org/10.1029/2018GL077870, 2018.

Michlmayr, G., Cohen, D., and Or, D.: Sources and Characteristics of Acoustic Emissions from Mechanically Stressed Geologic Granular Media — A Review, Earth-Science Reviews, 112, 97–114, https://doi.org/10.1016/j.earscirev.2012.02.009, 2012.

Occhiena, C., Coviello, V., Arattano, M., Chiarle, M., Morra di Cella, U., Pirulli, M., Pogliotti, P., and Scavia, C.: Analysis of Microseismic Signals and Temperature Recordings for Rock Slope Stability Investigations in High Mountain Areas, Nat. Hazards Earth Syst. Sci., 12, 2283–2298, https://doi.org/10.5194/nhess-12-2283-2012, 2012.

Olivier, G., Chaput, J., and Borchers, B.: Using Supervised Machine Learning to Improve Active Source Signal Retrieval, Seismological Research Letters, 89, 1023–1029, https://doi.org/10.1785/0220170239, 2018.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E.: Scikit-Learn: Machine Learning in Python, Journal of Machine Learning Research, 12, 2825–2830, 2011.

Perol, T., Gharbi, M., and Denolle, M.: Convolutional Neural Network for Earthquake Detection and Location, Science Advances, 4, e1700 578, https://doi.org/10.1126/sciadv.1700578, 2018.

Reynen, A. and Audet, P.: Supervised Machine Learning on a Network Scale: Application to Seismic Event Classification and Detection, Geophysical Journal International, 210, 1394–1409, https://doi.org/10.1093/gji/ggx238, 2017.

Ross, Z. E., Meier, M.-A., and Hauksson, E.: P Wave Arrival Picking and First-Motion Polarity Determination With Deep Learning, Journal of Geophysical Research: Solid Earth, 123, 5120–5129, https://doi.org/10.1029/2017JB015251, 2018.

Ruano, A. E., Madureira, G., Barros, O., Khosravani, H. R., Ruano, M. G., and Ferreira, P. M.: Seismic Detection Using Support Vector Machines, Neurocomputing, 135, 273–283, https://doi.org/10.1016/j.neucom.2013.12.020, 2014.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L.: ImageNet Large Scale Visual Recognition Challenge, International Journal of Computer Vision, 115, 211–252, https://doi.org/10.1007/s11263-015-0816-y, 2015.

Senfaute, G., Duperret, A., and Lawrence, J. A.: Micro-Seismic Precursory Cracks Prior to Rock-Fall on Coastal Chalk Cliffs: A Case Study at Mesnil-Val, Normandie, NW France, Natural Hazards and Earth System Sciences, 9, 1625–1641, 2009.

Spillmann, T., Maurer, H., Green, A. G., Heincke, B., Willenberg, H., and Husen, S.: Microseismic Investigation of an Unstable Mountain Slope in the Swiss Alps, Journal of Geophysical Research, 112, https://doi.org/10.1029/2006JB004723, 2007.

Springenberg, J. T., Dosovitskiy, A., Brox, T., and Riedmiller, M.: Striving for Simplicity: The All Convolutional Net, arXiv:1412.6806 [cs], 2014.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R.: Dropout: A Simple Way to Prevent Neural Networks from Overfitting, J. Mach. Learn. Res., 15, 1929–1958, 2014.

van Herwijnen, A. and Schweizer, J.: Monitoring Avalanche Activity Using a Seismic Sensor, Cold Regions Science and Technology, 69, 165–176, https://doi.org/10.1016/j.coldregions.2011.06.008, 2011.

Walter, F., Deichmann, N., and Funk, M.: Basal Icequakes during Changing Subglacial Water Pressures beneath Gornergletscher, Switzerland, Journal of Glaciology, 54, 511–521, https://doi.org/10.3189/002214308785837110, 2008.

Weber, S., Fäh, D., Beutel, J., Faillettaz, J., Gruber, S., and Vieli, A.: Ambient Seismic Vibrations in Steep Bedrock Permafrost Used to Infer Variations of Ice-Fill in Fractures, Earth and Planetary Science Letters, 501, 119–127, https://doi.org/10.1016/j.epsl.2018.08.042, 2018a.

Weber, S., Faillettaz, J., Meyer, M., Beutel, J., and Vieli, A.: Acoustic and Microseismic Characterization in Steep Bedrock Permafrost on Matterhorn (CH), Journal of Geophysical Research: Earth Surface, 123, 1363–1385, https://doi.org/10.1029/2018JF004615, 2018b.

Werner-Allen, G., Lorincz, K., Johnson, J., Lees, J., and Welsh, M.: Fidelity and Yield in a Volcano Monitoring Sensor Network, in: Proceedings of the 7th Symposium on Operating Systems Design and Implementation, OSDI '06, pp. 381–396, USENIX Association, Berkeley, CA, USA, 2006.

Withers, M., Aster, R., Young, C., Beiriger, J., Harris, M., Moore, S., and Trujillo, J.: A Comparison of Select Trigger Algorithms for Automated Global Seismic Phase and Event Detection, Bulletin of the Seismological Society of America, 88, 95–106, 1998.

Yuan, S., Liu, J., Wang, S., Wang, T., and Shi, P.: Seismic Waveform Classification and First-Break Picking Using Convolution Neural Networks, IEEE Geoscience and Remote Sensing Letters, 15, 272–276, https://doi.org/10.1109/LGRS.2017.2785834, 2018.