Machine Learning improves warning systems of debris flows

Małgorzata Chmiel[1], Fabian Walter[1], Michaela Wenner[1,2], Zhen Zhang[1,3,4], Brian W. McArdell[2], Clement Hibert[6]

[1]Laboratory of Hydraulics, Hydrology and Glaciology, ETH Zürich, Zürich, Switzerland
[2]Swiss Federal Institute for Forest, Snow and Landscape Research, Zürich, Switzerland
[3]Key Laboratory of Mountain Hazards and Surface Process, Institute of Mountain Hazards and Environment, Chinese Academy of Sciences, Chengdu, China
[4]University of Chinese Academy of Sciences, Beijing, China
[5]Institut de Physique du Globe de Strasbourg, CNRS UMR 7516, University of Strasbourg/EOST, 7 8 Strasbourg, France

Key Points:

- A novel debris-flow detector is developed using a machine-learning model and seismic data from a Swiss torrent.
- Signals of 22 debris flows recorded by six seismic stations are used to train and test the machine-learning model.
- A detector is running on the continuous real-time 2020 data stream detecting all 13 debris flows in 3 months and raising no false alarms.

Corresponding author: Małgorzata Chmiel, `chmielm@ee.ethz.ch`

Abstract

Automatic identification of debris flow signals in continuous seismic records remains a challenge. To tackle this problem we use a machine learning approach, which can be applied to continuous real-time data streams. We show that a machine learning model based on the random forest algorithm recognizes different stages of debris flow formation and propagation at the Illgraben torrent, Switzerland, with an accuracy exceeding 90 %. In contrast to typical debris flow detection requiring instrumentation installed directly in the torrent, our approach provides a significant gain in warning times of tens of minutes to hours. For real-time data streams from 2020, our detector raises alarms for all 8 independently confirmed Illgraben events and gives no false alarms. We suggest that our seismic machine-learning detector is a critical step towards the next generation of debris-flow warning, which increases warning times using both simpler and cheaper instrumentation compared to existing operational systems.

Plain Language Summary

Debris flows are fast-moving masses of mud, soil, fragmented rock, and water transporting large volume of material in mountainous areas. They pose a significant danger to human life, properties, and infrastructure. Thus, it is crucial to reliably detect debris flows early enough to send an alarm message to local communities. We propose a novel approach for debris-flow detections using recorded ground vibrations generated by 22 debris flows in Illgraben, Switzerland. We use a machine-learning algorithm that automatically learns to distinguish between debris flow generated ground vibrations and other seismic signals. This allows us to increase warning times by at least 42 min comparing to existing detection systems at Illgraben.

1 Introduction

Debris flows are mixtures of water and sediments of all sizes, which are mobilized by heavy precipitation in steep Alpine torrents. They move downstream with average velocities of several meters per seconds (Hürlimann et al., 2003) showing a flow behaviour in-between landslides and sediment transporting floods (Iverson, 1997). Debris flows have a high destructive potential, which is amplified at the flow front, where large cobbles and boulders concentrate (Iverson, 1997). The significant hazard to human life and infrastructure in Alpine regions, including Switzerland (e.g., Jakob & Hungr, 2005; Badoux et al., 2016) demands reliable warning systems to reduce risk in exposed terrain (e.g., Stähli et al., 2015).

Recently, modern seismic instrumentation has suggested new warning perspectives, because even at large distances (tens to hundreds of kilometers) seismometers can detect high-frequency (>1 Hz) ground unrest induced by debris flows (for a review see Allstadt et al., 2018). This may extend warning times compared to conventional instrumentation within or near debris flow torrents, which can only be installed in accessible terrain (e.g., Arattano & Marchi, 2008; Coviello et al., 2019).

Despite recent advances in our theoretical understanding of high-frequency debris flow seismograms (Cole et al., 2009; Kean et al., 2012; Lai et al., 2018; Farin et al., 2019), seismometers installed at larger distances from torrents have yet to be implemented in operational warning systems. Identification of debris flow signals in the presence of other seismic activity remains a challenge. Since seismic debris flow signals have moderate amplitudes, simple threshold-based detection criteria cannot distinguish them from cultural noise, earthquakes and other Alpine mass movements at a permissible false detection rate (Walter et al., 2017).

Here, we introduce a machine learning approach to detect debris flows based on their seismic signature. For the Illgraben torrent, Switzerland, seismic records from an 8-station

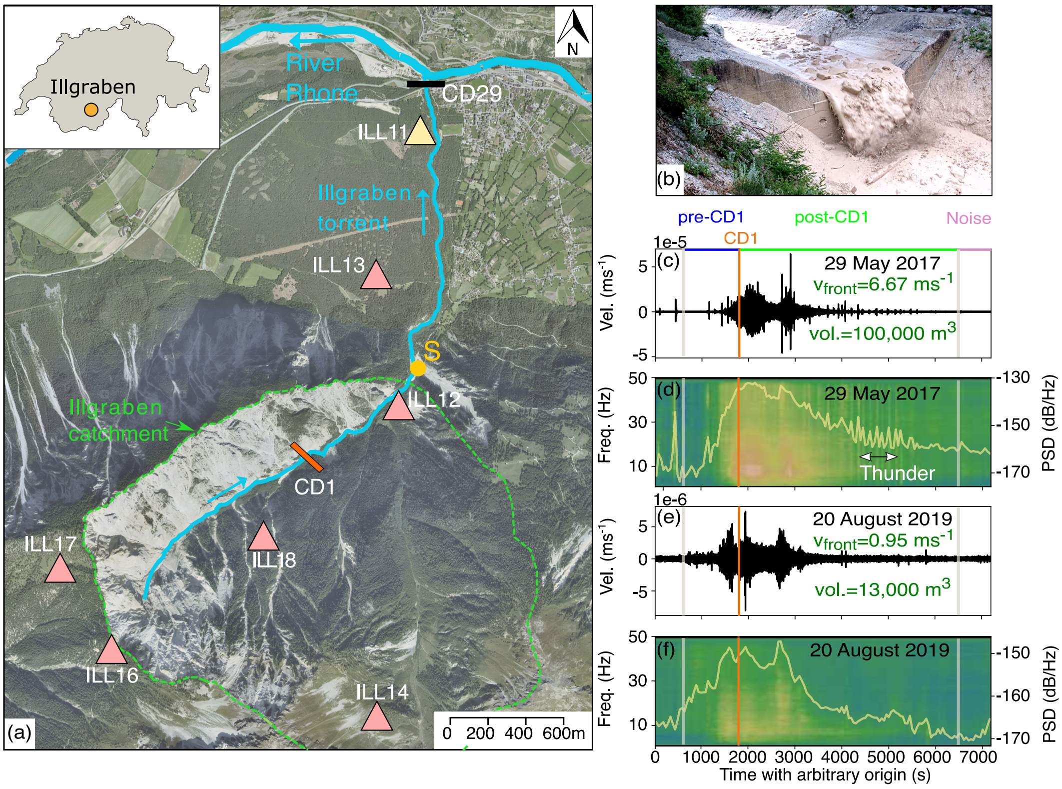

Figure 1. Study site. (a) Illgraben catchment is outlined with green dashed line (source: Swisstopo). Check dam (CD) 1 and CD29 are represented with orange and black bars. Seismometer locations are indicated with triangles. Connection of east side hillslope (Sagenschleif) with Illgraben channel is marked at Point S. Inset shows Illgraben's location in Switzerland. Station ILL11 not used in detection is marked in yellow and station ILL15 is located outside of the presented map. (b) Photo of Illgraben debris-flow passing CD29 (Source: WSL). (c) Vertical ground velocity recorded at ILL18 on 29 May 2017 (large and fast event) and the corresponding power spectral density (PSD) in (d). (e) Vertical ground velocity recorded at ILL18 on 20 August 2019 (small and slow event) and the corresponding PSD in (f). Arrival times of the debris flows at CD1 are marked with an orange line. In (d) and (f) PSDs averaged over 1-50 Hz are shown with a yellow line. Time windows between gray and orange lines divide the records in (c)-(f) into 3 signals classes.

network allow for debris flow detection in the upper catchment area, where instrument deployment is not possible. Trained with data from 20 events, our detection algorithm was subjected to real-time data streams from summer 2020 and identified all 13 debris flows with no false alarms. Our approach adds up to an hour of warning time to the earliest possible in-torrent detection at Illgraben.

2 Study site

Located in the southwestern part of Switzerland, Illgraben is one of the most active debris-flow torrents in the European Alps (Rickenmann et al., 2001) (Figure 1a). The catchment area extends over 10.4 km^2 from the summit of the Illhorn mountain [2716 m above sea level (asl)] to the Rhone River (610 m asl). The steep slopes (~40°) of the upper Illgraben catchment are characterized by rockfall and landslide activity (Berger et al., 2011b). The resulting sediments accumulate downslope or in the Illgraben channel and provide sliding material with volumes ranging from 500 to more than 4,000 m^3 (Schlunegger et al., 2009). During heavy precipitations and intense summer thunderstorms from April to October this material is regularly mobilized in form of debris flows (Badoux et al., 2009). The larger debris flow volumes (10^3 to $10^5\,m^3$) result from cumulative sediment mobilization and entrainment along the flow path and often reach the Rhone River. Like elsewhere, Illgraben debris flows have boulder-rich fronts resulting from particle sorting phenomena (Pierson, 1986; Johnson et al., 2012) followed by turbulent slurry with a large concentration of suspended sediments of variable granulometry and water content (Costa, 1984; Iverson, 1997; McCoy et al., 2010; Berger et al., 2011a, 2011b).

In 1961 a major landslide occurred in the upper Illgraben catchment and resulted in a debris flow destroying the bridge of the Cantonal highway along the Rhone river (Graf et al., 2007; Berger et al., 2011a). Consequently, a series of 30 Check Dams (CD; see Figure 1b for lowest CD29) was placed along the lower 3.4 km of the channel in order to stabilize the current debris flow path, expand discharge capacity, and minimize erosion.

As debris-flows still pose a hazard to people crossing the channel and to nearby infrastructure, an in-torrent warning system was commissioned by the governmental authorities and installed in 2007 (Badoux et al., 2009). The system consists of geophone detectors in check dams and flow depth measurements in the lower Illgraben part (Badoux et al., 2009; McArdell et al., 2007). Similar instruments and a recently re-deployed force plate form the debris flow observatory, which is operated for research purposes independently of the warning system since 2000 (Hürlimann et al., 2003; McArdell et al., 2007; Berger et al., 2011a). The observatory provides estimates of debris flow depth, volume and density (Schlunegger et al., 2009).

Illgraben's warning system has undergone different upgrades but is still based on the initial measurement principles. The earliest possible detection is provided by geophones installed inside check dam CD1 (Figure 1a), but this detection point is not deemed reliable as it is contingent upon solar power supply and mobile phone reception, which vary as a consequence of shadowing effects from the canyon walls. However, the CD1 arrival times can be downloaded in retrospect and are available for subsequent data analysis.

The present warning system in Illgraben requires instrument installation in direct contact with the torrent, which implies that detection is insensitive to sediment movement in the highly unstable and inaccessible upper catchment. This is a major weakness as debris flows mobilize in the upper catchment above CD1 (Schlunegger et al., 2009), where detection could increase warning times by tens of minutes.

2.1 Seismic debris-flow data

In past years, researchers have attempted to extend detection capabilities to the upper Illgraben catchment using seismological and acoustic measurements (Burtin et al., 2016; Walter et al., 2017; Schimmel et al., 2018; Marchetti et al., 2019; Wenner et al., 2019). In this context, since 2017, a seasonal seismometer network has been installed around the Illgraben catchment during spring/summer months (Figure 1a). The network consists of 8 3-component 1 Hz seismometers recording ground velocity continuously at a sampling frequency of 100 Hz.

All 8 stations are equipped with a modem to stream seismic data via the 4G mobile phone network to the Swiss Seismological Service from where they are made available via a seedlink server. A python client on a work station at the Institute of Hydraulics, Hydrology and Glaciology (VAW) at ETH Zürich receives data packages every 8-9 seconds, which are concatenated to continuous records. Disregarding station ILL18 (the only station within the Illgraben canyon), realtime data return reaches 95%, except during rare instrument malfunctioning events. Mobile phone reception at Ill18 was unstable and gappy in 2018 and 2019, but improved to 99%, during 2020 when the network provider was changed. Even when real-time data transfer is interrupted, data are locally recorded and transmitted when the mobile network connection is stable again.

Between 2017 and 2019, the seismic network recorded more than 22 debris flows. Figures 1c-f show vertical velocities seismograms and associated spectrograms of two debris flows. Figure 1c shows the largest recorded event (vol.= 100,000 m^3, velocity of the front v_{front}=6.67 m s^{-1}), and Figure 1e shows one of the smallest recorded events (vol.= 13,000 m^3, v_{front}=0.95 m s^{-1}). Both debris-flow signals show emergent onsets with dominant frequencies above 1 Hz reaching frequencies of 40-50 Hz. The signal emerges from the background noise at times that depend on the distance between the debris-flow front and the recording station (Walter et al., 2017). For the larger event (Figure 1c, d) we observe burst-like signals generated by thunder, not directly related to the debris-flow (Marchetti et al., 2019). Seismograms generated by all events recorded on eight stations are presented in Figures S1-S22.

3 Methods

We use the vertical-component seismograms of the 22 debris-flow events between 2017 and 2019 recorded on six stations (stations ILL12, ILL13, ILL14, ILL16, ILL17, ILL18 in Figure 1a) to train a machine learning model and test its detection capability. ILL15 and ILL11 were not used, because the former was deployed later in the season and the latter is located in the Rhone Valley recording strong anthropogenic noise signals. Debris-flow properties are shown in Table S1. The emergent nature of debris flow seismograms precludes use of standard event detectors and instead the spectral content of the continuous seismic signal is analyzed (Walter et al., 2017; Lai et al., 2018; Wenner et al., 2019, 2020). We split the debris flow seismograms into 100 s time windows with an overlap of 50 %. This window length is long enough to extract stable spectral characteristics and results in large enough set of training data. The overlap is chosen to further increase the number of samples in the data set.

3.1 Labeled data

We define labels for three seismic event classes:

1. *Pre-CD1*: debris flow signals before passage of CD1
2. *Post-CD1*: debris flow signals after passage of CD1
3. *Noise*: signals not associated with debris flows

Dividing the debris-flow signals into two classes caters to the need to detect debris flows in the upper catchment before CD1 passage. In the lower Illgraben part, the check dams interrupt the flow and possibly influence particle sorting which might change debris flow signals. For 20 events, the arrival times at CD1 are known from geophone installations within the check dam. For 2 events geophone detections were not available and instead we used estimates from amplitude source location (ASL), which traces the flow front location using time-varying amplitudes of the debris flow seismograms (Walter et al., 2017). The three different signal classes are indicated on Figures 1c-f and S23.

3.2 Catalog compilation and processing steps

The construction of our debris flow detector is a supervised machine-learning classification (Goodfellow et al., 2016), because we ask an algorithm to classify a signal of unknown origin based on a previously trained machine learning model. Training the model requires a labeled signal catalog with signals whose classes are known from independent observations. We compile such a labeled data set from debris flow seismograms defined by manually picked signal start and end times (Figure 1d, f; see TextS1 for details). Debris flow seismograms are defined to be those records lying between the earliest signal start and the latest signal end among all stations. Including all available stations, this yields 3,631 pre-CD1 time windows and 13,046 post-CD1 time windows. We randomly choose 550 100-second long noise time windows from 2017, 2018 and 2019 and several rainfall seismograms to compile the noise catalog. This provides 16,614 noise time windows.

We use a two-iteration training and testing approach: in the first iteration we confine ourselves to the 18 debris flows with the cleanest seismic signatures. From these we use all 100 second time windows from 15 events with both pre-CD1 and post-CD1 labels to train the model and test it on the seismograms from the remaining debris flows. We use 2/3 (11,076) randomly selected noise time windows for the training, and the rest (5,538) for the testing. In the second iteration we repeat this exercise with time windows from 20 debris flows for training and 2 debris flows for testing. We furthermore inject identified false positives (29,741 time windows) from the first iteration into the noise class. This increases the noise class to 46,355 time windows.

3.2.1 Detector implementation and performance

Rather than using the raw seismic signals, our classification algorithm operates on 70 statistical signal features. A feature is a scalar number, which describes waveform characteristics [e.g. root-mean-square amplitude (RMS), spectral content (e.g. mean and variance of the discrete Fourier transform), and signal variations throughout the network (e.g. ratio between maximum RMS and minimum RMS). The complete feature list is given in Table S2 in SI and Provost et al. (2017). 59 features are extracted for each station separately. Additional 11 network features are calculated based on all available stations.

We use the Random Forest (RF) supervised classifier (Breiman, 2001) as the machine learning algorithm, which comprises majority votes among an ensemble of randomized decision trees. The decision trees are formed by consecutive inequality operations, which determine if features of a signal are smaller or larger than predefined thresholds. These thresholds, the order and the number of the inequality operations are learned during the training phase, whereas hyperparameters (e.g. the maximum number of the inequality operations and the total number of decision trees) are determined as described in Text S1.

RF has proven useful in seismological applications (e.g., Rouet-Leduc et al., 2017, 2019) and mass movements detection (e.g., Hibert et al., 2017; Maggi et al., 2017; Provost et al., 2017). For our implementation we use Scikit-learn machine learning Python library (Pedregosa et al., 2011).

In the training phase the machine learning algorithm has access to the features and their associated labels (pre-CD1, post-CD1 and noise). Subsequently, the performance of the machine learning model is evaluated on testing set, which were not included in the training set. The true labels of the testing set are compared to the model predictions, which may or may not be correct (Figure 2).

The RF algorithm returns the feature importance which elucidates how the model reaches its predictions (Breiman, 2001). Figure 2a shows pairwise relations between the three most important features. In each subplot two features are plotted against each other and the univariate distributions of the same features are plotted on the diagonal with density plots. The three most important features are network features: 1. ratio between the maximum RMS and the minimum RMS in the network, 2. station number with maximum RMS, and 3. maximum coherence (normalized cross-correlation) between station pairs.

This shosw that: (1) The machine-learning model strongly relies on the relative RMS amplitudes throughout the network and the RMS amplitude ratio is the lowest for the pre-CD1 class. (2) Some noise time windows are highly correlated. (3) ILL18 has the largest

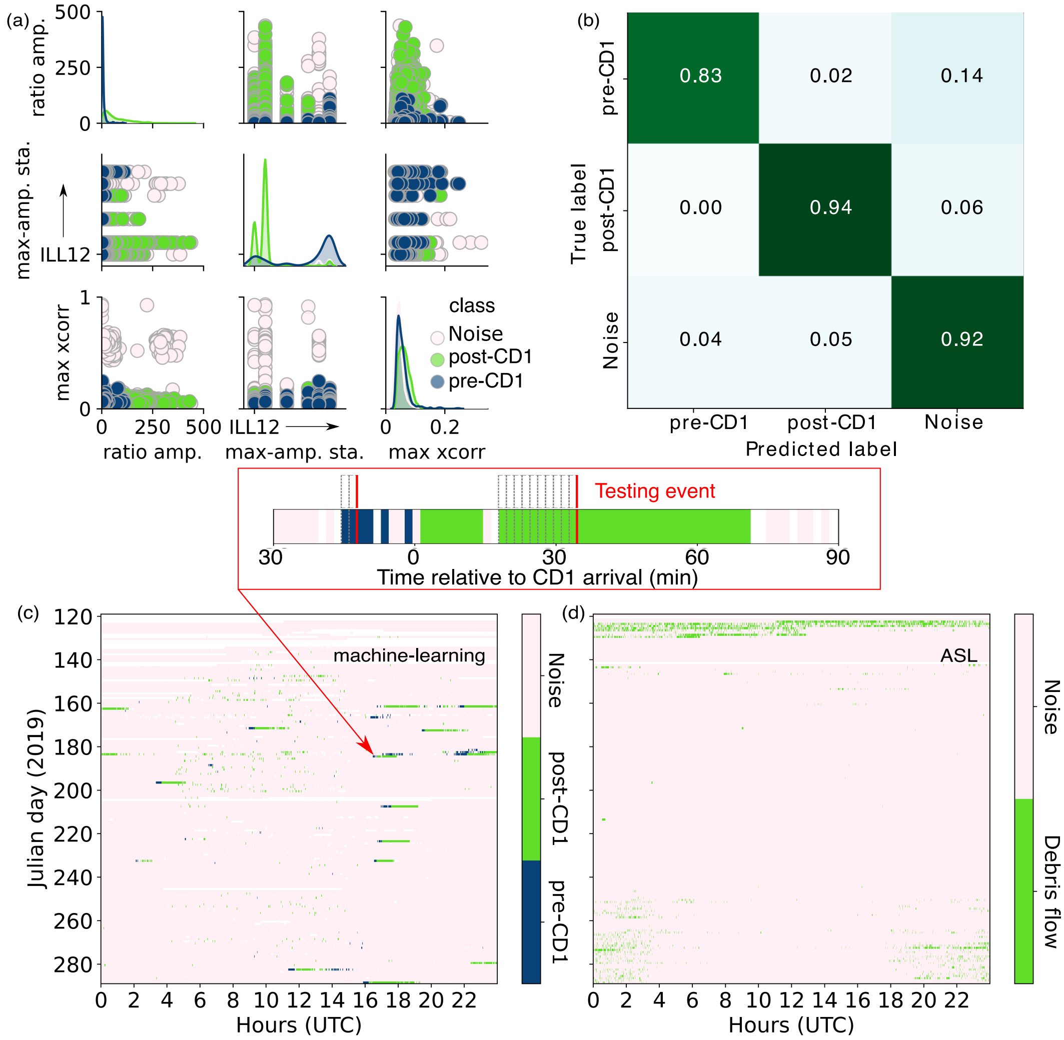

Figure 2. Machine-learning model evaluation (second iteration training). (a) Pairwise relations of the three most important features (see text for details). Features from each class are marked in different colors.(b) Normalized confusion matrix with true and predicted labels (columns and rows). (c) Results of the ML-based detector and (d) ASL-based detector applied to the 2019 continuous data. Inset in (c) shows a zoom on the testing debris flow, which was not part of the training set. Gray dashed lines denote individual detections in time windows and red line shows the alarm raised after a fixed number of subsequent detections.

RMS for the pre-CD1 class, while ILL12 and ILL13 show the largest RMS for the post-CD1 class.

We use a confusion matrix (Figure 2b) and Receiver Operating Characteristic (ROC) curve (Figure S25, in SI) to evaluate our model performance. The confusion matrix, also called an error matrix (Stehman, 1997), assesses classification performance in a table layout with true labels as columns and predicted labels as rows. For an ideal classifier all samples locate on the diagonal where the predicted label equals the true label and the diagonal values are normalized with 1.

We present within results of the second iteration, results of the first iteration are presented in Figures S24-S25. The confusion matrix on Figure 2b shows the highest misclassification for the pre-CD1 class with 14 % of pre-CD1 time windows classified as noise. However, we verified that ∼30 % of these "confused" time windows are the first three time windows of the pre-CD1 seismograms, and the normalized number of true positives increases to from 0.83 to 0.87 (pre-CD1—noise misclassification lowers to 0.11) if we remove these windows from the testing set. Whereas these initial samples are labeled as pre-CD1 they might still constitute noise for stations located further away from the torrent. Based on the scores on the confusion matrix diagonal we expect that our detector identifies debris flow signals at an accuracy near 90 %.

3.3 Detections and alarms

So far, we evaluated the performance of our machine learning model using the union of predictions from all stations. For an operational real-time alarm system we define a detector, which requires that more than half of the operational stations point towards the same class. If such a majority does not exist the detector does not make a prediction. Consequently, for real-time operation, we define "detection" and "alarm" as follows:

1. *Detection*: a single time window in which the majority vote over online stations predicts the pre-CD1 or post-CD1 class.
2. *Alarm*: > 2 subsequent detections for the pre-CD1 class, and > 10 subsequent detections for the post-CD1 class.

If no majority exists among online stations, the detector freezes the current alarm status and waits for the prediction from the next time window to update the alarm status. The alarm definition introduces a time lag between an initial debris flow detection (200 s for the pre-CD1 class and 16 min 40 s for the post-CD1 class, see the inset in Figure 2c for a visual representation). This time lag is small for the pre-CD1 class which is crucial for warning, and at the same time minimizes the number of false alarms. In future detector improvements, more advanced logical operations will likely reduce the time lag between initial detection and alarm, especially for the post-CD1 class.

4 Results

4.1 2019 continuous classification

Next, we run the detector over the 2019 archived data using 100 s time windows, this time without overlaps. In 2019 we monitored station up- and down times, which we now use to reproduce real-time station performance. The 2019 data contain 13 training events and one, which was part of the testing set.

Figure 2c shows the detector performance over 170 days in 2020. As expected, the debris flow detections (dark blue pixels for the pre-CD1 class, and green pixels for the post-CD1 class) are embedded in the noise windows (light pink). Debris flows consist of continuous detections, but isolated detection "pixels" illustrate numerous noise time windows falsely classified as debris flows.

We apply the alarm criterion to the 2019 detections and find that our debris flow detector raises a pre-CD1 alarm for 11 events, including the testing event, and misses only three small-volume events. For all 14 events a post-CD1 alarm was raised. 6 false positive alarms were raised (1 post-CD1 class, and 5 pre-CD1 class). For comparison, the ASL-based detector (e.g., Battaglia & Aki, 2003; Walter et al., 2017) catches 5 large debris flows but raises false alarms on 64 days; for some days (e.g.: Julian days 123-126, 270-280) it generates false alarms continuously (Figure 2d). Visual data inspection shows that ASL detection tends to fail when amplified noise signals resulting from electronic spikes or cultural activity are present on individual stations. The machine learning detector is less sensitive to these spurious signals.

We stress that for this 2019 debris-flow detection comparison, the machine learning approach is biased: The machine learning model learns 13 events in the training phase and only one event [marked with a red arrow (Figure 2c)] is independent from the training phase. This event happens to be missed by the ASL-detector. On the other hand, this test of the 2019 data demonstrates the drastic reduction of false alarms when moving from ASL to machine-learning based detection.

Finally, we test the machine learning detector on the 2017 and 2018 data (Figure S26), analogous to the 2019 test. The detector generates less than 3 false alarms per year and correctly raises pre-CD1 and post-CD1 alarms for the event not included in the training set (marked with red arrows in Figure S26). Moreover, the detector finds some previously unknown events (Figure S27) with either pre-CD1 or post-CD1 alarms. Based on signal strengths and characteristics, these alarms correspond to small debris flows, which did not trigger or reach the in-torrent detection system.

4.2 2020 continuous classification

The final realistic and rigorous test of our machine learning detector is the real-time classification of the 2020 data-stream. The 2020 seismic network was deployed at the end of May 2020 and the detector has been running continuously since 2 June 2020. In the first week of operation (3-9 June 2020) the detector correctly raised alarms for 5 debris flows triggered by high-intensity rainfalls [cumulative rainfall over one week=52.4 mm (Swiss Meteorological Service, Montana precipitation station)]. In total, during three months (3 June to 3 September 2020) the detector caught 13 debris flows and raised no false alarms. Figure 3b shows an example of detections and alarms, vertical records, and spectrograms during the initiation of the 29 June 2020 debris flow.

8 out of the 9 June and July 2020 alarms were independently confirmed by the WSL observatory, although the debris-flow observation station currently undergoes maintenance and does not provide CD1 arrival times. Nevertheless, we can compare our results with recordings from a video camera installed at the lowermost check dam CD29 near the Rhone River (Table S3). This comparison was not possible for events, which occurred at night or which stopped before reaching CD29. We also estimated arrival times at a nearby seismometer (ILL11) installed within a few meters from the torrent, which is not part of our detection system. Depending on their average flow velocities, most debris flows arrived at CD29 $\sim$ 1-2 h after our pre-CD alarm times (Figure 3c, Table S3). Given typical travel times between CD1 and CD29 of 20 minutes (Badoux et al., 2009; Walter et al., 2017), our system therefore provides additional warning time between 20 minutes and over 1.5 hours with respect to CD1.

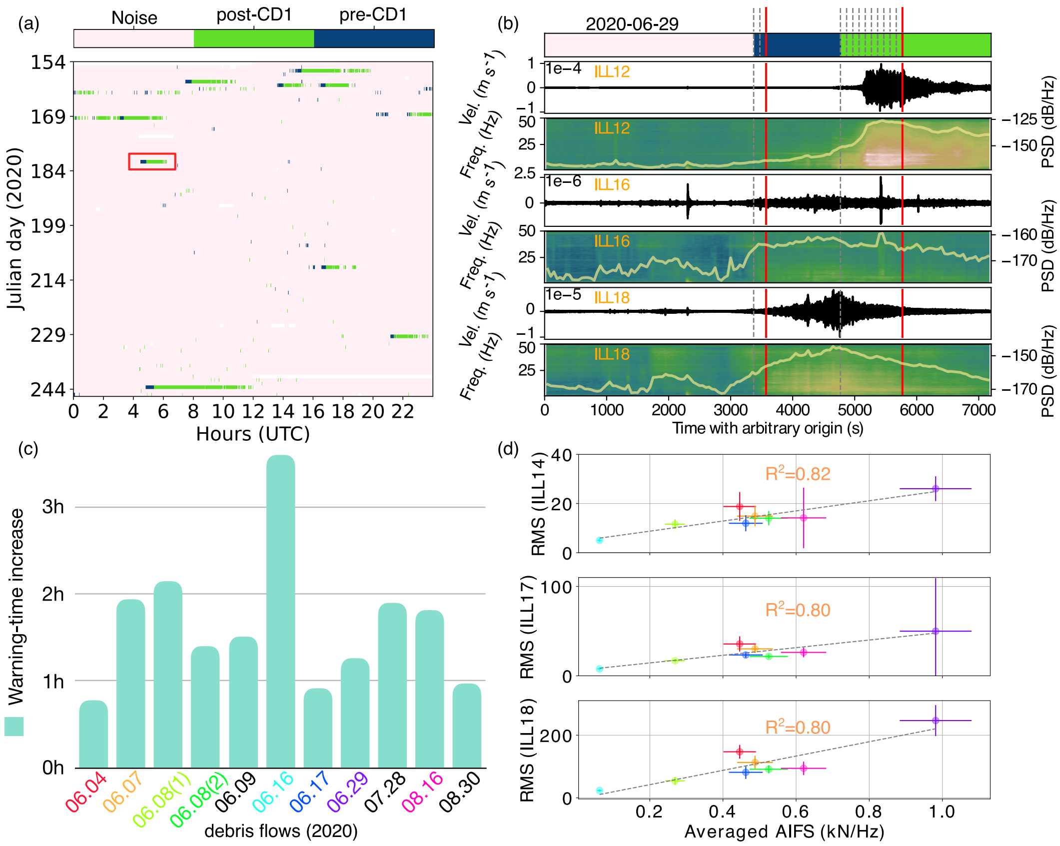

Figure 3. Debris flow detections in 2020. (a) Results of the ML-based detector run on the continuous real-time data stream from 2020. The 29 June 2020 event is marked with a red box. (b) Detections (gray dashed lines), alarms (red lines), vertical records of seismometers and spectrograms. The top horizontal color bar shows detection type (white: noise; blue: pre-CD1; green: post-CD1). (c) Warning time gain with respect to detection at CD29. (d) Relation between signal amplitudes near CD1 and averaged apparent impact force spectra (AIFS) calculated for the lowest Illgraben stretch (Zhang et al., 2020). The horizontal error bars are taken as 10 % from the averaged AIFS and the vertical error bars represent the standard deviation of RMS calculated over 10 consecutive post-CD1 detections. In Panels c and d events are indicated by the same color code.

5 Discussion and Conclusions

The central result of this study is that machine learning applied to real-time seismic data streams can detect debris flows in regions where conventional instrument deployment is not possible. This provides significant increases in warning times. All 8 events independently captured by the WSL debris flow observatory were detected with our approach. Several smaller debris flows, which did not reach in-torrent instrumentation but generated weak yet clear debris flow seismograms were also detected.

The performance in 2020 with no false positives or false negatives is encouraging but warrants modifications to the detector to automatically identify debris flows large enough to leave the upper catchment. This leads to the pivotal question whether our machine learning detector provides some quantitative measure of event size at the earliest alarm times, because this would allow warning against particularly destructive events. To this end we investigate if alarm-time seismic amplitudes scale with frequency-averaged apparent impact forces spectra (AIFS). The latter represent moment transfer of debris flow particles during ground impacts (Farin et al., 2019). We follow (Zhang et al., 2020) (see also Text S2) to calculate AIFS and their averages over the lowest Illgraben extent. These averages scale with boulder sizes accumulating at the flow front by the time it reaches the Rhone River near CD29 (Zhang et al., 2020) (Figure S28).

We do not find significant correlations between seismic amplitudes at the time of pre-CD1 alarms (R^2 varying between stations, from 0.01 to 0.38). However, for the earliest detection time window contributing to the post-CD1 alarms, there is a clear correlation between seismic amplitudes and AIFS (Figure 3d). Not all stations correlate equally, but ILL14, ILL17, and ILL18 have an R^2 of around 0.80. This shows that shortly after debris flow passage at CD1, seismic amplitudes can identify flow fronts with large boulder sizes, some 20 minutes before they arrive at CD29.

The poor correlation between seismic amplitudes during pre-CD1 alarms and AIFS raise questions about what seismogenic processes are detected at the very beginning of a debris flow. In general, initial sediment mobilization leading to debris flows may occur via pore water pressure increases or water drag forces leading to sediment failure on lateral slopes or within the torrent channel (Berti & Simoni, 2005; Godt & Coe, 2007; Gregoretti & Fontana, 2008). Our pre-CD1 detections identify time windows, when seismic amplitudes steadily increase (Figure 1d,f, 3b), rather than distinct bursts of seismic energy, which are observed in our records at other times (e.g. between 0 and 1000 s in Figure 1d). The steady increase in seismic energy argues for slow mobilization of debris flow material rather than sudden landslide failures on steep slope, which would be associated with burst-like signals.

The Illgraben site is an ideal natural laboratory to test debris flow detections, because regular event occurrence facilitates detector training. This is particularly important for machine learning algorithms relying exclusively on labeled training data. 22 training events used here can be considered a small training catalogue compared to most machine learning applications. Yet our practice to split signals into 100 second time windows increases the training data set by several orders of magnitudes to provide reliable detection. To transfer our Illgraben detector to other debris flow catchments, modifications are likely necessary to cope with fewer training events. We evaluated the accuracy of classification as a function of a number debris flow events used in a training set(Figure S29). The results show that machine-learning model trained even on a single event gives better results than a random guess, but a higher accuracy (>0.7) and stable predictions are obtained from 9 training events. Alternatively, it remains to be tested if the machine learning model trained at Illgraben could simply be applied in other geographic regions, i.e. if the model learned "general" characteristics of debris flow seismograms, which are independent of source-station distances and subsurface properties affecting seismic wave propagation. For machine learning algorithms applied to earthquake detection such detector transferability has already been confirmed (Ross et al., 2018).

Machine learning provides powerful tools for time series analysis and the approach presented here is only a first step to leverage this potential for natural hazard warning. Nevertheless, our relatively simple application already tackled the longstanding problem to reliably detect debris flows in an upper catchment area, which is inaccessible to existing detectors. Moreover, seismic data acquisition such as used here is a relatively cheap alternative to in-torrents instruments, which require major construction efforts. The combination of seismic monitoring and real-time data processing based on machine learning therefore provides significant advantages for Alpine mass movement detection, which have yet to be harnessed in operational warning schemes.

Acknowledgments

Seismometer installation was funded by WSL and the Canton Valais and supported by the Swiss Military. We thank John Clinton, Roman Racine, and Stefan Wiemer; the Swiss Seismological Service and its electronic laboratory (ELAB) for technical support. The data from the Illgraben network is collected under the network code XP (https://doi.org/10.12686/sed/networks/xp) and all seismic data will be openly available after a 2-year embargo (in 2022) via the archives in the Swiss Seismological Service, http://www.seismo.ethz.ch/en/research-and-teaching/products-software/waveform-data/ and the European Integrated Data Archive (EIDA), http://www.orfeus-eu.org/data/eida/. This work was funded by the Swiss National Science Foundation (SNSF) project Glacial Hazard Monitoring with Seismology (GlaHMSeis, grant PP00P2 157551) and Swisscom Broadcast AG. Obspy Python routines (www.obspy.org) were used to download waveforms and pre-process seismic data.

References

Allstadt, K. E., Matoza, R. S., Lockhart, A. B., Moran, S. C., Caplan-Auerbach, J., Haney, M. M., … Malone, S. D. (2018). Seismic and acoustic signatures of surficial mass movements at volcanoes. *Journal of Volcanology and Geothermal Research*, *364*, 76 - 106. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0377027317306261` doi: https://doi.org/10.1016/j.jvolgeores.2018.09.007

Arattano, M., & Marchi, L. (2008). Systems and sensors for debris-flow monitoring and warning. *Sensors*, *8*, 2436-2452.

Badoux, A., Andres, N., Techel, F., & C., H. (2016). Natural hazard fatalities in Switzerland from 1946 to 2015. *Nat. Hazards Earth Syst. Sci.*, *16*, 2747-2768. doi: https://doi.org/10.5194/nhess-16-2747-2016

Badoux, A., Graf, C., Rhyner, J., Kuntner, R., & McArdell, B. W. (2009). A debris-flow alarm system for the Alpine Illgraben catchment: Design and performance. *Nat. Hazards*, *49*(3), 1517?539. doi: doi:10.1007/s11069-008-9303-x

Battaglia, J., & Aki, K. (2003). Location of seismic events and eruptive fissures on the Piton de la Fournaise volcano using seismic amplitudes. *J. Geophys. Res.-Solid*, *108*. doi: 1029/2002JB002193

Berger, C., McArdell, B. W., & Schlunegger, F. (2011a). Direct measurement of channel erosion by debris flows, Illgraben, Switzerland. *Journal of Geophysical Research: Earth Surface*, *116*(F1). doi: 10.1029/2010JF001722

Berger, C., McArdell, B. W., & Schlunegger, F. (2011b). Sediment transfer patterns at the Illgraben catchment, Switzerland: Implications for the time scales of debris flow activities. *Geomorphology*, *125*(3), 421-432. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0169555X10004484` doi: https://doi.org/10.1016/j.geomorph.2010.10.019

Berti, M., & Simoni, A. (2005). Experimental evidences and numerical modelling of debris flow initiated by channel runoff. *Landslides*, *2*, 171–182.

Breiman, L. (2001). Random forests. *Mach. Learn.s*, *45*(1), 5-32. doi: 10.1023/A:1010933404324

Burtin, A., Hovius, N., & Turowski, J. (2016). Seismic monitoring of torrential and fluvial processes. *Earth Surf. Dynam.*, *4*, 285–307. doi: https://doi.org/10.5194/esurf-4-285-2016

Cole, S., Cronin, S., Sherburn, S., & Manville, V. (2009). Seismic signals of snow-slurry lahars in motion: 25 September 2007, Mt Ruapehu, New Zealand. *Geophys. Res. Lett.*, *36*. doi: https://doi.org/10.1029/2009GL038030

Costa, J. E. (1984). Physical geomorphology of debris flows. In J. E. Costa & P. J. Fleisher (Eds.), *Developments and applications of Geomorphology* (p. 268-317). Berlin: Springer.

Coviello, V., Arattano, M., Comiti, F., Macconi, P., & Marchi, L. (2019). Seismic characterization of debris flows: insights into energy radiation and implications for warning. *Journal of Geophysical Research: Earth Surface*, *124*, 1440-1463. doi: 10.1029/2018JF004683

Farin, M., Tsai, V., Lamb, M., & Allstadt, K. (2019). A physical model of the high-frequency seismic signal generated by debris flows. *Earth Surf. Process. Landforms*, *44*, 2529–2543. doi: https://doi.org/10.1002/esp.4677

Godt, J., & Coe, J. (2007). Alpine debris flows triggered by a 28 July 1999 thunderstorm in the central Front Range, colorado. *Geomorphology*, *84*, 80–97.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning.* MIT Press. (`http://www.deeplearningbook.org`)

Graf, C., Badoux, A., Dufour, F., Fritschi, B., McArdell, B., Rhyner, J., … Nigg, U. (2007). Alarmsystem für murgangfähige Wildbäche – Beispiel Illgraben. *Wasser Energie Luft*, *99*, 119–128.

Gregoretti, C., & Fontana, G. (2008). The triggering of debris flow due to channel-bed failure in some alpine headwater basins of the Dolomites: Analyses of critical runoff. *Hydrol. Process.*, *22*, 2248–2263.

Hibert, C., Provost, F., Malet, J.-P., Maggi, A., Stumpf, A., & Ferrazzini, V. (2017). Automatic identication of rockfalls and volcano-tectonic earthquakes at the Piton de la Fournaise volcano using a random forest algorithm. *Journal of Volcanology and Geothermal Research*, *340*, 130-142.

Hürlimann, M., Rickenmann, D., & Graf, C. (2003). Field and monitoring data of debris-flow events in the Swiss Alps. *Can. Geotech. J.*, *40*(1), 161-175. doi: doi:10.1139/t02-087

Iverson, R. (1997). The physics of debris flows. *Rev. Geophys*, *35*(3), 245-296. doi: 10.1029/97RG00426

Jakob, M., & Hungr, O. (2005). Introduction. In M. Jakob & O. Hungr (Eds.), *Debris-flow hazards and related phenomena* (p. 1-8). Berlin: Springer. doi: doi:10.1007/3-540-27129-5_7

Johnson, C. G., Kokelaar, B. P., Iverson, R. M., Logan, M., LaHusen, R. G., & T., G. J. M. N. (2012). Grain-size segregation and levee formation in geophysical mass flows. *Journal of Geophysical Research*, *117*(5A), F01032. doi: 10.1029/2011JF002185

Kean, J., Staley, D., Leeper, R., Schmidt, K., & Gartner, J. (2012). A low-cost method to measure the timing of postfire flash floods and debris flows relative to rainfall. *Water Resour. Res.*, *48*. doi: https://doi.org/10.1029/2011WR011460

Lai, V. H., Tsai, V. C., Lamb, M. P., Ulizio, T. P., & Beer, A. R. (2018). The Seismic Signature of Debris Flows: Flow Mechanics and Early Warning at Montecito, California. *Geophysical Research Letters*, *45*(11), 5528-5535. doi: 10.1029/2018GL077683

Maggi, A., Ferrazzini, V., Hibert, C., Beauducel, F., Boissier, P., & Amemoutou, A. (2017). Implementation of a multistation approach for automated event classication at piton de la fournaise volcano. *Seismological Research Letters*, *88*(3), 878-891.

Marchetti, E., Walter, F., Barfucci, G., Genco, R., Wenner, M., Ripepe, M., … Price, C. (2019). Infrasound array analysis of debris flow activity and implication for early warning. *Journal of Geophysical Research: Earth Surface*, *124*(2), 567-587. doi: 10.1029/2018JF004785

McArdell, B. W., Bartelt, P., & Kowalski, J. (2007). Field observations of basal forces and fluid pore pressure in a debris flow. *Geophys. Res. Lett.*, *34*, L07406. doi: doi:10.1029/2006GL029183

McCoy, S., Kean, J. W., Coe, J. A., Staley, D. M., Wasklewicz, T. A., & Tucker, G. E. (2010). Evolution of a natural debris flow: In situ measurements of flow dynamics, video imagery, and terrestrial laser scanning. *Geology*, *38*, 735?738. doi: doi:10.1130/G30928.1

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pierson, T. C. (1986). Flow behavior of channelized debris flows, Mount St. Helens, Washington. In A. D. Abrahams (Ed.), *Hillslope processes* (p. 269-296). Boston: Allen and Unwin.

Provost, F., Hibert, C., & Malet, J.-P. (2017). Automatic classication of endogenous landslide seismicity using the Random Forest supervised classier. *Geophysical Research Letters*, *44*(1), 113. doi: 10.1002/2016GL070709

Rickenmann, D., Hürlimann, M., Graf, C., Näf, D., & Weber, D. (2001). Murgang-Beobachtungsstationen in der Schweiz. *Wasser, Energie, Luft*, *93*, 1-8.

Ross, Z. E., Meier, M.-A., Hauksson, E., & Heaton, T. H. (2018). Generalized seismic phase detection with deep learning. *Bulletin of the Seismological Society of America*, *44*(5A), 2894?2901. doi: doi.org/10.1785/0120180080

Rouet-Leduc, B., Hulbert, C., Lubbers, N., Barros, K., Humphreys, C. J., & Johnson, P. A. (2017). Machine learning predicts laboratory earthquakes. *Geophysical Research Letters*, *44*(18), 9276-9282. doi: 10.1002/2017GL074677

Rouet-Leduc, B., Hulbert, C., Lubbers, N., Barros, K., Humphreys, C. J., & Johnson, P. A. (2019). Continuous chatter of the cascadia subduction zone revealed by machine learning. *Nature Geosci.*, *12*, 75-79. doi: 10.1038/s41561-018-0274-6

Schimmel, A., Húbl, J., McArdell, B. W., & Walter, F. (2018). Automatic Identification of Alpine Mass Movements by a Combination of Seismic and Infrasound Sensors. *Sensors*, *18*(5), 1658.

Schlunegger, F., Badoux, A., McArdell, B. W., Gwerder, C., Schnydrig, D., Rieke-Zapp, D., & Molnar, P. (2009). Limits of sediment transfer in an alpine debris-flow catchment, Illgraben, Switzerland. *Quat. Sci. Rev.*, *28*, 1097-1105. doi: doi:10.1016/j.quascirev.2008.10.025

Stähli, M., Sättele, M., Huggel, C., McArdell, B., Lehmann, P., Van Herwijnen, A., ... Springman, S. (2015). Monitoring and prediction in early warning systems for rapid mass movements. *Natural Hazards and Earth System Science*, *15*, 905-917.

Stehman, S. V. (1997). Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment*, *62*(1), 77 - 89. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0034425797000837` doi: https://doi.org/10.1016/S0034-4257(97)00083-7

Walter, F., Burtin, A., McArdell, B. W., Hovius, N., Weder, B., & M, T. J. (2017). Testing seismic amplitude source location for fast debris-flow detection at Illgraben, Switzerland. *Natural Hazards and Earth System Sciences*, *17*(6), 939-955. doi: https://doi.org/10.5194/nhess-17-939-2017

Wenner, M., Hibert, C., & Walter, F. (2020). Automatic near real-time classification of seismic signals in for slope failure detection in alpine environments. *submitted to Nat. Hazards Earth Syst. Sci.*.

Wenner, M., Walter, F., McArdell, B., & Farinotti, D. (2019). Deciphering debris-flow seismograms at Illgraben, Switzerland. In K. J. W., C. J. A., S. P. M., & G. B. K. (Eds.), *Association of environmental and engineering geologists special publication: Vol. 28. debris-flow hazards mitigation: mechanics, monitoring, modeling, and assessment* (p. 222-229). Colorado School of Mines: Association of Environmental and Engineering Geologists.

Zhang, Z., Walter, F., McArdell, B. W., Wenner, M., Chmiel, M., & He, S. (2020). Extracting dynamics of debris flows from their seismic signature. *GRL*, *in submission*.