

Geophysical Research Letters

RESEARCH LETTER

10.1029/2020GL090874

Key Points:

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

Supporting Information:

- Supporting Information S1
- Table S1

Correspondence to:

M. Chmiel,
chmiel@ee.ethz.ch

Citation:

Chmiel, M., Walter, F., Wenner, M., Zhang, Z., Mc Ardell, B. W., & Hibert, C. (2021). Machine learning improves debris flow warning. *Geophysical Research Letters*, 48, e2020GL090874. <https://doi.org/10.1029/2020GL090874>

Received 17 SEP 2020

Accepted 16 DEC 2020

Machine Learning Improves Debris Flow Warning

Małgorzata Chmiel¹ , Fabian Walter^{1,2}, Michaela Wenner^{1,2} , Zhen Zhang^{1,3,4}, Brian W. Mc Ardell², and Clement Hibert⁵ 

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

Abstract Automatic identification of debris flow signals in continuous seismic records remains a challenge. To tackle this problem, we use machine learning, which can be applied to continuous real-time data. 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 in the torrent, our approach provides a significant gain in warning times of tens of minutes to hours. For real-time data from 2020, our detector raises alarms for all 13 independently confirmed Illgraben events, giving no false alarms. We suggest that our seismic machine-learning detector is a critical step toward the next generation of debris-flow warning, which increases warning times using simpler 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, property, and infrastructure. It is crucial to detect debris flows early enough to send an alarm message to local communities. We propose a novel approach for debris-flow detection using recorded ground vibrations generated by 22 debris flows at the Illgraben torrent, Switzerland. We use a machine-learning algorithm that learns to distinguish between debris flow generated ground vibrations and other seismic signals. At the Illgraben torrent, this increases warning times by at least 20 min compared to existing detection systems.

1. Introduction

Debris flows are complex mixtures of water, fragmented rock and sediments of all sizes, which are mobilized by heavy precipitation in steep torrents. They move downstream with average velocities of several meters per seconds showing an unsteady and nonuniform flow behavior (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., Badoux et al., 2016; Jakob & Hungr, 2005) demands reliable warning systems to reduce risk in exposed terrain (e.g., Hürlimann et al., 2019; Stähli et al., 2015).

Recently, modern seismic instrumentation has suggested new warning perspectives. 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 torrents, which can only be installed in accessible terrain (e.g., Arattano & Marchi, 2008; Covello et al., 2019).

Despite recent advances in our theoretical understanding of high-frequency debris flow seismograms (Cole et al., 2009; Farin et al., 2019; Kean et al., 2015; Lai et al., 2018), 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 debris-flow seismograms have moderate amplitudes, simple threshold-based detection criteria cannot distinguish them from cultural noise, earthquakes and other mass movements at an acceptable false detection rate (Walter et al., 2017).

© 2020. The Authors.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

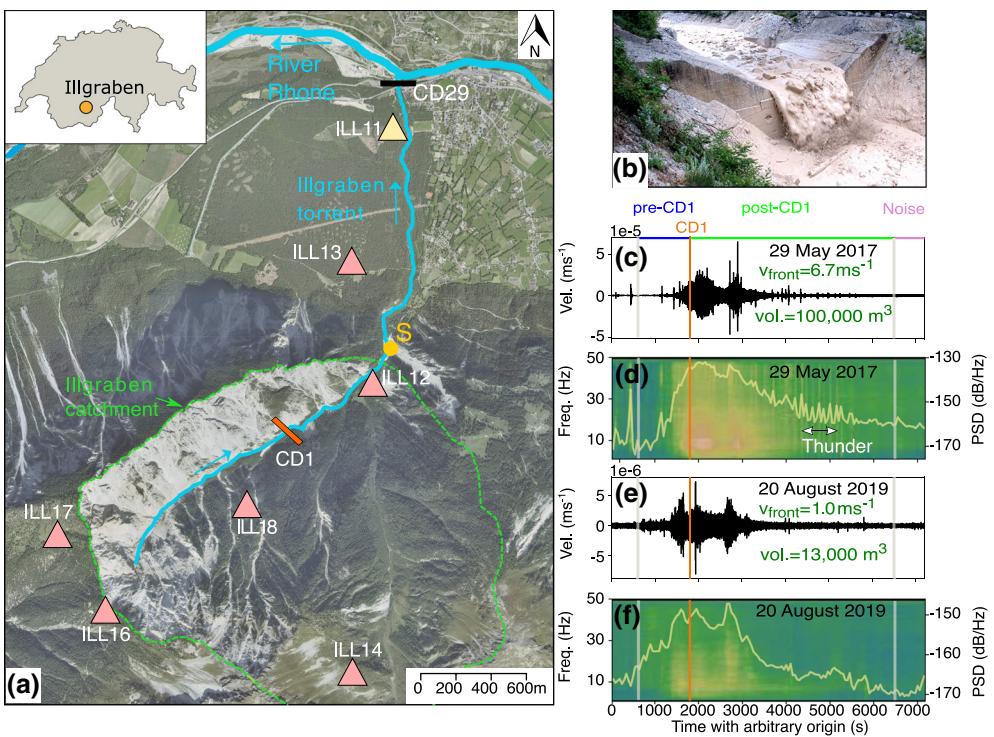


Figure 1. Study site. (a) Illgraben catchment is outlined with green dashed line (source: Swisstopo). Check dam (CD) 1 and CD29 are represented with orange and black bars. Triangles represent seismometer locations. Connection of east side hillslope (Sagenschleif) with Illgraben channel is marked at Point 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 map. (b) Photo of Illgraben debris flow passing CD29 (Source: WSL). (c) Vertical ground velocity recorded at ILL18 on May 29, 2017 (large, fast event) and the corresponding power spectral density (PSD) in (d). (e) Vertical ground velocity recorded at ILL18 on August 20, 2019 (small, 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 records in (c)–(f) into three signals classes.

Here, we introduce a machine-learning approach to detect debris flows and hazardous debris floods (e.g., Badoux et al., 2009) based on their seismic signature. For the Illgraben torrent, Switzerland, seismic records from an 8-station network allow for debris flow detection in the upper catchment area, where instrument deployment is not possible. Trained with 20 events, our detection algorithm was subjected to real-time data from 3 months in 2020. With no false alarms, we detected all 13 hazardous torrential events which include mostly debris flows, but also smaller floods with high water content. Our approach adds up to an additional hour of warning time to the earliest possible in-torrent detection.

2. Study Site

Located in the Southwest 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 km² from the summit of the Inkhorn mountain [2,716 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 providing sliding material with volumes ranging from 500 to more than 4,000 m³ (Schlunegger et al., 2009). During heavy precipitations and intense thunderstorms from April to October this material is mobilized in the form of debris flows and debris floods (hazardous sediment-transporting floods, Badoux et al. (2009)). The larger debris-flow volumes (10³–10⁵ m³) result from cumulative sediment mobilization and entrainment along the flow path and often reach the Rhone River. Like elsewhere, Illgraben debris flows often have boulder-rich fronts resulting from particle sorting phenomena (Johnson et al., 2012; Pierson & Mount, 1986) followed

by turbulent slurry with a high concentration of suspended sediments of variable granulometry and water content (Berger et al., 2011a, 2011b; Costa, 1984; Iverson, 1997; McCoy et al., 2010).

In 1961, a major rock avalanche occurred in the upper Illgraben catchment and resulted in a debris flow destroying the bridge of the Cantonal highway along the Rhone river (Berger et al., 2011a; Graf et al., 2007). Consequently, a series of 30 Check Dams (CD; see Figure 1b for a lower CD29) was placed along the lower 3.4 km of the channel in order to stabilize the debris flow path, reduce debris-flow discharge, 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; Mc Ardell et al., 2007). Similar instruments and a recently re-deployed force plate form the debris flow observatory of the Swiss Federal Institute for Forest, Snow and Landscape Research WSL, which is operated for research purposes independently of the warning system since 2000 (Berger et al., 2011a; Hürlimann et al., 2003; Mc Ardell et al., 2007). The observatory provides estimates of debris flow depth, volume, and density (Hürlimann et al., 2019; Schlunegger et al., 2009).

The present warning system and observatory at Illgraben require instrument installation in direct contact with the torrent and are insensitive to sediment movement in the unstable and inaccessible upper catchment. This is a major weakness as debris flows form 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; Marchetti et al., 2019; Schimmel et al., 2018; Walter et al., 2017; Wenner et al., 2019). Since 2017, a seasonal seismometer network has been installed around the 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 (for details on continuous data transmission see SI, Text S1).

Between 2017 and 2019, the seismic network recorded more than 22 events, which we classified as debris flows. Figure 1c–1f show vertical velocity seismograms and spectrograms of two debris flows. Figure 1c shows the largest recorded event (vol. = 100,000 m³, velocity of the front $v_{\text{front}} = 6.7 \text{ m s}^{-1}$), and Figure 1e shows one of the smallest recorded events (vol. = 13,000 m³, $v_{\text{front}} = 1 \text{ m s}^{-1}$). 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 flow front and the recording station (Walter et al., 2017). For the larger event (Figures 1c and 1d), we observe burst-like thunder signals, 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 and can be compared with rainfall and anthropogenic noise (Figures s23–s25).

3. Methods

We use the vertical-component seismograms of the 22 debris-flows between 2017 and 2019 recorded on 6 stations (stations ILL12, ILL13, ILL14, ILL16, ILL17, and 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. We split 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 provides large enough sets of training data.

3.1. Labeled Data

We define 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 of detection in the upper catchment before CD1 passage. We expect that seismic data distinguishes between pre-CD1 and post-CD1 signals, because seismic amplitudes and dominant frequencies depend on distances between flow front and recording stations (Walter et al., 2017; Wenner et al., 2019). Moreover, in the lower Illgraben part, the check dams cause several meters of free fall amplifying seismic signal generation as boulders pass across them. For 20 events, the arrival times at CD1 are known from geophones within the check dam. For two 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 Figure 1c–1f and S25 and signal partitioning into 100 s time windows is schematically presented in Figure S26.

3.2. Catalog Compilation and Processing

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 (Figures 1d and 1f; see Text S1 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-s 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: we first confine ourselves to the 18 debris flows with the cleanest seismic signatures. From these, we use all 100 s 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) of the 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 two debris flows for testing. We furthermore inject 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 raw seismic signals, our 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 and in Provost et al. (2017). Fifty-nine features are extracted for each station separately. Eleven network features are calculated based on all available stations (Figure S26b).

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 (Figure S27). The decision trees are formed by consecutive inequality operations, which determine if features 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; Wenner et al., 2020). 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

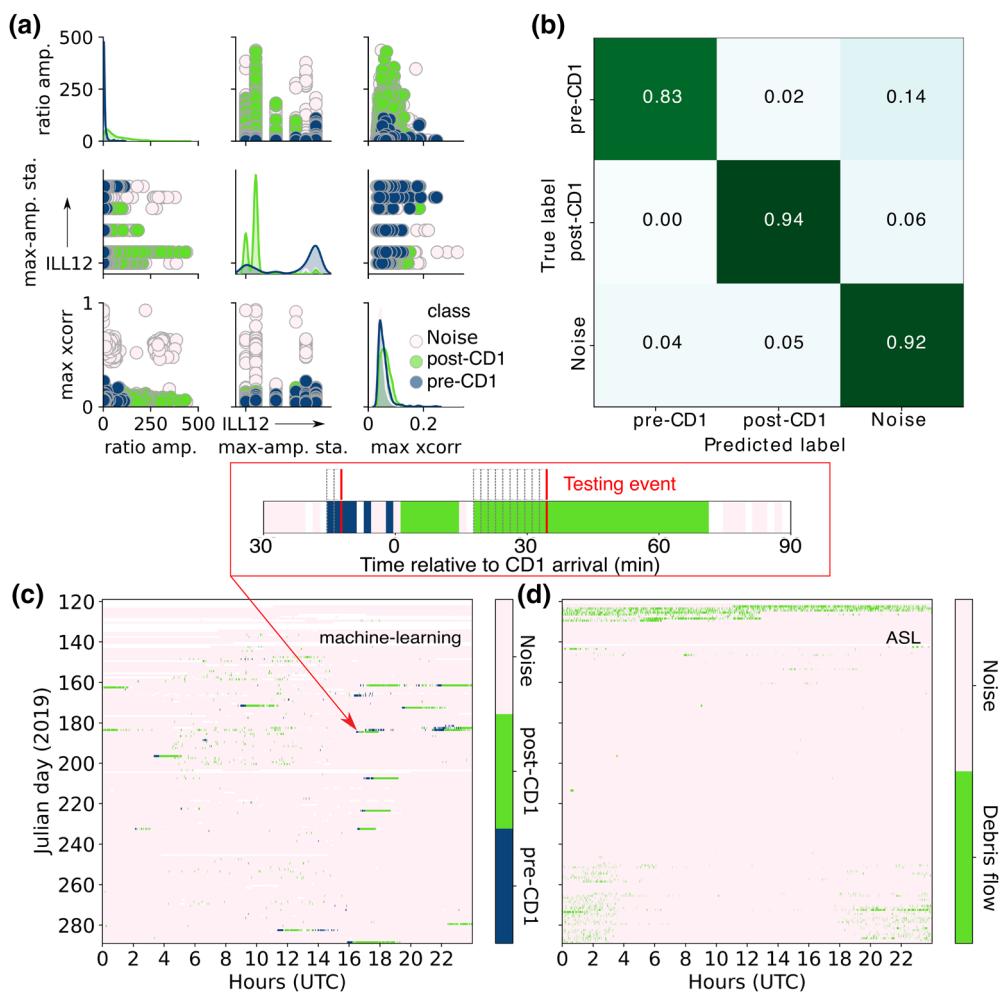


Figure 2. Machine-learning model evaluation (second iteration training). (a) Pairwise relations of the three most important features (see text for details). Features from each class are marked in different colors. (b) Normalized confusion matrix with true and predicted labels (columns and rows). (c) 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 and red line shows the alarm raised after a fixed number of subsequent detections.

on the testing data, which were not included in the training. The true labels of the testing data 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 shows 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 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 S30, in SI) to evaluate our model performance. The confusion matrix, also called 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 to 1.

In the main manuscript we present results of the second iteration, results of the first iteration are presented in SI, Figures S29–30. The confusion matrix in Figure 2b shows the highest misclassification for the pre-CD1 class with 14% of pre-CD1 time windows classified as noise. However, ~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 time windows from the testing set. Whereas these initial samples are labeled as pre-CD1 they may constitute noise for stations located further away from the torrent. Based 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 model performance 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 toward 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.

4. Results

4.1. 2019 Continuous Classification

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 2019. 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 numerous isolated false detections exist.

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. Six false positive alarms were raised (1 post-CD1 and 5 pre-CD1). 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 noise signals resulting from electronic spikes or cultural activity are present. 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 S31), analogous to the 2019 test. The detector generates less than three 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 S31). Moreover, the detector finds previously unknown events (Figure S32) 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 was running continuously between 2 June 2020 and end of September. 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 a 3 months test phase (3 June to 3 September 2020) the detector caught 13 events. All these alarms were confirmed with the WSL debris flow observatory (8 out of the 9 June and July alarms) and/or were visually verified on the seismic records. WSL staff are in the process of calculating event volumes, density and flow stage, but preliminary results indicate that at least one small event was a flood rather than a debris flow (Table S3). The set of 2020 events may thus more generally be referred to as "hazardous flow events" rather than debris flows. Nevertheless, all of our alarms are related to those events and thus our algorithm raised no false alarms. While several additional hazardous flow events occurred before and after the 3 months test phase, we are not aware of events, which did not trigger our alarm. Figure 3b shows an example of detections and alarms, vertical records, and spectrograms during the initiation of the June 29, 2020 debris flow.

We 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 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 events arrived at CD29 \sim 1–2 h after our preCD alarm times (Figures 3c and Table S3). Given typical travel times between CD1 and CD29 of 20 min (Badoux et al., 2009; Walter et al., 2017), our system therefore provides additional warning time between 20 min and over 1.5 h with respect to CD1.

5. Discussion and Conclusions

The central result of this study is that machine learning applied to real-time seismic data can detect debris flows in regions where conventional instrument deployment is not possible. This provides significant increases in warning times while the cost of installing a few seismometers is likely lower than building in-torrent structures such as check-dams for instrumentation. The majority of the detected hazardous flow events were independently captured by the WSL debris flow observatory and no known events were missed with our approach. Several smaller events, which did not reach in-torrent instrumentation but generated weak yet clear debris flow seismograms, were also detected.

The detector performance in 2020 is encouraging but warrants modifications 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 to warn 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 S33).

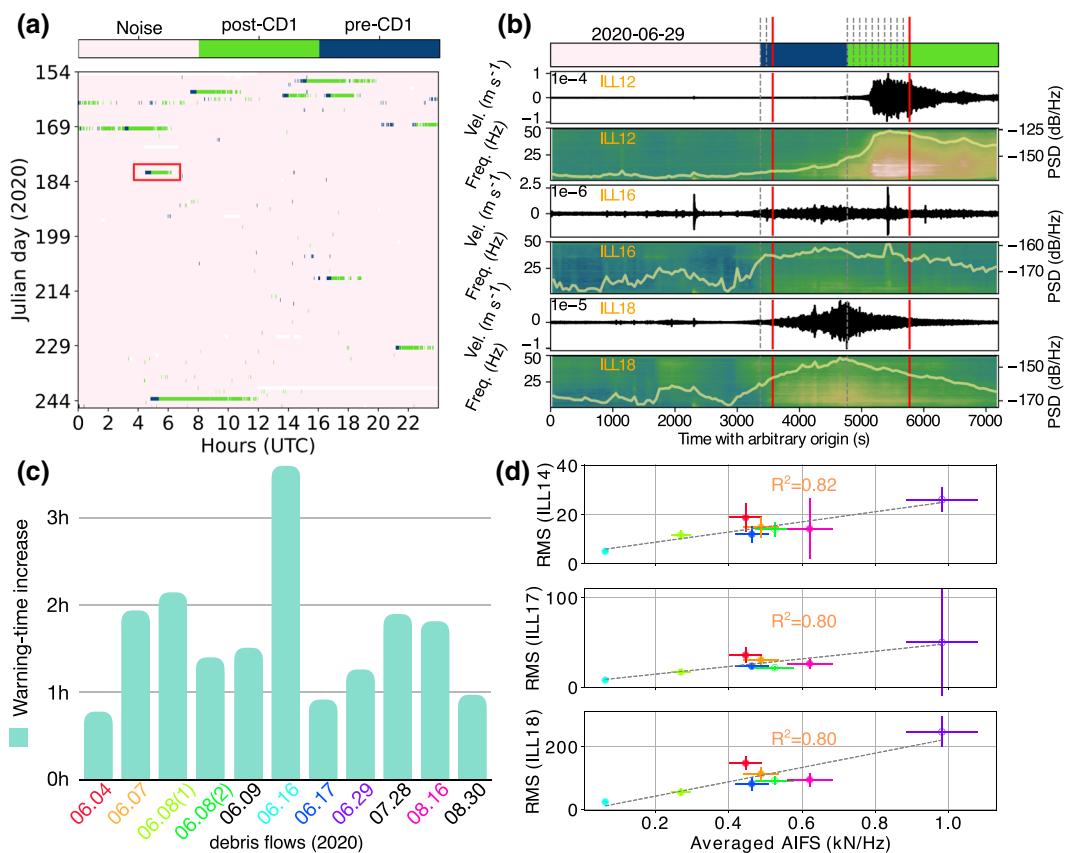


Figure 3. Debris flow detections in 2020. (a) Results of the machine-learning detector run on the continuous real-time data from 2020. The June 29, 2020 event is marked with a red box. (b) Detections (gray dashed lines), alarms (red lines), vertical records of seismometers, and spectrograms. The top 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. For events with black font in Panel (c) no AIFS were calculated for reasons given in Table S3.

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 min 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 (Figures 1d and 1f and 3b), rather than distinct bursts of seismic energy, which are observed in our records at other times (e.g. between 0 and 1,000 s in Figure 1d). The steady increase in seismic energy suggests slow mobilization of debris flow material rather than sudden landslide failures on steep slopes associated with burst-like signals. This slow sediment mobilization seems to be common to all events, which is an encouraging finding in view of operational early warning. In fact, even for the small flood event on June 9, 2020, pre-CD1 detections were made (Table S3). Apparently, sediments were mobilized although

they subsequently came to rest before reaching CD29. This, in turn, underlines our detector's sensitivity and flexibility with respect to event type, which during our three-months test phase varied between water-rich and sediment-rich flows (Figure S33).

The Illgraben site is an ideal natural laboratory to test debris flow detections, because the regular occurrence of variable events 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 catalog compared to most machine learning applications. Yet our practice to split signals into 100 s time windows increases the training data set by several orders of magnitudes to provide reliable detection. To transfer our detector to other catchments, modifications are likely necessary to cope with fewer training events. We evaluated the accuracy of classification as a function of number of debris-flow events used in a training set (Figure S34). The results show that a 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 nine training events.

In the future, the proposed algorithm and possible improvements should be investigated with data from other catchments with limited number of events and observations. First, a machine learning model trained at Illgraben could simply be applied to another geographic region to check 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 (Mousavi et al., 2020). Second, the Random Forest could be replaced with other algorithms to be used in transfer learning, such as Deep Neural Networks (e.g., Yosinski et al., 2014). In analogy to image recognition, typical debris flow characteristics could be learnt by the first layers, whereas site-specific attributes of the seismic signals may be learnt by lower layers. In that case only the lower layers would need to be retrained with limited seismic data from new catchments. Finally, we could test an anomaly detection, where the background state, which includes typical noise signals (anthropogenic and electronic noise, earthquakes) could be learned and debris flow signals would be recognized as an anomaly with respect to this background state (e.g., Ravanelli & Bengio, 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. 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.

Data Availability Statement

The data from the Illgraben network are 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 preprocess seismic data.

Acknowledgments

Seismometer installation was funded by WSL and the Canton Valais and supported by the Swiss Military. The authors thank Christoph Graf for explanations on Illgraben debris flows and John Clinton, Roman Racine, Stefan Wiemer, the Swiss Seismological Service and its electronic laboratory (ELAB) for technical support. The authors also thank the Associate Editor and the two anonymous reviewers for their useful comments and suggestions that allowed us to improve the manuscript.

References

- Allstadt, K. E., Matoza, R. S., Lockhart, A. B., Moran, S. C., Caplan-Auerbach, J., Haney, M. M., et al. (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., (2016). Natural hazard fatalities in Switzerland from 1946 to 2015. *Natural Hazards and Earth System Sciences*, 16, 2747–2768. <https://doi.org/10.5194/nhess-16-2747-2016>

Badoux, A., Graf, C., Rhyner, J., Kuntner, R., & Mc Ardell, B. W. (2009). A debris-flow alarm system for the Alpine Illgraben catchment: Design and performance. *Natural Hazards*, 49(3), 1517–1539. <https://doi.org/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. *Journal of Geophysical Research*, 108(B8), 2364. <https://doi.org/10.1029/2002JB002193>

Berger, C., Mc Ardell, B. W., & Schlunegger, F. (2011a). Direct measurement of channel erosion by debris flows, Illgraben, Switzerland. *Journal of Geophysical Research*, 116(F1), F01002. <https://doi.org/10.1029/2010JF001722>

Berger, C., Mc Ardell, 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 modeling of debris flow initiated by channel runoff. *Landslides*, 2, 171182.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>

Burton, A., Hovius, N., & Turowski, J. (2016). Seismic monitoring of torrential and fluvial processes. *Earth Surface Dynamics*, 4, 285307. <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. *Geophysical Research Letters*, 36, L09405. <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* (pp. 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. <https://doi.org/10.1029/2018JF004683>

Farin, M., Tsai, V., Lamb, M., & Allstadt, K. (2019). A physical model of the highfrequency seismic signal generated by debris flows. *Earth Surface Processes and Landforms*, 44, 25292543. <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, 8097.

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

Graf, C., Badoux, A., Dufour, F., Fritsch, B., Mc Ardell, B., Rhyner, J., et al. (2007). Alarm system fr murgangfhige Wildbche-Beispiel Illgraben. *Wasser, Energie, Luft*, 99, 119128

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. *Hydrological Processes*, 22, 22482263

Hibert, C., Provost, F., Malet, J.-P., Maggi, A., Stumpf, A., & Ferrazzini, V. (2017). Automatic identification 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., Coviello, V., Bel, C., Guo, X., Berti, M., Graf, C., et al. (2019). Debris-flow monitoring and warning: Review and examples. *Earth-Science Reviews*, 199, 102981. <https://doi.org/10.1016/j.earscirev.2019.102981>

Hürlimann, M., Rickenmann, D., & Graf, C. (2003). Field and monitoring data of debris-flow events in the Swiss Alps. *Canadian Geotechnical Journal*, 40(1), 161–175. <https://doi.org/10.1139/t02-087>

Iverson, R. (1997). The physics of debris flows. *Reviews of Geophysics*, 35(3), 245–296. <https://doi.org/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. https://doi.org/10.1007/3-540-27129-5_7

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

Kean, J. W., Coe, J. A., Coviello, V., Smith, J. B., McCoy, S. W., & Arattano, M. (2015). Estimating rates of debris flow entrainment from ground vibrations. *Geophys. Res. Lett.* 42, 6365–6372. <https://doi.org/10.1002/2015GL064811>

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. <https://doi.org/10.1029/2018GL077683>

Maggi, A., Ferrazzini, V., Hibert, C., Beauducel, F., Boissier, P., & Amemoutou, A. (2017). Implementation of a multistation approach for automated event classification 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., et al. (2019). Infrasound array analysis of debris flow activity and implication for early warning. *Journal of Geophysical Research: Earth Surface*, 124(2), 567–587. <https://doi.org/10.1029/2018JF004785>

Mc Ardell, B. W., Bartelt, P., & Kowalski, J. (2007). Field observations of basal forces and fluid pore pressure in a debris flow. *Geophysical Research Letters*, 34, L07406. <https://doi.org/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. <https://doi.org/10.1130/G30928.1>

Mousavi, S., Ellsworth, W., Zhu, W., Zhu, W., Chuang, L. Y., & Beroza, G. (2020). Earthquake transformer: An attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature Communications*, 11, 3952. <https://doi.org/10.1038/s41467-020-17591-w>

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (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. In A. D. Abrahams (Ed.), *Hillslope processes* (pp. 269–296). Mount St. Helens, Washington, SC: Allen and Unwin.

Provost, F., Hibert, C., & Malet, J.-P. (2017). Automatic classification of endogenous landslide seismicity using the random forest supervised classifier. *Geophysical Research Letters*, 44(1), 113. <https://doi.org/10.1002/2016GL070709>

Ravanelli, M., & Bengio, Y. (2018). Speaker recognition from raw waveform with SincNet. *2018 IEEE Spoken Language Technology Workshop (SLT)*, 1021–1028. IEEE. <https://doi.org/10.1109/SLT.2018.8639585>

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

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. <https://doi.org/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 Geoscience*, 12, 75–79. <https://doi.org/10.1038/s41561-018-0274-6>

Schimmel, A., Húbl, J., Mc Ardell, 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. *Quaternary Science Reviews*, 28, 1097–1105. <https://doi.org/10.1016/j.quascirev.2008.10.025>

Stähli, M., Sätttele, M., Huggel, C., McArdell, B., Lehmann, P., Van Herwijnen, A., et al. (2015). Monitoring and prediction in early warning systems for rapid mass movements. *Natural Hazards and Earth System Sciences*, 15, 905–917.

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

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. <https://doi.org/10.5194/nhess-17-939-2017>

Wenner, M., Hibert, C., Meier, L., & Walter, F. (2020). Near real-time automated classification of seismic signals of slope failures with continuous random forests. *Natural Hazards Earth System Science Discussion* [preprint]. <https://doi.org/10.5194/nhess-2020-200>

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

Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). *How transferable are features in deep neural networks?*

Zhang, Z., Walter, F., McArdell, B. W., Wenner, M., Chmiel, M., & He, S. (2020). Extracting dynamics of debris flows from their seismic signature. *Geophysical Research Letters*, e2020GL088994. <https://doi.org/10.1029/2020GL088994>