

Versuchsanstalt für Wasserbau
Hydrologie und Glaziologie
der Eidgenössischen
Technischen Hochschule Zürich

Mitteilungen **268**

Seismic Monitoring and Machine Learning for Gravitational Hazards in Alpine Terrain

Michaela Wenner

Zürich, 2022

Herausgeber: Prof. Dr. Fabian Walter

ETH zürich

Zitiervorschlag für VAW-Mitteilungen:

Michaela Wenner (2022).
Seismic Monitoring and Machine Learning
for Gravitational Hazards in Alpine Terrain.

VAW-Mitteilungen 268, Versuchsanstalt für Wasserbau, Hydrologie und Glaziologie (VAW),
(D. Farinotti, ed.), ETH Zürich, Schweiz.

Im Eigenverlag der
Versuchsanstalt für Wasserbau,
Hydrologie und Glaziologie
ETH Zürich
CH-8093 Zürich

Tel.: +41 - 44 - 632 40 91
Fax: +41 - 44 - 632 40 92
e-mail: info@vaw.baug.ethz.ch

ISSN 0374-0056

DOI 10.3929/ethz-b-000534895

In memory of Oma Lotte, who taught me how to savour life.

PREFACE

The surface of our planet is constantly changing. Alpine terrain, in particular, is prone to mass displacement driven by gravitation. Next to slow processes like creeping terrain and fluvial transport, catastrophic failures of unstable slopes move sediments to lower elevations. Such rapid mass movements pose a significant challenge to inhabitants of the Alpine space, since they are difficult to predict and may develop a tremendous destructive potential threatening roads, villages or even entire valleys in their runout stretches. Moreover, the triggering mechanisms remain incompletely understood as they depend on interactions between climate, fracture processes, friction, material fatigue and possible external forcings like seismic ground shaking or anthropogenic activity. In general, however, it is believed that current climate changes favor catastrophic failures of unstable slopes and demand new monitoring and warning approaches to cope with the unexpected.

Given the relatively long runout distance of rapid Alpine mass movements, alarms are often issued directly upon failure in order to warn affected settlements and infrastructure downslope. Compared to precursor detection, this is technically more feasible. However, the instrumentation widely used for mass movement detection is costly and its sensitivity is limited. For example, Doppler radars capable of recognizing even small rock falls, have limited fields of view and require access to stable power supply. In contrast, mass movements may affect wider regions in inaccessible terrain where solar radiation or batteries are the only power sources. When monitoring overall mass movement activity of a wider mountain range for constraining numerical model predictions, instruments that are sensitive at kilometer scales or longer are necessary.

In this thesis, Michaela Wenner addresses this need for new monitoring and warning approaches: Using seismic measurements she develops and tests real-time detection schemes for debris flows and smaller slope failures, which generate seismic signals that can be recorded over large distances. Debris flows, which are water-sediment mixtures typically mobilized by heavy precipitation, have a high destructive potential, while they move at moderate velocities of 10 m/s or less, often over large distances. Rapid detection is therefore particularly beneficial for debris flows and may yield warning times of minutes or even hours prior to their arrivals at endangered terrain.

Whereas seismic monitoring of mass movements has been the focus of many studies in the recent 1-2 decades, the novelty of the present thesis is the reliance on intelligent algorithms. In contrast to major earthquakes whose monitoring with seismometers is worldwide standard, mass movement seismograms may be weak and onsets rather emergent. This complicates the use of conventional triggering algorithms with threshold-based detection rules, which for mass movements lead to an impermissible amount of false positives. In contrast, intelligent algorithms like the Random Forest (RF) studied in this thesis, relax the role of user-defined thresholds and instead rely on a model, which is trained by example waveforms of previously recorded events. The computer thus learns to recognize mass movement seismograms based on dozens of "waveform features" instead of merely detecting high-amplitude ground shaking.

As a conceptual foundation for supervised machine learning approaches like RF, in Chapter 2 Michaela Wenner investigates the seismic signature of seven debris flows at Illgraben, a highly active torrent in Switzerland's Canton Valais. This allows testing theoretical predictions of seismogenisis in view of peak amplitudes and peak frequencies. The results underline the relevance of topographic steps for elastic energy transmission via particle-ground impacts. At Illgraben, a 48-m free fall after a check dam is therefore a prominent signal source. With the boulder-rich flow front acting as the "loudest" flow portion, the dominant frequency decreases as the flow front approaches a recording seismometer. Spectral characteristics of a debris flow seismogram are therefore diagnostic for mass movement detection and signal discrimination from other sources like earthquakes and anthropogenic noise.

Chapter 4 describes the implementation of the RF algorithm for automatically detecting slope failures at Illgraben and Val Bondasca in Eastern Switzerland. These two sites were chosen based on the geomorphic activity and to understand the value of abundant training data availability (Illgraben) versus limited training data with high signal-to-noise ratios (Val Bondasca). The analysis underlines the importance of training data availability with a highly imbalanced data set for Val Bondasca including only 5 events that were labeled as slope failures but practically an arbitrary amount of earthquake and noise signals. Michaela Wenner pays particular attention to this situation, which is common for dangerous mass movements. Whereas data augmentation techniques do not remedy the effect of limited training data, lowering the predictability threshold, a key parameter in the RF algorithm determining at which point a slope failure signal is classified as such, allows for reliable classification. Even with this imbalanced data set, the detection accuracy reaches a level that permits operational use.

The third core chapter of this thesis is once again centered on the seismic detection of debris flows. To this end, Michaela Wenner uses records from broadband seismometers installed within a few meters of a debris flow flume and the Illgraben channel. The broadband records capture ground tilt as passing debris flows elastically depress the ground material. An analytical solution of the elastic ground response to a point force on the Earth's surface links ground tilt to the instantaneous weight of the debris flow, which measurements at the flume and at Illgraben confirm. This experiment thus constitutes another use of relatively cheap and non-invasive seismic instrumentation for debris flow detection, although it requires installation within the immediate torrent vicinity.

While seismic mass movement monitoring is still subject to research and development, this thesis highlighted its value and potential. The processing steps proposed here have already found their way into prototype warning systems operating in real-time. High-quality observatory data like those from Illgraben are an invaluable asset for further algorithm development and testing. I therefore expect that the results of this work will soon find their way into the next generation of warning systems, which rely on data science techniques like seismic signal classifiers based on machine learning.

Zürich, March 2020

Prof. Dr. Fabian Walter

ABSTRACT

Gravitational hazards, such as rockfalls, landslides, and debris flows in steep mountain terrain, have a high destructive potential, and play a key role in erosion and landscape evolution. Furthermore, in response to climate change, such Alpine mass movements are likely to increase in number and volume in the future. Therefore, monitoring has to be improved to warn against mass movements and mitigate more significant damage. It is crucial to obtain an insight into both the process cascade and the step-wise formation of such events. As an alternative to many conventional approaches, seismology allows monitoring of mass movements at safe distances, provides estimates of event location and timing, and gives insights into flow dynamics and rheology of granular masses. This doctoral thesis aims to establish automated detection and classification of seismic signals and explores the extent to which seismic signatures contain information on mass movement parameters, including flow height and volume.

This first part of the thesis focuses on seismic data recorded between 2017 and 2020 at Illgraben, a steep canyon located in Switzerland, which is known for its high debris-flow and rockfall activity. High rainfall intensities in summer regularly trigger mass wasting with rock-slope failure and debris flows. The frequent debris-flow occurrence (up to ten events per year) makes the Illgraben an ideal site for cross-validating seismically derived mass-movement characteristics with ground-truth data, such as digital terrain models, and other in-torrent measurements. I first focus on the high-frequency (> 1 Hz) spectral content of the seismic signal and use an impact model to estimate particle sizes of debris flows. I show that by focusing on small time windows of the recorded signal, the power spectral density does correlate with particle sizes estimated from video footage. Next, I investigate the long period (> 5 s) seismic data recorded by a broadband seismometer at Illgraben and seismic data from a large-scale debris-flow experiment to estimate ground tilt caused by the weight of the passing mass. I correlate the observed tilt with bulk debris-flow parameters such as volume, vertical force, and maximum flow height and find a Pearson correlation coefficient larger than 0.76 for flow height and maximum tilt for all stations. This indicates that such parameters can successfully be inferred from sensors placed at a safe distance. I then use the solution of the Boussinesq problem for the elastic response of the ground to a surface force to model the elastic ground deformation induced by a distributed mass. I find that despite uncertainties in the model's elastic sub-surface properties, the tilt obtained from the seismic recordings lies within the model's predicted range. To demonstrate the transferability of the findings, I then use a simplified version of the model to make predictions of flow height of two debris flows that occurred at Volcán de Fuego, Guatemala, with sparse auxiliary data coverage.

To investigate monitoring and near-real-time warning capabilities, I use seismic data recorded near Pizzo Cengalo, a mountain located in eastern Switzerland that recently witnessed partial slope collapses up to 3.5×10^6 m^3. To train a machine-learning classifier intended to detect seismic signals of such slope failures, I use windowed seismic data of an imbalanced event catalog containing only five slope failure events, which leads to a highly biased training data set. By lowering the classification threshold for the minority

class (seismic signals labeled as slope failures), I manage to obtain a prediction accuracy of 80% for the slope failure class. In a real-world simulation, I demonstrate that by including a seismic reference station at a great distance to the area of interest (> 100 km), a misclassification of earthquakes as slope failures can be prevented. Moreover, after 1.5 years of operational use, I reevaluate the classifier's performance, during which six slope-failure alarms were issued. While one of the alarms successfully warned against the only reported event during this period, five alarms referred to Magnitude 3.4 - 4.4 tectonic earthquakes in northern Italy. I thus retrain the machine-learning classifier with the new events, which results in a 4% increase in performance for the slope failure class. Finally, the approach from my Pizzo Cengalo study was transferred to Illgraben by training a classifier to detect slope-failure signals at the channel head. The channel deposits are typically remobilized during debris-flow events. Data from summer 2017 is used to train the classifier, which is subsequently applied to summer 2019. I correlate the detected events with meteorological data, such as temperature, precipitation, and atmospheric pressure. In 2019, neither atmospheric pressure nor temperature variations can be identified as clear slope failure triggers. However, precipitation and slope failure frequency show a trend indicating a non-trivial relationship. I propose that freeze-thaw cycles in winter provide loose sediment to the gullies, which gets evacuated due to pore-pressure increase during the snowmelt season and intense rainfall events. Towards the end of the season, the availability of loose sediment limits the slope failure activity.

In summary, I evaluated the capability of seismic measurements in monitoring and real-time warning of gravitational hazards and found that seismology can close the gap between traditional point-in-space and point-in-time measurements. This enhances the understanding of mass-movement triggers and dynamics. Moreover, through careful parameter choices, I showed that machine-learning techniques could give reliable slope-failure detections, enhancing operational monitoring and early warning.

ZUSAMMENFASSUNG

Massenbewegungen wie Felsstürze, Erdrutsche und Murgänge in steilem Berggelände haben ein hohes Zerstörungspotenzial und spielen eine Schlüsselrolle bei der Erosion und der Landschaftsformung. Darüber hinaus werden solche alpinen Massenbewegungen als Reaktion auf den Klimawandel in Zukunft wahrscheinlich an Häufigkeit und Volumen zunehmen. Um grössere Schäden abzumildern, muss die Überwachung verbessert werden. Dafür ist ein besserer Einblick in die Prozesskaskade, also die schrittweise Entstehung solcher Ereignisse, zentral. Als Alternative zu vielen konventionellen Ansätzen ermöglicht die Seismologie die Überwachung von Massenbewegungen in sicherer Entfernung, liefert Schätzungen zu Ort und Zeitpunkt des Ereignisses und gibt Einblicke in die Fliessdynamik und Rheologie. Ziel dieser Doktorarbeit ist es, eine automatisierte Erkennung und Klassifizierung seismischer Signale zu etablieren und zu untersuchen, inwieweit seismische Signaturen Informationen über Parameter der Massenbewegung enthalten, wie z.B. Fliesshöhe und Volumen.

Der erste Teil der Arbeit konzentriert sich auf seismische Daten, die zwischen 2017 und 2020 am Illgraben aufgezeichnet wurden, einer steilen Schlucht in der Schweiz, die für ihre hohe Murgang- und Steinschlagaktivität bekannt ist. Hohe Niederschlagsintensitäten im Sommer lösen regelmäßig Massenabgänge mit Felsstürzen und Murgängen aus. Das häufige Auftreten von Murgängen (bis zu zehn Ereignisse pro Jahr) macht den Illgraben zu einem idealen Standort für den Vergleich von seismisch abgeleiteten Massenbewegungsmerkmalen mit Daten wie digitalen Geländemodellen, Fliesstiefenabschätzungen und anderen Messungen im Gelände.

Ich konzentriere mich zunächst auf den hochfrequenten (> 1 Hz) Spektralgehalt des seismischen Signals und verwende ein Aufprallmodell, um die Korngrössen des Murgangs zu schätzen. Ich zeige, dass durch die Konzentration auf kleine Zeitfenster des aufgezeichneten Signals die spektrale Leistungsdichte mit den aus dem Videomaterial geschätzten Partikelgrössen übereinstimmt. Als Nächstes untersuche ich seismische Daten mit langer Periode (> 5 s), die von einem Breitbandseismometer am Illgraben aufgezeichnet wurden, sowie seismische Daten aus einem gross angelegten Murgangexperiment, um die durch das Gewicht der vorbeifliessenden Masse verursachte Bodenneigung abzuschätzen. Ich korreliere die beobachtete Neigung mit Murgangparametern wie Volumen, Normalkraft und maximaler Fliesshöhe und finde einen Pearson-Korrelationskoeffizienten von mehr als 0,76 für die Fliesshöhe und der maximalen Neigung für alle Stationen. Dies deutet darauf hin, dass solche Parameter erfolgreich von Sensoren abgeleitet werden können, die in sicherer Entfernung installiert sind. Anschließend verwende ich die Lösung des Boussinesq-Problems für die elastische Reaktion des Bodens auf eine Oberflächenkraft, um die elastische Bodenverformung zu modellieren, die durch eine verteilte Masse verursacht wird. Ich stelle fest, dass die aus den seismischen Aufzeichnungen gewonnene Neigung trotz der Unsicherheiten in den elastischen Untergrundeigenschaften innerhalb des vom Modell vorhergesagten Bereichs liegt. Um die Übertragbarkeit der Ergebnisse zu demonstrieren, verwende ich anschliessend eine vereinfachte Version des Modells, um Vorhersagen über die Fliessshöhe von zwei Murgängen am Volcán de Fuego in Guatemala zu treffen, für die nur wenige andere Daten vorliegen.

Zur Untersuchung der Überwachungs- und Echtzeit-Warnmöglichkeiten verwende ich seismische Daten, die in der Nähe des Pizzo Cengalo aufgezeichnet wurden, einem Berg in der Ostschweiz, an dem es kürzlich zu partiellen Hangeinstürzen von bis zu $3{,}5 \times 10^6$ m^3 kam. Um mit maschninellem Lernen ein Model zur Klassifizierung zu trainieren, welches seismische Signale solcher Hangabbrüche erkennen soll, verwende ich einen Ereigniskatalog, der nur fünf Hangabbruchereignisse enthält, was zu einem stark verzerrten Trainingsdatensatz führt. Durch Herabsetzen des Klassifizierungsschwellenwerts für die am wenigsten repräsentierte Klasse (seismische Signale, die als Massenbewegung gekennzeichnet sind) gelingt es mir, eine Vorhersagegenauigkeit von 80% für die Klasse der Massenbewegungen zu erreichen. In einer realen Simulation zeige ich, dass durch die Einbeziehung einer seismischen Referenzstation in großer Entfernung zum Untersuchungsgebiet (> 100 km) eine Fehlklassifizierung von Erdbeben als Massenbewegung verhindert werden kann. Darüber hinaus bewerte ich nach 1.5 Jahren Betriebszeit die Leistung des Klassifikationsmodells neu, wobei sechs Massenbewegungsalarme ausgegeben wurden. Während einer der Alarme erfolgreich vor dem einzigen gemeldeten Ereignis in diesem Zeitraum warnte, bezogen sich fünf Alarme auf tektonische Erdbeben der Stärke 3.4 - 4.4 in Norditalien. Daher trainiere ich den Machine-Learning-Klassifikator mit den neuen Ereignissen neu, was zu einer Leistungssteigerung von 4% für die Klasse der Massenbewegungen führte.

Schließlich wurde der Ansatz aus meiner Pizzo-Cengalo-Studie auf den Illgraben übertragen, indem ein Klassifikator trainiert wurde, um Massenbewegungen am Gerinnekopf zu erkennen. Diese Ablagerungen werden typischerweise bei Murgängen remobilisiert. Mit Daten aus dem Sommer 2017 wurde der Klassifikator trainiert und anschliessend auf den Sommer 2019 angewendet. Ich korreliere die erkannten Ereignisse mit meteorologischen Daten, wie Temperatur, Niederschlag und Luftdruck. Im Jahr 2019 können weder Luftdruck- noch Temperaturschwankungen als eindeutige Auslöser für Hangversagen identifiziert werden. Niederschlag und Häufigkeit von Hangrutschungen stehen jedoch in einem komplexen Zusammenhang. Ich stelle die Hypothese auf, dass die Frost-Tau-Zyklen im Winter die Rinnen mit losem Sediment versorgen, das aufgrund des erhöhten Porendrucks während der Schneeschmelze und intensiver Niederschlagsereignisse mobilisiert wird. Gegen Ende der Saison begrenzt das Vorhandensein von Lockersedimenten die Aktivität der Massenbewegungen.

Zusammenfassend bewertete ich die Fähigkeit seismischer Messungen zur Überwachung und Echtzeitwarnung vor Massenbewegungen und stellte fest, dass die Seismologie die Lücke zwischen traditionellen Punktmessungen in Raum und Zeit schliessen kann. Dies kann dazu beitragen, ein vollständigeres Bild der Auslöser und der Dynamik von Massenbewegungen zu erhalten. Darüber hinaus habe ich gezeigt, dass maschinelle Lernverfahren durch eine sorgfältige Auswahl von Parametern zu einer zuverlässigen Erkennung von Massenbewegungen führen können, was die operative Überwachung und Frühwarnung ermöglicht.

CONTENTS

1 INTRODUCTION 1
1.1 Context . . . 1
1.2 Seismic signals of granular flows . . . 2
1.3 Detection and classification of slope failure signals . . . 3
1.4 Objectives and outline of this thesis . . . 4
2 DECIPHERING DEBRIS-FLOW SEISMOGRAMS AT ILLGRABEN, SWITZERLAND 7
2.1 Introduction . . . 8
2.2 Study Site . . . 9
2.3 Data: Debris-flow seismograms . . . 10
2.4 Peak frequency migration . . . 13
2.5 Debris-flow characteristics from seismic signals . . . 13
2.6 Conclusion . . . 15
3 SEISMOMETER RECORDS OF GROUND TILT INDUCED BY DEBRIS FLOWS 17
3.1 Introduction . . . 17
3.2 Study sites . . . 19
3.2.1 USGS debris-flow flume . . . 19
3.2.2 Illgraben, Switzerland . . . 20
3.3 Methods . . . 22
3.3.1 Tilts at frequencies above instrument low-corner frequency . . . 23
3.3.2 Tilts at frequencies below instrument low-corner frequency . . . 23
3.3.3 Modelling tilt of a distributed moving mass . . . 24
3.4 Results . . . 26
3.4.1 USGS debris-flow flume, Oregon, USA . . . 26
3.4.2 Illgraben, Switzerland . . . 27
3.4.3 Correlation with debris-flow parameters . . . 28
3.4.4 Modelled tilt signal . . . 29
3.5 Discussion . . . 32
3.5.1 Correlation with debris-flow parameters . . . 32
3.5.2 Model parameters . . . 34
3.5.3 Tilt measurements as an alternative to force plates? . . . 35
3.5.4 Transferability to other sites . . . 36
3.5.5 Low frequency data for early warning purposes . . . 38
3.6 Conclusion . . . 39
4 NEAR REAL-TIME AUTOMATED CLASSIFICATION OF SEISMIC SIGNALS 41
4.1 Introduction . . . 42
4.2 Study Sites and Data Set . . . 44
4.2.1 Illgraben, Switzerland . . . 44
4.2.2 Pizzo Cengalo, Switzerland . . . 45
4.2.3 Labeled Data Set . . . 45
4.3 Methodology . . . 48
4.3.1 Data Stream Handling . . . 48
4.3.2 Imbalanced Data Set Handling . . . 49
4.3.3 Training Process and Evaluation . . . 50

4.3.4 Simulated Real-Time Monitoring at Pizzo Cengalo 51
4.4 Results 52
4.4.1 Classifier Performance on Labeled Data Sets 52
4.4.2 ROC Analysis on Pizzo Cengalo Data Set 52
4.4.3 Classifier Performance with Classification Threshold on Pizzo Cengalo Data Set 54
4.4.4 Classifier Implementation on Pizzo Cengalo Seismic Data 55
4.4.5 Transferability to other study sites 55
4.5 Discussion 59
4.5.1 Data Set Comparison 59
4.5.2 Performance Enhancement 60
4.5.3 One-step vs. two-step method at Pizzo Cengalo 61
4.5.4 Test on 2019 Pizzo Cengalo data 62
4.5.5 Classifier Transferability 63
4.6 Conclusion 64
5 SUMMARY OF PUBLICATIONS 67
6 OUTLOOK 69
6.1 Pizzo Cengalo classifier 69
6.1.1 Classifier performance 69
6.1.2 Retraining 70
6.1.3 Discussion 71
6.2 Illgraben event detection and long-term monitoring 74
6.2.1 Long-term monitoring 74
6.2.2 Slope failure signals 77
6.2.3 Discussion 78
6.3 Implications and outlook 83
6.3.1 Implications for automatic signal detection 83
6.3.2 Implications for debris-flow monitoring 84
7 CONCLUSION 87
A APPENDIX 89
A.1 Supplementary material to chapter 2 - signals at other seismic stations . 89
A.2 Supplementary material to chapter 4 90

BIBLIOGRAPHY 96

ACKNOWLEDGEMENTS 105

CURRICULUM VITAE

1

INTRODUCTION

Was einmal gedacht wurde, kann nicht zurückgenommen werden.

Friedrich Dürrenmatt, Die Physiker

1.1 CONTEXT

In alpine regions, gravitational hazards such as rockfalls, rock- and snow avalanches, and debris flows constitute a severe threat to infrastructure and human life. With a changing climate, the resulting thawing of permafrost, deglaciation, and an increase of extreme meteorological events, these mass movements are expected to increase in frequency, thus playing a central role in the context of future land-use planning and hazard prevention (e.g., Allen & Huggel, 2013; Hock et al., 2019). While adequate prevention measures and suitable planning of infrastructures will be key to minimize the related risks, it is also crucial to monitor potentially endangered sites with high temporal and spatial resolution: Only in this way can dangers be minimized and accidents be prevented. Traditional monitoring techniques include, for example, in-channel geophone installations, extensometers, ultrasonic flow depth sensors, and terrestrial laser scanners (e.g., Badoux et al., 2009; Abellán et al., 2011). However, despite their widespread use, these methods have major drawbacks: In-channel or on-mountain-flank installations are at risk of being lost or destroyed during major events, and installation in challenging terrain is often dangerous. Additionally, many of these techniques are either point measurements in space or, for the case of terrestrial laser scanners, very expensive for long-term monitoring. Seismometers may bridge these gaps by recording ground motion at high temporal resolution and because of their sensitivity to surrounding areas. In recent years, seismology, with its cost-effective and straightforward instrument installations at a safe distance to unstable terrain or debris-flow torrents, has evolved into a standard tool to monitor and study mass movements (Larose et al., 2015; Allstadt et al., 2018).

Despite its benefits, mass movement monitoring with seismometers faces several complications. First, seismometers only record indirect properties of the mass movements. The observed signal is not only influenced by the seismic source but also by both the travel path of the seismic waves through the often heterogeneous sub-surface and the response of the instrument. Additionally, seismometers are sensitive to any seismogenic process, including earthquakes and anthropogenic noise, such as industrial machinery, nearby air traffic, and others. Often, these noise signals conceal signals of mass-movement events. Hence, to use the seismic signals of mass movements for monitoring purposes, one has to address these difficulties by (a) extracting process-related features of the seismic signal by deconvolving influences from the wave's travel path and the instrument response, and (b) developing algorithms to detect seismic signals of interest within a continuous data stream with an abundance of noise signals.

In this thesis, I try to find solutions for these problems. In particular, I investigate seismic signals of debris-flows and slope failures, whose high-frequency signals are generated by stochastic ground impacts of many individual particles.

1.2 SEISMIC SIGNALS OF GRANULAR FLOWS

When gravitational mass movements occur, part of the potential energy is released in the form of elastic waves that propagate through the earth and can be recorded up to thousands of kilometers away from the source. These signals are used to study slope failures in volcanic (e.g. Norris, 1994; Hibert et al., 2011; Hibert et al., 2017) and mountainous areas (e.g. Deparis et al., 2008; Vilajosana et al., 2008; Dammeier et al., 2011; Burtin et al., 2014; Dietze et al., 2017; Provost et al., 2017) or snow avalanches (e.g. Suriñach et al., 2005; Heck et al., 2018). The signals are characterized by an emergent onset with dominant frequencies of $5-10\,\mathrm{Hz}$ and no distinguishable seismic phases. Granular flows usually show a cigar-shaped form that is similar to tremors observed in volcanic environments (e.g., Hibert et al., 2011). The signal duration of mass movements varies between seconds and hours, depending on the type of event (Allstadt et al., 2018).

The recorded signals can generally be divided into two components: the high-frequency part and the low-frequency part. The high frequency signal ($> 1\,\mathrm{Hz}$) is generated by particle impacts against the earth's surface (e.g., Tsai et al., 2012; Allstadt et al., 2018). Owing to the numerous particle-bed and inter-particle collisions, source models of the high-frequency mass-movement signals are complicated. Particle impacts can be modeled as stochastically distributed in time, but the relation to grain size distributions, flow depth, and flow velocities, is subject to ongoing research and still not clear (e.g., Farin et al., 2019; Allstadt et al., 2020; Zhang et al., 2021).

Recently, Kean et al. (2015) and Lai et al. (2018) adapted an impact model for fluvial bedload transport introduced by Tsai et al. (2012) to explain the high-frequency seismic spectrum of debris flows. The model explains the seismic signal in terms of instantaneous Hertzian impacts of bedload particles on the ground. In this way, the impact force generating seismic waves results from a change in the particle's linear momentum (impulse), which is two times the product of the particle's initial mass and velocity (Tsai et al., 2012). The model shows that spectral amplitudes are primarily influenced by the grain sizes of the boulder-rich debris flow front. On the other hand, the peak frequencies are controlled by the source-receiver distance, with higher (lower) peak frequencies indicating shorter (longer) distances between the flow front and the seismometer. This is owed to the inelastic damping of the seismic wave, causing a high-frequency cutoff. By analyzing peak frequencies at different time steps, one can thus estimate the location and velocity of the flow (Lai et al., 2018). Kean et al. (2015) adapt the Hertzian impact model to investigate the influence on seismic signals of debris flows in sediment-filled channels. Instead of using the Hertzian impact model for the whole channel cross-section, only the exposed bedrock on the sides is modeled. In contrast to bedrock, sediment cover dampens the impacts on the channel's sediment-filled middle part. With progressing sediment entrainment, the area of exposed bedrock increases, and thus, the model allows estimating entrainment rates. The model confirms direct observations of sediment-covered channels, for which amplitudes of seismic signals are considerably smaller than for bare bedrock.

Farin et al. (2019) extended the Lai et al. (2018) model to consider not only the debris-flows boulder-rich snout but the whole debris flow in the source impact model. They find a power-law dependence between the apparent total impact force spectrum and the flow-front velocity as well as the effective particle diameter. Based on this model, Allstadt et al. (2020) find systematic trends between the seismically derived basal forces and debris-flow parameters such as flow thickness and normal stress; however, no direct constraints. Similarly, Zhang et al. (2021) find a weak correlation between impact force spectra inverted from seismic signals and particle sizes. Additionally, changes in the inverted spectra indicate changes in bed topography, hinting towards erosion during large events and deposition during small events.

In contrast to the high-frequency signals, at a great distance ($>$ tens of kilometers), the low-frequency signal ($< 5\,\mathrm{s}$) is modeled by the elastic response of the earth to acceleration and deceleration of the bulk mass of the flow (e.g. Allstadt, 2013; Ekström & Stark, 2013). Unfortunately, for smaller-scale mass movements ($< 10^5 m^3$), the force exerted on the earth is often too small to generate detectable low-frequency elastic waves at distances larger than several kilometers. At closer distances to the event ($<$ hundreds to thousands of meters), low-frequency signals can be interpreted as the instrument tilt caused by the elastic deformation of the earth due to the weight of the mass movement on the earth's surface.

1.3 DETECTION AND CLASSIFICATION OF SLOPE FAILURE SIGNALS

Analyzing the large amount of data acquired during continuous recordings of several stations requires the automation of seismic event detection. A well-known detection algorithm in classical seismology is the so-called "STA/LTA" trigger (e.g., Allen, 1982). It compares the average amplitudes of a short-time window (STA) to an adjacent and trailing long-time window (LTA). At an impulsive event onset, the short-time average will change drastically, whereas the long-time windows average will stay constant if seismic background noise is stationary. However, due to mass-movement signals emergent onsets, detection is often unreliable. Helmstetter and Garambois (2010) introduced a method similar to STA/LTA but applied it to the frequency rather than the time domain using the signal's spectrogram. However, such algorithms detect any signal with elevated power in a user-specified frequency band and do not distinguish between noise and events of interest.

Recently, machine learning algorithms have become a popular tool to perform such classification tasks in seismology (e.g., Beyreuther & Wassermann, 2008; Hibert et al., 2011; Maggi et al., 2017). These algorithms can be divided into supervised and unsupervised machine learning techniques. Supervised techniques are learned on training data with a set of signals that have been labeled by an expert, whereas unsupervised techniques do not require any pre-labeling. Here, labeling refers to the process of identifying the seismic data and adding a meaningful label, such as "earthquake". Several different types of algorithms have been developed and are openly available. For instance, Hibert et al. (2017) and Provost et al. (2017) use the Random Forest algorithm (RF), a supervised classifier, to automatically classify seismic events in volcanic and mountainous environments with more than 90% of correctly classified events compared to data sets labeled by experts. Hammer et al. (2013) and Dammeier et al. (2016) combined the detection and classification step by using Hidden Markov Models and showed that larger rockfalls are detectable

by a regional seismic network. Hammer et al. (2017) and Heck et al. (2018) use their approach to automatically detect snow avalanches. The long-term study ($\sim$ 100 days) of Heck et al. (2018) showed that a changing background noise strongly decreases the performance of their classifiers, and hence, the classifier needs to be updated regularly by including the current background noise. However, continuously retraining the algorithm can be computationally expensive and, therefore, unsuitable for real-time monitoring.

1.4 OBJECTIVES AND OUTLINE OF THIS THESIS

In this thesis, I investigate (1) what the seismic signals of more than four years of recorded debris-flow activity can tell us about debris-flow properties, and (2) how and to what extent data from a seismic network with only few recordings of prior slope failures can be used to train and apply a machine-learning model to automatically monitor nearby unstable slopes.

In chapter 2, I analyze seismic signals of seven debris flows recorded in 2017 and 2018 at Illgraben, Switzerland. I find that, due to inelastic damping, the frequency content of the recorded signal depends on the relative location between the debris-flow front and the seismometer. Additionally, I use the particle impact model developed for bedload transport (Tsai et al., 2012) and adapted to debris flows by Lai et al. (2018), to analyze the frequency spectrum of the recorded events. I find that the signals are dominated by two noise sources: (1) the tallest check dam in the Illgraben torrent, with a free-fall height of 48 m and (2) the part of the torrent closest to the station. Furthermore, I use observations from the seismic spectrum to compare to particle sizes derived by independent video analysis of the debris flows. I find that the spectral amplitudes of the seismic signal are smaller for flows with smaller particle sizes, which is in line with what the model by Tsai et al. (2012) suggests.

Instead of using the high-frequency signals of debris flows, in chapter 3, I investigate very low frequencies ($< 1/5$ Hz) recorded by a seismometer placed within tens of meters of the channel. This signal is caused by elastic ground deformation due to the weight of the debris flow. With a large-scale debris-flow experiment performed at the U.S. Geological Survey (USGS) debris-flow flume and by using direct tiltmeter observations, I confirm that seismometers are indeed sensitive to tilt. I additionally compare the maximum tilt obtained from the seismic recordings with debris-flow parameters such as volume, maximum flow height, and maximum vertical forces recorded at a closeby force plate at the USGS debris-flow flume and during 15 debris-flow events at Illgraben. Additionally, I use a distributed-load model based on the solution of the Boussinesq problem (e.g., Ling et al., 2002) to compare theoretical and observed tilt at both sits. I then test the model's applicability by comparing its output to observed tilt to predict the maximum flow height of two lahar events at Volcán de Fuego, Guatemala, for which only seismic recordings are available.

In chapter 4, I move from the seismically derived characteristics of mass-movement events to automatic signal classification. For this, I analyze the potential of a small seismic network installed at Pizzo Cengalo, Switzerland, to function as a real-time mass movement detector. First, I train a machine-learning model based on a manually compiled event catalog. The catalog consists of only five mass movements, making a statistical analysis challenging. I introduce an approach to continuously classify the seismic data by using progressing time windows, to facilitate real-time monitoring. Next,

I show how the classifier's performance compares against (a) an STA/LTA detection and a subsequent classification and (b) an event classification at Illgraben, where the catalog consists of almost a hundred mass-movement events, but with a lower signal to noise ratio. Lastly, I show how the performance can be improved by adding a classification output from another seismic station more than a hundred kilometers away from the network.

Chapter 2, 3 and 4 are self-contained chapters consisting of published or submitted scientific articles. Each chapter can be read on its own. In chapter 6, I revisit the classification output of the automatic classifier implemented at Pizzo Cengalo described in chapter 4. I analyze its performance on new data from 2020 and 2021, and retrain the classifier with meanwhile detected events. I find that retraining improves the classifier's performance on the test data set by 4% in the slope failure class. Additionally, I train a separate classifier on the Illgraben data set to detect slope failures at active slopes supplying the channel with sediment throughout the year. For example, I show the result of the detections in 2019 and compare the number of events with temperature, precipitation, and atmospheric pressure. The latter analysis aims at providing hints for the degree to which such parameters affect slope stability and hence mass-movement occurrence.

2

DECIPHERING DEBRIS-FLOW SEISMOGRAMS AT ILLGRABEN, SWITZERLAND

This chapter is slightly adapted from the published version by *Association of Environmental and Engineering Geologists* as:
Wenner, M., Walter, F., McArdell, B., & Farinotti, D. (2019). Deciphering debris-flow seismograms at Illgraben, Switzerland. *Contained in: Proceedings of the Seventh International Conference on Debris-Flow Hazards Mitigation, Golden, Colorado, USA, June 10-13, 2019*, 11(1), 1-10. http://dx.doi.org/10.25676/11124/173215.

ABSTRACT

Mass wasting, such as rockfalls, landslides and debris flows in steep mountain terrain, has a high destructive potential, and plays a key role in both erosion and landscape evolution. As an alternative to many conventional approaches, seismology allows monitoring of such mass movements at safe distances, provides estimates of event location and timing, and can give insights into dynamics and rheology granular flows. Here, we analyze seismic data recorded during the 2017 and 2018 debris-flow seasons at Illgraben, a steep canyon located in Switzerland. Yearly precipitation is controlled by summer rainstorms with high rainfall intensity during which mass wasting including rock-slope failure and debris flows occur regularly. The frequent debris-flow occurrence (on average three events per year) makes the Illgraben an ideal site for cross-validating a seismically-derived event catalog of mass movements with ground-truth data, such as digital terrain models, flow depths estimates and other in-torrent measurements. We present seismic frequency characteristics of the Illgraben debris-flow series and investigate how the seismic signature depends on actual debris-flow characteristics, such as grain sizes, and on propagation effects of the generated seismic waves. Whereas these two effects are usually difficult to separate, the source component contains valuable information on the flow's material composition. Stations that are close to the torrent, we find that dominant frequencies in the recorded signal reflect the distance to the dominant source. For one particular station, this is shown on recordings of several events, where a dominant frequency of about 5.5 Hz indicates the passing of the flow at a 48 m check dam. Power spectral densities at that instance give an estimate of the particle content of the debris flow. We also find that a jump in dominant frequency does not necessarily reflect the location of the flow front. Seismic studies of debris-flow dynamics and material composition should therefore not be limited to entire debris-flow seismograms, but instead focus on individual time windows and consider different sensors separately. The presented analysis underlines the use of seismic data in torrent and landscape studies.

2.1 INTRODUCTION

In mountainous areas, mass movements such as debris flows are a significant threat to infrastructure, properties and human life. Monitoring of debris-flow prone catchments is essential to improve our understanding of debris-flow dynamics, and needed for both designing mitigation measures and damage reduction. Standard granular flow models for debris flows assume a shallow propagating mass, in which the rheology is described by an effective friction (Mangeney-Castelnau et al., 2005; Mangeney et al., 2007; Christen et al., 2010). Field measurements are essential to constrain this effective friction and other parameters, including erosion, flow depth and flow density. Classical monitoring techniques, like radar altimeters, aerial imagery and geophones within the torrent, can give information about onset time, flow depth, discharge and erosion or deposition areas. However, they offer no direct measurements of flow characteristics such as effective friction and material density (for an overview in debris-flow monitoring instrumentation see Arattano and Marchi (2008)).

In recent years, seismology has evolved into a standard tool to study mass movements and their dynamics at high temporal resolution (Larose et al., 2015; Allstadt et al., 2018). Seismic signals generated by such events are often classified into low and high frequency content. The low frequency signal ($< 10\,s$) is modelled by the elastic response the earth to acceleration and deceleration of the bulk mass of the flow (e.g., Ekström & Stark, 2013). In contrast, higher frequency signals ($> 1\,Hz$) are generated by collisions of grains within the flow, and their impacts on the bed. For smaller scale mass movements, the force that is exerted on the earth is often too small to generate detectable low- frequency elastic waves. In contrast, high frequency signals can be recorded in such cases, and are used to study rockfalls and rock avalanches in both volcanic (e.g., Norris, 1994; Hibert et al., 2011; Hibert et al., 2017) and mountainous areas (e.g., Deparis et al., 2008; Vilajosana et al., 2008; Dammeier et al., 2011; Burtin et al., 2014; Dietze et al., 2017; Provost et al., 2017) or snow avalanches (e.g., Suriñach et al., 2005; Heck et al., 2018). Such signals are typically emergent with dominant frequencies of $5-10\,Hz$ and no distinguishable seismic phases. Signal durations vary between seconds to several tens of seconds, depending on the type of movement. Consisting of numerous and overlapping individual particle-bed and inter-particle collisions, source models of the high frequency mass movement signals are far more complicated than the low frequency component of the signal, and are hence less understood. Recently, Kean et al. (2015) and Lai et al. (2018) adapted a model of fluvial bedload transport to explain the high- frequency seismic spectrum of debris flows in sediment-filled channels, and during a debris flow in California, respectively. The model explains the seismic signal in terms of instantaneous Hertzian impacts of bedload particles on the ground. In this way, the impact force generating seismic waves results from a change in the particle's linear momentum (impulse), which is two times the product of the particle's initial mass and velocity (Tsai et al., 2012). To simplify the model for debris flows, these impacts are integrated over the boulder-rich flow front.

The adapted Tsai et al. (2012) model makes assumptions for particle velocities and seismic path effects describing the propagation of seismic waves from the river bed to a recording unit. Importantly, the calculated seismic spectrum is influenced by the poorly constrained grain size distributions of moving particles, as well as seismic velocities and attenuation of ground substrate influence. Nevertheless, the model explains observed seismic frequency spectra for bedload transport (Burtin et al., 2008). For debris flows,

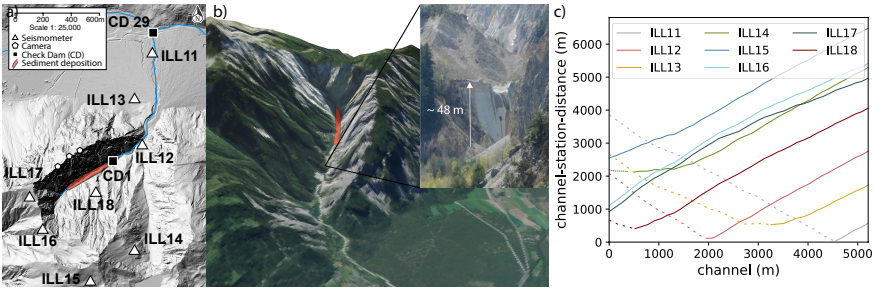


FIGURE 2.1: (a) Overview of the Illgraben catchment and the instrumentation installed during the debris-flow season (June-August) 2018. Seismometer locations (white triangles), the area of sediment supply for the debris flows (shaded red area), and the position of the 48 m tall check dam 1, and check dam 29 (black squares) are depicted (b) 3D orthophoto with sediment deposition area in red and check dam 1 shown on a photo. (c) Channel-receiver distances for all seismic stations installed in 2018. The y-axis represents the distance whereas the x-axis marks the position along the channel (origin at the head of the catchment; 5000 m corresponds to the Illbach inflow into the Rhone river). Dashed (continuous) lines show the section where the flow is approaching (moving away from) the station.

the model proposes that spectral amplitudes are primarily influenced by the grain sizes of the boulder-rich debris-flow front. Furthermore, the peak frequencies are controlled by the source-receiver distance, with higher (lower) peak frequencies indicating shorter (longer) distances between flow front and seismometer. By analysing peak frequencies at different time steps, Lai et al. (2018) propose that one can estimate location and velocity of the flow (Lai et al., 2018). Kean et al. (2015) used their adapted model to invert for sediment cover, and to estimate entrainment rates that compare well with observations.

Here, we study the high frequency content of seismic signals recorded at the debris-flow prone Illgraben torrent, located in Switzerland. This torrent is one of the most active catchments in the Alps (Hürlimann et al., 2003), producing several debris flows per year. In 2017 and 2018, we recorded a total of seven debris flows of different volumes, flow velocities and material composition. We relate spectral characteristics of the debris-flow seismograms to flow-receiver distances, and to topographic features within the torrent. The results show that the spectrum (1) cannot be analyzed as a whole, but has to be segmented in time in order to elucidate characteristics of the debris flow and (2) has to be treated independently for each sensor since the latter are sensitive to different stages of the debris flow.

2.2 STUDY SITE

The Illgraben catchment, located in the southwest of Switzerland, spans from its highest point, the Illhorn (2716 m asl), down to the Rhone Valley where its main torrent, the Illbach, flows into the Rhone River (610 m asl). Past activity within the 9.5 km^2 catchment and 5 km long torrent produced a large fan with a radius of about 2 km and a volume of about 500 x 10^6 m^3 (Hürlimann et al., 2003; McArdell et al., 2007). Illgraben

is characterized by a complex geology, with the northwest Illhorn face and the head of the trunk channel being dominated by highly fractured quartzite, and the southeast facing slope of the catchment mostly consisting of limestone (Schlunegger et al., 2009; Bennett et al., 2012). The fractured quartzite with erosion rates of tens of centimeters per year is the main contributor to the sediment transported via debris flows (Bennett et al., 2012). Precipitation patterns during summer are characterized by storms, with rainfall intensities of up to several tens of mm/h, and a duration of half an hour to one hour (Hürlimann et al., 2003). In such events, sediment deposited at the head of the trunk channel is mobilized by water gathered along the steep ($>$ 40 degrees) slopes (McArdell et al., 2007). This produces on average $3-5$ debris flows per year that contribute significantly to the sediment discharge of the Rhone River (Schlunegger et al., 2009). Typical events have volumes around 20,000 m^3 and velocities $3-4\,m/s$, and transport blocks up to several meters in diameter. In contrast, some events have lower concentration of sediment and fewer such large particles and are more appropriately classified as "debris floods" (e.g., Pierson et al., 1987). After a large ($3 \times 10^6\,m^3$) rock avalanche in 1961 in the upper catchment, a 48 m tall check dam (CD1, Figure 2.1) was built within the channel to stabilize the deposit and to prevent large debris flows (Hürlimann et al., 2003). Further downstream along the torrent, 29 additional check dams of several meters height were built to minimize vertical and lateral erosion, thereby stabilizing the channel at the present location. With these measures, most debris flows no longer leave the channel, and damage to infrastructure is rare. To our knowledge, the Illgraben carries little or no discharge in the summer between debris flows, and therefore considerable recreational activity takes place near and in the channel. For this reason, an early warning system was installed in 2000. It is maintained by the Swiss Federal Institute for Forest, Snow and Landscape research (WSL) and consists of radar flow-depth sensors and geophone sensors to provide automatic detection of flows. Upon detection, acoustic alarms are activated and information is sent to local authorities. The data from the early warning system compliments the existing scientific observation station (no warnings are generated) which consists of radar, laser and ultrasonic flow stage sensors at selected check dams, geophones mounted on the downstream facing wall of check dams, as well as video cameras and a 2 m x 4 m force plate (McArdell et al., 2007; Badoux et al., 2009). Recently, geophone and infrasound sensors have additionally been installed on the fan to increase early warning capabilities (Schimmel et al., 2018; Marchetti et al., 2019). An additional seismic network was installed throughout the Illgraben catchment between May and September in both 2017 and 2018 (Fig. 2.1). This network recorded seven debris flows, and consisted of eight stations with real-time data transfer via the mobile phone network. The interstation distance of the network is about 1.5 km, with an aperture of 5.5 km. Most stations operate with a Lennartz-1s sensor that has a flat response between 1 - 100 Hz and a sampling rate of 100 Hz. Station ILL11 is equipped with a Trillium compact sensor with a low-frequency corner at 120 seconds.

2.3 DATA: DEBRIS-FLOW SEISMOGRAMS

During the monitoring period in 2017, three large ($>$ 25,000 m^3) debris flows occurred, with one flow of 100,000 m^3 (Table 2.1). In 2018 four events were recorded. After the destruction of the force plate in July 2016, volume and flow depth are estimated at the instrumented wall, CD 29 (Fig. 2.1), situated about 10 m upstream of the location of the

TABLE 2.1: Characteristics of the seven debris flows recorded in 2017 and 2018 (n.a. denotes that estimates are not yet available for 2018). "CD1" stands for "check dam 1".

Date	Arrival time CD1 (UTC)	Volume (m^3)	Velocity (m/s)	Flow depth
2017-05-29	16:58:31	100000	6.7	4.8
2017-06-03	23:27:38	25000	5.1	3.3
2017-06-14	19:30:48	35000	7.1	3.4
2018-06-11	10:46:39	35000	7.0	3.5
2018-06-12	18:29:16	n.a,	n.a.	n.a.
2018-07-25	16:56:40	$<$ 50000	4.69	2.0
2018-08-08	17:49:25	$<$ 100000	6.70	n.a.

force plate (Berger et al., 2011). Due to the irregularity of the cross-section shape and the variability in the direction of the approaching flow at this location, flow depths tend to be over-estimated when compared to values previously reported at the force plate. Flow velocity is calculated from the travel time between in-torrent sensors along the channel as described in Schlunegger et al. (2009). As discussed in Section 2.5, video footage and power spectral densities of the signals suggest a relatively large water content without a boulder-rich front for the first two events in 2018.

Figure 2.2a shows the seismogram of a debris flow recorded on August 8, 2018. The signal shows the emergent onset and dominant frequencies above 5 Hz, typical for mass movements. The signal emerges from the background noise at a time that depends on the distance between the debris-flow front and the recording station (Walter et al., 2017). For the shown event, amplitudes at different stations vary between $1.5 \times 10^{-3}\,ms^{-1}$ at station ILL11, and $2 \times 10^{-6}\,ms^{-1}$ at ILL16. High amplitude impulsive signals on some stations within the network are of atmospheric origin, generated by thunder, and not directly related to the debris-flow signal (Marchetti et al., 2019). Between debris flows, seismic background noise at Illgraben is dominated by several almost discrete frequency bands (Fig. 2.3). Between 1 and 5 Hz, anthropogenic noise is present showing diurnal variation, as well as lower energy on both weekends and public holidays compared to work days. Though strong, this signal is unlikely to affect seismic detectability of debris flows (Walter et al., 2017). Another distinct frequency band of noise is found at about 15 Hz. Within this band, an abrupt decrease in power can be observed at the end of June of both 2017 and 2018. We suggest that this could either be related to water discharge in the catchment due to snow melt, or to a change in hydropower operations at a dam 3 km away from the seismic station. Here we concentrate on the frequency spectrum of the signal directly before and during the debris-flow events. We focus on station ILL12, because its near-torrent location implies a large range of distances to the flow front, both up and down-stream (Fig. 2.1c). Such distance variations are particularly important for investigating variations in frequency signature (Lai et al., 2018). The goal is, to separate source and path effects on the frequency content of the signal, to constrain source mechanisms.

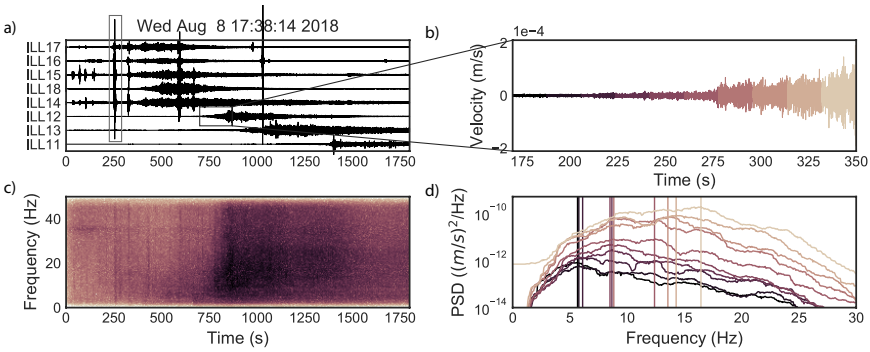


FIGURE 2.2: (a) Seismogram of August 8, 2018 debris flow recorded on all stations of the network. Amplitudes were normalized for each station. Impulsive signals (grey rectangle) are of atmospheric origin (thunderstorm). (b) Close up of onset of the debris-flow signal on station ILL12. Note the emergent onset over several tens of seconds. (c) Spectrogram of signal recorded at station ILL12. The colors show the energy of the frequencies at each time step in decibel, where darker colors represent a higher energy. (d) Spectra of ten time-windows of 18 seconds corresponding to the color-coded waveform in (b). The spectra are smoothed with a 1 Hz running average. The vertical lines mark the peak amplitude of the spectral power.

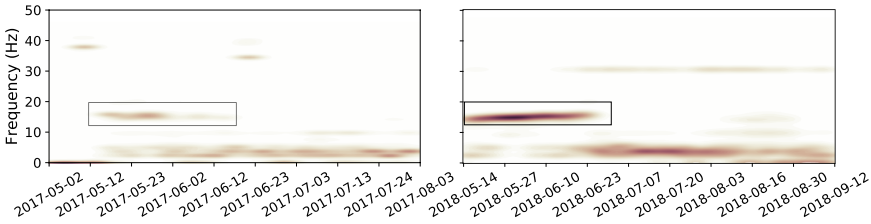


FIGURE 2.3: Density plot of peak frequencies at station ILL12 for both 2017 and 2018. The rectangles mark a discrete noise band at about 15 Hz. The noise-source stops in June for both years. We suggest this to be either related to snow melt or to a change in nearby hydropower operations.

2.4 PEAK FREQUENCY MIGRATION

For all 2017 and 2018 debris flows we computed the frequency spectrum for signal bins of 5 s with an overlap of 2.5 s. Figure 2.2b shows a close up of the initiation of the debris flow on August 8, 2018, with Figure 2.2d showing the corresponding color-coded spectra. As expected, the spectra show a peak migration from about 5.5 Hz towards higher frequencies for the flow front approaching station ILL12. Figure 2.4a shows the peak frequencies at each time step for all recorded events. As a result of the broadband nature of the debris-flow seismograms, peak frequencies are difficult to identify in conventional spectrograms (Fig. 2.2c). Only peak frequencies corresponding to a spectral power above $1x10^{-13}$ ms^{-2} Hz^{-1} were plotted, to eliminate low energy background noise caused by anthropogenic and discharge-related seismicity. In-torrent geophone-counts at CD1 of the August 8, 2018 event are shown additionally. The geophones, which measure the vertical velocity of ground vibration, start recording when the front of the debris flow reaches the check dam. The signal is represented as impulses that exceed a certain threshold (e.g., McArdell et al., 2007). In Figure 2.4a, all events were aligned according to the first recording of the geophone impulses, hence the arrival of the flow at CD1. The 48 m drop at CD1 leads to a shift in dominant frequencies from pre-event noise bands of 1 - 5 Hz and 15 Hz to about 5.5 Hz for all events and an overall increase in spectral power (Fig. 2.4a, dashed black line). Observations of flow over CD1 are not available, however, given that debris-flow velocities are relatively large, it is likely that the flow is largely detached from the face of the spillway and that it resembles free-fall conditions. Note that with the imposed energy threshold, the two small events of June 2018 do not emerge from the background noise. After the flow fronts pass CD1 and the 48 m free fall, the dominant frequencies rise to around 15 Hz when directly passing the station. After the initial rise, dominant frequencies shift between 12 and 25 Hz, but no time-dependent pattern can be observed (Fig. 2.4a, dotted black line). In contrast to the frequency domain, the signal of the free fall cannot be distinguished from background noise in the time domain as has been observed at other sites (Coviello et al., 2015; Schimmel et al., 2018). The evolution from low to higher peak frequencies can also be observed at station ILL11 and ILL13 (Appendix A.1). The stations have a minimum distance to the channel of 30 and 700 m, respectively. ILL18, which is closest to the initiation area of the flow, does not show a clear pattern in dominant frequencies. This can be explained by signal mixing of noise sources that are close to this station (i.e., precipitation and runoff concentration from the slopes, discharge, and 48 m channel step at CD1). For stations farther away from the channel, no pattern in dominant frequencies can be observed either. Specifically, stations that are farther away than 1 km from the channel (ILL14, 15, 16, 17) do not capture the peak frequency migration. For these stations, the 48 m free fall at CD1 may be too far to generate a signal that dominates over other seismic sources at the head of the trunk channel, closer to the stations.

2.5 DEBRIS-FLOW CHARACTERISTICS FROM SEISMIC SIGNALS

Next, we investigate the origin of the frequency signature of the flows in order to connect the findings with the debris-flow characteristics. The model by Lai et al. (2018), predicts

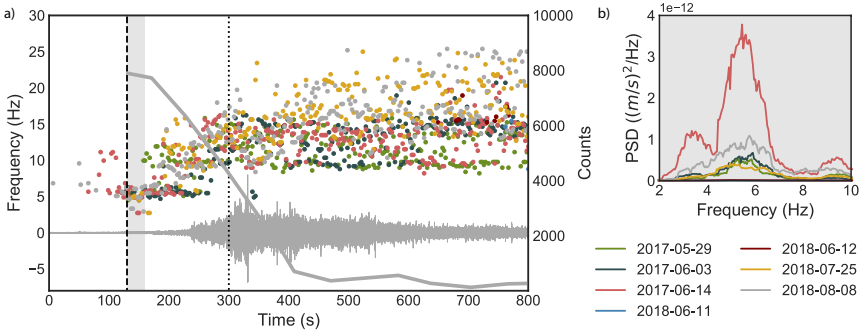


FIGURE 2.4: Peak frequencies for all events in 2017 and 2018, binned for time steps of 2.5 s (dots), together with geophone counts recorded at CD1 for August 8, 2018 event (grey line). The grey signal shows the waveform of the August 8, 2018 event recorded on station ILL12. The events were aligned according to the onset of the geophone measurements at CD1. The peak frequencies are plotted for power spectral densities that exceed $1e10^{-13}$ $(ms^{-1})^2\,Hz^{-1}$. The shaded area shows the time span used to calculate the spectra in (b). The dashed line marks the onset of the geophone recordings at CD1 and the corresponding frequency drop to 5.5 Hz. The dotted line marks the approximate passage of the debris flow at ILL12 which corresponds to a change in frequency pattern. Note the increase in dominant frequencies after the passage of CD1 until passing station ILL12. (b) Spectra of all events 30 seconds after passing CD1. The spectra are smoothed with a 1 Hz running average. The June 14, 2017 event clearly shows the highest power spectral density at that time.

the following power spectral density P as a function of frequency f of the seismic signal generated by the boulder-rich debris-flow front:

$$P \sim 1.9LWD^3u^3 \cdot \frac{f^{3+5\xi}}{v_c^5 r_0} e^{-\frac{8.8f^{1+\xi}r_0}{v_c Q}} \tag{2.1}$$

Using values of 500±100 m/s for Rayleigh-wave phase velocity v_c, 60±10 for the quality factor Q, and 0.417±0.05 for ξ (Tsai et al., 2012; Burtin et al., 2014; Burtin et al., 2016; Lai et al., 2018), equation (2.1) places the sources of the two dominant frequency bands of 5.5 Hz and 15 Hz at distances of about 1100±500 m and 300±150 m from ILL12, respectively. This corresponds to the distance between (a) the station and CD1, and (b) the smallest distance between the station and the channel, which lies for both cases within the uncertainties. Due to model simplifications and poorly constrained site effects, the distance is only a rough estimate. Nevertheless, it suggests that the main sources for the seismic waves recorded at ILL12 are the flow over the spillway at CD1 and the processes that happen at small distances from the station. Lai et al. (2018) use the continuous increase in peak frequencies to estimate flow velocity. At Illgraben, the two distinct dominant sources inhibit estimation of flow velocity, as the jump in frequency reflects the time when the dominance of the two signals changes, rather than a location of the flow front. Next, we compare the different event spectra recorded at Station ILL12 when the flow fronts pass CD1 (Fig. 2.4b). The spectra are computed over 30 s, and a running average of 1 Hz was applied. We assume that at these time intervals, the signals are dominated by seismic waves generated by boulder impacts at the base of the 48 m free fall at CD1. At this moment, projectile motion of the boulders implies that vertical flow (impact) velocities are the same for all events. Length of the front and width are therefore expected to be similar, assuming that the flow front impact lasts longer than the 30 s-long time window used to calculate the spectra. To only concentrate on the 5.5 Hz signals generated at CD1, the signal was filtered between 3 and 10 Hz. According to Equation (1), differences in spectra, in particular spectral amplitudes, are then attributable to grain size differences. The largest amplitude of power spectral density is generated by the event on June 14, 2017. Unfortunately, no video material is available for this event since it happened at night. Therefore, a cross validation to check whether the largest power is generated by larger boulder size in the debris-flow front, is not possible. When comparing the debris flow on May 29, 2017 with other events, one can observe that the energies are similar to the events on July 25 and August 8 in 2018. The videos indicate the presence of large boulders in all three flows, but a quantitative statement on the 94th percentile of the particle sizes is difficult (Fig. 2.5). The small events in June 2018 show the lowest spectral energies, which suggests that particle sizes for these flows are substantially smaller than for other events. Analysis of the videos for the 2018 events has not yet been completed, but preliminary analysis suggests that these events might be debris floods rather than debris flows (Fig. 2.5).

2.6 CONCLUSION

The source mechanism of seismic signals of debris flows is strongly dependent on particle sizes and topographic features of the flow path. Seismic studies of debris-flow dynamics and material composition should therefore not focus on the entire (temporally averaged) debris-flow seismograms, but on individual time windows. The spectral content of the



FIGURE 2.5: Pictures of five debris flows at check dam 29 at the lowermost part of the torrent. Pictures of the events of June 3 and June 14, 2017 do not give any insight on particle sizes, as they happened during night time. The pictures in the first row show the large event on May 29, 2017 (left), and the event on August 8, 2017 (right) that has been used as an example in Fig. 2.

signal can be dominated by large topographical features, such as the free fall behind a check dam in our case.

Large uncertainties of the seismic properties of the torrent catchment's sub-surface inhibit accurate models of the seismic propagation between flow front and recording unit. This complicates the interpretation of seismograms in terms of flow properties.

ACKNOWLEDGEMENTS

The project is funded by WSL's strategic initiative Climate Change Impacts on Alpine Mass Movements (CCAMM). FW's salary was funded by the Swiss National Science Foundation (GlaHMSeis Project PP00P2_157551). The municipally of Leuk and Anniviers generously provided permission and assistance for the seismometer deployment. We thank the Swiss Seismological Service who supported the project with seismic stations, real time data transfer and archiving facilities as well as manpower in the field. Alexandre Badoux and Christoph Graf assisted in the field campaign planning. We also thank the VAW and WSL personal, especially Martin Funk, for technical and scientific input and support in the field. The authors are grateful to Kate Allstadt and Velio Coviello for detailed and constructive reviews.

3

SEISMOMETER RECORDS OF GROUND TILT INDUCED BY DEBRIS FLOWS

This chapter is submitted for peer-review to the *Seismological Society of America* as:
Wenner, M., Allstadt, L., Weston, T., Lockhart, A., Hirschberg, J., McArdell, B.W. & Walter, F., (2021). Ground tilt caused by debris flows recorded on seismometers: A study on different scales. *Bulletin of the Seismological Society of America*.

ABSTRACT

A change in surface loading causes the earth's surface to deform. Mass movements, such as debris flows, can cause a tilt large enough to be recorded by nearby instruments, but the signal is strongly dependent on the mass loading and sub-surface parameters. Specifically designed sensors for such measurements (tiltmeters) are cumbersome to install. Alternatively, broadband seismometers record, in addition to ground motion, tilt signals at periods of tens to hundreds of seconds, with the horizontal components most sensitive to tilt. In this study, we show how to obtain tilt caused by the passing-by of debris flow from seismic measurements recorded within tens of meters of the flow and investigate the usefulness of this signal for flow characterization. We investigate the problem on three scales (1) large-scale laboratory experiments at the U.S. Geological Survey debris-flow flume where broadband seismometers and tiltmeters were installed for six 8-10 m^3 experiments, (2) the Illgraben torrent in Switzerland, one of the most active mass wasting sites in the European Alps, where a broadband seismometer placed within a few meters of the channel recorded 15 debris-flow events and (3) Volcán de Fuego, Guatemala, where a broadband seismometer recorded two lahars. We investigate how the tilt signals compare to debris-flow parameters such as mean normal stresses, usually measured by expensive force plates, and debris-flow height. We model the elastic ground deformation as the response of a layered elastic half-space to a moving surface load. Additionally, we use the model with some simplifications to determine maximum debris flow heights of Volcán de Fuego events, where no force-plate measurements are available. Finally, we address how and under what assumptions the relatively affordable and straightforward tilt measurements may be utilized to infer debris-flow parameters, as opposed to force plates and other complicated instrument setups.

3.1 INTRODUCTION

When large mass movements, such as debris-flows, lahars, and rock avalanches, occur and propagate towards lower elevation, some of the potential energy is converted into seismic energy. On one hand, particle impacts within the moving mass and their interaction with the ground generate high-frequency seismic signals (> 1 Hz) which are recorded up to hundreds of kilometers away from the mass-movement event (e.g., Suriñach et al.,

2005; Deparis et al., 2008; Allstadt et al., 2018). On the other hand, for rapid, massive (millions of m^3) landslides, the earth's response to acceleration and deceleration of the mass can be recorded in the long period seismic signal (> 10 s) up to many thousands of kilometers away (e.g., Kanamori & Given, 1982). In this very long period range (VLP), but significantly closer to the events (tens to thousands of meters), mass movements cause ground tilting. Mass-movement events exert stress on the earth's surface and cause a localized compression of the sub-surface that changes with time as the mass is redistributed. The earth's elastic response causes surrounding areas to tilt towards the compressed area.

Instruments specifically designed to record ground tilt are called tiltmeters. However, tiltmeters can be laborious to install while individual site effects vary widely and impact a tilt meter's ability to record signals of interest. Alternatively, broadband seismometers have been shown to record tilt in their low frequency response. Rodgers (1968) derive the response of the horizontal pendulum seismometer to ground tilt. Similarly, Graizer (2006) take the three-component (vertical Z, North-South N, East-West E) sensitivity of the seismometer to tilt and rotations into account and consider the full equation of pendulum motion. The author finds that the vertical component is only sensitive to translational ground motion, while the horizontal components are sensitive to translational and rotational ground motion including tilt. They conclude that if the spectrum in long periods is similar for the seismometer's vertical and horizontal components, the signal is dominated by translational ground motion and thus tilt can be neglected.

In volcanology, earth surface tilting is an often-observed phenomenon resulting from deformations caused by, for example, magma chamber expansion or collapse, or dike intrusions (e.g., Wielandt & Forbriger, 1999; Aoyama & Oshima, 2008; Genco & Ripepe, 2010; Lyons et al., 2012; Fontaine et al., 2014; Šindija et al., 2021). Wielandt and Forbriger (1999) use the tilt recorded by seismometers to invert for magma chamber location and volume expansion estimation focusing at long periods between 10 - 120 s. Aoyama and Oshima (2008) discover tilt caused by a volcanic crater subsidence at periods (> 60 s) below the corner frequency of their instrument. Chouet and Dawson (2013) finds tilt in VLP signals generated by rockfalls impacting a lava column. Beyond volcano-seismology, Anthony et al. (2018) suggest that at frequencies below 1 Hz, the seismic signals they recorded within 1 m of a river were actually dominated by oscillations of ground tilt. This tilt was due to viscoelastic deformation of the ground resulting from turbulent eddies on the banks, which induced rapid stress and pressure variations.

In this study, we investigate tilt caused by mass movements, specifically debris flows recorded on instruments within a few tens of meters or less of the channel. We discuss (1) what the tilt signal can tell us about debris-flow dynamics and material properties and (2) how useful the signals are for monitoring and early warning purposes. To answer these questions, we consider three different scales: (1) a controlled experimental setup at the U.S. Geological Survey (USGS) debris-flow flume with experiment volumes of 8 - 10 m^3, (2) a debris flow channel in Switzerland, with debris flow volumes up to $\sim 10^5$ m^3 and (3) two large-scale lahar events at Volcán de Fuego, Guatemala with sparse data coverage.

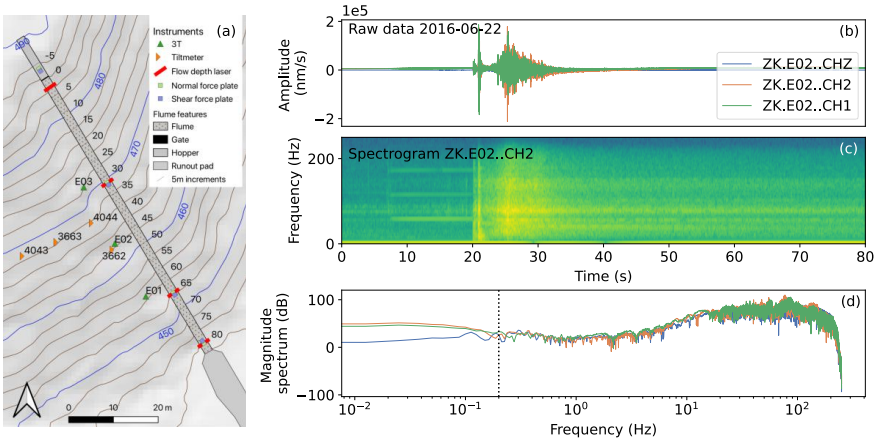


FIGURE 3.1: (a) Setup of 2016 debris-flow experiments at the USGS flume. Green triangles denote the locations of three broadband seismometers installed at a distance of 4.5 m to the edge of the flume. The orange triangles mark the location of the tiltmeters; we primarily use data from the one co-located with station E02 (3662). (b) Raw data of a debris-flow experiment on June 22, 2016, recorded at station E02. The initial pair of spikes around 20 s is from the gate opening and then hitting the flume wall (c) Spectrogram of channel 2. (d) Spectrum of all channels (vertical (Z) and two horizontals (CH1, CH2) of the June 22, 2016 experiment. The black dotted line at a period of 5 s marks the frequency, below which the horizontal components show larger energy than the vertical component, thus suggesting most of the signal is tilt.

3.2 STUDY SITES

We first focus on two different scales of ground tilt induced by debris flow passage for which we have high-quality data: (1) 6 USGS debris-flow flume experiments and (2) 15 natural debris flows at Illgraben, Switzerland. We choose these two sites for our initial considerations, as the events are well recorded with several independent measurements of bulk-debris-flow parameters. Later, we apply what we learned at these sites to a third site for which little independent data is available.

3.2.1 *USGS debris-flow flume*

The USGS debris-flow flume is an instrumented 95-m-long, 2-m-wide concrete flume built into a hillside that is used to conduct large-scale laboratory experiments on debris flows (Iverson et al., 2010). In 2016, six gate-release experiments, in which 8 to 10 m^3 of saturated debris is released from behind a gate, were conducted with seismic instrumentation and tiltmeters installed along with the typical flume instrumentation including laser flow-height sensors, force plates, and pore pressure sensors (Allstadt, 2016; Allstadt et al., 2019; Iverson & George, 2019; Iverson & Logan, 2017). Figure 3.1a shows the instrumental setup of the 2016 experiments that are relevant to this study. The

TABLE 3.1: Volumes and description of the 2016 debris-flow experiments at the USGS debris-flow flume

Date	Volume (m^3)	Peak Frontal Velocity (m^3)	Description
2016-06-14	8	8.7	Debris flow over bare bed, sand/gravel mix
2016-06-15	8	10.8	Debris flow over bare bed, sand/gravel mix
2016-06-16	8	9.3	Debris flow over bare bed, sand/gravel mix with large rocks
2016-06-21	8	10.3	Debris flow over bare bed, sand/gravel mix with large rocks
2016-06-22	10	12.9	Debris flow over bare bed, sand/gravel mix
2016-06-23	10	11.2	Debris flow over bare bed, sand/gravel mix

three broadband seismometers (Guralp CMG3T 120 s; E03, E02, and E01) were installed about 4.5 m from the flume edge in small 0.5-m deep vaults built out of buried Gatorade coolers emplaced on concrete pads. The horizontal channels (CH1 and CH2) are rotated at an azimuth of 350 degrees (E03) and 355 degrees (E02, E01) towards north and east. Additionally, four tilt sensors (Applied Geomechanics Model 712 dual-axis platform tiltmeters) were installed, one of which (3662) is located close to the seismic station E02. The tiltmeters were installed on concrete pedestals poured over 2-m long rebar that was hammered into the ground to improve coupling. The flume instrumental setup that is relevant to this paper includes several force plate installations and collocated flow height sensors (see Iverson et al. (2010) for instrumentation details). Two of these measurement units are located at the closest point of the flume to seismic stations E03 and E01.

By design, the 2016 experiments consisted of three replicate pairs of variations in debris composition and volume. The flows are more rapid and flow further for the larger (10 m^3) experiments (2016-06-22 and 2016-06-23) mainly because the geometry of the debris prism behind the gate is by necessity steeper and is such that a pond of 0.3 m^3 of water forms behind the debris. In contrast, the smaller experiments (8 m^3) stall partway down the flume until the flow front is overtaken by a subsequent surges. Videos of the experiments are available at https://pubs.usgs.gov/of/2007/1315/ (Logan et al., 2018). The experiments are summarized in Table 3.1. We refer the reader to Allstadt et al. (2019) and Iverson and George (2019) for more details on the experiments.

3.2.2 *Illgraben, Switzerland*

The Illgraben catchment, located in the Swiss south-western canton of Valais, is one of the most active mass wasting sites in the European Alps (e.g., Schlunegger et al., 2009). From its highest point, just below the Illhorn (2716 m asl) to its confluence with the Rhône river (610 m asl) the catchment spans around 9.5 km^2 and consists of two sub-catchments:

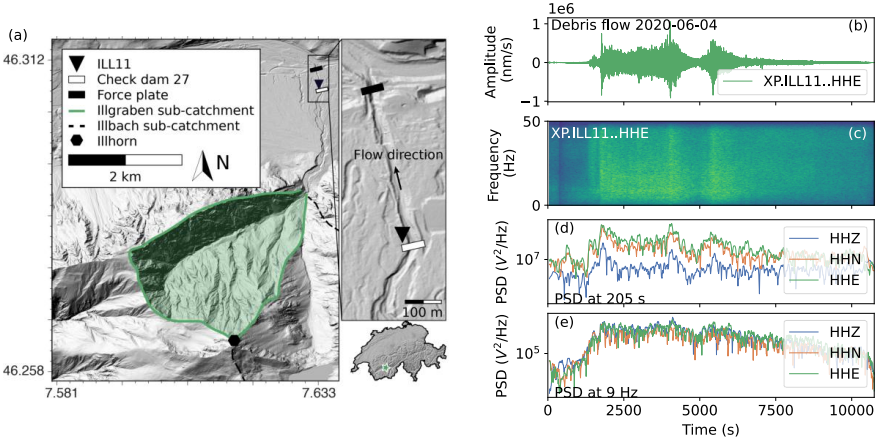


FIGURE 3.2: (a) Illgraben catchment in southwest Switzerland. The inset panel shows a zoom-in on the debris-flow channel. In the lower right corner, a map of Switzerland is shown, with the green star denoting the location of Illgraben. The green shaded area shows the Illgraben catchment, whereas the black dashed line shows parts of the Illbach catchment. The black hexagon denotes Illhorn (2716 m asl). The black triangle denotes the location of the broadband seismometer (ILL11) located 19 m away from the centerline of the debris-flow channel. Debris flows propagate from south to north, where they discharge into the Rhone river. The black square denotes the location of the force plate (620 m asl) installed 450 m downstream from station ILL11. The white square denotes check dam 27, which is located just upstream of the broadband seismometer (b) Three hours of seismic recordings of a debris flow on June 4, 2020 (east component). (c) Spectrogram of the east component of the seismic recordings shown in (b). (d), (e) Spectral density of all components in two different frequency bands: 205 s (d), 9 Hz (e).

Illgraben and Illbach (e.g., McArdell et al., 2007; Hirschberg et al., 2021). From these two sub-catchments, only the Illgraben (4.8 km^2) produces debris flows, mainly triggered by short-duration high-intensity or long-lasting rainfall events mobilizing loose sediment at the head of the channel (Hürlimann et al., 2003; Hirschberg et al., 2021). A large fan with a radius of about 2 km and a volume of roughly 500×10^6 m^3 was created by past activity inside the catchment (Hürlimann et al., 2003; McArdell et al., 2007). In 1961, a large rock avalanche (3×10^6 m^3) at the northern Illhorn slope deposited a significant amount of sediment in the Illgraben trunk channel from where several debris flows were triggered (Hürlimann et al., 2003). Authorities reacted with the construction of 30 check dams to minimize channel erosion and reduce the risk for the neighboring village of Leuk. In 2000, the Swiss Federal Research Institute for Forest, Snow and Landscape (WSL) installed an early warning system, including rain gauges, radar, laser and ultrasonic flow stage measurements, and geophones (Badoux et al., 2009). Additionally, at the confluence of the Illgraben channel and the Rhône river, a 2 m × 4 m force plate was installed to measure normal and shear forces at the base of debris flows (McArdell et al., 2007). This force plate was destroyed during a debris flow with exceptionally large boulders in 2016 and rebuilt in 2018. The new force plate has been operational since 2019. Since the start of the recordings, 3-5 debris flows per year have been observed on average (Schlunegger et al., 2009). However, since 2018, there has been an increase in debris-flow activity, with up to 10 events per year (Chmiel et al., 2021).

A campaign seismic network of eight seismic stations has been deployed annually within and around the Illgraben canyon during the debris-flow season since 2017 (e.g., Wenner et al., 2019). The network has been used for event detection, and the study of high-frequency (> 1 Hz) seismogenesis during debris flows (Chmiel et al., 2021; Meyer et al., 2021; Zhang et al., 2021). Since 2018, one out of the eight stations is a broadband seismometer (Trillium compact 120 s), located about 15 m away from the edge of the 8 m wide debris-flow channel (ILL11, black triangle Fig. 3.2a). Most other stations are equipped with a Lennartz-1s sensor. We use data from the broadband seismometer (ILL11) to investigate tilt generated by the passing of the debris flows. To compare the observed tilt amplitudes and waveforms, we restrict our study to years during which the force plate was also operating, i.e., 2019 and 2020. The force plate is located about 450 m downstream from the broadband seismometer.

Figures 3.2b and 3.2c show the raw seismic data of the June 4, 2020 event and its spectrogram, recorded on the east component, which is approximately perpendicular to the flow direction. Figure 3.2d and 3.2e show the spectral power of all three components at two different frequencies: 205 s and 9 Hz. We can see that when the debris flow passes the station, at 9 Hz the power spectral density (PSD) increases on all three components. However, at a period of 205 s, the increase of PSD is significantly higher on the horizontal components than the vertical. This increase is associated with ground tilt.

3.3 METHODS

The design of broadband seismometers allows them to record both translational and rotational ground motion with the latter including tilt. While generally seismologists tend to ignore and eliminate signals observed from tilt (e.g., Crawford & Webb, 2000; Šindija et al., 2021) we make use of such signals to estimate debris-flow parameters (e.g., Fig. 3.3a). Signals recorded on seismic sensors are a convolution of the source

time function of the seismic wave, the Green's function describing the impulse response between source and seismic station, and the response of the instrument. Generally, seismometers have a flat response to ground velocity in the frequency band between their natural frequency, also known as the lower corner frequency (f_c), and their upper corner frequency. However, depending on the type of tilt we want to observe (size and duration of passing debris-flow fronts), the frequency band of interest may be significantly lower than the lower corner frequency. Therefore, it is crucial to discriminate between two cases: tilt observed at frequencies above the instrument's lower corner frequency and tilt observed at frequencies below the instrument's lower corner frequency. In the following, we will describe the methods used to obtain tilt from seismometers in both cases.

3.3.1 *Tilts at frequencies above instrument low-corner frequency*

While the vertical component of the seismometer is predominantly sensitive to vertical ground acceleration, the horizontal components are sensitive to both displacement and tilt τ: $a_x(t) = \ddot{u}_x(t) - g\tau(t)$ where t is time, g is the gravitational acceleration, $a_x(t)$ is the apparent horizontal ground acceleration measured by the instrument and $u_x(t)$ is horizontal ground displacement (Wielandt & Forbriger, 1999) while two dots represent the second time derivative. The seismic signals and frequency analysis of the USGS flume experiments shown in Figure 3.1b-d helps us to define the frequency range of interest: (1) the seismic signal of an experiment lasts for about 20 s, therefore we can assume, that any signal below the corner frequency of the instrument (120 s) is not generated by the dynamic tilt due to the passing of the debris flow (Fig 3.1b), (2) the magnitude spectrum in Figure 3.1d clearly shows a divergence between the vertical (blue line) and horizontal (green and orange line) components below a period of 5 s (black dotted line). Following Graizer (2006), we choose the frequency of 1/5 Hz as characteristic frequency below which the magnitude spectrum is significantly higher on the horizontal components compared to the vertical component. The filtered signal between 1/120 Hz and 1/5 Hz is therefore assumed to be proportional to the tilt. Hence, in this frequency band, we ignore potential ground acceleration from translational ground motion, leading to the expected acceleration observation $a_x(t) = -g\tau(t)$. For this frequency band, we calculate the tilt from the recorded ground velocity data by (1) removing the mean, bandpass filtering the data between 120 s and 5 s (2) removing the instrument response and taking the derivative of the velocity output to obtain acceleration, (3) rotating the horizontal components so that the radial component points along the shortest distance to the debris-flow channel and (4) dividing the acceleration time series by $-g$.

3.3.2 *Tilts at frequencies below instrument low-corner frequency*

Following Rodgers (1968),Aoyama and Oshima (2008), Genco and Ripepe (2010), and Lyons et al. (2012) the instrument response of a seismometer to ground tilt $A_\tau(\omega)$ can be expressed as a function of angular frequency (ω) multiplied by the instrument response to ground displacement $A_u(\omega)$ and negative ground acceleration g:

$$A_\tau(\omega) = -\frac{g}{\omega^2} A_u(\omega) \tag{3.1}$$

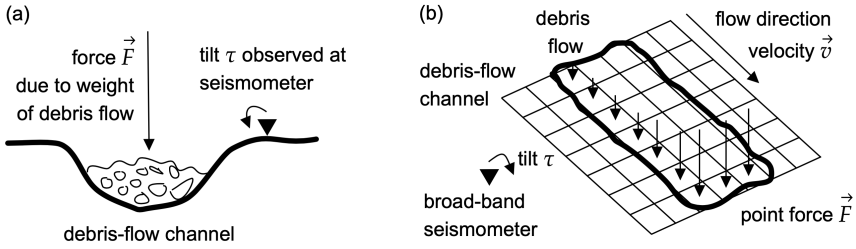


FIGURE 3.3: (a) Illustration of observed tilt due to the passing of a debris flow next to a broadband seismometer. The weight of the debris flow applies a force normal to the ground surface compressing it below the weight. This causes the seismometer to tilt over an angle, which depends on the distance to the debris flow. (b) Geometry for modeling ground tilt due to a force distribution, such as the weight of a moving debris flow. The area of the distributed mass is divided into grid cells. On each cell, a point force is applied on the channel surface. Tilt at the seismometer is then calculated for each cell at a distance to the grid cell. Following the principle of superposition, the responses to all force cells are summed up to obtain the overall tilt.

In contrast to the flat transfer function of a broadband seismometer to translational ground velocity above f_c, the instrument response to tilt is flat below f_c (Genco & Ripepe, 2010). Therefore, if we only use data from below the corner frequency and assume it features tilt only, the tilt time series $\tau(t)$ can be obtained by integrating the digitizer output $p(t)$ from velocity to displacement (equivalent to dividing by $-\omega^2$), multiplying it by the seismometer sensitivity S times $(2\pi f_c)^2$ and dividing it by the negative gravitational acceleration. The seismometer sensitivity value thereby represents the relation between digitizer output $p(t)$ and real ground velocity in the flat part of the transfer function (for a Trillium compact 120 s sensor and a Nanometrics Taurus digitizer: $S = 301696000$ Counts / unit of ground motion). Expressed in a single equation,

$$\tau(t) = -\frac{S(2\pi f_c)^2}{g} \int p(t)dt \tag{3.2}$$

For the frequency band below f_c, we thus derive the tilt from the seismic data by (1) rotating the horizontal components so that the radial component points along the shortest distance to the debris-flow channel, (2) integrating the digitizer output signal, (3) bandpass filtering the data between 2 hrs and 120 s using a two-corner butterworth zero-phase filter for optimal removal of noise and trends in the data, and minimizing filtering artifacts, and (4) multiplying the time series by $-\frac{S(2\pi f_c)^2}{g}$.

3.3.3 *Modelling tilt of a distributed moving mass*

We model the tilt response to a passing debris-flow at a distance d with the solution to the Boussinesq problem (e.g., Ling et al., 2002) describing the surface deformation of an

elastic half-space due to a concentrated vertical force. The vertical displacement u_z as a function of distance r from the point force F can be described as

$$u_z = \frac{F}{4\mu\pi}\left\{\frac{z^2}{r^3} + \frac{(3-4\nu)}{r} - \frac{(1-2\nu)}{r+z}(1+\frac{z}{r})\right\} \tag{3.3}$$

with shear modulus μ, Poisson's ratio ν, and vertical location z at distance r with respect to the surface. $r = \sqrt{x^2+y^2+z^2}$ represents the distance of the applied point force to the sensor with x,y being the horizontal distance and z being the depth. To estimate the elastic properties of the ground for the USGS site, which is located on altered tuffs and breccias, we used seismic velocity data from an active seismic experiment conducted along a seismic line perpendicular to the flume starting near tiltmeter 4044. With a thickness of 2 m, the upper layer of soil has a p-wave velocity (v_p) and s-wave velocity (v_s) of 220 m/s and 150 m/s respectively. Just underneath, a 4 m layer of weathered bedrock has seismic velocities of $v_p = 900m/s$ and $v_s = 600m/s$. We assume an uncertainty of 10 %. It is unclear whether the upper soil layer or the underlying weathered material is most influential on the tilt recorded by the seismometer. We therefore evaluate the results for uniform halfspaces consisting of each one of the two ground types as end member cases. We consider values between 1200 and 1750 kg/m^3 for soil density ρ (Weil & Brady, 2019) and 1800 to 2400 kg/m^3 for the bedrock, which are typical values for welded tuff (Wohletz & Heiken, 1992). We obtain the shear modulus μ and p-wave modulus M by multiplying v_s and v_p with ρ. The Poisson's ratio is then given by $\nu = (M-2\mu)/(2M-2\mu)$. For an error estimate we perform one hundred model runs with density, v_p and v_s randomly varying within the parameter range defined above.

Ground properties for the Illgraben debris-flow fan are not available. Therefore, we use values found by Salem (2000) for surface soils and shallow sediments as an estimate (Elastic modulus E: 0.3-41.9 GPa, ν: -0.15-0.47, ρ: 1600-2100 kg/m^3). Similar to the USGS site, we randomly vary parameters for ρ, E and ν one hundred times within these ranges to obtain a reasonable range of tilts at Illgraben.

Depending on the event scale, debris flow masses can be distributed over meters to kilometers along a channel and across different channel widths. This invalidates the point force assumption and therefore, we adjust the point force model to a model of distributed load by using superpositions of calculated vertical displacements (Domınguez, 1966). We divide the debris-flow channel surface into n grid cells of 1 m^2 (Fig. 3.3b). We then take the force recorded at (1) the collocated force plate at USGS site (Fig. 3.1a) and (2) the 450 m downstream located force plate at Illgraben (Fig. 3.2a). We do not account for the difference in along-stream location between the seismometer and the force plate although we cannot exclude the possibility that flow dynamics can change significantly over such distances. We convert the force plate time series of vertical forces into a spatial distribution with one sample per meter using the velocity of the debris flow at the time of the passage next to the seismometer. Furthermore, we assume that the force is the same over the whole channel cross-section. We then calculate $u_{zi}(r)$ for each grid cell i at a distance r_i to the seismometer. The superposition principle allows us to obtain the vertical deformation at the seismometer ($u_z(r)$) from the sum of deformations caused by the point forces on each grid cell:

$$u_z(r) = \sum_{i=1}^{n} u_{zi}(r) \tag{3.4}$$

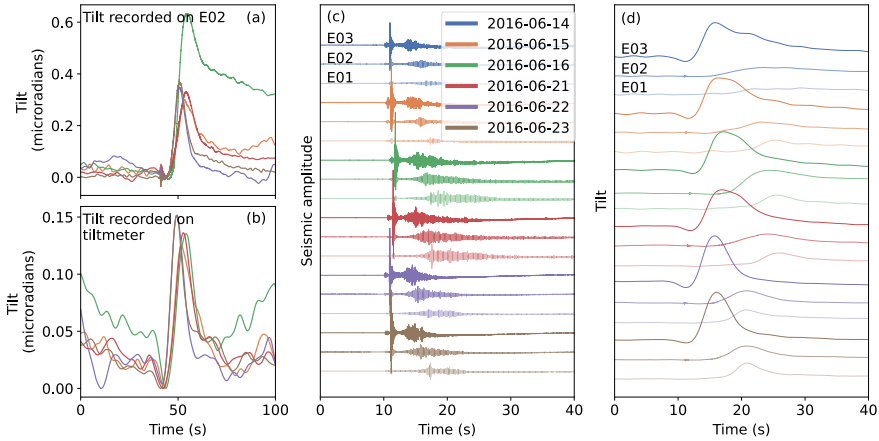


FIGURE 3.4: (a) Tilt obtained from seismometer recordings at the USGS seismic station E02. (b) Tilt recorded on collocated tiltmeter. Raw data (c) and tilt on the radial component (d) recorded for all 2016 experiments on all broadband seismometers (E03, E02, E01). The colors denote different experiments, whereas the transparency marks the recordings at different stations. The amplitudes are plotted as true amplitudes and are not normalized.

Additionally, we assume that while the debris flow passes the seismic station, the force profile of the flow does not change. We obtain a time series of the tilt by letting the force grid pass the station with the measured flow front velocity. The velocity of the debris-flow flume events was estimated from the videos. The velocity of Illgraben flows was estimated by the arrival of the debris-flow front recorded on geophones at two consecutive check dams downstream of station ILL11 (Badoux et al., 2009).

3.4 RESULTS

3.4.1 *USGS debris-flow flume, Oregon, USA*

Figure 3.4a shows the tilt signals of the five 2016 experiments in recorded on seismic station E02. For comparison Figure 3.4b shows the tilt signals recorded by the tiltmeter collocated with E02 for the same experiments. The June 14, 2016 record is not shown, as data from the tiltmeter are not available for this experiment. For both instruments, the larger volume experiments (June 22 and June 23, Fig 3.4a,b pink and brown line respectively) show a larger tilt than the lower volume experiments. Out of the three lower volume experiments, two of the recordings show good correlation in amplitudes relative to each other and the larger volume ones. However, the June 16 experiment (Fig 3.4a,b green line) shows a significantly higher amplitude on the tilt signal derived from the seismometer. Additionally, the tilt remains at about 0.4 microradians for this event after the flow passes the seismometer, whereas the tilt returns to zero for the other events. On average, the amplitudes of the tilt derived from the seismometer are 2.3 times higher than

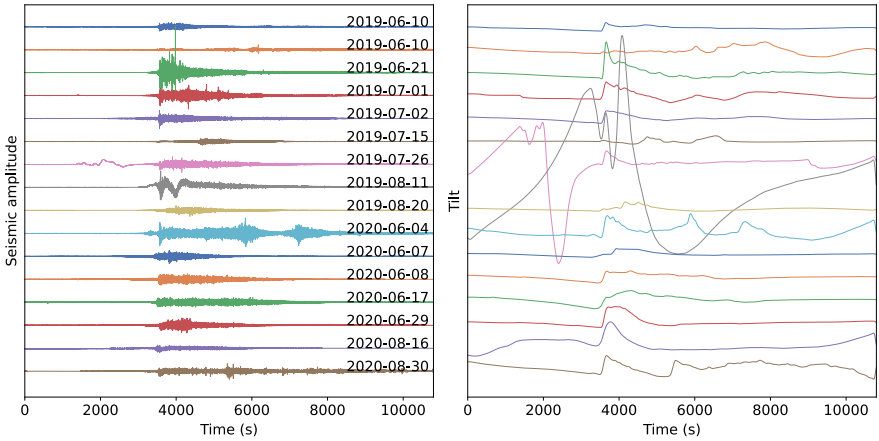


FIGURE 3.5: Raw data (a) and tilt measured on the radial component (b) of all debris-flows recorded at Illgraben in 2019 and 2020 on station ILL11 and the force plate. The colors denote different events. The events were aligned after the first peak in tilt observed at station ILL11. The amplitudes of the raw and the tilt signal are not normalized.

those recorded by the tiltmeter. Figures 3.4c and 3.4d show the raw data and derived tilt signals of experiments on seismic stations E01, E02, and E03. Different colors denote the different experiments, whereas the different transparency of the color denotes the station it was recorded on. The debris propagation downwards from station E03 to E01 is clearly seen by the later arrivals at the lower stations. The amplitudes in both Figure 3.4c and 3.4d are plotted as true amplitudes, i.e., the amplitudes are not normalized per track. The amplitude of the raw signal, as well as the tilt, decreases towards the lower station. However, the decrease in amplitude of the tilt is significantly larger than for the raw seismic signal, which is dominated by high-frequency energy. Whereas the June 16, 2016 event is an outlier in terms of amplitude on station E02, its amplitudes on station E03 and E01 are similar to the other experiments (Fig. 3.4d).

3.4.2 *Illgraben, Switzerland*

Figures 3.5a and 3.5b show the raw seismic data and derived tilt signals of all events recorded by both the seismometer and the force plate in 2019 and 2020 (in total 15 events). The amplitudes of the raw signal and the tilts are not normalized. We can see a clear tilt signal for all events, except the July 26, 2019, and August 11, 2019. For some events, such as the June 4, 2020 event, we can see multiple pulses of debris passing the station in both the high frequency and the tilt data. The maximum tilts generally coincide with times of maximum amplitudes at higher frequencies (> 1 Hz). In contrast, the tilt amplitude does not necessarily correlate with the amplitude in the higher frequencies. For example, the June 29, 2019 event clearly shows the highest amplitudes in the raw data and the tilt; however, the August 30, 2020 data shows relatively high amplitudes

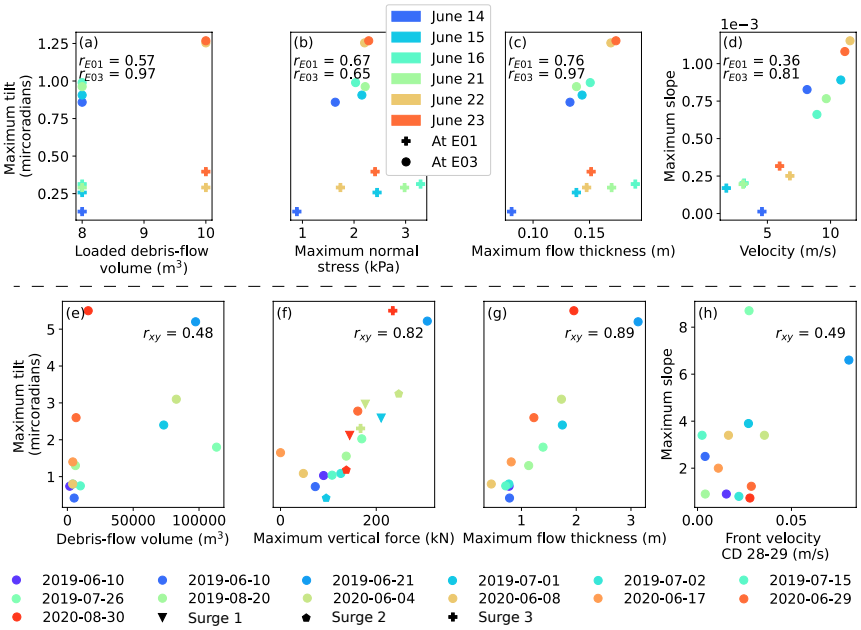


FIGURE 3.6: Panels (a), (b), (c), and (d) refer to data collected at the USGS debris-flow flume. Panels (e), (f), (g), and (h) refer to data collected at Illgraben. Correlation of maximum recorded tilt with debris-flow volume (a), maximum normal stress (b) and maximum flow thickness (c) at stations E01 and E03 at the USGS debris-flow flume. Correlation of the slope of the rise in tilt while the flow is approaching stations E01 and E03 and the velocity at the time of the passage (d). (e), (f), (g) Correlation of same parameters for Illgraben. r_{xy} gives the Pearson correlation coefficient of the parameters with maximum tilt. Different colors denote different experiments/events. Different symbols show (1) the correlations at the different stations (debris-flow flume) or (2) the correlations for different surges (Illgraben).

in the tilt signal compared to the other events but relatively small amplitudes in the high-frequency signal. For the July 26 and the August 11 events, a strong low-frequency signal is apparent even in the raw data. The tilt signal in the July 26 event is still visible in the data, although it is much smaller than the preceding low-frequency noise. For the August 11 event, the large amplitude of the low-frequency signal masks the tilt in the data. These low-frequency signals might arise from other noise sources close to the seismometer and will be discussed in the following section.

3.4.3 *Correlation with debris-flow parameters*

To assess if there is a correlation between the observed tilt and debris-flow parameters, we compare the maximum observed tilt with the debris-flow total volume, maximum normal stresses, and maximum flow thickness recorded at the USGS debris-flow flume (Fig. 3.6).

We use the loaded-debris volume of the experiment and data from the recording units, including force plates and flow depth sensors, smoothed with a 1s moving average. We choose to use the tilt recorded by the seismic stations E03 and E01 and their nearby recording unit, as there is no recording unit close to station E02. For station E03, the Pearson correlation coefficient (PCC) for maximum tilt and loaded-debris volume as well as maximum tilt and maximum flow thickness are both high with 0.97. In contrast, the correlation between maximum tilt and maximum normal stress is rather low with a PCC of 0.65 (Fig. 3.6a,b,c). The PCCs of station E01 are significantly lower for the correlation of the maximum tilt with the load debris-flow volume and flow height (0.57 and 0.76), whereas the PCC with maximum normal stress is slightly higher (0.67) compared to station E03.

We repeated the same parameter comparison for the Illgraben events, where the debris-flow volume is estimated following the approach of Schlunegger et al. (2009). Uncertainties in flow rheology may result in volume uncertainties of several percent; however, relative volume estimations should be precise. The maximum vertical force was measured at the force plate about 450 m downstream from the seismic station. We include maximum tilt and normal stresses from events with multiple surges as individual data points (i.e., 2019-07-01, 2020-06-04, 2020-08-30). The maximum flow depth is measured with a laser at the force plate. The PCC for the estimated debris-flow volume and the maximum tilt measured at the seismometer is 0.48 (Fig. 3.6e). The low PCC results largely from three events with a large estimated volume ($V > 70000\,m^3$) but low maximum tilt. For the correlation between the maximum tilt and maximum vertical force and between maximum tilt and flow thickness a PCC of 0.82 and 0.89 can be observed (Fig. 3.7f,g).

In Figure 3.4a,b, we can see that the larger and faster flume experiments of June 22, 2016, and June 23, 2016, show a narrower peak in the tilt signals and a higher slope of the peak between the debris flow arriving at the station and the maximum tilt. Figure 3.6d highlights this correlation for station E03 with a PCC of 0.81. For station E01, this correlation is less prominent; however, the June 14 event is the clear outlier for both cases. To assess if this correlation exists in the real-world case and the tilt signal can tell us something about the velocity of the debris-flow front, we inspect the slope of the tilt when the flow approaches the seismic station at Illgraben. Figure 3.6h shows the correlation of the estimated debris-flow front velocity and the slope of the tilt at the time the debris-flow arrives at station ILL11. The PCC of 0.49 illustrates that there is no clear correlation between these two parameters.

In addition to single value comparisons between the debris-flow parameters and the tilt signals, we also correlate the entire force plate time series with the tilt signals. Figure 3.7 shows the correlations for all events recorded at a force plate. For each time step, we plot the force recorded at the plate against the tilt signal. Cross-correlation coefficients (CC) between two signals show high correlations for the USGS experiments. CCs for the Illgraben events vary substantially between the events. We compute the CCs for both the entire flow (cc_{full}) and the first half of the flow (cc_{half}). For most events cc_{half} is higher than cc_{full}.

3.4.4 *Modelled tilt signal*

Figure 3.8a shows the tilt of the June 16 experiment recorded at station E03 at the USGS debris-flow flume (blue line). The tilt modeled with the distributed load is shown in

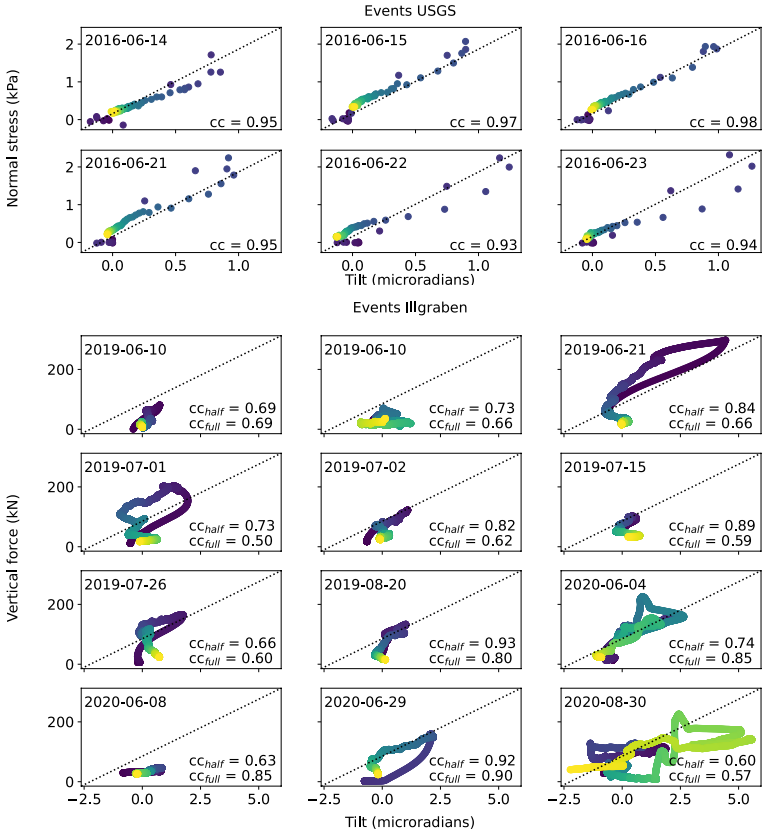


FIGURE 3.7: Correlation of forces recorded at force plates and tilt measured on the seismic station E03 at the debris-flow flume ("Events USGS") and at Illgraben ("Events Illgraben"). The correlation coefficient of the whole time series is shown in the lower right corner. For the Illgraben events, the correlation coefficient is shown for both the entire flow (cc_{full}) and first half of the flow (cc_{half}). The colormap represents the time of the plotted sample (from dark to light).

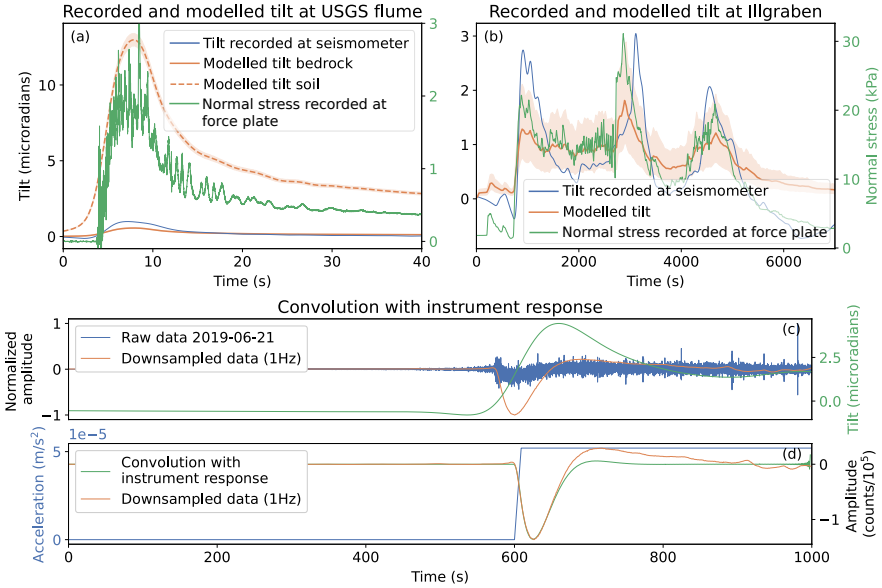


FIGURE 3.8: (a) Tilt recorded at station E03 at USGS debris-flow experiment of June 16, 2016 (blue), modelled tilt at station E03 with force history from collocated force plate as input (orange), vertical force recorded at force plate (green). The dashed orange line shows the mean of one hundred model runs with sub-surface parameters of soil. The solid line shows the mean of one hundred model runs with sub-surface parameters of weathered bedrock. The shaded orange area shows the 95 % confidence interval. (b) Tilt recorded at station ILL11 at Illgraben (blue), modelled tilt at station ILL11 with force history of the force plate 450 m downstream of station ILL11 (orange), vertical force recorded at the force plate (green) for the June 4, 2020 event. The solid line shows the mean of one hundred model runs with values from the parameter space found by Salem (2000). The shaded area shows the 95 % confidence interval. (c) Raw data (blue) and data lowpass-filtered below 1 Hz (orange) of June 21, 2019 event recorded at Illgraben. The green line shows the tilt obtained from the seismic data. (d) Numerical experiment for convolution of a step in acceleration (blue) with instrument response. The green line shows the convolution of the input step in acceleration and the instrument response. The orange line shows the downsampled data for comparison.

orange. The dashed line represents the mean tilt with soil parameters as input for the model. The orange shaded area shows the 95% confidence interval. The solid orange line shows the mean tilt obtained from weathered bedrock parameters. The model input, the normal stress time series, is plotted in green. The waveforms of the recorded and modeled tilt correlate well. For neither of the two cases (soil or bedrock), does the amplitude of the recorded tilt lie within the confidence interval. However, tilt obtained from the parameters of weathered bedrock is closer to the recorded tilt than the tilt obtained from soil parameters. Figure 3.8b shows the equivalent illustration for the June 4, 2020 debris flow at Illgraben. In contrast to the flume experiment, the amplitudes of the recorded and modeled tilt agree well. There is, however, a slight temporal offset between the recorded and the modeled tilt for the second peak. The input for the model is the vertical force history recorded at the force plate; hence, the peak of the force-time series and the modeled tilt correspond with each other.

Aside from the distributed load model, a simple approach to estimate sudden changes in acceleration is applied to the Illgraben data: When filtering the raw seismic data of the Illgraben data with a 1 Hz lowpass filter, a waveform emerges which resembles the instrument response of the seismometer to a step in acceleration (Fig. 3.8c, Wilson et al., 2017). We show this specific waveform for the example of the June 21, 2019 event. This "bump" is present for all recorded debris flows and sometimes even several times (for individual surges). We estimate the acceleration needed to generate such a bump in the raw seismic data by creating a step in acceleration of a specific amplitude and convolving this step with the instrument response. Figure 3.8d shows the step in acceleration with an amplitude of about 5 m/s^2 (blue line) which was then used for a convolution with the instrument response. The resulting signal (green line) correlates well with the observed signal at station ILL11 for the June 21, 2019 event (orange line).

3.5 DISCUSSION

3.5.1 *Correlation with debris-flow parameters*

3.5.1.1 *USGS debris-flow flume*

Figure 3.6 shows the correlation between the observed tilt at station E03 (dots) and E01 (crosses) and the bulk debris-flow parameters recorded at the nearby instrumentation unit. At the beginning of these large-scale experiments, when the mass of the debris flow is more concentrated, the correlation is particularly clear (at E03). At the time the flow arrives at E01 the correlation between the parameters is less pronounced. This can, at least partially, be attributed to the changing concentration of the debris flow as it becomes more elongated and agitated as it propagates down the flume and forms multiple surge fronts so the tilt signal may not be as clear. There also is a slight eastward tilt in the flume that causes the flows to migrate towards the eastern flume wall lower down. This means that the full flow thickness does not pass over the force plate or under the flow height sensor at the instrumental cross section near E01 (Logan et al., 2018), so the direct measurements are less representative of the conditions that may be causing the dynamic ground tilt at that location.

Furthermore, the tilt recordings on station E02 constitute an outlier in the June 16, 2016 experiment (Fig. 3.4a). However, the other seismic stations do not show unusual behavior

for this event. The tilt measured by the nearby tiltmeter does not show exceptional behaviour either. This leads us to conclude that this is a local phenomenon only appearing at seismic station E02. Therefore, we can exclude meteorological parameters as a possible explanation, as they would affect the whole experiment site and would not be restricted to one instrument. Potentially, a rock ejected from the flume and impacting the ground near station E02 could cause a larger amplitude during the experiment. However, we cannot find clear evidence for such an event in the high-frequency seismic data or the video documentation (Logan et al., 2018). Having a closer look at the outlier waveform, it becomes apparent that the large tilt emerges while the debris flow is approaching the station. After the passage of the station, however, the seismometer does not tilt back into its original position as it does for the other events but rather stays at a constant level around 0.2 mircoradians. This could indicate a static change in seismometer tilt, resulting from either a permanent consolidation of the ground or a permanent weight added close to the sensor.

Despite these outliers, the tilt data of the USGS experiments exhibit larger tilt amplitudes recorded by the seismometer for experiments with higher starting volumes. Moreover, tilt responds to the speed of the experimental debris flows with a narrower signal of higher volume events, which propagate faster towards the bottom of the flume (see Fig. 3.4a,b with the June 22, 2016 and June 23, 2016 events being the faster events). The correlation of the flow velocity and the slope of the rise in tilt while the flow is approaching the station at station E03 shows that the tilt can also be used to estimate frontal flow velocity, only using a single station. This is, however, less the case for the latter part of the experiment (at E01), where the correlation decreases drastically (Fig. 3.6d).

Therefore, at station E03, the tilt signals let us infer debris-flow parameters (at least relative to other events), such as volume and flow thickness. Additionally, whereas for velocity measurements, several sensors are typically needed, the maximum slope of the tilt signal is so well-correlated with the event velocity that we can derive a simple calibration to obtain flow front velocity with only one station. While the high-frequency seismic signal is also correlated with such bulk parameters, the specific relation is not as straightforward to determine (e.g., Allstadt et al., 2020).

Figure 3.7 shows the correlation of measured normal stress and tilt measured at station E03 for successive time steps. Correlation coefficients for the USGS experiment vary between 0.93 and 0.98 and validate the strong correlation between tilt and normal stresses. The aggregation of all data points on the diagonal shows that a single parameter can be used to transform tilt to normal stresses.

3.5.1.2 *Illgraben, Switzerland*

For the Illgraben debris flows, the tilt signal marks the passage of the debris flow past the broadband seismometer. The tilt signal also reflects the passage of secondary surges or roll waves, which arise from non-linear flow instabilities or secondary mobilization of sediment from a landslide or the material deposited within the channel (Zanuttigh & Lamberti, 2007). Surges represent parts of debris flows with large boulders and sediment content, similar to a debris-flow front (e.g., Zanuttigh & Lamberti, 2007). The tilt signal can therefore resolve flow heights of individual surges. The June 4, 2020, and August 30,

2020 events, clearly show such roll waves, whereas, for example, the June 21, 2019 event shows the signature of a debris flow with one dominant debris-flow front (Fig. 3.5).

When modeling the tilt signal observed at Illgraben, we use data from a force plate 450 m downstream of station ILL11. The differences between the recorded and modeled tilt may result from changing flow velocity, geometry, and/or composition during the propagation between the seismometer and the force plate. For example, we can see in Figure 3.8b, that the second surge seemed to have sped up in between those two recordings, whereas the third surge moved the same speed. Additionally, the geometry of the channel slightly differs between seismometer and force plate sites: aerial images show that the channel close to ILL11 is several meters wider than at the force plate. The channel width is included in the model by the distributed load, which could influence the model accuracy.

Figure 3.7h shows the correlation of the debris-flow-front velocities with the maximum slope of the step in acceleration. The PCC of 0.49 indicates that there is no correlation between those two parameters. This shows that the step in acceleration is not a result of a slowly and continuously approaching debris flow. If it was, the slope of the signal should correlate with the debris-flow front, as it does in the USGS experiments. Therefore, we assume that the step in acceleration is instead a result of the debris-flow mass passing over the nearby check dam, causing a step in acceleration due to the impact of the free-falling mass on the ground.

The time dependent correlation of vertical forces measured at the force plate and the tilt recorded at the seismic station (Figure 3.7) is significantly lower for Illgraben than the USGS experiment. Generally, the correlation between the first half of the flow is higher than considering the entire flow. This is reasonable, as the seismometer is sensitive to tilt caused by a large portion of the flow, whereas the force plate only records what is happening when the flow mass locates directly on the force plate. Therefore, the beginning of the tilt signal as the flow approaches the station and the flow front is close to the station, should correlate best with the force plate measurements. Later when the flow front has passed and propagates further downstream, the tilt is influenced by the smaller weight right next to the station and the heavier debris-flow front further away from the station. Additionally, the two different locations of the seismometer and the force plate can decrease correlations. As seen in Fig 3.8b, the travel distance of 450 m is large enough for the debris-flow dynamics to change and possible surges to decelerate and accelerate.

In Figure 3.8c and 3.8d we show that the Illgraben signal is dominated by the passage of the debris-flow over a check-dam, which causes a significant step in acceleration recorded at the seismometer. For the June 21, 2019 event shown here, a step in acceleration of 5 m/s^2 convolved with the instrument response constitutes the best fit to the data. By assuming that $a_x(t) = -g\tau(t)$ this translates to a maximum tilt of about 5 microradians, which is exactly what we obtain with our method described in section 3.3.2.

3.5.2 *Model parameters*

To define a quantitative scaling factor between debris-flow height and tilt observed at the seismometer, we can use the distributed-mass model introduced in this paper. First, however, for this model it is crucial that we know the elastic properties of the subsurface,

such as the Poisson's ratio and the shear modulus (see equation 3.3). These can be obtained from either geophysical measurements or laboratory experiments.

For the USGS debris-flow flume, ground velocities from an active seismic experiment allowed us to compare model output with the tilt observed at the seismometer. Figure 3.8a shows that the ground properties for the weathered bedrock give a better estimate for the observed tilt than the properties of soil. The model can also give insights into another issue: a discrepancy in amplitudes between seismic station E02 and tiltmeter 3662 is clearly shown in Figure 3.4c and d. The two sensors are located close to each other but nevertheless 1.15 m apart. Additionally, because of the rebar hammered into the ground under its concrete pedestal, the tilt sensor is coupled about 1.5 m deeper in the ground. We conduct a numerical experiment with a point mass and the actual positions of the tiltmeter and the seismometer to assess if the position of the sensors can explain the amplitude differences. An additional 90 cm of flume-sensor distance corresponding to the tiltmeter location causes a tilt amplitude decrease of about 30 %. Therefore, the location difference explains a significant part in tilt amplitude differences between the seismometer and the tiltmeter. The remaining and dominant part of the discrepancy may result from the sensor installation, which, for the seismometer, consists of a sensor merely installed on a concrete slab resting on top of the soil within a shallow vault. The tiltmeter is installed on a concrete base that is anchored with rebar to a larger depth, already considered weathered bedrock. A resulting difference in the depth and quality of the coupling to the ground could affect the tilt measurement. Especially the material in which the sensors were buried/installed might have an effect: Figure 3.8a shows the modeled tilt in soil and bedrock, whereby the tilt amplitude is an order of magnitude higher for soil than for bedrock. Consequently, we suggest that the difference in amplitude between the tilt measured by the tiltmeter and the seismometer is a combination of differing sensor locations, installation techniques and ground material coupled to the sensor.

For the Illgraben case, we do not have estimates for ground properties. This led us to use parameters from a broad spectrum of different soils found by Salem (2000). Even though the 95 % confidence interval of the modeled tilt spans a factor of two (Fig. 3.8) in tilt, the recorded tilt lies mainly within this range.

3.5.3 *Tilt measurements as an alternative to force plates?*

Figure 3.7 and 3.8a,b show that the tilt obtained with seismometers gives a measure of the force applied normal to the ground. This tilt therefore scales with the normal force measured at force plates. However, seismometers are significantly less cumbersome to install than force plates requiring construction in debris flow channels, especially at large-scale sites where the propagation path of the mass movement is not well defined. Force plates, by nature of being a direct measurement, are also prone to being damaged by the flows whereas seismometers can be installed at a safer distance. We highlight that by assessing the tilt caused by debris flows, we can estimate the debris-flow height and normal forces, given an accurate seismic ground model or scaling factor. Even without such a model, the tilt measurements give insights into these parameters relative to other recorded events. To access this information, we only need a single seismometer station. Compared to force plates, which in the case of Illgraben cost about one Million USD including installation costs (personal communication Brian McArdell), a seismic station containing a broadband sensor, state-of-the-art digitizer, and solar panels can be acquired

for about 25,000 USD. Additionally, while the force plate installation can, depending on its scale, take up to several months, our seismic broadband station can be installed within an hour. Clearly, another advantage of a broadband seismometer installation is the portability of the setup.

3.5.4 *Transferability to other sites*

To investigate if we can use the tilt signal to estimate debris-flow parameters at other sites, we consider data from two debris-flows (in volcanic environments also called lahars) at Volcán de Fuego in Guatemala. Fuego is one of the most active volcanoes in South America (e.g., Diaz-Moreno et al., 2020). The active lahar channel at the study site is about 90 m wide. A broadband seismometer (station FG14), operated by the Instituto Nacional de Sismologiía, Vulcanología, Meteorología y Hidrología (INSIVUMEH) is located approximately 30 m from the edge of the channel. Figure 3.9a shows a shaded digital elevation model (DEM) of Fuego (NASA JPL, 2020). The green line denotes the lahar channel, and the black triangle denotes the location of station FG14. The raw seismic signal of two lahars that occurred on October 16, 2019, and May 10, 2019, are shown in Figure 3.9b,d. We rotated the horizontal components (E,N) such that the radial component points towards the shortest distance to the channel (back-azimuth = 285 degrees) and derived the tilt shown in Fig. 3.9c,e. For the October 16 event, we can see five distinct tilting events on both the tilt signal and the raw seismic data (denoted with arrows in Fig. 3.9c). The first four consecutive main surges correlate well with high amplitudes in the raw data. After about half an hour, a fifth surge passes by the seismometer, also well seen in the raw seismic data. The May 10 event shows a high-frequency event preceding the main flow in the raw seismic data. This event does not cause a tilt signal on FG14 and only lasts for about 60 s (see arrow Fig. 3.9d). With our method described in section 3.3.2 we filter out such short signals, however we also think that this signal is not related to the lahar of interest as it passes by the station and may come from a more distant source. Following this short high-frequency signal, there is a build up in the energy of the raw signal likely reflecting the arrival of a flow front that lasts about 300 seconds that is mirrored by a rise in the tilt signal. Interestingly after a short decrease in both high frequency energy and tilt, the tilt then rises slowly over several minutes to a peak level of up to one microradians, but the high frequency signal only shows a much lower and broader peak in energy around this time compared to the initial flow front onset. The relationship between high frequency energy and flow height is not linear and is strongly dependent on flow style (e.g., Allstadt et al., 2020), so it is possible that this broad tilt that lasts for many minutes actually reflects a deeper, denser, less agitated flow than the initial flow onset. For example, in comparing several lahars at Mount Ruapehu, New Zealand, Cole et al. (2009) observed that plug-like, denser, more laminar flows had less high frequency energy than flows that were less dense, watery, and more agitated. Unfortunately no camera data was available for the May 10 event to confirm this hypothesis and the broad tilt signal could also be an artifact.

The tilt observed in May 10 event is more than two times larger than for the October 16 event. We assume that the May 10 event reached greater flow depths than the October 16 event based on the correlation of maximum tilt and flow depth at both the USGS experiments and Illgraben. To test this hypothesis, we use the distributed-load model to invert for the maximum debris-flow thickness. In contrast to the model setup for

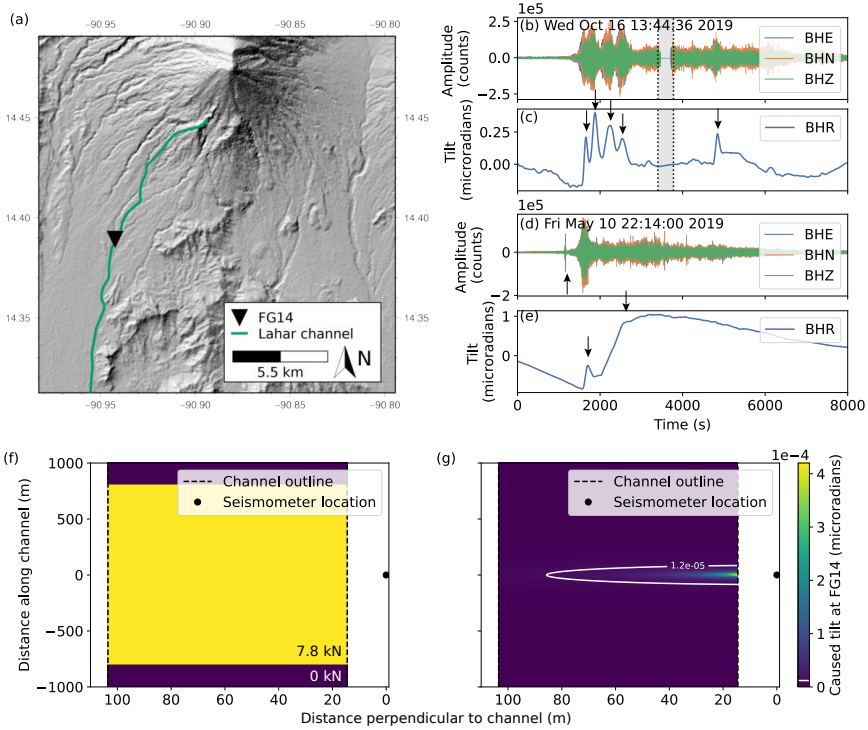


FIGURE 3.9: (a) Shaded DEM of Volcán de Fuego, Guatemala. The green line marks the location of the lahar channel. The black triangle marks the location of the broadband seismometer. (b) Raw seismic signal of all three components of the October 16, 2019 event at Volcán de Fuego. The grey window marks a gap in the data. The black arrows denote the five main surges. (c) Tilt obtained from seismic data. (d) Raw seismic signal of all three components of the May 10, 2019 event at Volcán de Fuego. (e) Tilt obtained from seismic data. The black arrows denote again the two main surges. (f) Input for sensitivity test: Vertical force exerted by a distributed debris-flow mass (yellow area, 7.8 kN per m^2) over the width of the channel (x-axis) and 800 m on both upstream and downstream (y-axis) of the seismometer location (black dot). Virtual seismometer location at point (0,0). (g) Vertical deformation caused at seismometer per grid point. Contour shows the threshold of 80 % of total caused tilt.

Illgraben and the debris-flow flume, we cannot use force-plate data in this case because they do not exist and we do not have flow depth data. To make assumptions on the flow depth, we conducted a simple sensitivity analysis (Fig 3.9f,g). We compute the normal force exerted by a distributed uniform load of a 0.4 m thick debris flow with a density of 2000 kg/m^3 (Vallance & Iverson, 2015). The mass is distributed over the whole channel width and along an 800 m distance both upstream and downstream of the seismometer location (Fig. 3.9f). Because of the unknown seismic velocities, we use sub-surface parameters proposed by Lyons et al. (2012) to model tilt at the upper cone of Volcán de Fuego: v_p = 3500 m/s, v_s = 2000 m/s, and ρ = 2650 m/s. This gives a Poisson's ratio of ν = 0.25 and a Young's modulus of E = 26.6 GPa. The tilt caused by each grid point at the seismometer is shown in Figure 3.9g. The sensitivity analysis shows that 80 % of the tilt is caused by a radius of 133 m around the seismometer. A uniform flow depth over a 1600 m stretch of channel as used in the model is highly unlikely. However, given that a large portion of the tilt is caused in a small radius around the seismometer, we used a uniform mass distribution 133 m around the seismometer as input for our model to obtain an estimate of the maximum flow thickness. With these model parameters, we estimate the maximum flow depth of the October 16 event as 0.4 m and the May 10 event as 1.1 m. Flow-depths for the October 16 event were estimated at about 1 - 2 m using camera images. Compared to these estimations, our model significantly underestimates the maximum debris-flow depth. Unfortunately, there is no camera footage of the May 10 event. However, given the correlation between debris-flow depth and tilt signals, we can give an estimate for the May 10 event: if the flow depth estimation by the camera images is taken as ground truth for the October 16 event, the May 10 events flow depth ranges between 2.8 and 5.6 m.

3.5.5 *Low frequency data for early warning purposes*

In the case of Illgraben, the raw long period seismic data (periods $>$ 10 s) of the horizontal components clearly shows the instrument response to a step in acceleration that we hypothesize is caused by the free fall of the flow front over the check dam that is adjacent to the station. As this signal is quite unique, we used a cross-correlation search to evaluate the use for early warning purposes. First, we mutually cross-correlated all available events and chose the event that achieves the highest CC with all other events, which was the July 15, 2019 event, as a template and cross-correlate it with the continuous seismic records from 2019 and 2020 (283 days total). We chose a cross-correlation coefficient threshold of 0.35 as an optimal trade-off to minimize the number of false positives and false negatives. In the 283 days of data, 14 out of 16 debris flows were found. All flows in 2020 were detected, whereas for two debris flows in 2019, the tilt signals were obscured by low-frequency noise (see Fig. 3.5b) or no clear debris-flow front was visible (no apparent step in acceleration). The cross-correlation also yielded three false positives, all of them in 2020. The detection might be attributed to the location of the seismometer, which is located only a few meters away from a trail frequented by hikers and horseback riders. These nearby noise sources can also cause a step in acceleration, which then gives a similar waveform as the step in acceleration caused by the debris flow overshooting the nearby check dam. With this simple cross-correlation search, we found almost 90 % of the events and only had three false positives. We argue that this result is satisfying for such a simple technique for flow detection that only requires one station and very

minimal computational power. However, it is clear that this correlation only works for cases similar to Illgraben, where the sensor is located close to a steep drop such as a check dam or a natural waterfall.

Apart from the correlation search, we have shown that all debris flows produce a tilt signal as the flow passes a near-channel broadband sensor. However, our method of deriving the tilt from broadband data is only applicable for an event re-analysis and not for real-time warning. The reason for this lies in the use of a low-pass filter below the corner frequency, which can produce large side lobes for finite time windows. Such side lobes are easily recognized and trimmed, but sufficient data is required before and after a suspected event. One potential solution could be the implementation of a recursive tilt "filter" converting seismic data to tilt as data streams in as developed by Thelen et al. (2015). These authors were able to recover tilt from streaming seismic data, but numerical instability limits the long-period sensitivity ($>600\,s$) of the filter. Existing machine learning detectors for mass movements such as discussed in Chmiel et al. (2021) and Wenner et al. (2021) operate on shorter time windows below 100 s. To include tilt as an additional characteristic for debris-flows, such algorithms would either need to be adjusted to predict mass movements on longer time scales or used as time shifted detection validation.

We have also shown in this paper that long period noise can conceal the tilt signal from a debris flow, and thus, site selection is essential. The sensor vault should be chosen to prioritize a short distance to the channel but also temperature stability of the sensor. On the other hand, unexpected problems can arise from a station too close to the channel: In the case of Nevado del Huila Volcano, a glacier-covered volcano in Colombia that produced three exceptionally large lahars (up to 300 million m^3) in 2007 and 2008 (Worni et al., 2012) with flow heights up to 50 m, the horizontal components of a seismometer located near this drainage clipped on the horizontal channels when the 300 million m^3 flow passed, making the record unusable to retrieve tilt. However, the vertical component did not clip and though tilt cannot be retrieved from that component, deformation modeling would be a possibility for such a large event. Additionally, investigations of the higher-frequency data on the vertical component of the July 26, 2019, and August 11, 2019 events at Illgraben have shown that the long-period noise might be caused by people close to the seismometer. It is, therefore favorable, to find a secluded spot away from hiking trails and roads.

3.6 CONCLUSION

In this study, we showed that we can record tilt on broadband seismometers caused by the passing of a debris flow. By comparing the tilt obtained from the seismometers with the output of a distributed-moving-force model, we can see that the waveform of the tilt signal is similar to the forces measured by force plates. The reconstruction of the amplitudes of the tilt signals is strongly dependent on sub-surface parameters. Relative to other recorded events, the tilt observed at the seismometers does correlate well with the debris-flow height and measured normal forces. Especially for hazard early warning and risk assessment, the flow depth is one of the most crucial parameters. Therefore, our method allows us to obtain information needed to evaluate the destructive potential of a debris flow event if (1) other events were recorded beforehand or (2) channel geometry and sub-surface properties at the site are known.

DATA AVAILABILITY

Seismic data of the Illgraben network are available via the Swiss Seismological Service, http://arclink.ethz.ch/webinterface. Seismic data of the USGS flume experiments are available via the IRIS Data Management Center (Network code: ZK; Allstadt, 2016). Forceplate and flow depth measurements are available via the USGS ScienceBase (Iverson & Logan, 2017). Video footage of all USGS flume experiments are available at https://pubs.usgs.gov/of/2007/1315 (Logan et al., 2018). Seismic data from Volcán de Fuego are available upon request by the Instituto Nacional de Sismología, Vulcanología, Meteorología y Hidrología.

ACKNOWLEDGEMENTS

This project was funded by the WSL Climate Change Impacts on Alpine Mass Movements (CCAMM) initiative and the U.S. Geological Survey. We would like to thank the Swiss Seismological Service and especially Pascal Graf and Raphael Moser for the active support during the seismometer installation at Illgraben and data management. The Canton of Valais and the municipalities of Leuk and Anniviers supported this project financially and granted permission for the seismometer installation at Illgraben. We would like to thank Dick Iverson, Maciej Obryk, Matt Logan, Chris Lockett, Aaron Rinehart, David George, Joel Smith, Jason Kean, Kelly Swinford, Francis Rengers, Maxime Farin and Luke McGuire for their many contributions to the 2016 experiments at the USGS debris flow flume that were used in this study. We'd also like to thank Rob Anthony who helped us investigate instrumentation issues related to measuring tilt. We thank our colleagues from colleagues the Instituto Nacional de Sísmologia, Vulcanología, Meteorología y Hidrología who installed and maintain the seismic stations at Volcán de Fuego and shared the data with us. Additionally we would like to thank the Obspy and PyGMT developer teams (Beyreuther et al., 2010; Uieda et al., 2021) for providing python frameworks for seismic and geospatial data handling. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

4

NEAR REAL-TIME AUTOMATED CLASSIFICATION OF SEISMIC SIGNALS

This chapter is published by the *European Geoscience Union* as:
Wenner, M., Hibert, C., van Herwijnen, A., Meier, L. & Walter, F., (2021). Near-real-time automated classification of seismic signals of slope failures with continuous random forests. *Natural Hazards and Earth System Sciencess*, 21(1), 339-361.

ABSTRACT

In mountainous areas, rockfalls, rock avalanches, and debris flows constitute a risk to human life and property. Seismology has proven a useful tool to monitor such mass movements while increasing data volumes and availability of real-time data streams demand new solutions for automatic signal classification. Ideally, seismic monitoring arrays have large apertures and record a significant number of mass movements to train detection algorithms. However, this is rarely the case, as a result of cost and time constraints and the rare occurrence of catastrophic mass movements. Here, we use the supervised random forest algorithm to classify windowed seismic data on a continuous data stream. We investigate algorithm performance for signal classification into noise (NO), slope failure (SF), and earthquake (EQ) classes and explore the influence of non-ideal though commonly encountered conditions: poor network coverage, imbalanced data sets, and low signal-to-noise ratios (SNR). To this end we use data from two separate locations in the Swiss Alps: data set (i), recorded at Illgraben, contains signals of several dozen slope failures with low SNR, data set (ii), recorded at Pizzo Cengalo, contains only five slope failure events albeit with higher SNR. The low SNR of slope failure events in data set (i) leads to a classification accuracy of 70% for SF, with the largest confusion between NO and SF. Although data set (ii) is highly imbalanced, lowering the prediction threshold for slope failures leads to a prediction accuracy of 80% for SF, with the largest confusion between SF and EQ. Standard techniques to mitigate training data imbalance do not increase prediction accuracy. The classifier of data set (ii) is then used to train a model for the classification of 176 days of continuous seismic recordings containing four slope failure events. The model classifies eight events as slope failures, of which two are snow avalanches, and one is a rock slope failure. The other events are local or regional earthquakes. By including earthquake detection of a permanent seismic station at 131 km distance to the test site into the decision-making process, all earthquakes falsely classified as slope failures can be excluded. Our study shows that even for limited training data and non-optimal network geometry, machine learning algorithms applied to high-quality seismic records can be used to monitor mass movements automatically.

4.1 INTRODUCTION

High mountain areas are particularly affected by climate change. Deglaciation and thawing of permafrost has implications on rock wall stability at high elevation and, consequently, on communities down-valley (e.g., Allen & Huggel, 2013; Phillips et al., 2017; Coe et al., 2018; Hock et al., 2019). The increasing threat to mountain communities, especially in densely populated areas, demands new monitoring techniques at high temporal resolution and broad spatial coverage to improve predictability, alarm time and post-event intervention. As a result of incomplete data and knowledge on relevant processes and triggering mechanisms, accurate prediction of rockfall events is still not possible (van Westen et al., 2006). However, an increase in slope activity (pre-event acceleration and increased frequency of small events) is a possible precursor to larger destructive events (Rosser et al., 2007). Existing methods to monitor slope failures include point measurements (e.g., extensometers) and large scale monitoring such as terrestrial laser scanners, interferometric radar, and video image recognition (e.g., Abellán et al., 2011). However, these techniques suffer from disadvantages like high operating costs, limited spatial coverage, and susceptibility to atmospheric conditions.

In the last decade, seismology has evolved into a method to monitor earth surface processes. Knowledge from wave propagation within the earth and generation mechanisms of seismic waves is transferred from its original study objectives, earthquakes, to the so-called field of environmental seismology (e.g., Burtin et al., 2008; Deparis et al., 2008; Helmstetter & Garambois, 2010; Gimbert et al., 2014; Hibert et al., 2014; Larose et al., 2015; Dietze et al., 2017; Allstadt et al., 2018; Lai et al., 2018). Seismic signals generated by mass movements are typically emergent with dominant frequencies of 5 – 10 Hz and few or no distinguishable seismic phases. Signals of large events ($> 10^5$ m^3) recorded at far distances are characterized by long-period seismic waves (< 0.1 Hz) (e.g., Allstadt, 2013). Signal duration varies between seconds and several minutes, depending on the type of slope failures and slope scales (e.g., Vilajosana et al., 2008; Hibert et al., 2011; Dietze et al., 2017).

Seismometers can record large mass movements up to hundreds of kilometers away from the source (e.g., Allstadt, 2013; Walter et al., 2020b) and allow continuous monitoring of large areas with real-time data transmission (e.g., Ekström & Stark, 2013). Additionally, the installation of seismometers is relatively low-cost and straightforward as no high-power supply and little or no or construction is needed, which is required, e.g., for interferometric radar. On the other hand, seismic sensors are sensitive to various sources like earthquakes, anthropogenic noise, atmospheric signals, runoff, and slope instabilities. Consequently, to distinguish signals of slope failures from other mechanisms, automated techniques for detection and classification are needed.

One approach that is often used to detect seismic signals is the well established short-term average over long-term average (STA/LTA) detection method, based on signal amplitudes (Allen, 1982). However, to classify source mechanisms, information on signal frequency content is often also required. For signals with similar frequency content, amplitudes and signal duration, such as earthquakes and slope failures, detection of signals from only one source mechanism with STA/LTA is therefore impossible. Additionally, parameter selection for optimizing STA/LTA is a tedious process that requires detailed knowledge of the data. Furthermore, seismic signals of slope instabilities are characterized by an emergent onset, making detection with STA/LTA difficult. For

this reason, Helmstetter and Garambois (2010) suggested an STA/LTA algorithm for use in the frequency domain that reliably detects seismic signals without impulsive onsets (e.g., Hibert et al., 2017). Nevertheless, the detector does not allow a distinction between different generation mechanisms. Therefore, Hibert et al. (2017) and Provost et al. (2017) used a supervised machine learning algorithm, random forest, to automatically classify local events detected with the adapted STA/LTA algorithm of Helmstetter and Garambois (2010). High accuracy (99 % and 93 %, respectively) emphasizes such algorithms' capability to classify seismic signals.

One downside of the combined STA/LTA approach by Hibert et al. (2017) and Helmstetter and Garambois (2010) is that this method requires two optimization steps - choosing the correct parameters for the STA/LTA algorithm and for the classifier. For this reason, this approach is called the two-step approach in the further course of the manuscript. Moreover, STA/LTA algorithms generally fail to detect signals which emerge over a timescale larger than the long term average window.

As an alternative to the two-step approach, a stochastic classifier, hidden Markov models (HMMs), has been used to automatically detect and classify a variety of seismic sources (e.g., Hammer et al., 2013; Dammeier et al., 2016; Heck et al., 2018). Hammer et al. (2013) and Dammeier et al. (2016) focused on a regional scale with larger rockfall volumes ($> 1000\ m^3$) detected tens to hundreds of kilometers away from the source. They show that HMMs successfully classify seismic signals on a continuous data stream. However, to minimize false detection and misclassification, careful retraining and post-processing steps were required. Dammeier et al. (2016) compared the classification output with an earthquake catalog and suggested that when using HMMs in an actual operational setting, the on-duty operator should manually inspect the signal and decide if the event is an earthquake or a slope failure. Yuan et al. (2019) used random forest to classify seismic signals in several minute-long windows of seismic data. The study focuses on multiple days of data recorded near a geyser to detect pre-eruption seismicity, which is hidden in the noise. In the following, signal classification on the continuous data stream that does not require separate event detection is called the one-step approach.

In this paper, we use the random forest algorithm (Breiman, 2001) to perform automatic signal classification on continuous data on a local scale and throughout an extended time period with a large variety of noise signals. Previous local-scale two-step approaches have used specialized networks designed to maximize monitoring capabilities (e.g. Provost et al., 2017). However, due to logistical and financial constraints, this is often not possible for potential hazard sites. Here, we compare the one-step approach's performance in non-ideal conditions applying it to a data set with (i) many small slope failure events with low signal-to-noise ratio (Illgraben) and (ii) few events but a higher signal-to-noise ratio (Pizzo Cengalo). Furthermore, we show that by adjusting our methodology to a network with a sub-optimal configuration and a data set with only a few recorded events available for training, automatic detection of potential slope failures is still possible.

The scope of this study is to a) test a system of continuous classification of windowed seismic data on two different types of data sets, b) test the influence of signal to noise ratios and an imbalanced training data set on classifier performance, c) discuss insights on the transferability of trained classifiers to other sites, and d) mimic operational conditions to assess our approach's capability as an alarm system for slope failures. The investigation of source mechanisms and processes of seismogenic mass movements

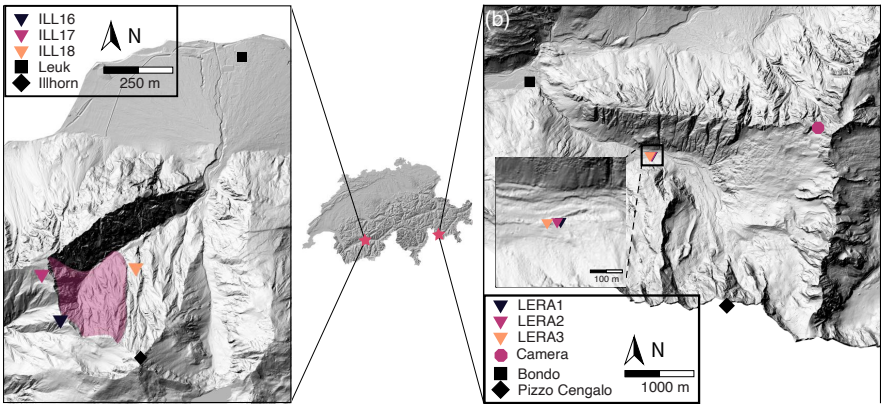


FIGURE 4.1: Illgraben catchment (left) and Bondasca valley (right) with locations in Switzerland (stars in middle panel).(a) Overview of the Illgraben catchment in Switzerland and locations of stations ILL16 – 18 depicted as colored triangles. (b) Overview of the Bondasca valley. Location of LERA network depicted as colored triangles. Zoom-in shows individual seismic stations (LERA1 – 3) in the upper right corner. Digital elevation models ©swisstopo.

is outside our study's scope. Such an endeavor should not be based on weak seismic signals, which make up a large part of our catalog.

4.2 STUDY SITES AND DATA SET

4.2.1 *Illgraben, Switzerland*

The Illgraben catchment in southwest Switzerland is one of the most active mass wasting sites in the European Alps (Fig. 4.1a). Yearly precipitation is controlled by summer rainstorms with high rainfall intensity during which mass wasting with rock-slope failures and debris flows occur regularly (e.g., Badoux et al., 2009). From its highest point at the Illhorn (2716 m.a.s.l), the Illgraben catchment reaches down to the Rhone Valley (600 m.a.s.l), where its main torrent flows into the Rhone River. The 9.5 km^2 Illgraben catchment is characterized by complex geology, where highly fractured quartzite is the dominating bedrock at the north-west face of the Illhorn and the head of the channel trunk and limestone at the southeast facing slope of the catchment (Schlunegger et al., 2009; Bennett et al., 2013). The fractured quartzite with erosion rates of tens of centimeters per year is the main contributor to sediments transported via debris flows (Bennett et al., 2013). A seismic network of 8 high-quality stations was installed throughout Illgraben between May and September 2017 to monitor rock-slope as well as debris-flow activity. Walter et al. (2017) used seismic data from a similar array to locate a debris-flows front as it propagated along the channel. In this study, we focus on slope activity at the Illhorn north-west face, which at peak times is characterized by several slope failures per day. Slope-failure volumes are to date not quantified, but direct field observations by the authors indicate volumes of tens to hundreds of cubic meters. We use seismic data

recorded by three seismic stations (ILL06, ILL07, ILL08) in 2017, located closest to the area of interest (Fig. 4.1a). The three-component seismometers (LE-3Dlite) with a lower cutoff frequency of 1 Hz and a sampling frequency of 100 Hz were installed with a mean inter-station distance of about 1 km and hundreds of meters from the Illhorn north-west face.

4.2.2 *Pizzo Cengalo, Switzerland*

Pizzo Cengalo is a mountain located in Val Bondasca in Eastern Swizerland's canton of Grisons (Fig. 4.1b) about 6 km south east of the down-slope village of Bondo near the Italian border. Pizzo Cengalo's slopes have been unstable for several decades, with multiple rock slope failures per year. Bergell Granite defines Pizzo Cengalo's geology and part of its north facing walls are covered by glaciers (Baer et al., 2017). After a large failure ($\sim 1.5 \times 10^6\,m^3$) in 2011, systematic monitoring started in 2012 (Baer et al., 2017). In 2017, an even larger rock avalanche ($> 3.5 \times 10^6\,m^3$) killed eight hikers and resulted in a series of debris flows that destroyed parts of the village of Bondo (Walter et al., 2020b). A warning had been issued weeks prior to the catastrophic failure because an acceleration of slope displacement was observed, as well as several smaller failure events before the rock avalanche. The large event in 2017 prompted an extension of the monitoring system, which included the installation of three seismometers (LERA1 – 3) close to the Bondasca river, the outlet of the catchment, some 3.5 km down-valley of Pizzo Cengalo and 2.5 km up-valley of Bondo.

The one component, short-period geophones (GeoSig 0.9 Hz), with a flat response from 0.9 Hz to 89 Hz and a sampling frequency of 200 Hz, were installed along the channel with a mean inter-station distance of about 20 m (Fig. 4.1b). An array in such a configuration can be used to detect debris flows based on amplitude differences while the flow approaches, passes by, and moves away from the stations (Coviello et al., 2019). Since the installation of the seismometers, seismic signals and spectrograms have been made available to stakeholders in hourly time windows in the online portal of the engineering company Geopraevent, to recognize an increase in rockfall activity by visually evaluating the seismic data. Until now, this required daily visual inspection of the seismic signals by employees of the canton of Grisons, who are not trained seismologists.

In November 2017, we installed an automatic camera facing Pizzo Cengalo (768411 / 132790, Fig. 4.1b). Images were taken every 30 minutes and transferred in real-time via a cellular connection. We used these images to validate detected events, and the images were also available to stakeholders on the online portal of Geopraevent.

4.2.3 *Labeled Data Set*

For supervised machine learning algorithms, a set of labeled data has to be provided. Here, we focus on seismic data from 3 sensors of the ILL array recorded in 2017 and the 3-sensor LERA array recorded in 2018 and 2019. We compiled an event database for both study sites by visual inspection of the seismic waveforms and spectrograms of seismic stations at Illgraben and Pizzo Cengalo and close-by stations of the Swiss Seismological Service (SED) (Illgraben: Stations CH.LKBD, CH.VANNI; Pizzo Cengalo: Stations XP.PICE1, CH.VDL, CH.FIESA). Additionally, we use a list of observed slope

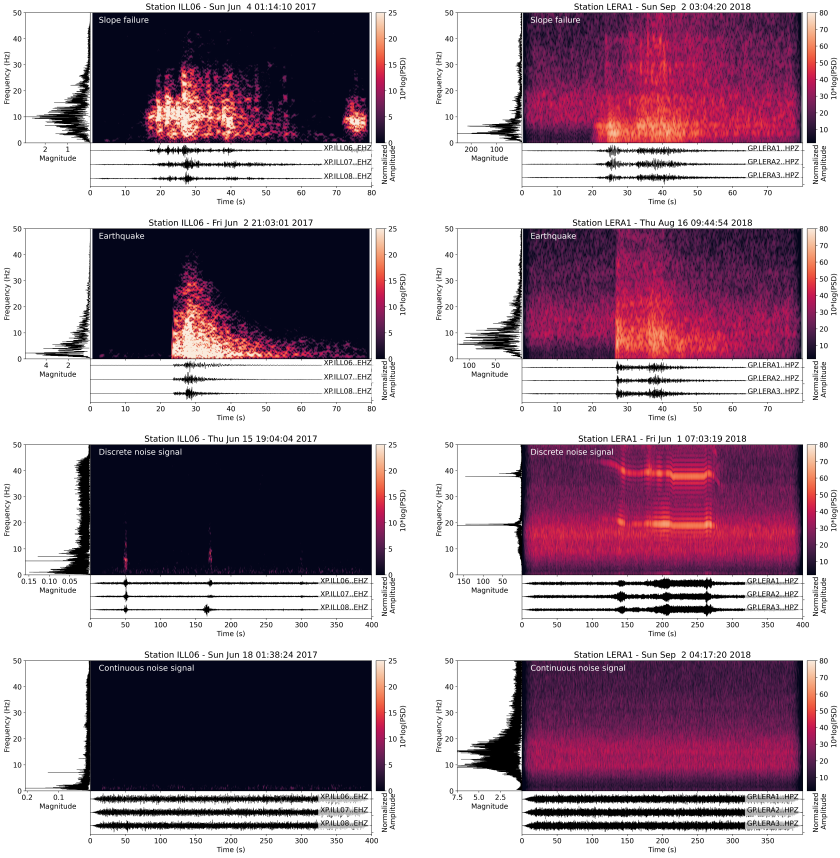


FIGURE 4.2: Waveforms, spectrogram and spectra of example events for each class recorded at Illgraben (left) and Pizzo Cengalo (right)

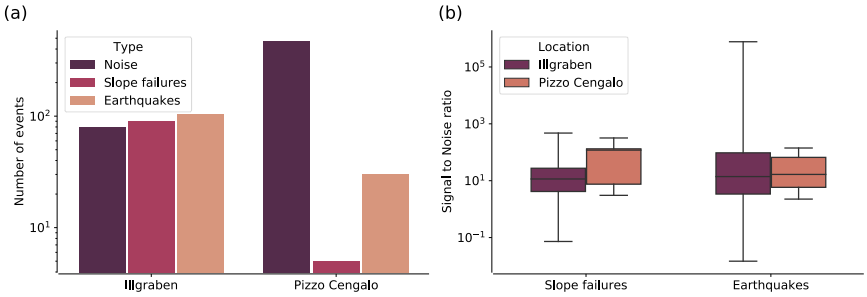


FIGURE 4.3: (a) Bar chart with events per class for both the Illgraben and the Pizzo Cengalo data set. (b) Boxplot of SNRs of all slope failure and earthquake events at Illgraben and Pizzo Cengalo. The boxes denote the quartiles (25-75%), the black horizontal line denotes the mean and the black vertical error bars indicate outliers.

failures made available by the canton of Grisons for Pizzo Cengalo and earthquake catalogs from SED and the European-Mediterranean Seismological Centre (EMSC). For the classification, we decided to use three different classes: noise (NO), slope failures (SF), and earthquakes (EQ). The NO class contains samples of continuous noise and noise signals of anthropogenic and atmospheric origin (Marchetti et al., 2019). We use the SF class as an umbrella term for all types of mass movements that might occur (e.g., snow/debris avalanches, rockfalls). We consider this assumption valid, as different granular flow types share the common seismogenesis of particle ground impacts (e.g., Suriñach et al., 2005; Farin et al., 2019), although length and amplitudes depend on runout distance and volume. We assume that differences in signal characteristics between EQ, NO, and SF class are more significant than discrepancies between different types of granular flow signals. The most critical parameters to manually classify an event as slope failure are: dominant frequencies of 5-10 Hz, emergent onset, no phase arrivals, not listed in earthquake catalog, not seen on surrounding seismic stations, or lower amplitude at surrounding stations.

The EQ class contains a set of local, regional, and teleseismic events. Extensive testing showed that lumping all earthquakes in one class does not negatively affect classifier performance (Fig. A.3). An example signal of each class is presented in Fig. 4.2. For the continuous noise, we choose random times over the year. For noise signals, earthquake signals, and slope failure signals, we manually picked the events' start-time and end-time when the signal exceeds the noise level. The number of events in each class for both study sites is presented in Fig. 4.3a and Tab. A.5. Note that there are only five events that are related to slope failures at Pizzo Cengalo in 2018. Due to instrument malfunction, these events were not captured by the automatic camera. The sparsity of recorded events leaves us with a poor data set for this class. This issue is addressed further in section 4.3.2. Figure 4.3b shows a boxplot of the signal to noise ratios of SF and EQ events at both sites. The signal-to-noise ratio for earthquakes ranges between 1 and 10^3, with a mean of 5 (Fig. 4.3b). The signal-to-noise ratio for slope failures varies between 1 and 30, with a mean of 4 for the Illgraben data set and 22 for the Pizzo Cengalo data set.

4.3 METHODOLOGY

A schematic illustration of our automatic one-step classifier applied to the Pizzo Cengalo data set is shown in Fig. 4.4a. We use random forest, a supervised ensemble machine learning algorithm (Breiman, 2001), to classify different seismic sources using a running window on a continuous data stream on all network stations. Random forest is based on the majority vote of several weak decision trees, where each decision tree is built on a random subset of features and training data set. Decision trees consist of nodes, branches, and final nodes. A split based on a threshold on a variable is performed on each node, resulting in one or two branches. This process continues until a classification result is obtained in a final node, a so-called leaf. A single weak tree performs poorly in the classification task, as it is only trained on a subset of features and the training data set. However, the performance improves as the aggregated decision trees perform a majority vote, where the proportion of trees that voted for one class gives the probability for the class. The time window is then labeled according to the class with the most votes, i.e., the highest probability. We choose random forest, because (i) it is a comprehensive machine learning algorithm that has shown to outperform other algorithms, like support vector machines and boosting ensembles, in a variety of cases (Fernández-Delgado et al., 2014) and (ii) it already was successfully used to classify rock slope failures (Hibert et al., 2017; Maggi et al., 2017; Provost et al., 2017; Malfante et al., 2018; Hibert et al., 2019). Moreover, random forest estimates the feature importance by measuring the impurity, which describes how many samples of how many different classes belong to one node after a split. Hence, if all samples in a node belong to one class, impurity is zero, and the classifier is perfect. The averaged impurity decrease from a feature over all decision trees then gives a ranking of the most discriminating feature. This allows a more detailed analysis of potential causes for misclassification. For the implementation of random forest we use scikit-learn, a python library for machine learning (Pedregosa et al., 2011).

4.3.1 *Data Stream Handling*

To avoid the extra step of detecting events with a trigger such as the STA/LTA algorithm, we classify a running window on the continuous data stream with an overlap of 2/3 of the window length. The overlap was chosen to avoid missing events on the window margins but was not tested for optimal performance. We transform the event catalog with start-times and end-times of all events into a catalog containing the times of all running windows that include an event for both the Illgraben and the Pizzo Cengalo data set. For the Pizzo Cengalo data set, we make use of the network configuration at the study site, in order to increase the number of training samples. At the frequency band of interest (1 - 10 Hz), associated wavelengths are larger than the inter-station distance, resulting in waveforms with only small differences. For earthquakes and slope failures, instead of using the same onset for the sliding windows on all stations, we choose a random onset with a maximum of 2/3 of the sliding window before the event start-time. This way, we catch different windows of the signal and increase the training data set by a factor of three without using the same window several times. This procedure is unnecessary for the Illgraben data set, as the source-receiver distances vary for each station. For discrete noise signals often recorded on only one station, we choose sliding windows on the

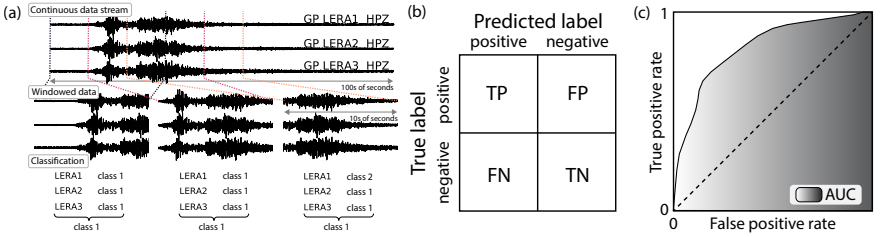


FIGURE 4.4: (a) One-step classification scheme with continuous data stream, windowed data, classification per station and label for window, (b) Confusion matrix for a two class problem (positive, negative), with true labels as rows and predicted labels as columns. True positives (TP) and true negatives (TN) on the diagonal and false negatives (FN) and false positives (FP) on the off diagonal elements. (c) ROC curve with true positive rate (TPR) on the y axis and false positive rate (FPR) on the x axis. Shaded area (AUC) measures model accuracy.

recording station, again with a random onset up to 2/3 of the sliding window before the event start-time. For the continuous noise, we choose a random station at each time step.

We divide the catalog with labeled events into a training and test data set, with 70% of all events as training data and 30% as test data. This partition was chosen to be able to meaningfully assess the algorithm performance for the small number of slope failures in the Pizzo Cengalo data set. This way, there are windows of three slope failure events in the training and validation data set and windows of two slope failure events in the test data set. We choose a window length of 20 seconds as an initial guess. Numbers for resulting training and test data set sizes for Pizzo Cengalo and Illgraben are presented in Tab. A.5. Following Provost et al. (2017), we then compute features of these sliding windows. As we do not use the entire waveform of the event, but only the parts that appear in the sliding window, we exclude features related to the entire waveform of the signal (e.g., duration and rise time). Additionally, the network configuration at Pizzo Cengalo does not allow network features (signals are too similar between stations), nor does it allow for polarity features (only vertical component available). For the sake of comparison, we also disregard network and polarity features for the Illgraben data set. We are left with a total number of 55 features, including waveform characteristics in the time and frequency domain (see Table A.1). These features have been proven significant for accurate seismic signal classification (e.g., Hibert et al., 2017; Provost et al., 2017). Additionally, we tested a python tool for automatic feature generation for time series (TSFRESH, Christ et al. (2016)), which did not improve classification results compared to the features proposed by Provost et al. (2017). Before feature calculation, we apply a four corners Butterworth bandpass filter (1 – 10 Hz). For feature generation after Provost et al. (2017), we choose frequency bands of 1 – 3 Hz, 3 – 6 Hz, 5 – 7 Hz, 6 – 9 Hz, and 8 – 10 Hz.

4.3.2 *Imbalanced Data Set Handling*

The limited amount of data and, more specifically, the small number of SF events that happened in 2018 at Pizzo Cengalo lead to an imbalanced data set. As shown in Fig. 4.3a, the number of events is unevenly distributed among classes. This poses a problem

for machine learning algorithms, as they generally optimize the score, i.e., the number of correctly labeled classes. In a highly imbalanced data set, the classification algorithm may be less sensitive to the minority class, as it does not drastically impair the score if it is labeled incorrectly. For our data set, with the events that we are most interested in being the minority class, it is particularly important to address this problem. Therefore, we introduce here different data augmentation and classifier tuning methods to improve the classifiers' performance on the imbalanced Pizzo Cengalo data set.

There are several possibilities to handle imbalanced data sets, either based on manipulating the training data set or on changes within the algorithm (Chawla, 2010). The most straightforward approaches are random undersampling (US) and naive oversampling (OS) training data. For random undersampling, only a random subset of training data of the majority class is chosen. This way, the data set becomes more balanced by reducing the samples in the majority classes. However, this might mean that important characteristics of the majority class are not captured. In contrast to undersampling, naive oversampling randomly multiplies samples in the minority class but thus increases the risk of overfitting, the lack of generalization, within the minority class.

A more sophisticated way of increasing training samples in the minority class is synthetic minority over-sampling (SMOTE) (Chawla et al., 2002). SMOTE is based on the idea of creating new training samples in the minority class by interpolating between a sample and a random set of its k-nearest neighbors in feature space. Therefore, a new sample is generated with features similar to already existing samples. This increases the sample size of the minority class but minimizes the problem of overfitting. On the algorithm level, random forest opens two possibilities for imbalanced data: setting a class weight on the minority class or undersample the training data for every single tree, a so-called balanced random forest (BRF) (Lemaître et al., 2017). Here, we use a BRF classifier that undersamples the majority class and weighs the classes inversely to the number of samples in each class.

4.3.3 *Training Process and Evaluation*

We use two different metrics to evaluate model accuracy, the confusion matrix and receiver operating characteristic (ROC) curves. The confusion matrix consists of the true label of each class's samples as rows and the classifier predicted label as columns (Fig. 4.4b). For a perfect classifier, all samples are located on the diagonal of the matrix. Using the confusion matrix, the classifier can be evaluated for each class separately. Furthermore, we normalize the confusion matrix, such that the sum of each row is 1.0. The ROC curve uses the true positive rate (TPR) and false positive rate (FPR) for different probability thresholds (Fawcett, 2006). TPR is defined as the number of true positives divided by the sum of true positives (TP) and false negatives (FN) (TPR = TP/(TP + FN)). FPR is defined as the number of false positives (FP) divided by the sum of false positives and true negatives (TN) (FPR = FP/(FP + TN)). Class prediction of random forest is based on the score of a class, i.e., its probability defined by the number of predictions out of all trees. By lowering the threshold for classification, i.e., the probability threshold for a class to be predicted, FPR and TPR increase as FN samples transition to TP and TN samples transition to FP. As an example, we consider an imaginary two-class problem with a decision threshold of probability > 0.5 for the "positive" class leading to a TP = 1, FP = 2 , TN = 3 and FN = 4 and resulting TPR = 1/5 and FPR = 2/5. When lowering

the probability threshold for the "positive" class to be predicted to, say, 0.2, TP will increase, but so will FP, giving TP = 4, FP = 3, TN, = 2, and FN = 1. This results in larger values of TPR and FPR (TPR = 4/5, FPR = 3/5). When plotting FPR against TPR for each probability threshold, one obtains the ROC curve with a monotonous increase. For a schematic drawing, see Fig. 4.4b). The best-case scenario is a TPR of one, and an FPR of zero, i.e., a step transition from coordinates (0,0) to (0,1), whereas a random classifier would result in a diagonal from (0,0) to (1,1). The area under the curve (AUC) can be used as a one value metric for model performance. The larger the area under the ROC curve, the better the model accuracy. Both metrics, confusion matrix and ROC/AUC, can directly be transferred into a multiclass environment. For the confusion matrix, this simply results in several columns and rows. The ROC curve can be computed for each class separately by bundling all other classes together.

We use the confusion matrix as an initial performance evaluation of random forest for both the Illgraben and the Pizzo Cengalo data set. We then use the ROC/AUC analysis to estimate the influence of the window length and the imbalanced data set handling techniques on the Pizzo Cengalo data set. For a more representative measure, we use k-fold cross validation when computing the ROC curves (e.g., Stone, 1974). As our training data set only contains three events in the minority class, we use 3-fold cross validation, with random 2/3 of each class in the training data set and 1/3 used for validation. This way, we obtain three ROC curves and AUC values per class trained and tested on three different random subsets. We then take the mean TPR and FPR to plot the ROC curves. We computed the 95% confidence level using Student's *t*-distribution for small sample sizes (n=3).

4.3.4 *Simulated Real-Time Monitoring at Pizzo Cengalo*

To test how the classifier performs in a real-life application, we use a model trained on the 2018 Pizzo Cengalo data set and classify more than a million time windows of recorded seismic data at Pizzo Cengalo in 2019. We train, validate, and test the model on 2018 data, containing five slope failure events, and then use the model to classify 2019 data. Seismograms and spectrograms of the training events are shown in Appendix A.2. For 2019, no full event catalog is available. We cross-check as slope failure classified events with earthquake catalogs, hiker reports, and pictures from the automated camera. If none of these methods give clarity, we manually classify the seismic signals based on typical characteristics of slope-failure events, such as dominant frequencies between 5 - 10 Hz, an emergent onset, and a duration of several tens of seconds (e.g., Hibert et al., 2011).

We benchmark our test with the two-step approach of STA/LTA detection in the frequency domain and classification of the detected events using random forest. After extensive testing, we define the parameters that provide accurate detection for our data set as an STA window length of 1s and an LTA window length of 18s. The detector turns on when the STA/LTA ratio exceeds four and turns off when the STA/LTA ratio becomes lower than two. Additionally, we use a coincidence trigger, with a threshold of three, which means that the STA/LTA threshold needs to be exceeded at all three stations.

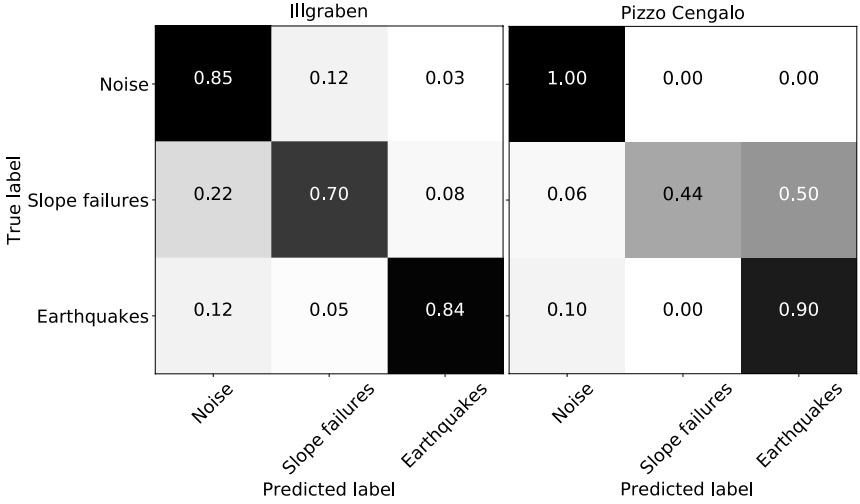


FIGURE 4.5: Confusion matrices of an initial test of the continuous random forest approach on 20 s windowed data. Left: Illgraben data set, right: Pizzo Cengalo data set.

4.4 RESULTS

4.4.1 *Classifier Performance on Labeled Data Sets*

For an initial evaluation, we tested the random forest classifier's performance on the Illgraben and the Pizzo Cengalo data set. We used a randomized grid search to obtain the best performing hyper-parameters such as number and depth of decision trees. The final parameters are presented in Tab. A.6. Normalized confusion matrices for both data sets are shown in Fig. 4.5. We trained the models on 20s window sizes, and no class balancing has been applied. For the Illgraben data set, the classifier correctly classifies between 70% and 85% for all classes. The largest confusion occurs between the NO and SF class. For the Pizzo Cengalo data set, the classifier correctly classifies between 90% and 100% for the NO and EQ class, but only 44% of the SF class. In this case, the largest confusion occurs between the SF and EQ class.

4.4.2 *ROC Analysis on Pizzo Cengalo Data Set*

We computed ROC curves and AUC values for different window sizes and under and oversampling techniques for the Pizzo Cengalo data set (Fig. 4.6). Figure 4.6a shows the AUC values for the SF class plotted in a heatmap. Rows show different window sizes (10s - 60s) and columns different techniques. Associated standard deviations are shown in Fig. 4.6b. Darker colors mark a larger AUC value and a smaller standard deviation. The different color maps highlight the difference between the AUC values and AUC standard deviations. Although almost all values lie within the confidence intervals, overall, the smallest window size and the largest window size give slightly better values with a

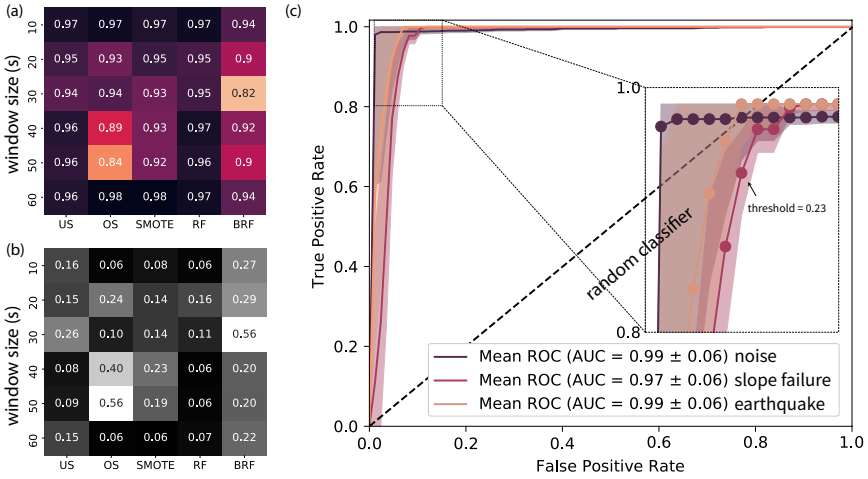


FIGURE 4.6: ROC analysis on the Pizzo Cengalo data set. (a) Heatmap of AUC values for SF class for different window sizes and different methods to handle imbalanced data sets, namely undersampling (US), oversampling (OS), synthetic minority oversampling (SMOTE), random forest with original data set (RF), balanced random forest (BRF). The darkest colors denote the largest AUC values. (b) Heatmap of 95% confidence interval of AUC values. Darker colors denote smaller standard deviation (c) ROC curve for 40 seconds time window and random forest. 3-fold cross validation with mean as solid line and 95% confidence interval. Zoom into corner with circles illustrates TPR and FPR values for different thresholds. Probability threshold for TPR rate of SF class > 0.9 is depicted.

small standard deviation. Additionally, the AUC value for a simple random forest and balanced random forest (BRF) is higher without the data manipulation techniques.

As a compromise between large AUC value and small standard deviation, we choose 40-second windows and classical random forest, i.e., without modifications for handling imbalanced data. The ROC curve of this configuration is shown in Fig. 4.6c. To make sure we correctly classify most of the windows containing a slope failure signal, we set the target TPR to > 0.9. From the 3-fold cross-validation ROC analysis, we obtain a mean probability threshold of 0.23 for a TPR > 0.9 in the SF class. As a next step, the optimal model parameters (i.e., number of decision trees, number of features chosen for each tree, maximum tree depth, ...) for the 40 seconds window size and random forest are chosen using the randomized cross-validation search of scikit-learn (Pedregosa et al., 2011). We use the obtained classifier to classify the test data set from the 2018 data, containing two SF events, which were not used for the model set-up and, therefore, an unbiased evaluation of the classifier. We set the probability threshold for the SF class to 0.23 as obtained from the ROC curve. Consequently, for probabilities higher than 0.23 for the SF class, the window will be classified as slope failures, even if another class has a higher probability.

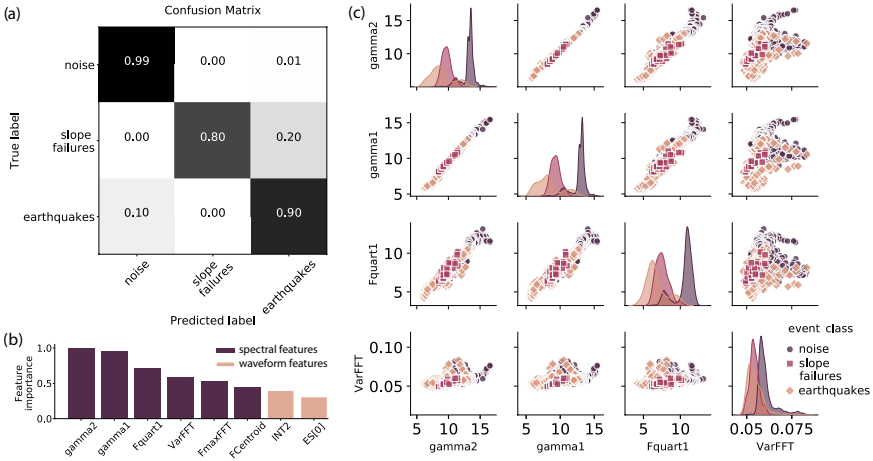


FIGURE 4.7: Confusion matrix and feature importance analysis of the optimized classifier for the Pizzo Cengalo data set. (a) Normalized confusion matrix of final model test. The darker the colors, the higher the values. For an ideal classifier, all samples would be located on the diagonal. (b) Eight most distinct features normalized to one. Dark columns mark spectral features (characteristics of signal in frequency domain), light columns mark waveform features (characteristics of signal in time domain). Labels: spectral gyration radius (gamma2), spectral centroid (gamma1), central frequency of the first quartile (Fquart1), variance of the normalized FFT (VarFFT), frequency at the maximum of the FFT (FmaxFFT), frequency at spectrum centroid (FCentroid), energy of the last 2/3 of the autocorrelation function (INT2), and the energy of the seismic signal in the frequency band of 1-3 Hz (ES[0]) Provost et al., 2017. (c) Pairplot of four most distinct features. Per cell two features plotted against each other, except for diagonal. Diagonal shows univariate distribution of the feature. Colors mark different event classes

4.4.3 *Classifier Performance with Classification Threshold on Pizzo Cengalo Data Set*

We used the low probability threshold from the ROC analysis to label all 40 seconds time windows of the Pizzo Cengalo test data set. The results of this set-up are shown in Fig. 4.7. The normalized confusion matrix (Fig. 4.7a) shows a misclassification of 20% for slope failures. Additionally, 10% of earthquakes are classified as noise. The misclassification rate of noise is however very small (1%). The most discriminating features are presented in Fig. 4.7b. The colors denote spectral and waveform features. Distinctive features are spectral gyration radius (gamma2), spectral centroid (gamma1), central frequency of the first quartile (Fquart1), variance of the normalized fast Fourier transform (FFT) (VarFFT), frequency at the maximum of the FFT (FmaxFFT), frequency at spectrum centroid (FCentroid), energy of the last 2/3 of the autocorrelation function (INT2), and the energy of the seismic signal in the frequency band of 1-3 Hz (ES[0]) (Tab. A.1, Provost et al., 2017). Figure 4.7b) shows that the by far most discriminating features are characteristics in the frequency domain.

4.4.4 *Classifier Implementation on Pizzo Cengalo Seismic Data*

As a next step, the model is used to classify seismic data recorded at Pizzo Cengalo in 2019, mimicking operational conditions of a near real-time classification. We first compute the signal features of 40 second time windows with an overlap of 2/3 of the window length for each station and perform a classification. Next, a majority vote of the stations is performed, and a label is assigned to the time window. This means that if more than one station assigns the time window to the same class, the end label is chosen accordingly. In case every seismic station classifies the same time window into a different class, the time window will be labeled as noise. We compare the results to an event catalog compiled from hiker reports, manual classification of seismic data and automatic camera images. The classification parameters for the manual classification of the seismic data are consistent with creating the training catalog.

In 176 days in 2019 (Julian Day 94 to 270), 21 days have at least one window classified as slope failure. To exclude misclassified windows because they only contain a small signal portion, we set a minimum threshold of three consecutive SF classifications. With this threshold, we limit the number of slope failure detections to eight. Out of these eight, three correspond to manually picked slope failures, two of which happened on 4 and 26 April and one on 16 July. However, the automatic camera shows that the April events are snow avalanches rather than rock slope failures (Fig. 4.8). Seismic waveforms and associated classifications for 16 July are shown in Fig. 4.9, with a zoom on a rockfall event (c-e) and a noise signal (f-h). Four events that were classified as slope failures are earthquakes on 22 May, 29 July, 8 August and 29 August. Two out of these earthquakes originate from the German lakeside of Lake Constance about 160 km north-west of Pizzo Cengalo, with a distance between the epicenters of about 3 km and magnitudes of 3.6 and 3.4 (EMSC catalog). The two other earthquakes are Magnitude 3.3 and 2.2 earthquakes with epicenters about 170 km south and 33 km west of Pizzo Cengalo, respectively (EMSC catalog). The last event classified as slope failure on August 13 is characterized by a duration of 10 s and is not listed in any earthquake catalog. All waveforms, spectrograms, and spectra are shown in Fig. 4.10.

To quantify the performance of the algorithm presented here, we benchmark the continuous approach against a two-step approach with an STA/LTA detection on the Pizzo Cengalo data set. The continuous approach correctly classifies three slope failures (TP), misses one slope failure (FN), and classifies four earthquakes as slope failures (FP). The two-step approach of STA / LTA detection correctly classifies two slope failures (TP), misses two slope failures (FN), and classifies six earthquakes as slope failures (FP). For a simple comparison, we can use the critical success index (CSI = TP / (TP + FN + FP)) which ignores all non events (TN). For the Pizzo Cengalo data set, we obtain a CSI of 0.375 for the continuous approach, whereas, for the STA/LTA approach, we obtain a CSI of 0.2.

4.4.5 *Transferability to other study sites*

We tested the transferability of a trained model to other study sites by assessing the performance of a model trained on one site and tested on another, and trained on both sites and tested on one site. The result of all possible combinations is shown in Fig. 4.11. A classifier trained on one site and tested on another reduces the mean score over each

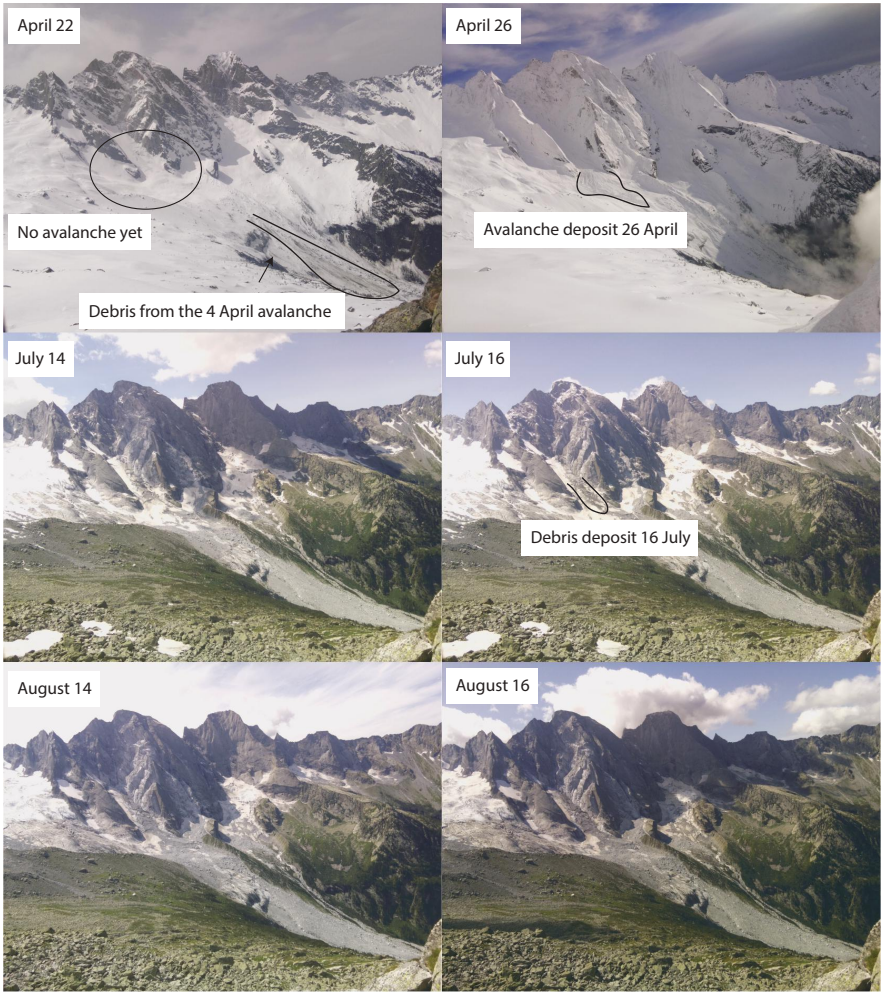


FIGURE 4.8: Automatic camera photos of slope failures in 2019. Photos from April show the avalanche deposits from the events classified as slope failures of 4 April and 26 April. Photos from July and August show photos before and after the events of 16 July and 14 August. For the 14 August event, no deposit is visible on the photos.

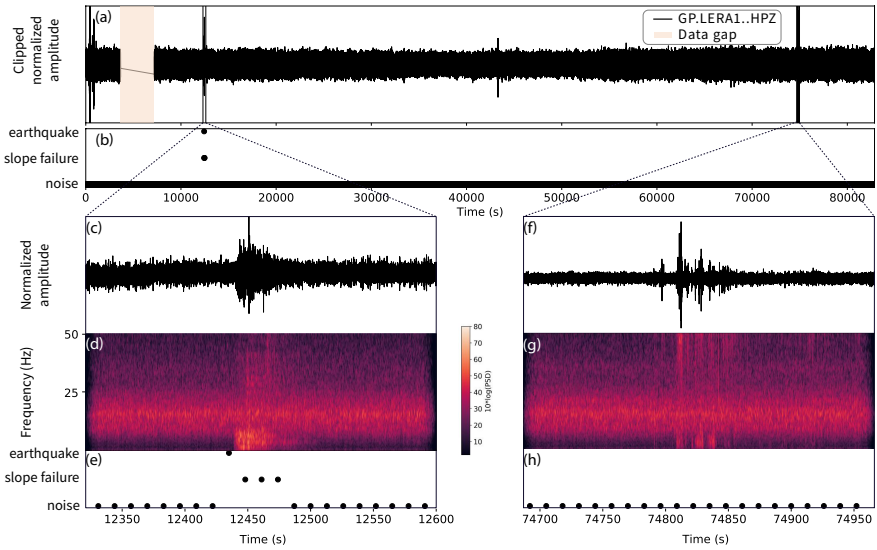


FIGURE 4.9: Classifier tested on one day of data from Pizzo Cengalo in 2019 (July 16). (a) Waveform of one day, with data gap (orange area) (b) label of each 40 seconds time window. (c) Waveform and (d) spectrogram with spectral power of a slope failure and (e) associated classifications. (f) Noise event with (g) spectrogram and (h) associated classifications.

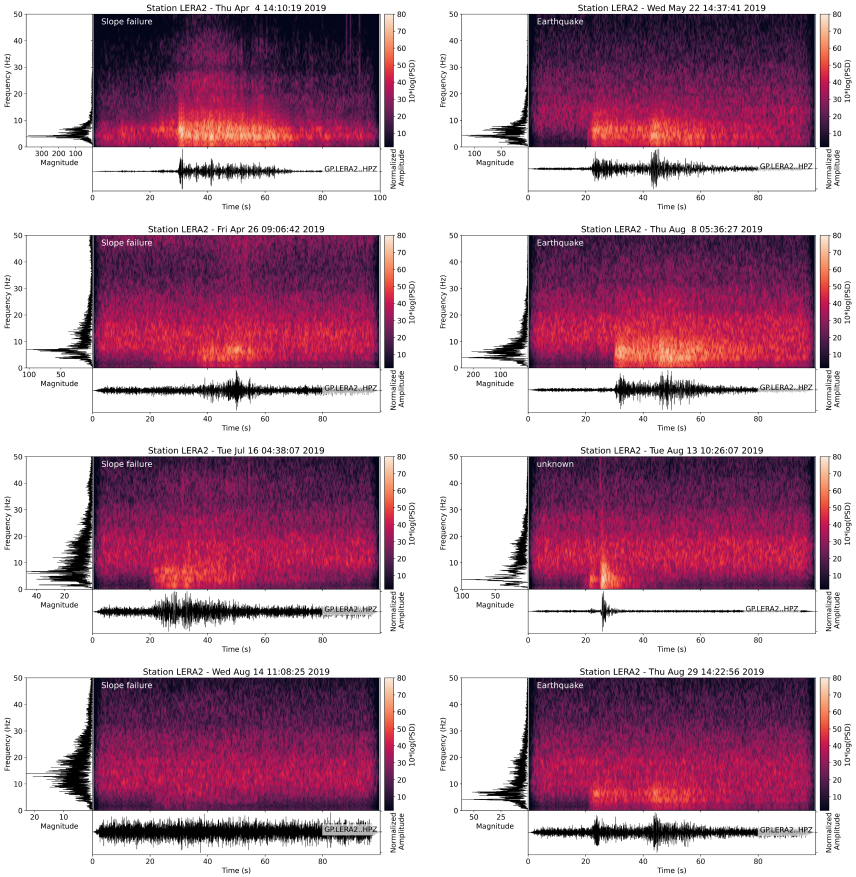


FIGURE 4.10: Waveforms, spectrograms and spectra of all events classified as slope failure in 2019 at Pizzo Cengalo, as well as the missed slope failure on August 14, 2019 (bottom left). White labels on spectrograms mark the type of event.

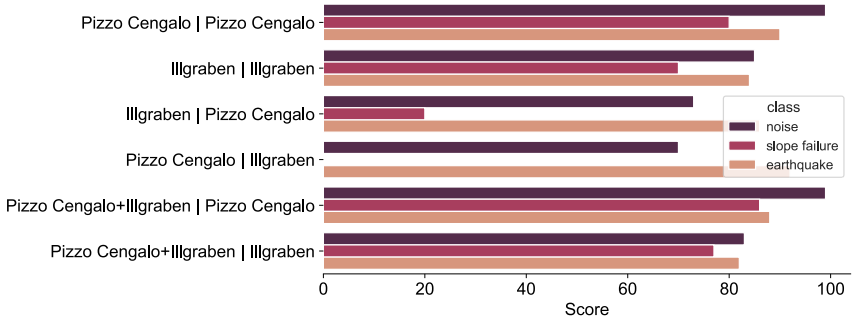


FIGURE 4.11: Scores for each class with different training and testing data sets. Labels on the y-axis show the data set(s) used for training before the vertical bar, and the data set used for testing after the vertical bar.

class by about 30 % for the Pizzo Cengalo and the Illgraben data set. Especially the score of the minority class, the slope failure class, is reduced to 20 % and even 0 % respectively. For both the Illgraben and the Pizzo Cengalo data set, a classifier trained on both data sets increases the mean test scores over all classes only marginally. However, in both cases, the slope failure class's test scores increase by nearly 10 % compared to a classifier only trained on the respective data set.

4.5 DISCUSSION

4.5.1 *Data Set Comparison*

We trained a random forest classifier on two different types of data sets: the Illgraben data set, with a balanced abundance of all types of classes, and the Pizzo Cengalo data set, with a highly imbalanced number of events. Both the Illgraben and the Pizzo Cengalo data sets were aggregated by manual inspection of the seismic data and, for the Pizzo Cengalo data set, direct observations at the site. Even though both data sets have been carefully examined, a misclassification of events cannot be excluded. For earthquakes, a misclassification is highly improbable, as all events were cross-checked with publicly available earthquake catalogs. However, cross-checking databases for rockfall events are rare, especially for events that have not affected infrastructure. Therefore, an inherent bias by manual labeling and falsely labeled events is possible but can hardly be avoided for these data sets.

The Illgraben data set contains only four weeks of manually labeled data, but in those four weeks, the number of recorded slope failures and earthquakes is nearly the same (Fig. 4.3a). The large number of slope failure events during that period can be explained by three extreme precipitation events that also triggered three debris flows (e.g., Wenner et al., 2019). Under these conditions, slope failure activity increases drastically.

The Pizzo Cengalo data set contains manually labeled data of the summer period (June - October). The LERA network was set-up in the aftermath of a large rock avalanche event

in 2017. Since then, slope activity has strongly decreased. This implies that automatic detection and classification is based on a small number of training events in the slope failure class and a comparably large number of events in other classes. The small number of recorded earthquakes can be attributed to general poor data quality. Many recorded earthquakes barely exceeded the noise level and were therefore not included in the catalog.

For slope failures, the Pizzo Cengalo data set includes fewer events than the Illgraben data set. However, the mean SNR is higher. Overall, SNR of events in both data sets compare to SNR found in other studies (e.g., Dammeier et al., 2016). The SNR distribution in both data sets is also reflected in the initial classification test (Fig. 4.5). In this test, no performance enhancement was applied. Whereas the classifier performs similarly well for each class for the Illgraben data set, 22% of the slope failure signals are misclassified as noise. The large number of misclassification could be related to the low SNR for slope failures in the data set. For the Pizzo Cengalo data set, the classifier performs significantly worse for the slope failure class, but the confusion is more pronounced between the slope failure class and earthquake class than slope failure and noise. Hence, the Pizzo Cengalo data set's misclassification might not result from poor SNR, but the underrepresentation of the slope failure class. We expect an improvement of the classifiers performance after adding slope failure seismic signals of future events to the training data set.

4.5.2 *Performance Enhancement*

One way to enhance the classifier's performance on the Illgraben data set could be to add network and polarity features. Provost et al. (2017) have shown that network features count among the most important features for classifying detected events. However, the set-up of the LERA network at Pizzo Cengalo has an unfavorable aperture to detect and classify detected seismic signals, as the stations are set-up in a line with interstation distances of 9 m and 31.5 m. This prohibits the usage of network characteristics, like arrival time differences and amplitude ratios.

To address the problem of the imbalanced Pizzo Cengalo data set and resulting misleading scores, we test several techniques to handle such data sets and use receiver operating characteristic curves for performance assessment. The area under the curve for SF is largest for a generic random forest. Further assessment shows that SF's true positives are largest when using a technique to handle imbalanced data sets, but leading to a substantial increase in EQ events being classified as SF. Using a generic random forest, SF is underrepresented, ending up with zero true positives but also zero false positives. By lowering the probability threshold for SF, the true positive rate increases, whereas the false positive rate stays low. Therefore, for this data set, we decided to ignore problems with imbalanced data sets and mitigate misclassifications by lowering the probability threshold. It remains to be seen if this approach performs equally well on other data sets, but in our case, it gives the best results by maximizing the number of true positives in SF and minimizing the number of EQ classified as SF. This does not mean, however, that imbalance countermeasures did not work. Compared to the initial test (Fig. 4.5b), any of the measures mentioned above improve the classification result.

We choose to use a 40s window as an operative window size to test on the 2019 data, even though 10s and 60s show equally high AUC values and low standard deviations. On a continuous data stream a 10s window with 6s of overlap does not leave enough

time to compute features and classify the event in real-time. The 60s window, on the other hand, results in a classification delay of one minute and we assume that chances are higher to miss smaller events which are masked by a large amount of noise in the 60s window.

Generally, an imbalanced data set can be tackled by increasing the amount of training data in the minority class. A classifier trained for an area that is more active or has been monitored during a longer period is expected to give better results with higher accuracy. Additionally, the small number of events in the slope failure class can lead to overfitting, i.e., an insufficient generalization of the model. Hence, small deviations in signal characteristics can lead to misclassification and undetected slope failures. However, as seen on tests on the Illgraben data set, solely a larger training data set will not give a perfect classifier either. We attribute the low SNR to the classifier's relatively poor performance of 30% misclassified events in the slope failure class at Illgraben.

4.5.3 *One-step vs. two-step method at Pizzo Cengalo*

Several studies have shown that classification algorithms accurately classify events detected with the STA/LTA approach (e.g., Hibert et al., 2017; Provost et al., 2017). The benchmark analysis performed on the Pizzo Cengalo data set indicates, that our continuous approach (CSI of 0.375) performs slightly better than a two-step approach (CSI of 0.2). However, the small number of events prohibits a statement on robustness. Interestingly, there is a large overlap in the earthquakes being misclassified as slope failures between the two approaches.

Feature importance analysis for the continuous one-step approach presented here shows that the classifier predominantly uses spectral features to distinguish between different classes. This is consistent with the fact that the windowing eliminates information from the entire waveform. Provost et al. (2017) showed that for the two-step approach, several waveform features, e.g., duration and the ratio between ascending and descending time of the signal, are powerful distinctive features of slope failures and earthquakes. These are, however, characteristics of the entire waveform of an event. In our case, constant window size with start and end regardless of event start and end sacrifices this information, and the classifier, therefore relies on spectral features. However, the spectral content of earthquakes and slope failures at our site is highly similar, which complicates a correct classification (Fig. 4.2). This is illustrated by the univariate distributions and correlations in Fig. 4.7c, that show large overlaps between the classes, even for the most discriminating features.

Despite the loss of several waveform features, the continuous approach outperforms the two-step approach in our case. We attribute the enhanced performance of the continuous approach to the reduced parameter tuning effort: In the two-step approach, the performance of the classifier is strongly influenced by the STA/LTA detectors accuracy. Tests show that a manually picked catalog achieves up to a 15% higher accuracy than the same catalog compiled with an STA/LTA algorithm.

Even though not a focus of this study, we note that STA/LTA detection algorithms tuned to detect short signals (several tens of seconds) miss events of long duration and gradual amplitude increase, such as debris flows, volcanic tremors, lahars, and glacier lake outburst floods. Coviello et al. (2019) show that with a window size of 10s and 100s for STA and LTA, respectively, debris flows can be detected, excluding other events like

earthquakes. However, this also excludes the detection of short slope-failure signals. The continuous approach is capable of detecting such events and is therefore applicable in multiple contexts and different sites. For example, intense precipitation raises the noise level by an increase in runoff and, consequently, seismogenic sediment transport (Tsai et al., 2012; Burtin et al., 2016). Similarly, snow cover and strong temperature fluctuations can affect the instrument itself and change the noise level. A preliminary implementation of a classifier with a fourth class called runoff trained on two days of increased water discharge (measured with gauges) found two additional days of peak discharge. Using the two step-method of STA/LTA, requires a second STA/LTA algorithm with its own parameters to detect these signals. Consequently, applying continuous random forest in different circumstances is potentially a low effort, as there is no need to fine-tune the detection algorithm while improving the overall results over the two-steps approach.

4.5.4 *Test on 2019 Pizzo Cengalo data*

Continuous random forest correctly classifies events in the test data with a high SNR. A slope failure that was observed by hikers on 14 August was classified as noise. Often events create dust clouds that are easily noticeable despite a small mass-movement volume. The misclassification likely results from a low SNR (Fig. 4.10), hence a probably relatively small volume of the event, as all windows containing the event were classified as noise. For most waveform and spectral features, especially the most discriminating ones (feature importance analysis), values of time windows containing the slope failure signal do not differ from that of windows that contain only noise. Filtering of the waveform to minimize noise is difficult, as the frequency band of the slope failure coincides with the primary noise. Furthermore, the automatic camera pictures show no apparent detachment zone (Fig. 4.8). This either means that the break-off happened outside of the camera field or validates our assumption of a small volume event.

Two of the events classified as slope failure are snow avalanches in April 2019 (Fig. 4.8). At this time, Pizzo Cengalo's active slope was partially covered by snow. As both avalanches happened during snowfall periods with obscured view from the automated camera to Pizzo Cengalo, the events were only validated by pictures several days after the actual event. The random forest classifier was only trained on data from the summer period, and therefore, snow avalanches were not part of the training data set. However, the seismic signature of snow avalanches is highly similar to those of other mass movements (e.g., Suriñach et al., 2001; van Herwijnen & Schweizer, 2011; Heck et al., 2018). This also validates your assumption that signal differences between different mass movements are less prominent than differences associated with noise and earthquakes. Hence, for training purposes, snow avalanches could be included in the slope failure class. This would allow the detection of slope failure events in winter and in summer month and increase the number of events in the slope failure class, resulting in a less imbalanced training data set.

We chose a minimal of three consecutive windows classified as slope failure for a slope failure event detection. If we were to increase the number to four, the number of slope failure event detections is lowered to four. Out of these four, only one is an actual slope failure and three are earthquakes. If we were to decrease the number of consecutive windows that need to be classified as slope failure to one, three more earthquakes would cause a slope failure event detection. Even by lowering the number of windows, the

missed slope failure event would still not be detected. Thus, the classifiers' performance in correctly classifying three slope failure events is overshadowed by the four false alarms despite the consecutive window threshold. For operational use, a false alarm ratio $>$ 50 % is not acceptable as operators will lose confidence in the classifier.

In a promising approach, Hammer et al. (2017) use Hidden Markov Models to detect snow avalanches by training a background model from continuous seismic data and using seismic signals of only one avalanche to distinguish noise from avalanches. However, the model was only tested on five days of data. Continuing with this approach Heck et al. (2018) classify more than 100 days and find that for a reliable classification with a small number of false alarms, a daily update of the background model, as well as extensive array-based post-processing is necessary. With our approach, using random forest, no retraining of the background model is necessary and thus, the computational cost for a single parameter tuning is negligible. For operational use at the Pizzo Cengalo site, array-based post-processing as proposed by Heck et al. (2018) is not possible. Amongst other array-based post-processing steps Heck et al. (2018) include detections from a seismic station 14 kilometers away from the test site in the decision process to discriminate between noise signals and avalanches. We explore this approach with our classification problem of earthquakes falsely classified as slope failures. We train a model for a seismic station of the Swiss Seismological Service (CH.FIESA) located 130 km away from Pizzo Cengalo. At this distance local and regional earthquakes that were misclassified as slope failures at Pizzo Cengalo are still recorded, but not slope failures. We use the EMSC earthquake catalog to compile a training data set containing 45 earthquakes and 513 noise samples in 2018. We then train the model on 20s windows of seismic data. For the decision-making process, we test the trained model on seismic data recorded when a slope failure was classified at Pizzo Cengalo in 2019. All four earthquakes falsely classified as slope failures at Pizzo Cengalo were classified as earthquakes at FIESA. All time windows around the slope failures at Pizzo Cengalo were classified as noise at FIESA. Therefore, a simple inclusion of a station at a regional distance to Pizzo Cengalo reduced the false alarms to zero.

Our approach is computationally inexpensive, as the classification model only has to be trained once. From there on, computational power is only needed to compute features of the 40s windowed seismic data. For this near-real-time simulation, we used a standard machine (2017, Inter Core i7), which was able to compute all features and classify the windowed data within 3 seconds. Combined with the consecutive window threshold, this allows for a short warning time within tens of seconds after the event start

4.5.5 *Classifier Transferability*

For future applications, a key question is how a classifier trained on one data set can be transferred to other environments and study sites. Especially for sites previously not monitored, a classifier trained on other sites could give a head start in seismic monitoring. Figure 4.11 clearly shows that a classifier trained on one site and tested on another performs poorly. However, for both the Illgraben and the Pizzo Cengalo data set, a classifier trained on both data sets increases the test scores for all classes, and especially the minority class, compared to a classifier only trained on the respective data set. Therefore, we suggest that cross-context training and data set amalgamation has the potential to improve classification outcome and should be investigated in future work.

4.6 CONCLUSION

In this study, we apply a random forest classifier on windowed seismic data to detect and distinguish between noise and seismic signals of slope failures and earthquakes. We test our workflow on two data sets and explore ways to improve its performance on the imbalanced data sets. The improvement techniques allow us to overcome an obstacle that often occurs in natural hazard detection: the deficiency of training data. The advantage of random forest compared to previously suggested approaches using Hidden Markov Models is its simplicity and relative ease of implementation in terms of parameter tuning, post-processing, and computational cost.

We developed a new method to process continuous data streams in near-real-time, which combines detection and identification of rare events. Additionally, this algorithm can outperform a two-step STA/LTA detector and event classifier. The high number of true positives gives us confidence to detect slope failures. However, sub-optimal network configuration, similar frequency content generated by different sources, and low SNR lead to a high false alarm rate. Inclusion of classification results from a second seismic station at regional distance to the study site significantly reduces false positives and is therefore advised for operational use.

We show that two different data sets run into two different types of problems: Low SNR of targeted class and an imbalanced data set. Training data manipulation or small adjustments to the classifier can both mitigate poor classification results due to an imbalanced data set.

Our approach enables us to detect the occurrence of rare events of high interest in a large data set of more than a million windowed seismic signals. Our model, trained on rock slope failures, also detected snow avalanches. It therefore seems that this method is well suited to detect mass movements in general.

Manual review of seismic data is a tedious task, and especially for non-experts, uncertainty and misclassification rates can be high. An automatic classifier, however, can run in the background on a standard machine (in our case 2017, Intel Core i7) and alert stakeholders in case of an event classified as slope failure. Our implementation of a machine learning algorithm for seismogenic mass movement detection may therefore in the future provide valuable support to natural hazard management.

AUTHOSCONTRIBUTION

LM with Geopraevent installed the three seismometers and assured data transmission. CH provided the feature computation code. MW processed and analyzed the data with the help of CH and FW. MW prepared the manuscript with contributions from all co-authors.

ACKNOWLEDGEMENTS

The project is funded by WSL's strategic initiative Climate Change Impacts on Alpine Mass Movements (CCAMM). FW's salary was funded by the Swiss National Science Foundation (GlaHMSeis Project PP00P2_157551 and PP00P2_183719). The canton of Grisons partly payed MW's salary. We thank Geopraevent Ltd., who installed and

maintain the seismic stations, provided the data, an online data portal and implemented the algorithm to run in real-time. We are thankful for constructive feedback from Velio Coviello and an anonymous reviewer on an earlier version of the manuscript. We also want to acknowledge the Obspy developer team Beyreuther et al., 2010; Krischer et al., 2015 for providing an easy to use framework for seismic data handling.

5

SUMMARY OF PUBLICATIONS

This work is composed of three independent publications. Here, I will briefly summarize the content of each of them.

DECIPHERING DEBRIS-FLOW SEISMOGRAMS AT ILLGRABEN, SWITZERLAND

In this work, I investigate the seismic signals of seven debris flows recorded at Illgraben in 2017 and 2018. I use the recordings of a seismic station close to the channel to compare spectral characteristics of the high-frequency signal (> 1 Hz) to debris flow dynamics and parameters. I find that depending on the distance of the most seismogenic part of the debris flow toward the station, the frequency content differs drastically. For example, in the case of Illgraben, a large 48 m check dam causes high power in the frequency band of ~ 5.5 Hz while the debris-flow front passes the check dam. Additionally, I use a model relating the power spectral densities of debris-flow seismograms to grain size in the flow to give an estimate of grain sizes. Videos of the debris flows mostly confirm what is reflected in the seismograms. I conclude that to estimate such parameters, it is crucial to concentrate on small-time windows instead of the whole seismogram, to account for the changing source-to-receiver distances.

SEISMOMETER RECORDS OF GROUND TILT INDUCED BY DEBRIS FLOWS

Away from the high-frequency component of the debris-flow seismograms, I here investigate low-frequency seismic data ($< 1/5$ Hz). First, I use data from the USGS debris-flow flume to show that seismometers do indeed record tilt. For this, I use seismic recordings from six debris-flow flume experiments for which a seismometer and a tiltmeter were installed close to each other. The signals on both instruments correlate well with each other when filtered below 5 s, confirming that the seismometers are sensitive to tilting. For the Illgraben data, I adjust my technique to detect tilt signals that are way below the corner frequency of the instrument. I then compare the maximum tilt recorded with the seismometer to debris-flow parameters, such as volume, flow height, vertical force, and flow velocity, recorded on nearby force plates. The Pearson correlation coefficients (PCCs) reach up to 0.97 for debris-flow thickness but is also as low as 0.36 for some stations and parameters (notably debris-flow velocity). In all cases, the debris-flow height correlates well with the recorded tilt (PCC > 0.76). Additionally, I model the tilt based on a distributed load model using superposition of vertical displacements modeled by the solution of the Boussinesq problem. The sub-surface parameters (shear modulus and Poisson's ratio) have to be defined as model input. As they are mostly unknown for both the flume and Illgraben, we include an uncertainty estimation. I find that for the USGS flume, the observed tilt fits better to mechanical parameters expected for weathered volcanic bedrock, rather than soil. For Illgraben, the observed tilt is well within the limits

of the model uncertainty. Lastly, I use a simplified version of the model to predict the flow height of two debris-flows at Volcán de Fuego , Guatemala.

NEAR REAL-TIME AUTOMATED CLASSIFICATION OF SEISMIC SIGNALS OF SLOPE FAILURES WITH CONTINUOUS RANDOM FORESTS

Instead of retrieving information on debris-flow dynamics and bulk parameters of the flow from the seismograms, I here use the seismic signals of various slope failures to create an automatic event detector. For this, I use data recorded at three seismic stations in Val Bondasca, close to Pizzo Cengalo. The automatic detection algorithm is based on a random forest classifier. Two main difficulties have to be faced while training the machine-learning algorithm: (1) data are available from one season only (April to September 2018), and contains only five slope failure events; (2) the network of stations is set up in one line, with very small inter-station distances (30 to 50 m). This results in similar seismic signals on all stations, as the wavelength is larger than the inter-station distance for the frequency band of interest, and precludes the use of network parameters, such as different arrival times of seismic waves. To tackle the problem of an imbalanced data set (small number of events in the class of interest), I explore several techniques to handle biased data sets. I find that a general random forest algorithm with lowered classification threshold for the class of interest, performs best on this data set. Additionally, I compare the performance of a classifier trained and tested on this data (with small number of slope failures but high signal to noise ratio) with the performance of a classifier trained and tested on an event catalog from Illgraben, with a larger number of slope failure events but a lower signal to noise ratio. I find that, without accounting for the imbalance in the Pizzo Cengalo data set, the classifier performance is better for the Illgraben data set. Lastly, in a simulated real-work application on data from 2019, I find that events falsely classified as slope failures are medium magnitude earthquakes (ML 3 - 4) at a distance < 300 km. I show that by (i) including a seismic station at great distance to the area of interest (> 100 km), (ii) training a separate model to detect earthquakes on this station, and (iii) cross-checking whether an event was detected on both the Pizzo Cengalo network and the other seismic station, these earthquakes can be detected and successfully discerned from local slope failures.

6

OUTLOOK

In this section, I present additional results and findings of my doctoral project, which have not yet been published in peer-reviewed journals. I first show a reevaluation of the Pizzo Cengalo seismic event classifier presented in chapter 4, given two years of additional data and operational mass-movement monitoring. Second, I present the possibilities, which machine-learning classifiers open for automatic long-term monitoring of sediment supply at Illgraben.

6.1 PIZZO CENGALO CLASSIFIER

6.1.1 *Classifier performance*

In chapter 4, I described the training and testing process of a supervised machine learning classifier to detect slope failures at Pizzo Cengalo. This classifier was implemented on a local machine situated close to the LERA array and running since February 2020 (setup described in chapter 4). Only one reported slope failure occurred over the following one and a half years (February 2020 to August 2021). The seismic signal, spectrogram, and amplitude spectrum of the event are shown in Figure 6.1. The event happened at the neighboring Piz dal Gal, about 4 km away from the LERA array, and its volume was estimated to be 10,000 - 20,000 m^3 (personal communication Lino Schmid, GeoPraevent). The detector correctly classified the signal and issued a slope failure alarm. However, over the entire period since February 2020, the system issued in total six slope-failure alarms. The other five false alarms were identified as earthquakes by manually cross-checking with the European-Mediterranean Seismological Centre (EMSC) earthquake catalog. Similar to misclassifications described in chapter 4, these earthquakes are all of relatively high magnitude (between 3.4 and 4.4 local magnitude), with the epicenter close to Pizzo Cengalo and located in northern Italy.

When the signal within a time window is classified as earthquake or slope failure, a plot is generated that showcases the signal, its spectrogram, and its amplitude spectrum. The plot facilitates the stakeholders' decision-making process as well as the supervision of the classifier output. This way, plots were generated for 638 events in the period since February 2020. Out of these events, 248 were classified as earthquakes whilst, in fact, they were periods of elevated noise. On the other hand, 313 earthquakes were correctly classified. After manual waveform inspection and cross-checking with earthquake catalogs, signals of 36 seismic events classified as earthquakes can most likely be attributed to small slope failure events. Eighteen slope failure events were correctly classified as slope failures, and 23 events classified as slope failures are actual earthquakes. No noise was classified as slope failure. All other periods were classified as noise and were not reevaluated. The upper panel in Figure 6.2 shows time windows classified as either earthquakes or slope failures during the recording period from February 2020 to July 2021. The dark vertical lines mark the slope failure alarms. The middle panel

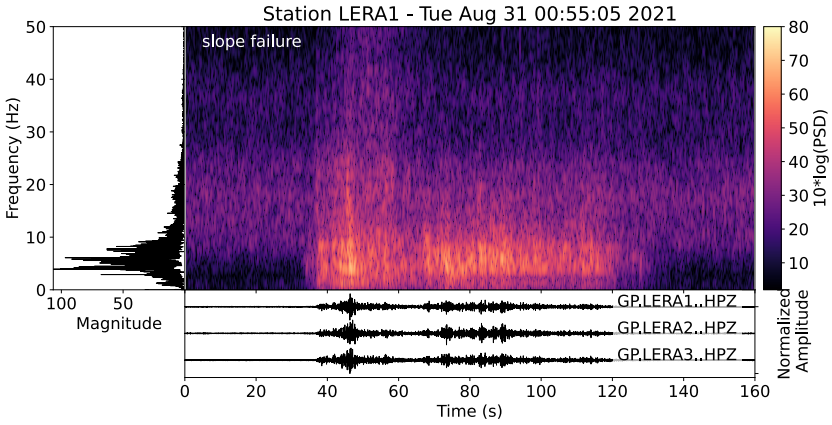


FIGURE 6.1: Spectrum, spectrogram, and waveform of the correctly classified slope failure event that happened on August 31, 2021 recorded at station LERA1-3 at Pizzo Cengalo.

in Figure 6.2 shows the manually corrected catalog by comparison with earthquake catalogs and evaluation of waveforms and spectrograms of the events. There is a clear trend towards fewer registered events during the summer months (May to October). To evaluate if fewer events happened during this time or if this gap is caused by sensor or detector performance, I compare the detected events to the EMSC earthquake catalog. For the comparison, we considered earthquakes with ML $<$ 4.5, and source-receiver distances up to 450 km. All earthquakes from this catalog are shown in the lower panel of Figure 6.2, revealing that there is no clear trend in the occurrence of earthquakes during this time. This hints towards an increase in noise during the summer months as the cause of reduced detector sensitivity and hence lower detection rate. This difference will be discussed in section 6.1.3.

Figure 6.3 shows the noise level from February 2020 until the end of July 2021, showing a substantial increase in background noise starting in May and lasting until November (Julian days 121 - 304). The increase is strongest in the frequency range between 5 to 20 Hz.

6.1.2 *Retraining*

After analyzing the past performance, I retrained the model by adding the new events to the training catalog. The original data set used in chapter 4 consisted of only five slope failure events. This causes a considerable bias in the training and testing process if not handled carefully. During the operational period in 2020 and 2021, only one confirmed slope failure was recorded. The other slope-failure events in the catalog are not confirmed by auxiliary observations and are often small events with a low signal-to-noise ratio. For this reason, I include only the single confirmed slope failure event in the training data. This has the disadvantage that the model becomes even more biased toward the noise and earthquake class. Despite this drawback, I perform the retraining to reduce

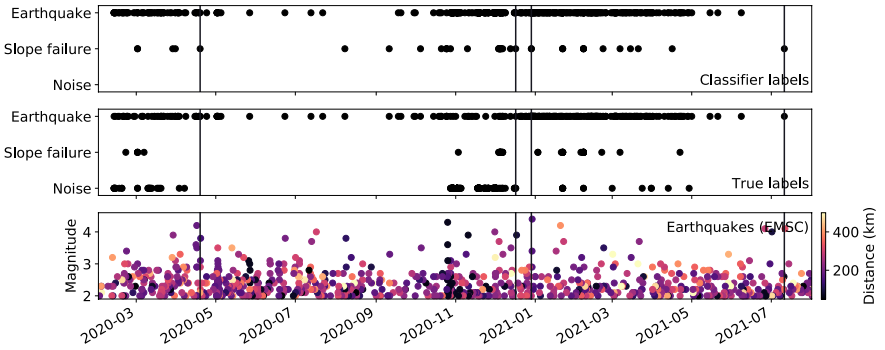


FIGURE 6.2: Upper panel: Classifier output of year 2020 and first half of 2021. Y-axis shows the labeled class: Noise, slope failures, and earthquakes. Time windows that were classified as noise were not plotted. Middle panel: Corrected labels (true labels) of slope failure and earthquake events detected by the classifier. Lower panel: Earthquake activity during the recording time. Y-axis shows the magnitude of the events; the color-scale denotes the distance to station LERA1. In all panels, vertical bars mark the issued slope failure alarms.

two types of false classification: (1) noise classified as earthquakes and (2) earthquakes classified as slope failures. Following the steps described in chapter 4, I first split the event catalog without the slope failure event in training and test data sets, with a data volume ratio of 7:3. Then, I convert the event start and end time catalog into a set of 40 s windows containing the event signals, with a 15 s overlap. I compute the features of these windowed data for each station. I then add the training data from the new event catalog to the training data used in chapter 4 and retrain the model using a conventional random forest classifier. The hyperparameters of the classifier are tuned to the new training data set. Figure 6.4a shows the receiver operating characteristic (ROC) curves of the now complete training data set obtained by 3-fold cross-validation. The AUC value for the slope failure class is the lowest (0.97) compared to the noise and earthquake class (0.99).

Figure 6.4b shows the normalized confusion matrix on a test data set consisting of test data from chapter 4. The accuracy of the noise class stays constant at 99%. In contrast, the slope failure and earthquake classes accuracy increased by 4% and 2%, respectively. Overall, the confusion between slope failures and earthquakes decreases by 4%. Figure 6.4c shows the eight most discriminating features used by the new model. Compared to the first training cycle, the most important features did not change. However, there is a slight change in relative importance, where the waveform features importance (energy in the frequency band 1-3 Hz, labeled "ES[0]" in Fig. 6.4c) increases compared to the importance of the frequency at spectrum centroid (labeled "FCentroid").

6.1.3 *Discussion*

Figure 6.2 shows a significant variation in the number of detected events during the summer period compared to winter. The spectrogram of February 2020 to August 2021

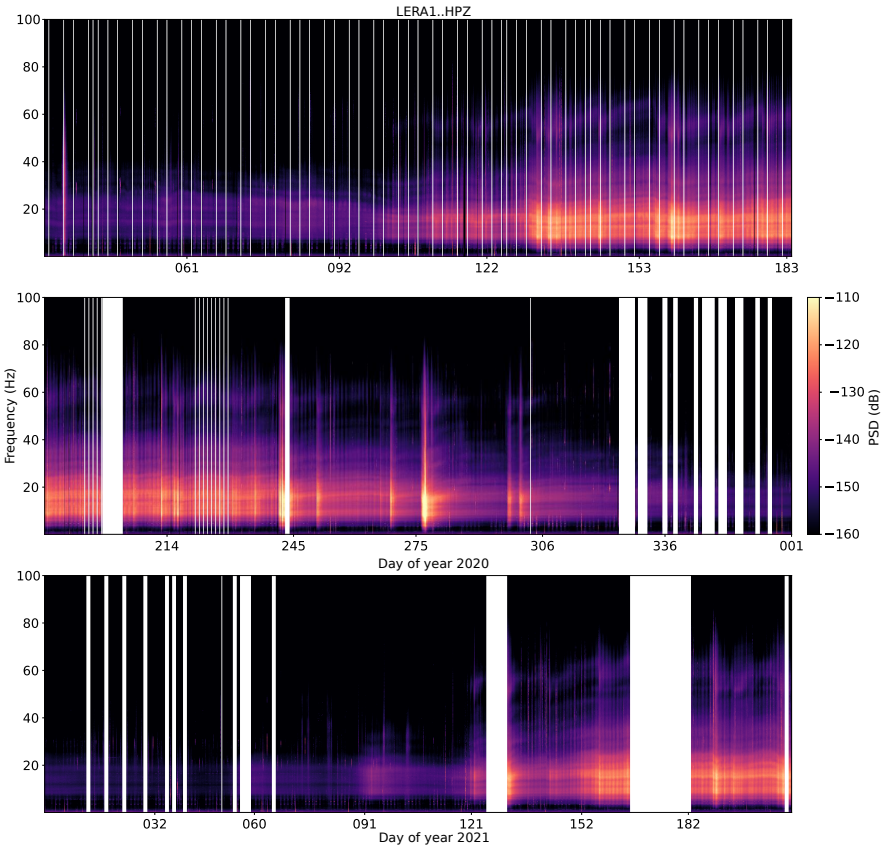


FIGURE 6.3: Long-term spectrogram over the recording period of 2020 and first half of 2021 at station LERA1. Time periods without recording are kept in white.

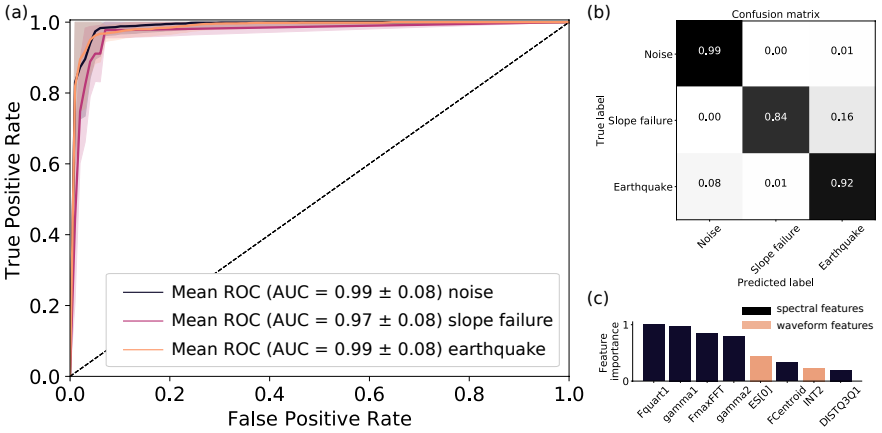


FIGURE 6.4: (a) ROC curve for all three classes using 3-fold cross-validation; noise (dark), slope failure (pink), earthquake (orange). The different colors represent different classes. The shaded area shows the 95 % confidence interval. The area under the curve (AUC) values are appended in the legend. (b) Normalized confusion matrix of the test data set. (c) Feature importance of the eight most distinctive features. Spectral features are denoted in a dark color; waveform features in skin color.

(Fig.6.3) illustrates that during the summer months, noise signals increase drastically, especially in the frequency band of interest (5 - 25 Hz). This is a known phenomenon for a near stream seismometer array, for which seismic noise levels grow with increased runoff (Burtin et al., 2008). The potential slope failure events were all recorded in the winter months, often several events per day. A reason for this might be snow avalanche activity classified as slope failures.

The ROC curves and AUC values for the new data only slightly changed compared to chapter 4. However, Figure 6.4b shows a slightly decreased confusion between seismic noise and earthquakes and, more importantly, between slope failures and earthquakes. Considering that the model was retrained on events that were confused exactly between these classes, this result is not surprising. However, the remaining confusion illustrates that some signals might have too many similarities to be distinguished with these 58 features. Potentially, unsupervised machine learning techniques could find a more discriminating feature set and thus enhance classification performance. However, the small number of slope-failure events make it challenging to train an unsupervised machine learning algorithm, as they generally require an abundance of data per class.

As discussed in chapter 4 more slope failure training events should increase the classifier's performance. However, the study site seems to have been more or less inactive since implementing the automated detection. Therefore, the issued alarm for the single observed slope failure event can already be considered a success. However, to minimize the number of misclassifications, I propose to include a different seismic station (e.g., FIESA) in the decision-making process (for details, see chapter 4). A seismic station at this distance ($>$ 100 km) will most likely not record the slope failure events of interest, but earthquakes that are confused as slope failure will be equally visible at a seismic

station at this distance. Hence, when an earthquake is detected on a reference station, the slope failure alarm issued by the algorithm at FIESA can be disregarded.

6.2 ILLGRABEN EVENT DETECTION AND LONG-TERM MONITORING

Since 2017, a seasonal array of eight seismic stations has been installed at Illgraben during the summer. The array has served two main purposes: (1) detection and location of debris flows (Walter et al., 2017; Chmiel et al., 2021); and (2) debris-flow parameter estimation (chapter 2,3; Zhang et al. (2021)). Three of the seismometers additionally have an ideal location for slope failure detection at the head of the channel (see Fig. 4.1a; Station ILL06, ILL07, ILL08). In this area, rockfall events are frequent but, as a result of the strongly fractured Illhorn north face, rarely with single large rocks impacting several times. Indeed, first-hand observations suggest that often, single rock movements cause a chain reaction, in which loose material and rocks of all sizes trigger slope failure and coalesce into a granular flow. In the following, I train a classifier to detect these slope failures based on data recorded in 2017 and apply it to data from the deployment season in 2019. I then discuss the relation between the number of detected windows containing slope failure signals with meteorological data, namely, precipitation, temperature, and atmospheric pressure.

6.2.1 *Long-term monitoring*

In chapter 4, I introduced the Illgraben slope failure data set, which consists of labeled seismic signals recorded in 2017. I use this data set to train a classifier and detect slope failures over the entire debris-flow season in 2019. Instead of keeping three different classes as in chapter 4, I combine the noise and earthquake class to only detect events related to slope failures. Hence, I end up with a two-class problem: (1) noise (including earthquakes) and (2) slope failures. This results in a training data set with 246 events in the noise class (comprising earthquakes, discrete noise signals, and background noise), and 27 events in the slope failure class. The test data set, consisting of 1/3 of the entire event catalog, contains 101 noise events and 17 slope failure events. Similar to chapter 4, I compute features for stations ILL06, ILL07, and ILL08 separately. I train a balanced random forest classifier and obtain a classification accuracy of 91% for the noise class, with 9% being classified as slope failure. The accuracy of the slope failure class is at 94%, with 6% of the events being misclassified as noise. I then apply the model to seismic data of stations ILL16, ILL17 and ILL18, acquired in 2019. I first use the model to classify the time windows separately for each station and then perform a majority vote over all stations. The classification algorithm provides windows that were classified as containing either a noise or a slope failure signal. The total number of detected events cannot be defined as it is not clear whether, for example, one event is detected in two or more time windows. The feature computation process set the number of windows per day to 8636 (40 s windows overlapping by 15 s). Hence, the classifier gives information about when windows with slope failure signals were detected and what amount of a day was dominated by slope failure signals. In the following, I address the windows classified as containing slope failure signals as slope failure events. Overall, the background noise level of the seismometers at Illgraben is significantly lower than the LERA array

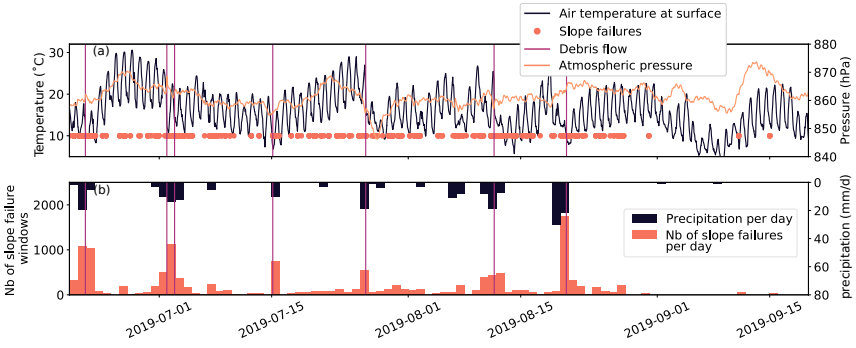


FIGURE 6.5: (a) Three months of air temperature (violet) and atmospheric pressure (orange) recorded at MeteoSwiss station Montana, about 11 km away from Illgraben. Pink dots denote slope-failure windows; pink lines denote debris flows. (b) daily precipitation (violet); total number of windowed seismic data classified as slope failure grouped per day (pink).

(difference $>$ 30 dB). Throughout the deployment season, the background noise does not change. However, as for all seismometers, the daily cycle of elevated anthropogenic noise during the day can be observed.

Figure 6.5a shows the comparison between slope failure detection and temperature, precipitation, and pressure data recorded at a MeteoSwiss station in Montana (south-facing side of the valley, 1427 m a.s.l, 11 km distance to Illgraben). Figure 6.5b shows the daily sum of precipitation and slope-failure windows. The plots show that the debris-flows are linked to high precipitation and a drop in temperature/pressure. At the same time, these parameters strongly correlate with the number of detected slope failures. All periods of high precipitation do show an increase in detected slope-failure windows. In contrast, high temperatures only rarely coincide with a higher or lower number of slope failures. For example, a long period without precipitation and maximum temperatures of up to 30 degrees was observed in June. However, only a small number of slope failures were detected during this time, and those were only on the warmest day. Similarly, during a dry and cold period in the beginning of September, almost no slope failures were detected. Just before the debris flow on July 26, a continuous increase in temperature and atmospheric pressure over several days was observed. During this time, also the number of slope failures was slightly increasing.

Figure 6.6 shows a zoom-in of four example debris-flow events in 2019. For all events, just before the debris-flow signal, several slope-failure windows were detected. The event with the least post-event precipitation (July 26) showed no slope failures after the debris-flow event. On July 2 and July 26, the detected slope failures before the debris-flow event coincide with high energy signals over the full frequency range (1-50 Hz). Slope failure signals, however, show a clear dominant frequency between 5-10 Hz. In contrast, these signals coincide with lightning events from the approaching thunderstorm. The move-out of these events (i.e., arrival time differences at all seismic stations of the array) shows wave velocities of about 330 m/s, indicating thunder as the main source (Marchetti et al., 2019).

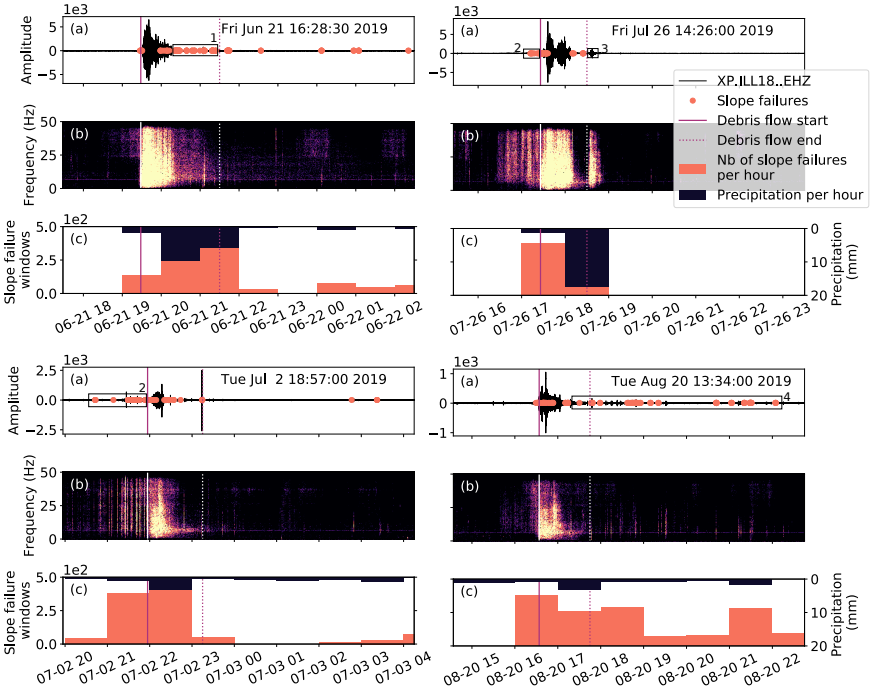


FIGURE 6.6: (a) Waveform and slope-failure windows (pink dots), (b) spectrogram and (c) hourly precipitation (violet); total number of windowed seismic data classified as slope failure grouped per hour (pink) of four debris flows that happened in 2019. Pink (white) vertical lines denote the start (solid line) and end (dotted line) of the debris-flow event. Black boxes in (a) and associated numbers mark periods of interest (see text).

Figure 6.7a shows a timeline of slope-failure windows and 6-hour precipitation sums of the full deployment period. Time intervals including debris flows are excluded to concentrate on the general relation between precipitation and slope failures. The relation between detected slope failure windows and precipitation is not immediately evident. The same amount of precipitation does not cause the same amount of detected slope failures throughout the season. Additionally, the response of slope failures to precipitation is present for many high precipitation events; however, the phase shift between precipitation and slope failures and relative amplitudes differ for each event. Especially in the first half of the season, high precipitation events are often followed by or coincide with an increased number of detected slope-failure windows. In contrast, on July 15, a peak in slope failures was observed without preceding precipitation. Several days later, a precipitation event did not cause any increase in slope failure windows. Towards the end of summer, detected slope-failure windows decline drastically, while high precipitation events still occur. While precipitation events of this intensity caused high numbers of detected slope failures at the beginning of the season, they now only reach numbers half or less than initially. The amount of precipitation and number of classified slope-failure windows per time of the day over the entire deployment season is shown in Figure 6.7b,c. I show two different cases: the integral over the entire season, transparent bars) and the integral over the entire season but excluding the debris-flow events (solid bars). I defined the excluded debris-flow periods as three hours before the start until one hour after the end of the debris-flow signal indicated in Figure 6.6. Generally, precipitation and slope failures happen at the same time of the day. The relative difference between the precipitation intensity and slope failure frequency with debris flows and without debris flows in precipitation and detected slope failures varies strongly. The period between 16:00 and 17:00 UTC poses the clearest outlier. While almost no precipitation is observed throughout the whole season during that time, the number of detected slope failures is high, including and excluding the debris-flow periods. Similarly, during 11:00 and 12:00 UTC, many slope-failure windows are detected, whereas little precipitation is observed. Between 03:00 and 04:00 UTC, almost no difference between the two precipitation integrals is observed. However, during the period including the debris flows, almost 50% more slope-failure windows are detected than without the debris flows.

6.2.2 *Slope failure signals*

Seismic signals of slope failures can be quite diverse, depending on the latter's location, dynamics, volume, and material properties. Provost et al. (2018) proposed a typology for seismogenic slope failures to standardize the classification of these events and enhance comparability. I manually identified slope failure seismograms, which were compiled in the 2017 Illgraben event catalog (see chapter 4). A selection of events is shown in Figure 6.8. Following Provost et al. (2018), the main types of events encountered at Illgraben are granular flows (Fig. 6.8a,b) and slopequakes (Fig. 6.8c,d). Accordingly, granular flows are characterized by their duration of several tens of seconds and their emergent onsets. In contrast, the slopequakes, likely caused by shearing and fracturing processes, only last for several seconds and have, except for the tremor-like class, a rather clear signal onset (Provost et al., 2018). Similarly, the tremor-like class is characterized by a short event duration, with the emergent onset resembling the granular-flow class. The

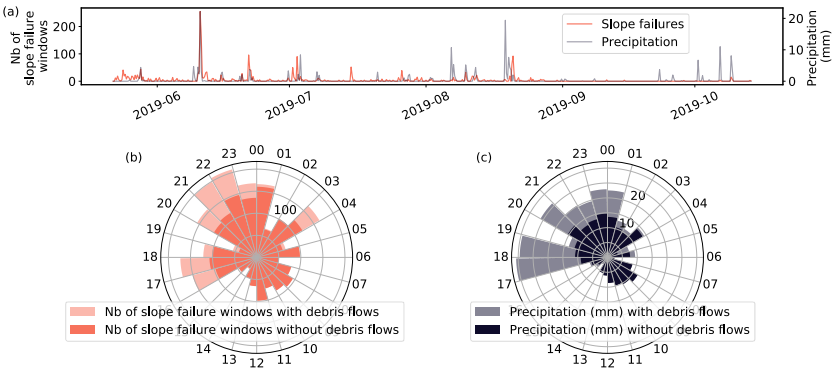


FIGURE 6.7: (a) Timeline of slope-failure windows (orange) and precipitation per hour (grey). Periods of debris-flow events are excluded. (b) Precipitation (mm) and (c) slope-failure windows detected over the whole deployment period integrated per hour of the day (UTC). The transparent bars show the total amount, whereas the solid bars exclude periods of debris flows.

seismic signals found whilst assembling the event catalog are characterized by a signal duration of several tens of seconds, an emergent onset, and relatively low signal-to-noise ratios (see chapter 4), indicating predominantly granular flow behavior at Illgraben. This coincides with my observations during fieldwork, where granular flows could be seen and heard for several minutes at a time. The slopequakes could indicate tensile cracks from extension and contraction of the sub-surface due to thermal variation that can induce slope failures (e.g., Gruber et al., 2004) and/or sliding processes along a shear fault parallel to the surface (e.g., Martel, 2004).

6.2.3 *Discussion*

Figure 6.5 shows a strong correlation between the number of detected slope failures and extreme precipitation events during which debris flows are released. However, the number of slope-failure windows is strongly influenced by two components: (1) Misclassification of thunder as slope failure and (2) debris-flow signals classified as slope failures. Figure 6.6 shows examples of four debris flows in 2019, the slope failure detection, and the precipitation measurements. Slope failure signals at Illgraben generally show similar frequency characteristics as debris flows, with the dominant frequencies being between 5-10 Hz. Debris flow signals are not included in the training process of the classifier because they are not the process of interest. Larger debris flows show very high power spectral densities, which the classifier associates with noise or earthquake signals; these periods are thus often classified as noise. However, for small debris flow events or while the debris flow propagates away from the seismic array, the power spectral density is significantly lower, and time windows are classified as slope failures (e.g., June 21 event, black rectangle (1), Fig. 6.6). Clearly, it is hard to say if the debris flow generates the signals or if a nearby slope failure exceeds the debris-flow signal. Chmiel

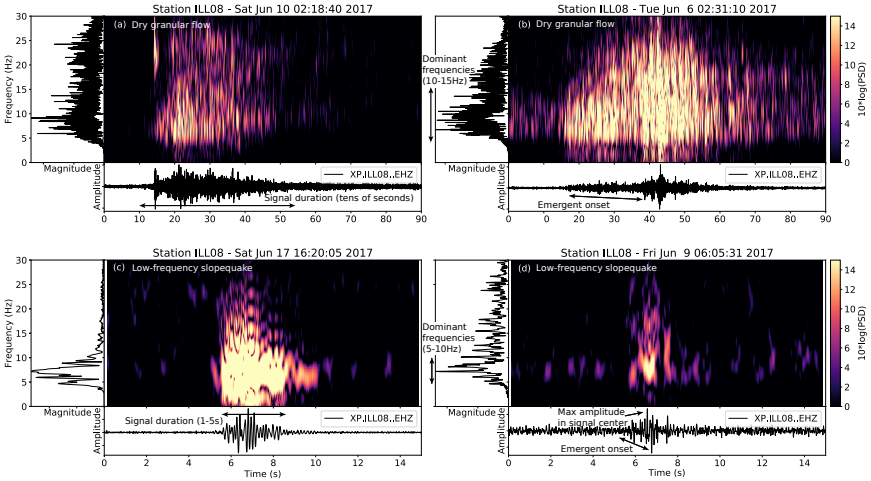


FIGURE 6.8: Waveform variety of slope failure signals recorded at station ILL08 at Illgraben. Signals were bandpass filtered between 1 and 30 Hz. (a) and (b) are tremor like or granular flow signal with a duration of several tens of seconds. (c) and (d) are examples of low-frequency slopequake signals. The classification follows (Provost et al., 2018).

et al. (2021) show that these early debris-flow detections can be distinguished from the later debris-flow signal; however, signal sources were not identified. Another potential misclassification source is the signal from thunders during arriving storms and the elastic signals generated from it in the ground. In the spectrograms, the thunder signals manifest themselves as high power bursts over the entire measured frequency range (e.g., July 2 and July 26, black rectangle (2), Fig. 6.6). Interestingly, not every thunder signal is classified as slope failure, and it remains unclear whether this is due to different intensities of the thunder events or whether, in some cases, the classifier detects actual slope failures that are masked by the thunder. As the training data set consists of several thunder signals, confusion between the signals suggests that further investigation is needed to discern between the two possible cases. So far, the high noise level caused by thunder and debris flows prevents a quantification of the number of slope failures that happen before and during the debris-flow events. What can be said is that no thunder occurred before the debris flow for the June 21, July 15, and August 20 events. For these events, the classifier detects up to 10 40s-long time windows with slope failures during the debris flow initiation. This is likely due to an increase in pore water pressure in response to extreme precipitation, which in turn might mobilize shallow landslides and materials deposited in both the hillslope gullies and the main channel (e.g., Iverson, 1997). However, also in this case, it is difficult to decide if the signals are generated by the individual slope failures or by the debris flow itself. In the five hours after the debris flow, slope failures are only observed if precipitation is still occurring (e.g., August 20 event, black rectangle (4), Fig. 6.6). This limits the influence of slope stabilization by debris-flow passages (Burtin et al., 2014) and indicates that rainfall triggers immediate slope failures,

rather than causing slope instabilities long after the rainfall event. This aligns with direct field observations made on July 26, 2018, i.e., the day after a debris flow on July 25, 2018. In that case, I could not see or hear a single slope failure at Illhorn's north face during the entire day. On July 26, another signal similar to a debris flow is recorded on station ILL18 (black rectangle (3), Fig. 6.6). This period was not classified as slope failure, as the signal is not visible on stations ILL16 and ILL17, whilst the design of the classifier requests a detection for at least two stations. The event, therefore, must have occurred close to station ILL18 and at a larger distance to ILL16 and ILL17. Potentially, the source location is an active gully next to ILL18, or the recently active Sagenschleif, an unstable gully about 2.5 km downstream from the channel head.

Strunden et al. (2015) find that for the Lauterbrunnen valley in the Swiss Alps, rockfalls happen predominantly in the winter months. They explain the pattern in rockfall activity with the freeze-thaw cycle, where water in cracks freezes and increases cracking, eventually leading to break-off from the rock wall. Strunden et al. (2015) find a two-month delay between the decreasing temperatures and the increasing rockfall occurrence and explain the difference with the delay by which heat penetrates the rock wall. Savi et al. (2021) find that in the eastern Italian Alps, rapid snowmelt and a temperature-triggered spread of the active permafrost layer cause an increase in rock-slope instabilities in early spring. Bennett et al. (2013) show that for Illgraben, sediment erosion follows similar patterns. Although only the top of Illhorn can be expected to be affected by permafrost (Kenner et al., 2019), the exposure of the slope to temperature variations and freezing temperatures has been suggested to be a dominant factor for the local erosion (Bennett et al., 2013). Bennett et al. (2013) estimated erosion rates based on a series of DEMs separated by $\sim$5-20 years. For these time periods, they did not find correlations between erosion and seismic activity or precipitation. However, they find that erosion likely increased after 1986 due to a decline in snow cover and the bedrock being increasingly exposed to atmospheric temperature fluctuations. A key limitation of their approach is that the exact timing of erosion events is unknown and therefore also the ultimate trigger. While the seismic data analysis does not give quantitative erosion measurements and while the seasonal installation of our seismic network at Illgraben (generally between May and October) does not allow assessment of erosion associated with slope failure over the entire year, patterns of triggers can be analyzed on a sub-minute timescale. For the patterns observed here (Fig. 6.5), during the deployment period of the seismometers, mechanisms related to freezing can be excluded as primary drivers for slope failures because temperatures do not fall below 10 degrees Celsius. However, this measurement is representative for the median basin altitude and therefore cryogenic processes can still occur in the higher parts. Photos from fixed cameras facing the Illhorn north face show, for example, that in 2019 this face became completely snow-free only at the beginning of July. In June, most gullies in the top part of Illhorn still contained snow, and it is known that rapid snowmelt can cause the ground to saturate and pore water pressure to increase, which in turn can trigger shallow slope failures (e.g., Wieczorek, 1996). This agrees with the higher resolution time series (6-hourly sums) of detected slope-failure windows (Fig. 6.7a). The time series in which the debris-flow events were excluded indicates a decrease in slope failures during the season. This suggests a supply limitation towards the end of the season, where material that is loose enough to be mobilized by increasing pore-water pressures is no longer present. The last weeks of May and the weeks at the beginning of June, instead, show the highest number of slope failures while

no precipitation is present. During this time, the snowmelt season at Illgraben is ongoing, again hinting at the fact that water availability is an essential condition for slope failure triggering.

Collins and Stock (2016) use thermodynamic models to show that fracture opening can be caused by high temperatures and temperature variations and suggest that this can cause rock detachment. This can potentially explain events where no other meteorological trigger was found. For the Illgraben case, the number of slope failures slightly increases during a period of continuously increasing temperatures, just before the debris flow on July 26. In contrast, during the hot period at the end of June, slope failures only occurred on the warmest day (June 26) and a small number on the day after. The relation between high temperatures and slope failures at Illgraben is thus not particularly obvious, and often, it is impossible to determine one single trigger mechanism. Although some indications exist, I conclude that the data at hand are insufficient to establish a strong causal relationship between the warm temperatures and slope failures during the summer period.

Figure 6.5 underlines the relation between slope failures and precipitation, whereas temperature and atmospheric pressure do not have a straightforward influence. Precipitation is a long-established trigger for slope failures (Wieczorek, 1996). Like snowmelt, high intensity and/or long-lasting precipitation events increase pore water pressure in sediments and soils, which can cause unstable slopes to fail eventually. This mechanism is especially effective on weak slopes with loose rocks (Wieczorek, 1996). A relation between high precipitation and the number of detected slope failures also becomes apparent in the higher resolution time series, where precipitation and slope failure events during debris flows were removed from the data (Fig. 6.7). Especially at the beginning of summer, strong precipitation events cause an increasing number of detected slope-failure windows. On June 10, 2019, 20 mm of precipitation during 6 hours triggered a considerable number of slope failures but did not lead to a debris flow. Still in the snowmelt season, increased pore-water pressure caused by already small precipitation events might exceed static equilibrium pressures of the already-saturated soils and sediments and trigger slope failure events (Iverson, 1997). On August 6, a slightly smaller but still relatively high precipitation event (12 mm during 6 hours) only triggered a marginal increase in slope failures compared to the days before. Similarly, on August 19, a high precipitation event did not trigger a debris flow, but instead another precipitation event on August 20 (see Fig. 6.5; debris-flow event not included in time series of Fig. 6.7.). The tendency for high precipitation events to trigger fewer slope failures towards the end of the summer season (August-October) is even more apparent than during spring. This is especially the case for September and October, where very few slope-failure windows are detected. This indicates that, indeed, a supply limit evolves throughout the season.

Figure 6.7 also shows slope-failure events during times where neither snowmelt nor precipitation is present. The most prominent peak is in the night between July 14, 2019, and July 15, 2019, just before the debris-flow event on July 15. Indeed, during this time, a storm passed several tens of kilometers north of Illgraben, with recordings of lighting (https://kachelmannwetter.com/ch/). Potentially, windows classified as slope failures were confused with signals from the passing thunderstorm. Throughout the whole period, many days exist on which no precipitation was recorded, but slope failures were detected. Compared to days with rainfall, the number of detected windows is small, except at the beginning of the season. As mentioned above, I explain this with

the ongoing snowmelt. Later in the season, no clear trigger mechanism can be identified. Hence, if the detected slope failure windows were not misclassified, and the precipitation measurements are precise despite the distance between Illgraben, and the MeteoSwiss station, another trigger mechanism such as thermal variations or seismic triggering through earthquakes might play a role and should be investigated in the future.

Based on the above observations, I propose the following main triggering mechanism at Illgraben: During winter, freezing temperatures enhance cracking, promote slope instabilities, and supply loose sediment to the many gullies on Illhorn's north face. At the beginning of the melt season, increasing pore water pressure due to snowmelt and precipitation mobilizes the so-deposited materials and evacuates them towards the main Illgraben channel. The materials, however, are not evacuated during one single event but rather during the course of the season. Towards the end of the season, less and less loose materials are available, causing intense precipitation events to trigger fewer slope failures than at the beginning of the season. frost-weathering is eventually needed again in order to re-supply the slopes with loose material for the next season.

In general, slope failure triggers such as frost-weathering, freeze-thaw cycles, snowmelt, precipitation, and thermal fracturing, are well-accepted concepts. However, many individual events remain without an obvious trigger mechanism (Collins & Stock, 2016). One reason for such poorly constrained triggers is the lack of precise and long-term observations. Continuous seismic records could fill this gap; automatic event detection is crucial to handle the resulting large amount of data. Here, I showed that a classifier trained and tested on three weeks of seismic data recorded in 2017 can be used to obtain an estimate of slope failure events in 2019. Model testing revealed good classifier performance, with a weighted accuracy of 92.5%. However, I cannot assess the classifier's performance separately for 2019, as no labeled data exist for that year. Additionally, uncertainties are introduced by the training data set, which may be inaccurate to an unknown degree: The data set was assembled to reflect my best knowledge, comparing the seismic records to many published events and cross-checking with colleagues. For this data set, no discrepancy between events labeled by peers and by myself occurred. Still, misjudgment in some cases is a possibility. The manual distinction between slope failures and earthquakes is not of concern, as dense earthquake catalogs are provided by the Swiss Seismological Service and EMSC, for instance. In contrast, the confusion of slope failure signals and noise is problematic. For example, thunders during the intense precipitation events triggering the debris flows generate seismic signals whose waveforms are similar to signals of granular flows. Generally, they can be distinguished from slope failures due to stronger energy in the higher frequencies, but still, individual misinterpretations might occur. In fact, the MeteoSwiss station at Montana also records lightning. In further work, the information content thereof could be used to exclude misclassification of signals originating from thunder. Additional noise sources at Illgraben include helicopters as well as private and military airplanes. While helicopters have a distinct frequency signature (see chapter 4), seismic signals of airplanes can be similar to slope failures (Heck et al., 2018). Despite these uncertainties, I found a multilayered relation between slope failure events and precipitation at Illgraben. Other meteorological triggers could not be confirmed but neither excluded.

For further examination, I propose extending the event catalog by including slopequakes as introduced above. Provost et al. (2018) show that slope quakes are often subject to repeating events with similar waveforms. Therefore, a simple cross-correlation search

could be valuable for finding slopequakes and investigating the temporal and spatial pattern of fracturing and shearing processes. Furthermore, their behavior could help gain a clear picture of the process cascade at Illgraben, from fracturing to slope failure to debris flows.

6.3 IMPLICATIONS AND OUTLOOK

6.3.1 *Implications for automatic signal detection*

A number of studies have shown that machine learning can successfully classify seismic signals of different origin (e.g., Hammer et al., 2013; Dammeier et al., 2016; Hibert et al., 2017; Provost et al., 2017; Heck et al., 2018). However, real-life applications for operational monitoring systems often pose additional requirements. In this thesis, I identified difficulties associated with external circumstances, such as network set-up and a limited number of recorded events (chapter 4). Notwithstanding these difficulties, I show that satisfying classification accuracy can be achieved with adequate consideration of the data set. Still, with this setup, the number of misclassifications cannot be reduced further without taking more seismic stations into account or without requiring manual intervention by an operator. A way to further improve the classification performance might be to take network features into account. In other words: it is critical to design the installed seismic network according to the methods intended to be used.

One of the largest challenges in machine-learning-based detection of mass movement within seismograms is the unbalanced data set, with on one hand only a few events of interest and on the other hand long time series of noise and other signals (see chapter 4). Potentially, slope failure events from other sites and similar network installations could be used to increase the number of training samples. To my knowledge, apart from the short discussion in chapter 4, to date no study explores if or how classifiers trained on one site can be applied at other sites and for different network set-ups. It, therefore, remains to be seen if a classifier's performance could be improved with enough data from various sites and different slope failure events. As a first step, Bahavar et al. (2019) started an "Exotic Seismic Events Catalog" for non-earthquake seismic events, openly available to the community. I envision a community effort to extend this event catalog and gather enough data to investigate the performance of machine-learning classifiers trained on seismic signals from various study sites and the classifier's performance on an unseen site. A robust performance would open new ways for mass movement monitoring. For example, with a classifier trained at another site, signal classifications would be available as soon as an instrument is installed, and recently active slopes could be monitored instantaneously without needing prior data.

A key component of the classifier is the machine learning algorithm itself. Machine learning or artificial intelligence is a dynamic field with continuously evolving methodologies. Hence, a continuous review of recently developed and potentially better-fitting algorithms should be carried out. Recently, Meyer et al. (2021) used contrastive learning, a self-supervised algorithm that learns similarities between two input data sets before even seeing the sample label. Theoretically, this increases model robustness for inaccurate and incomplete labeled data. Meyer et al. (2021) showed that applied on the Illgraben data set, the classification performance of contrastive learning is superior to the supervised random forest approach taken in this thesis. Of course, higher performance is desirable,

but I want to emphasize the simplicity of the machine learning approach, and the ease of implementation should also be considered. Training the model of Meyer et al. (2021) took several days, which requires significantly more computational power compared to my random forest classifier trained within less than one minute. This training time difference can be crucial for situations in which the model needs to be retrained regularly because of changing noise patterns.

Another way to improve classification performance could be to improve seismic data quality, i.e., by reducing the noise level. A lower noise level could potentially increase the number of detected events otherwise hidden in the noise. Eventually, a consequently larger training data set could lead to an enhanced classifier performance. In the last few years, noise reduction using artificial intelligence has developed into a widely studied field (e.g., Chen et al., 2019; Seydoux et al., 2020). For example, van den Ende et al. (2021) used a self-supervised deep learning approach to denoise data from a distributed acoustic sensing (DAS) experiment. Recently, DAS has become a widely used method to record seismic waves (e.g., Zhan, 2019). In DAS, back-scattered signals from laser pulses injected into a fiber-optic cable record strain rates in discrete steps along the cable several meters apart. This technology, therefore, outputs time series of strain rates every few meters, turning kilometer-long fiber optic cables into thousands of seismic sensors (Walter et al., 2020a). DAS thus delivers a large amount of data, usually with lower signal-to-noise ratios than conventional seismometer installations (e.g., Correa et al., 2018; Jousset et al., 2018) but with unmatched spatial coverage. Especially in places that are hard to access, such as high-alpine terrain, DAS opens new dimensions to near-real-time monitoring of mass movements with its ease of installation: it allows for acquiring data from a great number of sensors by simply lying out a cable, or even by using pre-installed fiberoptics communication networks.

6.3.2 *Implications for debris-flow monitoring*

The three core chapters (2, 3, 4) of this thesis have implications for state-of-the-art debris-flow monitoring. Based on the classification method developed in chapter 4, Chmiel et al. (2021) trained a model that successfully detects debris flows with the seismic network at Illgraben. Not only does it detect the event, but warning time was increased by up to several hours compared to in-torrent detectors at low elevation. The question remains on how classification techniques transfers to other study sites with typically less recorded events and/or smaller seismic networks.

I also showed that for the high-frequency component of the seismic signal, the model introduced by Tsai et al. (2012) and adapted to debris flows by Lai et al. (2018) can give realistic particle size estimates for Illgraben. Since then, models on the high-frequency signal of debris flows have been developed further (Farin et al., 2019; Allstadt et al., 2020; Zhang et al., 2021) and have been used to constrain the relation between seismically derived basal force fluctuations and direct force plate measurements. However, despite these advances, the link between bulk debris-flow parameters and the high-frequency content of emitted seismic waves is still challenging to quantify (Allstadt et al., 2020).

As an alternative to high-frequency seismic signals, I used tilt signals obtained from the low-frequency component of the seismic signal to obtain debris-flow parameters such as the maximum flow height. One drawback of using such low-frequency signals in real-time monitoring is the delay that comes with it: low-frequency signals require

lowpass filtering to suppress the dominant high-frequency signal content. For short time windows, that are required for real-time monitoring, the filtering can cause significant artifacts that potentially mask the tilt signal. To avoid these artefacts in real-time observation, a delay of several minutes needs to be included in the warning time (see chapter 3) in order to reliably determine the low-frequency signal content. An advantage compared to the high-frequency signal, which is influenced by many different factors (e.g., particle size, topography, flow velocity) that have to be simplified, is that the tilt is fully represents the weight and the spatial extent of the nearby debris-flow front. Therefore, tilt measurements provide a more direct characterization of the bulk debris-flow parameters. Thus, broadband seismometers can potentially be used as an alternative to more sensitive but significantly more expensive and harder to install force plates. A combined model of low- and high-frequency signal components may disclose a complete picture of debris-flow dynamics and bulk properties in future studies and applications.

7

CONCLUSION

In this thesis, I studied the use of seismological measurements in the context of monitoring and early warning of gravitational hazards. I mainly concentrated on the detection of seismic mass-movement signals and the relation between signal content and mass-movement parameters like particle size and flow height. I found that even though only limited seismic data of mass movements may be available in many cases, fine-tuning of machine-learning algorithms can result in satisfactory accuracy for automatic signal classification. Moreover, attentive pre-installation considerations and ancillary seismic reference stations further away from the area of interest can significantly improve classification results. The analysis of debris-flow seismic signals confirms the theoretical prediction that particle sizes influence seismogenisis, which should be leveraged in monitoring and early warning applications. Lastly, I show that in contrast to the high-frequency seismic signal, the low-frequency seismic signal of debris flows, i.e., the tilting due to the weight of the passing debris-flow's mass, gives reliable estimates on bulk parameters such as debris-flow height and vertical forces. Hence, by combining seismic studies of both high-frequency and low-frequency signal components, the full picture from small scale processes to large scale characteristics of debris flows could be revealed.

A

APPENDIX

A.1 SUPPLEMENTARY MATERIAL TO CHAPTER 2 - SIGNALS AT OTHER SEISMIC STATIONS

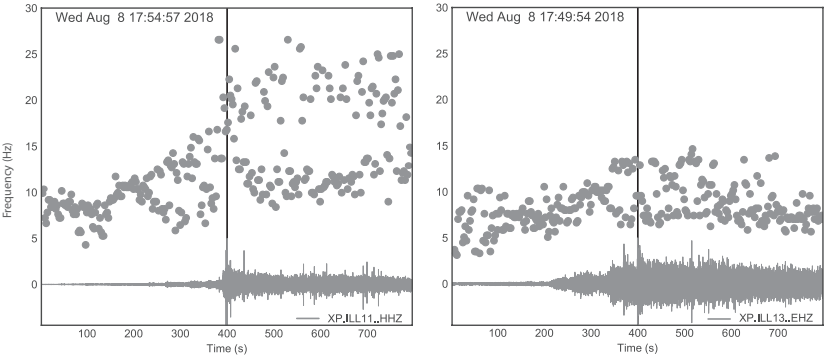


FIGURE A.1: Seismic signal and corresponding dominant frequencies of the August 8, 2018 event recorded at stations ILL11 and ILL13.

A.2 SUPPLEMENTARY MATERIAL TO CHAPTER 4

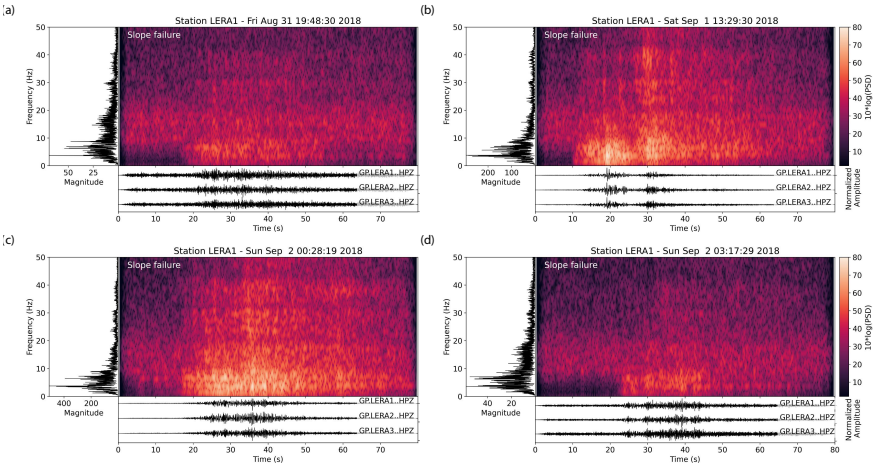


FIGURE A.2: Seismic signals, spectrograms and spectra of four additional slope failure events in 2018 used for training. Spectrograms have been computed with a window length of 128 samples, an overlap of 120 samples and an FFT length of 2048.

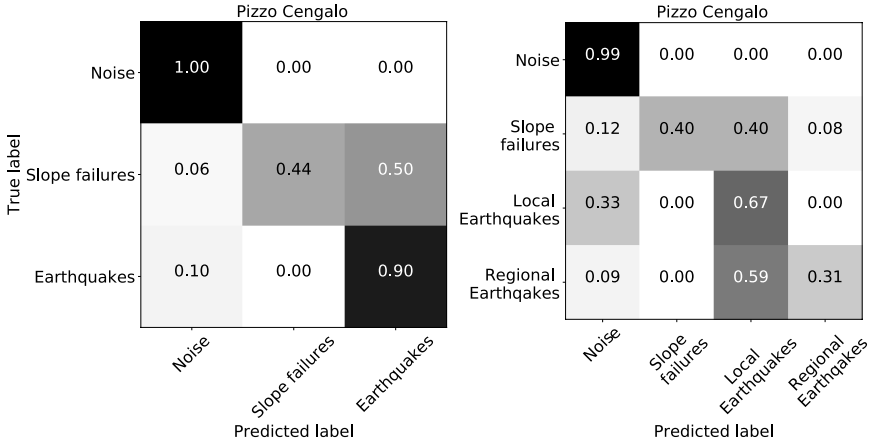


FIGURE A.3: Confusion matrix of initial test of random forest on the Pizzo Cengalo data set. Left: all earthquakes lumped into one class. Right: earthquakes divided in local and regional/teleseismic earthquakes.

Table A.1: Table with all features, slightly adjusted from Provost et al. (2017)

	Waveform Features:	
1	Ratio of the mean over the maximum of the envelope	–
2	Ratio of the median over the maximum of the envelope	–
3	Kurtosis of the raw signal (peakness of the signal)	$\frac{m_4}{\sigma^4}$, with m_4: fourth moment, σ: standard deviation
4	Kurtosis of the envelope	see 3
5	Skewness of the raw signal	$\frac{m_3}{\sigma^3}$, with m_3: third moment
6	Skewness of the envelope	see 5
7	Number of peaks in the autocorrelation function	–
8	Energy in the first third part of the autocorrelation function	$\int^{T/3} \mathrm{C}(\tau)d\tau$, with T: signal duration, C: autocorrelation function
9	Energy in the remaining part of the autocorrelation function	see 8
10	Ratio of 8 and 9	–
11 – 15	Energy of the signal filtered in 1 — 3 Hz, 3 — 6 Hz, 5 — 7 Hz, 6 — 9 Hz and 8 — 10 Hz	$\int_0^T y_f(t)dt$, with y_f: filtered signal in the frequency range [f1-f2]
16 – 20	Kurtosis of the signal in 1 — 3 Hz, 3 — 6 Hz, 5 — 7 Hz, 6 — 9 Hz and 8 — 10 Hz frequency range	see 3
21	Maximum of the envelope	–

	Spectral Features:	
22	Mean of the DFT	DFT: discrete Fourier transform
23	Max of the DFT	–
24	Frequency at the maximum	–
25	Frequency of spectrum centroid	–
26	Central frequency of the 1st quartile	–
27	Central frequency of the 2nd quartile	–
28	Median of the normalized DFT	–
29	Variance of the normalized DFT	–
30	Number of peaks (> 0.75 DFTmax)	$\mathrm{DFT}_{\max}$: maximum of the DFT
31	Mean value for the peaks	–
32 -- 35	Energy in [0,$\frac{1}{4}$]Nyf, [$\frac{1}{4}$,$\frac{1}{2}$]Nyf, [$\frac{1}{2}$,$\frac{3}{4}$]Nyf, [$\frac{3}{4}$,1]Nyf	$\int_{f_1}^{f_2} \mathrm{DFT}(f)df$ with f_1, f_2: the considered frequency range
36	Spectral centroid	$\gamma_1 = \frac{m_2}{m_1}$, with m_1 and m_2 the first and second moment
37	Gyration radius	$\gamma_3 = \sqrt{\frac{m_3}{m_2}}$, with m_3 the third moment
38	Spectral centroid width	$\sqrt{\gamma_1^2 - \gamma_2^2}$

Spectrogram Features:

39	Kurtosis of the maximum of all discrete Fourier transforms (DFTs) Kurtosis as a function of time t	Kurtosis $\left[\max_{t=0,..,T}(\mathrm{SPEC}(t,f))\right]$ with SPEC(t,f): spectrogram
40	Kurtosis of the maximum of all DFTs as a function of time t	see 39
41	Mean ratio between the maximum and the mean of all DFTs	mean $\left(\frac{\max(\mathrm{SPEC})}{\mathrm{mean}(\mathrm{SPEC})}\right)$
42	Mean ratio between the maximum and the median of all DFTs	see 41
43	Number of peaks in the curve showing the temporal evolution of the DFTs maximum	–
44	Number of peaks in the curve showing the temporal evolution of the DFTs mean	–
45	Number of peaks in the curve showing the temporal evolution of the DFTs median	–
46	Ratio between 43 and 44	–
47	Ratio between 43 and 45	–
48	Number of peaks in the curve of the temporal evolution of the DFTs central frequency	–
49	Number of peaks in the curve of the temporal evolution of the DFTs maximum frequency	–
50	Ratio between 48 and 49	–
51	Mean distance between the curves of the temporal evolution of the DFTs maximum frequency and mean frequency	–
52	Mean distance between the curves of the temporal evolution of the DFTs maximum frequency and median frequency	–
53	Mean distance between the 1st quartile and the median of all DFTs as a function of time	–
54	Mean distance between the 3rd quartile and the median of all DFTs as a function of time	–
55	Mean distance between the 3rd quartile and the 1st quartile of all DFTs as a function of time	–

TABLE A.5: Number of events and number of windowed seismic data (20s windows as an example) of the training and test data set of Pizzo Cengalo and Illgraben for classes noise (NO), slope failures (SF) and earthquakes (EQ).

		Events in data set		Number of 20s windows	
		Pizzo Cengalo	Illgraben	Pizzo Cengalo	Illgraben
Training data	NO	1025	222	2095	473
	SF	3	59	76	337
	EQ	17	66	438	674
Test data	NO	438	85	836	160
	SF	2	31	54	170
	EQ	13	34	373	373

TABLE A.6: Random forest parameters

Number of trees	2000
Split quality measure	Gini criterion
Minimum number of samples required to be a leaf node	4
Maximum depth of a tree	60
Minimum number of samples for an internal node to be split	2

BIBLIOGRAPHY

Abellán, A., Vilaplana, J. M., Calvet, J., García-Sellés, D., & Asensio, E. (2011). Rockfall monitoring by Terrestrial Laser Scanning - Case study of the basaltic rock face at Castellfollit de la Roca (Catalonia, Spain). *Natural Hazards and Earth System Science*.

Allen, R. (1982). Automatic phase pickers: Their present use and future prospects. *Bulletin of the Seismological Society of America*, *72*(6B), S225.

Allen, S., & Huggel, C. (2013). Extremely warm temperatures as a potential cause of recent high mountain rockfall. *Global and Planetary Change*.

Allstadt, K. (2013). Extracting source characteristics and dynamics of the August 2010 Mount Meager landslide from broadband seismograms. *Journal of Geophysical Research: Earth Surface*, *118*(3), 1472.

Allstadt, K. E., Farin, M., Iverson, R. M., Obryk, M. K., Kean, J. W., Tsai, V. C., Rapstine, T. D., & Logan, M. (2020). Measuring basal force fluctuations of debris flows using seismic recordings and empirical green's functions. *Journal of Geophysical Research: Earth Surface*, *125*(9), e2020JF005590.

Allstadt, K. E., Farin, M., Lockhart, A. B., McBride, S. K., Kean, J. W., Iverson, R. M., Logan, M., Smith, J. B., Tsai, V. C., & George, D. (2019). Overcoming barriers to progress in seismic monitoring and characterization of debris flows and lahars.

Allstadt, K. E., Matoza, R. S., Lockhart, A. B., Moran, S. C., Caplan-Auerbach, J., Haney, M. M., Thelen, W. A., & Malone, S. D. (2018). Seismic and acoustic signatures of surficial mass movements at volcanoes. *Journal of Volcanology and Geothermal Research*, *364*, 76.

Allstadt, K. (2016). Seismic signature of debris flume experiments.

Anthony, R. E., Aster, R. C., Ryan, S., Rathburn, S., & Baker, M. G. (2018). Measuring Mountain River Discharge Using Seismographs Emplaced Within the Hyporheic Zone. *Journal of Geophysical Research: Earth Surface*, *123*(2), 210.

Aoyama, H., & Oshima, H. (2008). Tilt change recorded by broadband seismometer prior to small phreatic explosion of Meakan-dake volcano, Hokkaido, Japan. *Geophysical Research Letters*, *35*(6), 1.

Arattano, M., & Marchi, L. (2008). Systems and sensors for debris-flow monitoring and warning. *Sensors*, *8*(4), 2436.

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. *Natural Hazards*, *49*(3), 517.

Baer, P., Huggel, C., McArdell, B. W., & Frank, F. (2017). Changing debris flow activity after sudden sediment input: a case study from the Swiss Alps. *Geology Today*.

Bahavar, M., Allstadt, K. E., Van Fossen, M., Malone, S. D., & Trabant, C. (2019). Exotic Seismic Events Catalog (ESEC) Data Product. *Seismological Research Letters*, *90*(3), 1355.

Bennett, G. L., Molnar, P., Eisenbeiss, H., & Mcardell, B. W. (2012). Erosional power in the Swiss Alps: Characterization of slope failure in the Illgraben. *Earth Surface Processes and Landforms*, *37*(15), 1627.

Bennett, G. L., Molnar, P., McArdell, B. W., Schlunegger, F., & Burlando, P. (2013). Patterns and controls of sediment production, transfer and yield in the Illgraben. *Geomorphology*, *188*, 68.

Berger, C., McArdell, B. W., & Schlunegger, F. (2011). Direct measurement of channel erosion by debris flows, Illgraben, Switzerland. *Journal of Geophysical Research: Earth Surface*, *116*(1).

Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y., & Wassermann, J. (2010). ObsPy: A python toolbox for seismology. *Seismological Research Letters*.

Beyreuther, M., & Wassermann, J. (2008). Continuous earthquake detection and classification using discrete Hidden Markov Models. *Geophysical Journal International*, *175*(3), 1055.

Breiman, L. (2001). Random forests. *Machine Learning*.

Burtin, A., Bollinger, L., Vergne, J., Cattin, R., & Nábělek, J. L. (2008). Spectral analysis of seismic noise induced by rivers: A new tool to monitor spatiotemporal changes in stream hydrodynamics. *Journal of Geophysical Research: Solid Earth*, *113*(5).

Burtin, A., Hovius, N., McArdell, B. W., Turowski, J. M., & Vergne, J. (2014). Seismic constraints on dynamic links between geomorphic processes and routing of sediment in a steep mountain catchment. *Earth Surface Dynamics*, *2*(1), 21.

Burtin, A., Hovius, N., & Turowski, J. M. (2016). Seismic monitoring of torrential and fluvial processes. *Earth Surface Dynamics*, *4*(2), 285.

Chawla, N. V. (2010). Data mining for imbalanced datasets: An overview. In O. Maimon & L. Rokach (Eds.), *Data mining and knowledge discovery handbook* (pp. 875–886). Springer US.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*.

Chen, Y., Zhang, M., Bai, M., & Chen, W. (2019). Improving the Signal-to-Noise Ratio of Seismological Datasets by Unsupervised Machine Learning. *Seismological Research Letters*, *90*(4), 1552.

Chmiel, M., Walter, F., Wenner, M., Zhang, Z., McArdell, B. W., & Hibert, C. (2021). Machine learning improves debris flow warning. *Geophysical Research Letters*, *48*(3), e2020GL090874.

Chouet, B., & Dawson, P. (2013). Very long period conduit oscillations induced by rockfalls at kilauea volcano, hawaii. *Journal of Geophysical Research: Solid Earth*, *118*(10), 5352.

Christ, M., Kempa-Liehr, A. W., & Feindt, M. (2016). Distributed and parallel time series feature extraction for industrial big data applications.

Christen, M., Kowalski, J., & Bartelt, P. (2010). Ramms: Numerical simulation of dense snow avalanches in three-dimensional terrain. *Cold Regions Science and Technology*, *63*(1), 1.

Coe, J. A., Bessette-Kirton, E. K., & Geertsema, M. (2018). Increasing rock-avalanche size and mobility in Glacier Bay National Park and Preserve, Alaska detected from 1984 to 2016 Landsat imagery. *Landslides*.

Cole, S. E., Cronin, S. J., Sherburn, S., & Manville, V. (2009). Seismic signals of snow-slurry lahars in motion: 25 september 2007, mt ruapehu, new zealand. *Geophysical Research Letters*, *36*(9).

Collins, B. D., & Stock, G. M. (2016). Rockfall triggering by cyclic thermal stressing of exfoliation fractures. *Nature Geoscience*, *9*(5), 395.

Correa, J., Dean, T., Zaanen, L. V., Tertyshnikov, K., Pevzner, R., & Bona, A. (2018). A comparison of a conventional borehole tool and distributed acoustic sensing at a dedicated field laboratory. *ASEG Extended Abstracts*, *2018*(1), 1.

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*.

Coviello, V., Arattano, M., & Turconi, L. (2015). Detecting torrential processes from a distance with a seismic monitoring network. *Natural Hazards*, *78*(3), 2055.

Crawford, W. C., & Webb, S. C. (2000). Identifying and removing tilt noise from low-frequency (<0.1 Hz) seafloor vertical seismic data. *Bulletin of the Seismological Society of America*, *90*(4), 952.

Dammeier, F., Moore, J. R., Hammer, C., Haslinger, F., & Loew, S. (2016). Automatic detection of alpine rockslides in continuous seismic data using hidden Markov models. *Journal of Geophysical Research: Earth Surface*.

Dammeier, F., Moore, J. R., Haslinger, F., & Loew, S. (2011). Characterization of alpine rockslides using statistical analysis of seismic signals. *Journal of Geophysical Research: Earth Surface*, *116*(4), 1.

Deparis, J., Jongmans, D., Cotton, F., Baillet, L., Thouvenot, F., & Hantz, D. (2008). Analysis of rock-fall and rock-fall avalanche seismograms in the French Alps. *Bulletin of the Seismological Society of America*, *98*(4), 1781.

Diaz-Moreno, A., Roca, A., Lamur, A., Munkli, B.-H., Ilanko, T., Pering, T. D., Pineda, A., & De Angelis, S. (2020). Characterization of acoustic infrasound signals at volcan de fuego, guatemala: A baseline for volcano monitoring. *Frontiers in Earth Science*, *8*, 469.

Dietze, M., Turowski, J. M., Cook, K. L., & Hovius, N. (2017). Spatiotemporal patterns, triggers and anatomies of seismically detected rockfalls. *Earth Surface Dynamics*, *5*(4), 757.

Domınguez, A. B. (1966). *Stresses and displacements in semi-infinite media.* (Doctoral dissertation). Massachusetts Institute of Technology.

Ekström, G., & Stark, C. P. (2013). Simple scaling of catastrophic landslide dynamics. *Science*, *339*(6126), 1416.

Farin, M., Tsai, V. C., Lamb, M. P., & Allstadt, K. E. (2019). A physical model of the high-frequency seismic signal generated by debris flows. *Earth Surface Processes and Landforms*, *44*(13), 2529.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*.

Fernández-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems? *Journal of Machine Learning Research*.

Fontaine, F. R., Roult, G., Michon, L., Barruol, G., & Muro, A. D. (2014). The 2007 eruptions and caldera collapse of the piton de la fournaise volcano (la réunion island) from tilt analysis at a single very broadband seismic station. *Geophysical Research Letters*, *41*(8), 2803.

Genco, R., & Ripepe, M. (2010). Inflation-deflation cycles revealed by tilt and seismic records at Stromboli volcano. *Geophysical Research Letters*, *37*(12), 1.

Gimbert, F., Tsai, V. C., & Lamb, M. P. (2014). A physicalmodel for seismic noise generation by turbulent flow in rivers. *Journal of Geophysical Research F: Earth Surface*.

Graizer, V. (2006). Tilts in strong ground motion. *Bulletin of the Seismological Society of America*, *96*(6), 2090.

Gruber, S., Hoelzle, M., & Haeberli, W. (2004). Permafrost thaw and destabilization of alpine rock walls in the hot summer of 2003. *Geophysical Research Letters*, *31*(13).

Hammer, C., Ohrnberger, M., & Fäh, D. (2013). Classifying seismic waveforms from scratch: A case study in the alpine environment. *Geophysical Journal International*.

Hammer, C., Fäh, D., & Ohrnberger, M. (2017). Automatic detection of wet-snow avalanche seismic signals. *Natural Hazards*, *86*(2), 601.

Heck, M., Hammer, C., Van Herwijnen, A., Schweizer, J., & Fäh, D. (2018). Automatic detection of snow avalanches in continuous seismic data using hidden Markov models. *Natural Hazards and Earth System Sciences*, *18*(1), 383.

Helmstetter, A., & Garambois, S. (2010). Seismic monitoring of Schilienne rockslide (French Alps): Analysis of seismic signals and their correlation with rainfalls. *Journal of Geophysical Research: Earth Surface*.

Hibert, C., Mangeney, A., Grandjean, G., Baillard, C., Rivet, D., Shapiro, N. M., Satriano, C., Maggi, A., Boissier, P., Ferrazzini, V., & Crawford, W. (2014). Automated identification, location, and volume estimation of rockfalls at Piton de la Fournaise volcano. *Journal of Geophysical Research: Earth Surface*.

Hibert, C., Mangeney, A., Grandjean, G., & Shapiro, N. M. (2011). Slope instabilities in Dolomieu crater, Réunion Island: From seismic signals to rockfall characteristics. *Journal of Geophysical Research: Earth Surface*, *116*(4), 1.

Hibert, C., Michéa, D., Provost, F., Malet, J. P., & Geertsema, M. (2019). Exploration of continuous seismic recordings with a machine learning approach to document 20 yr of landslide activity in Alaska. *Geophysical Journal International*.

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.

Hirschberg, J., Badoux, A., McArdell, B. W., Leonarduzzi, E., & Molnar, P. (2021). Evaluating methods for debris-flow prediction based on rainfall in an alpine catchment. *Natural Hazards and Earth System Sciences*, *21*(9), 2773.

Hock, R., Rasul, G., Adler, C., Cáceres, B., Gruber, S., Hirabayashi, Y., Jackson, M., Kääb, A., Kang, S., Kutuzov, S., Milner, A., Molau, U., Morin, S., Orlove, B., & Steltzer, H. I. (2019). Chapter 2: High Mountain Areas. *Ipcc special report on the ocean and cryosphere in a changing climate*.

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.

Iverson, R. M. (1997). The physics of debris flows. *Reviews of Geophysics*, *35*(3), 245.

Iverson, R. M., & George, D. L. (2019). Basal stress equations for granular debris masses on smooth or discretized slopes. *Journal of Geophysical Research: Earth Surface*, *124*(6), 1464.

Iverson, R. M., Logan, M., LaHusen, R. G., & Berti, M. (2010). The perfect debris flow? aggregated results from 28 large-scale experiments. *Journal of Geophysical Research: Earth Surface*, *115*(F3).

Iverson, R., & Logan, M. (2017). Sensor data from debris-flow experiments conducted in june, 2016, at the usgs debris-flow flume, hj andrews experimental forest, blue river, oregon: U.s. geological survey data release.

Jousset, P., Reinsch, T., Ryberg, T., Blanck, H., Clarke, A., Aghayev, R., Hersir, G. P., Henninges, J., Weber, M., & Krawczyk, C. M. (2018). Dynamic strain determination using fibre-optic cables allows imaging of seismological and structural features. *Nature Communications*, *9*(1), 2509.

Kanamori, H., & Given, J. W. (1982). Analysis of long-period seismic waves excited by the may 18, 1980, eruption of mount st. helens—a terrestrial monopole? *Journal of Geophysical Research: Solid Earth*, *87*(B7), 5422.

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. *Geophysical Research Letters*, *42*(15), 6365.

Kenner, R., Noetzli, J., Hoelzle, M., Raetzo, H., & Phillips, M. (2019). Distinguishing ice-rich and ice-poor permafrost to map ground temperatures and ground ice occurrence in the swiss alps. *The Cryosphere*, *13*(7), 1925.

Krischer, L., Megies, T., Barsch, R., Beyreuther, M., Lecocq, T., Caudron, C., & Wassermann, J. (2015). ObsPy: A bridge for seismology into the scientific Python ecosystem. *Computational Science and Discovery*.

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.

Larose, E., Carrière, S., Voisin, C., Bottelin, P., Baillet, L., Guéguen, P., Walter, F., Jongmans, D., Guillier, B., Garambois, S., Gimbert, F., & Massey, C. (2015). Environmental seismology: What can we learn on earth surface processes with ambient noise? *Journal of Applied Geophysics*, *116*, 62.

Lemaître, G., Nogueira, F., & Aridas, C. K. (2017). Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *Journal of Machine Learning Research*.

Ling, F. F., Lai, W. M., & Lucca, D. A. (2002). *Fundamentals of surface mechanics: With applications*. Springer Science & Business Media.

Logan, M., Iverson, R. M., & Obryk, M. K. (2018). Video documentation of experiments at the usgs debris-flow flume 1992–2017 (ver 1.4, january 2018). *US Geological Survey Open-File Report*, *1315*.

Lyons, J. J., Waite, G. P., Ichihara, M., & Lees, J. M. (2012). Tilt prior to explosions and the effect of topography on ultra-long-period seismic records at Fuego volcano, Guatemala. *Geophysical Research Letters*, *39*(8), 1.

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*.

Malfante, M., Dalla Mura, M., Metaxian, J. P., Mars, J. I., Macedo, O., & Inza, A. (2018). Machine Learning for Volcano-Seismic Signals: Challenges and Perspectives. *IEEE Signal Processing Magazine*.

Mangeney, A., Tsimring, L. S., Volfson, D., Aranson, I. S., & Bouchut, F. (2007). Avalanche mobility induced by the presence of an erodible bed and associated entrainment. *Geophysical Research Letters*, *34*(22).

Mangeney-Castelnau, A., Bouchut, F., Vilotte, J. P., Lajeunesse, E., Aubertin, A., & Pirulli, M. (2005). On the use of saint venant equations to simulate the spreading of a granular mass. *Journal of Geophysical Research: Solid Earth*, *110*(B9).

Marchetti, E., Walter, F., Barfucci, G., Genco, R., Wenner, M., Ripepe, M., McArdell, B., & Price, C. (2019). Infrasound Array Analysis of Debris Flow Activity and Implication for Early Warning. *Journal of Geophysical Research: Earth Surface*, *124*(2), 567.

Martel, S. (2004). Mechanics of landslide initiation as a shear fracture phenomenon. *Marine Geology*, *203*(3), 319.

McArdell, B. W., Bartelt, P., & Kowalski, J. (2007). Field observations of basal forces and fluid pore pressure in a debris flow. *Geophysical Research Letters*, *34*(7).

Meyer, M., Wenner, M., Hibert, C., Walter, F., & Thiele, L. (2021). Using system context information to complement weakly labeled data. *CoRR*, *abs/2107.10236*.

NASA JPL. (2020). Nasadem merged dem global 1 arc second v001 [data set].

Norris, R. D. (1994). Seismicity of rockfalls and avalanches at three Cascade Range volcanoes: implications for seismic detection of hazardous mass movements. *Bulletin - Seismological Society of America*, *84*(6), 1925.

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., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*.

Phillips, M., Wolter, A., Lüthi, R., Amann, F., Kenner, R., & Bühler, Y. (2017). Rock slope failure in a recently deglaciated permafrost rock wall at Piz Kesch (Eastern Swiss Alps), February 2014. *Earth Surface Processes and Landforms*.

Pierson, T. C., Costa, J. E., & Vancouver, W. (1987). A rheologic classification of subaerial sediment-water flows. *Debris Flows/Avalanches: Process, Recognition, and Mitigation. Reviews in Engineering Geology. Geological Society of America*, *7*, 1.

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.

Provost, F., Malet, J.-P., Hibert, C., Helmstetter, A., Radiguet, M., Amitrano, D., Langet, N., Larose, E., Abancó, C., Hürlimann, M., Lebourg, T., Levy, C., Le Roy, G., Ulrich, P., Vidal, M., & Vial, B. (2018). Towards a standard typology of endogenous landslide seismic sources. *Earth Surface Dynamics*, *6*(4), 1059.

Rodgers, P. (1968). The response of the horizontal pendulum seismometer to rayleigh and love waves, tilt, and free oscillations of the earth. *Bulletin of the Seismological Society of America*, *58*(5), 1385.

Rosser, N., Lim, M., Petley, D., Dunning, S., & Allison, R. (2007). Patterns of precursory rockfall prior to slope failure. *Journal of Geophysical Research: Earth Surface*.

Salem, H. S. (2000). Poisson's ratio and the porosity of surface soils and shallow sediments, determined from seismic compressional and shear wave velocities. *Géotechnique*, *50*(4), 461.

Savi, S., Comiti, F., & Strecker, M. R. (2021). Pronounced increase in slope instability linked to global warming: A case study from the eastern european alps. *Earth Surface Processes and Landforms*, *46*(7), 1328.

Schimmel, A., Hübl, J., McArdell, B. W., & Walter, F. (2018). Automatic identification of alpine mass movements by a combination of seismic and infrasound sensors. *Sensors (Switzerland)*, *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*(11-12), 1097.

Seydoux, L., Balestriero, R., Poli, P., Hoop, M. d., Campillo, M., & Baraniuk, R. (2020). Clustering earthquake signals and background noises in continuous seismic data with unsupervised deep learning. *Nature Communications*, *11*(1), 3972.

Šindija, D., Neuberg, J., & Smith, P. (2021). The complex, static displacement of a very long period seismic signal observed at soufrière hills volcano, montserrat, wi. *Journal of Volcanology and Geothermal Research*, *413*, 107209.

Stone, M. (1974). Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society: Series B (Methodological)*.

Strunden, J., Ehlers, T. A., Brehm, D., & Nettesheim, M. (2015). Spatial and temporal variations in rockfall determined from tls measurements in a deglaciated valley, switzerland. *Journal of Geophysical Research: Earth Surface*, *120*(7), 1251.

Suriñach, E., Furdada, G., Sabot, F., Biescas, B., & Vilaplana, J. M. (2001). On the characterization of seismic signals generated by snow avalanches for monitoring purposes. *Annals of Glaciology*, *32*, 268.

Suriñach, E., Vilajosana, I., Khazaradze, G., Biescas, B., Furdada, G., & Vilaplana, J. M. (2005). Seismic detection and characterization of landslides and other mass movements. *Natural Hazards and Earth System Science*, *5*(6), 791.

Thelen, W. A., Lyons, J. J., Haney, M. M., & Anderson, K. R. (2015). Broadband seismometer to tilt conversion for monitoring kılauea volcano. *AGU Fall Meeting Abstracts*, *2015*, G41A.

Tsai, V. C., Minchew, B., Lamb, M. P., & Ampuero, J. P. (2012). A physical model for seismic noise generation from sediment transport in rivers. *Geophysical Research Letters*, *39*(2).

Uieda, L., Tian, D., Leong, W. J., Schlitzer, W., Toney, L., Grund, M., Jones, M., Yao, J., Materna, K., Newton, T., Anant, A., Ziebarth, M., Magen, Y., & Wessel, P. (2021). *PyGMT: A Python interface for the Generic Mapping Tools* (Version v0.4.1). Zenodo.

Vallance, J. W., & Iverson, R. M. (2015). Chapter 37 - lahars and their deposits. In H. Sigurdsson (Ed.), *The encyclopedia of volcanoes (second edition)* (Second Edition, pp. 649–664). Academic Press.

van den Ende, M., Lior, I., Ampuero, J.-P., Sladen, A., Ferrari, A., & Richard, C. (2021). A self-supervised deep learning approach for blind denoising and waveform coherence enhancement in distributed acoustic sensing data.

van Herwijnen, A., & Schweizer, J. (2011). Monitoring avalanche activity using a seismic sensor. *Cold Regions Science and Technology*, *69*(2-3), 165.

van Westen, C. J., van Asch, T. W., & Soeters, R. (2006). Landslide hazard and risk zonation - Why is it still so difficult? *Bulletin of Engineering Geology and the Environment*, *65*(2), 167.

Vilajosana, I., Suriñach, E., Abellán, A., Khazaradze, G., Garcia, D., & Llosa, J. (2008). Rockfall induced seismic signals: Case study in Montserrat, Catalonia. *Natural Hazards and Earth System Science*, *8*(4), 805.

Walter, F., Gräff, D., Lindner, F., Paitz, P., Köpfli, M., Chmiel, M., & Fichtner, A. (2020a). Distributed acoustic sensing of microseismic sources and wave propagation in glaciated terrain. *Nature Communications*, *11*(1), 2436.

Walter, F., Amann, F., Kos, A., Kenner, R., Phillips, M., de Preux, A., Huss, M., Tognacca, C., Clinton, J., Diehl, T., & Bonanomi, Y. (2020b). Direct observations of a three million cubic meter rock-slope collapse with almost immediate initiation of ensuing debris flows. *Geomorphology*, *351*, 106933.

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

Weil, R. R., & Brady, N. C. (2019). *Elements of the nature and properties of soils*. Pearson.

Wenner, M., Hibert, C., van Herwijnen, A., Meier, L., & Walter, F. (2021). Near-real-time automated classification of seismic signals of slope failures with continuous random forests. *Natural Hazards and Earth System Sciences*, *21*(1), 339.

Wenner, M., Walter, F., McArdell, B., & Farinotti, D. (2019). Deciphering debris-flow seismograms at Illgraben, Switzerland. *Debris-Flow Hazards Mitigation: Mechanics, Monitoring, Modeling, and Assessment - Proceedings of the 7th International Conference on Debris-Flow Hazards Mitigation*, 222.

Wieczorek, G. F. (1996). Landslides: Investigation and mitigation. chapter 4-landslide triggering mechanisms. *Transportation Research Board Special Report*, (247).

Wielandt, E., & Forbriger, T. (1999). Near-field seismic displacement and tilt associated with the explosive activity of stromboli. *Annals of Geophysics*, *42*(3).

Wilson, D., Ringler, A. T., & Hutt, C. R. (2017). Detection and characterization of pulses in broadband seismometers. *Bulletin of the Seismological Society of America*, *107*(4), 1773.

Wohletz, K., & Heiken, G. (1992). *Volcanology and geothermal energy* (Vol. 432). University of California Press Berkeley.

Worni, R., Huggel, C., Stoffel, M., & Pulgarın, B. (2012). Challenges of modeling current very large lahars at nevado del huila volcano, colombia. *Bulletin of Volcanology*, *74*(2), 309.

Yuan, B., Tan, Y. J., Mudunuru, M. K., Marcillo, O. E., Delorey, A. A., Roberts, P. M., Webster, J. D., Gammans, C. N. L., Karra, S., Guthrie, G. D., et al. (2019). Using machine learning to discern eruption in noisy environments: A case study using CO 2-driven cold-water geyser in Chimayó, New Mexico. *Seismological Research Letters*, *90*(2A), 591.

Zanuttigh, B., & Lamberti, A. (2007). Instability and surge development in debris flows. *Reviews of Geophysics*, *45*(3).

Zhan, Z. (2019). Distributed Acoustic Sensing Turns Fiber-Optic Cables into Sensitive Seismic Antennas. *Seismological Research Letters*, *91*(1), 1.

Zhang, Z., Walter, F., McArdell, B. W., Wenner, M., Chmiel, M., de Haas, T., & He, S. (2021). Insights from the particle impact model into the high-frequency seismic signature of debris flows. *Geophysical Research Letters*, *48*(1), e2020GL088994.

ACKNOWLEDGEMENTS

This thesis would not have been possible without support from colleagues, friends, and family. In this section, I would like to express my gratitude to the most prominent contributors - directly and indirectly. Although I want to stress that I got to know many more inspirational people who left their mark on this journey.

First of all, I would like to thank my two thesis advisors, Fabian Walter, and Daniel Farinotti. During this theses, I was lucky to be supervised by two extraordinary scientists and individuals whose mentoring styles go hand in hand.

Fabian helped me shape this project according to my interests and skills, led me through many scientific pits, and helped me to always come out at the top. Without his encouragement in times of self-doubt, patience with my last-minute attitude, and help in theoretical aspects of my work, I am not sure where else I would be now. His untiring passion for science and advance inspired me and motivated me to strive for my best.

Even though Daniel was only partially scientifically involved in my project, I cannot express my gratitude enough to have had Daniel on board. Daniel's unprecedented managing skills helped me stay on track and fulfill the milestones I had set for myself. His positive outlook and trustworthy personality made it easy to confide in him, whether with a scientific problem or a personal one. His commitment to his group, an open ear for new ideas, and never-ending attempts (and success) to improve the group's everyday life is inspiring.

Thank you, Fabian and Daniel, for guiding me through this time, making me feel comfortable to ask stupid questions, always taking your time and supporting me in whatever it took.

I also would like to thank Alessia Maggi for co-examining my thesis and Ioannis Anastasopoulos for chairing my doctoral exam.

My deep gratitude also extends to my mentors during my studies, Joachim Wassermann, Heiner Igel, and Wayne Crawford. Your guidance and kindness set the foundation for my short scientific career.

This thesis would not have been possible without the help from my co-authors foremost, Clément Hibert and Kate Allstadt. Thank you for your many scientific discussions and valuable input. A special thanks to Clément for making my visit to Strasbourg so worthwhile and including me in social activities. I also want to thank my other co-authors, Alec van Herwijnen, Lorenz Meier, Weston Thelen, Andrew Lockhart, Jacob Hirschberg, and Brian McArdell, for their scientific input and providing data. Your contributions unequivocally improved the output of this thesis.

The basis is this thesis is the seismic data recorded at Illgraben and at Pizzo Cengalo. Without the Swiss Seismological Service (SED) and foremost Pascal Graf and Raphael Moser, I could not have conducted my fieldwork at Illgraben. Thank you, Pascal, for preparing the seismic set-up for Illgraben and being such a comfortable and fun fieldwork partner. Thank you to Raphael Moser for helping out when Pascal couldn't and always keeping me safe while going to Gorwetschgrat. I can honestly say that I never had a lousy fieldwork minute with you two. I also owe many thanks to Pierre Huguenin for staying calm and supporting me 100% during the probably worst field-day of my life.

Even though the circumstances were not ideal on that day, I very much enjoyed your company, as I did on the other field days we had together. Un merci tout particulier à Augustin Rion, ma personne de contact auprès de la commune du Val'Anniviers, qui m'a toujours aidée et inspirée par sa personnalité joyeuse. Furthermore, I want to express my gratitude to Martin Funk, Roger Lörtscher, Fabian Lindner, Dominik Gräff, Robert Tanner, Roman Racin, Jacob Hirschberg, Gosia Chmiel, Johanne Klahold, Lizzy Clyne, Zhen Zhang, and the many other people involved in my fieldwork. Last but not least, I want to thank Lino Schmid at Geopraevent, for the various iterations of data downloading and software updating he did for at Pizzo Cengalo.

This work was funded by the WSL Climate Change Impacts on Alpine Mass Movement initiative. I want to thank Alex Bast for managing the program and organizing the meetings. I very much enjoyed learning about the other projects. The Swiss military supported this theses by transports to Illgraben. The helicopter flights were always a highlight. Thank you to the European Cooperation in Science and Technology for funding my short-term scientific mission at Strasbourg.

I want to express my deep gratitude to all of my great colleagues at VAW and WSL. The prospect of seeing my colleagues and friends made it much easier to go to work with a smile on my face. Thank you to all (former) VAW glaciology employees who welcomed me at VAW when I started. You made it easy to feel integrated and welcome in the group.

A special thanks go to Dominik, Fabi, Gosia, and Jacob, who are not only my colleagues but dear friends I would not want to miss. Thank you for listening and being there for me, and thank you for the fun vacations and evenings we spent together.

Outside of work, I want to thank the earth-science Friday-beer team for offering their assistance to build a network amongst Zurich's earth scientists. I met so many inspiring and fun people there, and I want to express my gratitude to each one of them.

One of the most valuable things I found during my time as a Ph.D. is my friendship with Sölvi, Jesper, Lars, Nienke, Paul and Anna. There is never a dull moment with you people, and I am looking forward to many more experiences with you. I also want to thank BFG for the incredible time in Vienna and Leeds.

Vor allem möchte ich meiner Schwester und ihrem Mann, Dani und Stefan, und meinen Eltern, Roswitha und Günther danken. Ich kann nur hoffen, all die Unterstützung und Liebe die ihr mir entgegengebracht habt irgendwann zurückgeben zu können. Ohne euch wäre ich heute nicht hier.

Finally, I want to thank my partner Alex, who helped me to get through the last years of my Ph.D. Thank you being the incredible person that you are and for your support, patience, and love, in every aspect of life.

CURRICULUM VITAE

PERSONAL DATA

Name	Michaela Wenner
Date of Birth	February 15, 1994
Place of Birth	Dingolfing, Germany
Citizen of	Germany

EDUCATION

2017 - 2021	**Ph.D. Thesis**, *'Seismic Monitoring and Machine Learning for Gravitational Hazards in Alpine Terrain'*, ETH Zurich, Lab. of Hydraul., Hydrol. and Glaciol. (VAW), Supervisor: Fabian Walter
2017	**Master's Thesis**, *'Constraining the state and evolution of stress within Lucky Strike volcano, Mid-Atlantic Ridge, using a seafloor seismometer network'*, Insitut de Physique du Globe de Paris (IPGP), Marine Geosciences group (Wayne C. Crawford)
2016 - 2017	**M.Sc. Geophysics** (Geophysique terre solide), Institut du Physique du Globe de Paris, Université Paris 7
2015 - 2016	**M.Sc. Geophysics**, Joint degree of Technical University Munich (TUM) und Ludwig-Maximilians-University Munich (LMU)
2012 - 2015	**B.Sc. Geosciences**, Joint degree of Technical University Munich (TUM) und Ludwig-Maximilians-University Munich (LMU)

PUBLICATIONS

2021:

7. Chmiel, M., Walter, F., **Wenner, M.**, Zhang, Z., McArdell, B. W., & Hibert, C. (2021). Machine Learning improves debris flow warning. *Geophysical Research Letters, 48(3)*, e2020GL090874.

6. Meyer, M., **Wenner, M.**, Hibert, C., Walter, F., & Thiele, L. (2021). Using system context information to complement weakly labeled data. arXiv preprint arXiv:2107.10236.

5. **Wenner, M.**, Hibert, C., van Herwijnen, A., Meier, L., & Walter, F. (2021). Near-real-time automated classification of seismic signals of slope failures with continuous random forests. *Natural Hazards and Earth System Sciences, 21(1)*, 339-361.

4. **Wenner, M.**, Allstadt, K., Thelen, W., Hirschberg, J., McArdell, B. W., & Walter, F. (2021). Seismometer records of ground tilt induced by debris flows. Under review at *Bulletin of the Seismological Society of America*

3. Zhang, Z., Walter, F., McArdell, B. W., **Wenner, M.**, Chmiel, M., de Haas, T., & He, S. (2021). Insights From the Particle Impact Model Into the High-Frequency Seismic Signature of Debris Flows. *Geophysical Research Letters, 48(1)*, e2020GL088994.

2019:

2. 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.

1. **Wenner, M.**, Walter, F., McArdell, B., & Farinotti, D. (2019). Deciphering debris-flow seismograms at Illgraben, Switzerland. *Contained in: Proceedings of the Seventh International Conference on Debris-Flow Hazards Mitigation, Golden, Colorado, USA, June 10-13, 2019*, 11(1), 1-10.