Detection of debris-flow events from seismic signals using Benford's law

Qi Zhou[1, 2*], *Hui* Tang[1], *Jens* M. Turowski[1], *Jean* Braun[1, 2], *Ci-Jian* Yang[1], *Michael* Dietze[1, 3], *Sophie* Lagarde[1, 2], *Ahmed* Abdelwahab[1, 2], and *Fabian* Walter[4]

[1]Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, 14473 Potsdam, Germany
[2]Institute of Geosciences, University of Potsdam, 14473 Potsdam, Germany
[3]Faculty of Geoscience and Geography, Georg-August-Universität Göttingen, 37077 Göttingen, Germany
[4]Swiss Federal Institute for Forest, Snow and Landscape Research, 8903 Zürich, Switzerland

Abstract. The first step in building an early warning system using seismic signals is to automatically identify events of interest. Here, the first digit distribution of seismic signals generated by debris flows and other surface processes was calculated to validate compliance with Benford's law (BL). A detector model for debris flow events was introduced based on amplitude range and goodness of fit of BL. We show that seismic signals generated by debris flows, landslides, and bedload transport follow the BL. These events release more energy and last longer than rockfalls, which do not follow BL. In the test dataset with 1224 samples, the accuracy of the detector model in identifying debris flow events was 0.75.

1 Introduction

Debris flows are hazards that occur in channels (e.g., gullies, ravines, and valleys) with massive destructive power [1, 2]. Despite significant efforts to mitigate debris flow hazards through warning systems, risk assessment, and structural measures [3-5], extensive damage and casualties still cannot be prevented due to the complex geological conditions and dynamic processes governing debris flows [6]. Early warning is a promising approach to reducing debris flow hazards, which is becoming more precise and reliable as data acquisition and transmission improve.

Debris flow warning systems could employ two types of sensors to collect data: measuring the indicating parameters and monitoring the flow dynamics [7]. Traditional rainfall data-driven systems belong to the former class and require locally tailored thresholds to trigger alarms [8]. However, obtaining rainfall data and maintaining monitoring devices is not straightforward for catchments with large elevation differences and multiple sediment supply areas [9]. In the case of Illgraben (Switzerland), where we conducted the present study, the locally convective short-duration storms are difficult to monitor but are sufficient to trigger debris flows [10]. Continuous seismic signals belong to the second sensor type offering a relatively new opportunity to monitor debris flows with high temporal resolution. Coviello et al. [11] proposed an early warning system based on an automatic debris flow detector using seismic waveforms and a short-term average to long-term average ratio (STA/LTA) trigger algorithm. Chmiel et al. [12] combined the random forest algorithm and seismic waveform features to detect debris flow fronts and identify when they pass over check dams. In addition, seismic networks can contribute to detecting and locating large mass movements on a regional scale [13]. However, a seismic station senses all ground shaking within its bandwidth and records the signals, blending events of interest with ambient noise. Only a small amount of seismic data corresponds to geomorphic processes, which makes labelling events a time-consuming work requiring expert knowledge. The present seismic-driven early warning methods require labelled data to train a detection model. Transferability of these models to other study sites where little or no training data exist is not guaranteed.

Benford's law (BL) is a statistical phenomenon of the probability distribution of the first digit in a dataset. It originated from the observation that the first few pages of the logarithm tables were more worn out than others. Newcomb [14] declared that the probability of the first digit is such that the mantissae of their logarithms are equally probable. BL was rediscovered and tested with 20 different data domains by physicist *Frank Benford* [15] and named after him. BL gives the probability of the first digit:

$$P_D = \log_{10}(1+1/D) \qquad (1)$$

where P_D is the theoretical probability of the first non-zero digit. D ranges from 1 to 9. For example, 0.01, 100, -1 share the same first digit one with a probability of 0.301.

A classical mathematical theory to explain why the first digit of a dataset appear in BL is that after a sufficiently long computation in floating-point arithmetic, the occurring mantissas have a nearly

---

* Corresponding author: qi.zhou@gfz-potsdam.de

logarithmic distribution [16]. For more information, we refer to [17-18]. Even if BL is base and scale-invariant [19], not all collected datasets follow it. Durtschi et al. [20] summarized the guidelines on when data may conform to BL: (1) unrestricted (data from human height and weight is, e.g., restricted) and assigned data (e.g., check numbers and invoice numbers), and (2) large sets of data across several orders of magnitude (e.g. array from 1 to 10^7 with a sample size of 1.2×10^5). BL has been typically applied to examine data anomalies [21], and some researchers have adopted it to finance and sociology [22-23]. Recent work has applied BL to geoscience data, such as investigating natural hazard dataset homogeneity and anomalies [24-25]. It has been demonstrated that earthquakes and marsquakes are detectable with BL using seismic waveforms [26-27].

In this work, we calculate the first digit distribution of seismic signals (generated by debris flows, floods, and other surface processes [28-29]) and validate the compliance with BL. Then, an event detector model based on seismic signals and BL is introduced with the example of debris flow.

2 Study site and event catalog

2.1 Study site

The Illgraben catchment, located near the village of Leuk, in southwestern Switzerland (Fig 1), covers an area of about 9.5 km^2 and ranges in elevation from the Rhône River at 610 m to Illhorn mountain at 2716 m [30]. In the upstream channel trunk, the limestones (north side) and quartzites (south side) are susceptible to erosion and weathering, which provides debris flow material [31]. The Illgraben covers less than 0.2% of the Rhone Valley, but it contributes more than 5% of the annual sediment to the Rhône basin [32]. The local annual rainfall is concentrated from May to October, and debris flows are mainly triggered by short-duration storms [10]. Three to five debris flows and several floods are observed at the Illgraben catchment each year.
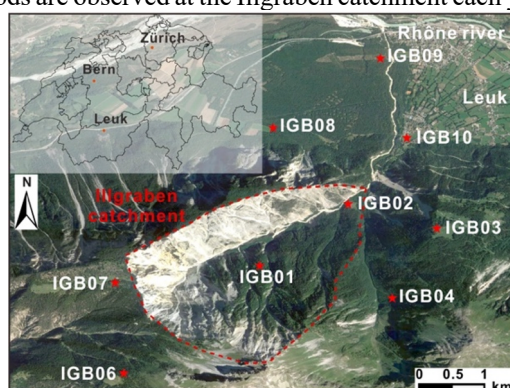

Fig. 1. Network of seismometers (red dot) in the Illgraben catchment.

For monitoring the surface processes, such as rockfall, landslide, and debris flow, ten three component (N-S, W-E, vertical) seismic stations were installed around the Illgraben (IGB01-10) during the period 2012 to 2014. The signal bandwidth and sampling frequency of these seismic stations are 1-100 Hz and 200 Hz,

respectively. Considering seismogenesis is concentrated at the debris flow front [33], the greater the distance between the flow front and the seismic station, the more significant the proportion of noise in the recorded signal. Here, only station IGB02, which is closest to the channel and far from the nearby residential area of Leuk, was selected as the data provider.

2.2 Event catalog

From 2013 to 2014, 24 debris flow events (*true positive event, TPE*) were collected, and one may be a flood. Ten of these events were observed by the Swiss Federal Institute for Forest, Snow and Landscape Research warning systems (henceforth referred to as "WSL"). A debris flow example with a WSL label is shown in Fig. 2. Based on amplitude durations in the waveform and spectral features between 1 and 50 Hz, the remaining 14 debris flow events were manually labelled by us (henceforth referred to as "GFZ"). Unfortunately, no other data for these 14 events were available for cross-validation. The statistics of the time accounted for by *TPE* (about 160 hours in total) and non-*TPE* has a ratio of 1:58. To create a true negative event (*TNE*) catalog (all non-debris flow events), we selected 1200 *TNE* (50 times of *TPE*) events at random start times and random duration (from 20 minutes to 6 hours) outside the *TPE* period. Finally, a test dataset containing 1224 events was created. Besides, one rockfall event (rockfall1) were collected in Illgraben, outside the Illgraben catchment, one landslide in Iceland [28], one rockfall (rockfall2) in Germany [29], and one bedload transport in Taiwan (ROC) were collected to test for compliance with BL.
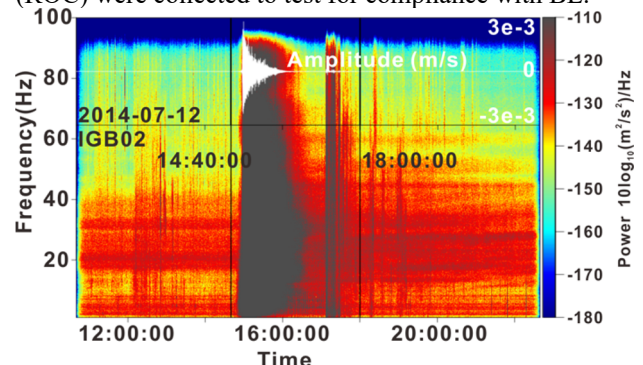
Fig. 2. Waveforms and spectrograms of a debris flow with WSL label (12th July, 2014). The vertical black lines are manually labelled start and end times (14:40, 18:00).

3 Methods

3.1 Validate natural hazard's conformity to BL

The vertical component of raw waveform data, without instrument response deconvolution, demeaning, detrending, filtering, or tapering, is utilized to validate the event conformity to BL. To this end, the first digits are counted with a 60s sliding window. The raw amplitude equal to zero for each moving window will be discarded. To compare the observed first digit distribution with the theoretical BL distribution (Equations in 2-3), we refer to the goodness of fit (Φ)

[26] and introduce the Shannon entropy (*H*) [34] to assist the assessment and detection:

$$\phi = \left[1 - \left(\sum_{d=1}^{9} \frac{(p_d - P_d)^2}{P_d} \right)^{1/2} \right] \times 100\% \qquad (2)$$

$$H(D) = -\sum_{d \in D} p(d) \log_b p(d) \qquad (3)$$

where P_d and p_d are the theoretical and observed probabilities of first digits; H(*D*) is the Shannon entropy H, $D=\{1, 2, …, 8, 9\}$. When $b=2$, the unit of H(*D*) is *bits* or *shannons*. Parameter H describes the relationship between information and uncertainty. The higher the occurrence probability, the less information an event carries. In our case, noise dominates most of the year, so H is expected to increase from a true negative event to a true positive event. The theoretical maximum of H is 3.17 when the first digit (1-9) is a uniform distribution with a probability of 1/9.

3.2 Detector with BL method

There are two indicators to design a debris flow detector with BL, the amplitude range *ar* and goodness of fit *Φ* of BL (Fig. 3a). The *ar* is more concentrated during the *TNE* (noise) (Fig. 3b), ranging only by 1-2 orders of magnitude, while the amplitude could cross several orders during the *TPE* (event). *ar* is defined as:

$$ar = Q_1 - Q_{100} \qquad (4)$$

where Q is the amplitude value from smallest to largest value within the sliding window, Q_1 and Q_{100} are the average within largest and smallest one percentage.
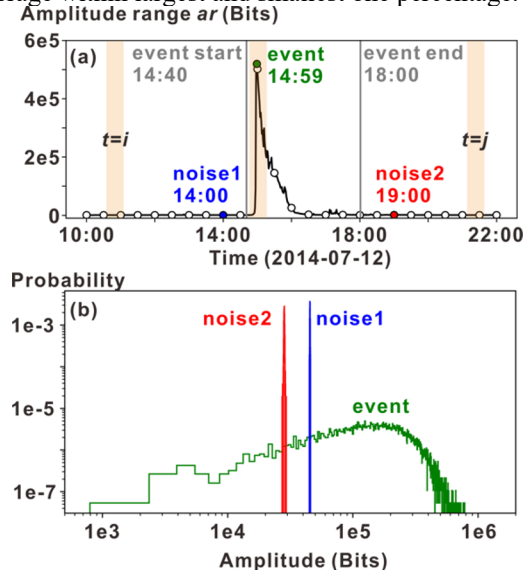
Fig. 3. Amplitude range *ar* over time (a) and amplitude probability distribution (b) of the 12th July, 2014 debris flow.

A sliding window (Fig. 3a) will scan *ar* and *Φ* at 60s intervals. If *ar* at the time *i* is greater than the *threshold start* (5000 bits), a potential event will be recorded until *ar* at time *j* is less than the *threshold off* (1500 bits). When the potential event duration (t_j-t_i) is greater than 20 and *Φ* is greater than a given *threshold g*, the potential event will be marked as *predicted positive*. If the above conditions are not met within the entire event, the test event will be marked as *predicted negative*. The 1224 events in the test dataset will be extracted one by one for model testing. The range of *ar*, *Φ* is set as [3×10³, 1×10⁵], [0, 80], and the interval is 10⁴ and 20,

respectively. We combine the *ar* and *Φ* orthogonally, then feed them into the model for testing.

We use a confusion matrix to assess our detector performance. The detector outputs were divided into four categories (Fig. 4). We employed positive ratio *TPR* and false positive ratio *FPR* to evaluate the detector model performance.

Fig. 4. The confusion matrix for evaluating detector model performance. TP, FP is True positive and False positive number. P, N is the Positive and Negative number.

4 Preliminary results and concluding remarks

Our preliminary results show that the first digit distribution of seismic signals generated by debris flows, landslides, and bedload transport follow BL, while our two rockfall cases do not conform to BL (Fig 5). Moreover, one flood case does not allow for rigorous conclusions to be drawn. When *Φ* is less than 0, and *SH* is close to 0, the event cannot be considered a high-energy mass movements like debris flows and landslides. The detector model has an optimal *TPR* of 0.75 and *FPR* of 0.01, and it could be improved by more parameter sensitivity analysis in the future.
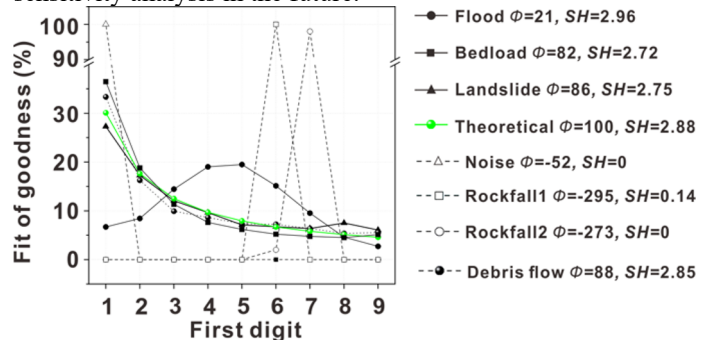
Fig. 5. The first digit distribution of different events. The bedload event is recorded by an acoustic sensor.

The seismic signals from long-duration and high-energy mass movements (e.g., debris flow and landslide) conform to BL (e.g., when the event starts in Fig. 2). This phenomenon provides a new potential approach for rapid and relatively accurate filtering events from seismic signals. In future work, we will explore why BL appears in seismic signals generated by some processes but not in others. Understanding conformity with BL will provide important insights into the performance of seismic mass movement monitoring. Furthermore, the debris flow detector is not yet efficient and reliable enough. It will be necessary to boost and compare our detector model to other approaches, such as the short-term average to long-term average ratio (STA/LTA) and random forest model in the future.

References

1. K. Hutter, B. Svendsen, D. Rickenmann. Debris flow modeling: A review. Continuum. Mech. Thermodyn. **8**, 1 (1994)

2. RM. Iverson. The physics of debris flows. Rev. Geophys. **35**, 97 (1997).

3. J. Huebl, G. Fiebiger. Debris-flow mitigation measures. (Debris-flow Hazards and Related Phenomena, Berlin Springer, 2007)

4. S. Fuchs, K. Heiss, J. Hübl. Natural Hazards and Earth System Sciences Towards an empirical vulnerability function for use in debris flow risk assessment. Hazards. Earth. Syst. Sci. **7**, 1 (2007)

5. RL. Baum, JW. Godt. Early warning of rainfall-induced shallow landslides and debris flows in the USA. Landslides. **7**, 3 (2010)

6. JW. Kean, DM. Staley, JT. Lancaster, et al. Inundation, flow dynamics, and damage in the 9 January 2018 Montecito debris-flow event, California, USA: Opportunities and challenges for post-wildfire risk assessment. Geosphere. **15**, 4 (2019)

7. M. Hürlimann, V. Coviello, C. Bel, et al. Debris-flow monitoring and warning: Review and examples. Earth-Science. Rev. **199** (2019)

8. F. Marra, EI. Nikolopoulos, JD. Creutin Space–time organization of debris flows-triggering rainfall and its effect on the identification of the rainfall threshold relationship. J. Hydrol. **541** (2016)

9. C. Tang, J. Zhu, WL. Li, et al. Rainfall-triggered debris flows following the Wenchuan earthquake. Bull. Eng. Geol. Environ. **68**, 2 (2009)

10. BW. McArdell, P. Bartelt, J. Kowalski. Field observations of basal forces and fluid pore pressure in a debris flow. Geophys. Res. Lett. **34**, 7 (2007)

11. V. Coviello, M. Arattano, F. Comiti, et al. Seismic Characterization of Debris Flows: Insights into Energy Radiation and Implications for Warning. J. Geophys. Res. Earth. Surf. **124**, 6 (2019)

12. M. Chmiel, F. Walter, M. Wenner, et al. Machine Learning Improves Debris Flow Warning. Geophys. Res. Lett. **48**, 3 (2021)

13. K. L. Cook, R. Rekapalli, M. Dietze, et al. Detection and potential early warning of catastrophic flow events with regional seismic networks. Science. **374** (2021)

14. S. Newcomb. Note on the Frequency of Use of the Different Digits in Natural Numbers. Am. J. Math. **4**, 1/4 (1881)

15. F. Benford. The Law of Anomalous Numbers. American Philosophical Society. **78** 551-572 (1938)

16. P. Schatte. On mantissa distributions in computing and Benford's law. Proc. Am. Math. Soc. **24**, 9 (1988)

17. TP. Hill. The Signifilcant Digit Phenomenon. Am. Math. Mon. **86**, (1998)

18. J. Omona, SJ. Miller, P. Fukofuka. Benford's Law: Theory and Applications Benford's Law : Theory and Applications. (Princeton University Press, 2015)

19. TP. Hill. Base-Invariance Implies Benford's Law. Proc. Am. Math. Soc. **123**, (1995)

20. C. Durtschi, W. Hillison, C. Pacini. The effective use of Benford's law to assist in detecting fraud in accounting data. J. Forensic. Account. **5**, 1 (2004)

21. WKT. Cho, BJ. Gaines. Breaking the (Benford) law: Statistical fraud detection in campaign finance. Am. Stat. **61**, 3 (2007)

22. G. Castañeda. La ley de Benford y su aplicabilidad en el análisis forense de resultados electorales. Polit. Y. Gob. **18**, 2 (2011)

23. E. Ley. On the Peculiar Distribution of the U.S. Stock Indexes' Digits. (American Statistician, 1996)

24. A. Geyer, J. Martí. Applying Benford's law to volcanology. Geology. **40**, 4 (2012)

25. R. Joannes-Boyau, T. Bodin, A. Scheffers, et al. Using Benford's law to investigate natural hazard dataset homogeneity. Sci. Rep. **5** (2015)

26. M. Sambridge, H. Tkalčić, A. Jackson. Benford's law in the natural sciences. Geophys. Res. Lett. **37**, 22 (2010)

27. W. Sun, H. Tkalčić. Repetitive marsquakes in Martian upper mantle. Nat. Commun. **13**, 1 (2022)

28. A. Schöpa, WA. Chao, BP. Lipovsky, et al. Dynamics of the Askja caldera July 2014 landslide, Iceland, from seismic signal analysis: Precursor, motion and aftermath. Earth. Surf. Dyn. **6**, 2 (2018)

29. M. Dietze, JM. Turowski, KL. Cook, et al. Spatiotemporal patterns, triggers and anatomies of seismically detected rockfalls. Earth. Surf. Dyn. **5**, 4 (2017)

30. A. Badoux, C. Graf, J. Rhyner, et al. A debris-flow alarm system for the Alpine Illgraben catchment: Design and performance. Nat. Hazards. **49**, 3 (2009)

31. F. Schlunegger, A. Badoux, BW. McArdell, et al. Limits of sediment transfer in an alpine debris-flow catchment, Illgraben, Switzerland. Quat. Sci. Rev. **28**, 11–12 (2009)

32. F. Schlunegger, K. Norton, R. Caduff. Hillslope Processes in Temperate Environments. Treatise. Geomorphol. **7** (2013)

33. M. Farin, VC. Tsai, MP. Lamb, et al. A physical model of the high-frequency seismic signal generated by debris flows. Earth. Surf. Process Landforms. **44**, 13 (2019)

34. CE. Shannon. A Mathematical Theory of Communication. Bell System Technical Journal. **27**, (1948)