Measuring bedload motion time at sub-second resolution using Benford's law on acoustic data

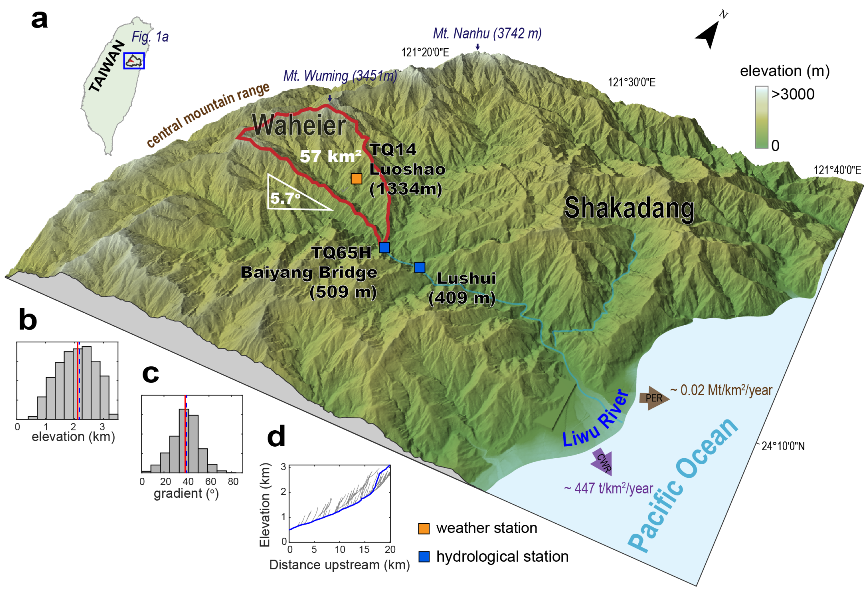
Ci-Jian Yang[1], Jens Martin Turowski[2], Qi Zhou (Chow)[3], Ron Nativ[4], Hui Tang[5], and Wen-Sheng Chen[6]

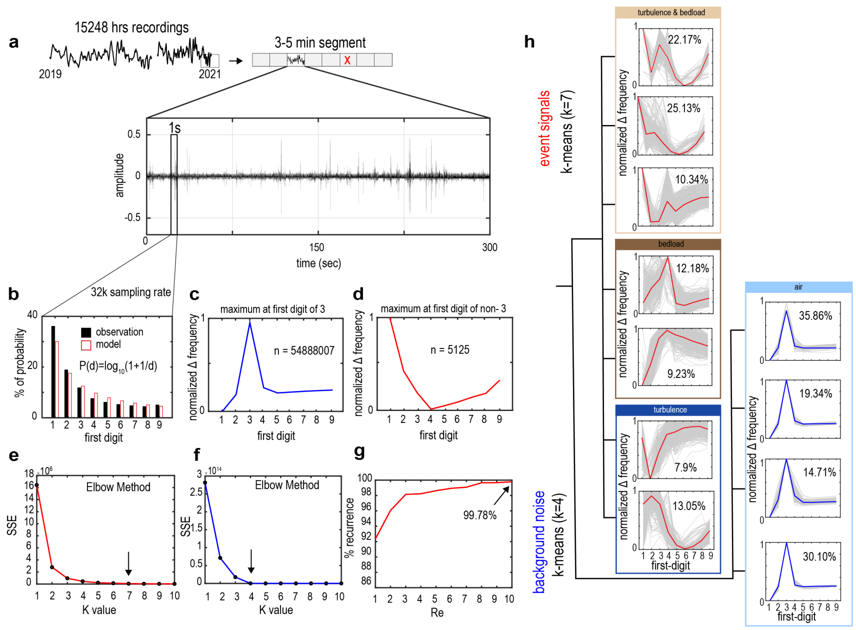
[1]National Taiwan University
[2]GFZ German Research Centre for Geosciences, Potsdam
[3]Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences
[4]Ben-Gurion University of the Negev
[5]German Research Centre for Geosciences (GFZ)
[6]Center for General Education, National Dong Hwa University

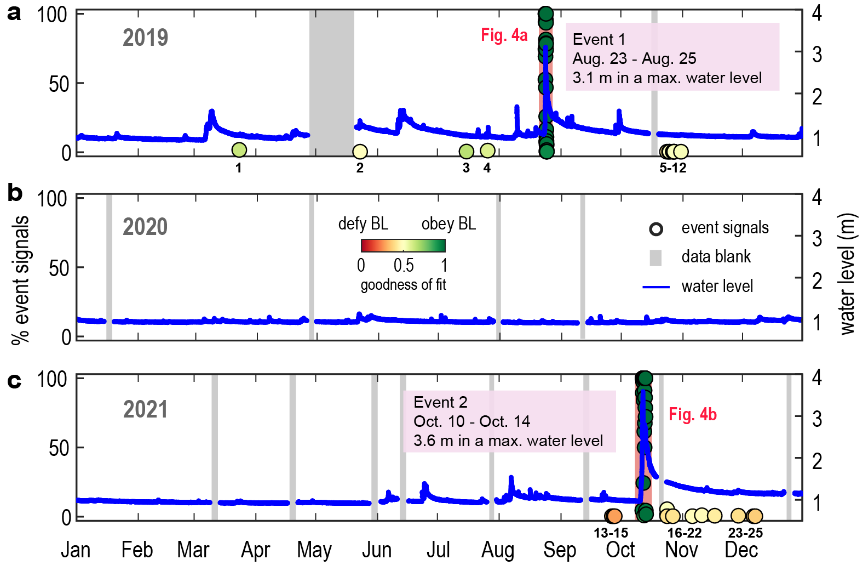
November 8, 2023

Abstract

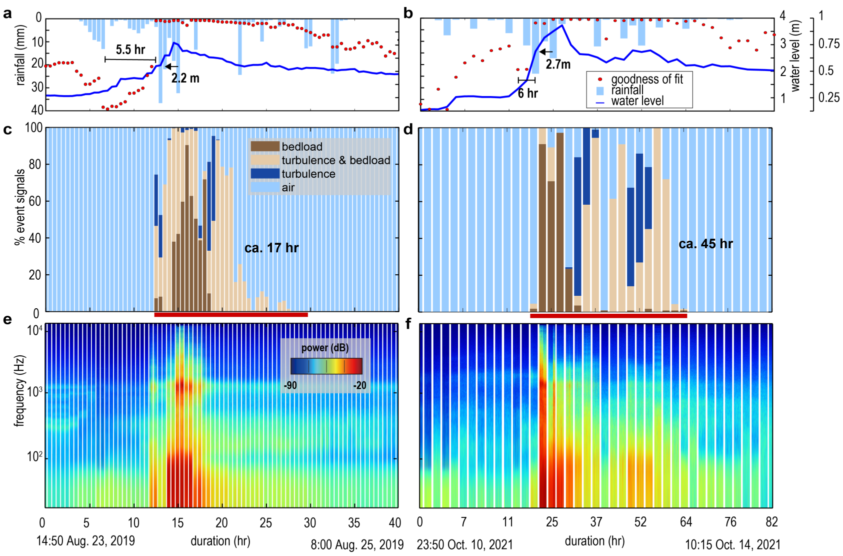
An important component of quantifying bedload transport flux is the identification of the onset of bedload motion. Bedload transport can be monitored with high temporal resolution using passive acoustic methods, e.g., hydrophones. Yet, an efficient method for identifying the onset of bedload transport from long-term continuous acoustic data is still lacking. Benford's Law defines a probability distribution of the first-digit of datasets and has been used to identify anomalies. We apply Benford's Law to the three years of acoustic recordings from a stationary hydrophone in the Taroko National Park, Taiwan. Our workflow allows for monitoring bedload motion in near-real-time, and it is convenient for others to reference. Two bedload transport events were identified during the examined period, lasting 17 and 45 hours, accounting for approximately 0.35% of the time per year.


a
TAIWAN
Fig. 1a
Mt. Nanhu (3742 m)
121°20'0"E
Mt. Wuming (3451m)
121°30'0"E
121°40'0"E
elevation (m)
>3000
0
central mountain range
Waheier
57 km²
TQ14
Luoshao
(1334m)
5.7°
Shakadang
TQ65H
Baiyang Bridge
(509 m)
Lushui
(409 m)
PER
~ 0.02 Mt/km²/year
Liwu River
CWR
~ 447 t/km²/year
Pacific Ocean
24°10'0"N
b
elevation (km)
c
gradient (°)
d
Elevation (km)
Distance upstream (km)
weather station
hydrological station


a
15248 hrs recordings
2019
2021
3-5 min segment
1s
amplitude
time (sec)
b
32k sampling rate
% of probability
observation
model
P(d)=log10(1+1/d)
first digit
c
maximum at first digit of 3
n = 54888007
normalized Δ frequency
first digit
d
maximum at first digit of non- 3
n = 5125
first digit
e
Elbow Method
SSE
K value
f
Elbow Method
SSE
K value
g
% recurrence
99.78%
Re
h
event signals
k-means (k=7)
turbulence & bedload
22.17%
25.13%
10.34%
bedload
12.18%
9.23%
turbulence
7.9%
13.05%
first-digit
background noise
k-means (k=4)
air
35.86%
19.34%
14.71%
30.10%
first-digit



a
2019
Fig. 4a
Event 1
Aug. 23 - Aug. 25
3.1 m in a max. water level
1
2
3
4
5-12
b
2020
defy BL
obey BL
0
0.5
1
goodness of fit
event signals
data blank
water level
% event signals
water level (m)
c
2021
Event 2
Oct. 10 - Oct. 14
3.6 m in a max. water level
Fig. 4b
13-15
16-22
23-25
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec



a
b
rainfall (mm)
5.5 hr
2.2 m
2.7m
6 hr
goodness of fit
rainfall
water level
water level (m)
σ
c
d
% event signals
bedload
turbulence & bedload
turbulence
air
ca. 17 hr
ca. 45 hr
e
f
frequency (Hz)
power (dB)
-90
-20
14:50 Aug. 23, 2019
duration (hr)
8:00 Aug. 25, 2019
23:50 Oct. 10, 2021
duration (hr)
10:15 Oct. 14, 2021

Measuring bedload motion time at sub-second resolution using Benford's law on acoustic data

Ci-Jian Yang[1*], Jens M. Turowski[2], Qi Zhou[2,3], Ron Nativ[2,3,4], Hui Tang[2], Wen-Sheng Chen[5]

[1]Department of Geography, National Taiwan University, Taipei
[2]Helmholtzzentrum Potsdam, GFZ German Research Center for Geosciences, Potsdam
[3]Institute of Geosciences, University of Potsdam, Potsdam
[4]Department of Earth and Environmental Sciences, Ben-Gurion University of the Negev, Be'er Sheva
[5]Center for General Education, National Dong Hwa University, Hualien

*Corresponding author: Ci-Jian Yang (cijianyang@ntu.edu.tw, No. 1, Sec. 4, Roosevelt Road, Taipei, 10617 Taiwan)

Key points:

- Long-term, high-frequency acoustic monitoring constitutes huge-volume datasets and an extremely small signal-to-noise ratio.
- The distinct first-digit distribution between signal and noise can used to filter out 99% of background noise from acoustic recordings.
- We tested the method for three year long acoustic data set in Baiyang, two identified bedload transportation events.

Abstract

An important component of quantifying bedload transport flux is the identification of the onset of bedload motion. Bedload transport can be monitored with high temporal resolution using passive acoustic methods, e.g., hydrophones. Yet, an efficient method for identifying the onset of bedload transport from long-term continuous acoustic data is still lacking. Benford's Law defines a probability distribution of the first-digit of datasets and has been used to identify anomalies. We apply Benford's Law to the three years of acoustic recordings from a stationary hydrophone in the Taroko National Park, Taiwan. Our workflow allows for monitoring bedload motion in near-real-time, and it is convenient for others to reference. Two bedload transport events were identified during the examined period, lasting 17 and 45 hours, accounting for approximately 0.35% of the time per year.

Plain Language Summary

Long-term, high-frequency monitoring of Earth surface processes brings huge datasets and an extremely small signal-to-noise ratio. Benford's Law defines the specific probability distribution of the first-digit of datasets and has been used to identify anomalies and high-energy events. We provide a workflow of applying Benford's Law to identify the onset of the motion of coarse sediment along the river bed at a time resolution of seconds. We identified three separate sound classes in the data related to the noise produced by the motion of pebbles, water flow, and air. The workflow could be referred for other different catchments, events, or datasets. Due to the influence of instrument and background noise on the regularity of the residuals of the first-digit, We recommend identifying the first-digit distribution of the background noise and ruling it out before implementing this workflow.

Keywords acoustic monitoring, bedload, first-digit, event indicator, early warning system

1. Introduction

Bedload transport driven by floods is one of the manifestations of natural processes that strongly affect the Earth's surface system. Bedload transport is a fundamental process in river corridors, with implications for channel stability (e.g. Turowski et al., 2009; Recking et al., 2016), sediment budgets (e.g., Theule et al., 2012), pollution transport (e.g., Stott et al., 2001), fluvial erosion (e.g., Turowski et al., 2008), and aquatic habitats (e.g., Snyder et al., 2009). Bedload transport increases river lateral migration or erosion and deposition, with potentially hazardous effects on downstream residents' lives and property (e.g., Krapesch et al., 2011, Bufe et al., 2019). In Switzerland, bedload transport caused cumulative financial losses of USD

5.3 billion from 1972 to 2011, about one-third of the total natural hazard damage during that period (Badoux et al., 2014). Reliable approaches for bedload monitoring are needed not only for hazard warning systems but also for quantifying fluvial processes.

Monitoring in extreme environments during storms can complement existing observations of fluvial processes, such as understanding temporal changes in bedload motion and calculating the proportion of total sediment flux. Yet, the estimations of bedload transport from long-term monitoring systems are limited. Passive acoustic methods, e.g., hydrophones, and seismometers, are sensitive to bedload motion (e.g., Geay et al., 2017; Burtin et al., 2016) and able to obtain the data at a safe distance. Acoustic data from hydrophones, where bedload impacts can be heard directly, provide a benchmark that is not usually available when using seismic data only (e.g., Roth et al., 2017). In addition, high-frequency acoustic monitoring allows for detecting bedload motion in realtime, which could be used for warning systems, improving over generic empirical values calibrated on previous events (Abancó et al., 2012; Baum & Godt, 2010; Badoux et al., 2014; Marra et al., 2016). However, an automatic and efficient method for constraining the onset of bedload transport events from long-term acoustic data is still lacking.

Benford's Law defines a specific probability distribution of the first-digit of datasets. It predicts that a first-digit of one occurs about 30% of the time in a given dataset, three times higher than the value of 1/9 expected from a uniform distribution. Benford's Law has been used to identify fraud in accounting or political votes (Nigrini, 1999). It appears in natural data as well. For example, nearly half of a million US annual average flows and the size of global lakes and wetlands follow Benford's Law (Nigrini and Steven, 2007). Benford's Law has also been used to distinguish noise from chaotic processes when the process causes higher energy events than baseline noise (Li et al., 2015). For example, the onset of earthquakes has been identified using Benford's Law on seismic amplitude data (Sambridge et al., 2010; Díaz et al., 2015). In addition, accurate and complete observational data on the traveled distance of tropical cyclones conform to Benford's Law. Thus, Benford's Law residuals become a tool for evaluating data quality and homogeneity (Joannes-Boyau et al., 2015).

In underwater acoustic recordings, the median power of bedload-generated noise in the frequency range between 10^3 Hz and 10^4 Hz is about 2.5 orders of magnitude higher than that of the low flow period at the same reach (Geay et al., 2017). Therefore, we hypothesize that the change in the first-digit distribution of acoustic amplitudes can properly identify high-energy events, and in principle, we expect that the first-digit distribution has the potential to be an indicator that can be used to separate sound categories, i.e., air, waterflow, and motion of pebbles. For example, the 95th percentile of power spectral density ranges from 10^4 to 5×10^4 (Geay et al., 2017). This half-order of magnitude data range results in a new first-digit distribution different from Benford's Law.

Here, we develop a simple statistical tool based on mathematical law that can automatically and efficiently identify bedload signals from long-term acoustic recordings. We apply the method to three years of underwater audio observations at Baiyang hydrometric station. We demonstrate the potential of Benford's Law in distinguishing sound categories, which we propose is significant for improving bedload flux calculations.

2. Materials and Methods

2.1 Benford's Law

Benford's Law (Benford, 1938) states that the probability of the first-digit is non-uniform but rather obeys Eq. (1):

$$P_D = \log_{10}\left(1 + {}^{1}/_{D}\right). \quad (1)$$

Here, P_D is the probability of the first-digit D occurring (D = 1, …,9). For example, the first-digit of -0.01, 1, or 1e8 are all 1. The law suggests that numbers beginning with a one occur about 30.1% of the time in some natural datasets, while those with the first- digit of two occur about 17.6% of the time, and so on, down to the first-digit of nine occurring about 4.6% of the time.

We use a least-squares misfit measure to quantify the discrepancy between the observed and theoretical probability of the first-digit (Joannes-Boyau, 2015). We subtract the misfit from one and define it as the goodness of fit (2):

$$\sigma = 1 - \sum_{(D=1)}^{9}\left(100\frac{n_D}{n} - P_D\right)^2, \quad (2)$$

where P_D is the theoretical probability of data with the first-digit D as given by Benford's Law, n_D is the number of data with the first-digit D, and n is the total number of data. The first-digit distribution can be independently assessed for the goodness of fit against theoretical values of Benford's Law, eliminating the need for other detecting methods, such as short-time average/long-time average (STA/LTA), which require long-term observations. In addition, we calculate the acoustic amplitude difference between the 75th and 25th percentile (interquartile range) for every second as an index of the data range.

2.2 Study site and monitoring

The Liwu catchment is located in eastern Taiwan (Figure 1a), experiencing high-frequency seismic activity and rapid tectonic uplift of 5.5 mm yr^{-1} (Petley et al., 1997). The mean annual rainfall is about 2.5 m, and typhoons are the dominant source of heavy rainfall, accounting for 66% of the annual discharge (Huang et al., 2012). This results in 20,000 t km^{-2} y^{-1} of physical denudation rate calculated from suspended sediment (Dadson et al., 2003) and 18 t km^{-2} y^{-1} derived from silicate weathering, which is one of highest measured so far in the world for felsic lithologies (Calmels et al, 2011). The Liwu provides a natural laboratory with active driving forces, relatively minor human influence, and a unique opportunity to investigate bedload dynamics from a typhoon-dominated system.

Baiyang hydrometric station is located on the outlet of Waheier catchment, a tributary of Liwu River, which drains 57 km^2. Elevation in the Waheier catchment spans from 509 to 3451 m with a mean of 2055 m (Figure 1b). The mean hillslope gradient is 39.5^o (Figure 1c), and the mean channel gradient is about 5.7%. The length of the mainstream is 20.8 km (Figure 1d). Baiyang hydrometric station was installed at Baiyang Bridge in April 2018. There, underwater acoustic noise has been continuously measured at a 32 kHz sampling rate using a broadband hydrophone, Aquarian H2a-XLR (Aquarian Audio, 2013). The hydrophone is protected by a 30 cm metal tube attached to the bedrock close to the water surface at a low flow of about ~1 m. Five-minute-resolution measurement of the water stage is measured using a Radar Level Sensor (RLS) with an accuracy of 10 mm. Half-hour time-lapse imagery is recorded by three D30 Canon cameras with different viewpoints. Within the same catchment, Luoshao station (Figure 1) provides minute-resolution rainfall measurements using an automatic weather station, WXT-536.

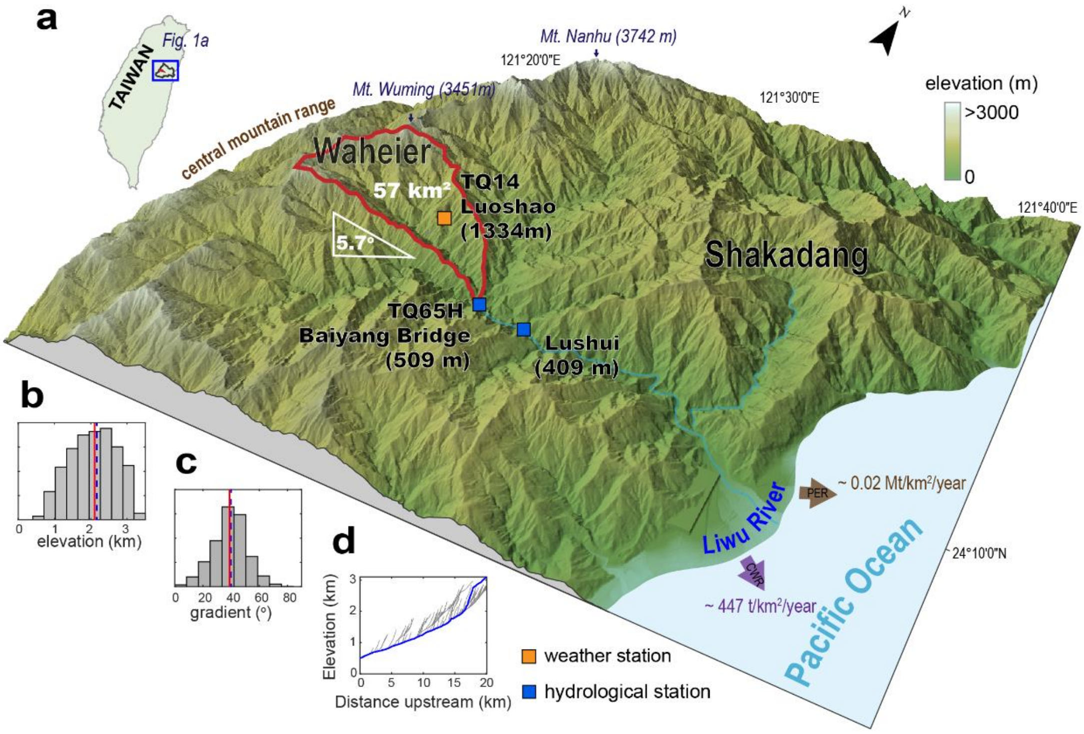


Figure 1. (a) Topographical 3D view of the Liwu catchment and the study site. In the outlet of the Waherier catchment, Baiyang hydrometric station (TQ65H) monitors river acoustic sounds and provides hydrometric data. Minute-resolution rainfall is obtained from the Luoshao (TQ14) weather station. (b) Histogram of elevation of Waheier catchment, red line denotes median value, and blue dash denotes mean value. (c) Histogram of hillslope gradient of Waheier catchment, red line denotes median value, and blue dash denotes mean value. (d) Longitudinal profile of the upstream from the Baiyang station

2.3 Data preparation and audio recording visualization

Signal processing, including detrending and deconvolution, may result in changes in acoustic amplitude, which may mask Benford's Law. Therefore, we did not pre-process the audio data. This has the further advantage of significantly reducing the computational cost of our method. Here, we used the acoustic recordings from the stationary hydrophone deployed from 2019 to 2022 (Figure 2a). The audio data was split into .mp3 files of three to five minutes in length. After removing damaged and short-period files (< 1 minute), we obtained a total of 15,248 hours of acoustic recordings. Each second of recording has 32,000 individual acoustic amplitude measurements, sufficient to calculate the probability distribution of the first-digit. To visualize audio recordings, we transformed the signals from the time domain to the frequency domain using a short-time Fourier transform to obtain the power spectral density.

2.4 Sound classification via residual probability distribution

To distinguish between different sound categories based on the probability of first-digit, our workflow contains three steps. First, we calculate the residual between the probability of first-digit for observed data and Benford's theoretical frequencies, and we categorize the residuals into two groups: event signals and background signals. Second, , we identify sound categories using the k-means clustering and determine the number of clusters using the Elbow method, along with the method to assess the clustering stability. Third, we calculate the time-series ratio of respective sound categories. These steps are described in detail in the supplementary.

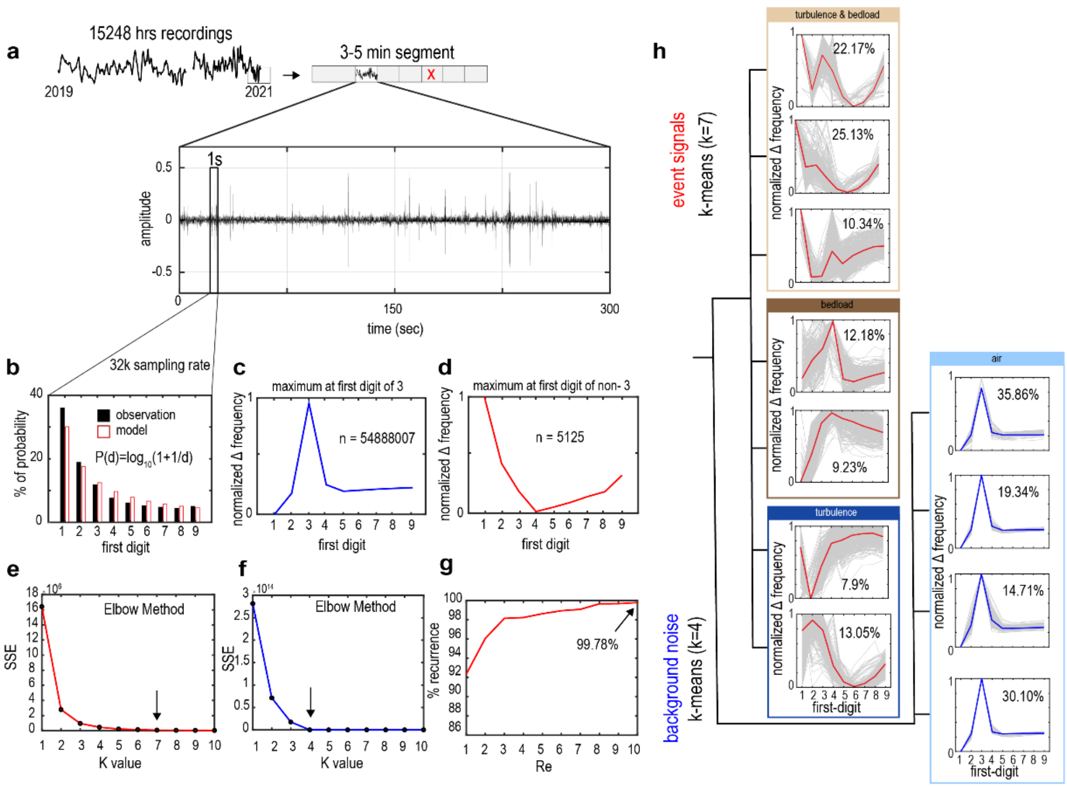


Figure 2. Workflow of the applied Benford's law to sound combinations. (a)

Schematic diagram of the acoustic amplitude along the entire study period. An acoustic data file (*.mp3) is generated for every 3 to 5 minutes of acoustic recordings. (b) A comparison of the probability distribution of Benford's Law model and observation in %, P is the probability, and D is the first-digit. (c) Schematic diagram of the category of normalized probability difference that maximum is not the first-digit with three. (d) The category of normalized probability difference that maximum is the first-digit with three. (e) Determining the k-value (number of clusters) of event noise according to the Elbow method. (f) Determining the k-value of background noise with the Elbow method. (g) Determining the parameter Re (number of times to repeat clustering). (h) Categories of normalized probability difference distribution, classified by the k-means method. Percentages represent proportions in the same group.

3. Results

3.1 Sound classification determined by k-means clustering

Our results from k-means clustering show seven classes for event signals (n=5125) and four classes for background noise (n=54888007). The Elbow method provides the k value to satisfy the statistical objective of minimizing within-cluster error in the k-means method, and it may lead to overfitting, surpassing the requirements for sound identification. For example, background noise can be separated into four classes, but they do not hold physical meaning. We found distinctive characteristics in the residual probability, where specific types of sounds exhibit the same largest residual position. For example, the largest residual value at the first-digit with a three is always an air sound; the largest residual value at the first-digit with a one is mainly the sound of turbulence with sediment impacts, which occurs about 57.6% of the total event signal; the largest residual value at the first-digit with a four is mainly the sound of sediment impacts that are inferred to be bedload transport, occurring at 21.41% of the total event signal. The other two classes accounted for 20.95% in total, mostly the sound of turbulence. Notably, the largest residuals of turbulence are not in the same position. To simplify the acoustic diversity, we merged them according to the location of the largest residual value into four classes of sounds, i.e., bedload motion, turbulence with bedload motion, turbulence, and air (Figure 2h).

3.2 The goodness of fit marks bedload transportation events

From 2019 to 2021, two bedload transport events occurred at Baiyang station. The first event happened on Aug. 24, 2019, with a maximum water level of 3.1 m. The goodness of fit is nearly one during this period, meaning that the first-digit distribution closely follows Bedford's law, and the ratio of event signal increases to 100% (Figure 3a). The second event happened on Oct. 10, 2021, with a maximum water level of 3.6 m. Similarly, the goodness of fit is nearly one during this period, and the ratio of event signal increases to 100% (Figure 3c). In 2020, the water level did not exceed 1.1 m, and bedload transport was negligible (Figure 3b). Apart from these two events, 25 audio files contain event signals, accounting for 28 seconds, 0.54 % of the total event signal. In addition, the mean amplitude difference (75^{th} – 25^{th}) of these 25 audio files is $0.007 \pm 3\times10^{-5}$, and the mean power calculated from the spectrogram is -85.21 ± 6.14 (Table S1). Given low values in duration, acoustic intensities, the goodness of fit, and the ratio of event signal, we ruled out these 25 audio recordings from bedload transport events.

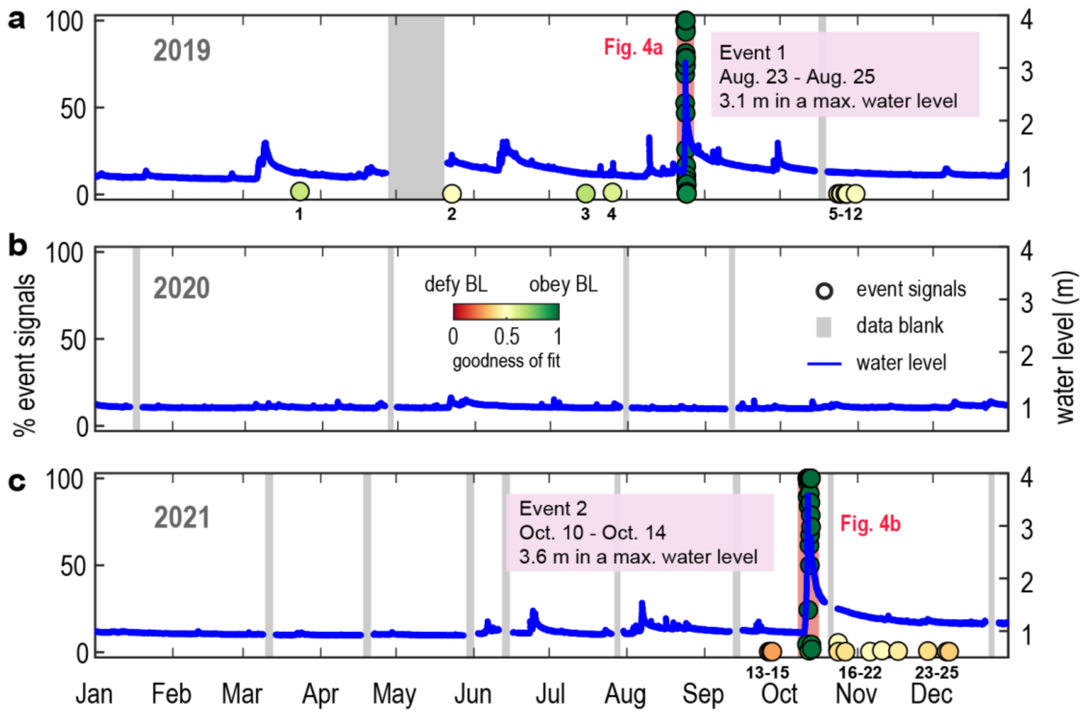


Figure 3. Three-year time series of event signal ratios, the goodness of fit, and river water levels. (a–c) represents the years from 2019 to 2021. Blue lines are water hydrographs, and circles denote event signals in %, colored by the goodness of fit. Numbers beside the circles mark the misidentified 25 audio files.

3.3 Changes in residual probability of the first-digit distribution during the two events

Our examination demonstrates that the hydrophone captures sounds emanating from various physical mediums, including air, water flow, and bedload motion throughout the monitoring period. In the first event, the ratio of bedload motion occurrence increased from 7.3% at 04:50 on Aug. 24, 2019, with a critical stage of 2.2 m to 90.1% after 3 hours, followed by a decrease to 9.9% at 10:50 on Aug. 24, about 6 hours later. Sounds of turbulence with sediment impact start with bedload motion but dominate the source of sound in the early and late stages of the event by over 52% of the five-minute sound contribution. Sounds reflecting sediment impact account for 82.5% of five-minute sound contribution during the peak of bedload motion. Eventually, the bedload motion ends at 21:50 on Aug. 24, while the dominant sound contributor becomes air (background noise) (Figure 4c).

During the second event, the ratio of bedload motion in five-minute sound contribution increased from 1.8% at 18:55 on Oct. 11, 2021, with a critical stage of 1.9 m, to 97.4% at 03:55 on Oct. 12 with a critical stage of 2.7 m. Contrary to the first event, the ratio of bedload motion lasted until 15:55, the end of the event on Oct. 13. At the time of the local low water stage of 2.4 m, bedload motion was halted. Then, the motion was re-activated at a higher water level of 2.5 m with the 1% ratio of bedload motion. Similarly, the occurrence of turbulence together with bedload transport dominates the sound source in the recession limb by over 60%. By 15:55 on Oct. 12, the sound is fully generated by air (Figure 4d). Based on the occurrence and end time of bedload signals, we calculate the duration of the two bedload transport events, yielding 17 and 45 hours, respectively, constituting roughly 0.35% of the time per year, which is equal to 30.7 hours/year.

4. Discussion

4.1 Applications of the acoustic and statistical method

We present an automatic and efficient workflow to identify the onset of bedload transport and reveal the dynamic sound combinations during sediment transport events. We have also proposed recommendations regarding data processing. The distribution of the first-digit in background noise may vary depending on the static voltage of the instrument, e.g., loggers, seismic or acoustic stations, and the type of noise. We propose visualizing short-term audio files and applying Benford's Law to establish a connection between background noise and the distribution of first-digit, which significantly reduced computational expenses.

The residual probability of bedload signals always appears at the location of the first-digit with four in this study, which may vary depending on the monitoring instrument, but can be verified through human listening and acoustic spectrograms.

Therefore, we recommend conducting short-term validations between the residual probability and the sound types. Although k-means clustering offers the advantage of fast computation, we encountered the issue of overfitting. we have merged 11 types of sounds into 4 types based on human listening. We recommend using supervised classification tools for distinguishing different sounds.

4.2 The sound combination determined by residual probability reflects bedload dynamics

Using the residual probability of the first-digit distribution, we classify sounds at a second timescale and accurately determine the timing and critical state for the onset of bedload motion. Sound combinations reflect dynamic flooding events where numerous processes may occur individually or concurrently (e.g., Fig. 4). Moreover, the critical state of the second event is 1.24 times higher than the first event. We infer that following the bedload transport event, the bed morphology was altered, As such, gravels inlaid with each other, forming higher critical shear stress for the onset of bedload motion (Turowski et al., 2011). In addition, the study in Erlenbach torrent shows that small to intermediate past flows contribute to the development of channel stability and high-magnitude flows decrease the critical shear stress (Masteller et al., 2019).

The ratio of bedload sound temporally coincides with the mean of the acoustic power calculated from the spectrogram (Figure S2). The spectrogram at Baiyang station on Aug. 23 to 25, 2019 (Figure 4e) shows that before the onset of the bedload motion (defined by the goodness of fit; Figure 4a), the acoustic power below 100 Hz is about two orders of magnitude higher than in other frequency bands, which can be attributed to the sound of flowing water. When the bedload transport begins, the acoustic power at frequency bands of ~1000 Hz increases by about five orders of magnitude. This increase lasts for about six to seven hours. The October 2021 spectrogram (Figure 4f) exhibits a similar pattern; the acoustic power increases by several orders of magnitude at high frequency. When the ratio of bedload sound decreases, the acoustic power also decreases.

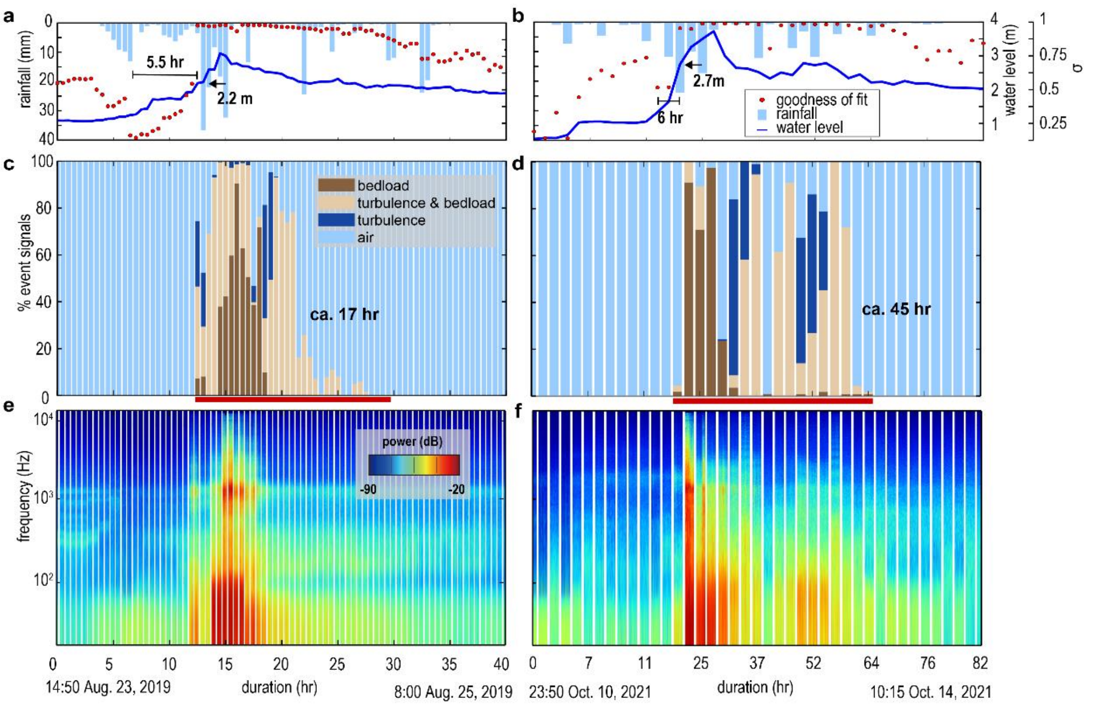


Figure. 4 Sound combinations of the two bedload transportation events. (a–b) Rainfall, water level, and goodness of fit. Periods denote the duration of the decline period in goodness of fit. (c–d) Time series of sound combinations. Colors represent the source of the sound (see legend). (e–f) Semilogarithmic spectrograms of acoustic signals.

4.3 Decreasing goodness of fit at incipient flooding

The goodness of fit not only identifies the onset of bedload transport but also has the potential to recognize changes in hydraulics. We found that decreasing goodness of fit and increasing water level are abrupt at incipient flooding (Figure 4a–4b). In the first event, 5.5 hours before the onset of bedload motion, the goodness of fit decreased from 0.63 to 0.45, and the water level increased conversely from 1.19 to 1.24 m. In the second event, 6 hours before bedload motion, the goodness of fit decreases from 0.79 to 0.63, and the water level increases conversely from 1.5 m to 1.7 m.

We found sound sources with sound durations shorter than one second which we consider as pulse-type sources (Figures S1a–S1b). The pulses may be caused by advancing flooding, where the surging water surface entrains a large number of air bubbles, making the hydrophone susceptible to a mechanical pulse sound. The sound increases amplitude by less than an order of magnitude, prohibiting the full application of Benford's Law and reducing the goodness of fit. Even though such pulse-type sound is defined as background noise in this study, it combines with the change in the goodness of fit, we could grasp this hydrological change. If such an abrupt decrease in the goodness of fit at the rising limb of the hydrograph is consistent throughout various study sites, it may constitute an important feature that can be utilized to improve early warning systems for Earth surface flows, including bedload transport and debris flows.

5. Conclusion

A method that can rapidly and accurately detect the onset of bedload transport in real-time is crucial for disaster warnings and calculating sediment flux. We use the probability change in first-digit distribution from the two bedload transport events to establish a workflow flow of event detection and sound classification. With our workflow, we were able to filter out >99% of the background noise from acoustic recordings and focus on flooding event acoustic signals that can further be separated into three sound classes by statistical clustering tools. We propose a statistical 'goodness of fit' between the theoretical Benford's Law and empirical data and find this parameter to match the onset of bedload motion. Hence, we propose that the operating timing of an expensive monitoring tool, e.g., an automatic river water sampler, can be initiated using this simple parameter.

Given that Benford's Law has demonstrated usefulness in acoustic amplitude analysis, and that Environmental Seismology has been widely used in monitoring fluvial processes (e.g., Burtin et al., 2016; Cook et al., 2021; Dietze et al., 2019, 2022; Walter et al., 2017). Therefore, we suggest that applying environmental seismology in parallel with Benford's Law can be useful in identifying anomalous events in any kind of real-time data series. We used the audio data at a sampling rate of 32 kHz, which is sufficient for Benford's Law calculation. Increasing the time resolution to sub-second resolutions is possible. However, since the common sampling rate of the seismometers is 200 Hz, which covers most environmental processes, reducing the time resolution to the minute scale is necessary to acquire a dataset with an adequate sample size and expected data range. Nonetheless, minute-scale observations are sufficient for early warning of fluvial disasters.

Data Availability Statement

All data and MATLAB code analyzed in this study are available at https://doi.org/10.6084/m9.figshare.24493273.v1.

References

Abancó, C., Hürlimann, M., Fritschi, B., Graf, C., and Moya, J. (2012). Transformation of ground vibration signal for debris flow monitoring and detection in alarm systems. *Sensors*, 12(4), 4870–4891.

Aquarian Audio (2013) Aquarian Audio Products H1a Hydrophone User's Guide: Anacortes, Wash. Aquarian Audio Products. https://www.aquarianaudio.com/ AqAudDocs/ H1a_manual.pdf.

Badoux, A., Andres, N., Turowski, J. M. (2014). Damage costs due to bedload transport processes in Switzerland. Nat.Hazards Earth Syst. Sci., 14, 279–294, https://doi:10.5194/nhess-14-279-2014

Baum, R. L., & Godt, J. W. (2010). Early warning of rainfall-induced shallow landslides and debris flows in the USA. Landslides, 7(3), 259–272. https://doi.org/10.1007/s10346-009-0177-0

Benford, F. (1938). The Law of Anomalous Numbers. Proceedings of the American Philosophical Society, 78(4), 551–572. http://www.jstor.org/stable/984802

Bufe, A., Turowski, J. M., Burbank, D. W., Paola, C., Wickert, A. D., and Tofelde, S. (2019). Controls on the lateral channel-migration rate of braided channel systems in coarse non-cohesive sediment. Earth Surf. Process. Landforms, 44: 2823–2836. https://doi.org/10.1002/esp.4710

Burtin, A., Hovius, N., & Turowski, J. (2016). Seismic monitoring of torrential and fluvial processes. Earth Surface Dynamics, 4(2), 285–307. https://doi.org/10.5194/esurf-4-285-2016

Calmels, D., Galy, A., Hovius, N., Bickle, M. J., West, A. J., Chen, M.-C., Chapman, H. (2011). Contribution of deep groundwater to the weathering budget in a rapidly eroding mountain belt, Taiwan. Earth and Planetary Science Letters, 303 (1–2), 48–58. https://doi.org/10.1016/j.epsl.2010.12.032

Cook, K., Rekapalli, R., Dietze, M., Pilz, M., Cesca, S., Purnachandra, R., et al. (2021). Early warning of catastrophic flow events using regional seismic networks. Science, 374(6563), 87–92. https://doi.org/10.1126/science.abj1227

Dadson, S. J., Hovius, N., Chen, H., Dade, W. B., Hsieh, M.-L., Willett, S. D., Hu, J.-C., Horng, M.-J., Chen, M.-C., Stark, C. P., Lague, D. Lin, J.-C. (2003). Links between erosion, runoff variability and seismicity in the Taiwan orogen. Nature, 426(6967), 648–651. https://doi.org/10.1038/nature02150

Díaz, J., Gallart, J., Ruiz, M. (2014). On the Ability of the Benford's Law to Detect Earthquakes and Discriminate Seismic Signals. *Seismological Research Letters*, 86 (1), 192–201. https://doi.org/10.1785/0220140131

Dietze, M., Lagarde, S., Halfi, E., Laronne, J., & Turowski, J. (2019). Jointsensing of bedload flux and water depth by seismic data inversion. Water Resources

Research, 55(11), 9892–9904. https://doi.org/10.1029/2019WR026072

Dietze, M., Hoffmann, T., Bell, R., Schrott, L., & Hovius, N. (2022). A seismic approach to flood detection and characterization in upland catchments. Geophysical Research Letters, 49, e2022GL100170. https://doi.org/10.1029/2022GL100170

Geay, T., Belleudy, P., Gervaise, C., Habersack, H., Aigner, J., Kreisler, A., Seitz, H., and Laronne, J. B. (2017). Passive acoustic monitoring of bed load discharge in a large gravel bed river. J. Geophys. Res. Earth Surf., 122, 528– 545. https://doi.org/10.1002/2016JF004112

Huang, J.-C., Yu, C.-K., Lee, J.-Y., Cheng, L.-W., Lee, T.-Y., and Kao, S.-J. (2012). Linking typhoon tracks and spatial rainfall patterns for improving flood lead time predictions over a mesoscale mountainous watershed. Water Resour. Res., 48, W09540. https://doi:10.1029/2011WR011508

Hung, C., Lin, G.-W., Kuo, H.-L., Zhang, J.-M., Chen, C.-W., Chen, H.-G. (2018). Impact of an Extreme Typhoon Event on Subsequent Sediment Discharges and Rainfall-Driven Landslides in Affected Mountainous Regions of Taiwan. Geofluids, 2018. https://doi.org/10.1155/2018/8126518

Joannes-Boyau, R., Bodin, T., Scheffers, A. Sambridge, M., May, S. M. (2015). Using Benford's law to investigate Natural Hazard dataset homogeneity. Sci Rep 5, 12046. https://doi.org/10.1038/srep12046

Krapesch, G., Hauer, C., and Habersack, H. (2011). Scale orientated analysis of river width changes due to extreme flood hazards. Nat.Hazards Earth Syst. Sci., 11, 2137–2147. https://doi:10.5194/nhess-11-2137-2011

Li, Q. Fu, Z., Yuan. N. (2015). Beyond Benford's Law: Distinguishing Noise from Chaos. PLoS ONE, 10(6), e0129161. https://doi:10.1371/journal.pone.0129161

Masteller, C. C., Finnegan, N. J., Turowski, J. M., Yager, E. M., & Rickenmann, D. (2019). History-dependent threshold for motion revealed by continuous bedload transport measurements in a steep mountain stream. Geophysical Research Letters, 46, 2583–2591. https://doi.org/10.1029/2018GL081325

Marra, F., Nikolopoulos, E. I., Creutin, J. D., & Borga, M. (2016). Space–time organization of debris flows-triggering rainfall and its effect on the identification of the rainfall threshold relationship. Journal of Hydrology, 541, 246–255. https://doi.org/10.1016/j.jhydrol.2015.10.010

Nigrini, M. (1999). I've got your number: How a mathematical phenomenon can help CPAS uncover fraud and other irregularities, Journal of Accountancy, 1-7.

Nigrini, M. J., Miller, S. J. (2007). Benford's Law Applied to Hydrology Data—Results and Relevance to Other Geophysical Data. Math Geol., 39, 469–490. https://doi.org/10.1007/s11004-007-9109-5

Petley D. N., Liu C-N., Liou Y-S. (1997). Geohazards in a Neotectonic Terrain, Taroko Gorge, eastern Taiwan. Memoir of the Geological Society of China, 40, 135–154.

Recking, A., Piton, G., Vazquez-Tarrio, D., and Parker, G. (2016). Quantifying the Morphological Print of Bedload Transport. Earth Surf. Process. Landforms, 41, 809– 822. https://doi.org/10.1002/esp.3869.

Roth, D. L., Finnegan, N. J., Brodsky, E. E., Rickenmann, D., Turowski, J., Badoux, A., Gimbert, F. (2017). Bed load transport and boundary roughness changes as competing causes of hysteresis in the relationship between river discharge and seismic amplitude recorded near a steep mountain stream. Journal of Geophysical Research, 122, 5, 1182—1200. http://doi.org/10.1002/2016JF004062.

Sambridge, M., Tkalčić, H., and Jackson, A. (2010). Benford's law in the natural sciences. Geophys. Res. Lett., 37, L22301. http://doi:10.1029/2010GL044830

Snyder, N. P., Castele, M. R., Wright, J. R. (2009). Bedload entrainment in low-gradient paraglacial coastal rivers of Maine, U.S.A. Implications for habitat restoration. Geomorphology, 103(3). https://doi.org/10.1016/j.geomorph.2008.07.013

Stott, T., Leeks, G., Marks, S., Sawyer, A. (2001). Environmentally sensitive plot-scale timber harvesting: impacts on suspended sediment, bedload and bank erosion dynamics. Journal of Environmental Management, 63(1). https://doi.org/10.1006/jema.2001.0459

Theule, J. I., Liébault, F., Loye, A., Laigle, D., and Jaboyedoff, M. (2012). Sediment budget monitoring of debris-flow and bedload transport in the Manival Torrent, SE France. Nat. Hazards Earth Syst. Sci., 12, 731–749. https://doi.org/10.5194/nhess-12-731-2012

Thorndike, R. L. (1953). Who belongs in the family?. Psychometrika, 18, 267–276. https://doi.org/10.1007/BF02289263

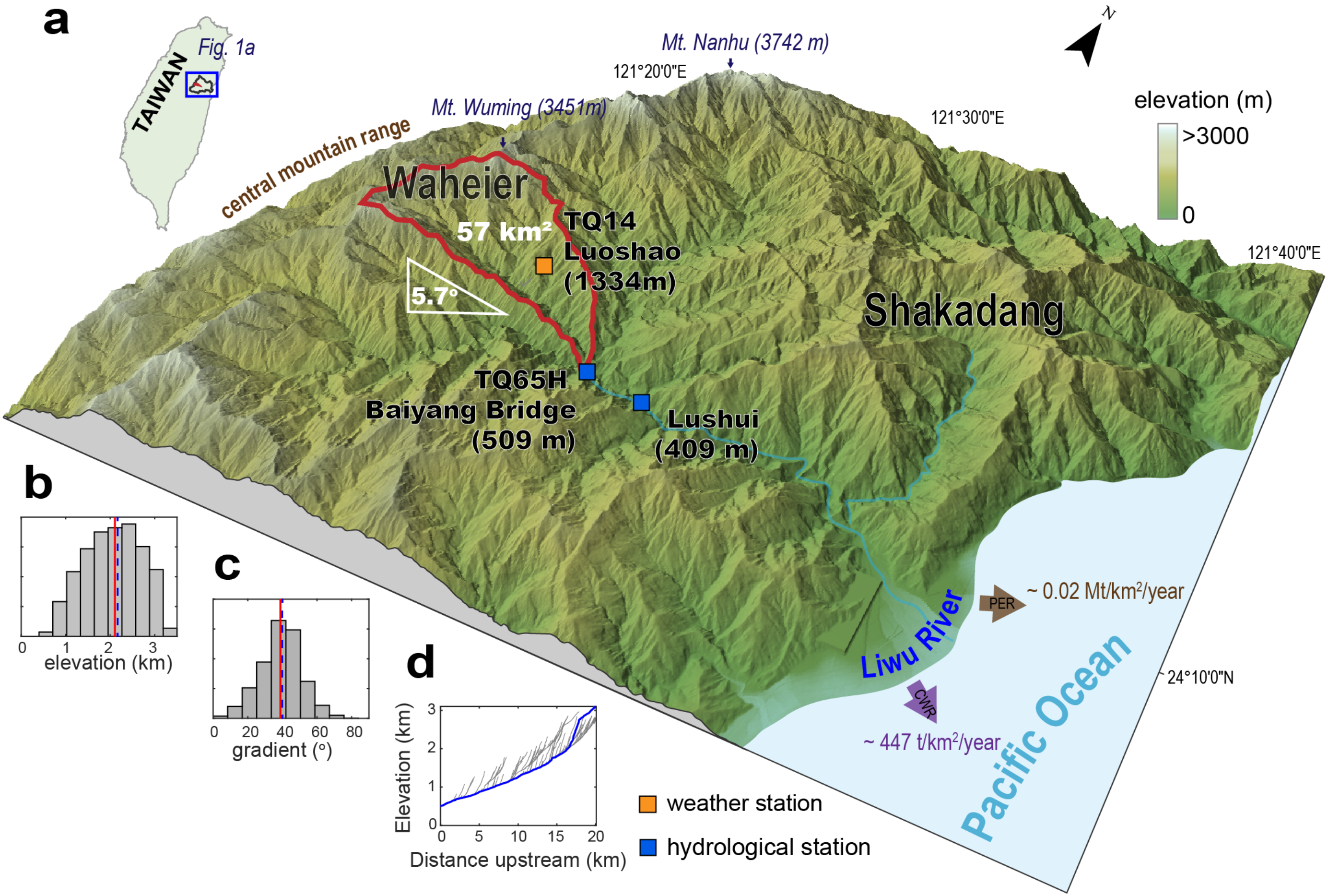
Turowski, J.M., Badoux, A., Rickenmann, D. (2011), Start and end of bedload transport in gravel-bed streams. Geophysical Research Letters, 38. https:// doi:10.1029/2010GL046558

Turowski, J.M., Hovius, N., Meng-Long, H., Lague, D. and Men-Chiang, C. (2008), distribution of erosion across bedrock channels. Earth Surf. Process. Landforms, 33: 353-363. https://doi.org/10.1002/esp.1559

Turowski, J.M., Yager, E.M., Badoux, A., Rickenmann, D. and Molnar, P. (2009). The impact of exceptional events on erosion, bedload transport and channel stability in a step-pool channel. Earth Surf. Process. Landforms, 34: 1661–1673. https://doi.org/10.1002/esp.1855

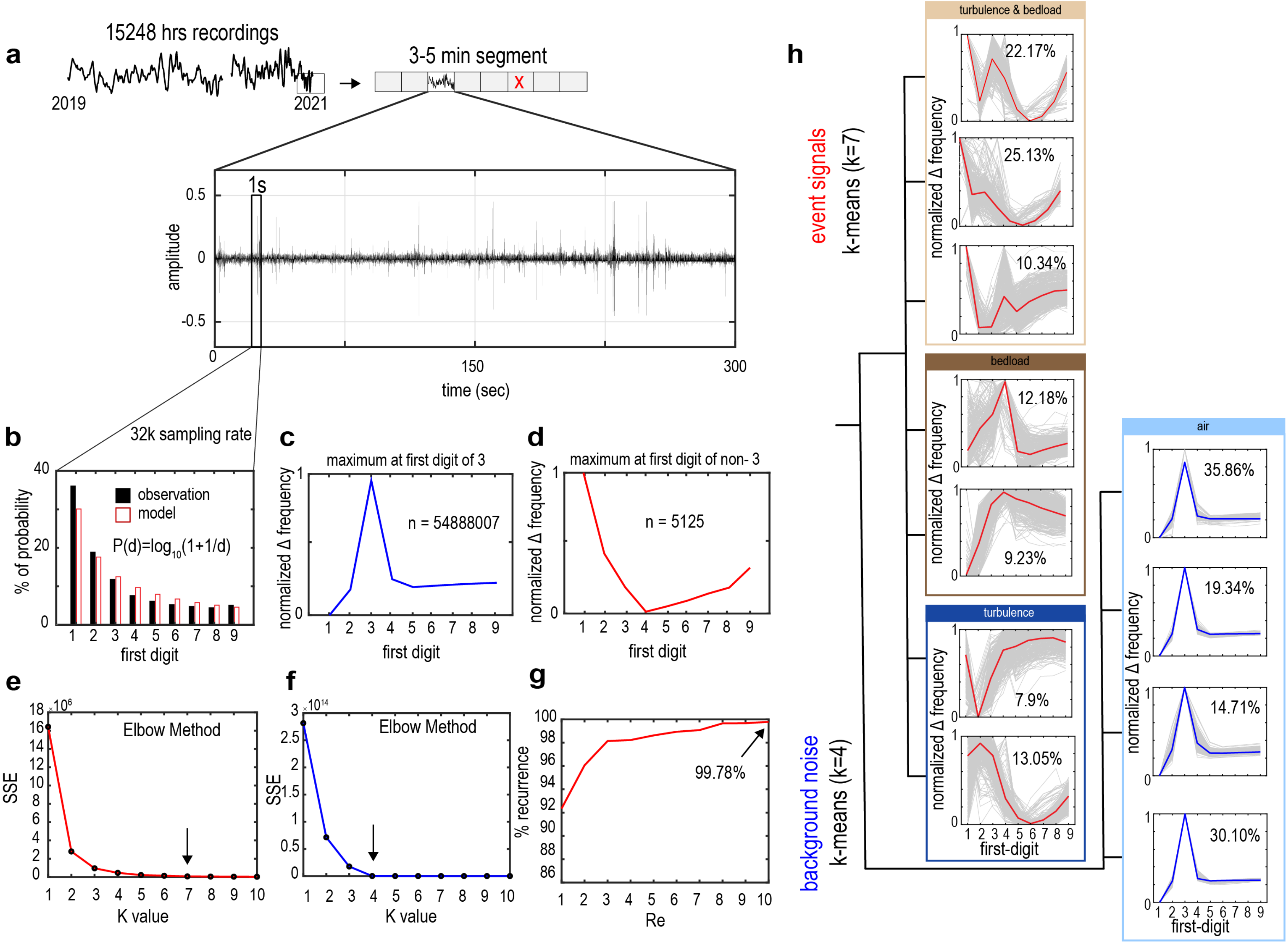
Walter, F., Burtin, A., McArdell, B., Hovius, N., Weder, B., & Turowski, J. (2017). Testing seismic amplitude source location for fast debris-flow detection at Illgraben, Switzerland. Natural Hazards and Earth System Sciences, 17(6), 939–955. https://doi.org/10.5194/nhess-17-939-2017

Figure1.



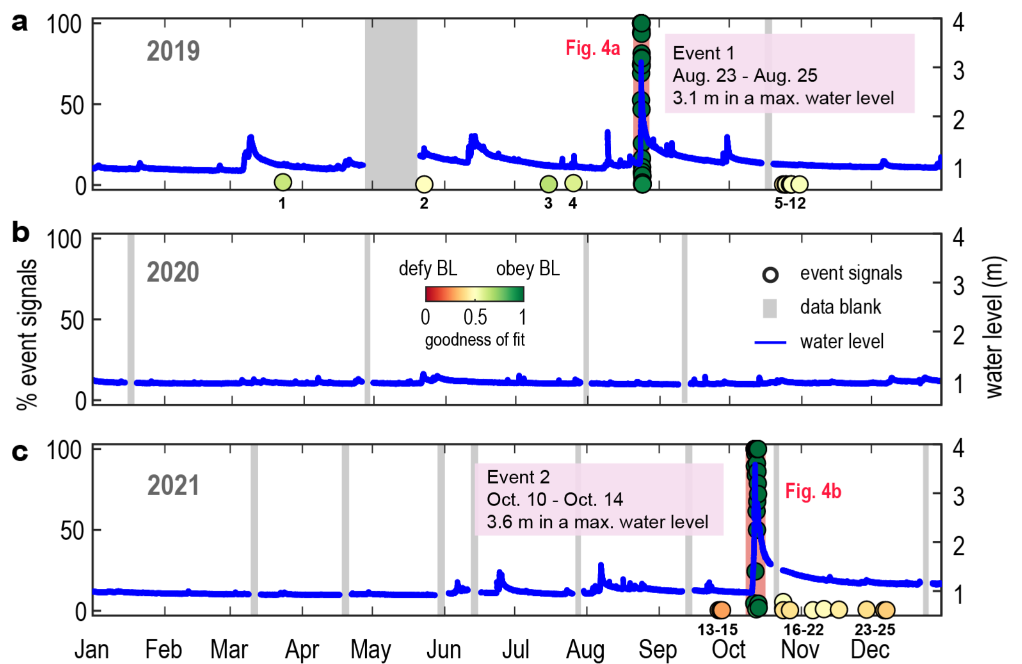
a
TAIWAN
Fig. 1a
Mt. Nanhu (3742 m)
121°20'0"E
Mt. Wuming (3451m)
121°30'0"E
elevation (m)
>3000
0
central mountain range
Waheier
57 km²
TQ14
Luoshao
(1334m)
5.7°
121°40'0"E
Shakadang
TQ65H
Baiyang Bridge
(509 m)
Lushui
(409 m)
b
elevation (km)
c
gradient (°)
PER
~ 0.02 Mt/km²/year
Liwu River
24°10'0"N
CWR
~ 447 t/km²/year
Pacific Ocean
d
Elevation (km)
Distance upstream (km)
weather station
hydrological station

Figure2.



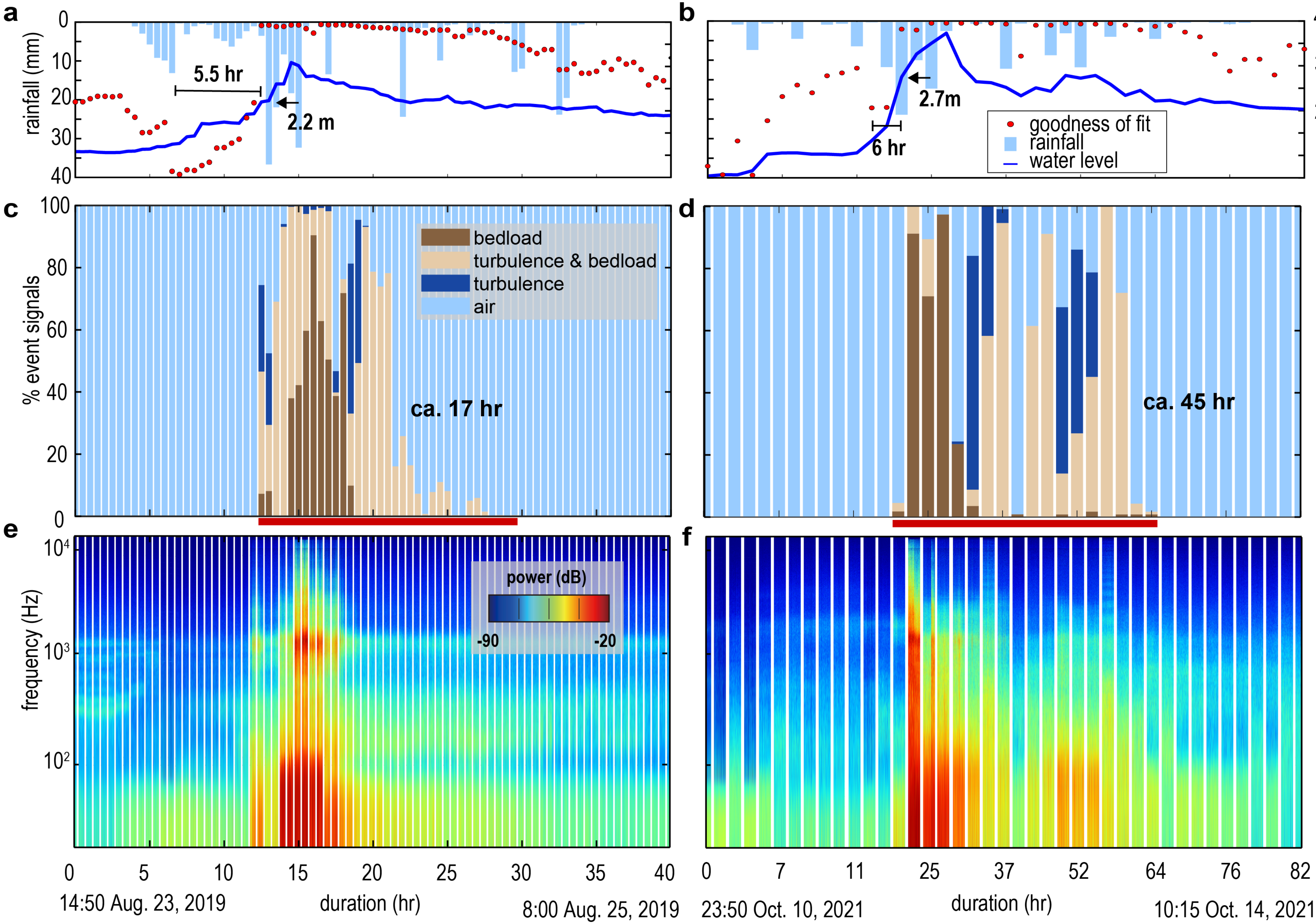
a
15248 hrs recordings
2019
2021
3-5 min segment
1s
amplitude
time (sec)
b
32k sampling rate
% of probability
observation
model
P(d)=log10(1+1/d)
first digit
c
maximum at first digit of 3
normalized Δ frequency
n = 54888007
first digit
d
maximum at first digit of non- 3
normalized Δ frequency
n = 5125
first digit
e
Elbow Method
SSE
K value
f
Elbow Method
SSE
K value
g
% recurrence
99.78%
Re
h
event signals
k-means (k=7)
turbulence & bedload
22.17%
25.13%
10.34%
bedload
12.18%
9.23%
turbulence
7.9%
13.05%
normalized Δ frequency
first-digit
background noise
k-means (k=4)
air
35.86%
19.34%
14.71%
30.10%
first-digit

Figure3.



a
b
c
2019
2020
2021
Fig. 4a
Fig. 4b
Event 1
Aug. 23 - Aug. 25
3.1 m in a max. water level
Event 2
Oct. 10 - Oct. 14
3.6 m in a max. water level
1
2
3
4
5-12
13-15
16-22
23-25
defy BL
obey BL
0
0.5
1
goodness of fit
event signals
data blank
water level
% event signals
water level (m)
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

Figure4.



a
b
c
d
e
f
rainfall (mm)
0
10
20
30
40
5.5 hr
2.2 m
2.7m
6 hr
goodness of fit
rainfall
water level
water level (m)
σ
1
0.75
0.5
0.25
4
3
2
1
% event signals
100
80
60
40
20
0
bedload
turbulence & bedload
turbulence
air
ca. 17 hr
ca. 45 hr
frequency (Hz)
10^4
10^3
10^2
power (dB)
-90
-20
0
5
10
15
20
25
30
35
40
14:50 Aug. 23, 2019
duration (hr)
8:00 Aug. 25, 2019
0
7
11
25
37
52
64
76
82
23:50 Oct. 10, 2021
duration (hr)
10:15 Oct. 14, 2021