**Measuring bedload motion time at sub-second using scale Benford's law from long-term acoustic recordings**

Ci-Jian Yang1\*, Jens M. Turowski2, Qi Zhou2,3, Ron Nativ2,3,4, Hui Tang2, Wen-Sheng Chen5

1Department of Geography, National Taiwan University, Taipei, Taiwan.

2Helmholtzzentrum Potsdam, GFZ German Research Center for Geosciences, Potsdam, Germany.

3Institute of Geosciences, University of Potsdam, Potsdam, Germany.

4Department of Earth and Environmental Sciences, Ben-Gurion University of the Negev, Be'er Sheva, Israel.

5Center for General Education, National Dong Hwa University, Hualien, Taiwan.

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

**Contents of this file**

Table S1

Figures S1 to S2

**Detail of sound classification via residual probability distribution**

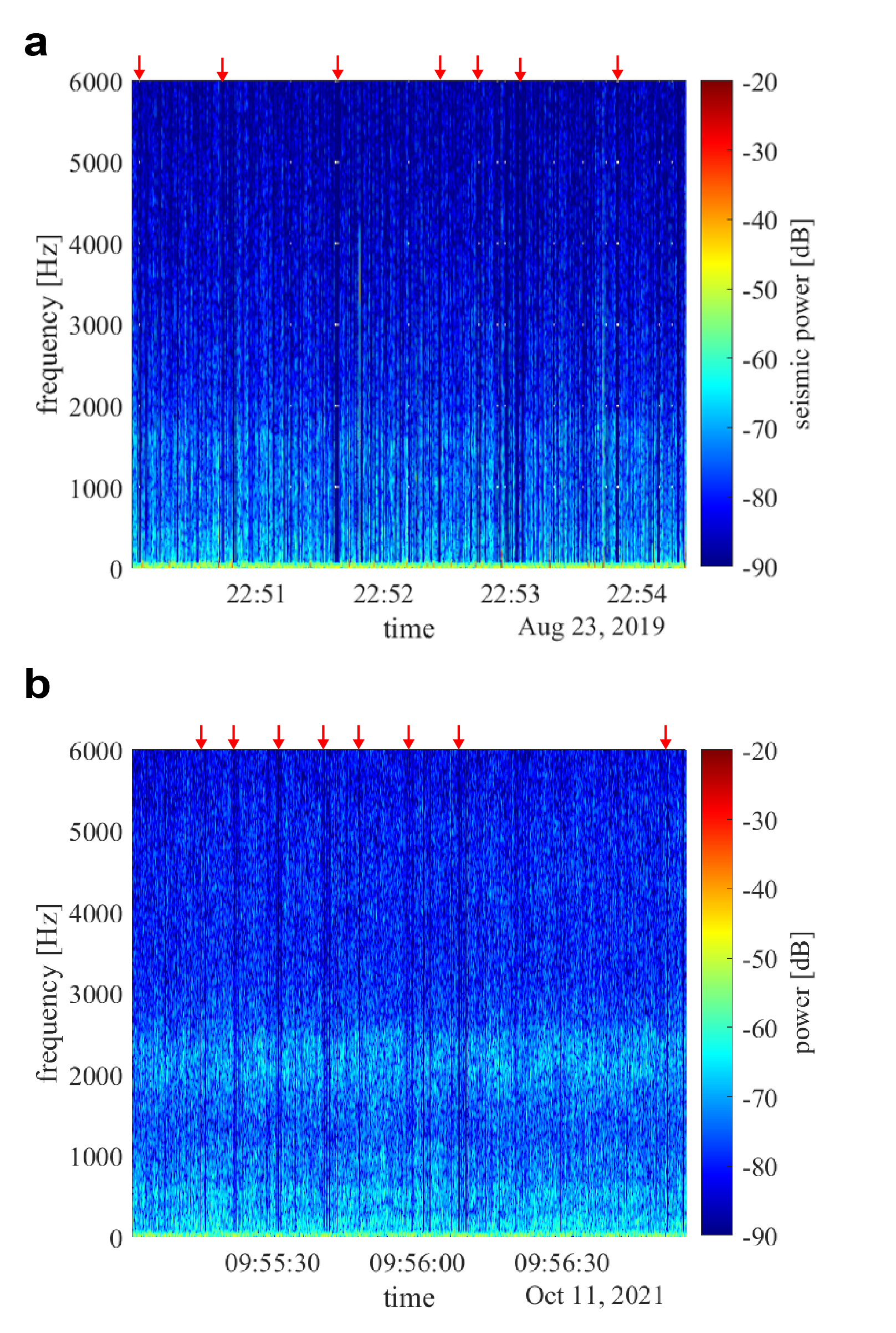
We calculate the first-digit probability of acoustic data every second and calculate the residuals of observed data with respect to Benford's Law (Eq. 1). We thus obtain a curve with nine residual values, each representing first-digit from one to nine, and normalize each residual probability to make the curve between zero and one. Based on a pre-test, the vast majority of residual probability curves occur under the condition where the maximum value is in the first-digit of three. This reflects the static voltage state of the instrument itself, which we thus define as the background noise (Figure 2c). The remaining types that we defined as anomalous (event) signals (Figure 2d).

Given the advantages of fast computation and not requiring training samples in handling vector data, we selected the method of k-means clustering, which can partition observations. For example, we can partition a vector into a given number of clusters (k) in which each observation belongs to the cluster with the minimum of the within-cluster sum of squares. To determine k values for each group, we used the Elbow method (Thorndike, 1953), which determines a turning point in curves of sum squared error (SSE) within the cluster as the cutoff point. Eventually, we gave k values of 7 and 4 to event signals and background noises, respectively (Figures 2e and 2f). Then, we designed a method to test the clustering stability. We used the entire dataset and iterated the number of times to repeat clustering (Re). Each specific condition (Re = 1, 2, 3, … 10) was repeated 101 times, and the recurrence rate is based on the first clustering of each data point. For example, the first clustering of a data point is class four, then we count the number of class four in the following 100 classifications. The test result shows that 99.78% of data points are classified into the same category as the first result (Figure 2g).

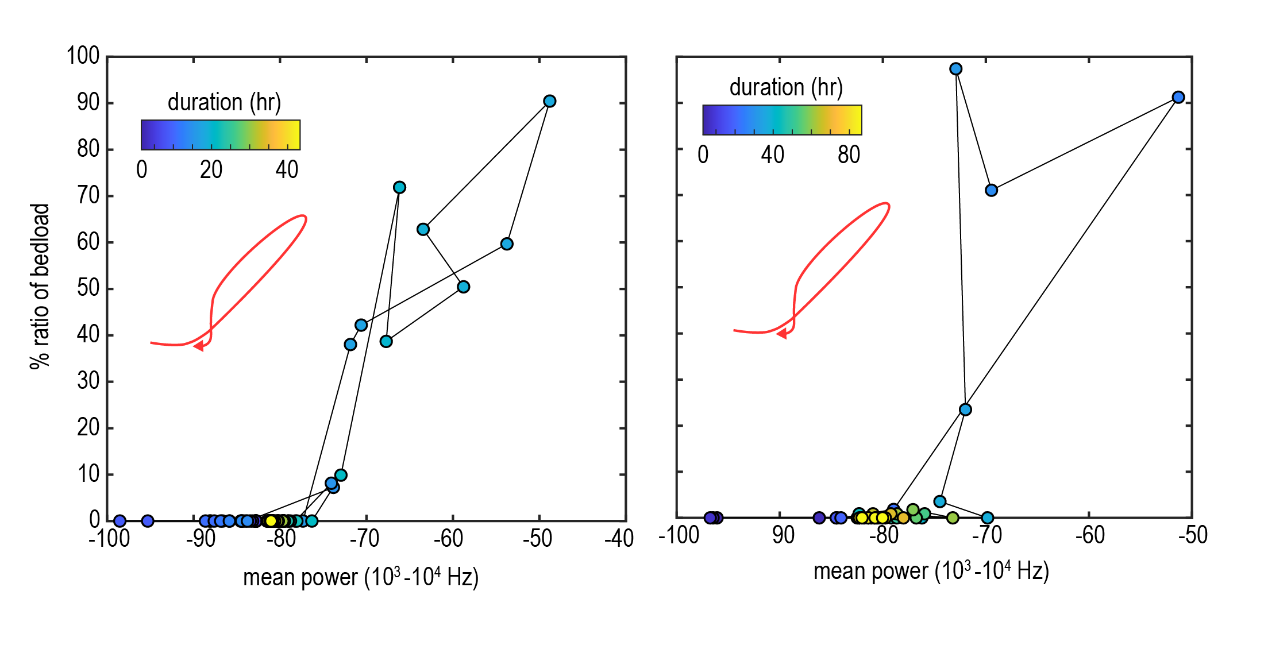
Then, we obtained the ratio of event signal (%) by calculating the duration of the event signal every five minutes. We further calculated sound combinations (%) at the same time resolution, which is determined by the proportion of time occupied by each type of sound, including event signals from specific classes and background noise. The time resolution of five minutes is determined based on the measurement resolution of the water level at Baiyang station. The definition of the duration of bedload transport event is the time difference between the starting point and ending point where the ratio of bedload signals exceeds 1%. If the time gap is less than 1 hour, it is considered the same event.

**Table S1** acoustic information of 25 misidentified audio files

|  |  |  |
| --- | --- | --- |
| ID | Amplitude difference (75th-25th) | Mean of power between 103–104 Hz (dB) |
| 1 | 0.007 | -85.213 |
| 2 | 0.007 | -86.485 |
| 3 | 0.007 | -84.345 |
| 4 | 0.007 | -85.617 |
| 5 | 0.007 | -96.441 |
| 6 | 0.007 | -93.040 |
| 7 | 0.007 | -92.932 |
| 8 | 0.007 | -97.371 |
| 9 | 0.007 | -97.045 |
| 10 | 0.007 | -97.903 |
| 11 | 0.007 | -99.445 |
| 12 | 0.007 | -93.654 |
| 13 | 0.007 | -102.728 |
| 14 | 0.007 | -102.587 |
| 15 | 0.007 | -106.093 |
| 16 | 0.007 | -88.380 |
| 17 | 0.007 | -88.937 |
| 18 | 0.007 | -88.297 |
| 19 | 0.007 | -88.081 |
| 20 | 0.007 | -86.010 |
| 21 | 0.007 | -88.766 |
| 22 | 0.007 | -88.022 |
| 23 | 0.007 | -88.967 |
| 24 | 0.007 | -89.891 |
| 25 | 0.007 | -89.013 |



**Figure S1.** Sectrograms of acoustic signals for (a) on 23 Aug. 2019 5.5 hours before the bedload motion, (b) on 10 Oct. 2021 6 hours before the bedload motion. The red arrows represent the positions of partial pulse-like signals.



**Figure S2.** Correlation of ratio of bedload signal in % with mean of acoustic power between 103–104 Hz, colored by time in hours. Red arrows indicate the direction of change.