



1 **Toward a general calibration of the Swiss plate geophone system for** 2 **fractional bedload transport**

3

4 Tobias Nicollier^{1,2}, Gilles Antoniazza^{3,1}, Lorenz Ammann¹, Dieter Rickenmann¹, James W. Kirchner^{1,2,4}

5 ¹Swiss Federal Research Institute WSL, Birmensdorf, 8903, Switzerland

6 ²Department of Environmental System Sciences, ETH Zürich, Zürich, 8092, Switzerland

7 ³Institute of Earth Surface Dynamics (IDYST), University of Lausanne, Lausanne, 1015, Switzerland

8 ⁴Department of Earth and Planetary Science, University of California, Berkeley, 94720, USA

9

10 *Correspondence to:* Tobias Nicollier, Swiss Federal Research Institute (WSL), Mountain Hydrology and Mass Movements,
11 8903 Birmensdorf, Switzerland. E-mail: tobias.nicollier@wsl.ch. Phone: +41 77 437 35 77

12

13

14 **Abstract.** Substantial uncertainties in bedload transport predictions in steep streams have triggered intensive efforts to
15 develop surrogate monitoring technologies. One such system, the Swiss plate geophone (SPG), has been deployed and
16 calibrated in numerous steep water courses, mainly in the Alps. Calibration relationships linking the signal recorded by the
17 SPG system to the transported bedload can vary substantially between different monitoring stations, likely due to site-
18 specific factors such as the flow velocity and the bed roughness. Furthermore, recent controlled experiments have shown that
19 site-specific calibration relationships can be biased by elastic waves resulting from impacts occurring outside the plate
20 boundaries. Motivated by these findings, here we present a hybrid calibration procedure derived from flume experiments and
21 an extensive dataset of 308 calibration measurements from four different field monitoring stations. Our main goal is to
22 investigate the feasibility of a general, site-independent calibration procedure for inferring fractional bedload transport from
23 the SPG signal. First, we use flume experiments to show that sediment size classes can be distinguished more accurately
24 using a combination of vibrational frequency and amplitude information than by using amplitude information alone. Second,
25 we apply this amplitude-frequency method to field measurements to derive general calibration coefficients for ten different
26 grain-size fractions. The amplitude-frequency method results in more homogeneous signal responses across all sites and
27 significantly improves the accuracy of fractional sediment flux and grain-size estimates. We attribute the remaining site-to-
28 site discrepancies to large differences in flow velocity, and discuss further factors that may influence the accuracy of these
29 bedload estimates.

30 **1 Introduction**

31 Flood events across Europe in the summer of 2021 have illustrated the threat of flood-related hazards like bedload transport
32 to human life and infrastructure, especially in small and steep mountainous catchments (Badoux et al., 2014; Blöschl et al.,
33 2020). Understanding sediment transport processes is also essential for efforts to return rivers to their near-natural state by
34 restoring their continuity and re-establishing balanced sediment budgets (e.g. Brouwer and Sheremet, 2017; Pauli et al.,
35 2018; Logar et al., 2019; Rachely et al., 2021). However, monitoring and predicting bedload transport still represents a
36 considerable challenge because of its large spatio-temporal variability (e.g. Mühlhofer, 1933; Einstein, 1937; Reid et al.,
37 1985; Rickenmann, 2018; Ancey, 2020). This is especially true for steep streams, because they are poorly described by
38 traditional bedload transport equations, which have mainly been developed for lower-gradient channels (e.g. Schneider et al.,



39 2016). Predicting sediment transport in steep channels is challenging, notably due to the presence of macro-roughness
40 elements affecting both the flow resistance and the flow energy (e.g. Manga and Kirchner, 2000; Yager et al., 2007, 2012;
41 Bathurst, 2007; Nitsche et al., 2011; Rickenmann and Recking, 2011; Prancevic and Lamb, 2015). It is further complicated
42 by a sediment supply that varies in both space and time, due in part to cycles of building and breaking of an armoring layer
43 at the riverbed (e.g. Church et al., 1998; Dhont and Ancey, 2018; Rickenmann, 2020; Piantini et al., 2021).

44 Bedload transport equations established for lower-gradient streams typically result in substantial errors when applied to
45 steep streams, motivating the development of new indirect monitoring techniques for steep mountain channels (e.g. Gray et
46 al., 2010; Rickenmann, 2017). Indirect monitoring techniques provide large spatial coverage of river transects at high
47 temporal resolution, reduce personal risk related to in-stream sampling, and enable consistent data collection at widely
48 varying flow conditions including floods (e.g. Gray et al., 2010; Rickenmann, 2017; Geay et al., 2020; Bakker et al., 2020;
49 Choi et al., 2020; Le Guern et al., 2021). The drawback of these monitoring technologies is that in order to provide
50 quantitative measurements, they require intensive calibration through direct bedload sampling with retention basins
51 (Rickenmann and McArdeell, 2008), slot samplers (e.g. Habersack et al., 2017; Halfi et al., 2020) or mobile bag samplers
52 (e.g. Bunte et al., 2004; Dell'Agnese et al., 2014; Hildale et al., 2015; Mao et al., 2016; Kreisler et al., 2017; Nicollier et al.,
53 2021a).

54 Among indirect monitoring techniques, the Swiss plate geophone (SPG) system has been deployed and tested in more
55 than 20 steep gravel-bed streams and rivers, mostly in the European Alps (Rickenmann, 2017). Typically, linear or power-
56 law calibration relationships have been developed between measured signal properties and bedload transport characteristics
57 (Rickenmann et al., 2014; Wyss et al. 2016a; Kreisler et al., 2017; Kuhnle et al., 2017). Such calibration equations facilitate
58 spatio-temporal estimates of bedload fluxes, absolute estimates of bedload fluxes and bedload grain-size distributions, and
59 the detection of the start and end of bedload transport. However, these equations have required calibration against
60 independent bedload transport measurements from each individual field site, because until now we have lacked generally
61 applicable signal-to-bedload calibration equations that are valid in multiple field settings. Although the similarities between
62 calibration relationships at various field sites are encouraging, it is not well understood why the linear calibration coefficients
63 for total mass flux can vary by about a factor of 20 among individual samples from different sites, or by about a factor of six
64 among the mean values from different sites (Rickenmann et al., 2014; Rickenmann and Fritschi, 2017). Given the substantial
65 field effort required for calibration campaigns, a generally applicable calibration equation would represent a significant
66 advance.

67 Numerous studies have reported successful calibration of impact plate systems in laboratory flumes (e.g. Bogen and
68 Møen, 2003; Krein et al., 2008; Tsakiris et al., 2014; Mao et al., 2016; Wyss et al., 2016b,c; Kuhnle et al., 2017; Chen et al.,
69 2021), although transferring these flume-based calibrations to the field remains challenging. Nonetheless, controlled flume
70 experiments are valuable because they allow us to systematically explore relationships between the recorded signal, the
71 transport rates of different sediment size fractions, and the hydraulic conditions. For example, the experiments of Wyss et al.
72 (2016b) showed that higher flow velocities induce a weaker SPG signal response per unit of transported sediment. More
73 recent controlled experiments have highlighted another important site-dependent factor influencing the SPG signal response,
74 namely the grain-size distribution (GSD) of the transported bedload (Nicollier et al., 2021a), where coarser grain mixtures
75 were shown to yield a stronger signal response per unit bedload weight.

76 Subsequent impact tests and flume experiments showed that this grain-size dependence arises because the impacts
77 plates are insufficiently isolated from their surroundings (Antoniazza et al., 2020; Nicollier et al., 2021b). The elastic wave
78 generated by an impact on or near a plate was found to propagate over several plate lengths, contaminating the signals
79 recorded by neighboring sensors within a multiple plate array. Nicollier et al. (2021b) introduced the notion of “apparent
80 packets” to define the portions of the recorded signal that were generated by such extraneous particle impacts.

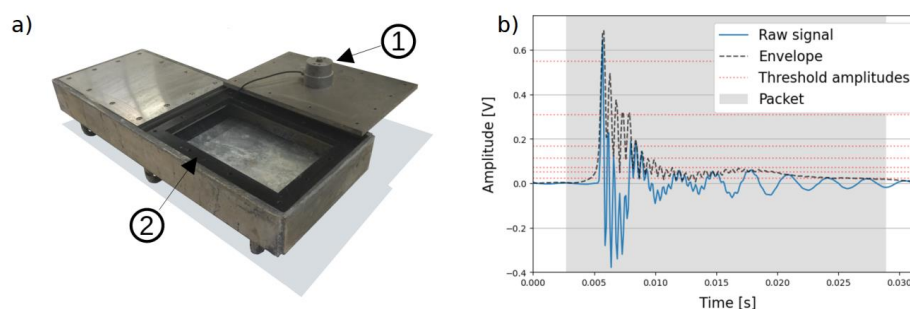


81 The main goal of this contribution is to examine the feasibility of a general, site-independent signal conversion
82 procedure for fractional bedload flux estimates. We follow a comprehensive hybrid signal conversion approach that
83 encompasses a set of full-scale controlled flume experiments conducted at an outdoor flume facility, as well as 308 field
84 calibration measurements performed with direct sampling methods at four different bedload monitoring stations in
85 Switzerland between 2009 and 2020. We present the amplitude-frequency (AF) method, aiming to reduce the bias introduced
86 by apparent packets in the relationship between the signal characteristics and the particle size. Finally, we compare the
87 performance of this novel AF method against the purely amplitude-histogram (AH) method developed by Wyss et al.
88 (2016a) for fractional and total bedload flux estimates as well as for characteristic grain-size estimates.

89 2 Methods

90 2.1 The SPG system

91 The Swiss plate geophone (SPG) consists of a geophone sensor fixed under a steel plate of standard dimensions 492 mm x
92 358 mm x 15 mm (Fig. 1a; Rickenmann, 2017). The geophone (GS-20DX by Geospace technologies; www.geospace.com)
93 uses a magnet moving inside an inertial coil (floating on springs) as an inductive element. The voltage induced by the
94 moving magnet is directly proportional to its vertical velocity resulting from particle impacts on the plate. Typically, a SPG
95 array includes several plates next to each other, acoustically isolated by elastomer elements and covering the river cross-
96 section. The array is either embedded in a concrete sill or fixed at the downstream face of a check dam. A detailed
97 description of the SPG system can be found in Rickenmann et al. (2014). Due to data storage limitations, field stations
98 usually do not continuously record the full raw 10 kHz geophone signal. Instead, it is typically preprocessed, and summary
99 values, such as the maximum amplitude and the number of impulses, are recorded at one-minute intervals. However, for the
100 relatively short duration of a single calibration measurement, ranging from a few seconds to one hour, the full raw signal is
101 recorded (Fig. 1b).



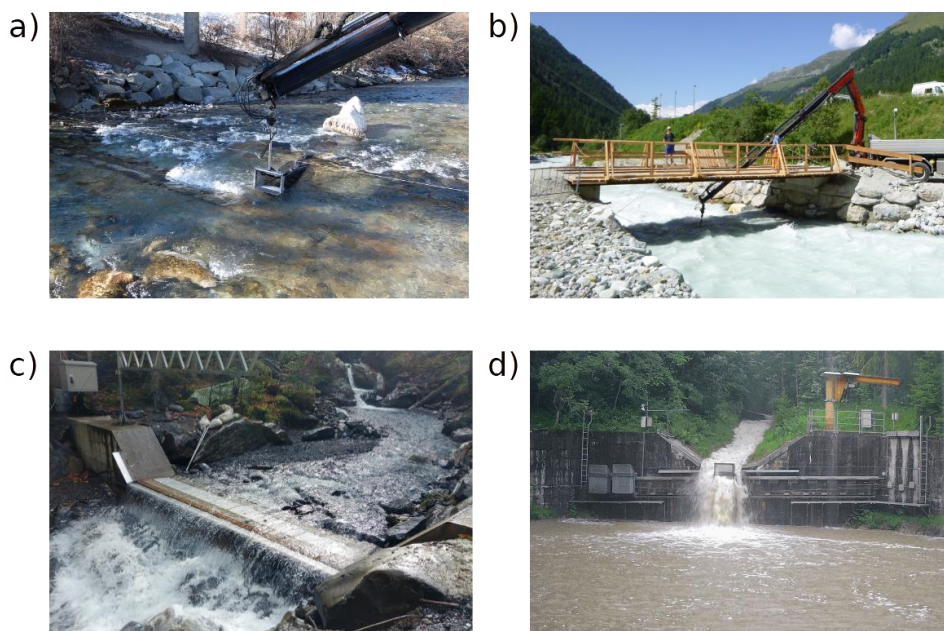
102
103 **Figure 1:** (a) Swiss plate geophone (SPG) system before installation (see Fig. 3). Each plate is equipped with a uniaxial geophone
104 sensor fixed in a watertight aluminum box (1) attached to the underside of the plate. The plates are acoustically isolated from each
105 other by elastomer elements (2). (b) Example of a packet (grey area) detected by the SPG system. A packet begins 20 time steps
106 (i.e., 2 ms) before the signal envelope crosses the lowest amplitude threshold of 0.0216 V and ends 20 time steps after the last
107 crossing of the lowest amplitude threshold (see Sect. 2.4).

108



109 **2.2 Field calibration measurements**

110 To test the AF and AH methods, this study uses 308 calibration measurements from four Swiss bedload monitoring stations
111 equipped with SPG systems (Fig. 2; Table 1). Field calibration samples were collected at the Albula, Navisence and
112 Avançon de Nant stations, and extensive calibration efforts have been undertaken at the fourth field station, Erlenbach, since
113 2009 (Rickenmann et al., 2012). The Erlenbach offers an interesting comparison with the other sites due to different channel
114 and flow characteristics upstream of the SPG plates. Field calibrations at each of the four sites consist of the following steps:
115 (i) direct bedload sampling downstream of an impact plate using either crane-mounted net samplers adapted from Bunte
116 traps (Bunte et al., 2004; Dell’Agnese et al., 2014; Nicollier et al., 2019; Fig. 2a, b), automated basket samplers
117 (Rickenmann et al., 2012; Fig. 2d) or manual basket samplers (Fig. 2c), (ii) synchronous recording of the raw geophone
118 signal, (iii) sieving and weighing of bedload samples using ten sieve classes (Table 3), and (iv) comparing the fractional
119 bedload mass of each sample to the packet histogram data to derive the corresponding calibration coefficient $k_{b,i,j}$. A more
120 detailed description of the sampling procedure is reported in Supporting Information S1.



121

122 **Figure 2: The four Swiss bedload monitoring stations at which raw Swiss plate geophone signals have been recorded during**
123 **calibration measurements. The stations are installed at the following streams: a) Albula, b) Navisence, c) Avançon de Nant and d)**
124 **Erlenbach. Pictures a) and c) were taken during low-flow conditions. Pictures b) and d) show calibration measurements with the**
125 **crane-mounted net sampler and the automated basket sampler, respectively, at high flows.**

126

127

128



129 **Table 1: Channel and flow characteristics based on *in situ* measurements during the calibration campaigns at the four field sites.**
130 **The year of the field calibration campaigns, the sampling technique and the number of collected samples are also indicated.**

Field site	Location (canton)	Bed slope [%] ^a	Mean flow velocity V_f [m s^{-1}] ^b	No. of plates	Year	Sampling technique	No. of samples
Albula ^c	Tiefencastel (Grisons)	0.7	2.6	30	2018	crane-mounted net sampler	51
Navisence ^c	Zinal (Valais)	3	3.2	12	2019	crane-mounted net sampler	80
Avançon de Nant ^d	Les Plans-sur-Bex (Vaud)	4	1.3	10	2019/2020	manual basket sampler	55
Erlenbach ^e	Alpthal (Schwyz)	16	5.0	2	Since 2009	automatic basket sampler	122

131 ^a Gradient measured upstream of the SPG plates. At the Erlenbach, this gradient is the slope of the artificial approach flow channel
132 upstream of the SPG system.

133 ^b Depth-averaged mean flow velocities measured during the calibration measurements.

134 ^c More information on the sites is available in Nicollier et al. (2021a).

135 ^d More information on the site is available in Antoniazza et al. (2021).

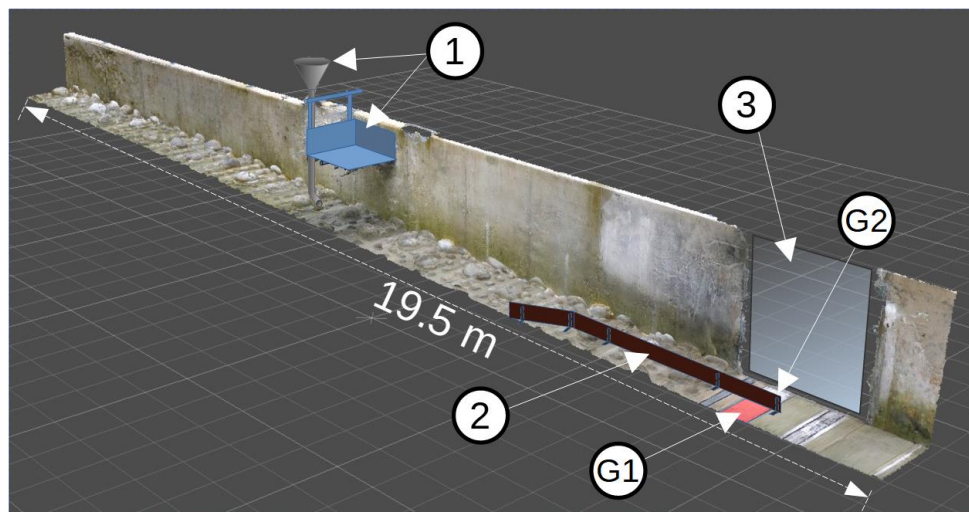
136 ^e More information on the site is available in e.g. Rickenmann et al. (2012), Wyss et al. (2016c), Rickenmann et al. (2018).

137 2.3 Controlled flume experiments

138 The first part of the signal conversion procedure described in this study is based on controlled flume experiments conducted
139 at the outdoor flume facility of the Oskar von Miller institute of TU Munich in Obernach, Germany. At this facility, we
140 reconstructed the bed characteristics of the Albula, Navisence and Avançon de Nant field sites, one after another, in a flume
141 test reach with dimensions of 24 m x 1 m equipped with two impact plates (Fig. 3). Each site reconstruction used bedload
142 material collected during field calibration measurements, and we adjusted the flow velocity, flow depth, and bed roughness
143 to match the respective field observations. A detailed description of the original flume setup and the performed experiments
144 can be found in Nicollier et al. (2020). In this paper, we primarily use the single-grain-size experiments conducted in 2018
145 with the flume configured to match conditions at the Albula field site (Table 2). Single-grain-size experiments consisted of
146 feeding the flume with a fixed number of grains for each of the ten particle-size classes described in Sect. 2.2 above. While
147 these particles were being transported over the SPG system, the full raw geophone signal was recorded. Up to 33 repetitions
148 were conducted until a representative range of amplitude and frequency values for each grain-size class were obtained
149 (Nicollier et al., 2021a). The same procedure was repeated for two different flow velocities ($V_f = 1.6 \text{ m s}^{-1}$ and 2.4 m s^{-1}). The
150 obtained information was then used to derive empirical relationships between the mean particle size $D_{m,j}$ and the packet
151 envelope's amplitude $MaxAmp_{env}$ and the ratio $MaxAmp_{env} / f_{centroid}$, as described in Sect. 2.5.2 below.

152 To illustrate the AF and AH methods and their respective performance, we use flume experiments that mimic the
153 Avançon de Nant field site, but with the addition of a 4 m wooden partition wall (Fig. 3) that shields one geophone plate
154 from impacting particles (Nicollier et al., 2021b). With this modified setup, single-grain-size experiments were run using
155 grains from each of the 10 particle-size classes, resulting in a total of 51 runs (Table 2). The flow velocity was set to 3 m s^{-1}
156 to facilitate particle transport through the narrower flume section and is therefore not representative for the Avançon de Nant
157 site, where typical flow velocities were roughly 1.3 m s^{-1} .

158



159
 160 **Figure 3:** Oblique view of the Obernach flume test reach with total length of 24 m and width of 1 m. The bed surface is paved with
 161 particles with diameters equaling the characteristic D_{67} and D_{84} sizes of the natural beds of the reconstructed sites. Grains were
 162 fed into the channel 8 m upstream from the SPG system location (G1 and G2) using either a vertical feed pipe or a tiltable basket
 163 (1). The sensor plate G1 (in red) was shielded from direct particle impacts by the 4 m long removable partition wall (2). The
 164 partition wall and the impact plates were decoupled from each other by a 2 mm vertical gap to prevent disturbances of the
 165 recorded signal. Plexiglas walls (3) on each side of the flume facilitated video recordings of the experiments.

166
 167 **Table 2:** Flume and hydraulic characteristics for the reconstruction of the Albula and the Avançon de Nant field sites.

Parameter	Units	Reconstructed field site setup	
		Albula (without partition wall)	Avaçon de Nant (with partition wall)
Flume width	m	1.02	1.02
Flume gradient of the natural bed	%	0.7	4.0
Bed surface D_{67} ^a	mm	120	200
Bed surface D_{84} ^a	mm	190	320
Number of D_{67} -particles/m ²	m ⁻²	15.0	5.0
Number of D_{84} -particles/m ²	m ⁻²	5.0	2.5
Min. water depth above SPG	m	0.79	0.35
Max. water depth above SPG	m	0.91	0.35
Min. flow velocity 10 cm above SPG ^b	m s ⁻¹	1.6	3.0
Max. flow velocity 10 cm above SPG ^b	m s ⁻¹	2.4	3.0
Min. unit discharge	m ² s ⁻¹	1.6	0.8
Max. unit discharge	m ² s ⁻¹	2.4	0.8
Number of different flow velocity settings	-	2	1
Total number of single-grain-size experiments	-	355	51
Total number of tested particles	-	10705	2485

168 ^a On the basis of line-by-number pebble counts at the natural site and a photo-sieving based granulometric analysis with BASEGRAIN
 169 software (Detert and Weitbrecht, 2013).

170 ^b Flow velocities measured with the OTT MF Pro magnetic-inductive flow meter.



171 2.4 The amplitude-histogram method

172 Wyss et al. (2016a) introduced the packet-based amplitude-histogram (AH) method to derive grain-size information from
 173 geophone signals. A packet is defined as a brief interval, typically lasting 5 to 30 milliseconds, reflecting a single impact of a
 174 particle on a plate (Fig. 1b); it begins and ends when the signal envelope crosses a threshold amplitude of 0.0216 V. The
 175 signal envelope is computed in Python with the Hilbert transform (Jones et al., 2002), yielding a continuous time series
 176 reflecting the total energy in the signal. Each packet's maximum amplitude is then used to assign it to a predefined amplitude
 177 class j (Table 3), yielding a packet-based amplitude histogram (e.g. Fig. 4 in Wyss et al., 2016a). Each amplitude class j is
 178 related to a corresponding grain-size class through the following relationship between the mean amplitude $A_{m,j}$ [V] and the
 179 mean particle size $D_{m,j}$ [mm]:

$$180 \quad A_{m,j} = 4.6 \cdot 10^{-4} \cdot D_{m,j}^{1.71} . \quad (1)$$

181 The coefficients in Eq. (1) were determined using 31 basket samples collected at the Erlenbach for which the maximum
 182 geophone amplitude was analyzed as a function of the B-axis of the largest particle found in the sample (Wyss et al., 2016a).
 183 The grain-size classes are delimited by the size of the meshes $D_{sieve,j}$ used to sieve the bedload samples from field
 184 calibration measurements. It is assumed that the number of packets between two thresholds is related to the fractional
 185 bedload mass between the respective sieve sizes (Wyss et al., 2016a). In the present study, we have extended the seven size
 186 classes used by Wyss et al. (2016a) to ten classes, in order to assess the performance of the AH and AF methods for larger
 187 particles.

188
 189 **Table 3: Size classes j derived from sieve mesh sizes $D_{sieve,j}$ (for classes 1 to 7) according to Wyss et al. (2016a), and mean particle
 190 diameter $D_{m,j}$, amplitude-histogram thresholds $th_{ah,j}$ derived from Eq. (1), and amplitude-frequency thresholds $th_{af,low,j}$ and $th_{af,up,j}$
 191 derived from Eq. (4) and (5), respectively. Particles in classes 8 to 10 were manually sorted on the basis of linearly extrapolated
 192 $D_{m,j}$ values. The value of $D_{m,j}$ for the largest class (10) in brackets is an estimate, because this size class is open-ended and thus the
 193 mean varied somewhat from site to site.**

Class j [-]	$D_{sieve,j}$ [mm]	$D_{m,j}$ [mm]	$th_{ah,j}$ [V]	$th_{af,low,j}$ [V]	$th_{af,up,j}$ [V Hz ⁻¹]
1	9.5	12.3	0.0216	0.0132	$1.55 \cdot 10^{-5}$
2	16.0	17.4	0.0527	0.0364	$2.33 \cdot 10^{-5}$
3	19.0	21.8	0.0707	0.0509	$4.45 \cdot 10^{-5}$
4	25.0	28.1	0.1130	0.0868	$7.67 \cdot 10^{-5}$
5	31.4	37.6	0.1670	0.1362	$1.78 \cdot 10^{-4}$
6	45.0	53.2	0.3088	0.2725	$3.93 \cdot 10^{-4}$
7	63.0	71.3	0.5489	0.5244	$7.05 \cdot 10^{-4}$
8	80.7	95.5	0.8378	0.8489	$1.56 \cdot 10^{-3}$
9	113.0	127.9	1.4919	1.6342	$2.79 \cdot 10^{-3}$
10	144.7	(171.5)	2.2760	2.6438	-

194

195 2.5 The amplitude-frequency method

196 In a recent study, Nicollier et al. (2021b) showed that the SPG system is sensitive to extraneous particle impacts despite the
 197 isolating effect of the elastomer. Extraneous signals at individual geophone plates can arise from impacts occurring on
 198 neighboring plates, or from impacts on the concrete sill surrounding the SPG array. The elastic waves generated by such
 199 impacts can reach multiple geophone sensors with enough energy to be recorded as “apparent” packets. Thus, packet



200 histograms (i.e. counts of the number of packets per class j) are subject to a certain bias, especially in the lower size classes.
201 The degree of bias was found to depend mainly on two factors. First, coarser grain sizes of transported bedload were shown
202 to generate more apparent packets. Second, more apparent packets were recorded, for a given bedload mass, at transects
203 containing more SPG plates. Nicollier et al. (2021b) showed that packet characteristics such as the start time, the amplitude
204 and the frequency help in identifying apparent packets and filtering them out from the final packet histograms. This filtering
205 method was subsequently applied to all four field calibration datasets (Albula, Navisence, Avançon de Nant and Erlenbach)
206 and helped to reduce the differences between the site-specific mean calibration relationships for the total bedload flux by
207 about 30% (Nicollier et al., 2021b). Based on these observations, the present study proposes an amplitude-frequency (AF)
208 method as an adaptation of the amplitude-histogram (AH) method presented by Wyss et al. (2016a). By introducing two-
209 dimensional (amplitude and centroid frequency) size class thresholds, the new method aims to reduce the effect of apparent
210 packets and improve the accuracy of fractional bedload flux estimates.

211 2.5.1 Centroid frequency

212 According to the Hertz contact theory, the frequency at which a geophone plate vibrates is controlled by the size of the
213 colliding particle (Johnson, 1985; Thorne, 1986; Bogen and Møen, 2003; Barrière et al., 2015; Rickenmann, 2017). In the
214 present study, the frequency spectrum of a packet is characterized by the spectral centroid f_{centroid} . It represents the center of
215 mass of the spectrum and is computed as

$$216 \quad f_{\text{centroid}} = \frac{\sum f_n A_{\text{FFT},n}}{\sum A_{\text{FFT},n}} \quad (2)$$

217 where $A_{\text{FFT},n}$ [V·s] is the Fourier amplitude (computed with the Fast Fourier Transform FFT) corresponding to the frequency
218 f_n [Hz]. Following Wyss et al. (2016b), before applying the FFT, each packet is preprocessed in two steps. First, a cosine
219 taper is applied at the edges of a max. 8 ms time window around the peak amplitude of each packet. Second, the signal
220 contained in this time window is zero-padded on either side to reach an optimal number of sample points n_{FFT} . The taper is
221 used to smooth the transition between the packet and the concatenated zeros, and to suppress spectral leakage, which results
222 in a more accurate amplitude spectrum. The value of n_{FFT} was set to 2^7 in order to adequately resolve the amplitude
223 spectrum of the raw signal contained in the max. 8 ms time window. This time window focuses on the first arrival waveform
224 to obtain a more accurate evaluation of the high-frequency content of the packet (Nicollier et al., 2021b). The single-sided
225 Fourier transform of the processed packet is then computed in order to extract A_{FFT} and derive f_{centroid} (Eq. 2). A decrease
226 in f_{centroid} with increasing particle size was observed for different bedload surrogate monitoring techniques (Belleudy et al.,
227 2010; Uher and Benes, 2012; Barrière et al., 2015). Furthermore, f_{centroid} has the advantage of showing weaker dependency
228 on the flow velocity and transport mode than the maximum registered packet amplitude (Wyss et al. 2016b; Chen et al.,
229 2021). As shown by Nicollier et al. (2021b), f_{centroid} also contains information about the impact location of a packet-
230 triggering particle. Because high frequencies are more rapidly attenuated than low frequencies along the travel path of a
231 seismic wave, (apparent) packets triggered by impacts on a given plate typically have higher f_{centroid} values than packets
232 triggered by impacts occurring beyond that plate's boundaries.

233 2.5.2 Flume-based amplitude-frequency thresholds

234 The transported bedload mass associated with an individual signal packet is strongly dependent on the size of the impacting
235 particle. Inferring sediment transport rates from SPG signals thus requires assigning each packet to a corresponding sediment
236 size class using threshold values of packet characteristics. Wyss et al. (2016a) derived size class thresholds (or AH
237 thresholds) of packet peak amplitude from field measurements (Eq. 1). In the present study, we derive size class thresholds



238 of packet amplitude and frequency from the single-grain-size experiments conducted at the flume facility using the Albula
239 setup (Nicollier et al., 2021a). For each class j , the lower threshold $th_{af,low,j}$ is based on the maximum amplitude of the
240 packet's envelope $MaxAmp_{env}$ [V] and the upper threshold $th_{af,up,j}$ is based on the ratio $MaxAmp_{env}/f_{centroid}$ [V Hz⁻¹].
241 Compared to the raw signal, the envelope has the advantage of returning the magnitude of the analytical signal and thus
242 better outlines the waveform by omitting the harmonic structure of the signal. Similar combinations of amplitude and
243 frequency have been used to infer particle sizes and improve the detectability of bedload particles in previous studies
244 involving impact plates (Tsakiris et al., 2014; Barrière et al., 2015; Wyss et al., 2016b; Koshiba and Sumi, 2018) and pipe
245 hydrophones (Choi et al., 2020).

246 The lower and upper amplitude-frequency (AF) thresholds are obtained as follows. First, all packets recorded during
247 the single-grain-size experiments (without the partition wall) are filtered with respect to the following criterion adapted from
248 Nicollier et al. (2021b):

$$249 \quad \text{Criterion: } f_{centroid} > a_c \cdot e^{(b_c \cdot MaxAmp_{env})}, \quad (3)$$

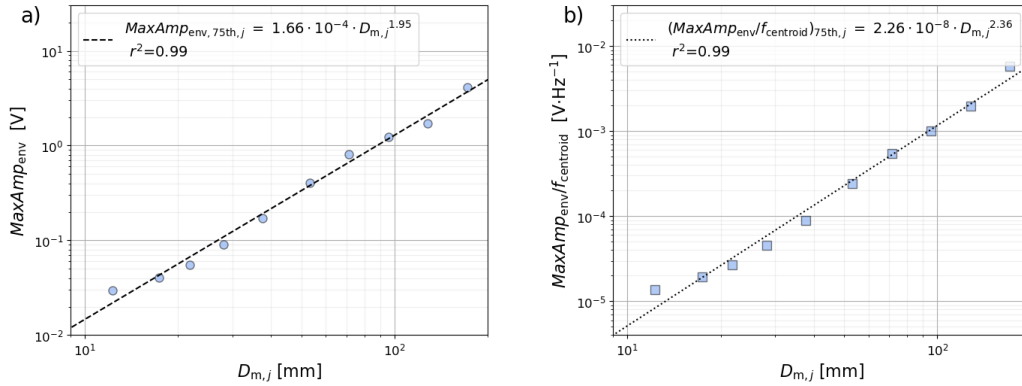
250 with $a_c = 1980$ Hz and $b_c = -1.58$ V⁻¹. Packets that do not meet this criterion are considered as apparent packets and are
251 ignored in the further analysis. The values for the linear coefficient a_c and the exponent b_c were obtained through an
252 optimization process discussed below. The next step consists in fitting a power-law least-squares regression line through the
253 75th percentile amplitude $MaxAmp_{env,75th,j}$ and amplitude-frequency $(MaxAmp_{env}/f_{centroid})_{75th,j}$ values of each class j
254 (Fig. 4), resulting in the following two equations:

$$255 \quad MaxAmp_{env,75th,j} = 1.66 \cdot 10^{-4} \cdot D_{m,j}^{1.95}, \quad \text{and} \quad (4)$$

$$256 \quad \left(\frac{MaxAmp_{env}}{f_{centroid}} \right)_{75th,j} = 2.26 \cdot 10^{-8} \cdot D_{m,j}^{2.36}. \quad (5)$$

257 Finally, the lower and upper threshold values $th_{af,low,j}$ and $th_{af,up,j}$ are obtained by replacing $D_{m,j}$ in Eq. (4) and (5)
258 with the lower ($D_{sieve,j}$) and upper ($D_{sieve,j+1}$) sieve sizes, respectively (Table 3 and triangles in Fig. 5). The advantage in
259 fitting functions such as Eq. (4) and (5) is that they allow the computation of thresholds for any classification of particle
260 (sieve) sizes.

261 Particularly for the largest particles, apparent packets can greatly outnumber real packets. Due to their relatively small
262 amplitudes, these apparent packets can substantially dilute the average signal response associated with the largest grain sizes
263 (see the red boxplots in Fig. 5). However, filtering out apparent packets reveals a clear relationship, which would otherwise
264 be obscured, between the mean particle size $D_{m,j}$ and both the amplitude $MaxAmp_{env}$ and the ratio $MaxAmp_{env}/f_{centroid}$
265 (see the blue boxplots in Fig. 5).



266

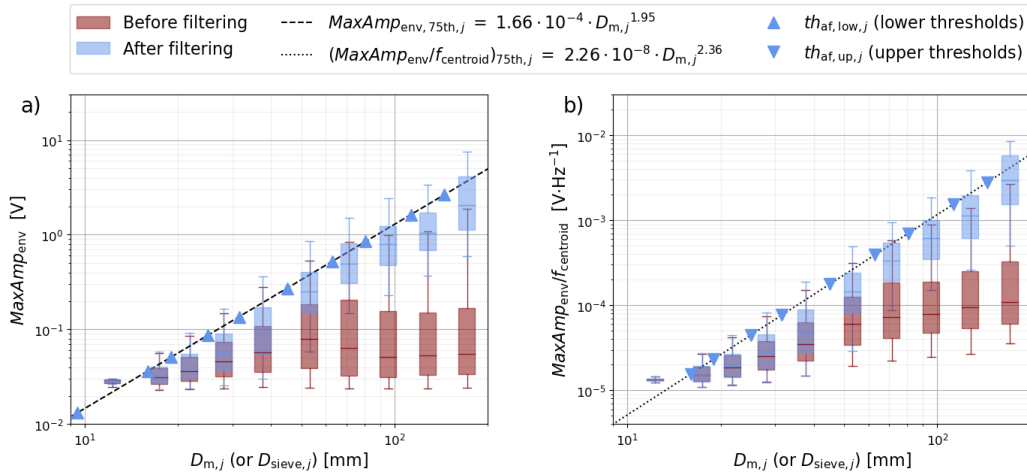
267

268

269

Figure 4: Power-law least-squares regression relationships between the mean particle diameter $D_{m,j}$ and the 75th percentile of the (a) amplitude $MaxAmp_{env,75th,j}$ and (b) amplitude-frequency $(MaxAmp_{env}/f_{centroid})_{75th,j}$ values obtained from the single-grain-size experiments after filtering out apparent packets using the filtering criterion in Eq. (3).

270



271

272

273

274

275

276

277

Figure 5: Range of signal responses obtained from the single-grain-size experiments before (red boxes) and after (blue boxes) filtering out apparent packets using the filtering criterion in Eq. (3), with (a) the maximum amplitude of the envelope $MaxAmp_{env}$ and (b) the ratio $MaxAmp_{env}/f_{centroid}$ as functions of the mean particle diameter $D_{m,j}$. In (a), the lower threshold values $th_{af,low,j}$ are obtained by replacing $D_{m,j}$ with the lower sieve sizes ($D_{sieve,j}$) in the equation of the dashed power-law regression line (Eq. 4). In (b), the upper threshold values $th_{af,up,j}$ are obtained by replacing $D_{m,j}$ with the upper sieve sizes ($D_{sieve,j+1}$) in the equation of the dotted power-law regression line (Eq. 5).

278 2.5.3 Application to field calibration measurements

279

280

281

282

The lower and upper thresholds $th_{af,low,j}$ and $th_{af,up,j}$ obtained from the filtered flume experiments can also be used for the field calibration datasets, if the SPG apparatus and the geophone data recording and preprocessing routines are identical in both cases. The following steps will now lead us to the final general calibration coefficients $k_{b,j,gen}$ (Fig. 6). First, for each field measurement i , the thresholds $th_{af,low,j}$ and $th_{af,up,j}$ are used for counting the number of packets per class j from the



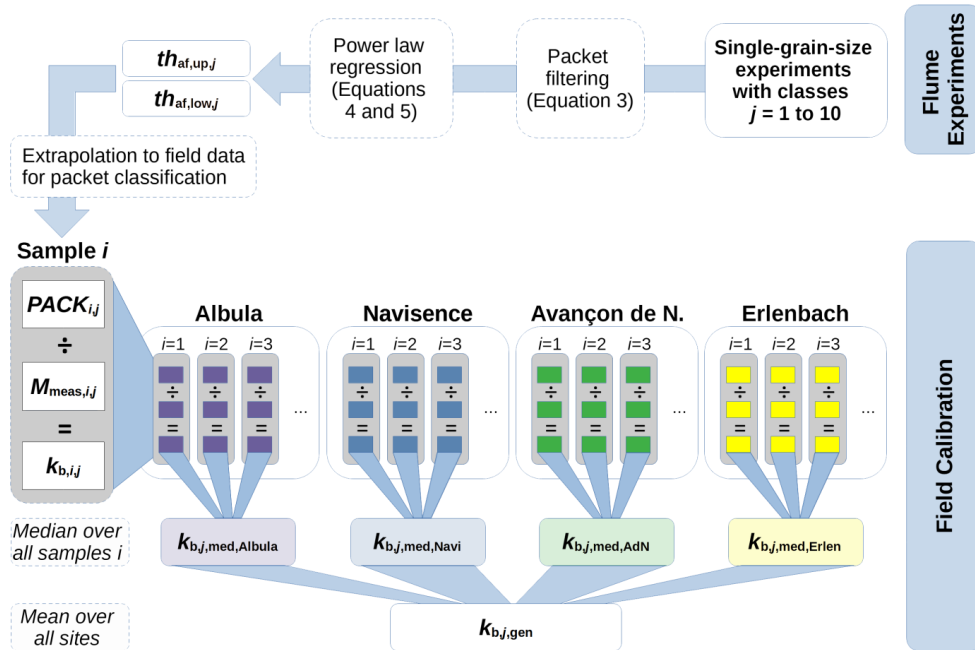
283 recorded geophone signal. Second, a sample- and class-specific calibration coefficient $k_{b,i,j}$ with units $[\text{kg}^{-1}]$ is obtained by
 284 dividing the number of recorded packets $PACK_{i,j}$ by the sampled fractional mass $M_{\text{meas},i,j}$ as follows:

$$285 \quad k_{b,i,j} = \frac{PACK_{i,j}}{M_{\text{meas},i,j}}. \quad (6)$$

286 Finally, the general calibration coefficient $k_{b,j,\text{gen}}$ is computed for each class j using

$$287 \quad k_{b,j,\text{gen}} = \frac{1}{N_{\text{stations}}} \sum_{\text{stations}} k_{b,j,\text{med,station}}, \quad (7)$$

288 where $k_{b,j,\text{med,station}}$ is the site-specific median calibration coefficient, and N_{stations} is the number of stations. Even though
 289 the number of calibration measurements differs from site to site, each coefficient $k_{b,j,\text{med,station}}$ in Eq. (7) is equally
 290 weighted in order to give the same importance to site-specific factors possibly affecting the signal response at each site.



291
 292 **Figure 6:** Workflow leading from the single-grain-size flume experiments with particles from ten size classes j (top right) to the
 293 final array of general calibration coefficients $k_{b,j,\text{gen}}$. Central elements are the lower and upper threshold values $th_{\text{af,low},j}$ and
 294 $th_{\text{af,up},j}$, the number of recorded packets $PACK_{i,j}$ per sample i and class j , the sampled fractional mass $M_{\text{meas},i,j}$, the sample- and
 295 class- specific calibration coefficient $k_{b,i,j}$, and finally the site-specific median calibration coefficient $k_{b,j,\text{med,station}}$. To enable a
 296 comparison with the AH method developed by Wyss et al. (2016a), the “Field Calibration” part of the workflow was also carried
 297 out with the AH thresholds $th_{\text{ah},j}$ (see Table 3).

298 At this point, the single array of calibration coefficients $k_{b,j,\text{gen}}$ is applied as follows to each field
 299 calibration measurement i in order to obtain fractional bedload mass estimates $M_{\text{est},i,j}$:

$$300 \quad M_{\text{est},i,j} = k_{b,j,\text{gen}} \cdot PACK_{i,j}. \quad (8)$$



301 Rickenmann and Fritschi (2017) showed that bedload mass estimates derived from SPG measurements are more accurate at
302 higher transport rates. The estimated fractional bedload mass $M_{est,i,j}$ can be converted to a unit fractional transport rate
303 $q_{b,est,i,j}$ [kg m⁻¹ s⁻¹] using:

$$304 \quad q_{b,est,i,j} = \frac{1}{w_p \cdot n_p} \cdot \frac{M_{est,i,j}}{\Delta t_i}, \quad (9)$$

305 where w_p is the standard width of an impact plate (0.5 m), n_p is the number of plates (which may include the whole transect,
306 or a section of particular interest), and Δt_i is the sampling duration in seconds. Finally, the estimated unit total bedload flux
307 $q_{b,tot,est,i}$ can be computed as follows:

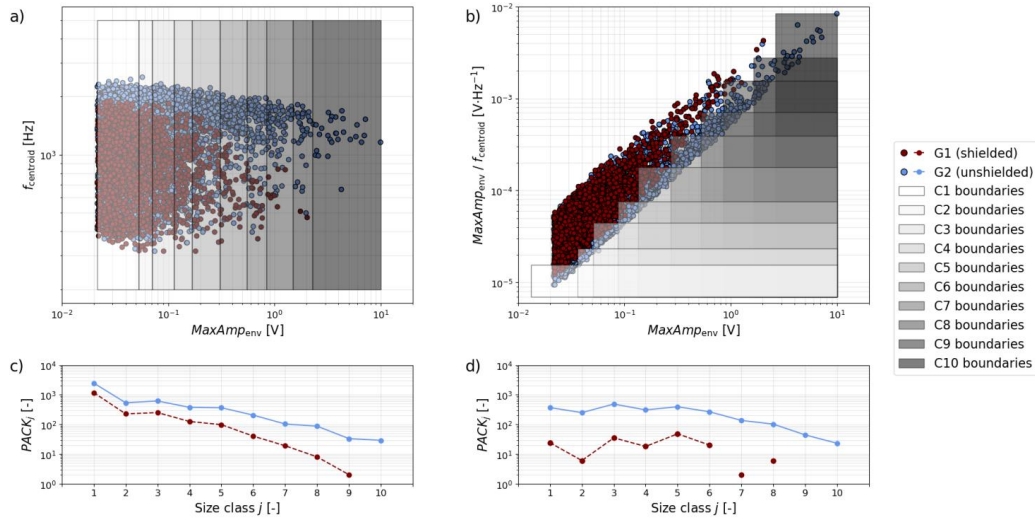
$$308 \quad q_{b,tot,est,i} = \sum_{j=1}^{10} q_{b,est,i,j} \quad (10)$$

309 Note that the exact same procedure was followed using the AH thresholds $th_{ah,j}$ derived from Wyss et al. (2016a) (Eq. 1;
310 Table 3) to compare the performance between the AH method and the new AF method.

311 3 Results

312 3.1 Flume experiments

313 The flume experiments performed in the modified Avançon de Nant setup with the partition wall help to illustrate the two
314 calibration methods. Fig. 7a and 7b show the amplitude and frequency characteristics of all packets detected by the SPG
315 system during these experiments. Packets detected by the shielded sensor G1 all originate from impacts that occurred either
316 on the concrete bed or on plate G2 (Nicollier et al., 2021b). Packets detected by the unshielded sensor G2 are considered as
317 apparent if they are located in the area of the amplitude-frequency graph (Fig. 7a) where G1 and G2 packets overlap. Such
318 packets are presumed to have been triggered by impacts on the concrete bed too. The remaining packets, detected by G2 and
319 located in the non-overlapping area of the amplitude-frequency graph, are considered as real, rather than apparent. The
320 difference in $f_{centroid}$ between real and apparent packets (Fig. 7a) reflects the faster attenuation of higher frequencies during
321 wave propagation, as mentioned earlier. Size class boundaries derived by the AH method of Wyss et al. (2016a) encompass
322 all of the packets, both apparent and real (Fig. 7a). This is because the boundaries are defined solely by AH thresholds
323 ($th_{ah,j}$). By contrast, in the AF method proposed here, the two-dimensional class boundaries given by $th_{af,low,j}$ and $th_{af,up,j}$
324 cover only a fraction of all detected packets (Fig. 7b). Applying the step-like AF thresholds leads to a strong reduction of the
325 number of packets $PACK_j$ within each size class j for plate G1 (shielded), particularly for the smaller classes. Meanwhile, the
326 AF thresholds had little effect on the number of detected packets for G2 (unshielded), except for a strong decrease for
327 classes $j = 1$ and 2, and a slight increase for classes $j = 6$ to 10 (Fig. 7c and 7d). Considering apparent packets as noise and
328 real packets as signal, applying the new AF method results in an increased signal to noise ratio, as shown by the larger
329 vertical separation between the blue (signal) and red (noise) lines in Fig. 7d compared to Fig. 7c.

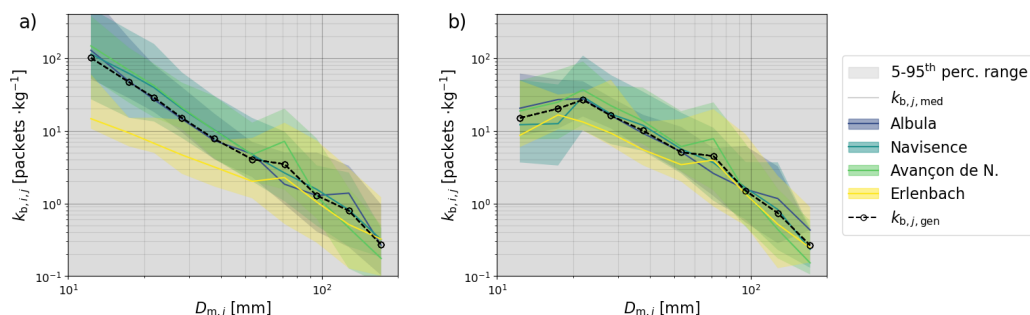


330

331 **Figure 7: Characteristics of the packets recorded during single-grain-size experiments conducted with the Avançon de Nant flume**
 332 **setup using the partition wall, with the maximum amplitude of the envelope $MaxAmp_{env}$ and the centroid frequency $f_{centroid}$.** The
 333 **red and blue dots correspond to packets recorded by the shielded plate G1 and the unshielded plate G2, respectively. The grey**
 334 **rectangles are the class boundaries delimited by the thresholds obtained for the AH method (a) and the AF method (b). In (a),**
 335 **$f_{centroid}$ is shown for information purposes only and is not incorporated in the thresholds. (c) and (d) represent the number of**
 336 **packets $PACK_j$ located within the class boundaries shown in (a) and (b), respectively. Missing markers signify that no packet was**
 337 **detected within the corresponding AH or AF thresholds.**

338 3.2 Field calibration coefficients

339 As discussed in the previous section, the number of packets $PACK_{i,j}$ detected for a given class j varies together with the
 340 thresholds $th_{ah,j}$, $th_{af,low,j}$ and $th_{af,up,j}$. Because the measured fractional bedload mass $M_{meas,i,j}$ remains constant, the
 341 calibration coefficients $k_{b,i,j}$ will depend on the number of packets detected, and thus on the thresholds that are used to
 342 classify them. We can make the following observations regarding the calibration coefficients $k_{b,i,j}$ obtained using the AF
 343 method (Fig. 8b) compared to the AH method (Fig. 8a). First, the $k_{b,i,j}$ coefficients of the smaller size classes are
 344 substantially lower, meaning that fewer packets per unit mass are detected. Second, for the larger size classes, slightly more
 345 packets are detected per unit mass. Third, considering all sites and all size classes j , the overall scatter of the $k_{b,i,j}$
 346 coefficients is smaller. This is reflected in the decrease of the mean coefficient of variation (CV) across all classes j and all
 347 sites from $CV = 1.17$ (in the AH method) to $CV = 0.93$ (in the AF method). Fourth, the scatter of the site-specific $k_{b,i,j}$
 348 coefficients is usually smaller. This is supported by the change of the mean CV across all classes from 0.89 to 0.54 for the
 349 Albula, from 0.83 to 0.75 for the Avançon de Nant and from 1.31 to 1.00 for the Erlenbach, between the AH and AF
 350 methods. The mean CV for the Navisence site however remains unchanged at 0.85. The general coefficients $k_{b,j,gen}$ obtained
 351 from the site-specific median coefficients $k_{b,j,med}$ using Eq. (7) are listed in Table 4.



352
 353 **Figure 8:** The $k_{b,i,j}$ calibration coefficients obtained with the AH method (a) and the AF method (b) for each field site. The colored
 354 areas indicate the range between the 5th and the 95th percentile $k_{b,i,j}$ values, the full lines indicate the site-specific median
 355 coefficients $k_{b,j,med}$ and the black dashed lines indicate the final general calibration coefficients $k_{b,j,gen}$ as a function of the mean
 356 particle diameter $D_{m,j}$ of each grain-size class j .

357

358 **Table 4:** General calibration coefficients $k_{b,j,gen}$ obtained for each grain-size class j with the AH method and the AF
 359 method using Eq. (7).

	Method	Units	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$	$j = 6$	$j = 7$	$j = 8$	$j = 9$	$j = 10$
$k_{b,j,gen}$	AH	kg^{-1}	100.67	46.43	28.68	15.03	7.76	4.04	3.47	1.29	0.79	0.27
	AF	kg^{-1}	14.97	20.15	26.65	16.15	10.06	5.05	4.49	1.50	0.74	0.27

360

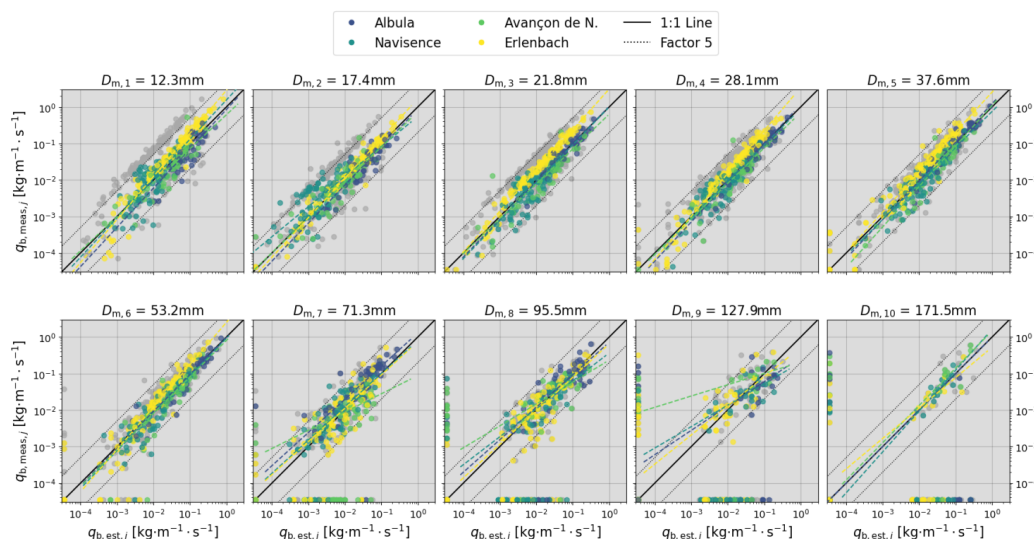
361 3.3 Bedload flux estimates

362 We can now insert the general calibration coefficients $k_{b,j,gen}$ in Eq. (8) to compute fractional bedload mass estimates $M_{est,i,j}$
 363 and subsequently the unit fractional flux estimates $q_{b,est,i,j}$ (Eq. 9) for every sample collected at the four field sites. Fig. 9
 364 illustrates the accuracy of the bedload flux estimates obtained with the AF method for each sample across the grain-size
 365 classes and the field sites. The results obtained with the AH method can be found in Supplementary Information S3, and
 366 Table 5 provides further information on the performance of the two methods. The dashed colored power-law regression lines
 367 shown in Fig. 9, described by the corresponding linear coefficient a and exponent b (Table 5), indicate possible trends in
 368 over/under-estimation at each field site. The coefficient of determination R^2 describes the accuracy of the estimates relative
 369 to the 1:1 line. The root-mean-square error ($RMSE$) quantifies the expected error of the estimates and is expressed in $[\text{kg m}^{-1}$
 370 $\text{s}^{-1}]$. When applied to the field calibration data, the AF method generally yields more accurate flux estimates than the AH
 371 method does. This is most notably reflected by the R^2 values and the percentages p_{factor_2} and p_{factor_5} of all detected
 372 samples whose estimated bedload fluxes differ by less than a factor of 2 and 5, respectively, from the measured values
 373 (Table 5). The five smallest grain-size classes were most strongly affected by these improvements, whereas the estimates for
 374 the largest fractions ($j = 7$ to 10) were only slightly improved.

375 Aside from these comparative observations, it is also worth mentioning the following more general findings that are
 376 valid for both methods: (i) for most size fractions, the relative scatter of the estimates (on the log-log plots) decreases with
 377 increasing transport rates; (ii) at low transport rates, mass fluxes are generally overestimated, while at high transport rates
 378 they are generally underestimated; (iii) mass fluxes for the Erlenbach closely follow the 1:1 line but tend to be slightly
 379 underestimated; (iv) the number of measured ($N_{samples,meas}$) and estimated ($N_{samples,est}$) samples both decrease with



380 increasing particle size. Samples for which either the measured or the estimated flux equals 0 are indicated as dots along the
 381 axes in Fig. 9. If the measured flux is zero but the estimated flux is positive, the sample can be regarded as false positive
 382 (Fawcett, 2006). The difference between $N_{\text{samples,meas}}$ and $N_{\text{samples,est}}$ in Table 5 indicates that the occurrence of such false
 383 positive samples increases with increasing particle size. Further performance metrics derived from the confusion matrix can
 384 be found in the Supporting Information (Table S2).



385
 386 **Figure 9:** Unit fractional transport rate estimates obtained with the AF method for each size class j and each station. The light grey
 387 dots in the background indicate the estimates obtained with the AH method and are represented in more detail in the Supporting
 388 Information (Fig. S1). Each frame is annotated with the mean particle size $D_{m,j}$ of the represented class. The solid black lines
 389 correspond to the reference 1:1 line while the dotted lines delimit factors of 5 above and below (from 0.2 to 5). The dashed colored
 390 lines are power-law regression lines; the mean coefficients over all four sites are listed in Table 5. The dots along the axes indicate
 391 samples for which either the measured or the estimated unit fractional flux equals 0. These samples are not considered for the
 392 computation of the trend lines.

393
 394 **Table 5:** Performance of the AH method and the AF method regarding fractional flux estimates for each class j with following
 395 parameters: the linear coefficient a , the exponent b and the correlation coefficient r of the power-law regression lines visible in Fig.
 396 9; the coefficient of determination R^2 ; the root-mean-square error $RMSE$; and the percentage of all detected samples for which the
 397 estimated value differs from the measured value by less than a factor of 2 and 5 p_{factor_2} and p_{factor_5} , respectively. These values
 398 were first computed for each site separately and then averaged over all four sites. The number of measured $N_{\text{samples,meas}}$ and the
 399 number of estimated samples $N_{\text{samples,est}}$ showing a positive unit fractional rate were summed over all four sites.

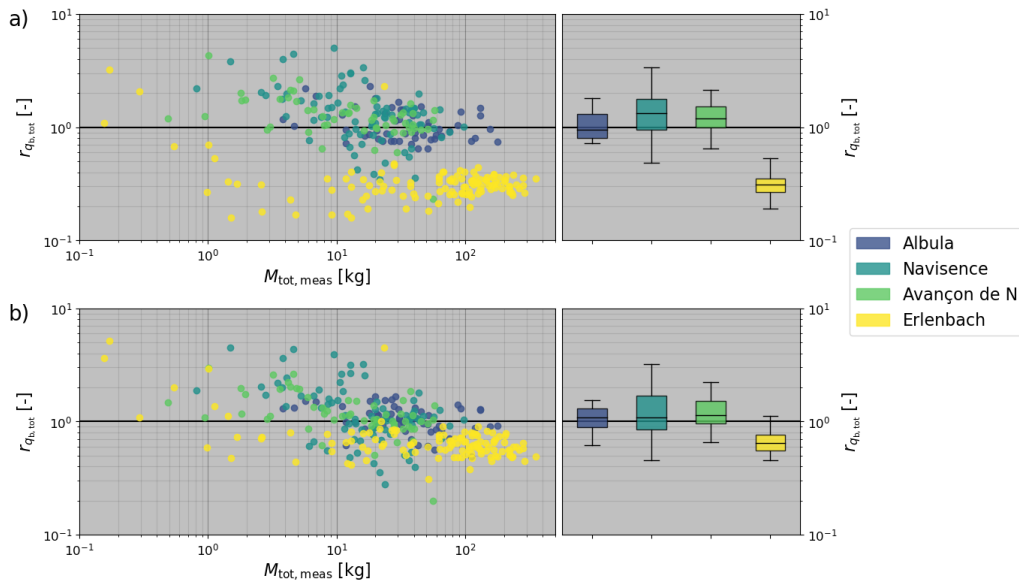
	Units	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$	$j = 6$	$j = 7$	$j = 8$	$j = 9$	$j = 10$
$N_{\text{samples,meas}}$	-	308	308	306	306	302	287	240	213	112	53
$N_{\text{samples,est}}$	-	308	305	307	301	299	289	267	237	149	117
r	-	0.77	0.83	0.87	0.88	0.91	0.89	0.73	0.75	0.53	0.46
a	-	3.6	2.02	1.95	2	1.39	1.54	0.85	0.53	0.42	0.58
b	-	0.94	0.95	1	1.05	1.01	1.05	0.83	0.83	0.64	0.6
R^2	-	0.4	0.51	0.64	0.70	0.78	0.81	0.36	0.57	-0.16	0.11
$RMSE$	$\text{kg} \cdot \text{m}^{-1} \cdot \text{s}^{-1}$	0.094	0.031	0.044	0.036	0.052	0.048	0.038	0.037	0.04	0.06
p_{factor_2}	%	50	54	54	58	64	72	50	58	37	57
p_{factor_5}	%	72	84	92	93	96	95	86	81	68	73



AF Method	$N_{\text{samples,est}}$	-	308	305	307	305	301	295	279	242	161	84
	r	-	0.79	0.82	0.89	0.91	0.93	0.93	0.81	0.78	0.52	0.61
	a	-	1.46	0.96	1.44	1.54	1.41	1.3	0.73	0.49	0.3	1.16
	b	-	1.07	0.98	1.03	1.05	1.06	1.05	0.81	0.79	0.59	0.74
	R^2	-	0.71	0.72	0.8	0.84	0.85	0.83	0.42	0.55	-0.08	0.59
	$RMSE$	$\text{kg} \cdot \text{m}^{-1} \cdot \text{s}^{-1}$	0.068	0.021	0.035	0.027	0.045	0.040	0.035	0.039	0.042	0.061
	p_{factor_2}	%	69	74	69	78	75	81	53	58	43	47
	p_{factor_5}	%	96	93	98	98	97	97	91	83	68	56

400

401 As indicated by Eq. (10), the unit total flux estimates are computed as the sum of the unit fractional flux estimates over
 402 all 10 classes. Fig. 10 shows the ratio $r_{q_{b,\text{tot}}}$ between the estimated total flux $q_{b,\text{tot,est}}$ and the measured total flux $q_{b,\text{tot,meas}}$
 403 for all 308 calibration samples, as a function of the sampled total mass $M_{\text{tot,meas}}$. Here, the estimates for the Albula, the
 404 Navisence and the Avançon de Nant sites are slightly more accurate with the AF method than with the AH method, whereas
 405 the estimates for the Erlenbach improve substantially, with the median $r_{q_{b,\text{tot}}}$ value increasing from 0.31 to 0.64. Note that
 406 the observations (i) to (iii) made earlier regarding the fractional flux estimates are also valid here. Fig. 10 also provides an
 407 interesting overview of the sampled masses at all four stations, reflecting the capacities of the different devices (automated
 408 and manual basket samplers and crane-mounted net sampler) used to collect the calibration samples.



409

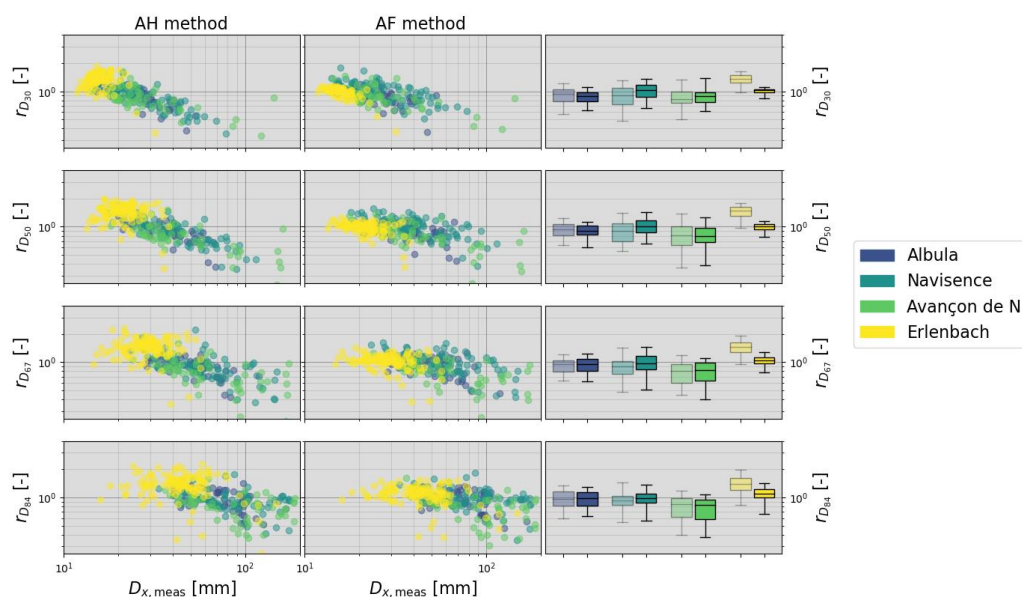
410 **Figure 10: Ratio $r_{q_{b,\text{tot}}}$ between the estimated and the measured unit total mass flux as a function of the total sampled mass**
 411 **$M_{\text{tot,meas}}$, for each collected sample i and each station, for the AH method (a) and the AF method (b). The boxplots on the right**
 412 **indicate the range of $r_{q_{b,\text{tot}}}$ values obtained for each station.**

413 3.4 Grain-size estimates

414 We can combine the SPG bedload flux estimates for all grain-size fractions and thus derive grain-size distributions, which
 415 can then be compared to the measured size distributions of each calibration sample. Fig. 11 compares the performance of the
 416 AH and the AF methods in estimating the characteristic grain sizes D_{30} , D_{50} , D_{67} and D_{84} (where D_x is the grain diameter for
 417 which x percent of the sampled bedload mass is finer). The accuracy of the estimates is indicated by the ratio r_{D_x} between the



418 estimated and the measured characteristic grain size D_x . Compared to the AH method, the AF method mainly improves the
 419 estimates of the four characteristic grain sizes for the Navisence and the Erlenbach sites, but has little effect at the other two
 420 sites. The largest improvement is achieved for the Erlenbach site, with the median $r_{D_{30}}$ changing from 1.37 to 1.02, the
 421 median $r_{D_{50}}$ changing from 1.48 to 1.01, the median $r_{D_{67}}$ changing from 1.46 to 1.05 and the median $r_{D_{84}}$ changing from
 422 1.39 to 1.10. The overall accuracy of the estimates decreases with increasing characteristic size D_x for both methods, and for
 423 every characteristic size D_x , the D_x tends to be overestimated for finer grain mixtures and underestimated for coarser grain
 424 mixtures.



425
 426 **Figure 11:** Ratio r_{D_x} between the estimated and the measured characteristic grain sizes D_{30} , D_{50} , D_{67} and D_{84} as a function of the
 427 measured grain diameter $D_{x, meas}$ for each collected sample i and each station using the AH method (column 1) and the AF method
 428 (column 2). D_x is the grain diameter for which x percent of the sampled bedload is finer. The boxplots in column 3 indicate the
 429 range of r_{D_x} values obtained for each station. The boxes in faded colors show the results obtained with the AH method and the
 430 boxes in brighter colors show the results obtained with the AF method.

431 4 Discussion

432 4.1 The hybrid calibration procedure

433 Recent studies have pointed out the difficulty of transferring flume-based calibrations of the impact plate system to field
 434 applications (e.g. Mao et al., 2016; Wyss et al., 2016c; Kuhnle et al., 2017). In the hybrid calibration approach presented
 435 here, we took advantage of controlled flume experiments, but only to obtain amplitude and amplitude-frequency thresholds
 436 for each particle-size class, which were subsequently applied to field calibration datasets to derive the general calibration
 437 coefficients $k_{b,j, gen}$.

438 Among the three sites reconstructed at the flume facility, only the experiments conducted in 2018 with the Albula setup
 439 were used for calibration purposes in the present study. Although the differences are small, the class thresholds derived from
 440 these experiments yielded slightly more accurate bedload flux estimates than the thresholds derived from the other site



441 reconstructions. A possible explanation for this is the lower bed roughness used for the Albula site reconstruction as
442 compared to the other two setups, which facilitated the transport of larger particles. The Albula setup was also less affected
443 by lateral sorting of small particles (mainly classes $j = 1$ to 4) toward the flume walls, which resulted in a weaker signal
444 response. Additionally, the flow velocities used in this setup ($V_f = 1.6$ and 2.4 m s^{-1}) lie between the velocities measured
445 during the field calibration campaigns at the Navisence and Avançon de Nant sites.

446 The entire hybrid calibration procedure was run iteratively until the optimal linear coefficient and exponent of the
447 criterion (Eq. 3) used to filter out apparent packets were found (Fig. 6). As objective function we used an equally weighted
448 combination of parameters describing the accuracy of bedload flux and grain-size estimates, i.e. r , R^2 , p_{factor_2} , p_{factor_5} , and
449 $RMSE$ as shown in Table 5, r_{D_x} as shown in Fig. 11, and the accuracy derived from the confusion matrix (Fawcett, 2006) as
450 shown in Table S2 in Supporting Information. We looked for two types of optimal calibrations. The first type is a general
451 calibration, for which we have presented the results in Sect. 3. This calibration combines all four stations in order to
452 investigate the feasibility of a general signal conversion procedure applicable to multiple sites equipped with SPG systems.
453 The second type is a site-specific calibration aiming to improve the accuracy of bedload transport rate estimates at a single
454 monitoring station, to be used for a more detailed analysis of bedload-related processes at a given site (details of these site-
455 specific calibrations are available in Supporting Information Sect. S4 and S5).

456 The biases introduced by apparent packets can be removed by site-specific calibration of the coefficients $k_{b,i,j}$, so the
457 AF and AH methods perform about equally well when calibrated separately to each individual site (see Supporting
458 Information Sect. S4 and S5). However, the abundance of apparent packets varies considerably from site to site, owing to
459 differences in the channel geometry, the bedload grain-size distribution, and the construction details of the individual SPG
460 installations. Because the AF method filters out a substantial fraction of these apparent packets, it yields substantially better
461 general calibrations than the AH method does (see Table 5).

462 We also tested the performance of an adapted version of the AH method introduced by Rickenmann et al. (2018). This
463 method was originally developed for the Erlenbach site and aimed to correct for the relationship between the signal response
464 and the transport rate. In the present study, we applied this method to each field site. The only notable improvement
465 introduced by the adapted AH method is the increased number of detected samples at the Erlenbach station, leading to more
466 accurate estimates of the various characteristic grain sizes D_x at this site (Tables S8 and S9 in Supporting Information); the
467 results for the other sites were not substantially improved.

468 4.2 Two-dimensional size class thresholds

469 To understand the performance of the new AF method it is worth taking a closer look at the role of the size class thresholds.
470 As shown in Fig. 7, replacing the upper amplitude thresholds with amplitude-frequency values results in the following two
471 important changes. First, a dimension is added, which facilitates focusing on the narrow range of signal responses
472 characteristic for real packets, and filtering out many of the apparent packets. Second, the areas of the amplitude-frequency
473 domain covered by two adjacent classes can now overlap. Packets located in overlapping areas are assigned once to each
474 class and therefore counted twice. This explains why both the number of detected packets $PACK_j$ (Fig. 7c and 7d) and
475 subsequently the $k_{b,j}$ values (Fig. 8) are slightly higher when the AF method (instead of the AH method) is applied to the
476 larger size classes. Counting such packets twice is not unreasonable, given that the ranges of signal responses recorded
477 during single-grain-size flume experiments for two contiguous grain-size classes significantly overlap, even after apparent
478 packets are filtered out (Fig. 5).

479 Through the reduced area covered by the new amplitude-frequency thresholds in Fig. 7b, a certain percentage of all the
480 packets recorded during the field calibration experiments is neglected for general calibration: 55% at the Albula site, 63% at
481 Navisence, 58% at Avançon de Nant and only 9% at Erlenbach. This suggests that the plates embedded at Erlenbach pick up



482 less noise from their surroundings. A similar trend was observed by Nicollier et al. (2021b) when comparing the maximum
483 amplitude registered by two adjacent plates for a given impact at the same location. This difference in noise detection levels
484 is possibly accentuated by the number of impacted plates during bedload transport events. The SPG array embedded in the
485 artificial U-shaped channel of the Erlenbach has the particularity that only 2 out of its 12 plates are usually impacted by
486 bedload particles during floods (and only sediment crossing these two plates is caught by the automatic basket sampler),
487 while at the other stations all 10 to 30 embedded plates are submerged by the flow and thus can potentially be impacted.

488 4.3 Sampling uncertainties

489 Even though the AF method improved the overall accuracy of flux estimates for most classes (Table 5), some trends
490 addressed in Sect. 3 suggest that factors other than the noise level also control the accuracy of the estimates. The dataset
491 presented in this study includes 308 calibration measurements and is in our knowledge the largest dataset gathered for an
492 impact plate system. Still, it appears that the number of collected samples is not sufficient to accurately assess the
493 performance of the two methods for the three largest particle-size classes (Fig. 9; Table 5). This is mainly due to the fact that
494 in typical sediment mixtures, large particles are rarer than fine particles (Rickenmann et al., 2014; Mao et al., 2016). Earlier
495 investigations have shown that a larger number of detected bedload particles reduces the scatter of total mass estimates by
496 averaging over stochastic factors such as the impact location on a given impact plate, the particle transport mode (sliding,
497 rolling, saltating, etc.), and the impact velocity (Rickenmann and McArdell, 2008; Turowski et al., 2013). A further
498 uncertainty arises because these larger particles are transported at higher bed shear stresses (Einstein, 1950; Wilcock and
499 Crowe, 2003), which also mobilize more total material and thus pose a serious challenge regarding the sampling efficiency
500 of the calibration bedload samplers. Bunte and Abt (2005) and Bunte et al. (2019) have demonstrated that reducing the
501 sampling duration with a bedload trap from 60 to 2 minutes decreases both the sampled unit total bedload flux $q_{b,tot}$ and the
502 sampled maximum particle size D_{max} by about half. In the present study, total bedload fluxes up to $4 \text{ kg m}^{-1} \text{ s}^{-1}$ were measured
503 with the net sampler, meaning that the measurement duration had to be minimized to avoid overloading the sampler. At the
504 Albula stream, for instance, only four samples contained particles of the largest class, and all four were sampled over a
505 duration ranging from 1 to 2 minutes. As a comparison, the longest sampling duration was reached at the Navisence site and
506 lasted 25 minutes. All this suggests that an optimal calibration of the SPG system requires balancing the sampling duration
507 and the number of collected particles. Flume experiments could potentially be used to assess the sampling efficiency of the
508 various calibration sampling methods, along with the detection efficiency of the SPG system.

509 4.4 Transport rate

510 Two further trends are evident in the unit fractional flux estimates obtained for the seven smallest classes, for which most
511 samples were detected ($N_{\text{samples,est}} / N_{\text{samples,meas}} > 96\%$; Table 5). First, the relative scatter (on the log-log plots) of the
512 fractional flux estimates around the power-law regression lines in Fig. 9 is smaller at higher transport rates. Second, both
513 total and fractional fluxes are generally overestimated at low transport rates and underestimated at high transport rates (Fig. 9
514 and 10), which also correspond to the largest calibration samples. These findings agree with results from previous calibration
515 campaigns with the SPG system (Rickenmann and Fritschi, 2017; Rickenmann et al., 2018) but a comprehensive explanation
516 for these trends is still missing. The following hypotheses can be put forward to explain the relationship between the mass
517 flux estimates and the transport rate q_b : (i) The SPG system may suffer from saturation when the transport rate is too high, as
518 has been document in the Japanese pipe microphone system (Mizuyama et al., 2011; Choi, 2020). In our SPG data, we have
519 observed long packets containing multiple large peaks corresponding to several impacts occurring so quickly after one
520 another that they were not detected as separate packets. One can expect that the probability of occurrence of such packets

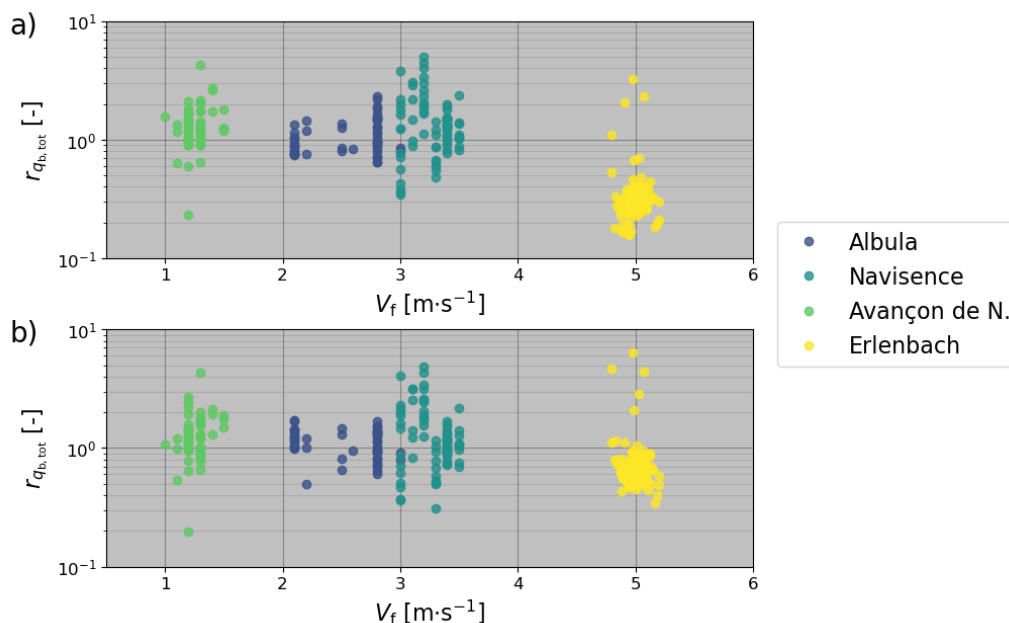


521 increases together with the transport rate, with the transport of large particles (which typically generate packets of longer
522 durations), and with the occurrence of sliding and rolling particles (Chen et al., 2021). The long packets take the place of
523 multiple shorter packets that would otherwise be individually counted; thus, they lead to underestimated mass fluxes for a
524 given $k_{b,j}$ value. (ii) Field observations of bedload sheets being transported over plates at high transport rates were made at
525 the Vallon de Nant site. In the presence of bedload sheets, one can expect that the detection rate of transported particles is
526 hampered by multiple particle layers (Rickenmann et al. 1997; Turowski and Rickenmann, 2009), kinetic sieving (e.g. Frey
527 and Church, 2011) or percolation processes (e.g. Recking et al., 2009).

528 We are not able to give a clear explanation for the overestimates of the characteristic grain size D_x for finer grain
529 mixtures and underestimates for coarser grain mixtures (as shown in Fig. 11). A similar trend was also observed by
530 Rickenmann et al. (2018) for calibration measurements originating from the Erlenbach. We speculate that the decrease of the
531 detection rate along with increasing transport intensity, as mentioned above, may partly explain this phenomenon.

532 4.5 Effect of the flow velocity

533 A recurrent feature in the results presented above is an offset between the estimates obtained for the Erlenbach and those
534 obtained for the three other stations. A similar offset was observed earlier for linear calibration relations for total bedload
535 mass between the Erlenbach and other field sites with more natural approach flow conditions (Rickenmann et al., 2014).
536 Although applying the new amplitude-frequency method has reduced the offset in the present study significantly, it remains
537 visible for both fractional and total bedload flux estimates (Fig. 9, 10, and 12). At the Erlenbach site, the last 35 meters
538 upstream of the SPG system consist of an artificial bed with a steep channel slope of 16%, consisting of large flat embedded
539 boulders (Roth et al., 2016). This explains the supercritical flow regime with a Froude number around 5.1 (Wyss et al.,
540 2016c) and a flow velocity V_f around 5 m s^{-1} at the check dam with the geophone sensors (Table S1). Bedload particle
541 velocity V_p was introduced by Wyss et al. (2016c) as a possible governing parameter affecting the number of particles
542 detected by the SPG system. For the present study, we used V_f as a proxy for V_p , even though bedload particles generally
543 travel more slowly than the fluid that surrounds them (Ancey et al., 2008; Chatanantavet et al., 2013; Auel et al., 2017). Past
544 flume experiments (Wyss et al., 2016a; Kuhnle et al., 2017) have shown that the calibration coefficient $k_{b,j}$ can vary with the
545 flow velocity V_f , such that a three-fold increase in V_f can lead to a two-fold decrease of $k_{b,j}$. Furthermore, bed morphology,
546 bed roughness and flow velocity play important roles in determining particle transport mode, i.e., sliding, rolling, or saltating
547 (e.g. Bagnold, 1973; Lajeunesse et al., 2010). Although high flow velocities generally favor the saltating mode (Ancey et al.,
548 2002), the shallow flow depths measured at the Erlenbach (in average 10 cm; Wyss et al. 2016b) may limit the hop height of
549 larger particles (Amir et al., 2017). Considering all these aspects, we hypothesize that the generally underestimated transport
550 rates observed for the Erlenbach site mainly arise from the exceptionally high flow velocity and the related transport mode
551 (Fig. 12). Continuous flow velocity measurements are lacking at the Albula and Navisence sites, hampering a more detailed
552 analysis of their relationships between flow velocities and detection rates.



553

554 **Figure 12:** Ratio $r_{q_b, tot}$ between the estimated and the measured unit total mass flux as a function of the mean flow velocity V_f , for
555 each collected sample and each station, for the AH method (a) and the AF method (b). The indicated flow velocity corresponds to
556 in situ measurements made during (or close in time to) the corresponding calibration measurement. For better readability, a
557 random scatter ranging from -0.2 m s^{-1} to 0.2 m s^{-1} was added to the stable flow velocity of 5 m s^{-1} measured at the Erlenbach site.

558 4.6 K-fold cross-validation

559 In a last stage, we tested the robustness of the AH and AF methods by splitting the dataset into calibration and validation
560 data. Given that the number of calibration measurements is relatively small and varies between stations, we applied a 4-fold
561 cross-validation technique (e.g. Khosravi et al., 2020). The field calibration measurements were distributed over four folds,
562 each containing an equal number of calibration measurements from each site (Supporting Information Fig. S4). One after
563 another, the folds were used as validation datasets while the remaining three folds were used for calibration. General
564 calibration coefficients $k_{b, j, gen}$ were obtained from the calibration dataset and subsequently applied to the validation data to
565 derive flux estimates. Even though each fold contains a total of only 48 samples (12 per site), the results obtained with the 4-
566 fold cross-validation procedure support our conclusion that including frequency information in the packet classification
567 procedure improves the mean accuracy of the estimates over all sites, in particular for the smaller five to six size classes j
568 (Supporting Information Table S10). Nicollier et al. (2021b) found that most apparent packets are detected as belonging to
569 smaller size classes than the particles that caused them, due to the attenuation of the vibrations as they propagate (see Fig. 7).
570 It is therefore reasonable that the AF method mainly improves the flux estimates for these smaller classes.

571 5 Conclusion

572 The Swiss plate geophone (SPG) is a bedload surrogate monitoring system that has been installed in several gravel-bed
573 streams and was calibrated using direct sampling techniques. While most site-specific calibration relationships for total mass
574 flux are robust across several orders of magnitude, the mean calibration coefficients can still vary by about a factor of six
575 between different sites. In this study, we derived a general procedure to convert SPG signals into fractional bedload fluxes



576 using an extensive dataset comprising controlled flume experiments as well as 308 field calibration measurements from four
577 field sites. The proposed hybrid approach is based on previous findings (Antoniazza et al., 2020; Nicollier et al., 2021b) that
578 the SPG system is biased by elastic waves that propagate through the apparatus and generate noise in the form of spurious
579 “apparent” packets. We introduced the amplitude-frequency (AF) method as an alternative to the amplitude-histogram (AH)
580 method developed by Wyss et al. (2016a). Packets recorded during single-grain-size flume experiments were first filtered to
581 exclude apparent packets, and then used to derive grain-size class thresholds for packet classification. We found that filtering
582 out apparent packets results in more consistent relationships between particle diameter and amplitude-frequency
583 characteristics of the SPG signal. Furthermore, we showed that including frequency information in size class thresholds
584 helps in excluding apparent packets and thus improves the signal-to-noise ratio. In a second stage, we applied these flume-
585 based thresholds to field calibration measurements and derived general calibration coefficients applicable at all four sites for
586 ten different grain-size fractions. The AH method, by contrast, requires site-specific calibration because it cannot account for
587 the site-to-site differences in the abundance of apparent packets. Averaged over the ten grain-size fractions, the bedload
588 masses of 69% and 96% of the samples were estimated within an offset of a factor of two and five, respectively, relative to
589 the measured sampled masses. The remaining discrepancies between the site-specific results are mainly attributed to large
590 differences in flow (and probably particle) velocity. Finally, the sampled mass, the transport rate and the sampling efficiency
591 were identified as further factors possibly influencing the accuracy of mass flux and grain-size estimates.

592 The presented results are highly encouraging regarding future applications of surrogate monitoring methods to
593 investigate bedload transport processes. The findings also underline the valuable contribution of flume experiments to our
594 understanding of the relationship between bedload transport and the recorded SPG signal. But above all, this study highlights
595 the requirements for obtaining calibrations that are transferable across sites: accurate and numerous direct sampling
596 measurements with long sampling durations and large sampled masses, sensors insulated from surrounding noise sources,
597 and highly resolved temporal information about the stream flow, to identify and account for variations in the transport
598 conditions.

599 Notation

600	a_c	Linear coefficient of the criterion
601	A_{FFT}	Fourier amplitude
602	$A_{m,j}$	Mean amplitude registered for particle-size class j
603	b_c	Linear coefficient of the criterion
604	Δt_i	Sampling duration
605	$D_{m,j}$	Mean particle diameter for particle-size class j
606	$D_{\text{sieve},j}$	Lower sieve size retaining particle class j
607	D_x	Characteristic grain size
608	f_{centroid}	Centroid frequency
609	i	Sample index
610	j	Particle-size class index
611	$k_{b,i,j}$	Sample- and class-specific calibration coefficient
612	$k_{b,j,\text{med,station}}$	Median calibration coefficient for particle-size class j and a given station
613	$k_{b,j,\text{gen}}$	General calibration coefficient for particle-size class j
614	$M_{\text{est},i,j}$	Estimated fractional mass per sample and per class
615	$M_{\text{meas},i,j}$	Sampled fractional mass per sample and per class
616	$\text{MaxAmp}_{\text{env}}$	Maximum registered amplitude within a packet



617	$N_{\text{samples,est}}$	Number of detected samples
618	N_{stations}	Number of stations
619	$PACK_{i,j}$	Number of recorded packets per sample and per class
620	p_{factor_x}	Percentage of all detected samples for which the estimated and the measured values differ from each other by less than a factor of x
621		
622	$q_{\text{b,est},i,j}$	Estimated unit fractional transport rate per sample and per class
623	$q_{\text{b,meas},i,j}$	Measured unit fractional transport rate per sample and per class
624	$q_{\text{b,tot,est},i}$	Estimated unit total bedload flux per sample
625	$q_{\text{b,tot,meas},i}$	Measured unit total bedload flux per sample
626	R^2	Coefficient of determination
627	r	Correlation coefficient
628	r_x	Ratio between estimated and measured values x
629	$th_{\text{ah},j}$	Amplitude-histogram thresholds
630	$th_{\text{af,low},j}$	Lower amplitude-frequency thresholds
631	$th_{\text{af,up},j}$	Upper amplitude-frequency thresholds
632	V_f	Mean flow velocity
633	w_p	Standard width of an impact plate

634 **Data availability**

635 The dataset presented in this paper is available online on the EnviDat repository
636 <https://www.envidat.ch/#/metadata/sediment-transport-observations-in-swiss-mountain-streams>.

637 **Author contribution**

638 Tobias Nicollier designed and carried out the field and flume experiments, developed the presented workflow and prepared
639 the manuscript with contributions from all co-authors. Gilles Antoniazza designed and carried out the field experiments at
640 the Vallon de Nant site. Lorenz Ammann helped developing the methodology and contributed to the formal analysis. Dieter
641 Rickenmann contributed to the conceptualization and the supervision of the presented work, contributed to the design of the
642 methodology, and provided support during the field and flume experiments. James W. Kirchner contributed to the
643 development of the methodology and significantly contributed to the preparation of the initial draft.

644 **Acknowledgements**

645 This study was supported by Swiss National Science Foundation (SNSF) grant 200021L_172606, and by Deutsche
646 Forschungsgemeinschaft (DFG) grant RU 1546/7-1. The authors are grateful to Arnd Hartlieb, to the students of the TU
647 Munich, and to the technical staff of the Oskar von Miller Institute for helping to set up and perform the flume experiments.
648 They also warmly thank Norina Andres, Mehdi Mattou, Nicolas Steeb, Florian Schläfli, Konrad Eppel and Jonas von
649 Wartburg for their efforts and motivation during the field calibration campaigns. Special thanks go to Andreas Schmucki,
650 who never gave up repairing the net sampler. Alexandre Badoux is further thanked for his valuable suggestions regarding an
651 earlier version of the manuscript.



652 **Competing interests**

653 The authors declare that they have no conflict of interest.

654 **References**

- 655 Amir, M., Nikora, V., and Witz, M.: A novel experimental technique and its application to study the effects of particle
656 density and flow submergence on bed particle saltation, *J. Hydraul. Res.*, 55, 101–113,
657 <https://doi.org/10.1080/00221686.2016.1233583>, 2017.
- 658 Ancey, C., Bigillon, F., Frey, P., Lanier, J., and Ducret, R.: Saltating motion of a bead in a rapid water stream, *Phys. Rev. E*,
659 66, p. 036306, <https://doi.org/10.1103/PhysRevE.66.036306>, 2002.
- 660 Ancey, C., Davison, A. C., Böhm, T., Jodeau, M., and Frey, P.: Entrainment and motion of coarse particles in a shallow
661 water stream down a steep slope, *J. Fluid Mech.*, 595, 83–114, <https://doi.org/10.1017/S0022112007008774>, 2008.
- 662 Ancey, C.: Bedload transport: a walk between randomness and determinism. Part 2. Challenges and prospects, *J. Hydraul.*
663 *Res.*, 58, 18–33, <https://doi.org/10.1080/00221686.2019.1702595>, 2020.
- 664 Antoniazza, G., Nicollier, T., Wyss, C. R., Boss, S., and Rickenmann, D.: Bedload transport monitoring in alpine rivers:
665 Variability in Swiss plate geophone response, *Sensors*, 20, <https://doi.org/10.3390/s20154089>, 2020.
- 666 Antoniazza, G., Nicollier, T., Boss, S., Mettra, F., Badoux, A., Schaeffli, B., Rickenmann, D., and Lane, S.: Hydrological
667 drivers of bedload transport in an Alpine watershed, ESSOAR [preprint], <https://doi.org/10.1002/essoar.10507461.1>, 12 July
668 2021.
- 669 Auel, C., Albayrak, I., Sumi, T., and Boes, R. M.: Sediment transport in high-speed flows over a fixed bed: 1. Particle
670 dynamics, *Earth Surf. Processes Landforms*, 42, 1365–1383, <https://doi.org/10.1002/esp.4128>, 2017.
- 671 Badoux, A., Andres, N. and Turowski, J.M.: Damage costs due to bedload transport processes in Switzerland, *Nat. Hazards*
672 *and Earth Syst. Sci.*, 14(2), 279–294, <https://doi.org/10.5194/nhess-14-279-2014>, 2014.
- 673 Bagnold, R. A.: The nature of saltation and of bed-load transport in water. *Proc. Royal Soc. A, London, England*, 332, 473–
674 504. <https://doi.org/10.1098/rspa.1973.0038>, 1973.
- 675 Bakker, M., Gimbert, F., Geay, T., Misset, C., Zanker, S., and Recking, A.: Field application and validation of a seismic
676 bedload transport model, *J. Geophys. Res.*, 125, e2019JF005416, <https://doi.org/10.1029/2019JF005416>, 2020.
- 677 Barrière, J., Krein, A., Oth, A., and Schenkluhn, R.: An advanced signal processing technique for deriving grain size
678 information of bedload transport from impact plate vibration measurements, *Earth Surf. Processes Landforms*,
679 <https://doi.org/10.1002/esp.3693>, 2015.
- 680 Bathurst, J. C.: Effect of coarse surface layer on bed-load transport, *J. Hydraul. Eng.*, 133(11), 1192–1205,
681 [https://doi.org/10.1061/\(ASCE\)0733-9429\(2007\)133:11\(1192\)](https://doi.org/10.1061/(ASCE)0733-9429(2007)133:11(1192)), 2007.
- 682 Belleudy, P., Valette, A., and Graff, B.: Passive hydrophone monitoring of bedload in river beds: First trials of signal
683 spectral analyses, *U.S. Geol. Surv. Sci. Invest. Rep.*, 2010-5091, 67–84, 2010.
- 684 Blöschl, G., Kiss, A., Viglione, A., Barriendos, M., Böhm, O., Brázdil, R., et al.: Current European flood-rich period
685 exceptional compared with past 500 years, *Nature*, 583(7817), 560–566, <https://doi.org/10.1038/s41586-020-2478-3>, 2020.



- 686 Bogen, J., and Møen, K.: Bed load measurements with a new passive acoustic sensor, in Erosion and Sediment Transport
687 Measurement in Rivers: Trends and Explanation, *IAHS Publications*, 283, 181-182, 2003.
- 688 Brouwer, R., and Sheremet, O. I.: The economic value of river restoration, *Water Resour. Eco.*, 17, 1-8.
689 <https://doi.org/10.1016/j.wre.2017.02.005>, 2017.
- 690 Bunte, K., Abt, S. R., Potyondy, J. P., and Ryan, S. E.: Measurement of coarse gravel and cobble transport using a portable
691 bedload trap, *J. Hydraul. Eng.*, 130(9), 879-893, [https://doi.org/10.1061/\(ASCE\)0733-9429\(2004\)130:9\(879\)](https://doi.org/10.1061/(ASCE)0733-9429(2004)130:9(879)), 2004.
- 692 Bunte, K., and Abt, S. R.: Effect of sampling time on measured gravel bed load transport rates in a coarse-bedded stream,
693 *Water Resour. Res.*, 41, W11405, <https://doi.org/10.1029/2004WR003880>, 2005.
- 694 Bunte, K., Abt, S. R., Cenderelli, D. A., Ettema, R., and Swingle, K. W.: Bedload traps and Helley-Smith Sampler Collect
695 Different Rates and Particle Sizes of Gravel Bedload, *Proceedings of the SEDHYD 2019 Conference, Federal Interagency
696 Sedimentation and Hydrologic Modeling Conference*, Reno, NV, 2019.
- 697 Chatanantavet, P., Whipple, K. X., Adams, M. A., and Lamb, M. P.: Experimental study on coarse grain saltation dynamics
698 in bedrock channels, *J. Geophys. Res.*, 118, 1161–1176, <https://doi.org/10.1002/jgrf.20053>, 2013.
- 699 Chen, Z., He, S., Nicollier, T., Ammann, L., Badoux, A., and Rickenmann, D.: Signal response of the Swiss plate geophone
700 monitoring system impacted by bedload particles with different transport modes, *Earth Surf. Dyn. Discussion [preprint]*,
701 <https://doi.org/10.5194/esurf-2021-72>, 2021.
- 702 Choi, J. H., Jun, K. W., and Jang, C. D.: Bed-Load Collision Sound Filtering through Separation of Pipe Hydrophone
703 Frequency Bands, *Water*, 12, 1875, <https://doi.org/10.3390/w12071875>, 2020.
- 704 Church, M., Hassan, M. A., and Wolcott, J. F.: Stabilizing self-organized structures in gravel-bed stream channels: Field and
705 experimental observations, *Water Resour. Res.*, 34(11), 3169–3179, <https://doi.org/10.1029/98WR00484>, 1998.
- 706 Dell'Agnese, A., Mao, L., and Comiti, F.: Calibration of an acoustic pipe sensor through bedload traps in a glacierized basin,
707 *CATENA*, 121, 222-231, <https://doi.org/10.1016/j.catena.2014.05.021>, 2014.
- 708 Detert, M., and Weitbrecht, V.: User guide to gravelometric image analysis by BASEGRAIN, *Adv. Sci. Res.*, S. Fukuoka, H.
709 Nakagawa, T. Sumi, H. Zhang (Eds.), Taylor and Francis Group, London, ISBN 978-1-138-00062-9, 1789-1795, 2013.
- 710 Dhont, B., and Ancey, C.: Are bedload transport pulses in gravel bed rivers created by bar migration or sediment waves?
711 *Geophys. Res. Lett.*, 45, 5501–5508. <https://doi.org/10.1029/2018GL077792>, 2018.
- 712 Einstein, H. A.: The Bedload Transport as Probability Problem, *Mitteilung der Versuchsanstalt für Wasserbau an der
713 Eidgenössischen Technischen Hochschule*, Zürich, Switzerland, 1937.
- 714 Einstein, H. A.: The Bedload Transport as Probability Problem, *Technical bulletin*, 1026, United States Department of
715 Agriculture, Soil Conservation Service, Washington, DC, 1950.
- 716 Fawcett, T.: An introduction to ROC analysis, *Pattern Recognit. Lett.*, 27 (8), 861-874,
717 <https://doi.org/10.1016/j.patrec.2005.10.010>, 2006.
- 718 Frey, P., and Church, M.: Bedload: a granular phenomenon, *Earth Surf. Processes Landforms*, 36, 58-69,
719 <https://doi.org/10.1002/esp.2103>, 2011.



- 720 Geay, T., Zanker, S., Misset, C., and Recking, A.: Passive Acoustic Measurement of Bedload Transport: Toward a Global
721 Calibration Curve?, *J. Geophys. Res.*, *125*, <https://doi.org/10.1029/2019JF005242>, 2020.
- 722 Gray, J. R., Laronne, J. B., Marr, J. D. G. (eds): Bedload-surrogate Monitoring Technologies, *U.S. Geol. Surv. Sci. Invest.*
723 *Rep.*, 2010–5091, US Geological Survey: Reston, VA, <http://pubs.usgs.gov/sir/2010/5091/>, 2010.
- 724 Habersack, H., Kreisler, A., Rindler, R., Aigner, J., Seitz, H., Liedermann, M., and Laronne, J. B.: Integrated automatic and
725 continuous bedload monitoring in gravel bed rivers, *Geomorphology*, *291*, 80–93,
726 <https://doi.org/10.1016/j.geomorph.2016.10.020>, 2017.
- 727 Halfi, E., Paz, D., Stark, K., Yogev, U., Reid, I., Dorman, M., and Laronne, J. B.: Novel mass-aggregation-based calibration
728 of an acoustic method of monitoring bedload flux by infrequent desert flash floods, *Earth Surf. Processes Landforms*, *45*,
729 3510-3524, <https://doi.org/10.1002/esp.4988>, 2020.
- 730 Hilldale, R. C., Carpenter, W. O., Goodwiller, B., Chambers, J. P. and Randle, T. J.: Installation of impact plates to
731 continuously measure bed load: Elwha River, Washington, USA, *J. Hydraul. Eng.*, *141*(3),
732 [https://doi.org/10.1061/\(ASCE\)HY.1943-7900.0000975](https://doi.org/10.1061/(ASCE)HY.1943-7900.0000975), 2015.
- 733 Johnson, K.: Contact Mechanics, Cambridge: Cambridge University Press, <https://doi.org/10.1017/CBO9781139171731>,
734 1985.
- 735 Jones, E., Oliphant, T., and Peterson, P.: SciPy: Open source scientific tools for Python [Cited 2021 December 29],
736 Available from: <http://www.scipy.org>, 2002.
- 737 Khosravi, K., Cooper, J. R., Daggupati, P., Thai Pham, B., and Bui, D. T.: Bedload transport rateprediction: Application of
738 novel hybrid data mining techniques, *J. Hydrol.*, *585*, 124774, <https://doi.org/10.1016/j.jhydrol.2020.124774>, 2020.
- 739 Koshiha , T., and Sumi, T.: Application of the wavelet transform to sediment grain sizes analysis with an impact plate for
740 bedload monitoring in sediment bypass tunnels, *E3S Web of Conferences*, *40*, 04022,
741 <https://doi.org/10.1051/e3sconf/20184004022>, 2018.
- 742 Krein, A., Klinck, H., Eiden, M., Symader, W., Bierl, R., Hoffmann, L., and Pfister, L.: Investigating the transport dynamics
743 and the properties of bedload material with a hydro-acoustic measuring system, *Earth Surf. Processes Landforms*, *33*, 152–
744 163, <https://doi.org/10.1002/esp.1576>, 2008.
- 745 Kreisler, A., Moser, M., Aigner, J., Rindler, R., Tritthard, M., and Habersack, H.: Analysis and classification of bedload
746 transport events with variable process characteristics, *Geomorphology*, *291*, 57–68,
747 <https://doi.org/10.1016/j.geomorph.2016.06.033>, 2017.
- 748 Kuhnle, R., Wren, D., Hilldale, R. C., Goodwiller, B., and Carpenter, W.: Laboratory Calibration of Impact Plates for
749 Measuring Gravel Bed Load Size and Mass, *J. Hydraul. Eng.*, *143*, [https://doi.org/10.1061/\(ASCE\)HY.1943-7900.0001391](https://doi.org/10.1061/(ASCE)HY.1943-7900.0001391),
750 2017.
- 751 Lajeunesse, E., Malverti, L., and Charru, F.: Bed load transport in turbulent flow at the grain scale: Experiments and
752 modeling, *J. Geophys. Res.*, *115*, F04001, <https://doi.org/10.1029/2009JF001628>, 2010.
- 753 Le Guern, J., Rodrigues, S., Geay, T., Zanker, S., Hauet, A., Tassi, P., et al.: Relevance of acoustic methods to quantify
754 bedload transport and bedform dynamics in a large sandy-gravel-bed river, *Earth Surf. Dyn.*, *9*, 423–444,
755 <https://doi.org/10.5194/esurf-9-423-2021>, 2021.



- 756 Logar, I., Brouwer, R., and Paillex, A.: Do the societal benefits of river restoration outweigh their costs? A cost-benefit
757 analysis, *J. Environ. Manage.*, 232, 1075–1085, <https://doi.org/10.1016/j.jenvman.2018.11.098>, 2019.
- 758 Manga, M., and Kirchner, J. W.: Stress partitioning in streams by large woody debris, *Water Resour. Res.*, 36(8), 2373–2379,
759 <https://doi.org/10.1029/2000WR900153>, 2000.
- 760 Mao, L., Carrillo, R., Escauriaza, C., and Iroume, A.: Flume and field-based calibration of surrogate sensors for monitoring
761 bedload transport, *Geomorphology*, 253, 10–21, <https://doi.org/10.1016/j.geomorph.2015.10.002>, 2016.
- 762 Mizuyama, T., Hirasawa, R., Kosugi, K., Tsutsumi, D., and Nonaka, M.: Sediment monitoring with a hydrophone in
763 mountain torrents, *Int. J. Erosion Control Eng.*, 4(2), 43–47, <https://doi.org/10.13101/ijece.4.43>, 2011.
- 764 Mühlhofer, L.: Untersuchungen über die Schwebstoff und Geschiebeführung des Inn nächst Kirchbichl (Tirol), *Die*
765 *Wasserwirtschaft*, 1(6), 23 pp, 1933.
- 766 Nicollier, T., Rickenmann, D., and Hartlieb, A.: Field calibration of the Swiss plate geophone system at the Albula stream
767 and comparison with controlled flume experiments, 8 pp., Paper presented at the SEDHYD 2019 Conference, Reno, NV,
768 2019.
- 769 Nicollier, T., Rickenmann, D., Boss, S., Travaglini, E., and Hartlieb, A.: Calibration of the Swiss plate geophone system at
770 the Zinal field site with direct bedload samples and results from controlled flume experiments, in River Flow 2020,
771 *Proceedings of the 10th Conference on Fluvial Hydraulics*, 901–909, <https://doi.org/10.1201/b22619>, 2020.
- 772 Nicollier, T., Rickenmann, D., and Hartlieb, A.: Field and flume measurements with the impact plate: Effect of bedload
773 grain-size distribution on signal response, *Earth Surf. Processes Landforms*, 17 pp., <https://doi.org/10.1002/esp.5117>, 2021a.
- 774 Nicollier, T., Antoniazza, G., Rickenmann, D., Hartlieb, A., and Kirchner, J. W.: Improving the calibration of impact plate
775 bedload monitoring systems by filtering out acoustic signals from extraneous particle impacts, *ESSOAR* [preprint],
776 <https://doi.org/10.1002/essoar.10507726.2>, 13 August 2021b.
- 777 Nitsche, M., Rickenmann, D., Turowski, J. M., Badoux, A., and Kirchner, J. W.: Evaluation of bedload transport predictions
778 using flow resistance equations to account for macro-roughness in steep mountain streams, *Water Resour. Res.*, 47, W08513,
779 <https://doi.org/10.1029/2011WR010645>, 2011.
- 780 Pauli, M., Hunzinger, L., and Hitz, O.: More bed load in rivers. Achieving a sediment balance close to the natural state, *J.*
781 *Appl. Water Eng. Res.*, 6(4), 274–282, <https://doi.org/10.1080/23249676.2018.1497554>, 2018.
- 782 Piantini, M., Gimbert, F., Bellot, F., and Recking, A.: Triggering and propagation of exogenous sediment pulses in mountain
783 channels: insights from flume experiments with seismic monitoring, *Earth Surf. Dyn.*, 9, 1423–1439,
784 <https://doi.org/10.5194/esurf-9-1423-2021>, 2021.
- 785 Prancevic, J. P., and Lamb, M. P.: Unraveling bed slope from relative roughness in initial sediment motion, *J. Geophys. Res.*,
786 120, 474–489, <https://doi.org/10.1002/2014JF003323>, 2015.
- 787 Rachely, C., Friedl, F., Boes, R. M., and Weitbrecht, V.: Morphological response of channelized, sinuous gravel-bed rivers
788 to sediment replenishment, *Water Resour. Res.*, 57, e2020WR029178, <https://doi.org/10.1029/2020WR029178>, 2021.
- 789 Recking, A., Frey, P., Paquier, A., and Belleudy, P.: An experimental investigation of mechanisms involved in bed load
790 sheet production and migration, *J. Geophys. Res.*, 114, F03010, <https://doi.org/10.1029/2008JF000990>, 2009.



- 791 Reid, I., Frostick, L. E., and Layman, J.T.: The incidence and nature of bedload transport during flood flows in coarse-
792 grained alluvial channels, *Earth Surf. Processes Landforms*, 10, 33-44, <https://doi.org/10.1002/esp.3290100107>, 1985.
- 793 Rickenmann, D.: Bed-load transport measurements with geophones and other passive acoustic methods, *J. Hydraul. Eng.*,
794 143(6), 03117004-1-14, [https://doi.org/10.1061/\(ASCE\)HY.1943-7900.0001300](https://doi.org/10.1061/(ASCE)HY.1943-7900.0001300), 2017.
- 795 Rickenmann, D.: Effect of sediment supply on cyclic fluctuations of the disequilibrium ratio and threshold transport
796 discharge, inferred from bedload transport measurements over 27 years at the Swiss Erlenbach stream, *Water Resour. Res.*,
797 56, e2020WR027741, <https://doi.org/10.1029/2020WR027741>, 2020.
- 798 Rickenmann, D., and McArdell, B. W.: Calibration measurements with piezoelectric bedload impact sensors in the Pitzbach
799 mountain stream, *Geodin. Acta*, 21, 35–52, <https://doi.org/10.3166/ga.21.35-52>, 2008.
- 800 Rickenmann, D., and Recking, A.: Evaluation of flow resistance in gravel-bed rivers through a large field data set, *Water*
801 *Resour. Res.*, 47, W07538, <https://doi.org/10.1029/2010WR009793>, 2011.
- 802 Rickenmann, D., and Fritschi, B.: Bedload transport measurements with impact plate geophones in two Austrian mountain
803 streams (Fischbach and Ruetz): system calibration, grain size estimation, and environmental signal pick-up, *Earth Surf.*
804 *Dyn.*, 5(4): 669-687, <https://doi.org/10.5194/esurf-5-669-2017>, 2017.
- 805 Rickenmann, D., Hofer, B., and Fritschi, B.: Geschiebemessung mittels Hydrophon, *Österreichische Wasser- und*
806 *Abfallwirtschaft*, 49(11/12). 219-228, 1997.
- 807 Rickenmann, D., Turowski, J. M., Fritschi, B., Klaiber, A., and Ludwig, A.: Bedload transport measurements at the
808 Erlenbach stream with geophones and automated basket samplers, *Earth Surf. Processes Landforms*, 37(9), 1000–1011,
809 <https://doi.org/10.1002/esp.3225>, 2012.
- 810 Rickenmann, D., Turowski, J. M., Fritschi, B., Wyss, C., Laronne J.B., Barzilai, R., et al.: Bedload transport measurements
811 with impact plate geophones: comparison of sensor calibration in different gravel-bed streams, *Earth Surf. Processes*
812 *Landforms*, 39, 928– 942, <https://doi.org/10.1002/esp.3499>, 2014.
- 813 Rickenmann, D., Steeb, N., and Badoux, A.: Improving bedload transport determination by grain-size fraction using the
814 Swiss plate geophone recordings at the Erlenbach stream, in River Flow 2018, *Proceedings of the 9th Int. Conference on*
815 *Fluvial Hydraulics*, 8 pp., <https://doi.org/10.1051/e3sconf/20184002009>, 2018.
- 816 Roth, D. L., Brodsky, E. E., Finnegan, N. J., Rickenmann, D., Turowski, J.M., and Badoux, A.: Bed load sediment transport
817 inferred from seismic signals near a river, *J. Geophys. Res.*, 121, 725-747, <https://doi.org/10.1002/2015JF003782>, 2016.
- 818 Thorne, P. D.: Laboratory and marine measurements on the acoustic detection of sediment transport, *J. Acoust. Soc. Am.*, 80,
819 899–910, <https://doi.org/10.1121/1.393913>, 1986.
- 820 Tsakiris, A. G., Papanicolaou, A. N., and Lauth, T.: Signature of bedload particle transport mode in the acoustic signal of a
821 geophone, *J. Hydraul. Res.*, 52, 185–204, <https://doi.org/10.1080/00221686.2013.876454>, 2014.
- 822 Turowski, J.M., and Rickenmann, D.: Tools and cover effect in the Pitzbach, Austria, *Earth Surf. Processes Landforms*, 34,
823 26–37, <https://doi.org/10.1002/esp.1686>, 2009.



824 Turowski, J. M., Bockli, M., Rickenmann, D., and Beer, A. R.: Field measurements of the energy delivered to the channel
825 bed by moving bed load and links to bedrock erosion, *J. Geophys. Res.*, 118, 2438–2450,
826 <https://doi.org/10.1002/2013JF002765>, 2013.

827 Uher, M., and Benes, P.: Measurement of particle size distribution by acoustic emission method, paper presented at XX
828 IMEKO World Congress, Busan, South Korea, 2012.

829 Wilcock, P.R., and Crowe, J.C.: A surface-based transport model for sand and gravel, *J. Hydraul. Eng.*, 129(2), 120-128,
830 [https://doi.org/10.1061/\(ASCE\)0733-9429\(2003\)129:2\(120\)](https://doi.org/10.1061/(ASCE)0733-9429(2003)129:2(120)), 2003.

831 Wyss, C. R., Rickenmann, D., Fritschi, B., Turowski, J., Weitbrecht, V., and Boes, R.: Measuring bed load transport rates by
832 grain-size fraction using the Swiss plate geophone signal at the Erlenbach, *J. Hydraul. Eng.*, 142(5),
833 [https://doi.org/10.1061/\(ASCE\)HY.1943-7900.0001090,04016003](https://doi.org/10.1061/(ASCE)HY.1943-7900.0001090,04016003), 2016a.

834 Wyss, C. R., Rickenmann, D., Fritschi, B., Turowski, J., Weitbrecht, V., and Boes, R.: Laboratory flume experiments with
835 the Swiss plate geophone bed load monitoring system: 1. Impulse counts and particle size identification, *Water Resour. Res.*,
836 52, 7744–7759, <https://doi.org/10.1002/2015WR018555>, 2016b.

837 Wyss, C. R., Rickenmann, D., Fritschi, B., Turowski, J., Weitbrecht, V., Travaglini E, et al.: Laboratory flume experiments
838 with the Swiss plate geophone bed load monitoring system: 2. Application to field sites with direct bed load samples, *Water*
839 *Resour. Res.*, 52, 7760–7778, <https://doi.org/10.1002/2016WR019283>, 2016c.

840 Yager, E. M., Kirchner, J. W., and Dietrich, W. E.: Calculating bed load transport in steep boulder bed channels, *Water*
841 *Resour. Res.*, 43, W07418, <https://doi.org/10.1029/2006WR005432>, 2007.

842 Yager, E. M., Dietrich, W. E., Kirchner, J. W., and McArdeil, B. W.: Prediction of sediment transport in step-pool channels,
843 *Water Resour. Res.*, 48, W01541, <https://doi.org/10.1029/2011WR010829>, 2012.

844

845

846

847

848