Supporting Information for

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

Tobias Nicollier[1,2], Gilles Antoniazza[3,1], Lorenz Ammann[1], Dieter Rickenmann[1], James W. Kirchner[1,2,4]

[1]Swiss Federal Research Institute WSL, Birmensdorf, 8903, Switzerland

[2]Deptartment of Environmental System Sciences, ETH Zürich, Zürich, 8092, Switzerland

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

[4]Deptartment of Earth and Planetary Science, University of California, Berkeley, 94720, USA

Contents of this file

S1.   Field Calibration

The field calibration measurements of the Swiss plate geophone (SPG) system at the Albula and Navisence sites were conducted based on the concept developed by Kreisler et al. (2017), using a crane-mounted net sampler adapted from the Bunte bedload trap (Bunte et al., 2010; Nicollier et al., 2019). The net sampler consists of a steel frame, a sampler bag and steel bar (see Fig. 4a and 4b in the main text). The 3 m long sampler bag has a mesh size of 8 mm × 8 mm, corresponding to the size of the smallest particle size that can be sampled. The frame on which the net is fixed has an intake of 500 mm width and 300 mm height in order to cover the whole width of a steel plate. In addition, a thin tilted metal plate was welded at the bottom of the intake to ensure a good coupling with the concrete sill. Note that the presence of a smooth concrete sill increases the sampling accuracy by enabling an accurate positioning of the sampler and preventing any disturbance of the natural bed (Bunte et al., 2004). The steel bar is mounted centrally on the upper part of the intake frame and connected to a crane over a hydraulic rotator. This system can withstand the fluvial forces exerted on the sampler and helps to position the aperture of the frame parallel to the steel plate. Three additional elements were necessary to ensure accurate sampling. First, a cable with markers indicating the correct position of the sensor plates was stretched from one bank to the other. Second, two static ropes attached on each side of the frame and handled from the banks gave support to the hydraulic rotator to correct for fluvial forces at high discharges. Third, a metallic tube was fixed horizontally at the top of the steel bar (which always remained above the water surface) to visually indicate the positioning of the frame parallel to the sensor plates in turbid water.

A calibration measurement started as soon as the frame was placed on the concrete bed on the downstream side of an impact plate. The duration of each run had to be carefully matched with the current discharge in order to avoid overloading the sampling bag. After direct sampling downstream of an impact plate and synchronous recording of the raw geophone signal at a sampling rate of 10 kHz, each bedload sample was sieved and weighed per grain-size fraction following the ten sieve classes presented in Table 1 in the main text. The large capacity of the net proved its value in collecting bedload

samples with masses ranging from 0.82 kg to 556 kg. Also, having the sampling system fixed on a mobile crane made it possible to vary the sampling location along a given SPG array from measurement to measurement.

Different direct sampling techniques were used to calibrate the Erlenbach and the Avançon de Nant streams (Fig. 2 in the main text). At the Erlenbach site, the bedload samples were collected using automated basket samplers covering the width of two impact plates (Rickenmann et al., 2012). The basket samplers were automatically set into motion when both the water discharge and the number of impulses recorded by the SPG system crossed predefined thresholds. The first basket was then pulled towards the lower edge of the concrete sill below the SPG array, until it was aligned with the two center plates of the SPG system. There, the basket sampled bedload for a maximum of 10 minutes or until a given load was reached. The loaded basket was then moved further to the opposite bank, and if the triggering criteria mentioned earlier were still met, the next basket was moved in (up to a total of three such baskets). At the Avançon de Nant, the SPG system was embedded in a concrete weir and the calibration measurements were conducted using a one-plate-length sediment basket mounted with rollers on a rail fixed on the downstream side of the concrete weir, and (for some measurements) the basket was extended with a net to increase the sampling volume. Using a system of ropes and pulleys, the sediment basket could be moved along the rail and placed directly downstream of a given plate (Antoniazza et al., 2021). Apart of the sampling technique, the calibration procedure was similar to the one followed at the Albula and Navisence sites.

S2.   Flow conditions for the Field and Flume Measurements

Table S1: Minimum and maximum values of the mean flow velocity, flow depth and Froude number measured during calibration measurements at the Albula, Navisence, Avançon de Nant and Erlenbach field sites as well as during flume experiments with the Albula, Navisence and Avançon de Nant site setups.

			Field				Flume		
			Albula [a]	Navisence [a]	Avançon de Nant [b]	Erlenbach [c]	Albula [a]	Navisence [a]	Avançon de Nant [a]
Mean flow velocity	min.	$m\,s^{-1}$	2	3	1.2	5	2	2.6	1.75
	max.		3.1	3.5	1.56	5	2.7	3.3	2.8
Flow depth	min.	m	0.7	0.4	0.15	0.08	0.79	0.24	0.12
	max.		0.95	0.65	0.23	0.15	0.91	0.54	0.5
Froude number	min.	-	0.76	1.51	0.99	5.64	0.72	1.69	1.61
	max.		1.02	1.39	1.04	4.12	0.9	1.43	1.26

[a] Streamflow data derived from a magnetic-inductive flow meter OTT MF Pro (Nicollier et al, 2021a)
[b] Streamflow data derived from a flow height sensor Vegapuls WL 61 radar (Antoniazza et al., 2021)
[c] Streamflow data derived from a 2-D laser sensor TiM551 by SICK AG© (Wyss et al., 2016b,c)

S3.   Fractional bedload flux estimates

Here, we first present the fractional flux estimates obtained using the AH method calibrated for the four field sites together (Fig. S1). The general calibration coefficients $k_{b,j,gen}$ used to compute these estimates are listed in Table 4 in the main text.

Due to the presence of samples with zero values in both the measured unit fractional flux and the estimated unit fractional flux, we can also evaluate the performance of the AH and the AF methods based on a confusion matrix and the accuracy computed from it (Fawcett, 2006). Regarding the unit fractional flux estimates, we distinguish the following four possible outcomes: (i) if the measured fractional flux for a given sample i is nonzero and is estimated as nonzero, the sample is counted as a true positive (*TP*); (ii) if the measured fractional flux is nonzero but is estimated as zero, the sample is counted as a false negative (*FN*); (iii) if the measured fractional flux is zero and is estimated as zero, the sample is counted as a true negative (*TN*); (iv) if the measured fraction flux is zero but is estimated as nonzero, the sample is counted as a false

positive (*FP*). The accuracy is computed as the sum of the true positive and true negative samples divided by the total number of samples $N_{\text{samples,meas}}$ (Fawcett, 2006). As Table S2 shows, false positives and false negatives were rare in the smallest six size classes for both the AH and AF methods, but could range up to 28 percent (for the AH method) and 23 percent (for the AF method) in the largest four size classes.
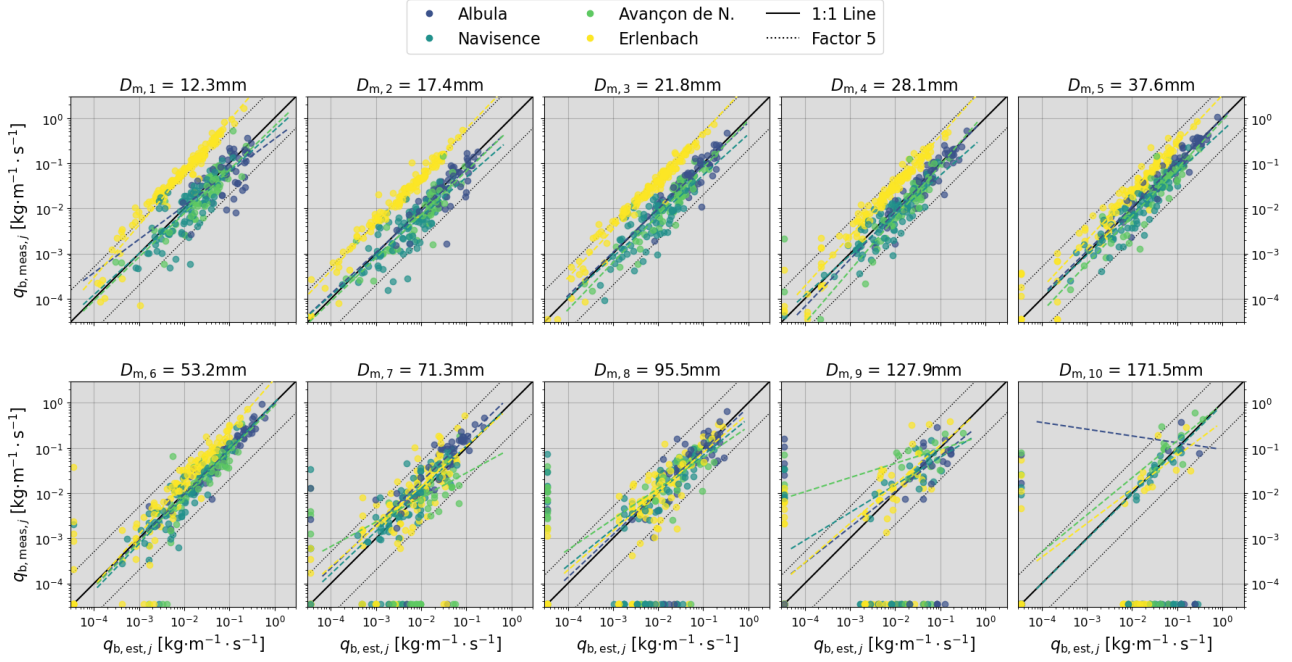
Figure S1: Unit fractional mass flux estimates obtained with the AH method for each size class j and each station. Each frame is annotated with the mean particle size $D_{\text{m},j}$ of the represented class. The solid black lines correspond to the reference 1:1 line while the dotted lines delimit factors of 5 above and below (from 0.2 to 5). The dashed colored lines are power-law regression lines; the mean coefficients over all four sites are listed in Table 5 in the main text. The dots along the axes indicate samples for which either the measured or the estimated unit fractional flux equals 0. These samples are not considered for the computation of the trend lines.

Table S2: Performance of the AH method and the AF method regarding the detection accuracy of samples for each class j with following parameters: the number of measured samples $N_{\text{samples,meas}}$, the true positive *TP*, false positive *FP*, true negative *TN*, and false negative *FN* samples, and the accuracy. Each parameter was computed for the four field sites jointly.

		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
AH method	TP	-	308	305	306	299	297	281	228	190	85	43
	FP	-	0	0	1	2	2	8	39	47	64	74
	TN	-	0	0	1	0	4	13	29	48	132	181
	FN	-	0	3	0	7	5	6	12	23	27	10
	Accuracy	%	100	99	100	97	99	96	82	77	72	73
AF method	TP	-	308	305	306	304	298	285	233	193	86	34
	FP	-	0	0	1	1	3	10	46	49	75	50
	TN	-	0	0	1	1	3	11	22	46	121	205
	FN	-	0	3	0	2	4	2	7	20	26	19
	Accuracy	%	100	99	100	99	99	97	82	77	68	77

S4. Site-specific bedload flux estimates

In this section we present results from site-specific calibrations that were intended to improve the accuracy of bedload estimates at a single monitoring station and to enable a more detailed analysis of bedload transport processes in the future. The following tables list, for each site, the optimal coefficients for the filtering criterion (Eq. 3 in the main text), the median calibration coefficients $k_{b,j,\mathrm{med}}$, and various parameters describing the accuracy of fractional bedload flux estimates represented in Fig. S2.
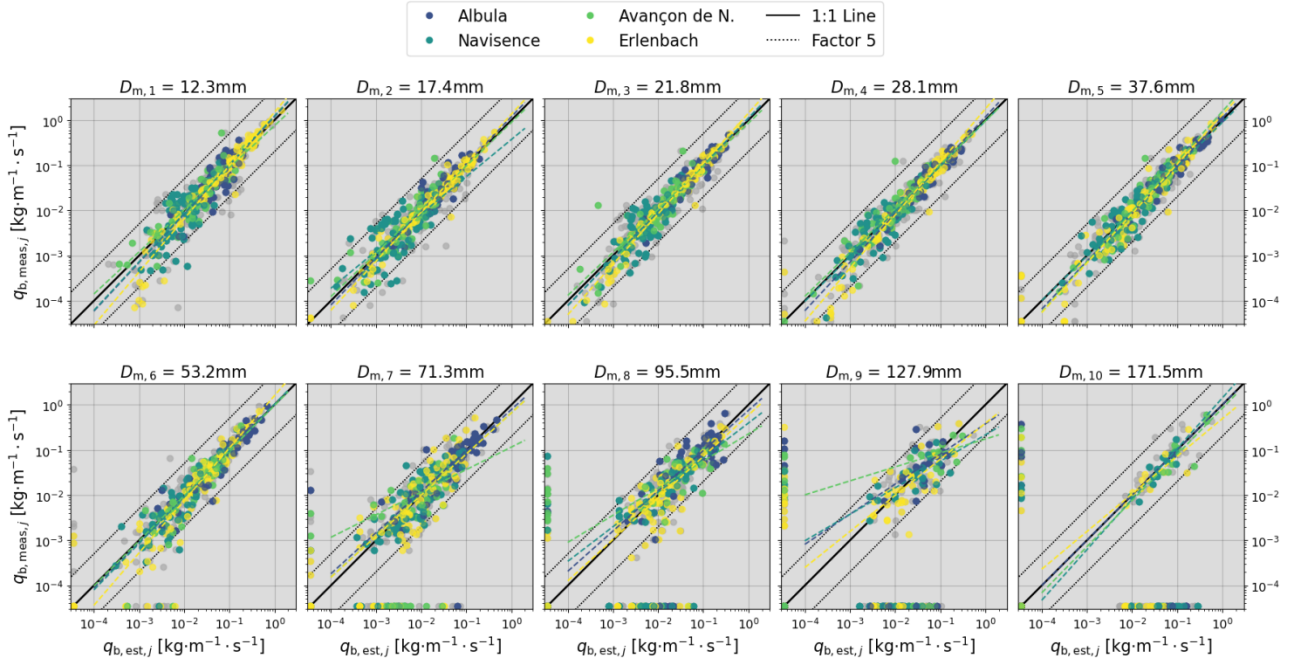
Figure S2: Unit fractional mass flux estimates obtained with the site-specific optimized amplitude-frequency method for each size class j and each station. The light grey dots in the background indicate the estimates obtained with the amplitude-histogram method. Each frame is annotated with the mean particle size $D_{\mathrm{m},j}$ of the represented class. The solid black lines 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 lines are power-law regression lines; the coefficients obtained for each site sites are listed in Tables S5 to S8. The dots along the axes indicate samples for which either the measured or the estimated unit fractional flux equals 0. These samples are not considered for the computation of the trend lines.

Table S3: Linear coefficient a_c and exponent b_c of the filtering criterion (see Eq. 3 in the main text) resulting in the optimal site-specific calibration.

	a_c	b_c
Albula	2220	-1.88
Navisence	1940	-1.83
Avançon de Nant	1630	-2.33
Erlenbach	2040	-3.37

Table S4: Median calibration coefficients $k_{b,j,med}$ with units [$PACK_j$ kg^{-1}] for each class j in the optimal site-specific calibration.

		$k_{b,1,med}$	$k_{b,2,med}$	$k_{b,3,med}$	$k_{b,4,med}$	$k_{b,5,med}$	$k_{b,6,med}$	$k_{b,7,med}$	$k_{b,8,med}$	$k_{b,9,med}$	$k_{b,10,med}$
AH method	Albula	127.78	47.69	26.92	14.52	7.36	4.80	1.84	1.29	1.39	0.27
	Navisence	113.27	61.12	39.25	20.12	10.30	4.68	2.63	1.56	0.82	0.32
	Avançon de N.	146.98	67.59	41.82	20.83	10.17	4.69	7.14	1.26	0.46	0.18
	Erlenbach	14.65	9.33	6.72	4.65	3.19	2.01	2.27	1.05	0.51	0.32
AF method	Albula	32.47	18.03	23.49	13.23	8.17	4.50	2.46	1.48	1.17	0.65
	Navisence	11.16	10.86	28.93	18.39	13.10	6.50	4.47	2.14	0.99	0.32
	Avançon de N.	13.44	24.06	40.36	24.20	15.53	8.67	8.97	2.71	0.85	0.25
	Erlenbach	8.76	15.43	12.99	9.87	6.58	4.34	5.15	2.29	0.91	0.44

Table S5: Performance of the AH method and the AF method in generating fractional flux estimates for each class j at the Albula site. Both methods were optimized for, and applied, to the Albula site, yielding the following parameters: the linear coefficient a, the exponent b and the correlation coefficient r of the power-law regression lines visible in Fig. S2; the coefficient of determination R^2 relative to the 1:1 line; the root-mean-square error $RMSE$; and the percentage of all detected samples for which the 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 were first computed for each site separately and then averaged over all four sites. The number of measured samples $N_{samples,meas}$, the true positive TP, false positive FP, true negative TN, and false negative FN samples, and the accuracy were computed for all four sites jointly.

		Units	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$	$j=7$	$j=8$	$j=9$	$j=10$
$N_{samples,meas}$		-	51	51	51	51	51	51	49	43	22	6
AH method	r	-	0.62	0.81	0.88	0.88	0.94	0.9	0.9	0.68	0.27	-0.23
	a	-	0.41	0.62	0.75	0.99	0.93	1.13	0.81	0.80	0.63	0.09
	b	-	0.74	0.92	0.96	1.05	0.98	1.01	0.96	0.94	0.79	-0.15
	R^2	-	0.02	0.50	0.71	0.76	0.88	0.90	0.78	0.57	-0.41	-1.57
	$RMSE$	kg $\cdot$m$^{-1}\cdot$s^{-1}	0.064	0.027	0.043	0.045	0.063	0.087	0.043	0.078	0.053	0.094
	p_{factor_2}	%	56	82	82	86	90	88	75	65	54	33
	p_{factor_5}	%	90	96	98	96	100	100	95	90	72	66
	TP	-	51	51	51	51	51	51	47	41	18	4
	FP	-	0	0	0	0	0	0	2	5	11	16
	TN	-	0	0	0	0	0	0	0	3	18	29
	FN	-	0	0	0	0	0	0	2	2	4	2
	$Accuracy$	%	100	100	100	100	100	100	92	86	71	65
AF method	r	-	0.76	0.8	0.92	0.93	0.96	0.94	0.92	0.63	0.23	0
	a	-	1.03	0.96	1.10	1.06	1.13	0.86	0.90	0.76	0.38	1.00
	b	-	1.04	0.96	1.00	1.03	1.04	0.96	0.96	0.90	0.67	0.00
	R^2	-	0.53	0.72	0.89	0.91	0.90	0.87	0.81	0.62	-0.02	-0.02
	$RMSE$	kg $\cdot$m$^{-1}\cdot$s^{-1}	0.052	0.027	0.038	0.034	0.052	0.059	0.037	0.083	0.054	0.065
	p_{factor_2}	%	74	84	92	96	94	88	79	60	54	16
	p_{factor_5}	%	98	98	100	100	100	98	95	93	81	16
	TP	-	51	51	51	51	51	51	48	42	19	1
	FP	-	0	0	0	0	0	0	2	3	13	12
	TN	-	0	0	0	0	0	0	0	5	16	33
	FN	-	0	0	0	0	0	0	1	1	3	5
	$Accuracy$	%	100	100	100	100	100	100	94	92	69	67

Table S6: Performance of the AH method and the AF method in generating fractional flux estimates for each class *j* at the Navisence site. Both methods were optimized for, and applied, to the Navisence site, yielding the following parameters: the linear coefficient *a*, the exponent *b* and the correlation coefficient *r* of the power-law regression lines visible in Fig. S2; the coefficient of determination R^2 relative to the 1:1 line; the root-mean-square error *RMSE*; and the percentage of all detected samples for which the estimated value differs from the measured value by less than a factor of 2 and 5 ($p_{\text{factor\_2}}$ and $p_{\text{factor\_5}}$), respectively. These values were first computed for each site separately and then averaged over all four sites. The number of measured samples $N_{\text{samples,meas}}$, the true positive *TP*, false positive *FP*, true negative *TN*, and false negative *FN* samples, and the accuracy were computed for all four sites jointly.

		Units	j = 1	j = 2	j = 3	j = 4	j = 5	j = 6	j = 7	j = 8	j = 9	j = 10
	$N_{\text{samples,meas}}$	-	80	80	80	80	80	79	72	50	26	17
AH method	*r*	-	0.85	0.85	0.88	0.89	0.91	0.9	0.8	0.85	0.49	0.86
	a	-	0.60	0.46	0.58	0.54	0.68	1.18	0.74	0.53	0.26	1.05
	b	-	0.91	0.87	0.91	0.89	0.92	1.04	0.94	0.82	0.61	0.99
	R^2	-	0.38	0.48	0.60	0.70	0.80	0.77	0.68	0.72	0.50	0.84
	RMSE	kg·m^{-1}·s^{-1}	0.012	0.005	0.007	0.008	0.014	0.021	0.013	0.017	0.021	0.047
	$p_{\text{factor\_2}}$	%	68	66	67	65	72	70	68	57	46	88
	$p_{\text{factor\_5}}$	%	91	93	92	97	98	97	91	88	76	88
	TP	-	80	80	80	79	80	78	69	45	20	15
	FP	-	0	0	0	0	0	1	5	23	20	27
	TN	-	0	0	0	0	0	0	3	7	34	36
	FN	-	0	0	0	1	0	1	3	5	6	2
	Accuracy	%	100	100	100	99	100	98	90	65	68	64
AF method	*r*	-	0.87	0.85	0.92	0.93	0.94	0.93	0.83	0.85	0.37	0.86
	a	-	1.23	0.44	1.15	1.02	0.91	1.07	0.82	0.43	0.19	1.05
	b	-	1.08	0.85	1.04	1.01	0.98	1.03	0.95	0.78	0.55	0.99
	R^2	-	0.21	0.49	0.58	0.68	0.77	0.75	0.69	0.71	0.48	0.84
	RMSE	kg·m^{-1}·s^{-1}	0.012	0.005	0.006	0.006	0.013	0.019	0.011	0.017	0.026	0.046
	$p_{\text{factor\_2}}$	%	60	56	65	72	75	78	69	66	57	88
	$p_{\text{factor\_5}}$	%	96	93	97	97	100	97	98	90	88	88
	TP	-	80	78	80	80	80	79	71	46	24	15
	FP	-	0	0	0	0	0	1	7	24	30	27
	TN	-	0	0	0	0	0	0	1	6	24	36
	FN	-	0	2	0	0	0	0	1	4	2	2
	Accuracy	%	100	98	100	100	100	99	90	65	60	64

Table S7: Performance of the AH method and the AF method in generating fractional flux estimates for each class j at the Avançon de Nant site. Both methods were optimized for, and applied, to the Avançon de Nant site, yielding the following parameters: the linear coefficient a, the exponent b and the correlation coefficient r of the power-law regression lines visible in Fig. S2; the coefficient of determination R^2 relative to the 1:1 line; the root-mean-square error $RMSE$; and the percentage of all detected samples for which the estimated value differs from the measured value by less than a factor of 2 and 5 ($p_{\text{factor\_2}}$ and $p_{\text{factor\_5}}$), respectively. These values were first computed for each site separately and then averaged over all four sites. The number of measured samples $N_{\text{samples,meas}}$, the true positive TP, false positive FP, true negative TN, and false negative FN samples, and the accuracy were computed for all four sites jointly.

		Units	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$	$j = 6$	$j = 7$	$j = 8$	$j = 9$	$j = 10$
	$N_{\text{samples,meas}}$	-	55	55	55	54	55	51	30	43	19	15
AH method	r	-	0.63	0.67	0.72	0.79	0.84	0.81	0.49	0.56	0.72	0.77
	a	-	0.95	0.88	1.32	1.91	1.16	1.00	0.14	0.29	0.16	0.67
	b	-	0.97	0.96	1.05	1.16	1.06	1.00	0.54	0.68	0.31	0.82
	R^2	-	0.77	0.79	0.79	0.67	0.76	0.68	0.11	0.35	0.18	0.19
	$RMSE$	kg $\cdot$m$^{-1}\cdot$s^{-1}	0.064	0.018	0.028	0.019	0.024	0.030	0.022	0.030	0.056	0.059
	$p_{\text{factor\_2}}$	%	74	87	81	79	76	80	50	48	36	73
	$p_{\text{factor\_5}}$	%	96	94	96	92	96	96	76	65	68	86
	TP	-	55	55	55	52	55	51	26	30	14	13
	FP	-	0	0	0	1	0	3	20	3	6	6
	TN	-	0	0	0	0	0	1	5	9	30	34
	FN	-	0	0	0	2	0	0	4	13	5	2
	$Accuracy$	%	100	100	100	95	100	95	56	71	80	85
AF method	r	-	0.55	0.59	0.73	0.83	0.88	0.89	0.77	0.73	0.77	0.87
	a	-	0.76	0.76	0.97	0.94	1.34	1.33	0.10	0.44	0.30	0.51
	b	-	0.93	0.93	0.98	0.99	1.09	1.07	0.49	0.80	0.58	0.74
	R^2	-	0.79	0.75	0.72	0.84	0.86	0.70	0.48	0.61	0.36	0.38
	$RMSE$	kg $\cdot$m$^{-1}\cdot$s^{-1}	0.065	0.020	0.026	0.017	0.021	0.027	0.017	0.023	0.031	0.063
	$p_{\text{factor\_2}}$	%	76	74	78	83	89	84	46	60	52	80
	$p_{\text{factor\_5}}$	%	96	94	94	96	98	94	86	79	78	86
	TP	-	55	55	54	54	55	51	30	35	15	14
	FP	-	0	0	0	1	0	4	23	8	10	7
	TN	-	0	0	0	0	0	0	2	4	26	33
	FN	-	0	0	1	0	0	0	0	8	4	1
	$Accuracy$	%	100	100	98	98	100	93	58	71	75	85

S5. Correction procedure applied to the Erlenbach calibration dataset

Tables S8 and S9 include results obtained for the Erlenbach site using the adapted AH method introduced by Rickenmann et al. (2018). This method was originally developed for the Erlenbach site and aims to correct for the relationship between the signal response (alpha values) and the transport rate, representing a fine-tuning of the original AH method presented by Wyss et al. (2016a). The transport rate is approximated using the total unit impulse rate $IMPT_{\text{tot}}$ with units [impulses m^{-1} s^{-1}], with an impulse being a positive peak of the raw geophone signal above a threshold of 0.0216 V. Here, the signal response is given as the alpha value $\alpha_{\text{imp},i,j}$ computed as follows:

$$\alpha_{\text{imp},i,j} = \frac{IMP_{i,j}}{N_{\text{meas},i,j}}, \tag{S1}$$

where $IMP_{i,j}$ is the number of impulses detected for an amplitude class j and $N_{\text{meas},i,j}$ is the number of sampled grains for a given sample i. Following Wyss et al. (2016a), $N_{\text{meas},i,j}$ is obtained from the fractional mass $M_{\text{meas},i,j}$ using the class-specific mean particle mass $G_{\text{m},j}$ given as:

$$G_{\text{m},j} = 0.00219 \cdot D_{\text{m},j}{}^{2.88}, \tag{S2}$$

with $D_{\text{m},j}$ the mean particle diameter for particle-size class j. Once $\alpha_{\text{imp},i,j}$ is obtained for each sample i, the next step consists in fitting a power-law regression line computed with the least-squares solution through the log-transformed unit impulse rate $IMPT_{\text{tot},i}$ and $\alpha_{\text{imp},i,j}$ values of each class j (Fig. S3). The equations of the power-law regression equation are then used to derive a new $\alpha_{\text{imp},i,j}$ value as a function of the unit impulse rate $IMPT_i$ for each sample i and each particle-size class j. Subsequently, the new $\alpha_{\text{imp},i,j}$ values are introduced in Eq. (S1) to obtain estimates of the number of particles $N_{\text{est},i,j}$ and ultimately also estimates of the unit fractional flux $q_{\text{b,est},i,j}$. In the present study, we applied the adapted AH method to each field site. The only notable improvement introduced by this method with regard to the AF method is the increased number of detected samples at the Erlenbach station, potentially leading to more accurate estimates of the various characteristic grain sizes D_x (Tables S8 and S9).
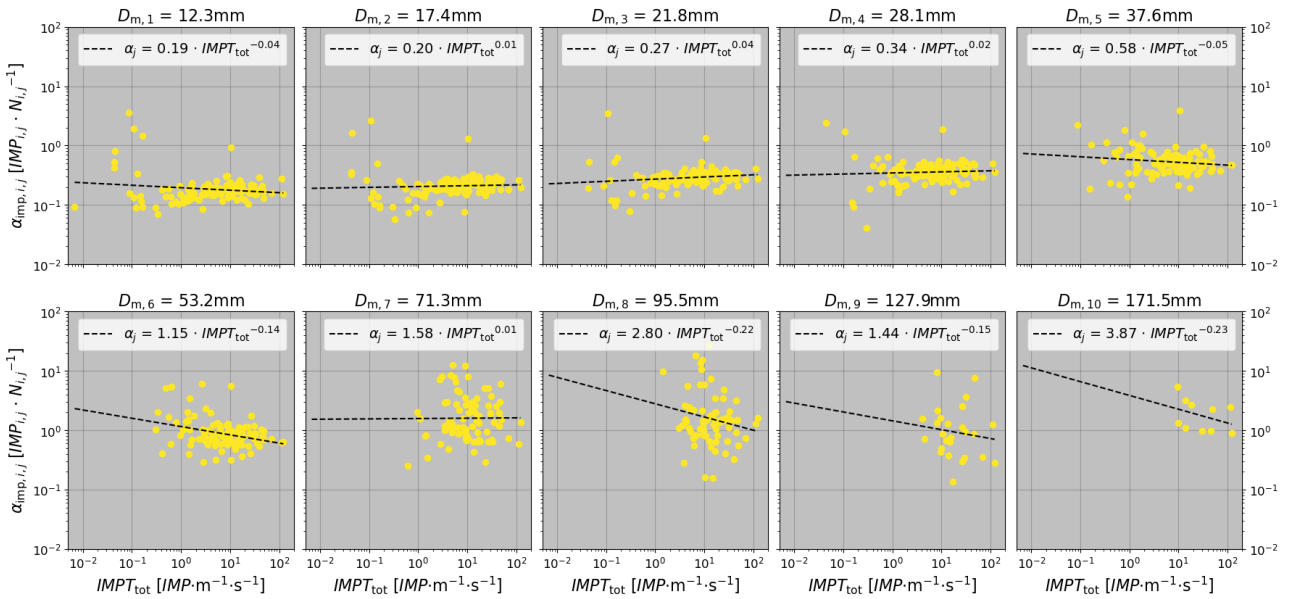
Figure S3: The class-specific relationships between the signal response $\alpha_{\text{imp},i,j}$ and the total unit impulse rate $IMPT_{\text{tot},i}$ for the calibration measurements conducted at the Erlenbach site. The black dashed lines represent the power-law regression lines used to compute new $\alpha_{\text{imp},i,j}$ values, following the adapted AH method presented by Rickenmann et al. (2018).

Table S8: Performance of the AH method, the adapted AH method developed by Rickenmann et al. (2018) and the AF method in generating fractional flux estimates for each class *j* at the Erlenbach site. Both methods were optimized for, and applied, to the Erlenbach site, yielding the following parameters: the linear coefficient *a*, the exponent *b* and the correlation coefficient *r* of the power-law regression lines visible in Fig. S2; the coefficient of determination R^2 relative to the 1:1 line; the root-mean-square error *RMSE*; and the percentage of all detected samples for which the 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 were first computed for each site separately and then averaged over all four sites. The number of measured samples $N_{samples,meas}$, the true positive *TP*, false positive *FP*, true negative *TN*, and false negative *FN* samples, and the accuracy were computed for all four sites jointly.

		Units	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$	$j=7$	$j=8$	$j=9$	$j=10$
	$N_{samples,meas}$	-	122	122	120	121	116	106	89	77	45	15
AH method	*r*	-	0.98	0.98	0.99	0.98	0.97	0.96	0.75	0.93	0.66	0.46
	a	-	1.38	1.21	1.19	1.43	1.24	1.51	0.58	0.49	0.57	0.43
	b	-	1.16	1.05	1.07	1.11	1.07	1.15	0.89	0.87	0.86	0.75
	R^2	-	0.84	0.94	0.93	0.91	0.91	0.82	0.41	0.39	0.19	0.27
	RMSE	kg $\cdot$m$^{-1}\cdot$s^{-1}	0.052	0.019	0.015	0.018	0.040	0.033	0.046	0.019	0.041	0.025
	p_{factor_2}	%	91	91	93	87	85	81	56	59	40	46
	p_{factor_5}	%	95	95	99	94	94	92	85	83	64	60
	TP	-	122	119	120	117	111	101	86	74	33	11
	FP	-	0	0	1	1	2	4	12	16	27	25
	TN	-	0	0	1	0	4	12	21	29	50	82
	FN	-	0	3	0	4	5	5	3	3	12	4
	Accuracy	%	100	98	99	96	94	93	88	84	68	76
Adapted AH method (Rickenmann et al., 2018)	*r*	-	0.96	0.98	0.97	0.96	0.97	0.98	0.83	0.89	0.76	0.97
	a	-	1.00	0.96	0.95	0.85	0.89	0.93	0.65	0.33	0.49	0.79
	b	-	1.00	0.99	0.98	0.96	0.97	0.98	0.91	0.74	0.82	0.94
	R^2	-	0.93	0.93	0.95	0.92	0.92	0.91	0.62	0.48	0.49	0.72
	RMSE	kg $\cdot$m$^{-1}\cdot$s^{-1}	0.068	0.017	0.025	0.024	0.029	0.021	0.037	0.024	0.044	0.021
	p_{factor_2}	%	90	88	90	86	86	84	53	50	28	60
	p_{factor_5}	%	96	95	97	90	94	96	85	80	51	66
	TP	-	122	120	118	113	111	104	83	70	26	10
	FP	-	0	0	0	0	2	6	14	13	16	5
	TN	-	0	0	2	1	4	10	19	32	61	102
	FN	-	0	2	2	8	5	2	6	7	19	5
	Accuracy	%	100	98	98	93	94	93	84	84	71	92
AF method	*r*	-	0.98	0.97	0.99	0.97	0.95	0.96	0.83	0.93	0.68	0.67
	a	-	1.48	1.24	1.38	1.61	1.61	1.72	0.77	0.86	0.39	0.51
	b	-	1.18	1.06	1.11	1.15	1.14	1.17	0.96	1.02	0.79	0.82
	R^2	-	0.87	0.94	0.95	0.90	0.90	0.81	0.54	0.50	0.48	0.67
	RMSE	kg $\cdot$m$^{-1}\cdot$s^{-1}	0.055	0.021	0.022	0.020	0.048	0.041	0.042	0.023	0.039	0.022
	p_{factor_2}	%	91	91	92	87	84	79	60	67	60	73
	p_{factor_5}	%	94	95	100	94	95	92	89	90	86	80
	TP	-	122	119	120	120	113	103	87	76	42	13
	FP	-	0	0	1	1	5	7	22	23	38	38
	TN	-	0	0	1	0	1	9	11	22	39	69
	FN	-	0	3	0	1	3	3	2	1	3	2
	Accuracy	%	100	98	99	98	93	92	80	80	66	67

Table S9: Performance of the AH method, the adapted AH method developed by Rickenmann et al. (2018) and the AF method regarding estimates of characteristic grain sizes D_x optimized for and applied to the Erlenbach site. The accuracy is indicated by the root-mean-square error (*RMSE*) between the estimated and the measured characteristic grain sizes D_x over all samples.

Method	Units	*RMSE* D_{30}	*RMSE* D_{50}	*RMSE* D_{67}	*RMSE* D_{84}
AH method	mm	2.73	8.07	10.69	21.19
Adapted AH method	mm	2.73	4.44	7.59	18.98
AF method	mm	3.30	9.01	10.76	21.15

S6.   Validation procedures

In addition to the 4-fold cross validation procedure (Fig. S4a), two further calibration-validation procedures have been followed to test a hypothetical extrapolation of a calibration relationship to uncalibrated sites. In the one case, the calibration relationship was first derived from three sites and subsequently applied to the fourth site (Fig. S4b). In the other case, the calibration relationship was first derived from one site and subsequently applied to the three remaining sites (Fig. S4c).
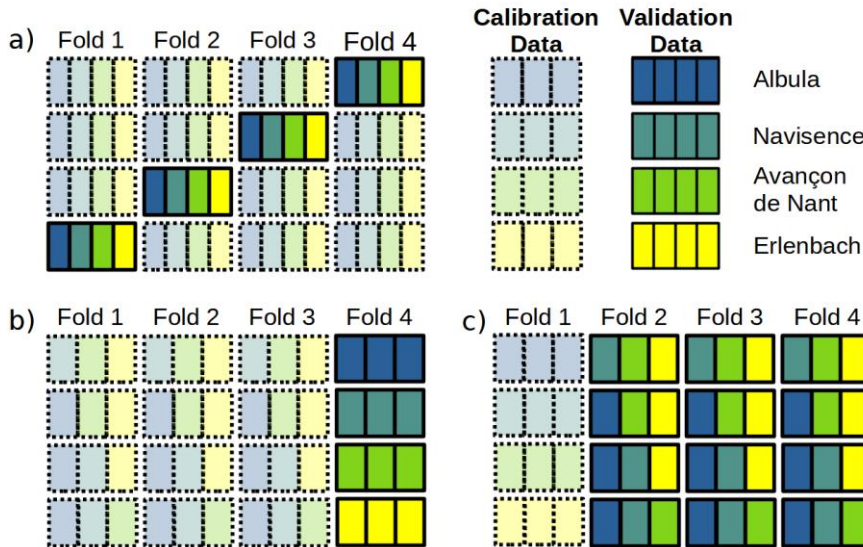
Figure S4: Validation procedures used to test the robustness of the AH method and the AF method; the 4-fold cross-validation in (a) uses the data from all four sites equally distributed over four folds. Each fold is used once as the validation dataset and three times as part of the calibration dataset (results reported in Table S10). In (b) the calibration coefficients $k_{b,j,gen}$ are derived from three sites only and are subsequently applied to the fourth site. This is repeated for all four sites (results reported in Table S11). In (c) the calibration coefficients $k_{b,j,gen}$ are derived from one site only and are subsequently applied to three remaining sites. This is repeated for all four sites (results reported in Table S11).

Table S10: Accuracy of estimates resulting from the 4-fold cross-validation procedure using the AH method and the AF method for each grain-size class j of the validation dataset. The listed values were computed as the mean over all four stations, with the coefficient of determination R^2, the root-mean-square error $RMSE$, and the percentage of all detected samples for which the estimated value differs from the measured value by less than a factor of 2 (p_{factor_2}).

	Units	Method	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$	$j=7$	$j=8$	$j=9$	$j=10$
R^2		AH	-0.07	0.25	0.46	0.53	0.70	0.70	0.01	0.35	-0.84	0.12
		AF	0.47	0.58	0.69	0.77	0.75	0.75	0.28	0.41	-0.91	0.18
$RMSE$	kg $\cdot$m$^{-1}\cdot$s^{-1}	AH	0.077	0.025	0.036	0.032	0.046	0.042	0.033	0.038	0.048	0.084
		AF	0.056	0.017	0.030	0.024	0.040	0.034	0.030	0.039	0.051	0.059
p_{factor_2}	%	AH	50	54	54	60	65	74	56	66	50	65
		AF	69	74	69	79	76	82	57	65	50	75

Table S11: Accuracy of estimates resulting from the alternative validation procedures using either three calibration sites and one validation site or one calibration site and three validation sites, for each grain-size class j. The listed values, with units [kg m^{-1} s^{-1}], are the mean $RMSE$ values over all stations used for validation (either one or three stations).

	Calibration/validation station	Method	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$	$j=7$	$j=8$	$j=9$	$j=10$
3 Calibration stations/1 Validation Station	Albula	AH	0.063	0.027	0.055	0.054	0.097	0.086	0.085	0.087	0.088	0.181
		AF	0.057	0.032	0.045	0.040	0.088	0.072	0.079	0.091	0.092	0.090
	Navisence	AH	0.014	0.005	0.007	0.007	0.013	0.021	0.017	0.021	0.035	0.030
		AF	0.017	0.007	0.007	0.007	0.011	0.019	0.015	0.023	0.045	0.033
	Avançon de N.	AH	0.062	0.018	0.027	0.019	0.026	0.031	0.043	0.035	0.058	0.143
		AF	0.064	0.019	0.026	0.020	0.025	0.026	0.034	0.039	0.065	0.158
	Erlenbach	AH	0.238	0.078	0.102	0.078	0.110	0.081	0.061	0.027	0.087	0.082
		AF	0.152	0.034	0.081	0.054	0.095	0.068	0.054	0.025	0.068	0.057
1 Calibration Station/3 Validation stations	Albula	AH	0.091	0.029	0.041	0.032	0.044	0.037	0.033	0.027	0.051	0.064
		AF	0.072	0.019	0.032	0.024	0.036	0.029	0.029	0.026	0.045	0.073
	Navisence	AH	0.107	0.036	0.053	0.047	0.068	0.056	0.038	0.050	0.057	0.114
		AF	0.077	0.036	0.040	0.031	0.057	0.045	0.036	0.050	0.056	0.075
	Avançon de N.	AH	0.092	0.032	0.051	0.044	0.068	0.054	0.054	0.042	0.078	0.098
		AF	0.069	0.021	0.043	0.037	0.066	0.048	0.050	0.045	0.093	0.051
	Erlenbach	AH	0.452	0.141	0.200	0.144	0.166	0.097	0.034	0.052	0.075	0.082
		AF	0.077	0.021	0.069	0.049	0.089	0.051	0.032	0.054	0.074	0.052