



*Supplement of*

## **Development of a machine learning model for river bed load**

**Hossein Hosseiny et al.**

*Correspondence to:* Claire C. Masteller (cmasteller@gmail.com)

The copyright of individual parts of the supplement might differ from the article licence.

# Supplement

## Contents of this file

Text S1 and S2

Figures S1 to S5

Tables S1 to S5

## Introduction

This supporting information provides five supplementary figures to illustrate the range of data used to train and test the ANN model (Fig. 1-2), the distributions of model predictions of bedload flux (Fig. 3) and the associated model errors (Fig. 4,5). We include five tables that summarize the data used to train and test the ANN model (Table 1-2), a description of the inputs to all bedload transport models used in the study (Table 3), and a summary of model errors (Table 4-5). We provide supporting text describing the data used (Text S1) and an expanded description of how bedload transport predictions were carried out using four previously published bedload models (Text S2).

All data associated with this manuscript, the trained ANN model, and a sample Jupyter notebook for model implementation are published on Zenodo under a GNU General Public License at <https://zenodo.org/record/7641313#.Y-vfbezMIeY>.

## Text S1

Figure S1 shows the distribution of the data and the correlation between variables. It also shows that the data is not normally distributed. Further, it can be visually confirmed that the bedload does not vary monotonically with any individual input variable, suggesting a non-linear relation between the input and output variables. This visual observation is also supported quantitatively (Table S2). The original data used in this study are skewed toward lower values (Figure S1).

However, applying the logarithmic functions tends to stretch the data toward normal distribution (Figure S2).

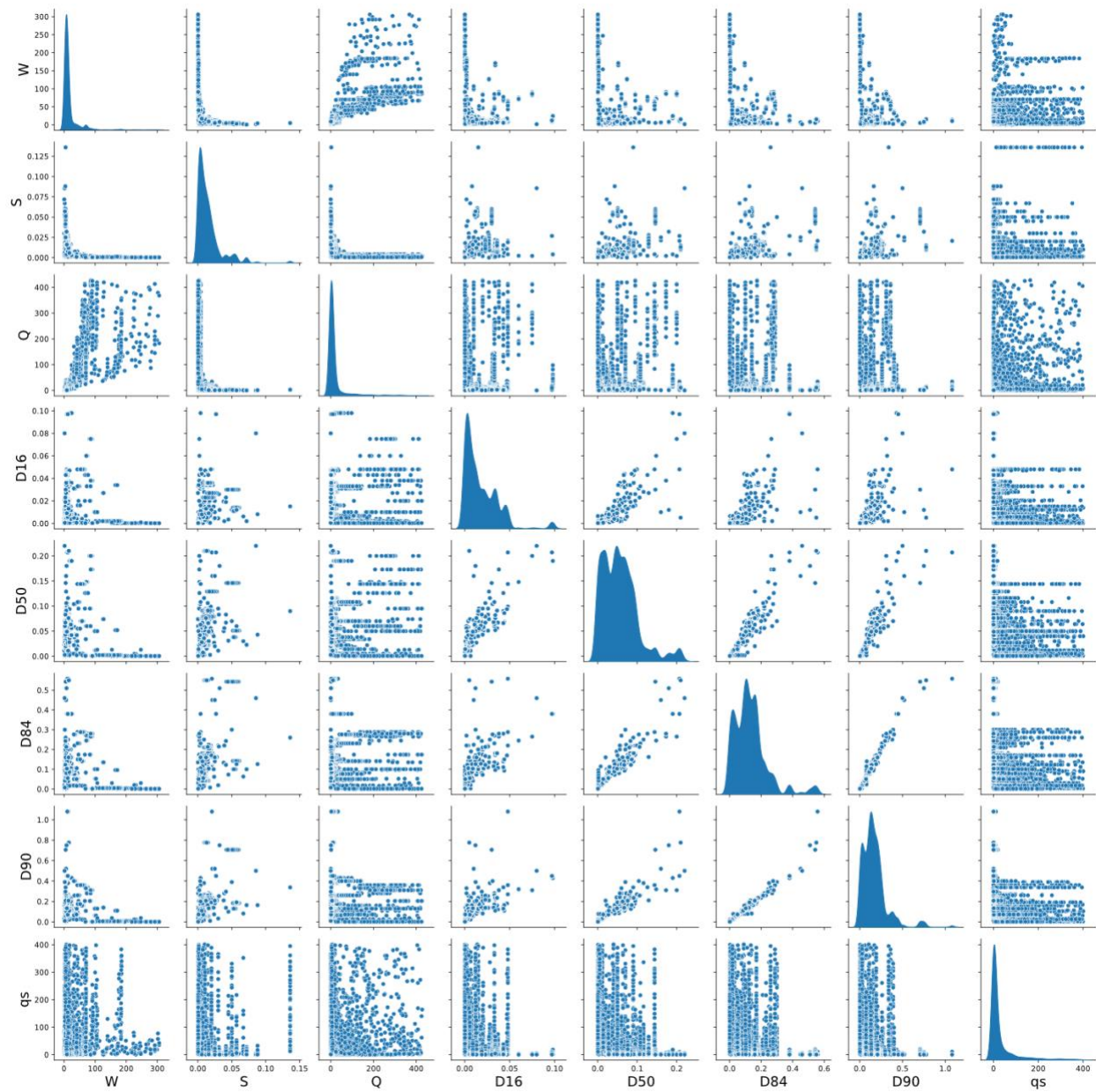


Figure S1. Illustration of the original data used in this study.

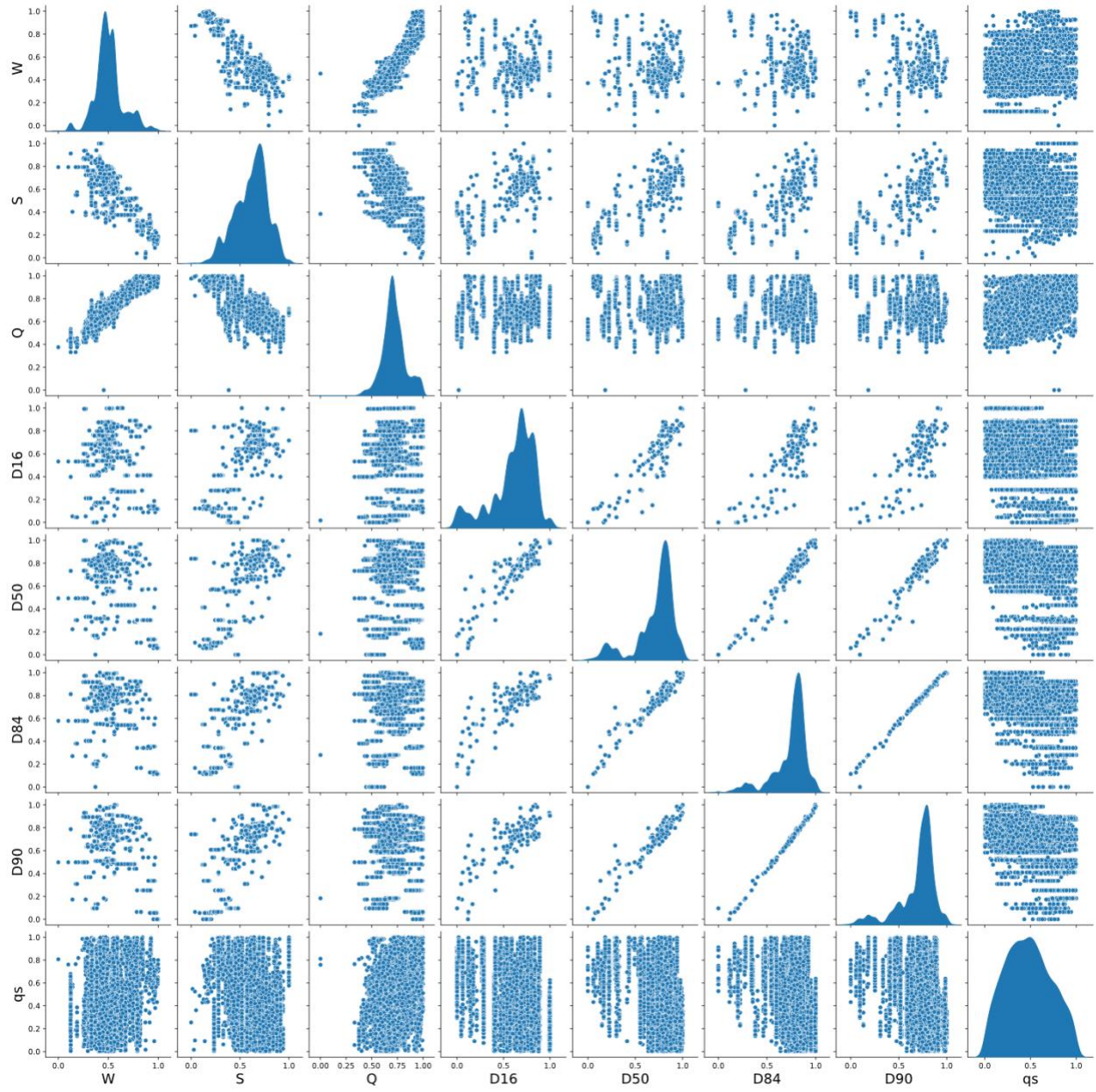


Figure S2. Illustration of the log-transformed, normalized data used in this study.

Table S1. Statistical indices of all data used in this study.

	<b>Width (m)</b>	<b>Slope (m/m)</b>	<b>Q (m<sup>3</sup>/s)</b>	<b>D16 (m)</b>	<b>D50 (m)</b>	<b>D84 (m)</b>	<b>D90 (m)</b>	<b>qb (g/s/m)</b>
<b>Mean</b>	18.176	0.015	23.740	0.015	0.057	0.135	0.171	31.934
<b>Standard Deviation</b>	30.344	0.018	61.142	0.017	0.044	0.104	0.134	65.883
<b>Min</b>	0.300	0.00009	0.00005	0.00013	0.0003	0.0004	0.001	0.108
<b>Max</b>	306.000	0.136	427.580	0.098	0.219	0.543	0.776	401.000
<b>25th percentile</b>	6.100	0.004	1.708	0.005	0.023	0.058	0.072	1.140
<b>50th Percentile</b>	9.000	0.010	4.040	0.012	0.051	0.125	0.152	5.150
<b>75th percentile</b>	14.600	0.019	11.360	0.021	0.076	0.176	0.228	24.800
<b>Skewness</b>	4.841	2.877	3.955	2.513	1.211	1.543	1.660	3.125

Table S2. Correlation coefficients for all data used in this study.

<b>Variable</b>	<b>W</b>	<b>S</b>	<b>Q</b>	<b>D16</b>	<b>D50</b>	<b>D84</b>	<b>D90</b>	<b>qs</b>
<b>W</b>	1.00	-0.29	0.76	-0.07	-0.07	-0.10	-0.10	0.17
<b>S</b>	-0.29	1.00	-0.25	0.09	0.23	0.36	0.37	-0.09
<b>Q</b>	0.76	-0.25	1.00	0.06	0.10	0.05	0.03	0.22
<b>D16</b>	-0.07	0.09	0.06	1.00	0.79	0.58	0.51	-0.23
<b>D50</b>	-0.07	0.23	0.10	0.79	1.00	0.89	0.86	-0.26
<b>D84</b>	-0.10	0.36	0.05	0.58	0.89	1.00	0.98	-0.23
<b>D90</b>	-0.10	0.37	0.03	0.51	0.86	0.98	1.00	-0.22
<b>qs</b>	0.17	-0.09	0.22	-0.23	-0.26	-0.23	-0.22	1.00

## Text S2

### Details of Previous bedload Models

We selected four bedload transport models with varying approaches and degrees of complexity to compare to and build intuition for the predictions of the ANN model. We selected: (1) a probabilistic model developed by Einstein (1950), (2) a physics-based model developed by Wilcock and Crowe (2003), and (3, 4) two empirical models from Wong and Parker (2006) and Recking (2013). This section describes how each of these models were developed to estimate mass (or volume) bedload rate ( $q_b$ ) based on measured data obtained from *bedload.web*.

### **Einstein (1950)**

Einstein bedload model (1950) is probabilistic model that relates the flow intensity to the bedload. To estimate the volumetric bedload flux ( $q_{bv}$ ; m<sup>2</sup>/s), the Einstein model was used as,

$$1 - \frac{1}{\sqrt{\pi}} \int_{-(0.413/\tau^*)-2}^{(0.413/\tau^*)-2} e^{-t^2} dt = \frac{43.5 q^*}{1 + 43.5 q^*} \quad (\text{S-1})$$

where  $\tau^*$  is the dimensionless shear stress for uniform flow (Shields stress),  $t$  is the integral parameter, and  $q^*$  is the dimensionless bedload transport rate (or Einstein bedload number) defined as

$$q^* = \frac{q_{bv}}{D\sqrt{gRD}} \quad (\text{S-2})$$

where  $q_{bv}$  is the volumetric bedload transport rate (m<sup>2</sup>/s),  $g$  is the gravitational acceleration (m/s<sup>2</sup>),  $R$  is the dimensionless submerged specific gravity which was set to 1.65 in this research (Garcia, 2007), and  $D$  is  $D_{50}$  (m). To solve the left-hand-side of Eq. (S-1), the dimensionless shear stress ( $\tau^*$ ) was estimated by

$$\tau^* = \frac{HS}{RD} \quad (\text{S-3})$$

where  $H$  is the water depth (m), and  $S$  is the slope (m/m). The water depth in Eq. S-3 can be estimated by using the Manning Equation defined as

$$Q = \frac{1}{n} A R_h^{2/3} S^{1/2} \quad (\text{S-4})$$

where  $Q$  is the river discharge (m<sup>3</sup>/s),  $A$  is the flow area (m<sup>2</sup>),  $R_h$  is the hydraulic radius (flow area divided by the wetted perimeter; m), and  $n$  is the Manning coefficient (m<sup>-1/3</sup>s). Assuming a rectangular channel and by the trial and error, the water depth can be obtained from the Manning equation in the form of

$$Q = \frac{1}{n} \frac{(BH)^{5/3}}{(B+2H)^{2/3}} S^{1/2} \quad (\text{S-5})$$

where  $B$  is the bottom width of the channel (in this study assumed equal to the flow width (W); m). The manning coefficient ( $n$ ) in Eq. S-5 was obtained from the Manning-Strickler Equation (Garcia, 2007; Hosseiny & Smith, 2019) defined as

$$n = \frac{D_{50}^{1/6}}{21.1} \quad (\text{S-6})$$

### Wong and –Parker (2006)

Wong and Parker (2006) reanalyzed the data used to develop the foundational Meyer-Peter and Muller (MPM) equation (Meyer-Peter & Müller, 1948) and found a better fit to data resulting in the following equation:

$$q^* = 3.97(\tau^* - 0.0495)^{3/2} \quad (\text{S-7})$$

where the  $\tau^*$  is the non-dimensional shear stress (Eq. S-3) and the exponent is fixed at 3/2. The MPM equation is similar in form, but tends to overpredict bedload at higher discharges (Barry et al., 2004). Experimentally, bedload flux is well-described by Eq. S-7 and similar models employ excess shear stress raised to a 3/2 power, however application within different rivers typically requires that both the coefficient and threshold shear stress be treated as fitting parameters (Mueller et al., 2005; Phillips & Jerolmack, 2019). Here, for the sake of comparison, we have applied this equation using fixed coefficient and thresholds shown in Eq. S-7 as it was not possible to estimate these parameters at each site in the database. Once  $q^*$  was obtained, using Eq. S-2 and measured  $D_{50}$ , the volumetric bedload rate ( $q_{bv}$ ) was estimated.

### Wilcock and -Crowe (2003)

Wilcock and Crowe (2003) is a physics-based transport law formulated to predict transport rates of individual grain size fractions,  $D_i$ , as a function of the grain size distribution and associated mobility of each fraction. The volumetric transport rate per unit width ( $\text{m}^2/\text{s}$ ) for each size fraction,  $q_{bvi}$ , is defined by the following function:

$$q_{bvi} = \frac{F_i u^{*3} q_i^*}{Rg} \quad (\text{S-8})$$

where  $F_i$  is the proportion of fraction  $i$  in surface size distribution (set to 0.16, 0.5, 0.84, and 0.9 for four measured sediment sizes used in this research),  $R$  is the dimensionless submerged specific gravity, assumed 1.65 in this study, and  $u^*$  is the shear velocity ( $\text{m/s}$ ) defined as

$$u^* = \sqrt{\tau/\rho} \quad (\text{S-9})$$

in which  $\tau$  is the shear stress ( $\text{N/m}^2$ ; Pa) and  $\rho$  is the water density ( $\text{kg/m}^3$ ).

We calculated the dimensionless bedload transport rate for each size fraction,  $q_i^*$  using

$$q_i^* = \begin{cases} 0.002 \varphi^{7.5} & \varphi < 1.35 \\ 14 \left(1 - \frac{0.894}{\sqrt{\varphi}}\right)^{4.5} & \varphi \geq 1.35 \end{cases} \quad (\text{S-10})$$

where  $\varphi$  is defined as

$$\varphi = \left( \frac{\tau}{\tau_{ri}} \right) \quad (\text{S-11})$$

in which  $\tau_{ri}$  is the reference shear stress of size fraction  $i$ , defined as

$$\tau_{ri} = \tau_{rm} \left( \frac{D_i}{D_{sm}} \right)^{b_i} \quad (\text{S-12})$$

where  $\tau_{rm}$  is the reference shear stress of mean size of bed surface,  $D_{sm}$  is the geometric mean sediment grain size (m) defined as

$$D_{sm} = \sqrt{D_{84}D_{16}} \quad (S-13)$$

and  $b_i$  is the exponent defined as

$$b_i = \frac{0.67}{1 + e^{(1.5 - \frac{D_i}{D_{sm}})}}, \quad (S-14)$$

We calculated the average shear stress applied on particles,  $\tau$  in Eq. S-11 (Pa), following Gaeuman et al., (2009)

$$\tau = 0.052 \rho (g S D_{65})^{0.25} U^{1.5} \quad (S-15)$$

where  $U$  is the cross sectionally averaged velocity (m/s), and  $D_{65}$  is the 65<sup>th</sup> percentile of the grain size distribution (m) (Gaeuman et al., 2009).  $D_{65}$  was assumed to be equal to the average of  $D_{50}$  and  $D_{84}$ . Cross sectionally averaged velocity,  $U$ , was estimated by dividing the discharge by the area of the flow as

$$U = \frac{Q}{BH} \quad (S-16)$$

where  $Q$  is the flow discharge (m<sup>3</sup>/s). We calculated the reference shear stress of mean size of bed surface,  $\tau_{rm}$  in Eq. S-12 (Pa) as

$$\tau_{rm} = \tau^*_{rm} R \rho g D_{sm}, \quad (S-17)$$

where  $\rho$  is the water density (kg/m<sup>3</sup>). To do that, we calculated the non-dimensional mean reference shear stress,  $\tau^*_{rm}$ , for the geometric mean sediment diameter,  $D_{sm}$  (m), as

$$\tau^*_{rm} = 0.021 + 0.015 e^{(-20Fs)}, \quad (S-18)$$

where  $Fs$  is the bed sand fraction. To estimate the bed sand fraction based on four measured sediment sizes ( $D_{16}$ ,  $D_{50}$ ,  $D_{84}$ , and  $D_{90}$ ), a hyperbolic tangent function was fitted to the four sediment sizes ( $Tanh_{fit}$ ) to generate a sediment grain size distribution model. This function was selected because it generally follows the common S-shape of the sediment grain size distribution. As such, the input to the  $Tanh_{fit}$  function was the sediment size and the output was the percent finer. The sand fraction (the fraction of the bed materials between sediment sizes of  $62 \times 10^{-6}$  m and  $2 \times 10^{-3}$  m) for each measured datapoint then was estimated by

$$F_s = Tanh_{fit}(2 \times 10^{-3}) - Tanh_{fit}(62 \times 10^{-6}) \quad (S-19)$$

In this research, the transport rates for four classes of  $D_{16}$ ,  $D_{50}$ ,  $D_{84}$ , and  $D_{90}$  were calculated and summed up as an estimate for the total bedload.

### Recking (2013)

Recking bedload model (2013) was developed based on the two equations that were developed by Recking (2010). The model can be used for sand and gravel mixtures and was developed based on 6,319 field observations and 1,317 flume measurements (Recking, 2010). In this model, the sediment size  $D_{84}$  is considered to control sediment mobility. Recking (2013) is formulated as

$$\phi = 14 \tau_{84}^{*2.5} / [1 + (\tau_m^* / \tau_{84}^*)^4] \quad (S-20)$$



where  $\phi$  is the Einstein dimensionless bedload parameter ( $\phi = qb_v / \sqrt{gRD_{84}^3}$ ) based on the volumetric bedload rate ( $qb_v$ ). Further,  $\tau_m^*$  is the non-dimensional mobility Shields stress related to the transition from partial to full mobility, and can be estimated by

$$\begin{cases} (5S + 0.06)(D_{84}/D_{50})^{4.4\sqrt{S}-1.5} & \text{for gravel} \\ 0.045 & \text{for sand} \end{cases} \quad (\text{S-21})$$

The non-dimensional Shields stress related to  $D_{84}$  ( $\tau_{84}^*$ ) in this method is characterized as

$$\tau_{84}^* = \frac{SR_h}{RD_{84}} = \frac{S}{RD_{84}[\frac{2}{W} + 74 p^{2.6} (gS)^p q^{-2p} D_{84}^{3p-1}]} \quad (\text{S-22})$$

where  $p = 0.23$  when  $q / \sqrt{gSD_{84}^3} < 100$  and  $p = 0.3$  otherwise, and  $q$  is the flow discharge per unit width (Recking, 2013).

Finally, the volumetric transport rate ( $qb_v$ ) can be converted into mass transport rate by multiplying it by the density of the sediment (assumed  $2650 \text{ kg/m}^3$  for natural sediment).

Table S3. A comparison between input variables into bedload models used in this study. Bolded variables represent direct measurements used from bedload.web.

Model Name	Input Variables from Bedload.web Database	Inputs Derived from Measured Data
Einstein	$Q, W, S, D_{50}$	$\tau^*, n, R_h, A, H$
Wong-Parker	$Q, W, S, D_{50}$	$\tau^*, n, R_h, A, H$
Wilcock-Crowe	$Q, W, S, D_{16}, D_{84}$	$F_i, u^*, q^*, \tau, \tau_{ri}, \tau_{rm}, D_{sm}, b, D_{65}, U, \tau_{rm}^*, F_S$
Recking 2013	$Q, W, S, D_{50}, D_{84}$	$\tau_m^*, \tau_{84}^*$
ANN	$Q, W, S, D_{16}, D_{50}, D_{84}, D_{90}$	-

Table S4. Statistical indices of error ratio (Eq. S-23) for different methods used to estimate bedload for sand and gravel bed rivers in the test data.

	Einstein	Wilcock and Crowe	Recking	Wong and Parker	ANN
Minimum Error Ratio	-1.00	-1.00	-1.00	-1.00	-0.93
Maximum Error Ratio	63.43	24620.18	365490.89	488.84	362.80
Mean Error Ratio	-0.07	202.52	5557.42	12.14	1.00
Std. Dev. Error Ratio	4.02	858.48	16203.00	43.10	13.39

Relative error (error ratio) was defined as

$$\text{Error Ratio} = \frac{(\text{prediction} - \text{measured})}{\text{measured}} \quad (\text{S-23})$$

Table S5- Statistical indices for bedload flux predictions ( $q_s$ ) for the sand and gravel bed rivers in the test data used in this study relative to measurements.

	<b>MAE log(data)</b>	<b>Mean</b>	<b>Standard Deviation</b>	<b>10th Percentile</b>	<b>25<sup>th</sup> Percentile</b>	<b>50<sup>th</sup> Percentile</b>	<b>75<sup>th</sup> Percentile</b>	<b>90<sup>th</sup> Percentile</b>	<b>Skewness</b>
<b>Observed</b>	N/A	31.6127	66.0306	0.4000	1.1881	5.1700	24.4565	94.0600	3.2194
<b>Einstein</b>	11.8459	31.0738	151.4993	0.0000	0.0000	0.0000	0.0458	61.7413	11.3005
<b>Wilcock and Crowe</b>	2.2295	855.7139	2317.9800	0.0000	0.0000	138.1887	934.8568	2477.8117	12.0938
<b>Recking</b>	6.5890	13278.4576	24011.4344	409.3240	1538.2942	5267.5620	14530.4607	30093.5401	4.5742
<b>Wong and Parker</b>	1.8128	104.2200	1424.9227	0.3434	1.9363	11.3479	45.3444	128.3584	23.4941
<b>ANN</b>	0.7045	25.5962	48.2562	0.5148	0.8810	4.6893	27.2363	78.2982	2.9634

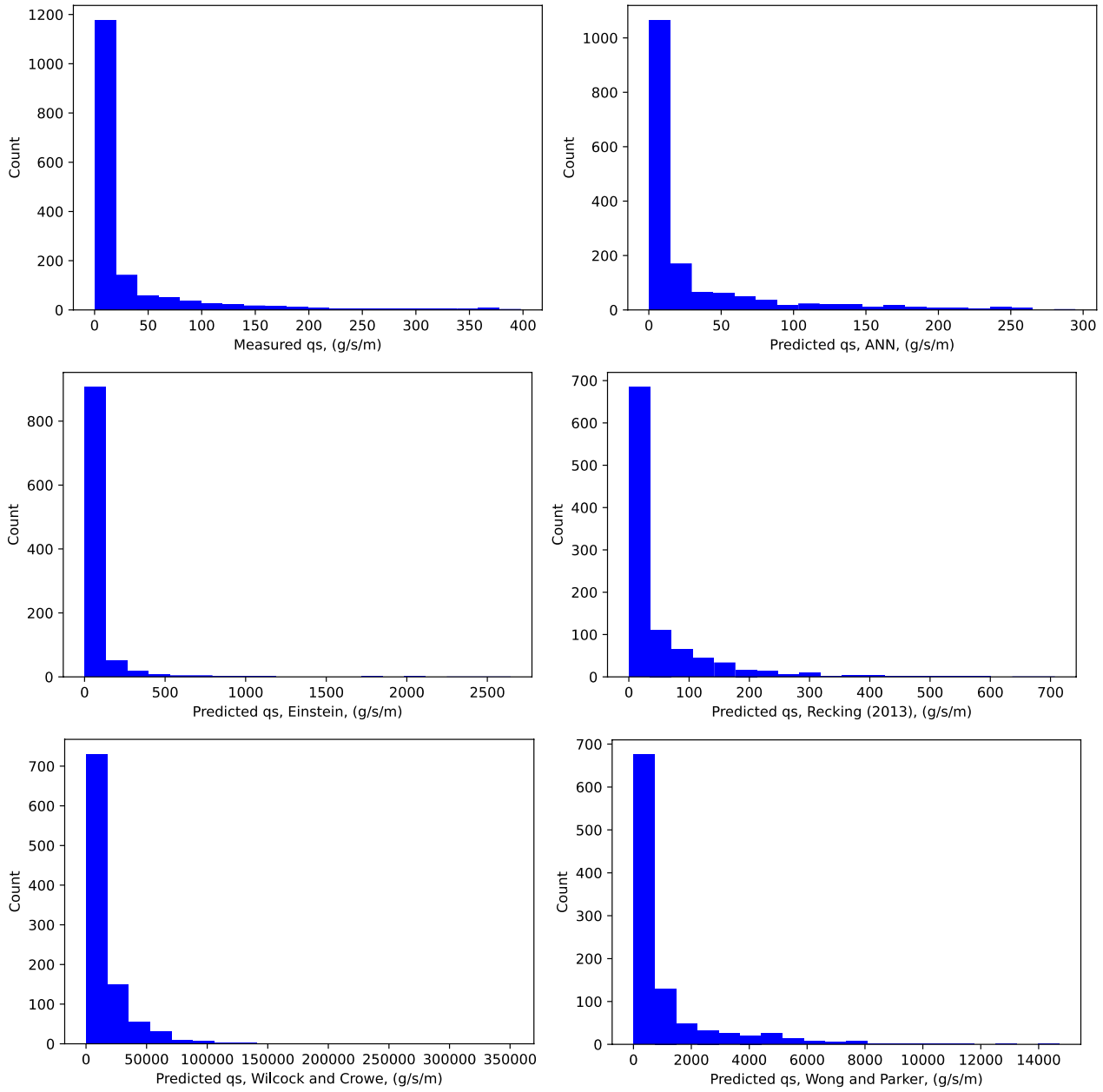


Figure S3- The distribution of the predictions for the sand and gravel bed rivers obtained from different models.

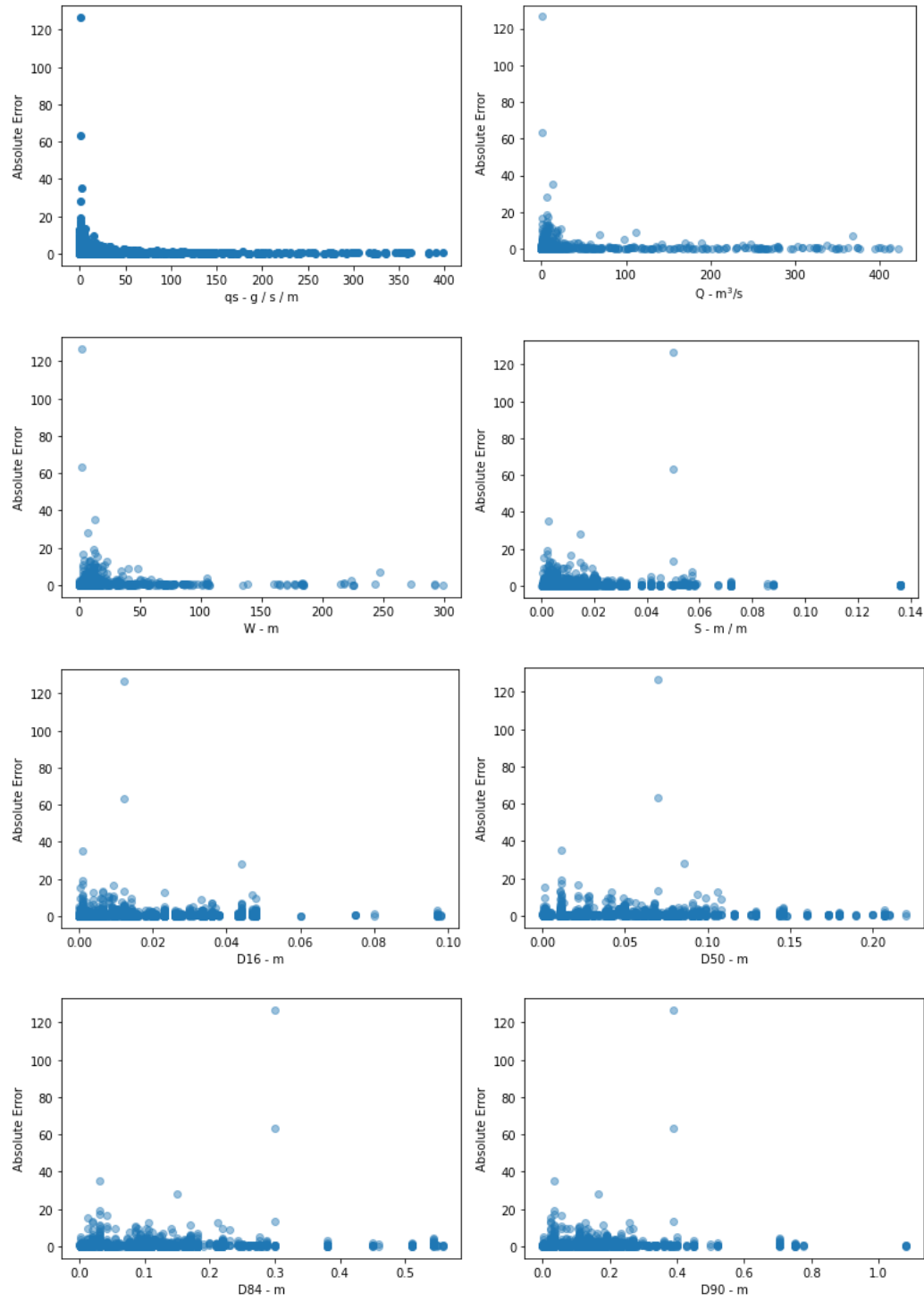


Figure S4. Variations in the absolute error ratio, Eq. S-23, of predicted bedload discharge obtained from the ANN model with the input variables

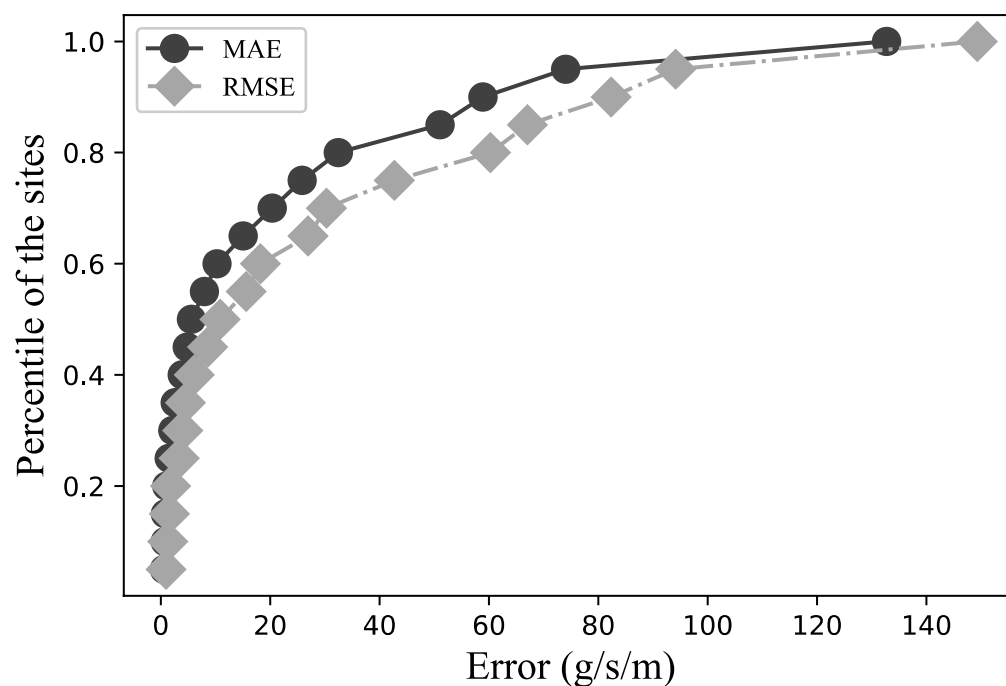


Figure S5. Variations in prediction errors obtained from trained ML model for all river sites used in this study.

## References

- Barry, J. J., Buffington, J. M., & King, J. G. (2004). A general power equation for predicting bed load transport rates in gravel bed rivers. *Water Resources Research*, 40(10), 1–22. <https://doi.org/10.1029/2004WR003190>
- Einstein, A. H. (1950). The Bed-Load Function for Sediment Transportation in Open Channel Flows. In *Soil Conservation Series, U.S. Dept. of Agriculture, Technical Bulletin* (Issue 1026). <https://naldc.nal.usda.gov/download/CAT86201017/PDF>
- Gaeuman, D., Andrews, E. D., Kraus, A., & Smith, W. (2009). Predicting fractional bed load transport rates: Application of the Wilcock-Crowe equations to a regulated gravel bed river. *Water Resources Research*, 45(6), 1–15. <https://doi.org/10.1029/2008WR007320>
- Garcia, M. H. (Ed.). (2007). *Sedimentation Engineering: Processes, Measurements, Modeling, and Practice* (No. 110). ASCE. <https://doi.org/10.1061/9780784408148.ch02>
- Hosseiny, H., & Smith, V. (2019). Two Dimensional Model for Backwater: Darby Creek, PA. *Water*, 11(2204). <https://www.mdpi.com/2073-4441/11/11/2204>
- Meyer-Peter, E., & Müller, R. (1948). Formulas for bed load transport. *2nd Meeting, Int. Assoc. for Hydroaul. Environ. Eng. and Res.*, 39–64.

- Mueller, E. R., Pitlick, J., & Nelson, J. M. (2005). Variation in the reference Shields stress for bed load transport in gravel-bed streams and rivers. *Water Resources Research*, 41(W04006). <https://doi.org/10.1029/2004WR003692>
- Phillips, C. B., & Jerolmack, D. J. (2019). Bankfull Transport Capacity and the Threshold of Motion in Coarse-Grained Rivers. *Water Resources Research*, 55(12), 11316–11330. <https://doi.org/10.1029/2019WR025455>
- Recking, A. (2010). A comparison between flume and field bed load data and consequences for surface-based bed load transport prediction. *Water Resources Research*, 46(W03518). <https://doi.org/10.1029/2009WR008007>
- Recking, A. (2013). Simple Method for Calculating Reach-Averaged Bed-Load Transport. *Journal of Hydraulic Engineering*, 139(1), 70–75. [https://doi.org/10.1061/\(asce\)hy.1943-7900.0000653](https://doi.org/10.1061/(asce)hy.1943-7900.0000653)
- Wilcock, P. R., & Crowe, J. C. (2003). Surface-based Transport Model for Mixed-Size Sediment. *Journal of Hydraulic Engineering*, 9429. [https://doi.org/10.1061/\(ASCE\)0733-9429\(2003\)129](https://doi.org/10.1061/(ASCE)0733-9429(2003)129)
- Wong, M., & Parker, G. (2006). Reanalysis and Correction of Bed-Load Relation of Meyer-Peter and Müller Using Their Own Database. *Journal of Hydraulic Engineering*, 132(11), 1159–1168. [https://doi.org/10.1061/\(asce\)0733-9429\(2006\)132:11\(1159\)](https://doi.org/10.1061/(asce)0733-9429(2006)132:11(1159))