Earth Surface Dynamics

Supporting information for

Development of a machine learning model for river bedload

Hossein Hosseiny[1], Claire Masteller[1], and Colin Phillips[2]

[1]Department of Earth and Planetary Sciences, Washington University in St. Louis, St. Louis, MO, 63130, USA

[2]Department of Civil and Environmental Engineering, Utah State University, Logan, UT, USA.

Correspondence to: Hossein Hosseiny (syd.hosseini@gmail.com)

Contents of this file

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).

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).

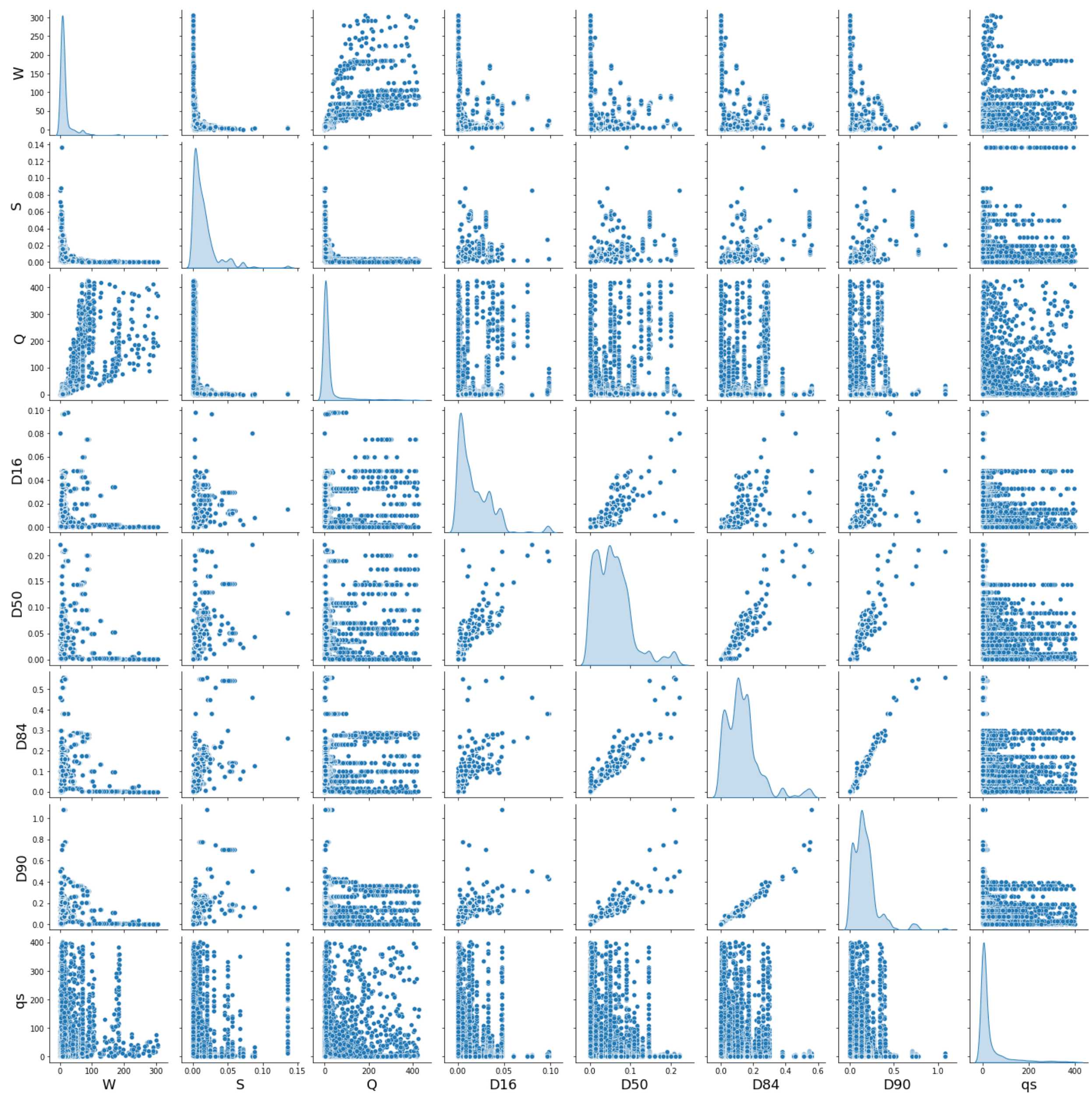

Figure S1. Illustration of the original data used in this study.

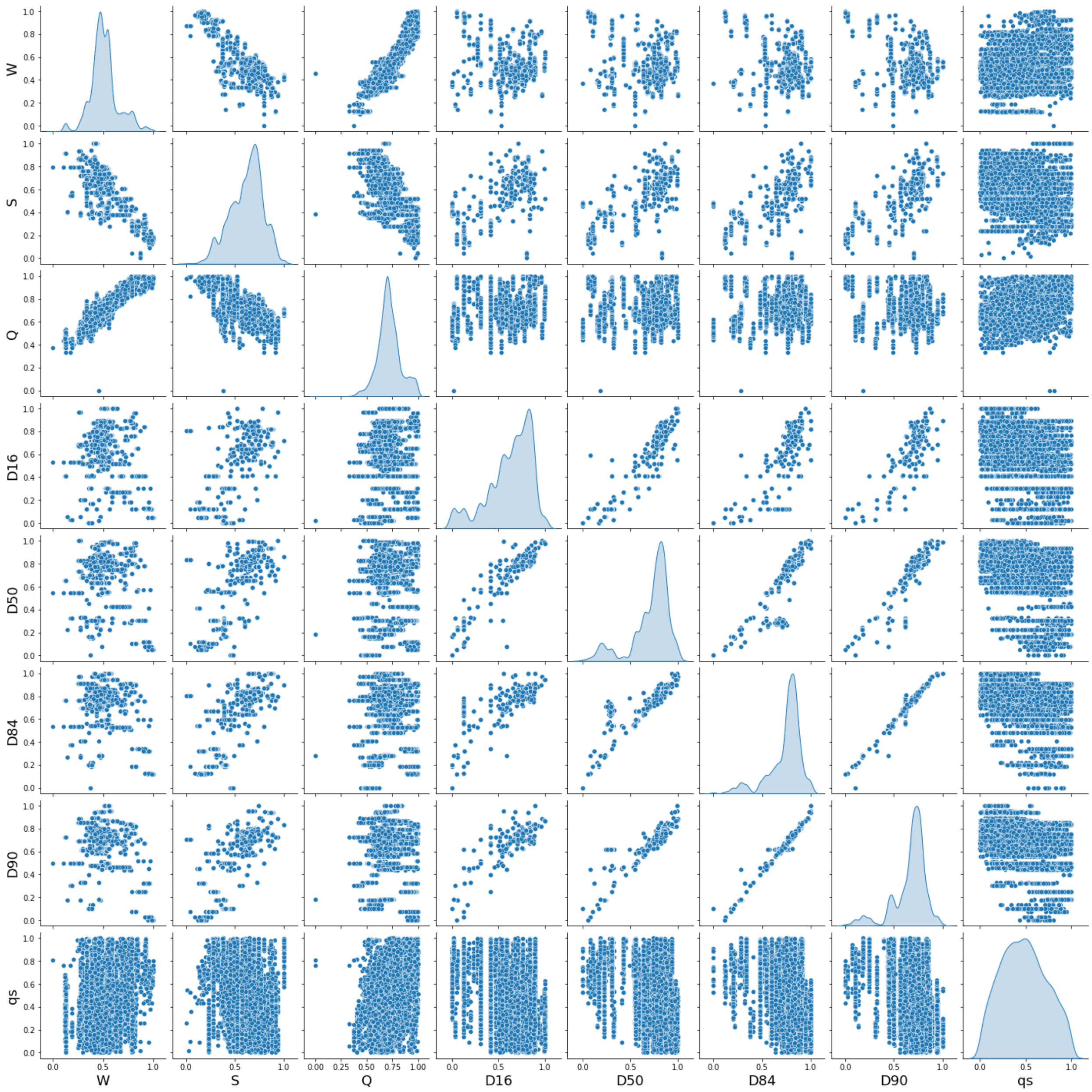

Figure S2. Illustration of the log-transformed, normalized data used in this study.

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

Index	W (m)	S (m / m)	Q (m^3/s)	D_{16} (m)	D_{50} (m)	D_{84} (m)	D_{90} (m)	q_s (g / s / m)
mean	18.1	0.01529	23.7	0.0176	0.0568	0.1321	0.1671	31.9
std	30.3	0.01813	61.1	0.0184	0.0451	0.1055	0.1435	65.8
min	0.3	0.00009	0.00005	0.0001	0.0003	0.0004	0.0011	0.1
max	306.0	0.13600	427.5	0.0980	0.2200	0.5580	1.0800	401.0
25%	6.1	0.00380	1.7	0.0040	0.0220	0.0550	0.0754	1.1
50%	9.0	0.01000	4.0	0.0120	0.0500	0.1120	0.1400	5.1
75%	14.6	0.01900	11.3	0.0280	0.0790	0.1700	0.2171	24.8
skew (original data)	4.8	2.8	3.9	1.8	1.2	1.6	2.3	3.1
skew (log-scaled)	0.5	-0.4	0.1	-0.9	-1.4	-1.5	-1.5	0.1

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

Variable	W	S	Q	D_{16}	D_{50}	D_{84}	D_{90}	q_s
W	1.00	-0.29	0.76	-0.07	-0.07	-0.10	-0.10	0.17
S	-0.29	1.00	-0.25	0.09	0.23	0.36	0.37	-0.09
Q	0.76	-0.25	1.00	0.06	0.10	0.05	0.03	0.22
D_{16}	-0.07	0.09	0.06	1.00	0.79	0.58	0.51	-0.23
D_{50}	-0.07	0.23	0.10	0.79	1.00	0.89	0.86	-0.26
D_{84}	-0.10	0.36	0.05	0.58	0.89	1.00	0.98	-0.23
D_{90}	-0.10	0.37	0.03	0.51	0.86	0.98	1.00	-0.22
qs	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^2/s), the Einstein model was used as,

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

where τ^* 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

$$q^* = \frac{q_{bv}}{\boldsymbol{D}\sqrt{gR\boldsymbol{D}}} \tag{S-2}$$

where q_{bv} is the volumetric bedload transport rate (m^2/s), g is the gravitational acceleration (m/s^2), 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 (τ^*) was estimated by

$$\tau^* = \frac{H\boldsymbol{S}}{RD} \tag{S-3}$$

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

$$Q = \frac{1}{n}\, A\, {R_h}^{2/3} S^{1/2} \tag{S-4}$$

where Q is the river discharge (m^3/s), A is the flow area (m^2), R_h is the hydraulic radius (flow area divided by the wetted perimeter; m), and n is the Manning coefficient (m$^{-1/3}$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

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

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

$$n = \frac{{D_{50}}^{\frac{1}{6}}}{21.1} \tag{S-6}$$

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:

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

where the τ^* 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 (m^2/s) for each size fraction, q_{bvi}, is defined by the following function:

$$q_{bvi} = \frac{F_i\, {u^*}^3 {q^*}_i}{Rg} \qquad \text{(S-8)}$$

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 (m/s) defined as

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

in which τ is the shear stress (N/m^2; Pa) and ρ is the water density (kg/m^3).

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

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

where φ is defined as

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

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

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

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

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

and b_i is the exponent defined as

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

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

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

where U is the cross sectionally averaged velocity (m/s), and D_{65} is the 65th 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

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

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

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

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

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

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 62x10E-6 m and 2x10E-3 m) for each measured datapoint then was estimated by

$$F_S = Tanh_{fit}(2x10E-3) - Tanh_{fit}(62x10E-6) \quad \text{(S-19)}$$

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

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

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

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

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

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

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

Finally, the volumetric transport rate (q_{bv}) can be converted into mass transport rate by multiplying it by the density of the sediment (assumed 2650 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, D₅₀***	τ^*, n, R_h, A, H
Wong-Parker	*Q, W,* ***S, D₅₀***	τ^*, n, R_h, A, H
Wilcock-Crowe	*Q, W,* ***S, D₁₆, D₈₄***	$F_i, u^*, q^*, \tau, \tau_{ri}, \tau_{rm}, D_{sm}, b, D_{65}, \mathrm{U}, \tau^*{}_{rm}, F_S$
Recking 2013	*Q, W,* ***S, D₅₀, D₈₄***	τ_m^*, τ_{84}^*
ANN	***Q, W, S, D₁₆, D₅₀, D₈₄, D₉₀***	-

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.

Method	Min	Max	Std	Mean	Mean Absolute
Einstein	-1.0	133.2	7.0	0.3	1.9
Wong and Parker	-1.0	24620.1	1012.6	223.9	224.5
Wilcock and Crowe	-0.9	328482.3	17702.5	5441.9	5441.9
Recking (2013)	-0.9	169.6	13.2	4.0	4.7
ANN	-0.9	35.4	2.4	0.7	1.1

Relative error (error ratio) was defined as

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

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.

Method	Average (g/s/m)	Std (g/s/m)	90% (g/s/m)	75% (g/s/m)	50% (g/s/m)	25% (g/s/m)	10% (g/s/m)	Skew	MAE (g/s/m)	RMSE (g/s/m)
Measured Data	42.04	74.40	133.55	41.47	9.46	2.00	0.60	2.63	-	-
ANN	39.55	60.10	132.10	53.95	11.06	1.84	0.62	2.03	16.51	44.00
Einstein	48.38	202.20	126.20	5.29	7.35E-05	0.00	0.00	8.67	54.98	189.78
Wong and Parker	1044.66	1832.19	3535.31	1138.42	316.99	0.00	0.00	2.98	1007.58	2073.96
Wilcock and Crowe	16377.56	28458.14	39188.88	18390.10	8017.39	2515.76	662.07	5.34	16335.69	32802.60
Recking (2013)	46.54	81.09	138.43	55.03	13.50	2.02	0.37	3.39	40.30	78.83

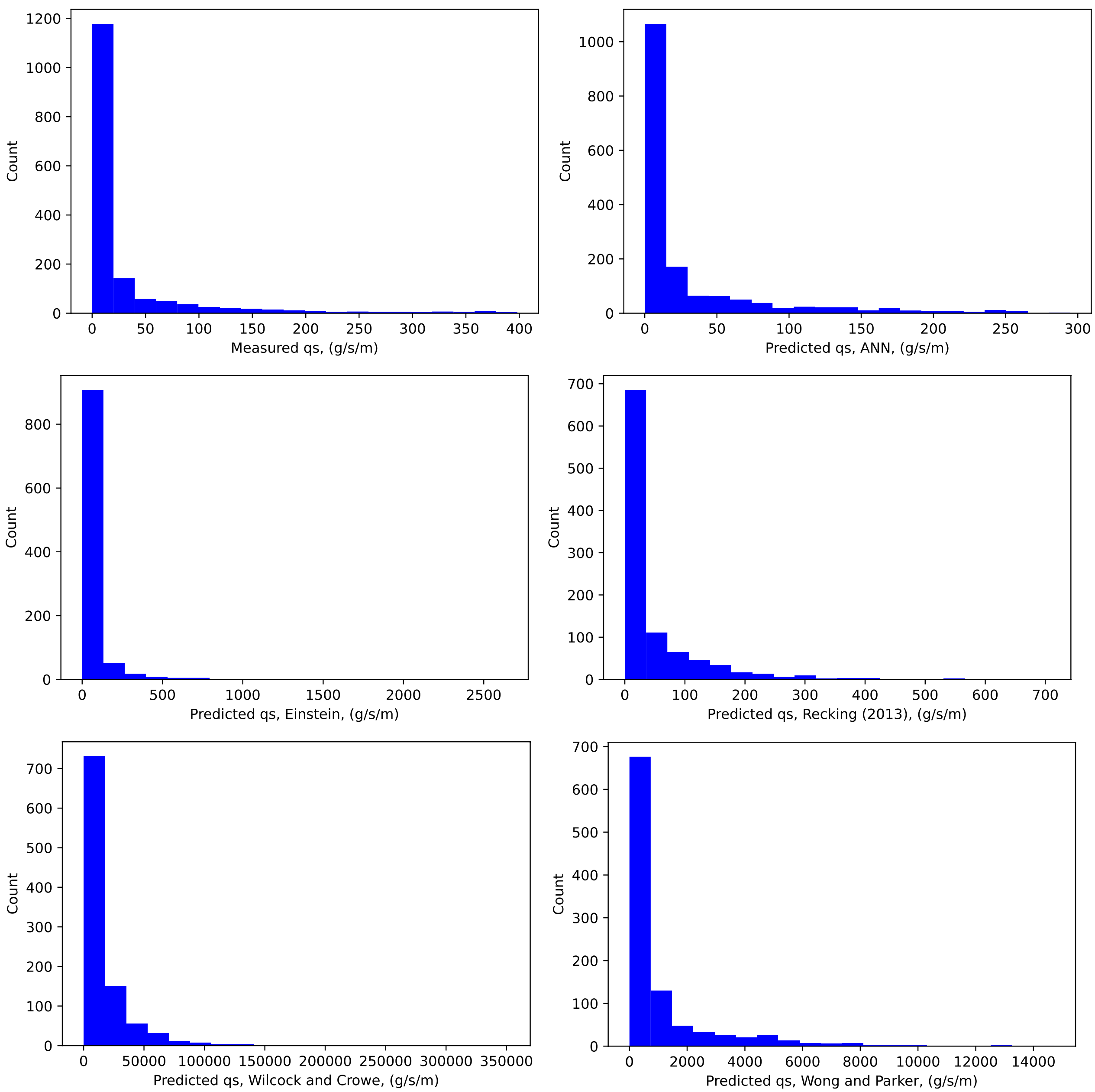

Figure S3- The distribution of the predictions for the sand and gravel bed rivers obtained from different models.

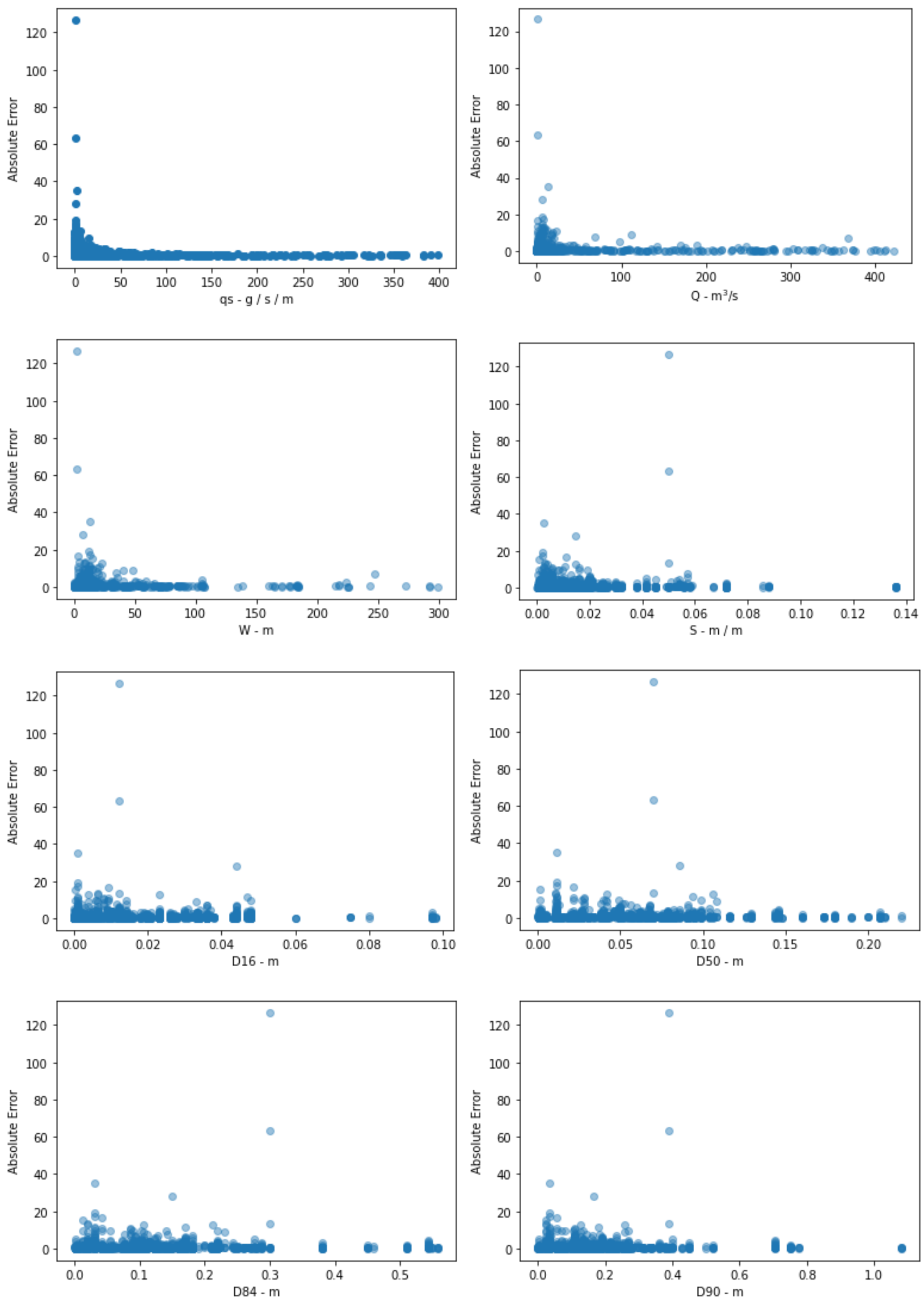

Figure S4. Variations in the absolute error ratio, Eq. S-23, of predicted bedload discharge obtained from the ANN model with the input variables

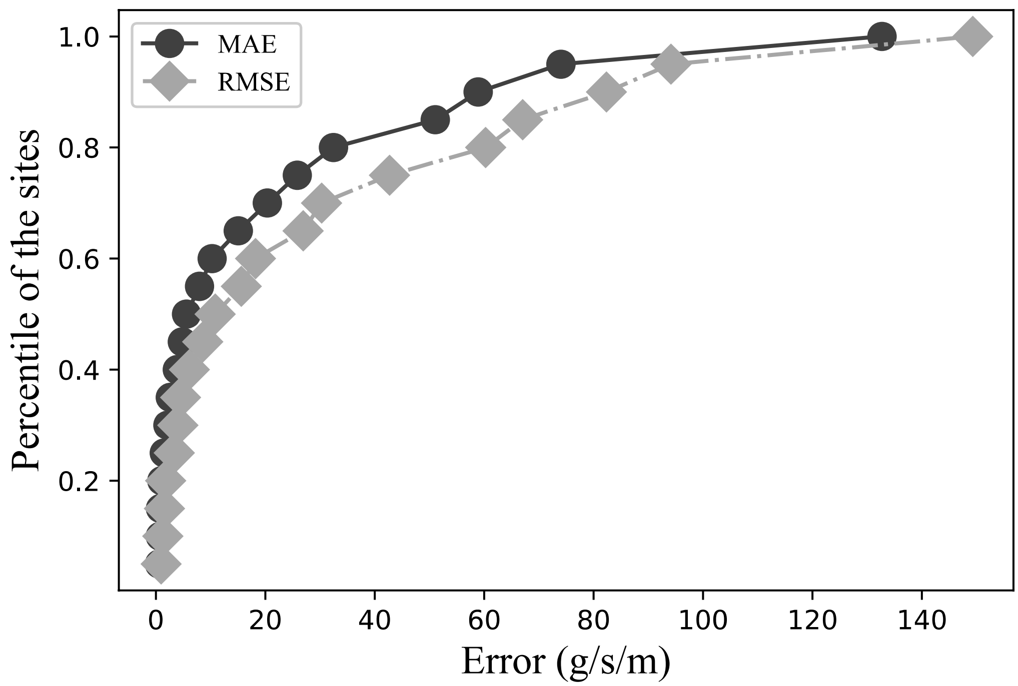

Figure S5. Variations in prediction errors obtained from trained ML model for all river sites used in this study.

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/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

Wilcock, P. R., & Crowe, J. C. (2003). Surface-based Transport Model for Mixed-Size Sediment Surface-based Transport Model for Mixed-Size Sediment. *Journal of Hydraulic Engineering*, *9429*. 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)