Development of a machine learning model for river bedload
Hossein Hosseiny¹, Claire Masteller¹, and Colin Phillips²

¹Department of Earth and Planetary Sciences, Washington University in St. Louis, St. Louis, MO, 63130, USA
²Department of Civil and Environmental Engineering, Utah State University, Logan, UT, USA.

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

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).

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.

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

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

<table>
<thead>
<tr>
<th>Variable</th>
<th>(W)</th>
<th>(S)</th>
<th>(Q)</th>
<th>(D_{16})</th>
<th>(D_{50})</th>
<th>(D_{84})</th>
<th>(D_{90})</th>
<th>(q_s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(W)</td>
<td>1.00</td>
<td>-0.29</td>
<td>0.76</td>
<td>-0.07</td>
<td>-0.07</td>
<td>-0.10</td>
<td>-0.10</td>
<td>0.17</td>
</tr>
<tr>
<td>(S)</td>
<td>-0.29</td>
<td>1.00</td>
<td>-0.25</td>
<td>0.09</td>
<td>0.23</td>
<td>0.36</td>
<td>0.37</td>
<td>-0.09</td>
</tr>
<tr>
<td>(Q)</td>
<td>0.76</td>
<td>-0.25</td>
<td>1.00</td>
<td>0.06</td>
<td>0.10</td>
<td>0.05</td>
<td>0.03</td>
<td>0.22</td>
</tr>
<tr>
<td>(D_{16})</td>
<td>-0.07</td>
<td>0.09</td>
<td>0.06</td>
<td>1.00</td>
<td>0.79</td>
<td>0.58</td>
<td>0.51</td>
<td>-0.23</td>
</tr>
<tr>
<td>(D_{50})</td>
<td>-0.07</td>
<td>0.23</td>
<td>0.10</td>
<td>0.79</td>
<td>1.00</td>
<td>0.89</td>
<td>0.86</td>
<td>-0.26</td>
</tr>
<tr>
<td>(D_{84})</td>
<td>-0.10</td>
<td>0.36</td>
<td>0.05</td>
<td>0.58</td>
<td>0.89</td>
<td>1.00</td>
<td>0.98</td>
<td>-0.23</td>
</tr>
<tr>
<td>(D_{90})</td>
<td>-0.10</td>
<td>0.37</td>
<td>0.03</td>
<td>0.51</td>
<td>0.86</td>
<td>0.98</td>
<td>1.00</td>
<td>-0.22</td>
</tr>
<tr>
<td>(q_s)</td>
<td>0.17</td>
<td>-0.09</td>
<td>0.22</td>
<td>-0.23</td>
<td>-0.26</td>
<td>-0.23</td>
<td>-0.22</td>
<td>1.00</td>
</tr>
</tbody>
</table>
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,

\[
1 - \frac{1}{\sqrt{\pi}} \int_{-0.413/\tau^*}^{0.413/\tau^*} e^{-t^2} dt = \frac{43.5 q^*}{1 + 43.5 q^*}
\]

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

\[
q^* = \frac{q_{bv} D^{1/2} g R}{D^{1/2}}
\]

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 (\( \tau^* \)) was estimated by

\[
\tau^* = \frac{H S}{R D}
\]

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

\[
Q = \frac{1}{n} A R_h^{2/3} S^{1/2}
\]

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

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

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

\[
n = \frac{D_{50}^{1/2}}{21.1}
\]
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:

\[ q^* = 3.97 (\tau^* - 0.0495)^{3/2} \]  

(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) 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, \( q_{bv i} \), is defined by the following function:

\[ q_{bv i} = \frac{F_i u^3 q_i^*}{\rho g} \]  

(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

\[ u^* = \sqrt{\frac{\tau}{\rho}} \]  

(S-9)

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

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

\[ q_i^* = \begin{cases} 
0.002 \phi^{7.5} & \phi < 1.35 \\
14 \left(1 - \frac{0.894}{\sqrt{\phi}}\right)^{4.5} & \phi \geq 1.35 
\end{cases} \]  

(S-10)

where \( \phi \) is defined as

\[ \phi = \left(\frac{\tau}{\tau_{ri}}\right) \]  

(S-11)

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

\[ \tau_{ri} = \tau_{rm} \left(\frac{D_i}{D_{sm}}\right)^{b_i} \]  

(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
\[ D_{sm} = \sqrt{D_{84}D_{16}} \]  
(S-13)

and \( b_l \) is the exponent defined as

\[ b_l = \frac{0.67}{1 + e^{(1.5 - \frac{D_{l}}{D_{sm}})}} \]  
(S-14)

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

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

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

\[ U = \frac{Q}{BH} \]  
(S-16)

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

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

where \( \rho \) 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

\[ \tau^*_{rm} = 0.021 + 0.015 e^{(-20F_s)} \]  
(S-18)

where \( F_s \) 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 62x10\(^{-6}\) m and 2x10\(^{-3}\) m) for each measured datapoint then was estimated by

\[ F_s = Tanh_{fit}(2x10E - 3) - Tanh_{fit}(62x10E - 6) \]  
(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

\[ \phi = 14 \frac{\tau^*_{84}^{2.5}}{[1 + (\tau^*_{m}/\tau^*_{84})^4]} \]  
(S-20)
where $\phi$ is the Einstein dimensionless bedload parameter ($\phi = \frac{q b_v}{\sqrt{g R D_{84}^3}}$) 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

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

(S-21)

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

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

(S-22)

where $p = 0.23$ when $q / \sqrt{g S D_{84}^3} < 100$ 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 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.

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

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.

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

Relative error (error ratio) was defined as

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

(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.

<table>
<thead>
<tr>
<th>Method</th>
<th>Average (g/s/m)</th>
<th>Std (g/s/m)</th>
<th>90% (g/s/m)</th>
<th>75% (g/s/m)</th>
<th>50% (g/s/m)</th>
<th>25% (g/s/m)</th>
<th>10% (g/s/m)</th>
<th>Skew</th>
<th>MAE (g/s/m)</th>
<th>RMSE (g/s/m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measured Data</td>
<td>42.04</td>
<td>74.40</td>
<td>133.55</td>
<td>41.47</td>
<td>9.46</td>
<td>2.00</td>
<td>0.60</td>
<td>2.63</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ANN</td>
<td>39.55</td>
<td>60.10</td>
<td>132.10</td>
<td>53.95</td>
<td>11.06</td>
<td>1.84</td>
<td>0.62</td>
<td>2.03</td>
<td>16.51</td>
<td>44.00</td>
</tr>
<tr>
<td>Einstein</td>
<td>48.38</td>
<td>202.20</td>
<td>126.20</td>
<td>5.29</td>
<td>7.35E-05</td>
<td>0.00</td>
<td>0.00</td>
<td>8.67</td>
<td>54.98</td>
<td>189.78</td>
</tr>
<tr>
<td>Wong and Parker</td>
<td>1044.66</td>
<td>1832.19</td>
<td>3535.31</td>
<td>1138.42</td>
<td>316.99</td>
<td>0.00</td>
<td>0.00</td>
<td>2.98</td>
<td>1007.58</td>
<td>2073.96</td>
</tr>
<tr>
<td>Wilcock and Crowe</td>
<td>16377.56</td>
<td>28458.1</td>
<td>39188.8</td>
<td>18390.1</td>
<td>8017.39</td>
<td>2515.76</td>
<td>662.07</td>
<td>5.34</td>
<td>16335.6</td>
<td>32802.60</td>
</tr>
<tr>
<td>Recking (2013)</td>
<td>46.54</td>
<td>81.09</td>
<td>138.43</td>
<td>55.03</td>
<td>13.50</td>
<td>2.02</td>
<td>0.37</td>
<td>3.39</td>
<td>40.30</td>
<td>78.83</td>
</tr>
</tbody>
</table>
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


