Using machine learning to improve predictions and provide insight into fluvial sediment transport

J. William Lund1,2 | Joel T. Groten1 | Diana L. Karwan2 | Chad Babcock2

Abstract
A thorough understanding of fluvial sediment transport is critical to addressing many environmental concerns such as exacerbated flooding, degradation of aquatic habitat, excess nutrients, and the economic challenges of restoring aquatic systems. Fluvial sediment samples are integral for addressing these environmental concerns but cannot be collected at every river and time of interest. Therefore, to gain a better understanding for rivers where direct measurements have not been made, extreme gradient boosting machine learning (ML) models were developed and trained to predict suspended sediment and bedload from sampling data collected in Minnesota, United States (U.S.), by the U.S. Geological Survey. Approximately 400 watershed (full upstream area), catchment (nearby landscape), near-channel, channel, and streamflow features were retrieved or developed from multiple sources, reduced to approximately 30 uncorrelated features, and used in the final ML models. The results indicate suspended sediment and bedload ML models explain approximately 70% of the variance in the datasets. Important features used in the models were interpreted with Shapley additive explanation (SHAP) plots, which provided insight into sediment transport processes. The most important features in the models were developed to normalize streamflow by the 2-year recurrence interval and quantify the rate of change in streamflow (slope), which helped account for sediment hysteresis. Generally, this study also showed a combination of mostly watershed and catchment geospatial features were important in ML models that predict sediment transport from physical samples. This study is a promising step forward in making fluvial sediment transport predictions using machine learning models trained by physically collected samples. The approach developed here can be used wherever similar datasets exists and will be useful for landscape and water management.

KEYWORDS
bedload, hysteresis, machine learning, sediment rating curve, sediment transport, suspended sediment

1 | INTRODUCTION

An understanding of fluvial sediment processes and transport rates is important for managing rivers and their surrounding watersheds. However, fluvial sediment transport is difficult to understand because of the multitude of factors controlling the potential sources, delivery, mechanics, and storage of sediment in aquatic systems. Furthermore, sediment is difficult to manage because an excess or limited supply of...
sediment can shift a river into disequilibrium. Excess suspended sediment can impair rivers by adversely affecting aquatic habitats, degrading water quality, transporting harmful contaminants, diminishing recreational opportunities, and depositing sediment in navigable waterways (Alexander et al., 2012; Minnesota Pollution Control Agency, 2021a). In contrast, dams and other river modifications can reduce sediment transport and contribute to loss of native fish species and riparian ecosystems, subsidence and loss of wetlands, and decreased nutrient delivery to downstream receiving waters (Drauté et al., 2011; Kondolf, 1997; Schmidt & Wilcock, 2008). Therefore, accurate and cost-effective estimates of sediment loading are needed to manage riverine sediment transport at a multitude of scales (Ellison et al., 2016); also needed are methods to estimate sediment transport at sites where little or no physical samples have been collected (Gray & Simões, 2008).

Physically collected sediment samples provide the most accurate data to inform understanding of fluvial sediment processes and transport. The most accurate suspended sediment sampling methods are equal-width increment (EWI) or equal-discharge increment (EDI) and use depth integrating, isokinetic samplers (Davis, 2005) to represent nearly the entire water column in a river cross section (Edwards & Glysson, 1999). The most accurate laboratory method analyzes the entire suspended sample for suspended-sediment concentration (SSC; American Society for Testing and Materials, 2000; Guy, 1969) and accounts for both fines and sands. Pressure difference bedload (BL) samples collect larger sized particles (sands, gravel, and cobbles) moving along the bottom of the water column, and one method used to sample BL transport is the single EWI (SEWI; Edwards & Glysson, 1999). However, EWI and EDI sampling analysed for SSC, and SEWI analysed for BL is time-consuming and costly, requiring specialized training and equipment to collect samples. In fact, the number of U.S. Geological Survey (USGS) daily-record sediment-monitoring stations was reduced by more than 67% from 1981 to 2005 (Larsen et al., 2010). The number of BL stations is even fewer (U.S. Geological Survey, 2021b).

Less accurate grab samples only represent the top of the water column at one location and are analysed for total suspended solids (TSS; Clesceri et al., 1999). Unlike SSC, TSS is determined from an extracted sub-sample. These methods are inexpensive and faster to collect, require less specialized equipment and training, and meet regulatory requirements for evaluating sediment-related impairments established by the U.S. Environmental Protection Agency (Minnesota Pollution Control Agency, 2021b). However, TSS methods have been shown to substantially underestimate the suspended sand component in streams and rivers (Ellison et al., 2014; Gray et al., 2000; Groten & Johnson, 2018). TSS methods can only inform understanding of suspended fines while EWI and EDI analysed for SSC and SEWI for BL transport can inform the transport and processes for total sediment load (fines, sands, gravels, and cobbles) in rivers.

Sediment rating curves (SRCs) are empirical relations developed between streamflow and a set of discrete physical sediment samples and are used to provide estimates of sediment transport and loads when samples are unavailable. Past research has found that correlations between the SRCs’ slope and intercept parameters, river basin morphology, and climate helped define physical controls on sediment loads in rivers (Syvitski et al., 2000). More recent research has used machine learning (ML) and geospatial datasets to predict SRCs’ parameters in order to make inferences on sediment transport controls (e.g., Vaughan, Belmont, et al., 2017) or to make predictions at ungauged locations (e.g., Atieh et al., 2015). Having accurate estimates of the total sediment will provide insight on the processes controlling transport to help diagnose and restore fluvial systems.

Bankfull streamflows are the most geomorphically active streamflows in the channel before streamflow spills over its banks and loses energy to the floodplain (Biedenharn et al., 2008; Lane, 1955). However, bankfull streamflows can be difficult to determine, and a thorough site evaluation is often needed (Leopold et al., 1964; Rosgen, 1994, 1996). Generally, bankfull streamflows have a recurrence interval (RI) range from 1 to 2 years depending on the site (Simon et al., 2004). Previous sediment transport studies have used bankfull streamflow to normalize data while developing dimensionless sediment rating curves (DSRCs; Ellison et al., 2016; Rosgen, 2010).

Streamflow and sediment transport are not always strongly correlated, and hysteresis in the relation can cause inaccurate predictions. Hysteresis occurs when the relation between sediment and streamflow changes based upon the history of the system such that different measured values of suspended sediment or BL can occur in the same stream at the same streamflow but at different times. There are multiple types of hysteresis, and only a selection of the possible types will be presented. The two most common types of hysteresis are clockwise (type-1) and counterclockwise (type-3), with clockwise hysteresis being more common (Gells, 2013). With clockwise hysteresis there are higher values of sediment transport on the rising limb than the falling limb of the hydrograph. Clockwise hysteresis can be caused by a source of in-stream sediment that is readily mobilized during the rising limb of the hydrograph and corresponding increase in shear stress, which then becomes exhausted on the falling limb as sources are depleted (Gells, 2013; Smith & Dragovich, 2009). Alternatively, counterclockwise hysteresis is the opposite and can be caused by a delayed delivery from a sediment source. The source of sediment can be made available from an upstream tributary or due to a saturated riverbank collapse after the river receded (Gells, 2013; Kelly & Belmont, 2018). Therefore, SRCs should be carefully examined for stream locations with available data before applying methods to estimate sediment at stream locations of interest that have similar watershed characteristics and lack physical data. Some studies have tried to account for hysteresis in SRCs by using categorical variables to classify the rising and falling limbs or developing a SRC based on where the sample was collected on the streamflow hydrograph (Asselman, 2000; Vaughan, Belmont, et al., 2017).

The application of different ML approaches to estimate sediment transport has grown over the past two decades (Afan et al., 2016). ML has multiple benefits over traditional approaches, such as SRCs, with increased prediction accuracy of suspended sediment at specific sites while having the ability to learn complex linear and non-linear relations (Cisty et al., 2021; Francke et al., 2008; Khan et al., 2021; Zounemat-Kermani et al., 2020). Another benefit of ML is the ability to interpret these complex relations with the important features used
in the model (Breiman, 2001; Cutler et al., 2007). Vaughan, Belmont, et al. (2017) found that random forest ML models used to predict SRCs' parameters indicated that watershed features corresponded to TSS concentrations at low and moderate streamflows while near-channel features were strongly related to the SRCs' slopes. The study showed the utility of ML to distinguish between complex watershed and near-channel features to inform TSS SRCs' parameters at selected sites.

Random forest and extreme gradient boosting (XGBoost) both belong to the ensemble tree family of supervised ML models.

FIGURE 1 Location of the study area and sampling sites in the State of Minnesota, United States of America
(Breiman, 2001; Chen & Guestrin, 2016). Machine learning uses computer algorithms to learn complex interactions among linear and nonlinear data without the user needing to program the exact interaction (Bortnik & Camporeale, 2021). Random forest and XGBoost models learn by building many decision trees (learners) on random subsets of the full dataset and combine them to estimate the target outcome in a repeated process (ensembles). Random forest and XGBoost models can use many features as input to make predictions due to the averaging of trees which reduces the risk of overfitting while being impervious to noise (Bortnik & Camporeale, 2021; Breiman, 2001; Fox et al., 2017; Hastie et al., 2009). Both models can calculate permutation feature importance by removing features from the model to calculate if the model error increases when the feature was omitted (Breiman, 2001). Recent advances in data science incorporate game theory techniques to compute Shapley additive explanations (SHAP) values (Lundberg & Lee, 2017) which allocate credit across all marginal contributions and model ensembles. SHAP values can be used to show what the models have learned from the training datasets and how predictions were made (Molnar, 2019). XGBoost was chosen over random forest because at the time of writing, the authors could not find an efficient way to calculate SHAP values for random forest models in R (R Core Team, 2021).

There were three primary objectives of this study. The first objective was to use representative physical samples, streamflow, and publicly available geospatial datasets that describe watershed, catchment, near-channel, and channel features to develop methods to provide estimates of total sediment at stream locations where little or no physical samples have been collected. The second objective was to develop features from dimensionless streamflow to better account for hysteresis in the streamflow-sediment relation because streamflow data are more commonly available than sediment data. The third objective was to interpret XGBoost models with SHAP values to assess how predictions were made while making connections to known processes controlling sediment transport.

1.1 Study area

Minnesota has a complex glacial history, which resulted in diverse landforms and surface water conditions (Figure 1; Ellison et al., 2016; Ojakangas & Matsch, 1982; Sims & Morey, 1972). The southwest (SW) region received drained water from glacial Lake Agassiz (not shown) approximately 10 000 years ago which resulted in incised valleys and highly erodible knickpoints that influence the region’s current sediment regimes (Gran et al., 2009; Minnesota Pollution Control Agency, 2011). The southeast (SE) karst region was relatively untouched by the glaciers and has higher relief than the other regions of Minnesota (Figure 1; Lively, 2020). The predominantly forested northeast (NE) region has shallow bedrock, and steeper gradient rivers flow toward Lake Superior (Ojakangas & Matsch, 1982; Sims & Morey, 1972). Lakes and wetlands dominate the middle (MID) region, while intensively cultivated lands cover south, western, and northwest (NW) regions (Ellison et al., 2016). Overall Minnesota has diverse landscapes and complex sediment transport processes with differing sediment transport regimes, supplies, and controls that are representative of low-relief glaciated regions around the World (Ellison et al., 2016; Vaughan, Belmont, et al., 2017).

2 DATA AND METHODS

Supervised ML models were developed with physically collected sediment samples, streamflow, and geospatial datasets (Figure 2) to predict SSC in milligrams per litre (mg/L) and BL discharge in tons per day (tons/day; Edwards & Glysson, 1999). Two new streamflow features were calculated to better account for hysteresis (Section 2.1.1). Approximately 400 features (Lund & Groten, 2022) were tested for correlation before the final selected features were included in the models, as described in Section 2.2.1. Model development is described in Section 2.3. Analysis of results are described in Section 2.4 and includes testing of streamflow control models (Section 2.4.1), interpretation of SHAP plots (Section 2.4.2), and comparison of cumulative daily loads to in-situ surrogate datasets (Section 2.4.3).
2.1 Sediment transport and streamflow data

Sediment data were obtained from the USGS National Water Information System (NWIS; U.S. Geological Survey, 2021b). EWI and EDI SSC samples were primarily used along with 85 grab samples that were collected during low flows (less than 2 ft per second) when water velocities were too slow for isokinetic samplers and EWI and EDI methods (Edwards & Glysson, 1999). Two samples greater than 6400 mg/L were omitted from model development to reduce bias because they were collected at sites impacted by major floods and were deemed as extreme sediment transport events not representative of the entire dataset (Lund & Groten, 2022). Fourteen SSC samples were removed from the dataset due to having concentrations of greater than 80% sand, likely due to field crews inadvertently sampling the streambed. A total of 1382 SSC samples from 56 sites and 638 bedload samples (collected with SEWI methods [Edwards & Glysson, 1999]) from 43 sites were included in the final dataset (Lund & Groten, 2022).

Streamflow data were obtained from NWIS (U.S. Geological Survey, 2021b) and the Minnesota Department of Natural Resources/Minnesota Pollution Control Agency Cooperative Streamgauging (Minnesota Department of Natural Resources, 2021). Recurrence intervals (1.5 and 2-year) were obtained from USGS StreamStats (U.S. Geological Survey, 2021a). When instantaneous streamflow (15-min increment) was available, it was matched with SSC and BL data. Daily mean streamflow was used when instantaneous streamflow was unavailable (102 SSC samples and 21 BL samples). Instantaneous and daily streamflow values were normalized by the 2-year RI at each site which created a dimensionless streamflow dataset.

2.1.1 Development of streamflow slope

The dimensionless streamflow dataset was used to calculate two slope features that provide a magnitude and direction of the rate of change of streamflow (Figure 3). These slope features facilitate ML models in understanding changing streamflow conditions around the time of sample collection and possibly better account for sediment hysteresis. The first slope feature was calculated from dimensionless streamflow during sample collection and dimensionless streamflow 24-h after (hereafter will be referred as ‘streamflow slope [24-h before]’). The second slope feature was calculated from dimensionless streamflow during sample collection and dimensionless streamflow 24-h after (hereafter will be referred to as ‘streamflow
Table 1

| Category         | Data type | Definition               |
|------------------|-----------|--------------------------|
| Watershed        | Static    | Full drainage area upstream of the site |
| Catchment        | Static    | Local sub-drainage area closest to the site |
| Near channel     | Static    | Riparian zone – 100 m buffer along the mainstem NHDPlusV2 flowline of the upstream watershed |
| Channel specific | Static    | Describes site specific and channel features |
| Streamflow       | Variable  | Describes developed dimensionless streamflow features |
| Sample specific  | Variable  | Describes sample specific features |

Abbreviation: NHD, National Hydrography Dataset; V2, version two.

*Definitions, descriptions, and sources of all features used are available in a U.S. Geological Survey data release (Lund & Groten, 2022).

2.2 Geospatial data

Geospatial features used in model development were sourced from USGS StreamStats basin characteristics (U.S. Geological Survey, 2021a), the U.S. Environmental Protection Agency’s (USEPA) StreamCat dataset (U.S. Environmental Protection Agency, 2021b; Hill et al., 2016), and the U.S. Department of Energy’s U.S. Stream Classification System dataset (McManamay & Derolph, 2019; Oak Ridge National Laboratory, 2021). The Stream Classification System feature river size class was modified by combining Small Creek sites with Large Creek sites to form a new class called Creek (C). Five physiographic sediment regions were developed by comparing sediment transport rates in the USEPA level IV ecoregions (U.S. Environmental Protection Agency, 2021a) and merging ecoregions into the five sediment regions (Figure 1). Geospatial features were organized using the USGS/USEPA’s National Hydrography Dataset Plus Version 2 framework (McKay et al., 2012; Moore & Dewald, 2016). Approximately 400 static geospatial watershed, catchment, near channel, and channel features (Table 1) were included in the initial dataset (Lund & Groten, 2022).

2.2.1 Removing geospatial features from the dataset

While ML models are robust to datasets including many input features (Breiman, 2001; Fox et al., 2017), it can be difficult to assess feature importance if there are multiple highly correlated features (Molnar, 2019). To reduce correlation in our dataset and improve model interpretation, two Pearson correlation matrixes were built. The first correlation matrix was sorted manually to remove geospatial features that described similar characteristics (Lund & Groten, 2022). The second correlation matrix (Kuhn, 2008, 2021; R Core Team, 2021) was used to search a second correlation matrix and remove features with an absolute-value pair-wise correlation greater than 0.50 (Lund & Groten, 2022). The remaining features were used in model development as described in Section 2.3.

2.3 Model development

XGBoost ML models were trained and tested in R using the XGBoost package (Chen et al., 2021; R Core Team, 2021). Data were split into training (80% of the data) and testing (20% of the data) datasets using a stratified random split to capture equal proportions of samples per site in each dataset. The training data were used in a grid search to build over 2000 5-fold cross validation (CV) XGBoost models to determine the best set of tuning parameters (Lund & Groten, 2022). The final evaluation error for a CV XGBoost model is the average of the five folds, so all the training data were used to validate performance. The CV XGBoost models were compiled by organizing their evaluation errors from lowest to highest (Lund & Groten, 2022).

The top 10 sets of tuning parameters from the CV XGBoost model grid search were used to build 10 XGBoost models to find the best overall set of tuning parameters for the final model (Lund & Groten, 2022). The learning rate was reduced from 0.10 to 0.01 to slow down the model and force it to be more conservative and help prevent overfitting (Chen & Guestrin, 2016). A watchlist metric from the XGBoost package was used to calculate the error statistic for the testing dataset on the trained model after every tree addition. Early stopping rounds were set to 10 to stop the addition of trees and abort the model training once the evaluation error stopped optimizing. The model with the best evaluation error from testing data was selected, and its optimal number of trees were used for the final model (Lund & Groten, 2022).

Machine learning regression models can have biased results as they typically overpredict on the low end and underpredict on the high end (Belitz & Stackelberg, 2021). Three goodness-of-fit (GOF) statistics were used as evaluation errors during the CV grid search process to test if one method could train less biased and more accurate models compared to the other. The first GOF statistic used was root mean squared error (RMSE):

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{x}_i)^2}
\]

where, RMSE is the root mean squared error, \(x_i\) are the measured values, \(\hat{x}_i\) are the predicted values, and \(n\) is the number of observations used. The second GOF statistic used was the Nash-Sutcliffe efficiency (NSE):
\[
\text{NSE} = 1 - \frac{\sum_{i=1}^{n} (x_i - x'_i)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2},
\]

where, NSE is the Nash-Sutcliffe efficiency value, \( n \) is the number of observations used, \( x_i \) is the measured value for observation \( i \) (SSC in mg/L or BL in tons/day), \( x'_i \) is the predicted value for observation \( i \) (SSC in mg/L or BL in tons/day), and \( \bar{x} \) is the mean of the measured values (SSC in mg/L or BL in tons/day). The third GOF statistic used was bias corrected correlation coefficient (\( bR^2 \)):

\[
bR^2 = \begin{cases} 
|b|R^2, & \text{if } b \leq 1 \\
R^2, & \text{if } b > 1 
\end{cases}
\]

where, \( b \) is the slope of the regression line between predicted and measured values, and \( R^2 \) is the coefficient of determination.

2.4 | Analysis of results

2.4.1 | Comparison of streamflow control and final models

In order to test the accuracy gained with the development of streamflow features (dimensionless streamflow and streamflow slopes), ‘control models’ were developed for SSC and BL. In each control model, dimensionless streamflow was replaced with the corresponding streamflow value, and streamflow slopes were replaced with categorical features to describe the rising limb (1) or falling limb (0) on the streamflow hydrograph. The control and final models’ GOF statistics were compared.

2.4.2 | Shapley additive explanation (SHAP) plots

SHAP summary and dependence plots were used to interpret trends in the final SSC and BL models’ top 10 important features. The relationship between feature values and their corresponding SHAP values can show what the models learned from the training datasets and how they made their predictions (Molnar, 2019). All SHAP plots were made with the SHAPforxgboost R package (Liu & Allan, 2021; version 0.1.1).

SHAP summary plots were created by ordering the top 10 features’ absolute mean SHAP value. A higher absolute mean of the SHAP value indicated greater contribution to the prediction and importance. SHAP summary plots show the most important feature on the top left of the y-axis, and the nine other features are listed below in descending order. Adjacent to each feature is a numeric value which is the absolute mean SHAP value. SHAP values are coloured by the features’ numerical values from low to high.

SHAP dependence plots were created for each of the top 10 features in both SSC and BL ML models to see the effect each feature has on the model predictions. A positive SHAP value indicates the feature observation had a higher impact on predicting a target value greater than the mean of the observed values. A negative SHAP value indicates the feature observation impacted a prediction that was lower than the mean of observed values (Lundberg et al., 2018). SHAP dependence plots were constructed with SHAP values on the \( y \)-axis and feature values on the \( x \)-axis. A locally estimated scatterplot smoothing (LOESS) is presented as a red line on the SHAP dependence plots to indicate the global trend (Liu & Allan, 2021; version 0.1.1).

2.4.3 | Comparison of cumulative daily suspended-sediment loads

Cumulative daily mean suspended-sediment loads (SSLs) were calculated at four sites from SSC ML model outputs and compared to cumulative daily mean SSLs and 90% prediction intervals estimated from linear regression model outputs that used in-situ sediment surrogates (Groten, 2017a, 2017b, 2017c; Groten et al., 2019). Site-specific ML models were trained with the same methods used to train the final SSC ML model, but the site-specific ML models did not include the SSC data collected at each surrogate site. This comparison was used to test the ML model’s ability to predict SSLs at a site not included in training or testing datasets.

Three sites used optical turbidity as a surrogate for SSC (Rasmussen et al., 2009): the Knife River near Two Harbors, Minnesota (USGS station number 04015330), Blue Earth River in Mankato, Minnesota (USGS station number 05321995), and Zumbro River at Kellogg, Minnesota (USGS station number 05374900). The Minnesota River at Fort Snelling, Minnesota site (USGS station number 05330920) used mean sediment corrected backscatter from an acoustic Doppler velocity meter as a surrogate for SSC (Landers et al., 2016). The sediment surrogates were used to compare against the ML models because of the strong relations between the surrogates and SSCs. The sites that used turbidity as a surrogate for SSC had \( R^2 \) greater than 0.95 (Groten, 2017a, 2017b, 2017c). The site that used sediment corrected backscatter as a surrogate for SSC had a \( R^2 \) greater than 0.8 (Groten et al., 2019). The cumulative daily SSLs estimated from the site-specific ML models were considered to be relatively validated if they were between the corresponding upper- and lower-90% prediction interval from each surrogate site.

3 | RESULTS

3.1 | Streamflow normalization

Bankfull streamflow values were available for 30 of the 56 sites, so these bankfull streamflows were compared to 1.5 and 2-year RIs to select the RI most representative of bankfull streamflow. The 2-year
RI was selected to normalize streamflow because it had a stronger relation (higher $R^2$) with bankfull streamflow values (Figure 4). Normalizing streamflow gives a systematic way to compare the magnitude of streamflow across sediment regions and river size class as a value of 1, now represents the 2-year RI for every site (Figure 5). Normalizing streamflow allowed for larger datasets to train and test the SSC and BL ML models rather than having to subgroup datasets into separate models (i.e., sediment regions or river size class).

**3.2 Summary statistics**

The sediment transport dataset represented different river sizes in Minnesota’s five sediment regions. Summary statistics demonstrate the range of sediment transport across the state (Table 2). The SE and then the SW regions had the highest mean SSC followed by the NW, NE, and MID regions, respectively. Mean BL was highest in the SW and then the SE followed by NE, NW, and MID regions, respectively. Overall, southern Minnesota (SW and SE) had the highest mean values of SSC and BL.
### TABLE 2  Summary statistics by region and river size class adapted from McManamay & Derolph, 2019

| Region          | NW | NE  | MID | SE  | SW  |
|-----------------|----|-----|-----|-----|-----|
| Number of sites | 7  | 10  | 4   | 12  | 23  |
| Number of samples | 143 | 205 | 79  | 332 | 623 |
| Minimum         | 2  | 2   | 2   | 2   | 5   |
| Maximum         | 944 | 6400 | 107 | 4820 | 5830 |
| Median          | 63 | 35  | 14  | 90  | 195 |
| Mean ± standard deviation | 129 ± 159 | 142 ± 555 | 20 ± 20 | 335 ± 662 | 308 ± 442 |

| Bedload, in metric tons per day |
|----------------------------------|
| Number of sites                  | 4  | 10  | 4   |
| Number of samples                | 62 | 85  |
| Minimum                          | 1.1 | 0.67 |
| Maximum                          | 347.4 | 1011.2 |
| Median                           | 17.56 | 9.64 |
| Mean ± standard deviation        | 54 ± 83 | 66 ± 177 |

| River size class                  |
|-----------------------------------|
| C                                 |
| SR                                |
| MR                                |
| MS                                |
| LR                                |
| GR                                |
| Number of sites                  | 11 | 20  | 11  | 5   | 5   | 4   |
| Number of samples                | 162 | 490 | 234 | 153 | 243 | 100 |
| Minimum                          | 2  | 2   | 2   | 21  | 2   |
| Maximum                          | 6400 | 5830 | 1840 | 3880 | 1620 |
| Median                           | 94 | 66.5 | 76  | 264 | 222 |
| Mean ± standard deviation        | 283 ± 720 | 230 ± 581 | 202 ± 310 | 477 ± 572 | 275 ± 232 |

| Bedload, in metric tons per day |
|----------------------------------|
| Number of sites                  | 7  | 19  | 8   | 3   | 3   |
| Number of samples                | 70 | 274 |
| Minimum                          | 0.18 | 0.63 |
| Maximum                          | 841.9 | 1011.2 |
| Median                           | 3.67 | 21.75 |
| Mean ± standard deviation        | 37 ± 139 | 52 ± 93 |

| Streamflow, in cubic meters per second |
|----------------------------------------|
| Minimum                                | 0.09 | 0.06 | 0.76 | 4.45 | 10.25 | 110.2 |
| Maximum                               | 104.8 | 154.9 | 360.1 | 755.7 | 2183.7 | 2941.1 |
| Median                                | 2.35 | 9.71 | 32.15 | 82.4 | 370.3 | 934.9 |
| Mean ± standard deviation             | 6.3 ± 13.1 | 14.8 ± 16.8 | 55.7 ± 63.6 | 116.9 ± 109.4 | 483.0 ± 424.4 | 1037.1 ± 535.5 |

| Dimensionless streamflow             |
|--------------------------------------|
| Minimum                              | 0.003 | 0.002 | 0.008 | 0.03 | 0.024 | 0.068 |
| Maximum                              | 2.19 | 3.58 | 3.14 | 3.94 | 5.44 | 1.81 |
| Median                               | 0.14 | 0.28 | 0.45 | 0.33 | 0.9 | 0.53 |
| Mean ± standard deviation            | 0.25 ± 0.35 | 0.41 ± 0.48 | 0.66 ± 0.63 | 0.48 ± 0.51 | 1.19 ± 1.05 | 0.55 ± 0.30 |

| Dimensionless streamflow slope (24 h before) |
|----------------------------------------------|
| Minimum                                      | −1.04 | −0.91 | −0.49 | −0.48 | −0.54 | −0.12 |
| Maximum                                      | 1.95 | 1.44 | 0.92 | 0.99 | 0.87 | 0.16 |
| Median                                       | 0.01 | 0 | 0.01 | 0 | 0 | 0.01 |
| Mean ± standard deviation                    | 0.07 ± 0.29 | 0.01 ± 0.21 | 0.04 ± 0.17 | 0.06 ± 0.19 | 0.03 ± 0.16 | 0.01 ± 0.05 |

Abbreviations: ±, plus or minus; C, creek; GR, great river; LR, larger river; MID, middle; MR, medium river; MS, mainstem; NE, northeast; NW, northwest; SE, southeast; SR, small river; SW, southwest definition.
Dimensionless streamflow values ranged from 0.1 to 8.63 indicating samples were collected in varying streamflow conditions. For river size class (McManamay & Derolph, 2019; Oak Ridge National Laboratory, 2021) mean SSC was highest for mainstem (MS) followed by large river (LR), C, medium river (MR), small river (SR), and great river (GR), respectively. Mean BL was highest in MS followed by MR, LR, SR, and C, respectively (Table 2). Streamflow at smaller sites (C and SR) had a greater rate of change with greater minimum and maximum streamflow slopes.

### 3.3 Extreme gradient boosting models

Two XGBoost models were developed to predict SSC and BL from geospatial and streamflow features. The final SSC XGBoost model included 1106 samples for training, 276 samples for testing, and 32 selected features (Table 3). The BL XGBoost model included 511 samples for training, 127 samples for testing, and 25 selected features (Table 3). Final SSC and BL ML models had 21 selected features in common (Table 3).
### Table 4: Goodness-of-fit results from extreme gradient boosting models

| Model | RMSE | NSE  | bR²  |
|-------|------|------|------|
| SSC   | 329.4 | 0.69 | 0.45 |
| Bedload | 149.4 | 0.78 | 0.63 |

Abbreviations: bR², bias correlation coefficient; NSE, Nash-Sutcliffe efficiency; RMSE, root mean squared error; SSC, suspended-sediment concentration.

#### 3.3.1 Model goodness-of-fit

Three different evaluation error methods were tested when training the models, and the final models were selected for having the best combination of GOF statistics (RMSE, NSE, bR²). The final SSC XGBoost model was trained for RMSE because it was more accurate (lower RMSE, higher NSE) and less biased (higher bR²) (Table 4) than using the other GOF statistics. The final BL XGBoost model was trained for bR² because it was more accurate (higher NSE), RMSEs were similar, and less biased due to having a higher bR² (Table 4).

#### 3.3.2 Comparison of streamflow control and final models

Results comparing the SSC-streamflow control ML model to the final SSC ML model showed that normalizing streamflow by the 2-year RI and calculating the streamflow slopes improved RMSE values from 377.6 to 329, NSE values from 0.59 to 0.69, and bR² values from 0.35 to 0.45 (Table 5). Results comparing the BL-streamflow control ML model to the final BL ML model showed that normalizing streamflow by the 2-year RI and calculating the streamflow slopes improved RMSE values from 178.9 to 150.3 and NSE values from 0.68 to 0.78 while bR² values slightly decreased from 0.69 to 0.67 (Table 5).

#### 3.3.3 Extreme gradient boosting model results

The final SSC and BL XGBoost ML models were able to predict sediment transport at a variety of river sizes across the state of Minnesota. The SSC and BL ML models overpredicted on the low end (approximately less than 10 for SSC and BL) and underpredicted on the high end (approximately greater than 1000 for SSC and BL; Figure 6). The results had more variability on the low end, as shown by the greater range in scatter between predicted and observed SSCs (Figure 6a) and BL (Figure 6b) at the lower end of respective data ranges. A closer look at important ML model features and how they affect sediment transport prediction is facilitated by SHAP plots (Sections 3.4.1 and 3.4.2).

### Table 5: Goodness-of-fit results from comparison of machine learning streamflow control models to final models

| Model                          | RMSE  | NSE  | bR²  |
|--------------------------------|-------|------|------|
| SSC-streamflow control         | 377.6 | 0.59 | 0.35 |
| SSC-final                      | 329.4 | 0.69 | 0.45 |
| BL-streamflow control          | 178.9 | 0.68 | 0.69 |
| BL-final                       | 150.3 | 0.78 | 0.67 |

Abbreviations: BL, bedload; bR², bias correlation coefficient; NSE, Nash-Sutcliffe efficiency; RMSE, root mean squared error; SSC, suspended-sediment concentration.

#### 3.4 Shapley additive explanation (SHAP) plots

##### 3.4.1 Suspended sediment Shapley additive explanation (SHAP) plots

The 10 most important features in the SSC ML model included streamflow, watershed, catchment, sample specific, and channel features (Figure 7). Three streamflow features ranked 1, 2, and 9 (streamflow slope [24-h before], dimensionless streamflow, and streamflow slope [24-h after], respectively). Three watershed features ranked 3, 5, and 8 (percent grassland/herbaceous land cover, percent clay soils, and mean elevation, respectively). Two catchment features ranked 4 and 6 (percent deciduous forest land cover and percent grassland/herbaceous land cover, respectively). One sample specific feature ranked 7 (month). One channel feature ranked 10 (river width area).

The two most important features were streamflow slope (24-h before) and dimensionless streamflow. When streamflow slope was near zero (streamflow is stable) or negative (streamflow is falling) the feature had a negative SHAP value, indicating lower SSC transport (Figure 8a). Conversely, positive streamflow slopes corresponded with an increase in SHAP values, indicating that a quickly rising streamflow had a positive impact on higher SSC transport. When dimensionless streamflow was low (near zero), the feature generally had a negative SHAP value (Figure 8b). As dimensionless streamflow increased, the SHAP value increased until dimensionless streamflow was around one and a half. Observations with dimensionless streamflow near one had the highest SHAP values. Once dimensionless streamflow was greater than one and a half times the 2-year RI the SHAP values are scattered around zero.

The streamflow slope (24-h after) results were different than the streamflow slope ([24-h before] Figure 8i,a). Streamflow slopes near zero generally had negative SHAP values, but as streamflow slopes...
(24-h after) increased SHAP values were both positive and negative. SHAP values were higher when the streamflow slopes (24-h after) were higher.

Sites with less than 1% of their watershed in agricultural land on slopes ≥ 10% generally resulted in negative SHAP values; whereas sites with greater than 1% of their watershed covered by agricultural land on slopes ≥ 10% generally had positive SHAP values (Figure 8c). Sites with greater than 22% of their watershed in clay soils had positive SHAP values while sites with less clay soils had negative SHAP values (Figure 8e). Sites with watershed mean elevation between 300 and 360 m generally had positive SHAP values while greater mean watershed elevations resulted in negative SHAP values (Figure 8h).

3.4.2 | Bedload Shapley additive explanation (SHAP) plots

The 10 most important features in the BL ML model included streamflow, watershed, catchment, sample specific, and near channel features (Figure 9). Streamflow features ranked 2, 5, and 7 (dimensionless streamflow, streamflow slope [24-h before], and streamflow slope [24-h after]). Two watershed features ranked 1 and 8 (drainage area and density of dams, respectively). Three catchment features ranked 4, 6, and
percent glacial outwash and glacial lake sediment [coarse textured], percent evergreen forest land cover, and mean soil erodibility on agricultural land, respectively. One sample specific feature ranked 10 (month). One near channel feature ranked 3 (river size class).

Dimensionless streamflow BL SHAP dependence plot showed a similar shape as the SSC plot but showed different values on the x-axis as breakpoints (Figures 10b and 8b). When dimensionless streamflow was near zero, the feature generally had a negative SHAP value similar to the SSC model. When dimensionless streamflow features were above 0.75, SHAP values were positive and increased until about two and a half times the 2-year RI indicating higher BL transport (Figure 10b). Like the SSC dependence plot, the streamflow slope (24-h before) had several high SHAP values that correspond to positive streamflow slopes (Figure 10e), indicating that rising streamflow was...
important in the prediction of higher BL transport. A negative streamflow slope (24-h before) generally received SHAP values near zero. Positive streamflow slope (24-h after) had negative SHAP values. As streamflow slope (24-h after) became more negative, the associated SHAP value increased (Figure 10g), indicating negative streamflow slope (24-h after) had an impact on greater BL transport, and positive streamflow slope (24-h after) had an impact on less BL transport.

Samples collected from sites with a watershed drainage area greater than 2860 km² had positive SHAP values, while sites with smaller watersheds had negative SHAP values (Figure 10a). Sites with greater than 0.005 density of dams had positive SHAP values, while below this dam density the majority of the SHAP values were negative (Figure 10h).

Sites with more than 1% glacial outwash and glacial lake sediment (coarse textured) in their catchment, had mostly negative SHAP values (Figure 10d). Percent evergreen forest in the catchment was mostly associated with negative SHAP values (Figure 10f). Most of the samples had positive SHAP values when mean soil erodibility on agricultural land was between 0.05 and 0.1 (Figure 10i). Similar to the SSC model, June had the highest positive SHAP values indicating higher BL transport (Figure 10j). The one channel feature, river size class, had positive SHAP values for river size classes MS and LR (Figure 10c).

3.5 Comparison of cumulative daily suspended-sediment loads

Comparison of site-specific ML model output to in-situ sediment surrogate model outputs (Table 6) provided the opportunity to validate this ML modelling approach. The site-specific ML cumulative daily SSLs were within the sediment surrogate 90% prediction intervals at all four sites (Table 6). On shorter time intervals (e.g., one to multiple months) the site-specific ML model predicted higher SSLs than the surrogate’s upper 90% prediction interval at the Knife River in 2017 and 2018, the Zumbro River in 2017, and the Minnesota River in 2018 (Figure 11). However, the site-specific ML model estimates were within the surrogate’s 90% prediction intervals after the shorter time intervals ended (greater than one to multiple months) for the previously mentioned sites and years.

4 DISCUSSION

Advancements in data science and ML allowed for enhanced data driven sediment transport modelling, prediction accuracy, and interpretation techniques. Normalizing streamflow around the 2-year RI reduced variability and allowed for larger datasets to train and test the SSC and BL ML models. Calculating two new streamflow slope features helped to better account for hysteresis and improved the accuracy of the models. Geospatial datasets that account for local, near-channel, and watershed features helped improve predictions by allowing the model to learn more complex processes related to sediment transport. Comparing ML model SSLs to in-situ surrogate model SSLs highlighted the utility of ML model’s ability to learn and apply complex relations when making predictions at sites without physically collected samples.
Utility of a machine learning approach

ML's ability to use large datasets and learn complex relationships has helped the SSC and BL ML models make more accurate predictions without needing to define site-specific streamflow-sediment transport relationships. The ML models were able to learn sediment-transport processes while traditional methods such as SRCs and DSRCs do not have that ability (Cisty et al., 2021; Francke et al., 2008; Zounemat-Kermani et al., 2020). XGBoost's ability to add custom evaluation metrics and additional parameter tuning helped reduce variance and bias (Table 4) (Lund & Groten, 2022). SHAP values provided a quantitative way to interpret, present, and support the model by displaying the relation and interaction of feature values and prediction output. SHAP value interpretation, facilitated by SHAP summary and dependence

4.1 | Utility of a machine learning approach

FIGURE 10 Shapley additive explanation (SHAP) dependence plot showing the top 10 important features from the bedload machine learning model. SHAP values on the y-axis and features observed values on the x-axis, each subplot has different scales. (a) Drainage area (watershed), in square kilometres, (b) dimensionless streamflow, (c) river size class (channel specific), (d) percent glacial outwash and glacial lake sediment, coarse-textured (catchment), (e) streamflow slope (24-h before), (f) percent evergreen forest land cover (catchment), (g) streamflow slope (24-h after), (h) density of dams (watershed), dams per square kilometre, (i) mean soil erodibility on agricultural land (catchment), and (j) month
plots, provided insight into how ML models made predictions and the processes controlling sediment transport. Comparing the ML models’ outputs with in-situ sediment surrogate estimates showed that the ML models are comparable to proven SSC estimation techniques and can be used to predict SSC at sites without physically collected samples.

### 4.2 Normalizing streamflow

Normalizing streamflow reduced variability in the dataset and allowed for the development of one model for each constituent (SSC and BL) rather than developing multiple models (e.g., developing models for each sediment region or river size class). Developing one model for each constituent allowed for larger datasets to be used for training and testing. The dimensionless streamflow SHAP dependence plots showed the highest SHAP values were near the 2-year RI, which indicates higher sediment transport near bankfull streamflows (Figures 8b and 10b). These results are consistent with bankfull streamflows being the most geomorphically active streamflows in the channel before streamflow spills over its banks and loses energy to the floodplain (Biedenharn et al., 2008; Lane, 1955). Additionally, the observed decrease in SHAP values at very high streamflows could be connected to sediment deposition in the channel and floodplain caused by the river overbanking and subsequent loss of energy to transport sediment (Wilcock et al., 2009). The observed decrease in SHAP values might also be explained by the depletion of upstream sediment sources after long flood durations (Gellis, 2013; Smith & Dragovich, 2009).

Ultimately, normalizing streamflow by the 2-year RIs had multiple benefits. First, normalizing streamflow by the 2-year RI reduced the variability in the dataset near bankfull streamflow levels which is known to be an important sediment transport index. Second, the model used dimensionless streamflow to learn complex relations across the dataset of varying river sizes by systematically quantifying the degree that 2-year RIs were less than, matched, or exceeded. Third, the model learned from the relation between sediment transport, streamflow, and the other features used in the models. Fourth, prediction accuracies increased considerably when streamflow was normalized by the 2-year RI when comparing the control and final models GOF statistics (Table 5).

#### 4.3 Accounting for hysteresis with streamflow slope

Streamflow slopes (24-h before and after) were features developed from the dimensionless streamflow dataset to provide the model with more insight into the complex relation between streamflow and sediment transport. The streamflow slope provided a rate and direction (rise is a positive value and fall is a negative value) of streamflow in the channel rather than just classifying streamflow's position on the hydrograph with simple categorical features. The SSC and BL ML models learned from the streamflow slope by differentiating the complex rate of change in streamflow events. A streamflow event that is quickly or slowly changing could be related to storm intensity, measured as peakflow divided by total runoff, which is positively correlated with sediment transport (Guy, 1964). As seen from the control model tests (Table 5), the final models’ GOF statistics improved considerably compared to the control models.

The results from the ML models suggest that the streamflow slope features helped to reduce uncertainty between streamflow and sediment transport across varying river sizes and sediment regions by supplying the ML models with more information to understand complex relations between sediment source and transport dynamics. This finding directly supports objective number two and connects with previous work that shows ML can make more accurate predictions than SRGs (Cisty et al., 2021; Francke et al., 2008; Zounemat-Kermani et al., 2020).

#### 4.4 Geospatial datasets

Geospatial features in the SSC and BL ML models can be used to make inferences about sediment sources and sediment transport processes. The modelling framework facilitated the use of geospatial features without fully relying on them to predict sediment transport, and highly correlated features were removed to aid in the interpretation...
of important features and to prevent the SSC and BL ML models from overfitting (Molnar, 2019).

There were more watershed and catchment features in the SSC and BL ML models than near channel and channel features (Table 3). The BL ML model had more geospatial features related to the streamflow regime (drainage area, river size class, and density of dams), which relates to river power and the ability to transport available bed sediment (Gray & Simões, 2008), in the top 10 important features than

FIGURE 11 Cumulative suspended-sediment loads from four site-specific machine learning models and in-situ sediment surrogates at the Knife River near Two Harbors, Minnesota, in 2017 (a) and 2018 (b), Blue Earth River in Mankato, Minnesota, in 2017 (c) and 2018 (d), Zumbro River at Kellogg, Minnesota, in 2017 (e) and 2018 (f), and Minnesota River at Fort Snelling, Minnesota, in 2017 (g) and 2018 (h)
the SSC ML model which had more watershed and catchment features related to land use and land cover (percent agricultural land on slopes ≥ 10%, percent deciduous forest land cover, and percent grassland/herbaceous land cover). The geospatial features percent of agricultural land on ≥ 10% slopes (SSC ML model) and mean soil erodibility on agricultural land (BL ML model) were expected because other studies have shown the connection between agricultural land use and excess sediment transport (Foufoula-Georgiou et al., 2015; Schottler et al., 2014). More specifically, the SSC ML SHAP values indicate a higher percent of agricultural land on ≥ 10% slopes in the watershed area impact higher SSC predictions (Figure 8c), which could be connected to other landscape and land-use features such as increased stormflow from tile drainage and erosion (Belmont et al., 2011; Schottler et al., 2014). An unexpected catchment feature in the SSC ML model was percent deciduous forest land cover because the SSC SHAP dependence plot showed an increase in SSC transport with an increase in percent deciduous forest land cover (Figure 8d). A possible explanation is that higher percentage of deciduous forest land cover only represents the local catchment area around the site, so an increase in SSC transport could be due to other surrounding land use features represented by the upstream watershed.

Looking more closely at potential sediment sources and grain sizes, BL SHAP values showed that the lower percent of coarse textured glacial outwash and glacial lake sediments (Figure 10d) in the catchment impacts higher BL transport while the SSC SHAP values showed that high percent clay in the watershed impacts higher SSC transport (Figure 8e). The percent of coarse textured glacial outwash and glacial lake sediments in the catchment could be connected to potential sources of BL in the channel; however, a possible explanation is difficult to determine without having better geospatial datasets representing bed material type in the channel. The percent clay in the watershed could be teaching the SSC model about a potential suspended sediment source since fine particles can make up a considerable amount of SSCs. Altogether, these results show that geospatial predictors are helping the ML models account for complex sediment source and transport processes at various scales which are difficult to account for with SRCs and DSRCs (Atieh et al., 2015; Ellison et al., 2016; Francke et al., 2008; Vaughan, Belmont, et al., 2017).

4.5 Possible model improvements

Further development of features and additional sites and samples could improve ML models. Improvements to ML models could include calculation of higher resolution in-channel or near-channel features that are known to be sediment transport controls. The current models were developed from publicly available datasets available for the entire state, and the resolution could be too coarse. A more efficient method of locating important features from the vast amount of available geospatial data while accounting for correlation to other features could help model interpretation and prediction accuracy. Additional continuous time-series datasets could be used to add more features like gridded rainfall patterns, precipitation intensities, and antecedent soil moisture calculated for the upstream catchment and watershed (Essou et al., 2016). Alternative methods to calculate dimensionless streamflow could be explored. Using different streamflow slope time intervals could increase prediction accuracy since varying river sizes respond differently (streamflow rising and falling at different time intervals) to snowmelt and storm events. Additional analyses of streamflow data could include a time-since-last-event feature that could teach the model sediment source and storage controls to help better account for hysteresis (Gellis, 2013; Smith & Dragovich, 2009). ML is a complex and ever-changing field of study; additional work could be done exploring other methods including artificial neural networking (ANN) and, more specifically, hybrid wavelet and neural networking (WANN), which produced accurate results in sediment transport prediction studies (Afan et al., 2016; Khan et al., 2021). Lastly, the models could be improved with additional sites and additional samples to better represent sediment transport.

4.6 Comparison of streamflow feature accuracy and loads

In-situ sediment surrogates are a proven and accurate method to estimate SSC and were used to relatively validate ML model outputs in this study. This validation showed that this ML approach can be used at sites that did not have physically collected samples available to train the ML models. The results validated the use of the ML models to estimate cumulative daily SSLs. The ML models predicted higher SSLs during shorter time intervals at some sites. Because these time intervals generally had lower streamflow, ML models tend to overpredict at lower streamflow, as described in Section 3.3.3 (Figure 6), and the surrogate models tended to underpredict at lower streamflow. Future work could include using these sediment surrogate model outputs to test if the ML model is better at accounting for the processes controlling sediment transport as more accurate and representative features are calculated. The sediment surrogate model outputs could also be used to calculate a hysteresis-index to quantify rising and falling limb hysteresis trends (Liu et al., 2021; Lloyd et al., 2016; Vaughan, Bowden, et al., 2017) to improve the ML models.

5 CONCLUSIONS

This research was possible because local, state, and federal natural resource managers realized the importance in collecting physical sediment samples across the state of Minnesota. This study elucidates the potential of supervised ML models paired with geospatial datasets and more accurate streamflow features to increase prediction accuracy and provide a better understanding of the relative roles that landscape, near-channel, and in-stream conditions play in sediment transport and was achieved through three objectives.

The first objective was achieved by comparing SSLs from trained ML models to SSL datasets from in-situ sediment surrogates at four sites across Minnesota that were not included in model training.
Results from ML models were mostly within the 90% prediction intervals of the surrogates at all four sites, supporting the idea that ML can learn from complex relations and apply those relations to sites with few to no physical samples. The second objective was achieved by the normalization of streamflow by the 2-year RI, which trained the models to learn where the samples were collected in relation to a geomorphically active stream system, and the streamflow slopes allowed the ML models to learn from changing streamflow conditions before and after sample collection. These developed streamflow features helped account for hysteresis and improved the prediction accuracy of the ML models. The third objective was achieved by using SHAP values to interpret how the ML models were making predictions and learning from the complex relations between sediment transport, streamflow, watershed, catchment, and near-channel features. Finally, these findings are useful for natural resource managers, stream practitioners, and anyone interested in fluvial sediment transport because they can help improve sediment load estimations, enhance restoration design and priority, identify streams that depart from reference conditions, help improve sediment load estimations, enhance restoration design and priority, identify streams that depart from reference conditions, and any one interested in fluvial sediment transport because they can help improve sediment load estimations, enhance restoration design and priority, identify streams that depart from reference conditions, and help evaluate effectiveness of sediment reduction strategies.

ACKNOWLEDGEMENTS
The authors would like to thank the Minnesota Department of Natural Resources, Minnesota Pollution Control Agency, Minnesota’s Clean Water Fund, and U.S. Geological Survey Cooperative Matching Funds for their financial assistance with this study. J. William Lund was partially supported by the Peter F. Frolliott Fellowship in the Department of Forest Resources, University of Minnesota and USGS Cooperative Agreement G20AC00427. Diana L. Karwan acknowledges support from the Minnesota Agricultural Research Station (Project MIN-42-080), Christopher Ellison, Erin Coenen, and Gerald Storey of the U.S. Geological Survey are acknowledged for assistance with project planning, data collection, and database management. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

DATA AVAILABILITY STATEMENT
Data used to train and test the machine learning models, R scripts, model outputs, and computed loads are available in a U.S. Geological Survey data release (Lund & Groten, 2022).

ORCID
J. William Lund https://orcid.org/0000-0002-8830-4468
Joel T. Groten https://orcid.org/0000-0002-0441-8442
Diana L. Karwan https://orcid.org/0000-0002-4529-0369

REFERENCES
Afan, H. A., El-shafie, A., Mohtar, W. H. M. W., & YaseenPast, Z. M. (2016). Present and prospect of an artificial intelligence (AI) based model for sediment transport prediction. Journal of Hydrology, 541B, 902–913. https://doi.org/10.1016/j.jhydrol.2016.07.048
Alexander, J. S., Wilson, R. C., & Green, W. R. (2012). A brief history and summary of the effects of river engineering and dams on the Mississippi River system and delta. U.S. Geological Survey Circular 1375 (43 p.). https://pubs.usgs.gov/circ/1375/

American Society for Testing and Materials. (2000). Standard test methods for determining sediment concentration in water samples (pp. 395–400). American Society for Testing and Materials International. D3977-97, v. 11.02, Water (II).

Asselman, N. E. M. (2000). Fitting and interpretation of sediment rating curves. Journal of Hydrology, 234(3–4), 228–248. https://doi.org/10.1016/S0022-1694(00)00253-5

Atieh, M., Mehltretter, S. L., Garbaragh, B., & Rudra, R. (2015). Integrative neural network model for prediction of sediment rating curve parameters for ungauged basins. Journal of Hydrology, 531, 1095–1107. https://doi.org/10.1016/j.jhydrol.2015.11.008

Belitz, K., & Stackelberg, P. E. (2021). Evaluation of six methods for correcting bias in estimates from ensemble tree machine learning regression models. Environmental Modelling & Software, 139, 105006. https://doi.org/10.1016/j.envsoft.2021.105006

Belmont, P., Gran, K. G., Schottler, S. P., Wilcock, P. R., Day, S. S., Jennings, C., Lauer, J. W., Viparelli, E., Willenbring, J. K., Engstrom, D. R., & Parker, G. (2011). Large shift in source of fine sediment in the upper Mississippi River. Environmental Science & Technology, 45, 8804–8810. https://doi.org/10.1021/es2019109

Biedenharn, D. S., Watson, C. C., & Thorne, C. R. (2008). In M. H. Garcia (Ed.), Chapter 6. Sedimentation engineering, processes measurement, modeling, and practice (pp. 355–386). Ph.D. American Society of Civil Engineers, Practice No. 110. ASCE. https://doi.org/10.1061/9780784408148.ch06

Bortnik, J., & Camporeale, E. (2021). Ten ways to apply machine learning in earth and space sciences. Eos, 102, 1–8. https://doi.org/10.1029/2021EO160257

Breiman, L. (2001). Random forests. Machine Learning, 45, 5–32. https://doi.org/10.1023/A:1010933404324

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining, 13-17-August-2016 (pp. 785–794). https://doi.org/10.1145/2939672.2939785

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Mu, L., Xie, J., Lin, M., Geng, Y., & Li, Y. (2021). XGBoost: Extreme gradient boosting. R package version 1.4.1.1. https://cran.r-project.org/web/packages/xgboost/

Cisty, M., Soldanova, V., Cyprich, F., Holubova, K., & Simor, V. (2021). Suspended sediment modelling with hydrological and climate input data. Journal of Hydroinformatics, 23(1), 192–210. https://doi.org/10.2166/hydro.2020.116

Clesceri, L. S., Greenberg, A. E., & Eaton, A. D. (1998). Standard methods for the examination of water and wastewater (20th ed.). American Public Health Association, American Water Works Association, Water Environment Federation.

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Belitz, K., & Stackelberg, P. E. (2021). Evaluation of six methods for correcting bias in estimates from ensemble tree machine learning regression models. Environmental Modelling & Software, 139, 105006. https://doi.org/10.1016/j.envsoft.2021.105006

Davies, B. E. (2005). A guide to the proper selection and use of federally approved sediment and water-quality samplers. U.S. Geological Survey, Open File Report 2005-1087 (20 p.). https://pubs.usgs.gov/of/2005/1087/

Draut, A. E., Logan, J. B., & Mastin, M. C. (2010). Channel evolution on the dammed Elwha River, Washington, USA. Geomorphology, 127(1–2), 71–87. https://doi.org/10.1016/j.geomorph.2010.12.008

Edwards, T. K., & Glysson, G. D. (1999). Field methods for measurement of fluvial sediment. U.S. Geological Survey Techniques of Water-Resources Investigations, 03-C2 (89 p.). https://doi.org/10.3133/twi03C2

Ellison, C. A., Groten, J. T., Lorenz, D. L., & Koller, K. S. (2016). Application of dimensionless sediment rating curves to predict suspended-sediment concentrations, bedload, and annual sediment loads for Rivers in Minnesota. U.S. Geological Survey Scientific Investigations Report 2016-5146 (68 p.). https://pubs.er.usgs.gov/publication/sir20165146
Ellison, C. A., Savage, B. E., & Johnson, G. D. (2014). Suspended-sediment concentrations, loads, total suspended solids, turbidity, and particle-size fractions for selected rivers in Minnesota, 2007 through 2011. U. S. Geological Survey Scientific Investigations Report, 2013-5205 (43 p.). https://pubs.usgs.gov/sir/2013/5205/

Essou, G. R. C., Sabarly, F., Lucas-Picher, P., Brissette, F., & Poulin, A. (2016). Can precipitation and temperature from meteorological reanalysis be used for hydrological modeling? Journal of Hydrometeorology, 17(7), 1929–1950. https://doi.org/10.1175/JHM-D-15-0138.1

Foufoula-Georgiou, E., Takbir, Z., Czuba, J. A., & Schwenk, J. (2015). The change of nature and the nature of change in agricultural landscapes: Hydrologic regime shifts modulate ecological transitions. Water Resources Research, 51, 6649–6671. https://doi.org/10.1002/2015WR017637

Fox, E. W., Hill, R. A., Leibowitz, S. G., Olsen, A. R., Thornbrugh, D. J., & Weber, M. H. (2017). Assessing the accuracy and stability of variable selection methods for random forest modeling in ecology. Environmental Monitoring and Assessment, 187(7), 316. https://doi.org/10.1007/s10661-017-6025-0

Francke, T., López-Tarazon, J., & Schröder, B. (2008). Estimation of suspended sediment concentration and yield using linear models, random forests and quantile regression forests. Hydrological Processes, 22, 4892–4904. https://doi.org/10.1002/hyp.7110

Ellison, C. A., Savage, B. E., & Johnson, G. D. (2014). Suspended-sediment concentration hysteresis at station 04015330 Knife River near Two Harbors, MN. U.S. Geological Survey Water-Resources Investigations Report 00-4191 (p. 14). https://pubs.usgs.gov/wri/wri004191/

Gray, J. R., & Simões, J. F. M. (2008). Estimating sediment discharge: Appendix D. In Sedimentation engineering: processes, measurements, modelling, and practice. American Society of Civil Engineers. (pp. 1067–1088). https://doi.org/10.1061/9780784401848筷

Groten, J. T., Belmont, P., Day, S. S., Jennings, C., Johnson, A., Perg, L., & Fox, E. W. (2017b). Assessment of soft computing techniques for the prediction of suspended sediment loads in rivers. Applied Sciences, 11(18), 8290. https://doi.org/10.3390/app11188290

Kondolf, G. M. (1997). Hungry water: Effects of dams and gravel mining on river channels. Environmental Management, 21(4), 533–551. https://doi.org/10.1007/s002679900048

Kuhn, M. (2008). Building predictive models in R using the caret package. Journal of Statistical Software, 28(5), 1–26. https://doi.org/10.18637/jss.v028.i05

Kuhn, M. (2021). Caret: Classification and regression training. R package version 6.0-90. https://CRAN.R-project.org/package=caret

Landers, M. N., Straub, T. D., Wood, M. S., & Domanski, M. M. (2016). Sediment acoustic index method for computing continuous suspended-sediment concentrations. U.S. Geological Survey Techniques and Methods, 03-C5 (63 p.). https://doi.org/10.3133/tm3C5

Lane, E. W. (1955). The importance of fluvial morphology in hydraulic engineering. Proceedings of the American Society of Civil Engineers, 81(art.745), 1–17.

Larsen, M. C., Gellis, A. C., Glysson, G. D., Gray J. R., & Horowitz, A. J. (2010). Fluvial sediment in the environment: A national challenge. In Proceedings of the 9th federal interagency sedimentation conference, Las Vegas, Nevada, June 27–July 1, 2010 (15 p.). https://pubs.er.usgs.gov/publication/70120717

Leopold, L. B., Wolman, M. G., & Miller, J. P. (1964). Fluvial processes in geomorphology (p. 544). Dover Publications, Inc. https://pubs.er.usgs.gov/publication/70185663

Liu, W., Birgand, F., Tian, S., & Chen, C. (2021). Event-scale hysteresis metrics to reveal processes and mechanisms controlling constituent export from watersheds: A review. Water Research, 200, 117254. https://doi.org/10.1016/j.watres.2021.117254

Liu, Y., & Allan, J. (2021). SHAPforxgboost: SHAP plots for XGBoost’ R package version 0.1.1. https://github.com/liuyanggu/SHAPforXGBoost/

Lively, R. S. (2020). Minnesota at a glance coves in Minnesota. Minnesota Geological Survey Report. University of Minnesota Digital Conservancy. https://hdl.handle.net/11299/59431

Lloyd, C. E. M., Freer, J. E., Johnes, P. J., & Collins, A. L. (2016). Testing an improved index for analyzing storm discharge–concentration hystereseis. Hydrology and Earth System Sciences, 20(2), 625–632. https://doi.org/10.5194/hess-20-625-2016

Lund, J. W., & Groten, J. T. (2022). Extreme gradient boosting machine learning models, suspended sediment, bedload, and geospatial data, Minnesota, 2007–2019. U.S. Geological Survey data release. https://doi.org/10.5066/P9VP0S51

Lundberg, S. M., Erion, G. G., & Lee, S. I. (2018). Consistent individualized feature attribution for tree ensembles. (2). https://doi.org/10.48550/arXiv.1802.03888

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4766–4775. https://doi.org/10.48550/arXiv.1705.07874

Mackay, L., Bondeld, T., Dewald, T., Johnston, J., Moore, R., & Rea, A. (2012). NHDPPlusVersion 2: User guide. U.S. Environmental Protection
Agency. http://www.horizon-systems.com/NHDPlus/NHDPlusV2/home.php

McManamy, R. A., & Derolph, C. R. (2019). A stream classification system for the conterminous United States. Scientific Data, 6, 1–18. https://doi.org/10.1038/sdata.2019.17

Minnesota Department of Natural Resources [MNDNR]. (2021). MNDNR/Minnesota Pollution Control Agency cooperative stream gaging. Minnesota Department of Natural Resources, digital data. https://www.dnr.state.mn.us/waters/csg/index.html

Minnesota Pollution Control Agency. (2011). An integrated sediment budget for the Le Sueur River Basin. https://www.pca.state.mn.us/sites/default/files/wq-lw7-29o.pdf

Minnesota Pollution Control Agency. (2021a). Total maximum daily load (TMDL) projects. https://www.pca.state.mn.us/water/water-quality-standards

Minnesota Pollution Control Agency. (2021b). Water quality standards. https://www.pca.state.mn.us/water/total-maximum-daily-load-tmdl-projects#approved-5ecc5337

Molnar, C. (2019). Interpretable machine learning: A guide for making black box models explainable. https://christophm.github.io/interpretable-ml-book/

Moore, R. B., & Dewald, T. G. (2016). The road to NHDPlus–Advancements in digital stream networks and associated catchments. Journal of the American Water Resources Association, 52(4), 890–900. https://doi.org/10.1111/1752-1688.12389

Oak Ridge National Laboratory. (2021). U.S. stream classification system. Oak Ridge National Laboratory. https://hydrosource.ornl.gov/tool/us-stream-classification-system

Ojakangas, R. W., & Matsch, C. L. (1982). Minnesota’s geology (p. 255). University of Minnesota Press.

R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/

Rasmussen, P. P., Gray, J. R., Glysson, G. D., & Ziegler, A. C. (2009). Guide- lines and procedures for computing time-series suspended-sediment concentrations and loads from in-stream turbidity-sensor and streamflow data. U.S. Geological Survey Techniques and Methods 3-C4 (53 p.). https://pubs.usgs.gov/tm/tm3c4/

Rosgen, D. L. (1994). A classification of natural rivers. Catena, 22-03, 169–199. https://doi.org/10.1016/0341-8162(94)90001-9

Rosgen, D. L. (1996). Applied river morphology. In Wildland hydrology. Pagosa Springs.

Rosgen, D. L. (2010). The application and validation of dimensionless sediment rating curves. In 2nd Joint federal interagency conference, Las Vegas, Nevada, June 27–July 1, 2010 (11 p.).

Schmidt, J. C., & Wilcock, P. R. (2008). Metrics for assessing the downstream effects of dams. Water Resources Research, 44(4), W04004. https://doi.org/10.1029/2006WR005092

Schottler, S. P., Ulrich, J., Belmont, P., Moore, R., Lauer, J. W., Engstrom, D. R., & Almdendler, J. E. (2014). Twentieth century agricultural drainage creates more erosive rivers. Hydrological Processes, 28(4), 1951–1961. https://doi.org/10.1002/hyp.9738

Simon, A., Dickerson, W., & Heins, A. (2004). Suspended-sediment transport rates at the 1.5-year recurrence interval for ecoregions of the United States: Transport conditions at the bankfull and effective discharge? Geomorphology, 58(1–4), 243–262. https://doi.org/10.1016/j.geomorph.2003.07.003

Sims, P. K., & Morey, G. G. (1972). Geology of Minnesota – A centennial volume (p. 632). Minneapolis, Minnesota Geological Survey, University of Minnesota. https://hdl.handle.net/11299/57318

Smith, H. G., & Dragovich, D. (2009). Interpreting sediment delivery processes using suspended sediment-discharge hysteresis patterns from nested upland catchments, South-Eastern Australia. Hydrological Processes, 23(17), 2415–2426. https://doi.org/10.1002/hyp.7357

Syvitski, J. P., Morehead, M. D., Bahr, D. B., & Mulder, T. (2000). Estimating fluvial sediment transport: The rating parameters. Water Resources Research, 36(9), 2747–2760. https://doi.org/10.1029/2000WR000133

U.S. Environmental Protection Agency. (2021a). Level III and IV ecoregions by state. U.S. Environmental Protection Agency. https://www.epa.gov/eco-research/level-iii-and-iv-ecoregions-state

U.S. Environmental Protection Agency. (2021b). StreamCat dataset. U.S. Environmental Protection Agency. https://www.epa.gov/national-aquatic-resource-surveys/streamcat-dataset-0

U.S. Geological Survey. (2021a). StreamStats. U.S. Geological Survey. https://streamstats.usgs.gov/ss/

U.S. Geological Survey. (2021b). USGS water data for the nation. U.S. Geological Survey National Water Information System Database. https://doi.org/10.5066/F7PP5KJN

Vaughan, A. A., Belmont, P., Hawkins, C. P., & Wilcock, P. (2017). Near-channel versus watershed controls on sediment rating curves. Journal of Geophysical Research: Earth Surface, 122(10), 1901–1923. https://doi.org/10.1002/2016JF004180

Vaughan, M. C. H., Bowden, W. B., Stanley, J. B., Vermilyea, A., Sleeper, R., Gold, A. J., Pradhanang, S. M., Inamdar, S. P., Levi, D. F., Andres, A. S., Birgand, F., & Schroth, A. W. (2017). High-frequency dissolved organic carbon and nitrate measurements reveal differences in storm hysteresis and loading in relation to land cover and seasonality. Water Resources Research, 53(7), 5345–5363. https://doi.org/10.1002/2017WR020491

Wilcock, P., Baskfield, P., Birr, A., Cooper, P., Engstrom, D., & Schottler, S. (2009). Identifying sediment sources in the Minnesota River Basin. Minnesota River Sediment Colloquium report published by Minnesota Pollution Control Agency (p. 1–16). https://www.pca.state.mn.us/sites/default/files/wq-b3-36.pdf

Zounemat-Kermani, M., Mahdavi-Meymand, A., Alizamir, M., Adarsh, S., & Yaseen, Z. M. (2020). On the complexities of sediment load modeling using integrative machine learning: Application of the great river of Loiza in Puerto Rico. Journal of Hydrology, 585, 124759. https://doi.org/10.1016/j.jhydrol.2020.124759

How to cite this article: Lund, J. W., Groten, J. T., Karwan, D. L., & Babcock, C. (2022). Using machine learning to improve predictions and provide insight into fluvial sediment transport. Hydrological Processes, 36(8), e14648. https://doi.org/10.1002/hyp.14648