Analysis of Variables Influencing Scour on Large Sand-Bed Rivers Conducted Using Field Data

Harasti, Antonija; Gilja, Gordon; Adžaga, Nikola; Žic, Mark

Download date / Datum preuzimanja: **2025-03-17**

Analysis of Variables Influencing Scour on Large Sand-Bed Rivers Conducted Using Field Data

Antonija Harasti [1], Gordon Gilja [1,*], Nikola Adžaga [2] and Mark Žic [3]

[1] Department of Hydroscience and Engineering, Faculty of Civil Engineering, University of Zagreb, Fra Andrije Kacica Miosica 26, 10000 Zagreb, Croatia; antonija.harasti@grad.unizg.hr
[2] Department of Mathematics, Faculty of Civil Engineering, University of Zagreb, Fra Andrije Kacica Miosica 26, 10000 Zagreb, Croatia; nikola.adzaga@grad.unizg.hr
[3] Division of Materials Physics, Ruder Boskovic Institute, Bijenicka Cesta 54, 10000 Zagreb, Croatia; mark.zic@irb.hr
* Correspondence: gordon.gilja@grad.unizg.hr

Abstract: Throughout the lifespan of a bridge, morphological changes in the riverbed affect the variable action-imposed loads on the structure. This emphasizes the need for accurate and reliable data that can be used in model-based projections targeted for the identification of risk associated with bridge failure induced by scour. The aim of this paper is to provide an analysis of scour depth estimation on large sand-bed rivers under the clear water regime, detect the most influential (i.e., explanatory) variables, and examine the relationship between them and scour depth as a response variable. A dataset used for the analysis was obtained from the United States Geological Survey's extensive field database of local scour at bridge piers, i.e., the Pier-Scour Database (PSDB-2014). The original database was filtered to exclude the data that did not reflect large sand-bed rivers, and several influential variables were omitted by using the principal component analysis. This reduction process resulted in 10 influential variables that were used in multiple non-linear regression scour modeling (MNLR). Two MNLR models (i.e., non-dimensional and dimensional models) were prepared for scour estimation; however, the dimensional model slightly overperformed the other one. According to the Pearson correlation coefficients (r), the most influential variables for estimating scour depth were as follows: Effective pier width ($r = 0.625$), flow depth ($r = 0.492$), and critical and local velocity ($r = 0.474$ and $r = 0.436$), respectively. In the compounded hydraulic-sediment category, critical velocity had the greatest impact (i.e., the highest correlation coefficient) on scour depth in comparison to densimetric Froude and critical Froude numbers that were characterized by correlation coefficients of $r = 0.427$ and $r = 0.323$, respectively. The remaining four variables (local and critical bed shear stress, Froude number, and particle Reynolds number) exhibited a very weak correlation with scour depth, with $r < 0.3$.

Keywords: bridge scour; sand-bed; principal component analysis; multiple non-linear regression; PSDB-2014

1. Introduction

The courses of large European rivers have been altered over the centuries of urbanization to provide support for development, such as flood protection, fairway routes, energy production, and water use for domestic and industrial consumption, agriculture, or recreation [1]. The proximity of rivers and large cities inevitably means that river flow interacts with the built environment, most often with hard and critical infrastructure [2]. Bridges stand out from other structures in the infrastructure since their elements are typically placed in the main channel, and thus, they are continuously exposed to river flow. In this context, bridges are subjected to multiple hazards of various severity that can occur simultaneously, such as flooding, scouring, and seismic loading, resulting in bridge failure.

Throughout the lifespan of the bridge, the riverbed morphology changes, which in turn affects the variable action-imposed loads on the structure [3] and increases the risk of failure [4]. Numerous researchers have identified scour as the most common cause of failure among all other hydraulic-related causes [5–7]. Climate change has a significant impact on bridge safety since their long life span does not allow a reliable estimation of the impact of climate change [8,9]. The annual costs of adaptation to reduce the risk of bridge scour under projected future climate change scenarios are estimated at €541 million in Europe for the period 2040–2070 [10]. Therefore, potential threats to bridge safety have been identified during maintenance, which ensures their resilience [11,12] and reduces potential direct and indirect losses [13]. Obstacles to reliable estimation of potential scour are numerous: the unknown accuracy of field data, the uncertainty of laboratory data upscaled to the prototype scale, and an overall lack of understanding of the interaction between the turbulent flow and the erodible riverbed. Scour is initiated when the flow velocity exceeds a threshold for entrainment of sediment, and it lasts until after the flood peak [14]. Thereafter, as the flood stage recedes, the scour hole refills and its depth decreases [15,16]. As scour is the most common cause of bridge failure, there is an apparent need for accurate and reliable data that can be used in model-based projections of the risks associated with bridge scour.

Most of the existing scour prediction models available in the literature are derived from laboratory data [17]. Using a pump and a tail gate in the laboratory provides controlled hydraulic conditions for experimental data collection. Establishing a specific flow environment in a laboratory limits the range of laboratory conclusions [18]. Furthermore, upscaling to field prototype bridge piers is a daunting task that can lead to unreliable extrapolated data. It is fair to say that laboratory-based scour models are adequate only for the limited range of variables for which they were developed. With the aim of avoiding extrapolation issues, the accuracy of these models should be additionally confirmed with field measurements. Therefore, it is necessary to gain a perspective on scour potential under field conditions [19]. Regardless of the data collection method, most of the scour prediction equations are based on conventional regression methods to provide explicit expressions that are easy-to-use and readily applicable to practice. However, conventional methods are limited to conditions similar to those under which the data were collected. This leads to biased results that often overestimate scour depth, i.e., a so-called conservative approach [20,21]. To describe the scour process, the scour database should be extensive and structured because many local variables are required to capture the complex flow environment in the bridge opening.

Recently, artificial intelligence (AI) has become a prominent tool that has the capability to capture complex scour processes governed by a large number of influencing variables. Recent studies have used AI techniques for scour prediction. Dong et al. [22] developed four machine-learning models: a back propagation neural network (BP NN), a genetic algorithm neural network (GA NN), a convolutional neural network (CNN), and a deep belief network (DBN NN) to predict local scour based on a combination of laboratory tests and field observations. Pandey et al. [23] developed scour equations using a genetic algorithm (GA) on a large experimental dataset. Rady [24] applied the adaptive neural fuzzy inference system (ANFIS) and genetic programming (GP) methods for pier scour assessment. Muzzammil et al. [25] implemented gene expression programming (GEP) to derive a new equation for pier scour estimation in a bed of a mixture of clay and sand. To evaluate the performances of developed AI models, the abovementioned authors conducted comparisons with traditional methods (regression technique, dimensional analysis, and functional relations) and concluded that AI methods showed better scour depth prediction. Although AI-based models have proven to be more accurate than conventional models, they have a tendency to overfit data. Generally, AI models learn to fit the selected data set as closely as possible, but learning the model for a particular data set might make it inappropriate for unseen data. Moreover, in the

case of insufficient scouring of input data, it might occur that the AI model has a biased outcome due to overfitting issues.

The United States Geological Survey (USGS) has published an extensive field database of local scour at bridge piers, "A Pier-Scour Database" [26], collected for 270 bridges spanning over 433 rivers in 29 countries, resulting in a total of 1858 measurements of flow, bridge geometry, and river morphology. The database is structured as a digital spreadsheet with measurements organized as rows and values of variables under each measurement organized in columns, e.g., scour depth around bridge piers (d_s), pier width (b), pier length (l), angle of flow attack (θ), effective pier width (b_{ef}), pier type and nose shape, approach flow depth (y), approach flow velocity (v), bed material type, median sediment size (d_{50}), geometric standard deviation of sediment particles (σ_g), approximate recurrence interval for measured flow rate (RI), stream slope (S), date, type, time, and location of measurements, etc.

As this extensive PSDB-2014 database has been available since 2014, it has been used in several studies focusing on pier scour. Benedict and Caldwell [18] aimed to define the maximum scour depth in South Carolina by developing an envelope curve. They evaluate the upper bound of the flow-scour relationship by combining pier scour data from other sources for effective pier widths up to 9 m. Benedict and Knight [27] evaluated the HEC-18 model by comparing the PSDB-2014 experimental database since the HEC-18 model was developed on a scarce laboratory dataset. Rathod and Manekar [28] designed a new universal scour model utilizing the GEP method on the PSDB-2014 database. They compared the new GEP model with five existing conventional scour equations and found a good correlation with Jain and Fischer model Equation (14). Pandey et al. [29] used the PSDB-2014 database to modify laboratory-based Melville and Coleman's equation and make it applicable to field conditions. The authors filtered the PSDB-2014 database to retain only non-uniform gravel beds in clear-water conditions and introduced new *K*-factors. Ali and Günal [30] collected a new laboratory dataset and complemented it with the corresponding PSDB-2014 laboratory subset to train the ANN model with the Levenberg-Marquardt algorithm. Shahriar et al. [31] used the PSDB-2014 database to compare four pier scour models and assess their performance through error statistics and the probabilistic distribution of predictions.

This paper presents a preliminary analysis of the variables influencing the pier scour on large sand-bed rivers under the clear-water regime. Due to the lack of scour equations based on field measurements and the low accuracy of the laboratory-derived equations deployed for field data, this study uses a comprehensive PSDB-2014 database. The objective of this paper is to identify the most influential variables affecting scour on large sand-bed rivers using the structure as presented by the flow chart (Figure 1). First, the collection of a large database is required, i.e., in our case, obtaining the PSDB-2014 database. Second, filling the original database with more potentially important variables that were not originally included. Herein, these variables are designed as additionally calculated variables, such as Froude number, shear stress, and other similar variables that cannot be directly measured. Third, filter measurements from the collected database that are not related to specific environmental conditions, i.e., eliminate measurements to obtain a relevant dataset. Fourth, reduce unnecessary variables to the filtered subset with only influencing variables using principal component analysis (PCA). Fifth, perform sensitivity analysis to examine the strength of correlation between the influencing variables and the scour depth using the Pearson correlation matrix. At this point, a reliable scour prediction model can be designed and tested. In this paper, a dimensional regression-based model is proposed for scour depth estimation in large sand-bed rivers in clear water regimes. Several already existing models applicable to the aforementioned conditions were selected and compared with the model developed in this study to put this investigation in a relative perspective.
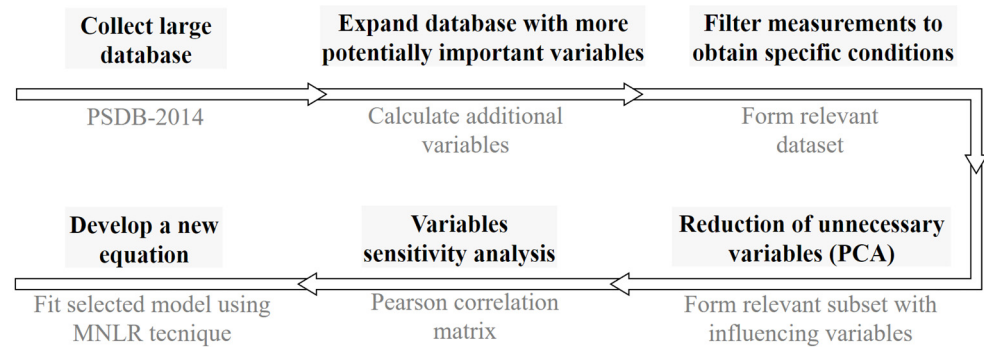
Figure 1. Flowchart of the proposed methodology to identify the most influential variables before developing a new model.

2. Methodology

2.1. Influencing Variables

The original PSDB-2014 database covers a wide range of directly measured flow, sediment, and geometry variables. From all measured variables in the original PSDB-2014 database, the following 13 variables were selected for the purpose of this study: d_s, b, l, θ, b_{ef}, pier type, nose shape, y, v, d_{50}, σ_g, RI, and S. Since this dataset contains only directly measured variables, it is of keen interest to include several other variables commonly used in scour prediction equations. Thus, this study introduces 8 additional variables calculated using the measured data from the PSDB-2014 database: channel width (B), Froude number (Fr), local bed shear stress (τ), critical bed shear stress (τ_c), critical velocity (v_c), critical Froude number for incipient motion (Fr_c), densimetric Froude number (Fr_d), and particle Reynolds number (Re_p) (Table 1). By including the aforementioned additional variables, information about the flow regime, clear-water or live-bed scour, and particle entrainment threshold conditions are taken into account. These additional variables are not given in the original PSDB-2014 database, but they are often used in scour research (e.g., [32,33]). With this addition, a comprehensive dataset of 21 independent variables was created with essential information on flow, sediment, and morphology at each location.

Table 1. Additional variables included in the analysis.

Additional Variable	Equation	
Froude number	$Fr = \dfrac{v}{\sqrt{g \cdot y}}$	(1)
local shear stress	$\tau = \gamma_w \cdot y \cdot S$	(2)
critical shear stress [34]	$\tau_c = 0.25 \cdot \left(d_{50} \cdot \left[\dfrac{\left(\frac{\gamma_s}{\gamma_w} - 1 \right) \cdot g}{v^{2*}} \right]^{\frac{1}{3}} \right)^{-0.6} \cdot g \cdot (\rho_s - \rho_w) \cdot d_{50} \cdot tan(\Phi^*)$	(3)
critical velocity [35]	$v_c = \sqrt{\rho_{rel}^* \cdot g \cdot d_{50}} \cdot \left(0.0024 \cdot \left(\frac{y}{d_{50}} \right) + 2.34 \right)$	(4)
critical Froude number	$Fr_c = \dfrac{v_c}{\sqrt{g \cdot y}}$	(5)
densimetric Froude number	$Fr_d = \dfrac{v}{\sqrt{(\rho_{rel}^* - 1) \cdot g \cdot d_{50}}}$	(6)
particle Reynolds number	$Re_p = \dfrac{d_{50} \cdot \sqrt{\rho_{rel}^* \cdot g \cdot d_{50}}}{v^*}$	(7)

* v is kinematic viscosity of fluid. Φ is the angle of repose of the particle [36]. ρ_{rel} is the submerged relative mass density of sediment particles.

2.2. Data Filtering

After the expansion of the subset (containing 13 initial variables) with the additional computed 8 variables, data filtering was conducted to retain only data relevant for the aim of the paper, i.e., to study measurements referring to large sand-bed rivers. To obtain this

data, several sets of filtering criteria were defined that correspond to the flow regime and riverbed morphology of the large rivers.

The first set of filtering criteria is related to the flow regime. The flow of large rivers is subcritical due to the nature of the flow in lowland areas; hence, the Froude number can be used to distinguish subcritical and supercritical flow regimes. In open-channel flow, the critical flow regime is lower than its theoretical value ($Fr = 1$) since the flow transition from wide channel to contraction is gradual. Different studies identified the transition between subcritical and supercritical flow regimes to be within the interval $0.7 < Fr < 0.85$ for field conditions. For example, in the study by Azamathulla et al. [37], Fr values varied from 0.7 to 0.8, as in several other reports [18,29,38]. In this study, measurements with $Fr > 0.75$ were filtered out, i.e., applying Rady's approach [24]. Additionally, erroneous data related to Froude number were also filtered out—measurements of approach flow velocity (v) with missing or zero values were removed.

The second set of filtering criteria was applied to preserve the data corresponding to large sand-bed rivers. To restrain the dataset to sand-bed rivers, the following criteria were used: sediment particle median size (d_{50}) ranging between 0.0625 mm and 2 mm. The geometrical standard deviation of the sediment (σ_g) presents a measure of the non-uniformity of the sediment [28]. Uniform sediment gradation associated with uniform sand is considered to be less than 1.3 [27] because larger σ_g values would indicate the presence of coarser gravel particles. After applied filtration, the average value of σ_g is 3.0, which shows slightly non-uniform sediments, which indicates potential bed armoring and consequently reduces scour depth [39]. To restrain the dataset to large rivers, measurements of flow depth (y) lower than 1.5 m were excluded from further analysis. The original PSDB-2014 database [26] contains no information regarding the river widths. Benedict and Knight [40] have applied basin data to correlate known channel widths against drainage area, flow depth, and stream slopes. To validate filtering criteria for large rivers, correlation by Benedict and Knight [40] was used herein to calculate channel widths, of which 69% are greater than 30 m.

Finally, the filter that takes into account the scour depth error margin range has been applied. If only large rivers are considered, the accuracy of riverbed morphology measurement has to be taken into account, especially when equipment limitations do not allow detailed riverbed mapping. Therefore, from the original PSDB-2014 database, an additional 140 measurements were excluded where scour depth was in the error margin range of less than 0.2 m. Moreover, the original PSDB-2014 database contains multiple measurements collected at the same location under different hydrologic events, including multiple scour depth values. To prepare the database for the analysis of cause-and-effect relationships, it is important to maintain the independence of the measurements, i.e., only one measurement should be retained for each bridge. Therefore, for the second set of filtering criteria, only the measurements with the largest scour depth were considered (568 exclusions). The association of scour depth with hydraulic conditions in the PSDB-2014 database is not extensively documented. For some multiple measurements, scour depth and flow depth are inversely proportional, which may be a consequence of an undocumented hydraulic event, e.g., a recent flood. In addition to the second set of filtering criteria, only the most recent measurements have been retained (291 exclusions). Some measurements date from the early 1900s, so their reliability is low due to technical limitations such as the development of guidelines and instruments for bathymetric and hydraulic surveys. Overall, a total of 859 measurements were excluded.

After filtering the original PSDB-2014 database, a total of 348 independent measurements have remained relevant for estimating maximum scour depth on large, alluvial, and sand-bed rivers. Since scour depth is 10% greater in clear water than in live-bed conditions [41], the clear-water regime is the desired condition for estimating scour depth in this study. To distinguish whether flow conditions are clear-water or live-bed scour, the expression $Fr - Fr_c > 0.2$ by Jain and Fischer is used for validation [38]. Interestingly, after filtering by using the aforesaid expression, 98% of the 348 remaining measurements

fall within the clear-water scour conditions, even though the clear-water criterion was not applied in the filter, which validates the reliability of the applied data filters. The range of original and additional variables before and after filtration is shown in Table 2.

Table 2. Range of original and additional variables with their symbols, measurement units, ranges, average values, and standard deviations before and after filtering.

Variable		Before Filtering			After Filtering		
		Range	Average	Standard Deviation	Range	Average	Standard Deviation
Original	d_s	0–10.4	1.1	1.3	0.21–7.8	1.5	1.3
	b_{ef}	0.21–28.7	2.3	2.5	0.24–11.6	2.0	1.9
	b	0.21–19.5	1.6	1.6	0.24–11.6	1.3	1.4
	l	0.21–39.6	6.3	5.3	0.37–25.3	6.3	4.7
	θ	0–85.0	6.1	10.3	0–600	6.6	9.6
	d_{50}	0.001–228.6	14.7	25.0	0.06–1.82	0.59	0.41
	y	0–22.5	3.9	3.2	1.5–22.5	5.6	3.4
	v	0–5.4	1.4	0.8	0.20–3.9	1.4	0.71
	RI	1–500	53.6	50.9	1–500	63.1	69.4
	σ_g	1.2–20.3	3.3	2.8	1.4–20.3	3.0	1.2
	S	0.00007–0.02	0.00086	0.00152	0.00007–0.0036	0.00052	0.00044
Additional	B	5.4–692.5	71.5	67.1	10.5–692.5	75.3	80.7
	v_c	0.15–55.7	2.4	3.2	0.76–11.6	2.8	1.7
	τ	0–180	21.3	21.8	0.015–1.7	0.25	0.21
	τ_c	0.025–3.9	0.87	0.77	0.13–0.56	0.32	0.10
	Fr	0–1.98	0.28	0.21	0.039–0.55	0.20	0.10
	Fr_c	0.19–5.77	0.40	0.34	0.19–1.0	0.38	0.13
	Fr_d	0–629	11.1	23.4	1.6–88.9	16.0	9.9
	Re_p	0.0025–274,834	8222	19,074	1.2–195.2	42.1	43.7
Original	Pier type	Single and group			Not affected		
	Pier nose shape (drag coefficient) [42]	Cylindrical (1.2), Round (1.33), Square (2.0), Sharp (1.0), Triangular (1.72)					

2.3. Variable Reduction

Estimation of scour depth requires the detection of the most significant input variables relevant to the complex scour process in large sand-bed rivers. This can be achieved using principal component analysis (PCA), which has already been proven to be an effective decision-making tool for dimensionality reduction when many variables are involved [43]. PCA determines which variable influences scour depth the most by analyzing how much each variable contributes to the variance. Therefore, PCA was applied to the filtered dataset of 348 measurements and a total of 21 variables (including the observed scour depth d_s). PCA interprets data with principal components—linear combinations of the input variables that are orthogonal to each other. In this study, there are 21 principal components as well as 21 input variables. The first principal component explains the largest amount of variability in the dataset; the second principal component accounts for the next largest variance; etc. The principal components are lines in the coordinate system with a corresponding eigenvalue—the sum of the squared distances between the orthogonally projected observed data onto the line and the origin of the coordinate system.

Variable reduction in PCA analysis is based on the number of retained principal components without diminishing the total variance of the scour dataset. A commonly used criterion for reducing the number of variables is the Kaiser-Guttman criterion, which considers only those principal components whose eigenvalues are greater than 1. Applying the Kaiser-Guttman criterion for variable reduction would keep the first 7 principal components and thus eliminate only 2 variables. Since the Kaiser-Guttman criterion was previously proven to be inaccurate because it overestimates the number of principal components [44], it was decided to use a more stringent criterion. If we discard the principal components whose eigenvalues are less than 2, then the first three principal components remain. The first three principal components explain 56% of the variance of the entire dataset. The variables that contribute the most to the first three principal components

can be identified as influential (b_{ef}, y, d_{50}, v, v_{cr}, τ, τ_{cr}, Fr, Fr_c, Fr_d, Re_p, B, b, l, and S), while other variables (σ_g, θ, RI, pier type, and nose shape) can be excluded from further analysis. The results of the PCA analysis are presented in the form of a loading plot, in which the variable vectors are positioned with respect to the first two principal components (Figure 2a). Variables located near the center of the coordinate system have the smallest impact on the variance of the dataset (within the inner dashed circle) and are therefore eliminated from further analysis. When two variable vectors are perpendicular to each other, they are not correlated at all, while variables whose vectors are close to each other are highly correlated. The contribution of each variable to the first three principal components can be evaluated by the squared cosine values (Figure 2b). Since the principal components are axes of the rotated coordinate system, the squared cosine represents the quality of the variables after rotation. The values in bold are the maximum squared cosine values for each variable. Therefore, only the variables that have the maximum squared cosine for one of the first three principal components are retained for further analysis.
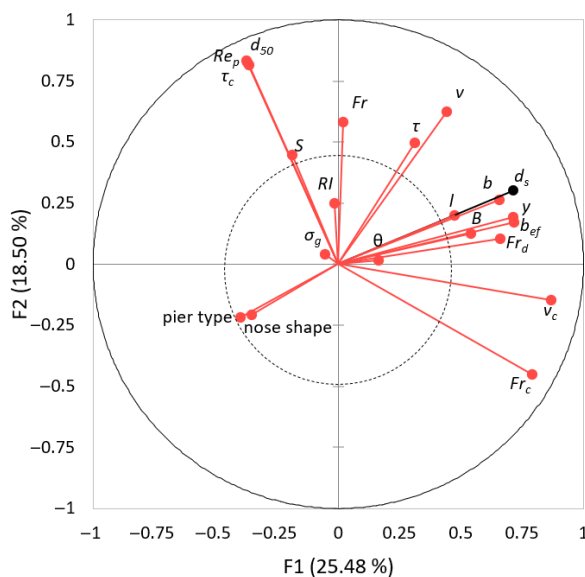
	F1	F2	F3
d_s	**0.513**	0.091	0.000
b_{ef}	**0.514**	0.030	0.009
y	**0.508**	0.038	0.120
d_{50}	0.140	**0.691**	0.123
v	0.196	**0.389**	0.154
v_{cr}	**0.754**	0.022	0.013
τ	0.097	**0.248**	0.126
τ_{cr}	0.139	**0.680**	0.126
Fr	0.000	0.338	**0.387**
Fr_c	**0.623**	0.203	0.005
Fr_d	**0.436**	0.011	0.337
Re_p	0.133	**0.665**	0.115
B	**0.294**	0.016	0.128
b	**0.430**	0.070	0.000
l	**0.224**	0.039	0.174
S	0.035	0.200	**0.297**
σ_g	**0.003**	0.002	0.001
θ	**0.026**	0.000	0.003
RI	0.000	0.062	**0.220**
Pier	**0.160**	0.048	0.119
Nose	**0.125**	0.044	0.037

(**a**) 　　　　　　　　　　　　　　　　　　　　　 (**b**)

Figure 2. Results of PCA analysis: (**a**) Loading plot that presents a total of 21 variables in the coordinate system with the first two principal components on the axes; (**b**) Contribution of each variable to the retained first three principal components throughout the squared cosines of the variables. Values in bold correspond for each variable to the principal component for which the squared cosine is the largest.

Based on the PCA results, 5 variables were discarded from further analyses, and all remaining variables can be classified into the following categories: pier geometry (b, l, and b_{ef}), hydraulic (y, v, τ, and Fr), compounded hydraulic-sediment variables (v_c, Fr_c, and Fr_d), sediment properties (d_{50}, τ_c, and Re_p), and channel properties (B and S). Previous studies have concluded that d_{50} can be eliminated from the development of the scour equation since its range of values is small when compared to the other variables, which consequently leads to biased results [19]. Since only field data are used in this study, d_{50} is expected to be negligible due to reaching large values of the ratio b/d_{50} [45,46]. In order to obtain information on sediment properties, only τc and $Re p$ were retained for developing the equation, while d_{50} was discarded.

Similarly, the variables describing pier geometry (θ, b, l and b_{ef}) are also close together. The angle of flow attack (θ) is not a reliable variable for scour prediction because flow direction depends on the water level and changes over time. This makes the measurement of θ arbitrary and should be used only when it can be reliably estimated [24]. The former debate was supported by the PCA results, which suggested that θ and l should not be used in scour prediction, leaving b and b_{ef} as the only descriptors of pier geometry. Taking into consideration the difficult approximation of complex pier geometry (in the case of a group of piers, pile caps, etc.), this study uses b_{ef} as the pier geometry variable, which is reduced to b when no other data is known but can be adjusted for pier alignment when necessary.

Both variables from the channel properties category (B and S) can be excluded. Although river width was an important variable for validating the filter applied to obtain a database corresponding to the large rivers, it is not necessary to include it in the development of the equation since this paper focuses on local scour. Stream slope is a local variable, originally measured in PSDB-2014 near the bridge. Since S was only used for estimating shear stress, a variable that estimates the drag force of flowing water, it can be eliminated from scour equation development.

The remaining variables are a combination of commonly used variables in scour prediction as well as variables that are rarely used in scour research on field conditions (τ, τ_c, Re_p). Although τ has been previously recognized as an influential variable for scour prediction [47], it is usually omitted from conventional scour equations due to difficulties in obtaining direct measurements in a complex flow environment [19]. However, τ and v are both variables that reflect drag forces in front of the pier. Since it is challenging to measure v in situ during the flood, perhaps it can be replaced by τ. Locally measured variables (y and S) in the original PSDB-2014 database allow τ to be estimated for each pier. Since it is not sufficiently investigated how much local value of τ contributes to d_s compared to v, its importance for the local scour process will be evaluated in this study. The incipient motion of sediment starts when τ exceeds the critical shear stress (τ_c). Instead, in equations for estimating scour depth, v_c is mainly used because it is easier to measure and is proportional to τ_c. In this study, it is decided to retain both critical variables τ_c and v_c, in order to evaluate their contribution to d_s estimation. Particle Reynolds number (Re_p), which is a measure of eddy currents around particles, is an important variable for estimating the transport of sediment mixtures [48]. Previous research by Vonkeman and Basson [49] demonstrates that replacing v_c with Re_p in the HEC-18 scour equation can improve scour depth prediction.

After the variable reduction analysis performed on the filtered dataset, 10 influential variables remained (Table 3), forming a filtered subset relevant to the development of functional dependencies. Since the focus of this study is on the functional dependencies between scour depth and selected variables, dependent variables were retained with the aim of testing which of them has a greater impact on scour depth. Full variable names can be seen at the end of a paper in the list of symbols.

Table 3. Influential variables selected for developing new regression-based models.

Influential variables	b_{ef}	y	v	v_c	τ	τ_c	Fr	Fr_c	Fr_d	Re_p

3. Results

3.1. MNLR—Multiple Nonlinear Regression

Multiple nonlinear regression (MNLR) is a mathematical method that establishes a functional dependence between variables to predict the target variable. The optimization process finds the best fit for the model based on the principle of least squares. In Section 2.3, 10 influential variables were chosen as suitable for scour analysis on large sand-bed rivers. In order to model the data between the response d_s and the explanatory influential variables, the following dimensional and non-dimensional models were proposed:

$$d_s = a \cdot \left(b_{ef}\right)^b \cdot (y)^c \cdot (v)^d \cdot (v_c)^e \cdot (\tau)^f \cdot (\tau_c)^g \cdot (Fr)^h \cdot (Fr_c)^i \cdot (Fr_d)^j \cdot \left(Re_p\right)^k, \tag{8}$$

and non-dimensional model:

$$\frac{d_s}{y} = a \cdot \left(\frac{b_{ef}}{y}\right)^b \cdot \left(\frac{v}{v_c}\right)^d \cdot \left(\frac{\tau}{\tau_c}\right)^f \cdot (Fr)^h \cdot (Fr_c)^i \cdot (Fr_d)^j \cdot \left(Re_p\right)^k. \tag{9}$$

Both models were fitted to the filtered dataset by using MNLR, and the extracted parameters (*a*, *b*, *c*, *d*, *e*, *f*, *g*, *h*, *i*, *j*, and *k*) were used to form two new regression-based models presented in Table 4.

Table 4. Parameters obtained applying MNLR technique on both dimensional and non-dimensional models.

	a	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*	*j*	*k*
dimensional	3.29	0.49	1.19	−0.91	−0.99	−0.011	0.019	0.38	0.26	0.97	−0.44
non-dimensional	0.002	0.48		−0.90		−0.047		−0.33	−2.50	1.16	−0.094

The performance of MNLR dimensional and non-dimensional models was validated with a commonly used statistical measure, i.e., the coefficient of determination (R^2). R^2 shows how closely the scour depth, predicted by the MNLR dimensional and nondimensional equations, resembles the observed scour depth with values of 0.51 and 0.45, respectively. The same observation was made by Ali and Günal [30], who noticed that dimensional scour data were more accurate than those based on dimensionless data. Based on R^2, the dimensional MNLR model was adopted for further analysis. Too many influencing variables reduce the number of degrees of freedom and also increase R^2. Hereby, the adjusted R^2 is useful because it corrects the original R^2 in such a way that if the number of variables increases, the number of degrees of freedom decreases, as does the adjusted R^2. For the developed model, the adjusted $R^2 = 0.50$ does not change significantly in comparison with the original R^2, which confirms that the number of selected variables is sufficient and not overused.

The dependency between measured scour depth values (d_s) and those predicted by the MNLR dimensional model ($d_{s,pred}$) is presented in Figure 3. The gray continuous line shows ideal agreement between measured and predicted values, i.e., $d_{s,pred} = d_s$. Data points that lie below the line of agreement indicate overprediction (54% of predicted values are greater than actual measured values), while data points that lie above the abscissa represent underprediction (46% of predicted values are lower than measured values).
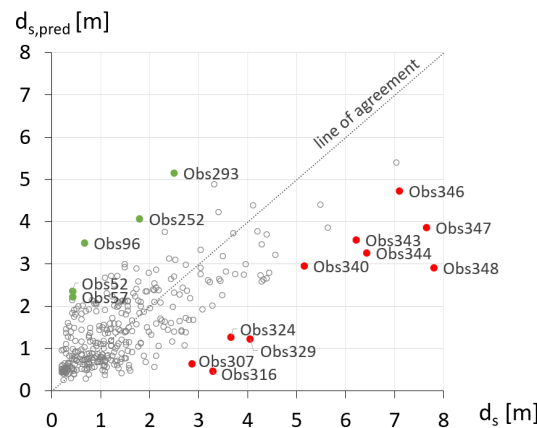

Figure 3. Scattered plot of measured (d_s) and scour depth predicted ($d_{s,pred}$) by MNLR dimensional model. Isolated points are outliers where red dots present underprediction and green dots present overprediction.

A prediction interval of 95% was calculated using the following equation:

$$= d_s \mp t_{95} \cdot SE \cdot \sqrt{1 + \frac{1}{n} + \frac{\left(d_s - \overline{d_s}\right)}{SS_{xx}}} \qquad (10)$$

where t_{95} is the two-tailed Student's T-Distribution for 95% fit, SE is the standard error, which considers the squared ratio of the residual sum of squares by degrees of freedom, n is the number of entries, $(\overline{d_s})$ is the average of all measured scour depths, and finally SS_{xx} is the sum of squares of the deviations of measured scour depths from their mean value.

The individual dependance of each variable on the residual (the difference between measured and predicted scour depth) is presented as a series of graphs merged into Figure 4. Upper and lower limits represent the 95% prediction intervals in Equation (10) illustrated by the red dashed lines for the purpose of detecting potential outliers. This means that 95% of all measurements would deviate from predicted values within the region between the dashed red lines. If the data falls outside the 95% prediction interval region, it can be marked as a potential outlier. The red dots show the data with the best prediction—residuals within 5% of the total deviation. A common feature for all the graphs is the homoscedasticity of residuals over the entire range of explanatory variables, indicating that there is no bias in the model in any part of the data range.
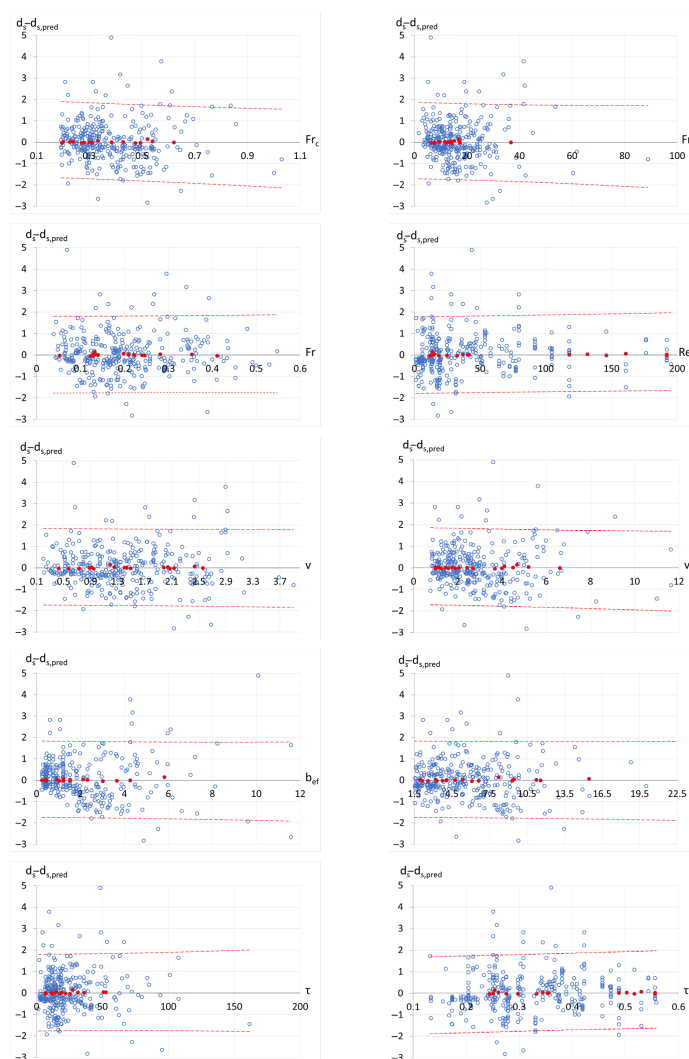

Figure 4. Dependence of influential variables on the scour depth residual. Upper and lower limits illustrated by the red dashed lines represent the 95% prediction intervals.

Since all the scour data only contains measured variables, there is no additional metadata that could be used to expand the filtered dataset, and therefore only collected variables were analyzed in order to detect outliers. All points lying outside of the prediction interval were considered potential outliers. There are 5 outliers detected based on model overprediction, whose residuals have a negative value, and 10 outliers based on model underprediction, whose residuals have a positive value. Outliers are detected based on prediction intervals in Figure 4 and presented as isolated points with associated IDs in Figure 3, where green dots illustrate overprediction and red dots underprediction. Potential outliers and their corresponding influential variable values are presented in Table 5, sorted from minimal to maximum residual. The lack of data does not allow the detection of true outliers, and therefore, outliers can only be detected based on their common features. The only identifiable common feature of all potential outliers is the Froude number value. One of the data points (#Obs348) has a very low Froude number value ($Fr = 0.07$), making it unreliable, while the other one (#Obs293) is at the boundary conditions for clear-water scour ($Fr - Fr_c = 0.06$). Therefore, of the initial 15 potential outliers, two were detected as true outliers and removed from further analyses. When outliers were removed from the model, the performance of the dimensional model increased from $R^2 = 0.50$ to $R^2 = 0.54$.

Table 5. List of potential outliers with corresponding measured and calculated influential variable values.

ID	b_{ef}	y	v	v_c	τ	τ_c	Fr	Fr_c	Fr_d	Re_p	d_s	$d_{s,pred}$	Residuals
Obs96	4.88	9.81	2.13	5.11	38.51	0.27	0.22	0.52	27.57	17.90	0.67	3.50	−2.83
Obs293	11.58	4.85	2.68	2.30	95.08	0.31	0.39	0.33	29.82	28.11	2.50	5.15	−2.65
Obs252	5.54	13.47	2.35	7.44	72.45	0.26	0.20	0.65	32.61	14.39	1.80	4.08	−2.28
Obs52	9.63	3.60	0.79	1.30	7.76	0.49	0.13	0.22	5.46	117.86	0.43	2.36	−1.93
Obs57	2.50	9.33	1.55	4.09	17.38	0.33	0.16	0.43	16.63	31.55	0.43	2.22	−1.79
Obs340	5.97	8.35	1.22	3.28	11.47	0.37	0.13	0.36	11.37	47.57	5.15	2.95	2.20
Obs307	0.61	2.77	1.13	1.14	5.17	0.42	0.22	0.22	8.86	79.52	2.87	0.65	2.22
Obs346	6.10	22.52	2.43	9.11	66.29	0.34	0.16	0.61	24.65	36.96	7.10	4.73	2.37
Obs324	1.07	4.54	1.77	2.17	53.46	0.31	0.26	0.33	19.65	28.11	3.66	1.27	2.38
Obs343	4.33	5.67	2.93	3.32	32.81	0.25	0.39	0.45	41.99	13.07	6.22	3.57	2.65
Obs329	1.07	4.15	1.71	2.00	48.80	0.31	0.27	0.31	18.97	28.11	4.05	1.23	2.82
Obs316	0.61	2.19	0.67	0.97	4.52	0.42	0.14	0.21	5.27	79.52	3.29	0.47	2.82
Obs344	4.37	5.21	2.44	2.98	16.62	0.26	0.34	0.42	33.88	14.39	6.43	3.26	3.17
Obs347	4.27	9.78	2.90	5.62	9.60	0.25	0.30	0.57	41.55	13.07	7.65	3.86	3.79
Obs348	10.09	9.00	0.65	3.60	48.40	0.36	0.07	0.38	6.24	43.61	7.80	2.90	4.90

3.2. Comparison with Different Scour Models

To evaluate the relative performance of the dimensional model developed on specifically filtered data in the present study, a comparison with already developed empirical scour models was conducted (Figure 5). The link between selected models (Table 6) is their applicability for scour depth prediction in sand-bed rivers during the clear-water regime. However, there are some differences between them based on: variables used for scour prediction; source of large-scale data (field, combination of laboratory and field, and numerical data); and equation development technique (1 conventional MNLR equation, 2 recent MNLR equations, and 2 GEP methods). The conventional MNLR technique is an approach to developing functional dependencies using dimensional analysis and the Buckingham PI theorem to attain dimensional consistency. GEP is one of the evolutionary algorithms that can create simple and explicit equations and is therefore often used for developing scour depth expressions. Two governing variables that are included in all scour equations are water depth (y) and pier width (b), as presented in Table 6.

The comparisons between previously developed empirical models (Table 6) and the dimensional model developed in this study, Equation (17), are presented in Figure 5, where the line of agreement indicates the perfect match between observed and predicted data, i.e., $d_{s,pred} = d_s$. The circle points, obtained by using our dimensional model and the filtered dataset, are the same as in Figure 3. However, Figure 5 also displays the lines that were constructed by using predicted $d_{s,pred}$ data obtained by utilizing different models from Table 6. The following discussions will explain the performance of all models from Table 6.
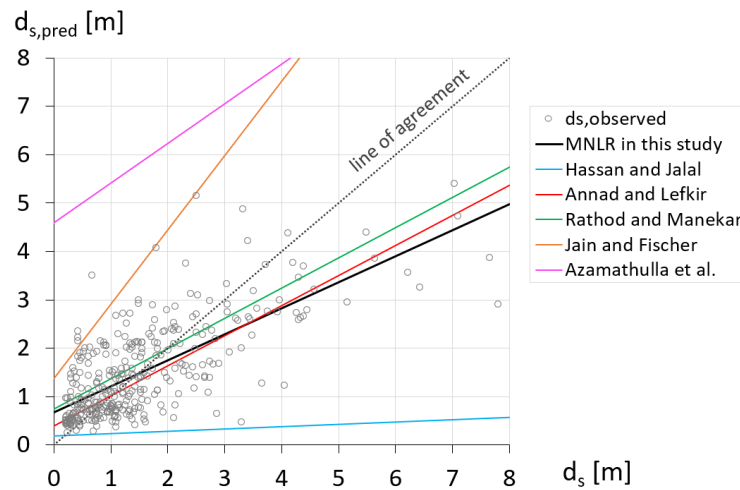
Figure 5. Comparison of dimensional model developed in this study with various previously developed scour equations (Table 6). The displayed data (i.e., circles) are obtained by using our non-dimensional model and the filtered data. The lines present data computed by different models when using the filtered data.

Table 6. Summary of selected previously developed scour models and two models developed in this paper.

Author	Dataset	Method	Equation	
Annad and Lefkir, 2022 [19]	field (PSDB-2014)	MNLR	$d_s = 0.318 \cdot K_1^* \cdot K_2^* \cdot b^{0.76} \cdot y^{0.515}$	(11)
Rathod and Manekar, 2022 [28]	field and laboratory (PSDB-2014)	GEP	$d_s = y \cdot \left[\left(\frac{b}{b+1.6 \cdot y} \right) + \left(\frac{b \cdot v}{y \cdot (b+v)} \right) + \left(\frac{\frac{b}{y}}{\frac{34.12 \cdot y}{Fr_d} + 2\sigma_g} \right) \right]$	(12)
Hassan and Jalal, 2021 [50]	numerical on a large scale	GEP	$d_s = b \cdot \left[7.24 \cdot \frac{v}{v_c} \cdot \frac{b}{B} - \left(\frac{v}{v_c} \right)^2 \cdot \frac{b}{B} - \frac{v}{v_c} \cdot \left(\frac{b}{B} \right)^2 + \frac{\frac{b}{B}}{\frac{b}{B} \cdot \frac{v}{v_c} - \frac{y}{b} - \frac{b}{B}} + \left\{ \frac{y}{b} \cdot \frac{b}{B} \cdot Fr_{pier}^* + \left(\frac{b}{B} \right)^2 \cdot Fr_{pier}^* \right\} \cdot K_2^* \cdot \frac{v}{v_c} \right]$	(13)
Jain and Fischer, 1979 [38]	field and laboratory	conventional MNLR	$d_s = 1.84 \cdot b \cdot \left(\frac{y}{b} \right)^{0.3} \cdot (Fr_c)^{0.25}$	(14)
Azamathulla et al., 2010 [37]	field	MNLR	$d_s = 1.82 \cdot y \cdot \left(\frac{d_{50}}{y} \right)^{0.042} \cdot \left(\frac{b}{y} \right)^{-0.28} \cdot \left(\frac{L}{y} \right)^{-0.37} \cdot Fr^{0.42} \cdot \sigma_g^{-0.031}$	(15)
Our non-dimensional model	field (PSDB-2014)	MNLR	$\frac{d_s}{y} = 0.002 \cdot \left(\frac{b_{ef}}{y} \right)^{0.48} \cdot \left(\frac{v}{v_c} \right)^{-0.90} \cdot \left(\frac{\tau}{\tau_c} \right)^{-0.047} \cdot (Fr)^{0.33} \cdot (Fr_c)^{-2.5} \cdot (Fr_d)^{1.16} \cdot (Re_p)^{-0.094}.$	(16)
Our dimensional model	field (PSDB-2014)	MNLR	$d_s = 3.29 \cdot (b_{ef})^{0.49} \cdot (y)^{1.19} \cdot (v)^{-0.91} \cdot (v_c)^{-0.99} \cdot (\tau)^{-0.011} \cdot (\tau_c)^{0.019} \cdot (Fr)^{0.38} \cdot (Fr_c)^{0.26} \cdot (Fr_d)^{0.97} \cdot (Re_p)^{-0.44}$	(17)

* K_1 is 1.2 for live-bed scour and 1 for clear-water scour. K_2 is 1.1 for round, 1 for square, and 0.9 for a sharply pierced nose. Fr_{pier} is the pier Froude number, calculated as follows: $Fr_{pier} = \frac{v}{\sqrt{g \cdot b_{ef}}}$.

Annad and Lefkir [19] (AL abbreviated) proposed a new scour equation obtained by the same dimensional MNLR technique and the identical PSDB-2014 database as used in this study. The authors have retained almost all observed measurements (1249 of 1858), except for those that lack information regarding bridge pier shape. However, they have included only 4 variables: *b*, *y*, pier shape nose correction factor (K_2), and live-bed vs. clear water correction factor (K_1). Considering that their approach was similar to ours (the same database and the same modeling technique), their equation almost matches the dimensional model developed in this study (Equation (17)). According to Figure 5, there are similarities between the performance of our model and that of the AL model that can be assigned to the fact that in the AL model there was no restrictive data filtering, i.e., almost the whole span of measurements was considered.

Rathod and Manekar [28] (RM abbreviated) developed the GEP model for scour depth estimation using the PSDB-2014 database. In order to develop a unique and universal scour equation, the authors integrated laboratory and field datasets to obtain larger variable ranges. Even though the RM model was developed using the GEP method and on the

combined nature of the data (lab and field), the model has a similar performance in comparison to the model developed in this study, probably due to the application of the PSDB-2014 database. Assuming that the type of technique selected for model development does not affect scour prediction performance as much as the filtering of the dataset, the performances of the RM and AL models are expected to be very close, as shown in Figure 5.

Hassan and Jalal [50] (HJ abbreviated) numerically simulated scour around the bridge pier on a real scale to evaluate the performance of predicting scour depth with the GEP model. They used a numerical dataset consisting of 243 observations collected in a clear-water regime and calibrated with the results from the Melville laboratory model [51]. The authors carried out a sensitivity analysis of influential variables and showed that y has the greatest influence on the predicted d_s, followed by the ratio of velocities v/v_c, the ratio of pier width B/b, the pier Froude number Fr_{pier}, and finally the pier shape factor K_s with the weakest influence. According to Figure 5, the HJ model tends to underpredict scour depth. Unfortunately, only ranges of nondimensional variables are given in their study, so the real span of variable values remains unknown. However, it can be assumed that Hassan and Jalal [50] collected numerical data in terms of lower water depths since flow depth turned out to be the most significant variable. This assumption is supported by the work of Melville and Sutherland [52], as they claim that flow depth does not play an important role in scour when the y/b_{ef} ratio is above 2.6. The final difference is that they have taken into account the ratio of pier width B/b, so it can be assumed that contraction scour has a more important role in their dataset.

Jain and Fischer [38] (JF abbreviated) is one of the many existing regression-based scour equations. The JF model was selected for this comparison process among other regression-based models because previous research [28], which applied the PSDB-2014 database, claims that the JF model is superior to the other conventional models. The JF model is developed in conditions of higher Fr values and sand-bed sediment properties (0.25–2.5 mm with a median of about 1.5 mm). Even though Jain and Fischer collected data in similar conditions that were applied for data filtration in this study, it is still an overly predictive and conservative method. The first reason for overprediction in JF model performance is related to including laboratory-based data, and the second reason is the traditional methodology of developing equations where coefficients were generated to form an envelope for all collected data.

Azamathulla et al.'s [37] (Az abbreviated) model presented the lowest performance amongst all equations selected for comparison in this study. Data in [37] were collected from different studies available in the literature to develop their MNLR model based on 398 field measurements collected over a non-cohesive and uniform riverbed. A possible reason for such low performance could be the larger range of sediments, with d_{50} values ranging from 0.12 to 108 mm and a median of 54 mm. Although information regarding the angle of flow attack is not available in their study, the authors included pier length in their equation. However, it can be assumed that piers were skewed, as otherwise, it would have no effect on scour depth [53]. Azamathulla et al. also provided an explicit equation for the GEP model, but when the PSDB-2014 measurements were input into the equation, the computed results were unreasonable, such as negative values of scour depth. In addition, parts of an equation of the developed MNLR model are vague, such as the exponent of the variable σ_g.

3.3. Variable Sensitivity Analysis

The Pearson correlation coefficient (r) (Figure 6) interprets the strength of the correlation between selected variables in the filtered subset and measured scour depth (d_s). The Pearson correlation coefficient varies in range from −1 to 1. A value of −1 indicates a perfect negative correlation, a value of 1 a perfect positive correlation, and a value of 0 indicates no correlation at all. If variables are classified into four categories: pier geometry (b_{ef}), hydraulic (y, v, τ, and Fr), compounded hydraulic-sediment variables (v_c, Fr_c, and Fr_d), and sediment (τ_c, and Re_p), then the pier geometric variables are the most influential, fol-

lowed by compounded hydraulic-sediment, hydraulic, and finally the sediment variables. Based on Pearson correlation coefficients, b_{ef} proved to be the most influential variable for estimating d_s ($r = 0.625$), which is the same observation that was already stated in previous investigations [28,30]. The second most influential variable is y ($r = 0.492$), followed closely by v_c and v, $r = 0.474$ and $r = 0.436$, respectively. In the compounded hydraulic-sediment category, v_c has the greatest impact on scour depth in comparison to Fr_d and Fr_c, which are defined by $r = 0.427$ and $r = 0.323$, respectively. The remaining four variables (τ, τ_c, Fr, and Re_p) show a very weak correlation with scour depth, with $r < 0.3$.
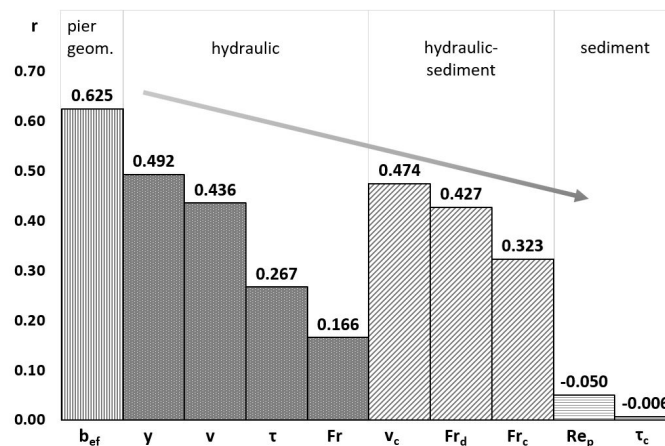

Figure 6. Pearson correlation coefficients (r) for selected influential variables that are classified into four variable categories. The selected variables are taken from the filtered dataset.

4. Discussion

In the present study, large, alluvial, and sand-bed rivers were taken into account with the intention of evaluating the significance of several influential variables in estimating maximum scour depth. Comparing four different categories, i.e., pier geometry (b_{ef}), hydraulic (y, v, τ, and Fr), sediment (τ_c, and Re_p), and compounded hydraulic-sediment variables (v_c, Frc, and Fr_d), it was elucidated that pier geometry variables are the most influential ones. Rating pier geometry variables above hydraulic variables was expected because more than 70% of the data used in this study exceeds the threshold value of $y/b_{ef} = 2.6$, over which the flow depth is no longer significant for the pier scour process [52].

The approach velocity and the flow depth have the next largest impact on scour depth, while local shear stress and Froude number remain the least important variables. The latter confirms the traditional approach assumptions that the equilibrium state of scour depth primarily depends on both the pier width and approach flow velocity, while flow depth is just an indirect effect of the downflow magnitude [38]. The introduction of new hydraulic variables such as τ and Fr is supposed not to significantly contribute to the variance of scour depth due to their unreliable estimation and small ranges of values. A similar conclusion can be drawn regarding the sediment variables. In this study, d_{50} was excluded from analysis, suspecting that it has a negligible effect, so two new sediment variables (τ_c, Re_p) were included in the scour depth equation. Even though it was expected that Re_p would play a more significant role owing to a larger standard deviation than d_{50}, both sediment variables τ_c and Re_p yielded the lowest Pearson correlation coefficients, which shed light on sediment properties in general (d_{50}, τ_c, and Re_p) as being insignificant on scour depth variance. In order to consider the effect of sediment properties, it was recommended to use compound hydraulic-sediment variables such as v_c, Fr_c, and Fr_d that consist of both the acting forces of the flow and sediment properties. In this study, the compounded hydraulic-sediment variables and hydraulic variables had a similar impact on the scour prediction because the average values of their Pearson correlation coefficients were 0.408 and 0.340, respectively. A comparable observation was made by Török et al. [54], who deem that shear Reynolds number, a function of grain size and shear velocity, is a more

adequate variable for evaluating sediment transport than Re_p, d_{50}, or other variables whose evaluation is based only on sediment properties. Furthermore, at the beginning of this paper, there was a doubt concerning which variable should be used to assess the incipient motion of sediment, v_c or τ_c. The doubt originates from the fact that v_c is a function of y and d_{50}, while τ_c is a function of d_{50}. The assumption that the compound hydraulic-sediment variable would be more significant was verified by Pearson correlation coefficients of v_c and τ_c with values of 0.474 and -0.006, respectively. After v_c, which proved to be the most influential variable in the category of compounded hydraulic-sediment variables, Fr_d and Fr_c are next in the sequence. It was expected that Fr_d would show a stronger association with scour depth since it has been previously determined as the most influencing variable in the group of variables that describe sediment properties [55–57].

The introduction of sediment variables into the scour depth equation showed negligible effects. However, this observation does not mean that they can be completely excluded from the scour analysis, as sediment properties have a significant effect on the scour process. Equations developed for different compositions of riverbed sediments behave differently, as evidenced by the comparison process performed in this work. The Az model showed the lowest performance and was based on sediment ranges that tend toward gravel grain sizes. The fact that sediment variables play an important role indicates that it is necessary to take sediment size into account when filtering data sets to achieve certain environmental conditions. However, if sediment properties are selected as one of the scour variables in the equation, it is better to include them through compounded hydraulic-sediment variables than to take them directly because of their small contribution to the variance of the data set.

Recently, regression-based AI models have become widely used in scour studies because of their simple structure and their ability to find functional dependencies among a large number of scour-related variables. However, AI-based models tend to overfit the training dataset, and they are only appropriate for the range of variable values taken for the training. To avoid the aforesaid issues, it is important to appropriately choose specific environmental variables (flow regime, complex pier geometry, pier alignment, sediment uniformity and grain sizes, bridge in the bend, etc.) and their value ranges in order to filter the dataset adequately for AI-model training. If the data are too dispersed, the development of a unique and universal scour equation remains an option. For many years, researchers have been struggling to find a universal equation, which means finding an envelope for a dispersed cloud of data that will predict scour oversafely, i.e., the so-called conservative method.

To avoid overprediction, a best-fitting curve that passes through a cloud of datasets while minimizing the sum of the squared distances should be created. However, it is unlikely to develop such a curve with a machine learning algorithm trained to match the data as closely as possible to a dispersed and scattered dataset. For example, Rathod and Manekar [28] developed two GEP models: one based on a combination of field and laboratory data and one based on laboratory data only. The GEP model, whose development was based only on laboratory data, showed much better performance owing to smaller variable ranges and, consequently, a less scattered dataset. Furthermore, comparisons performed in this work (Figure 5) showed that the AL and RM models are superior to all other scour models. Although the models utilize different techniques such as MNLR and GEP, the deviation between them is quite small, and they have a high similarity in scour prediction. This possibly stems from the same dataset used (PSDB-2014), indicating that any prediction model's performance highly depends on the data and less on the method used for its development. The same can be said for the selection of influencing variables. In the AL and RM models, only 4 and 5 variables, respectively, were introduced, while in this study, 10 variables were considered for the development of the equation without improving the performance of the model. Without the detection of environmental conditions and the application of a specific filter, the data set will be too dispersed or too gathered, and neither the complex curve of the AI technique nor the trivialized envelope curve of the regression model will provide a more accurate prediction.

However, it must be emphasized that the dimensional MNLR model Equation (17) provided a low R^2 value ($R^2 = 0.5$) due to the presence of scatter in the dataset. Previous equations that used the same PSDB-2014 dataset (AL and RM) performed better, as indicated by higher R^2 or lower RMSE values. The reason could be that they used almost all field measurements without excluding multiple measurements for the same bridge. Such a collection of data with similar measurements could lead to overfitting. However, in this study, the independence of the measurements was maintained in the final set of filtration criteria by retaining only one measurement for each bridge, which eventually led to the exclusion of 46% from the original dataset. Another reason for the scatter is that the filtered dataset contains only field data where the scour regime is unknown; it is difficult to determine whether some measurements have reached an equilibrium or maximum state.

5. Conclusions

Most of the existing scour prediction equations available in the literature are derived from laboratory data since flume experiments provide controlled hydraulic conditions and straightforward data collection. On the other hand, the limited range of flow conditions in the flume limits the wider application of derived scour equations, i.e., their application to the prototype conditions. The USGS dataset PSDB-2014 provides a wide range of on-site measured flow, sediment, and geometry variables related to scour; thus, this dataset was the most representative source of scour data for large sand-bed rivers. The measured variables from the dataset were expanded to include eight additional calculated variables commonly used in scour prediction equations and consequently filtered to retain measurements taken only for large sand-bed rivers. After filtering the original dataset, 98% of the remaining measurements were within the clear-water scour conditions, even though the clear-water criterion was not applied in the filter, which validates the reliability of the applied data filter.

Several influential variables were removed by using the PCA, as well as two commonly used variables, θ and d_{50}, as they showed low impact on the scour depth. Since the measurement of θ is arbitrary and changes with flow severity, it should be used only when it can be reliably estimated, which is not often the case. Since the pier length was also eliminated by the PCA and taking into consideration the difficult approximation of complex pier geometry with a single variable, this study selected b_{ef} as the only pier geometry variable, combining the information of pier geometry and alignment with the flow, if available. PCA has eliminated the characteristic sediment size as well, which can be explained by focusing on sand-bed rivers where sand is uniform and therefore has a significantly smaller range than other variables. To take into account riverbed composition, compounded hydraulic-sediment variables (v_c, Fr_c, and Fr_d) were retained, and consequently, 10 influential variables were classified into four categories: pier geometry (b_{ef}), hydraulic (y, v, τ, and Fr), sediment (τ_c, and Re_p), hydraulic-sediment variables (v_c, Fr_c, and Fr_d), and compounded. Afterward, the variables from these categories were used to determine the scour model variables.

The proposed dimensional MNLR model designed for scour estimation on the selected data subset has a firm similarity with the two other models (AL and RM) developed using the same PSDB-2014 dataset but with different filters or methods. The comparison performed in this work indicates that the filtering method has a greater influence on the model's performance than the model type (MNLR or GEP). However, the MNLR model developed in this study, Equation (17), yielded a low R^2 value of 0.5 due to the scatter of the dataset, which could be a consequence of excluding multiple measurements for the same bridge locations. Since, at the time of the field measurement, it is not known whether the equilibrium or maximum state has been reached, dispersion in the results is expected.

Furthermore, the selection of influential variables that can be reliably measured or estimated is crucial when creating a database for scour estimation. Based on the Pearson correlation coefficient, bef proved to be the most influential variable for estimating ds in this study, and the second most influential variable is y, followed closely by v_c, v, Fr_d, and Fr_c. The remaining four variables (τ, τ_c, Fr, and Re_p) exhibit a very weak correlation with

scour depth, probably resulting from errors in the measurement of variables used for their calculation. The fact that compounded hydraulic-sediment variables were highly influential for our dimensional model indicates that it is necessary to use sediment size when filtering data to reduce the uncertainty associated with the acquisition of a representative bed sample.

Author Contributions: Conceptualization, A.H. and G.G.; methodology, A.H., M.Ž. and N.A.; validation, A.H. and G.G.; formal analysis, A.H. and N.A.; investigation, G.G.; data curation, A.H.; writing—original draft preparation, A.H. and G.G.; writing—review and editing, G.G., N.A. and M.Ž.; visualization, A.H.; supervision, G.G.; project administration, G.G. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Glossary

Symbol	Unit	Description
d_s	[m]	scour depth
v	[m/s]	local approach flow velocity (upstream of the pier)
v_c	[m/s]	critical velocity
y	[m]	approach water depth (upstream of the pier)
b	[m]	nominal pier width
l	[m]	pier length
b_{ef}	[m]	effective pier width normal to the flow
θ	[°]	angle of attack
B	[m]	the channel width
S	[1]	stream slope
τ	[Pa]	local bed shear stress
τ_c	[Pa]	critical bed shear stress
Fr	[1]	Froude number
Fr_c	[1]	critical Froude number for incipient motion
Fr_d	[1]	densimetric Froude number
Re_p	[1]	particle Reynolds number
RI	[years]	recurrence interval for measured flow rate
g	[m/s^2]	gravitational acceleration
d_{50}	[mm]	sediment median grain size
d_{95}	[mm]	the size at which 95% of the sediment particles are smaller
σ_g	[1]	geometrical standard deviation of sediment (measure of non-uniformity)
ρ_{rel}	[1]	submerged relative mass density of sediment particles ($\rho_{rel} = [(\rho_s - \rho)/\rho] - 1 = 1.65$)
ρ_s	[kg/m^3]	mass density of sediment particles (equal to 2650 kg/m^3)
ρ_w	[kg/m^3]	mass density of water (equal to 1000 kg/m^3)
γ_s	[N/m^3]	specific gravity of sediment (equal to 25,996.5 N/m^3)
γ_w	[N/m^3]	specific gravity of water (equal to 9810 N/m^3)
K_1	[1]	the live-bed vs. clear-water correction factor
K_2	[1]	the pier shape correction factor
ν	[m^2/s]	kinematic viscosity of fluid ($\nu = 1.6 \times 10^{-6}$ m^2/s)
Φ	[°]	angle of repose for sediments
r	[1]	Pearson correlation coefficient

References

1. Grizzetti, B.; Pistocchi, A.; Liquete, C.; Udias, A.; Bouraoui, F.; Van de Bund, W. Human pressures and ecological status of European rivers. *Sci. Rep.* **2017**, *7*, 205. [CrossRef]
2. Sholtes, J.S.; Ubing, C.; Randle, T.J.; Fripp, J.; Cenderelli, D.; Baird, D.C. *Managing Infrastructure in the Stream Environment*; Advisory Committee on Water Information Subcommittee on Sedimentation: Austin, TX, USA, 2017; p. 65.
3. Lee, M.; Yoo, M.; Jung, H.-S.; Kim, K.H.; Lee, I.-W. Study on Dynamic Behavior of Bridge Pier by Impact Load Test Considering Scour. *Appl. Sci.* **2020**, *10*, 6741. [CrossRef]

4. Kallias, A.N.; Imam, B. Probabilistic assessment of local scour in bridge piers under changing environmental conditions. *Struct. Infrastruct. Eng.* **2016**, *12*, 1228–1241. [CrossRef]

5. Imhof, D. Risk Assessment of Existing Bridge Structures. Ph.D. Thesis, University of Cambridge, Cambridge, UK, 2004.

6. Schaap, H.S.; Caner, A. Bridge collapses in Turkey: Causes and remedies. *Struct. Infrastruct. Eng.* **2022**, *18*, 694–709. [CrossRef]

7. Yao, C.; Briaud, J.-L.; Gardoni, P. Risk Analysis on Bridge Scour Failure. In Proceedings of the International Foundations Congress and Equipment Expo, San Antonio, TX, USA, 17–21 March 2015; pp. 1936–1945.

8. Nasr, A.; Björnsson, I.; Honfi, D.; Larsson Ivanov, O.; Johansson, J.; Kjellström, E. A review of the potential impacts of climate change on the safety and performance of bridges. *Sustain. Resilient Infrastruct.* **2021**, *6*, 192–212. [CrossRef]

9. Kundzewicz, Z.W.; Pińskwar, I. Are Pluvial and Fluvial Floods on the Rise? *Water* **2022**, *14*, 2612. [CrossRef]

10. Nemry, F.; Demirel, H. *Impacts of Climate Change on Transport: A Focus on Road and Rail Transport Infrastructures*; European Commission Joint Research Centre: Luxembourg, 2012.

11. Badroddin, M.; Chen, Z. Lifetime Resilience Measurement of River-Crossing Bridges with Scour Countermeasures under Multiple Hazards. *J. Eng. Mech.* **2021**, *147*, 04021058. [CrossRef]

12. Tubaldi, E.; White, C.J.; Patelli, E.; Mitoulis, S.A.; de Almeida, G.; Brown, J.; Cranston, M.; Hardman, M.; Koursari, E.; Lamb, R.; et al. Invited perspectives: Challenges and future directions in improving bridge flood resilience. *Nat. Hazards Earth Syst. Sci.* **2022**, *22*, 795–812. [CrossRef]

13. Imam, B.M.; Chryssanthopoulos, M.K. Causes and Consequences of Metallic Bridge Failures. *Struct. Eng. Int.* **2012**, *22*, 93–98. [CrossRef]

14. Borghei, S.M.; Kabiri-Samani, A.; Banihashem, S.A. Influence of unsteady flow hydrograph shape on local scouring around bridge pier. *Proc. Inst. Civ. Eng. -Water Manag.* **2012**, *165*, 473–480. [CrossRef]

15. Hung, C.-C.; Yau, W.-G. Behavior of scoured bridge piers subjected to flood-induced loads. *Eng. Struct.* **2014**, *80*, 241–250. [CrossRef]

16. Lu, J.-Y.; Hong, J.-H.; Su, C.-C.; Wang, C.-Y.; Lai, J.-S. Field Measurements and Simulation of Bridge Scour Depth Variations during Floods. *J. Hydraul. Eng.* **2008**, *134*, 810–821. [CrossRef]

17. Harasti, A.; Gilja, G.; Potočki, K.; Lacko, M. Scour at Bridge Piers Protected by the Riprap Sloping Structure: A Review. *Water* **2021**, *13*, 3606. [CrossRef]

18. Benedict, S.T.; Caldwell, A.W. Upper Bound of Pier Scour in Laboratory and Field Data. *Transp. Res. Rec.* **2016**, *2588*, 145–153. [CrossRef]

19. Annad, M.; Lefkir, A. New Formula for Calculating Local Scour around Bridge Piers. *Adv. Eng. Forum* **2022**, *45*, 57–64. [CrossRef]

20. Gaudio, R.; Grimaldi, C.; Tafarojnoruz, A.; Calomino, F. Comparison of formulae for the prediction of scour depth at piers. In Proceedings of the First European IAHR Congress, Edinburgh, UK, 4–6 May 2010; pp. 6–12.

21. Zhang, G.; Hsu, S.A.; Guo, T.; Zhao, X.; Augustine, A.D.; Zhang, L. *Evaluation of Design Methods to Determine Scour Depths for Bridge Structures*; FHWA/LA.11/491; Louisiana State University: Baton Rouge, LA, USA; Federal Highway Administration: Washington, DC, USA, 2013.

22. Dong, H.; Sun, Z.; Li, Z.; Chong, L.; Zhou, H. Artificial Intelligence for Predicting Local Scour Depth around Piers Based on Dimensional Analysis. *J. Coast. Res.* **2020**, *111*, 21–25. [CrossRef]

23. Pandey, M.; Zakwan, M.; Sharma, P.K.; Ahmad, Z. Multiple linear regression and genetic algorithm approaches to predict temporal scour depth near circular pier in non-cohesive sediment. *ISH J. Hydraul. Eng.* **2018**, *26*, 96–103. [CrossRef]

24. Rady, R.M.A.E.-H. Prediction of local scour around bridge piers: Artificial-intelligence-based modeling versus conventional regression methods. *Appl. Water Sci.* **2020**, *10*, 57. [CrossRef]

25. Muzzammil, M.; Alama, J.; Danish, M. Scour Prediction at Bridge Piers in Cohesive Bed Using Gene Expression Programming. *Aquat. Procedia* **2015**, *4*, 789–796. [CrossRef]

26. Benedict, S.T.; Caldwell, A.W. *A Pier-Scour Database: 2427 Field and Laboratory Measurements of Pier Scour*; Data Series 845; U.S. Geological Survey: Reston, VA, USA, 2014; p. 32.

27. Benedict, S.T.; Knight, T.P. Use of Laboratory and Field Data to Evaluate the Pier Scour Equation from Hydraulic Engineering Circular 18. *Transp. Res. Rec.* **2017**, *2638*, 113–121. [CrossRef]

28. Rathod, P.; Manekar, V.L. Gene expression programming to predict local scour using laboratory and field data. *ISH J. Hydraul. Eng.* **2022**, *28*, 143–151. [CrossRef]

29. Pandey, M.; Oliveto, G.; Pu, J.H.; Sharma, P.K.; Ojha, C.S.P. Pier Scour Prediction in Non-Uniform Gravel Beds. *Water* **2020**, *12*, 1696. [CrossRef]

30. Ali, A.S.A.; Günal, M. Artificial Neural Network for Estimation of Local Scour Depth Around Bridge Piers. *Arch. Hydro-Eng. Environ. Mech.* **2021**, *68*, 87–101. [CrossRef]

31. Shahriar, A.R.; Ortiz, A.C.; Montoya, B.M.; Gabr, M.A. Bridge Pier Scour: An overview of factors affecting the phenomenon and comparative evaluation of selected models. *Transp. Geotech.* **2021**, *28*, 100549. [CrossRef]

32. Qi, M.; Li, J.; Chen, Q. Comparison of existing equations for local scour at bridge piers: Parameter influence and validation. *Nat. Hazards* **2016**, *82*, 2089–2105. [CrossRef]

33. Guo, J.; Suaznabar, O.; Shan, H.; Shen, J. *Pier Scour in Clear-Water Conditions with Non-Uniform Bed Materials*; FHWA-HRT-12-022; Federal Highway Administration: Washington, DC, USA, 2012; p. 62.

34. Shields, A. *Application of Similarity Principles and Turbulence Research to Bed-Load Movement*; Hydrodynamics Laboratory: Washington, DC, USA, 1936; p. 47.

35. Shahmohammadi, R.; Afzalimehr, H.; Sui, J. Assessment of Critical Shear Stress and Threshold Velocity in Shallow Flow with Sand Particles. *Water* **2021**, *13*, 994. [CrossRef]

36. Julien, P.Y. *Erosion and Sedimentation*; Cambridge University Press: Cambridge, UK, 1995.

37. Azamathulla, H.M.; Ghani Aminuddin, A.; Zakaria Nor, A.; Guven, A. Genetic Programming to Predict Bridge Pier Scour. *J. Hydraul. Eng.* **2010**, *136*, 165–169. [CrossRef]

38. Jain, S.C.; Fischer, E.E. *Scour around Circular Bridge Piers at High Froude Numbers*; FHWA-RD-79-104; Federal Highway Administration: Washington, DC, USA, 1979; p. 70.

39. Dey, S.; Raikar, R.V. Clear-Water Scour at Piers in Sand Beds with an Armor Layer of Gravels. *J. Hydraul. Eng.* **2007**, *133*, 703–711. [CrossRef]

40. Benedict, S.T.; Knight, T.P. Benefits of Compiling and Analyzing Hydraulic-Design Data for Bridges. *Transp. Res. Rec.* **2021**, *2675*, 1073–1081. [CrossRef]

41. Garde, R.C.J.; Kothyari, U.C. Scour around bridge piers. *PINSA 64* **1998**, *4*, 569–580.

42. Subedi, A.S.; Sharma, S.; Islam, A.; Lamichhane, N. Quantification of the Effect of Bridge Pier Encasement on Headwater Elevation Using HEC-RAS. *Hydrology* **2019**, *6*, 25. [CrossRef]

43. Harasti, A.; Gilja, G.; Adžaga, N.; Škreb, K.A. Principal Component Analysis in development of empirical scour formulae. In Proceedings of the 7th IAHR Europe Congress, Athens, Greece, 7–9 September 2022; pp. 271–272.

44. Zwick, W.R.; Velicer, W.F. Comparison of five rules for determining the number of components to retain. *Psychol. Bull.* **1986**, *99*, 432–442. [CrossRef]

45. Breusers, H.N.C.; Raudkivi, A.J. *Scouring*, 1st ed.; Taylor and Francis Group: London, UK, 1991; p. 152.

46. Laursen, E.M. Scour at Bridge Crossings. *J. Hydraul. Div.* **1960**, *86*, 39–54. [CrossRef]

47. Kiraga, M.; Popek, Z. Bed Shear Stress Influence on Local Scour Geometry Properties in Various Flume Development Conditions. *Water* **2019**, *11*, 2346. [CrossRef]

48. Parker, G. Transport of Gravel and Sediment Mixtures. In *Sedimentation Engineering*; Garcia, M., Ed.; American Society of Civil Engineers: Reston, VA, USA, 2008; pp. 165–251.

49. Vonkeman, J.K.; Basson, G.R. Evaluation of empirical equations to predict bridge pier scour in a non-cohesive bed under clear-water conditions. *J. S. Afr. Inst. Civ. Eng.* **2019**, *61*, 2–20. [CrossRef]

50. Hassan, W.H.; Jalal, H.K. Prediction of the depth of local scouring at a bridge pier using a gene expression programming method. *SN Appl. Sci.* **2021**, *3*, 159. [CrossRef]

51. Melville, B.W. *Local Scour at Bridge Sites*; University of Auckland: Auckland, New Zealand, 1975; p. 227.

52. Melville, B.W.; Sutherland, A.J. Design Method for Local Scour at Bridge Piers. *J. Hydraul. Eng.* **1988**, *114*, 1210–1226. [CrossRef]

53. Richardson, E.V.; Davis, S.R. *Evaluating Scour at Bridges*, 4th ed.; FHWA NHI 01-001; Hydraulic Engineering Circular No. 18; Federal Highway Administration: Washington, DC, USA, 2001; p. 378.

54. Török, G.T.; Józsa, J.; Baranya, S. A Shear Reynolds Number-Based Classification Method of the Nonuniform Bed Load Transport. *Water* **2019**, *11*, 73. [CrossRef]

55. Oliveto, G.; Hager, W.H. Temporal Evolution of Clear-Water Pier and Abutment Scour. *J. Hydraul. Eng.* **2002**, *128*, 811–820. [CrossRef]

56. Oliveto, G.; Hager, W.H. Further Results to Time-Dependent Local Scour at Bridge Elements. *J. Hydraul. Eng.* **2005**, *131*, 97–105. [CrossRef]

57. Tan, S.M.; Lim, S.-Y.; Wei, M.; Cheng, N.-S. Application of Particle Densimetric Froude Number for Evaluating the Maximum Culvert Scour Depth. *J. Irrig. Drain. Eng.* **2020**, *146*, 04020020. [CrossRef]