

Integrating machine learning and empirical approaches for scour depth estimation at sluice gates: evaluating tree-based models, hyperparameter tuning, and proposing new formulas

Xuan-Hien Le, Le Thi Thu Hien*

Faculty of Water Resources Engineering, Thuyloi University, 175 Tay Son, Dong Da, Hanoi 10000, Vietnam.

*Corresponding author. E-mail: lethuhien.thuyluc@tlu.edu.vn

Abstract: Accurate prediction of maximum scour depth (MSD) at sluice gates is critical for guaranteeing the stability and safety of hydraulic systems. Traditional empirical formulas often fail to capture the non-linear interactions between flow dynamics, sediment characteristics, and structural configurations. This study addresses these limitations by leveraging advanced machine learning (ML) techniques, specifically tree-based models, to enhance predictive accuracy. The performance of three tree-based models—Extra Trees (ERT), CatBoost (CAT), and Histogram-Based Gradient Boosting (HGB)—was examined using two hyperparameter tuning methods: Tree-Structured Parzen Estimator (TPE) and Particle Swarm Optimization (PSO). The models underwent 100 simulations to quantify uncertainty and variability in performance metrics. The results indicate that CAT_PSO (optimized with PSO) exhibits superior predictive performance compared to empirical formulas and other ML techniques. CAT_PSO achieved the highest mean CORR (correlation coefficient) of 0.9644 and mean NSE (Nash-Sutcliffe Efficiency) of 0.9272. HGB models demonstrated slightly lower performance compared to CAT and ERT, with higher variability in predictions. Further analysis explored the influence of individual input factors on model performance. The inclusion of more variables, such as tailwater depth and sediment size, generally enhanced model performance. The study also developed new empirical equations for MSD estimation by considering both multiplicative and additive models, progressively incorporating additional input features. These new formulas show improved predictive accuracy over empirical methods, though they still fall short of the performance achieved by the ML models.

Keywords: Extra Trees (ERT); Histogram-Based Gradient Boosting (HGB); Hyperparameter Tuning; Particle Swarm Optimization (PSO); Scour depth estimation; Tree-structured Parzen Estimator (TPE).

1 INTRODUCTION

Scour at sluice outlets is a critical phenomenon in hydraulic engineering, posing significant risks to the stability and safety of hydraulic structures. The erosion caused by the high-velocity water flow downstream of sluice gates can undermine foundations, leading to potential structural failures (Melville, 2014). Accurate estimation of maximum scour depth (MSD) is therefore essential for the design, construction, and maintenance of these structures. Effective prediction methods enable engineers to implement adequate protective measures, ensuring the longevity and integrity of hydraulic infrastructure (Sharafati et al., 2021).

Predicting scour depth is inherently complex due to the multifaceted interactions between flow dynamics, sediment characteristics, and structural configurations (Palermo et al., 2021). The factors influencing scour depth include flow velocity, tailwater depth, sediment size, and sluice gate geometry (Lantz et al., 2022; Sarathi et al., 2008). These variables interact in non-linear ways, making the scour process highly sensitive to changes in conditions (Eghbalzadeh et al., 2018). Traditional empirical formulas, although practical, often oversimplify these interactions and could fail to capture the full complexity of the scouring process (Le et al., 2022).

Empirical formulas are typically derived from specific experimental conditions, limiting their applicability across different scenarios (Laucelli and Giustolisi, 2011). Empirical formulas for MSD estimation have been extensively developed

from foundational studies like Aderibigbe and Rajaratnam (1996), who contributed significantly to studying scour due to impinging jets. These formulas, derived from experimental data, offer simplified relationships between scour depth and influencing parameters such as flow velocity, sediment size, and structural geometry. They may neglect critical factors such as tailwater depth or sediment heterogeneity, leading to inaccurate predictions. Additionally, the variability in experimental setups and conditions across different studies results in a wide range of empirical formulas, each with its strengths and limitations (Aamir and Ahmad, 2016). This variability complicates the selection of an appropriate formula for a given situation, further challenging accurate scour depth prediction.

Despite their widespread use, empirical formulas often face limitations due to their derivation from specific experimental setups. They may not adequately capture the complex, non-linear interactions between various factors influencing scour depth. For instance, formulas proposed by Aamir and Ahmad (2017) and Chatterjee et al. (1994) each reflect the conditions under which they were developed, which may not be universally applicable. The limitations of empirical formulas are evident in their dependency on the scope of experimental data. Many studies, such as those by Dey and Sarkar (2006) and Sarkar and Dey (2005), primarily consider non-cohesive soil types and often exclude auxiliary structures like headwalls and aprons, limiting their applicability. The state of a submerged hydraulic jump remained through every experiment where the tailwater depth surpassed that of a free hydraulic jump (Aamir et al., 2022).

Moreover, the inherent complexity of scour phenomena, influenced by a multitude of interdependent variables, often renders these formulas less reliable when applied outside their original experimental contexts.

The advent of computational advancements has facilitated the application of machine learning (ML) algorithms in predicting MSD, presenting a promising alternative to traditional empirical methods (Guan et al., 2023; Hien et al., 2024). ML algorithms are particularly well-suited for tasks that involve multiple interdependent variables due to their ability to identify complex patterns and relationships within large datasets (Le et al., 2024; Seyedian and Kisi, 2024). This capability has been harnessed in numerous studies to enhance the accuracy of MSD predictions. Studies by Azmathullah et al. (2006) and Kambekar and Deo (2003) were among the early adopters of neural networks and genetic programming to estimate MSD. These methods demonstrated superior performance compared to empirical equations, owing to their ability to learn from extensive datasets. Subsequent research by Lee et al. (2007) and Bateni et al. (2007) further validated the efficacy of ML models in this domain. Firat and Gungor (2009) and Guven and Azamathulla (2012) expanded on these findings, using advanced neural networks and gene expression programming (GEP) to estimate MSD with high accuracy.

More recent studies have explored a range of ML approaches, including support vector machines (SVM), group method of data handling (GMDH), and adaptive neuro-fuzzy inference systems (ANFIS). For instance, Karbasi and Azamathulla (2017) compared several soft-computing techniques, concluding that ANN models often outperform other approaches in predicting MSD. Similarly, studies by Najafzadeh et al. (2016) and Najafzadeh and Barani (2011) demonstrated the effectiveness of GMDH and NF-GMDH models in providing accurate scour predictions. The comparative analysis of various ML models reveals that while each technique has its strengths, the choice of model significantly impacts predictive accuracy. For example, Qaderi et al. (2021) reported that ANFIS outperformed other models, including ANN, GMDH, and GEP, in estimating scour depth around bridge piers. Parsaie et al. (2019) highlighted the slight edge of SVM over ANN and ANFIS, emphasizing the importance of selecting appropriate ML models based on the specific characteristics of the data. Le and Le (2024) compared several empirical formulas with ML algorithms like SVM, Ridge Regression, XGBoost, and CatBoost (CAT) for estimating MSD behind the sluice outlets. They concluded that ML algorithms, especially XGBoost and CAT, are superior.

Despite the advancements offered by these various ML techniques, tree-based models, including Extra Trees (ERT), Histogram-Based Gradient Boosting (HGB), and CAT, have not been widely used. These models are known for their robustness, interpretability, and efficiency in handling complex, high-dimensional datasets. ERT, for instance, excels in reducing variance through ensemble learning, while CAT leverages ordered boosting to handle categorical features efficiently. HGB combines the strengths of gradient boosting with histogram-based techniques to improve computational efficiency and predictive performance. Previous applications of tree-based models have shown promising results, suggesting their potential in MSD estimation. However, to fully harness their potential, it is crucial to optimize these models through effective hyperparameter tuning.

Hyperparameter tuning is an important part of optimizing ML algorithms, directly influencing their performance and generalizability. The process involves selecting the most effective set of hyperparameters that govern the learning process

of the model. Proper tuning ensures that the model can effectively learn from the data, avoiding issues like overfitting or underfitting. Common hyperparameter tuning methods include grid search, random search, and more sophisticated techniques like Tree-Structured Parzen Estimator (TPE) and Particle Swarm Optimization (PSO). TPE and PSO were chosen for benchmarking in this study because of their effectiveness in optimizing complex ML models. A probabilistic model of the objective function is constructed by TPE, which is then employed to identify the most promising hyperparameters. In contrast, PSO, which is motivated by the social behavior of birds, iteratively improves the hyperparameters by considering the best solutions found so far. These methods offer a balance between exploration and exploitation, enabling the efficient identification of optimal hyperparameter configurations. By comparing these advanced tuning techniques, the study aims to determine their impact on the predictive performance of tree-based models in MSD estimation.

This study aims to advance the predictive modeling of MSD at sluice gates by addressing the limitations of traditional empirical formulas and leveraging the strengths of ML techniques. The specific objectives of this research are: 1) to evaluate the performance of selected tree-based ML models (ERT, CAT, and HGB) in predicting MSD at the outlet of sluices; 2) to compare the efficacy of hyperparameter tuning methods, specifically TPE and PSO, in optimizing these model's performance; 3) to assess the influence of individual input variables on model accuracy, examining how different combinations of predictors impact performance; 4) to quantify the uncertainty and variability in model predictions through extensive simulations, ensuring a robust evaluation of model reliability; 5) to develop new empirical equations for MSD estimation based on comprehensive data analysis and incorporating multiple dimensionless variables; and 6) to systematically compare these newly developed empirical equations with traditional formulas and ML models to establish their relative accuracy and reliability.

By fulfilling these objectives, the study seeks to provide hydraulic engineers with more accurate and reliable tools for MSD estimation, thereby contributing to safer and more efficient hydraulic engineering practices. The subsequent sections will detail the methodologies employed, followed by a thorough analysis of the results obtained from both the ML techniques and the empirical equations.

2 BACKGROUND AND EMPIRICAL APPROACHES

The scour phenomena downstream of sluice gates involve a complicated interaction of geometric, hydraulic, and sedimentary factors. The complexity of the phenomenon is rooted in the interplay between these factors and their influence on the development and dimensions of the scour hole behind the sluice gate (Mutlu Sumer, 2007). Fig. 1 is a schematic illustrating scour depth geometry.

Scouring downstream of sluices is influenced by a multitude of factors, which can be broadly categorized into hydraulic conditions, geometric configurations, and sediment properties (Farooq and Ghumman, 2019). The primary metric of interest is the equilibrium or MSD (d_s), which is crucial for understanding and mitigating the erosion effects. Variables like the apron length (L), tailwater depth (d_t), sluice gate opening height (a), and input velocity (V) play significant roles in the scouring process (Dey and Westrich, 2003). Additionally, sediment characteristics, including the mean grain size (D_{50}), soil density (ρ_s), and geometric standard deviation (σ_g), as well as fluid properties like

water density (ρ_w) and kinematic viscosity (ν), are critical in determining the scour depth (Najafzadeh and Lim, 2015). The relationship between these factors and the MSD may be stated as:

$$d_s = f(V, a, L, d_t, D_{50}, \sigma_g, \rho_s, \rho_w, g, \nu) \quad (1)$$

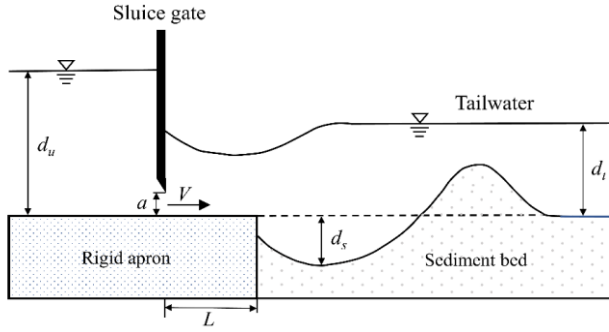


Fig. 1. Diagram of the scour depth geometry behind the sluice gate (Le and Le, 2024)-

In scour modeling, the use of dimensionless parameters has been demonstrated to increase model predictive capacity (Azamathulla and Ghani, 2011). The dimensionless approach ensures that the predictive models are not biased by the scale of the measurements and can be applied universally across different conditions and setups (Guvén and Gunal, 2008). In addition, the dimensionless parameters can enable the generalization of experimental results to various scales and conditions, which reduces the complexity of the modeling process (Manes and Brocchini, 2015). Moreover, they facilitate the comparison of results across different studies and experimental setups, improving the robustness and applicability of prediction models. In order for dimensionless analysis, this approach employs the Buckingham π theorem, a fundamental principle in dimensional analysis, to reduce complex physical phenomena into simpler, dimensionless relationships. This theorem minimizes the number of variables by combining them into dimensionless groups, known as π terms, which capture the essential physics of the problem. For this problem, the repeating variables were identified as V , a , and ρ_w , while the non-repeating variables included D_{50} , d_t , L , ρ_s , d_s , ν , and σ_g . The following dimensionless π terms were constructed from these variables:

$$\pi_1 = F = \frac{V}{\sqrt{g \cdot a}} \text{ is Froude number;}$$

$$\pi_2 = F_d = \frac{V}{\sqrt{\left(\frac{\rho_s}{\rho_w} - 1\right) \cdot g \cdot D_{50}}} \text{ the densimetric Froude number;}$$

$$\pi_3 = \frac{D_{50}}{a}; \pi_4 = \frac{d_t}{a}; \pi_5 = \frac{L}{a}; \pi_6 = \sigma_g; \pi_7 = \frac{V \cdot a}{\nu} \text{ is the}$$

$$\text{Reynolds number; } \pi_8 = \frac{d_s}{a}$$

In addition, empirical studies have shown that the Reynolds number (π_7) has an insignificant effect on MSD under turbulent conditions despite its importance in determining the flow regime. Therefore, the relative MSD (d_s/a) may be calculated using six π terms in the functions shown below:

$$\frac{d_s}{a} = f\left(F_d, \frac{D_{50}}{a}, \frac{d_t}{a}, \frac{L}{a}, F, \sigma_g\right) \quad (2)$$

From Eq. 2, the empirical formulas developed over the years provide various ways to estimate d_s/a . Table 1 summarizes these formulas, illustrating their reliance on different parameters.

These formulas, derived from experimental observations, often simplify the complex interactions governing the scouring process. By reducing the problem to dimensionless terms, researchers can derive more generalizable and scalable relationships. For instance, some formulas consider the d_s as a function of a single dimensionless parameter (see Eqs. 4-5), such as the F or F_d . Such simplifications are advantageous in scenarios where comprehensive data is not available, but they may fail to capture the full range of influencing factors (Xie and Lim, 2015). Several studies below demonstrate the influence of factors like tailwater depth (d_t), apron length (L), or sediment size (D_{50}).

The influence of flow dynamics on MSD has been extensively studied. Ali and Neyshaboury (1991) examined the influence of F_d on the size of the scour hole and found that equilibrium scour depth rises with F_d . This finding was corroborated by Dey and Sarkar (2006), who noted a similar relationship. F_d , reflecting the relative velocity of the flow to the sediment transport capacity, is vital in determining the equilibrium scour depth. The authors also provided further insights by plotting d_s/a as a function of d_t/a for various F_d values and apron lengths. They identified a sagging pattern in the change of d_s/a with d_t/a , indicating a crucial d_t corresponding to a minimum d_s/a . Verma and Goel (2005) observed that the MSD increases with the submergence ratio (d_t/a) or the apron length ratio (L/a). Balachandar et al. (2000) reported that as the d_t increases, the initial magnitude and rate of scour hole development decrease, suggesting a stabilizing effect of higher tailwater depths on the scour process.

Research by Dey and Sarkar (2006) demonstrated that the dimensionless maximum equilibrium scour depth (d_s/a) decreases as the ratio of the mean grain size to the sluice opening (D_{50}/a) increases. This observation indicates that coarser sediments and larger sluice openings result in shallower scour depths. Kells et al. (2001) found similar trends, attributing the decrease in scour depth with increasing sediment grain size to the greater resistance offered by larger grains to shear stress. Their findings highlighted that MSD occurs in the range between medium and coarse sand and increases with tailwater depth.

The roughness of the apron significantly affects scour depth. Hamidifar et al. (2011) found that a rough apron may lower the MSD by up to 59% for the highest F_d value with the smallest roughness and more than 93% for the largest roughness. This reduction is due to the increased turbulence and energy dissipation induced by the rough surface, which limits the scouring action. Ali et al. (2014) explored the impact of corrugated aprons on scour due to submerged jets, identifying optimal corrugation wavelengths that minimize scour. Their study indicated that strategically spaced corrugated aprons effectively disrupt the flow patterns, reducing the erosive potential of the jet.

Additionally, some parameters, like the F_d based on D_{95} or the critical velocity (u_{*c}), can be challenging to measure, limiting the practical applicability of specific formulas. The variability in experimental conditions and datasets used to develop these formulas means their efficacy can fluctuate when applied to new datasets (Aamir and Ahmad, 2019). For example, the empirical formulas may not account for the full range of hydraulic and sedimentary conditions encountered in different field scenarios. Thus, while empirical formulas provide a helpful starting point, advanced modeling approaches, including ML techniques, offer the potential for improved accuracy by incorporating a broader range of variables and leveraging large datasets.

Table 1. Several empirical formulas for estimating MSD.

Researcher	Equation
(Breusers and Raudkivi, 1991)	$\frac{d_s}{a} = 0.008 \left(\frac{v}{u_{*c}} \right)^2$; here u_{*c} is critical bed shear velocity (3)
(Chatterjee et al., 1994)	$\frac{d_s}{a} = 0.775F$ (4)
(Aderibigbe and Rajaratnam, 1998)	$\frac{d_s}{a} = -6.11 + 3.35F_{d(95)}$; here $F_{d(95)}$ is a F_d based on D_{95} (5)
(Lim and Yu, 2002)	$\frac{d_s}{a} = 1.04\sigma_g^{-0.69}F_d^{1.47} \left(\frac{D_{50}}{a} \right)^{0.33} e^{-0.04\beta \left(\frac{L}{a} \right)^{1.4}}$; here $\beta = \sigma_g^{-0.5}F_d^{-0.35} \left(\frac{D_{50}}{a} \right)^{-0.5}$ (6)
(Hopfinger et al., 2004)	$\frac{d_s}{a} = 0.43 \left(\frac{D_{50}}{a} \right)^{0.11} F_d^{1.1} - 0.2$ (7)
(Sarkar and Dey, 2005)	$\frac{d_s}{a} = 0.42F_d^{0.49} \left(\frac{L}{a} \right)^{-0.36} \left(\frac{d_t}{a} \right)^{1.08}$ (8)
(Dey and Sarkar, 2006)	$\frac{d_s}{a} = 2.59F_d^{0.94} \left(\frac{L}{a} \right)^{-0.37} \left(\frac{d_t}{a} \right)^{0.16} \left(\frac{D_{50}}{a} \right)^{0.25}$ (9)
(Aamir et al., 2022)	$\frac{d_s}{a} = 2.35F_d^{1.80} \left(\frac{L}{a} \right)^{-0.76} \left(\frac{d_t}{a} \right)^{0.25} \left(\frac{D_{50}}{a} \right)^{0.49} \left(\frac{k_s}{a} \right)^{-2.79}$; here k_s is the roughness (10)

3 MATERIALS AND METHODS

3.1 Data Collection and Preprocessing

The data utilized in this research were sourced from several works, including Chatterjee et al. (1994) (hereafter Chatterjee, 1994), Sarkar and Dey (2005) (hereafter Sarkar, 2005), Dey and Sarkar (2006) (hereafter Dey, 2006), and Aamir et al. (2022) (hereafter Aamir, 2022). These datasets were chosen due to their comprehensive coverage of the variables influencing scour depth and their relevance to the empirical equations under consideration. The dataset includes six input variables: densimetric Froude number (F_d), mean grain size ratio (D_{50}/a), apron length ratio (L/a), tailwater depth ratio (d_t/a), Froude

number (F), and geometric standard deviation (σ_g). The target variable is the relative scour depth (d_s/a).

Given that the data come from experimental studies, it inherently includes uncertainties and potential outliers. To address this, a preprocessing step was conducted to remove extreme values that could skew the model training and evaluation. Specifically, data points with values of d_s/a higher than 10.0 or lower than 1.0 were excluded, resulting in a final dataset of 357 samples. Table 2 summarizes the results of many test runs, including comprehensive metrics. This study only looks at smooth, rigid aprons; their roughness is insignificant.

Table 2. Overview of collected experiment datasets.

Investigator	Samples	F_d	D_{50}/a	L/a	d_t/a	F	σ_g	d_s/a	Note
Aamir 2022	68	2.70	0.02	20.00	6.67	1.46	1.11	1.00	Min
		12.56	1.34	100.00	30.00	12.06	1.45	9.80	Max
		6.14	0.33	44.12	12.83	3.86	1.19	3.92	Mean
		2.21	0.34	18.25	4.88	2.50	0.10	1.98	STD
Dey 2006	223	3.31	0.02	26.67	6.97	2.37	1.00	1.47	Min
		22.12	0.44	55.00	13.86	4.87	3.92	8.16	Max
		8.37	0.14	34.66	10.60	3.41	1.16	4.15	Mean
		3.37	0.09	6.49	1.47	0.74	0.56	1.25	STD
Sarkar 2005	40	4.49	0.03	33.33	9.07	2.58	1.71	1.69	Min
		13.59	0.20	33.33	10.87	3.15	3.92	4.13	Max
		8.34	0.10	33.33	9.97	2.87	2.79	2.61	Mean
		3.01	0.07	0.00	0.90	0.29	0.74	0.59	STD
Chatterjee 1994	26	3.56	0.02	13.20	5.82	1.34	1.22	1.08	Min
		18.18	0.22	33.00	15.50	5.46	1.43	4.10	Max
		8.15	0.09	21.24	9.58	2.56	1.34	2.05	Mean
		3.95	0.06	7.22	3.23	1.06	0.10	0.78	STD

3.2 Tree-Based Models

Tree-based models are a type of ML techniques that employ decision trees as its core building component. These models are highly effective for both classification and regression tasks since they can capture complicated non-linear relationships and interactions between features. For this study, three advanced

tree-based models were selected: ERT, CAT, and HGB. These models were chosen for their diverse approach to tree construction and ensemble learning, their robustness in handling different types of data, and their proven performance in various predictive modeling tasks.

3.2.1 Extra Trees (ERT)

ERT (or Extremely Randomized Trees) is an ensemble learning technique that aggregates the predictions of multiple randomized decision trees (Geurts et al., 2006). Unlike traditional random forests, which construct each tree using a bootstrap sample of the training data, ERT uses the entire dataset for each tree. Moreover, it introduces additional randomness by selecting cut points for each feature randomly rather than optimizing the split based on a criterion like Gini impurity or entropy. This additional randomness aims to reduce overfitting and improve generalization. Mathematically, the algorithm builds a forest of T trees, where each tree t_i is built by splitting the data at randomly chosen thresholds for each feature. Given a feature vector x , the prediction $\hat{y}$ is obtained by averaging the estimations of all trees:

$$\hat{y} = \frac{1}{T} \sum_{i=1}^T \hat{y}_{t_i}(x) \quad (11)$$

Where $\hat{y}_{t_i}(x)$ represents the estimation from the i -th tree. The algorithm's complexity is primarily driven by the number of trees T and the depth of each tree, which determines the number of splits.

3.2.2 CatBoost (CAT)

CAT, short for Categorical Boosting, is a gradient-boosting technique designed to handle categorical features and mitigate target leakage effectively (Prokhorenkova et al., 2018). It employs a novel ordered boosting technique that prevents overfitting by using permutations of the dataset. CAT also incorporates an efficient implementation of oblivious trees, where all splits at each level of the tree are made using the same feature and threshold. The prediction at the m -th iterations is updated iteratively as follows:

$$\hat{y}^{(m)}(x) = \hat{y}^{(m-1)}(x) + \eta \cdot f_m(x) \quad (12)$$

where η represents the learning rate, and f_m is the decision tree built at iteration m . The objective is to minimize a loss function L , such as the mean squared error for regression:

$$L = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (13)$$

where y_i is the true value and $\hat{y}_i$ is the estimated value. The ordered boosting and handling of categorical features significantly enhance CAT's performance on datasets with categorical variables.

3.2.3 Histogram-Based Gradient Boosting (HGB)

HGB is a variant of gradient boosting that improves computational efficiency by using histograms to approximate the continuous distribution of feature values (Ke et al., 2017). This method discretizes the features into a fixed number of bins and uses these binned values to construct decision trees. This approach significantly speeds up the training process while maintaining high predictive performance, making it suitable for large datasets. The algorithm iteratively builds trees to minimize a loss function L . At each iteration m , the prediction $\hat{y}^{(m)}(x)$ is updated by adding the weighted sum of the estimations from the new tree:

$$\hat{y}^{(m)}(x) = \hat{y}^{(m-1)}(x) + \eta \cdot g_m(x) \quad (14)$$

where $g_m(x)$ denotes the estimation of the m -th tree, and η is the learning rate. The use of the histograms allows for efficient computations of split points, significantly reducing the computational cost associated with the traditional gradient boosting methods. The algorithm's complexity is influenced by the number of bins used and the depth of each tree.

3.3 Hyperparameter tuning approaches

3.3.1 Dataset and validation methodology

The experimental dataset was collected from four previous research studies. This comprehensive collection covers a wide range of experimental conditions, ensuring the dataset's diversity and robustness. The inclusion of various experimental setups and conditions enhances the generalizability of the models developed in this study. To mitigate the effects of any inherent biases and to ensure a fair distribution of the data between training and testing sets, the entire dataset was shuffled randomly 10,000 times. This rigorous shuffling process ensured that the data distribution in both the training and testing sets was representative of the overall dataset, thereby enhancing the reliability of the model evaluation.

The data was then split into training and testing sets using a 75-25% split. Specifically, 75% of the data (268 samples) was employed for training the models and tuning the hyperparameters, while the remaining 25% (89 samples) was reserved for testing the model's performance. This split provided a substantial amount of data for both training and testing purposes, enabling the models to learn effectively while retaining enough data for a robust evaluation of their performance. To further enhance the robustness of the validation process, cross-validation (CV) was implemented using a 5-fold strategy. In this approach, the training data was divided into five folds or subsets. Each subset was utilized once as a validation dataset, whereas the remaining four subsets were employed for training. This procedure was performed five times, ensuring that each subset of the data was used for both training and validation across different folds. CV prevents overfitting and offers a more reliable assessment of model performance by validating the model on multiple sets of data.

In addition, the performance of each ML model was evaluated through 100 simulations to capture the uncertainty and variability inherent in the model's predictions. To enable a thorough assessment of the model performance, the dataset was randomly divided into training (75%) and testing (25%) sets 100 times. This iterative splitting process allowed for different subsets of the data to be employed for training and testing in each simulation, thus providing a robust evaluation framework. By varying the training and testing sets, the models were exposed to different data distributions, capturing a wide range of possible scenarios and ensuring the robustness of the performance metrics. Each model was trained 100 times, with hyperparameters optimized for each training iteration. Following training, the model was evaluated on the corresponding testing set, generating predictions for which three key performance metrics were calculated: CORR, NSE, and Mean Absolute Error (MAE). These metrics were chosen because they are widely accepted and valuable in hydro-dynamic modeling, providing a comprehensive assessment of model accuracy, predictive power, and overall performance. The mathematical expressions for these metrics are as follows:

$$ORR = \frac{\sum (o_i - \bar{o})(p_i - \bar{p})}{\sqrt{\sum (o_i - \bar{o})^2 \cdot \sum (p_i - \bar{p})^2}} \quad (15)$$

$$NSE = 1 - \frac{\sum(O_i - P_i)}{\sum(O_i - \bar{O})^2} \quad (16)$$

$$MAE = \frac{1}{n} \sum |O_i - P_i| \quad (17)$$

Where O_i and P_i represent observed and predicted values, respectively; $\bar{O}, \bar{P}$ are their mean values; n denotes the total number of observations.

MAE measures the average magnitude of errors in the predictions, providing insight into the typical error size. Lower MAE values indicate more accurate estimations, making it a critical metric for evaluating the precision of the model's estimates. CORR measures the linear relationship between predicted and observed values. A strong positive correlation is indicated by values closer to 1, a strong negative correlation by values closer to -1, and no linear relationship by values near 0. This metric is crucial for understanding the degree of alignment between the estimated and observed data trends. NSE evaluates the predictive skill of the model relative to the mean of the observed values. An NSE value of 1 indicates perfect model predictions, values between 0 and 1 signify that the model's predictions are better than the mean of the observed data, and negative values suggest that the mean of the observed data would be a better predictor than the model.

After completing 100 simulations, each metric—MAE, CORR, and NSE—yielded 100 values. The mean and standard deviation (STD) of these values were then calculated to provide a summary of the model's performance. The mean values offer a central tendency measure of model performance across all simulations, while the STD provides insight into the variability and stability of the model's predictions.

Hyperparameter tuning is an important part of developing ML models to achieve superior predictive performance. Given the complexity and non-linearity of scour depth estimation at sluice outlets, TPE and PSO were chosen for their strengths in handling high-dimensional hyperparameter spaces and their ability to converge to optimal solutions efficiently.

3.3.2 Tree-Structured Parzen Estimator (TPE)

TPE is a Bayesian optimization technique that models the distribution of hyperparameters to identify the optimal set (Bergstra et al., 2011). Unlike traditional grid or random search methods, TPE builds two probabilistic models: one for the hyperparameters associated with high performance and another for the rest. It then samples from these models to suggest new hyperparameter configurations (Watanabe, 2023). In this study, TPE was implemented using the Optuna library to optimize the hyperparameters for three tree-based models (Akiba et al., 2019). The objective function for each model was designed to minimize the RMSE using a 5-fold CV.

For the ERT, the hyperparameters tuned were the number of estimators ($n_{estimators}$), the maximum depth of trees (max_depth), and the maximum number of features evaluated for splitting ($max\_features$). The objective function for these hyperparameters was defined to minimize the RMSE across the 5-fold CV. The search space for these hyperparameters was described as follows:

$$n_{estimators} \in [50, 200], max\_depth \in [3, 10], max\_features \in [1, 6] \quad (18)$$

For CAT, the hyperparameters optimized were the number of iterations ($iterations$), learning rate ($learning\_rate$), and maximum depth ($depth$):

$$iterations \in [50, 200], depth \in [3, 10], learning\_rate \in [0.01, 0.3] \quad (19)$$

Similarly, for HGB, the hyperparameters tuned were the maximum number of iterations (max_iter), learning rate ($learning\_rate$), and maximum depth (max_depth):

$$max\_iter \in [50, 200], max\_depth \in [3, 10], learning\_rate \in [0.01, 0.3] \quad (20)$$

The TPE algorithm's ability to focus on promising portions of the hyperparameter space substantially boosts the efficiency of the tuning process. The Optuna study was configured to minimize the RMSE through 50 trials for each of the 100 simulations. The best hyperparameters identified were used to fit the model on the training dataset, and performance was evaluated on the testing dataset.

3.3.3 Particle Swarm Optimization (PSO)

PSO is a population-based optimization approach influenced by the social behavior of birds flocking or fish schooling (Kennedy and Eberhart, 1995). In PSO, each particle represents a potential solution (i.e., a set of hyperparameters) and moves through the hyperparameter space influenced by its own experience and the experience of neighboring particles. The movement of the particles is governed by the following equations:

$$v_i^{t+1} = \omega v_i^t + c_1 r_1 (p_i^{best} - x_i^t) + c_2 r_2 (g^{best} - x_i^t) \quad (21)$$

$$x_i^{t+1} = x_i^t + v_i^{t+1} \quad (22)$$

Where v_i^t and x_i^t represent the velocity and position (set of hyperparameters) of the particle i at iteration t ; p_i^{best} and g^{best} denotes the best position found by particle i and entire swarm; ω is the inertia weight; c_1 and c_2 are acceleration coefficients with random values r_1 and r_2 sampled from a uniform distribution.

In this study, PSO was applied to optimize the same set of hyperparameters for the three tree-based models. The objective function evaluated the RMSE using a 5-fold CV for each candidate solution proposed by the particles. The hyperparameter bounds for PSO were defined similarly to those used in TPE, ensuring a fair comparison between the two methods. Each particle swarm consisted of 20 particles, and the optimization was run for 50 iterations in each of the 100 simulations. The best hyperparameters identified by PSO were then used to train the models on the training data, and their performance was assessed on the testing data.

3.3.4 Development of new empirical equations

The development of new empirical formulas for predicting MSD at sluice gates is a critical task in hydraulic engineering. These equations aim to capture the complex interplay of hydraulic and sedimentary factors affecting scour depth using dimensionless parameters. This study proposes several empirical models based on extensive data collected from previous research and experimental studies. The models were formulated using a combination of key dimensionless parameters and were designed

to enhance the predictive accuracy and robustness of scour depth estimations.

Two types of empirical equations were considered: multiplicative (M_i) and additive (A_i) models. The models were progressively expanded by including additional input features, resulting in a maximum of 10 formulations (five multiplicative and five additive). The multiplicative models (M1–M5) assume that the impact of each predictor on the target variable (d_s/a) can be represented by a power-law relationship, while the additive models (A1–A5) combine linear and non-linear terms to account for the interactions between variables. Each type was formulated with an increasing number of input features, ranging from two to six (Eqs. 23–27). The multiplicative and additive models take general forms (Eqs. 28–29).

$$M1, P1 \quad \frac{d_s}{a} = f\left(F_d, \frac{D_{50}}{a}\right) \quad (23)$$

$$M2, P2 \quad \frac{d_s}{a} = f\left(F_d, \frac{D_{50}}{a}, \frac{d_t}{a}\right) \quad (24)$$

$$M3, P3 \quad \frac{d_s}{a} = f\left(F_d, \frac{D_{50}}{a}, \frac{d_t}{a}, \frac{L}{a}\right) \quad (25)$$

$$M4, P4 \quad \frac{d_s}{a} = f\left(F_d, \frac{D_{50}}{a}, \frac{d_t}{a}, \frac{L}{a}, F\right) \quad (26)$$

$$M5, P5 \quad \frac{d_s}{a} = f\left(F_d, \frac{D_{50}}{a}, \frac{d_t}{a}, \frac{L}{a}, F, \sigma_g\right) \quad (27)$$

$$M_i \text{ Models} \quad \frac{d_s}{a} = a_1 \cdot F_d^{a_2} \cdot \left(\frac{D_{50}}{a}\right)^{a_3} \cdot \left(\frac{d_t}{a}\right)^{a_4} \cdot \left(\frac{L}{a}\right)^{a_5} \cdot F^{a_6} \cdot \sigma_g^{a_7} \quad (28)$$

$$A_i \text{ Models} \quad \frac{d_s}{a} = a_1 + a_2 \cdot F_d^{b_2} + a_3 \cdot \left(\frac{D_{50}}{a}\right)^{b_3} + a_4 \cdot \left(\frac{d_t}{a}\right)^{b_4} + a_5 \cdot \left(\frac{L}{a}\right)^{b_5} + a_6 \cdot F^{b_6} + a_7 \cdot \sigma_g^{b_7} \quad (29)$$

For each type of model, multiple formulations were verified to identify the optimal set of parameters that best fit the data. All formula forms were fitted using the available data, and the parameters were estimated using non-linear least squares optimization through the curve fitting technique. The initial parameter guesses were set to one to ensure a standard starting point for the optimization process. The performance of these equations was measured by three metrics: MAE, CORR, and NSE. For each model, the parameters were optimized using the curve_fit function from the SciPy optimize library (Virtanen et al., 2020), which minimizes the sum of the squared residuals between the observed and predicted data. The initial parameter guesses were set to 1 for simplicity. The optimization was constrained to a maximum of 10,000 iterations to ensure convergence.

4 RESULTS AND DISCUSSION

4.1 Performance of tree-based models

The performance of tree-based models in estimating scour depth was evaluated using 100 simulations to capture the uncertainty and variability inherent in the model's estimations. The models were optimized using two hyperparameter tuning methods: TPE and PSO. Table 3 summarizes the mean and STD

of the CORR, NSE, and MAE values for each model and tuning method combination. Fig. 2 visually represents the mean CORR and its variability for each model tuned with TPE or PSO.

The findings in Table 3 and Fig. 2 indicate that the models generally exhibited high predictive accuracy, as evidenced by the strong CORR and NSE values. Among the models, CAT consistently demonstrated a higher performance. CAT tuned with PSO achieved the highest mean CORR of 0.9644 and the highest mean NSE of 0.9272, along with the lowest mean MAE of 0.2619. These results suggest that CAT, particularly when optimized with PSO, has a remarkable ability to capture the relationships between the input features and the scour depth, providing highly accurate predictions.

ERT models also performed well, with ERT optimized using PSO slightly outperforming ERT tuned with TPE. The mean CORR values for ERT_PSO and ERT_TPE were 0.9613 and 0.9591, respectively, while the mean NSE values were 0.9200 and 0.9163, respectively. The mean MAE values for these models were 0.2830 for ERT_PSO and 0.2885 for ERT_TPE. The relatively low standard deviations for both CORR and NSE indicate that the ERT models provided stable and consistent predictions across different simulations.

Table 3. Performance metrics of tree-based models with TPE and PSO hyperparameter tuning in MSD estimation (100 simulations per scenario).

Algorithms	CORR		NSE		MAE	
	Mean	STD	Mean	STD	Mean	STD
ERT_TPE	0.959	0.016	0.916	0.031	0.288	0.037
ERT_PSO	0.961	0.014	0.920	0.028	0.283	0.036
CAT_TPE	0.962	0.017	0.923	0.033	0.269	0.037
CAT_PSO	0.964	0.014	0.927	0.029	0.262	0.037
HGB_TPE	0.949	0.021	0.896	0.042	0.313	0.038
HGB_PSO	0.949	0.021	0.897	0.043	0.312	0.037

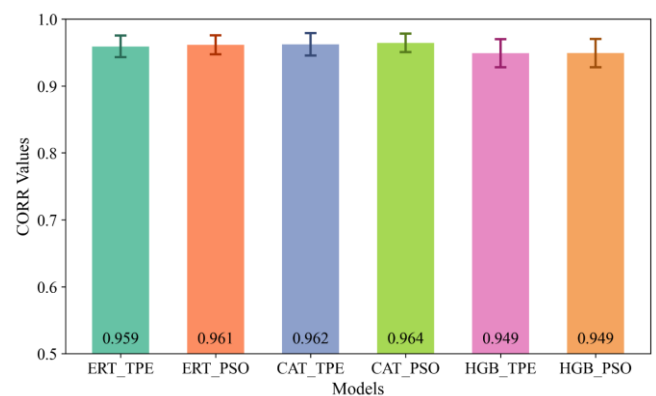


Fig. 1. CORR values of tree-based models with TPE and PSO hyperparameter tuning. Each bar represents the mean CORR and STD CORR values for a model (ERT, CAT, HGB) tuned with TPE or PSO.

For HGB, these models showed slightly lower performance compared to CAT and ERT. HGB tuned with PSO had a mean CORR of 0.9492 and a mean NSE of 0.8968, with a mean MAE of 0.3119. Similarly, HGB optimized using TPE achieved a mean CORR of 0.9488 and a mean NSE of 0.8961, with a mean MAE of 0.3131. The higher standard deviations observed for the

HGB models suggest a greater variability in their predictions, indicating that these models might be more sensitive to changes in the training dataset or the initial hyperparameter settings.

When comparing the tuning methods, PSO generally resulted in slightly better performance metrics than TPE across all models. This trend is consistent with the hypothesis that PSO, by simulating the social behavior of particles and optimizing the search process, can more effectively explore the hyperparameter space, leading to better model configurations. The differences, however, were relatively small, suggesting that both tuning methods are effective, but PSO might have a slight edge in finding more optimal hyperparameter settings.

Generally, the results demonstrate that CAT is the most effective model for predicting MSD, particularly when optimized using PSO. The ERT model also shows strong performance, while HGB, although slightly less accurate, remains a robust option. The use of advanced hyperparameter tuning methods such as PSO and TPE plays a critical role in enhancing model performance, with PSO showing a slight advantage in this study.

4.2 Impact of individual variables on model performance

This part evaluates the effect of individual components on the performance of tree-based models in predicting MSD. Understanding the significance of each input variable is crucial, especially in scenarios where collecting all parameters might be challenging. By identifying the most influential variables, the study aims to provide guidance on which parameters are essential for accurate predictions when some data are unavailable.

To measure the impact of different variables, various data scenarios were evaluated using combinations of the six dimensionless input variables. Each model was subjected to 100 simulations using TPE hyperparameter tuning to measure the uncertainty and variability in performance metrics. Table 4 summarizes the mean and STD of CORR and NSE values for each model across different data scenarios. Fig. 3 illustrates the

comparison of NSE values for ERT, CAT, and HGB models across various input variable scenarios, providing a visual representation of the model's performance.

The results indicate that the inclusion of additional input variables generally enhances model performance. For instance, the simplest scenario (S1) using only F_d and D_{50}/a resulted in lower CORR and NSE values compared to more complex scenarios. In S1, ERT achieved a mean CORR of 0.9048 and a mean NSE of 0.8135, while CAT and HGB had mean CORR values of 0.9023 and 0.8396, and mean NSE of 0.8096 and 0.6975, respectively. The relatively higher STDs in this scenario suggest greater variability in model performance due to the limited input information. As more variables were included, model performance improved significantly. In scenario S3, which incorporated F_d , D_{50}/a , d_t/a , and L/a , both ERT and CAT showed substantial improvements, with mean CORR values of 0.9374 and 0.9367, and mean NSE values of 0.8743 and 0.8744, respectively. HGB also showed improved performance with a mean CORR of 0.8803 and a mean NSE of 0.7684. These results, depicted in Fig. 3, highlight the importance of considering additional geometric and hydraulic parameters to capture the complexity of the scour phenomenon more effectively.

The inclusion of the Froude number (F) in scenario S4 further enhanced model performance. ERT and CAT models demonstrated slight improvements in their performance metrics, with CAT achieving a mean NSE of 0.8746 and a mean CORR of 0.9369. HGB showed a mean NSE of 0.7799 and a mean CORR of 0.8877, indicating that the inclusion of F provided additional valuable information for the models. The most comprehensive scenario (S5), which included all six input variables, yielded the best performance across all models. ERT, CAT, and HGB achieved mean CORR values of 0.9592, 0.9621, and 0.9488, respectively, and mean NSE values of 0.9161, 0.9232, and 0.8961, respectively. The lower standard deviations in this scenario, as shown in Fig. 3, reflect the model's stability and consistency in performance when provided with a complete set of input variables.

Table 4. Performance of ML algorithms with 100 simulations across various data scenarios.

Model	Input Variables	Mean CORR	STD CORR	Mean NSE	STD NSE
S1_ERT	$F_d, D_{50}/a$	0.9048	0.0236	0.8135	0.0434
S1_CAT	$F_d, D_{50}/a$	0.9023	0.0258	0.8096	0.0473
S1_HGB	$F_d, D_{50}/a$	0.8396	0.0332	0.6975	0.0594
S2_ERT	$F_d, D_{50}/a, d_t/a$	0.9046	0.0266	0.8125	0.0481
S2_CAT	$F_d, D_{50}/a, d_t/a$	0.8977	0.0305	0.8006	0.0575
S2_HGB	$F_d, D_{50}/a, d_t/a$	0.8369	0.0386	0.6891	0.0752
S3_ERT	$F_d, D_{50}/a, d_t/a, L/a$	0.9374	0.0193	0.8743	0.0360
S3_CAT	$F_d, D_{50}/a, d_t/a, L/a$	0.9367	0.0196	0.8744	0.0379
S3_HGB	$F_d, D_{50}/a, d_t/a, L/a$	0.8803	0.0366	0.7684	0.0691
S4_ERT	$F_d, D_{50}/a, d_t/a, L/a, F$	0.9339	0.0207	0.8669	0.0386
S4_CAT	$F_d, D_{50}/a, d_t/a, L/a, F$	0.9369	0.0230	0.8746	0.0444
S4_HGB	$F_d, D_{50}/a, d_t/a, L/a, F$	0.8877	0.0378	0.7799	0.0739
S5_ERT	$F_d, D_{50}/a, d_t/a, L/a, F, \sigma_g$	0.9592	0.0151	0.9161	0.0291
S5_CAT	$F_d, D_{50}/a, d_t/a, L/a, F, \sigma_g$	0.9621	0.0160	0.9232	0.0317
S5_HGB	$F_d, D_{50}/a, d_t/a, L/a, F, \sigma_g$	0.9488	0.0208	0.8961	0.0418

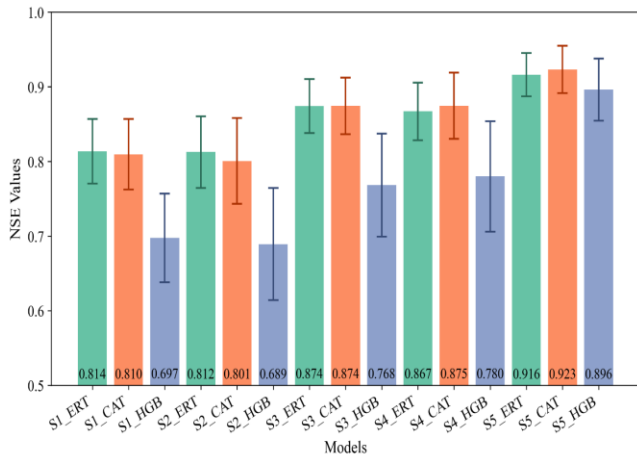


Fig. 3. Comparison of NSE values for ERT, CAT, and HGB models across various input variable scenarios (100 simulations each). Data scenarios vary by including different combinations of input variables.

The analysis demonstrates that the addition of variables (d_r/a , L/a , and σ_g) significantly improves the predictive capabilities of the algorithms. It also underscores the importance of

incorporating a wide range of relevant parameters to accurately model complex hydraulic processes like scour depth. The detailed performance metrics indicate that while ERT and CAT generally outperform HGB, the latter still provides valuable insights and robust predictions, particularly when all input variables are considered. The results clearly show that the inclusion of more comprehensive input variables enhances the performance of tree-based models in predicting MSD. The use of advanced ML techniques, combined with a thorough understanding of the underlying physical processes, allows for the development of highly accurate and reliable predictive models.

4.3 Comparison with empirical formulas

The performance of various empirical formulas for predicting MSD was evaluated and compared using data sourced from multiple studies, including Chatterjee, 1994, Sarkar, 2005, Dey, 2006, and Aamir, 2022. The empirical equations from these studies were applied to the dataset to assess their predictive accuracy. Table 5 summarizes the comparative performance metrics, and Fig. 4 (4a–4e) provides a visual representation through scatter plots of empirical predictions versus observed data.

Table 5. Comparative performance metrics of empirical equations for MSD estimation.

Metrics	Chatterjee 1994	Hopfinger 2004	Sarkar 2005	Dey 2006	Aamir 2022
MAE	1.2363	1.0656	0.8343	0.706	0.8783
CORR	0.7297	0.6716	0.7091	0.775	0.7448
NSE	-0.0703	0.1894	0.4182	0.565	0.2509

The analysis of the empirical formulas reveals distinct variations in performance. The formula proposed by Dey_2006 demonstrated the highest accuracy among the empirical methods, with an MAE of 0.7058, a CORR of 0.7748, and an NSE of 0.5649. These metrics indicate that Dey's formula most closely aligns with the observed scour depth values, suggesting a robust predictive capability under the experimental conditions considered. Figs 4a–4e further illustrate the performance of these empirical formulas through scatter plots. The scatter plots show the distribution of estimated versus observed scour depth values, with marginal distribution lines highlighting the spread of predictions. The visual comparison indicates that Dey's formula has the closest alignment between predicted and observed values, with the points clustering more tightly around the 1:1 line, indicating better predictive accuracy.

Chatterjee, 1994 and Hopfinger, 2004 exhibited relatively higher MAE values of 1.2363 and 1.0656, respectively, indicating larger average errors in their predictions. Their CORR values of 0.7297 and 0.6716 reflect a moderate linear relationship between predicted and observed values. In contrast, their NSE values of -0.0703 and 0.1894 highlight a lower predictive skill, with Chatterjee's formula performing worse than the mean of the observation data. Sarkar, 2005 and Aamir, 2022 showed intermediate performance. Sarkar's formula achieved a CORR of 0.7091 and an NSE of 0.4182. Aamir's formula had a CORR of 0.7448 and an NSE of 0.2509. These results suggest that while these formulas provide reasonable predictions, their accuracy is lower compared to Dey's formula.

Comparing these empirical formulas with the ML models, it is evident that ML methods typically outperform the empirical

formulas. For instance, the best-performing ML model, CAT optimized with PSO, achieved a mean CORR of 0.9644 and a mean NSE of 0.9272, which are significantly better than the metrics for any empirical formula. This stark difference underscores the advantage of using ML techniques that can capture complex non-linear relationships and interactions between predictors, leading to more accurate and reliable predictions.

The trends observed in the data highlight several vital insights. Empirical formulas, while useful for their simplicity and ease of application, often fall short of capturing the full complexity of the scouring process. They tend to be calibrated for specific experimental conditions and may not generalize well to different scenarios. The variability in performance among the empirical formulas suggests that their applicability is limited by the particular datasets and conditions under which they were developed.

In contrast, ML models, with their ability to learn from large and diverse datasets, offer superior predictive performance. Advanced hyperparameter tuning methods such as TPE and PSO further enhance the model's accuracy by optimizing the search for the best model configurations. This comprehensive approach allows ML models to adapt to different data distributions and complexities, making them more robust and generalizable than traditional empirical formulas. Though empirical formulas provide a valuable baseline for predicting MSD, the results demonstrate that ML models offer significantly improved accuracy and reliability.

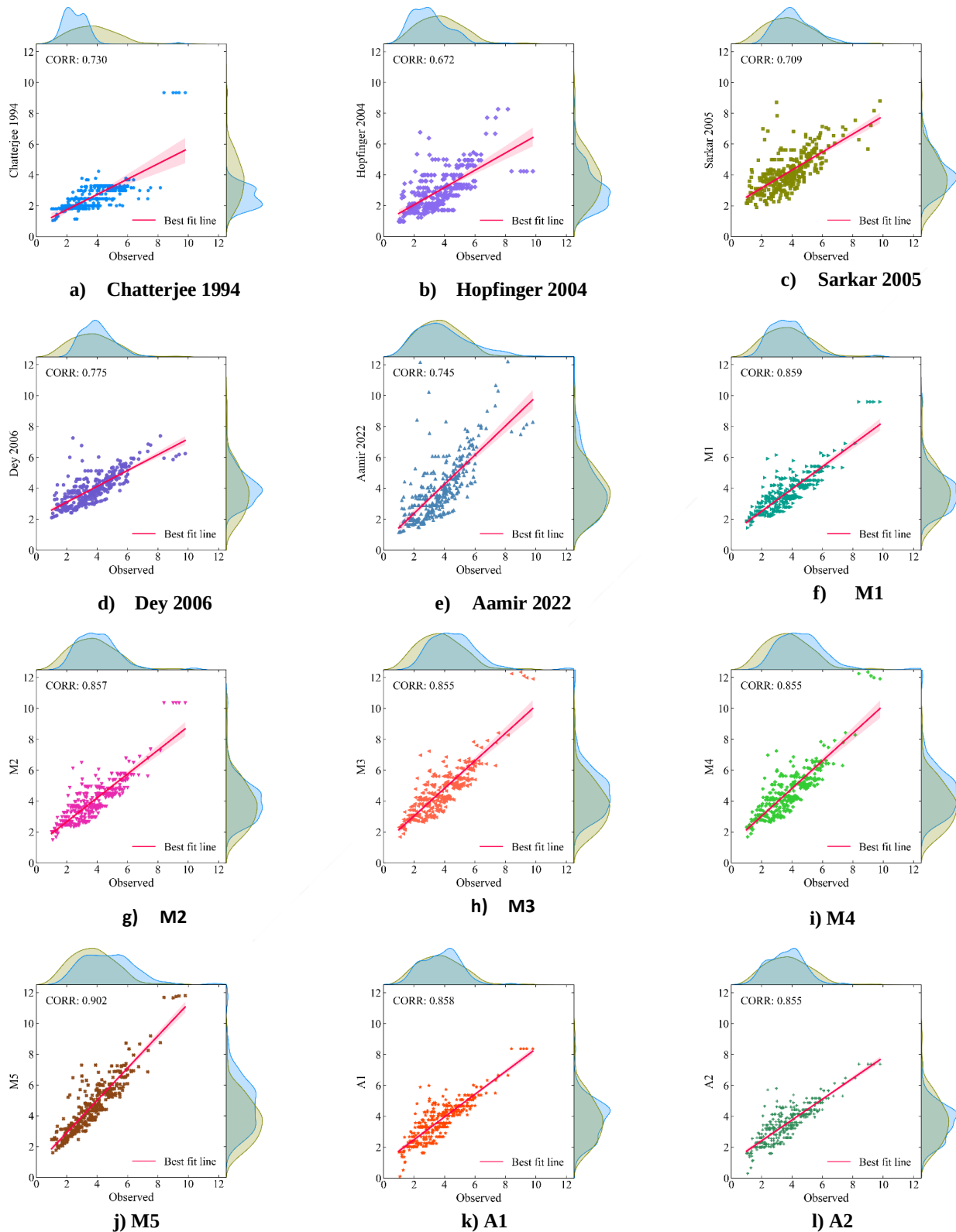


Fig. 2. Comparative scatter plots of empirical equations vs. observed data. Each figure features two distinct distribution lines along the margins. The distributions for the predicted values and observed values are blue and green, respectively.

4.4 New empirical equations

This section presents the development and evaluation of new empirical equations for estimating MSD at sluice outlets. These equations were derived using a dataset consisting of 357 samples, encompassing six input features. Two types of

empirical equations were considered: multiplicative (M) and additive (A) models. Each type was progressively expanded by including additional input features, resulting in seven formulations (five multiplicative and two additive). Table 6 and Fig. 4 (4f–4l) present the developed empirical equations and their

comparative scatter plots against observed data, respectively. The models were evaluated using the entire dataset.

The performance of the seven empirical equations was further evaluated through 100 simulations, where the dataset was randomly split into training (75%) and testing (25%) sets. This approach ensured a comprehensive evaluation by capturing the

inherent uncertainty and variability in model predictions. Table 7 summarizes the statistical performance metrics (mean and STD) for CORR, NSE, and MAE across these simulations and estimations on the testing dataset. Fig. 5 provides a comparative analysis of CORR values for the multiplicative and additive models, illustrating their accuracy and variability.

Table 6. Developed empirical equations for estimating the MSD at sluice outlets.

Code	Proposed Empirical Equations	Eq.
M1	$\frac{d_s}{a} = 0.805 \cdot F_d^{1.135} \cdot \left(\frac{D_{50}}{a}\right)^{0.353}$	(30)
M2	$\frac{d_s}{a} = 0.728 \cdot F_d^{1.105} \cdot \left(\frac{D_{50}}{a}\right)^{0.335} \cdot \left(\frac{d_t}{a}\right)^{0.053}$	(31)
M3	$\frac{d_s}{a} = 0.914 \cdot F_d^{1.149} \cdot \left(\frac{D_{50}}{a}\right)^{0.364} \cdot \left(\frac{d_t}{a}\right)^{0.079} \cdot \left(\frac{L}{a}\right)^{-0.09}$	(32)
M4	$\frac{d_s}{a} = 0.901 \cdot F_d^{1.092} \cdot \left(\frac{D_{50}}{a}\right)^{0.335} \cdot \left(\frac{d_t}{a}\right)^{0.079} \cdot \left(\frac{L}{a}\right)^{-0.09} \cdot F^{0.057}$	(33)
M5	$\frac{d_s}{a} = 134.385 \cdot F_d^{22.770} \cdot \left(\frac{D_{50}}{a}\right)^{11.172} \cdot \left(\frac{d_t}{a}\right)^{0.221} \cdot \left(\frac{L}{a}\right)^{0.03} \cdot F^{-21.844} \cdot \sigma_g^{-0.411}$	(34)
A1	$\frac{d_s}{a} = -48.589 + 22.546 \cdot F_d^{0.151} + 25.521 \cdot \left(\frac{D_{50}}{a}\right)^{0.072}$	(35)
A2	$\frac{d_s}{a} = -6939.4 + 14.015 \cdot F_d^{0.209} + 6942.55 \cdot \left(\frac{D_{50}}{a}\right)^{0.0002}$	(36)

The multiplicative models (M1–M5) showed varying degrees of success. Model M1, which included only F_d and D_{50}/a , achieved a mean CORR of 0.8548, a mean NSE of 0.7263, and a mean MAE of 0.6265. As additional variables were incorporated, the performance of the models improved slightly but not substantially until the inclusion of all six variables in Model M5. Model M5 demonstrated the highest accuracy among the multiplicative models, with a mean CORR of 0.9036 and a mean NSE of 0.8136. These results indicate that the inclusion of more variables generally enhances the model's predictive capability, aligning with the observed trends in previous sections.

Table 7. Statistical performance of newly developed empirical equations for scour depth estimation (100 simulations per model).

Equations	CORR		NSE		MAE	
	Mean	STD	Mean	STD	Mean	STD
M1	0.855	0.036	0.726	0.061	0.626	0.043
M2	0.855	0.037	0.726	0.062	0.627	0.043
M3	0.856	0.039	0.728	0.065	0.618	0.044
M4	0.856	0.039	0.728	0.065	0.618	0.044
M5	0.904	0.028	0.814	0.052	0.462	0.043
A1	0.854	0.035	0.725	0.061	0.611	0.045
A2	0.860	0.035	0.737	0.061	0.596	0.043

The additive models (A1–A2) provided a different approach by combining linear and non-linear terms. Model A1, which

included F_d and D_{50}/a , achieved a mean CORR of 0.8541 and a mean NSE of 0.7254. Model A2, intended to include dt/a , showed similar performance with a mean CORR of 0.8605 and a mean NSE of 0.7369. The comparable performance of A1 and A2 is due to the negligible weight of d/a in the Eq. A2, effectively making A2 similar to A1. The additive model's performance metrics indicate that they are competitive with the multiplicative models, offering a viable alternative for empirical predictions.

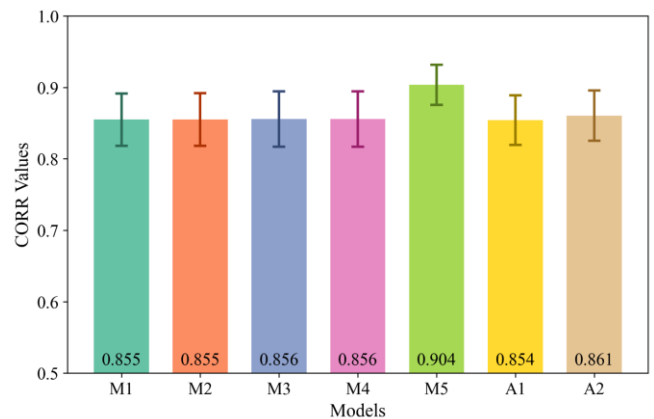


Fig. 3. Comparison of CORR values for multiplicative and additive empirical models in scour depth prediction.

Comparing the newly developed empirical equations with previous empirical formulas (subsection Comparison with Empirical Formulas), it is evident that the new models offer superior predictive accuracy. For instance, the best-performing empirical formula from previous studies (Dey_2006) had a mean

CORR of 0.7748 and a mean NSE of 0.5649. In contrast, the new formulas, particularly M5 and A2, demonstrated significantly better performance metrics.

When compared to ML models, the new empirical equations still lag in terms of predictive accuracy. The best-performing ML algorithm, CAT optimized with PSO, achieved a mean CORR of 0.9644 and a mean NSE of 0.9272, which are considerably higher than those of the new empirical equations. This comparison underscores the advanced capability of ML techniques in capturing complex non-linear relationships and interactions between variables.

The trends observed in the performance metrics highlight several vital insights. The gradual improvement in model accuracy with the inclusion of additional variables emphasizes the importance of considering a comprehensive set of input parameters. This is particularly evident in the performance of Eq. M5, which includes all six variables and shows the highest accuracy among the multiplicative models.

The comparison between multiplicative and additive models reveals that while both approaches have their merits, the multiplicative models generally perform slightly better when all relevant variables are included. This suggests that the power-law relationships assumed in the multiplicative models may more effectively capture the underlying physical processes governing scour depth.

The development of new empirical equations for MSD estimation has resulted in models that significantly improve upon previous empirical formulas. Although ML algorithms still offer superior predictive performance, the proposed empirical equations provide valuable insights and practical tools for scenarios where collecting comprehensive datasets may be challenging. The detailed evaluation and comparison underscore the importance of incorporating a wide range of relevant variables and leveraging advanced modeling techniques to enhance predictive accuracy in hydraulic engineering applications.

5 CONCLUSIONS

The study aimed to enhance the predictive accuracy of MSD estimation at sluice outlets by employing advanced tree-based ML models and developing new empirical equations. The analysis of the ML models, including ERT, CAT, and HGB, demonstrated superior performance compared to traditional empirical formulas. Specifically, CAT optimized with PSO achieved the highest accuracy, with a mean CORR of 0.9644 and mean NSE of 0.9272, highlighting the model's robust predictive capabilities.

The evaluation of individual variables revealed that incorporating a comprehensive set of input parameters enhances model performance. Models including all six variables- F_d , D_{50}/a , L/a , d/a , F , and σ_g -exhibited the highest accuracy, underscoring the importance of considering multiple influencing factors in MSD prediction. The new empirical equations, particularly the multiplicative model M5 and additive model A2, demonstrated improved performance over previous empirical formulas, with M5 achieving a mean CORR of 0.9036 and mean NSE of 0.8136.

Comparing the new empirical equations with traditional formulas and ML models highlighted the advanced capabilities of modern computational techniques. While the new empirical equations offered improved accuracy over previous formulas, ML models remained superior, effectively capturing complex non-linear relationships and interactions between variables.

The quantification of uncertainty in the ML models was achieved through 100 simulations, where the dataset was randomly split into training and testing sets. This rigorous evaluation approach ensured that the model's predictions were robust and reliable, capturing the inherent variability in the data. The lower STD in the performance metrics for the best-performing models indicated high consistency and stability in their predictions.

Despite these advancements, several limitations must be acknowledged. The use of laboratory-based experimental data constrains the study's generalizability, as controlled conditions may not fully reflect real-world scour dynamics at sluice gates. Additionally, while hyperparameter tuning enhances predictive accuracy, it demands significant computational resources, presenting a challenge for applications involving large datasets or real-time analysis.

In conclusion, the study provided significant advancements in the predictive modeling of MSD at sluice outlets. The integration of tree-based models and the development of new empirical equations demonstrated the potential for more accurate and reliable MSD estimates. These findings underscore the value of leveraging comprehensive datasets and advanced modeling techniques in hydraulic engineering applications. The study's methodology and results offer a robust framework for future research and practical implementations in the field.

REFERENCES

- Aamir, M., Ahmad, Z., 2016. Review of literature on local scour under plane turbulent wall jets. *Physics of Fluids*, 28(10). <https://doi.org/10.1063/1.4964659>
- Aamir, M., Ahmad, Z., 2017. Prediction of Local Scour Depth Downstream of an Apron Under Wall Jets, In *Proceedings of Development of Water Resources in India*, Cham; pp. 375-385. 10.1007/978-3-319-55125-8_32
- Aamir, M., Ahmad, Z., 2019. Estimation of maximum scour depth downstream of an apron under submerged wall jets. *J. Hydroinformatics*, 21(4): 523-540. <https://doi.org/10.2166/hydro.2019.008>
- Aamir, M., Ahmad, Z., Pandey, M., Khan, M.A., Aldrees, A., Mohamed, A., 2022. The Effect of Rough Rigid Apron on Scour Downstream of Sluice Gates. *Water*, 14(14): 2223. <https://doi.org/10.3390/w14142223>
- Aderibigbe, O., Rajaratnam, N., 1998. Effect of Sediment Gradation on Erosion by Plane Turbulent Wall Jets. *J. Hydraul. Eng.*, 124(10): 1034-1042. [https://doi.org/10.1061/\(ASCE\)0733-9429\(1998\)124:10\(1034\)](https://doi.org/10.1061/(ASCE)0733-9429(1998)124:10(1034))
- Aderibigbe, O.O., Rajaratnam, N., 1996. Erosion of loose beds by submerged circular impinging vertical turbulent jets. *Journal of Hydraulic Research*, 34(1): 19-33. <https://doi.org/10.1080/00221689609498762>
- Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M., 2019. Optuna: A Next-generation Hyperparameter Optimization Framework, In *Proceedings of The 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Anchorage, AK, USA; pp. 2623-2631. 10.1145/3292500.3330701
- Ali, H.M., El Gendy, M.M., Mirdan, A.M.H., Ali, A.A.M., Abdelhaleem, F.S.F., 2014. Minimizing downstream scour due to submerged hydraulic jump using corrugated aprons. *Ain Shams Eng. J.*, 5(4): 1059-1069. <https://doi.org/10.1016/j.asej.2014.07.007>

- Ali, K.H.M., Neyshaboury, A.A.S., 1991. Localized scour downstream of a deeply submerged horizontal jet. *Proceedings of the Institution of Civil Engineers*, 91(1): 1-18. <https://doi.org/10.1680/iicep.1991.13579>
- Azamathulla, H.M., Ghani, A.A., 2011. ANFIS-Based Approach for Predicting the Scour Depth at Culvert Outlets. *J. Pipeline Syst. Eng. Pract.*, 2(1): 35-40. [https://doi.org/10.1061/\(ASCE\)PS.1949-1204.0000066](https://doi.org/10.1061/(ASCE)PS.1949-1204.0000066)
- Azmathullah, H.M.D., Deo, M.C., Deolalikar, P.B., 2006. Estimation of scour below spillways using neural networks. *Journal of Hydraulic Research*, 44(1): 61-69. <https://doi.org/10.1080/00221686.2006.9521661>
- Balachandar, R., Kells, J.A., Thiessen, R.J., 2000. The effect of tailwater depth on the dynamics of local scour. *Can. J. Civ. Eng.*, 27(1): 138-150. <https://doi.org/10.1139/99-061>
- Bateni, S.M., Borghei, S.M., Jeng, D.S., 2007. Neural network and neuro-fuzzy assessments for scour depth around bridge piers. *Eng. Appl. Artif. Intell.*, 20(3): 401-414. <https://doi.org/10.1016/j.engappai.2006.06.012>
- Bergstra, J., Bardenet, R., Bengio, Y., Kégl, B., 2011. Algorithms for Hyper-Parameter Optimization, In *Proceedings of Advances in Neural Information Processing Systems*, 2011.
- Breusers, H.N.C., Raudkivi, A.J., 1991. *Scouring*. Taylor & Francis, 52 Vanderbilt Avenue, New York, 152 pp.
- Chatterjee, S.S., Ghosh, S.N., Chatterjee, M., 1994. Local Scour due to Submerged Horizontal Jet. *J. Hydraul. Eng.*, 120(8): 973-992. [https://doi.org/10.1061/\(ASCE\)0733-9429\(1994\)120:8\(973\)](https://doi.org/10.1061/(ASCE)0733-9429(1994)120:8(973))
- Dey, S., Sarkar, A., 2006. Scour Downstream of an Apron Due to Submerged Horizontal Jets. *J. Hydraul. Eng.*, 132(3): 246-257. [https://doi.org/10.1061/\(ASCE\)0733-9429\(2006\)132:3\(246\)](https://doi.org/10.1061/(ASCE)0733-9429(2006)132:3(246))
- Dey, S., Westrich, B., 2003. Hydraulics of Submerged Jet Subject to Change in Cohesive Bed Geometry. *J. Hydraul. Eng.*, 129(1): 44-53. [https://doi.org/10.1061/\(ASCE\)0733-9429\(2003\)129:1\(44\)](https://doi.org/10.1061/(ASCE)0733-9429(2003)129:1(44))
- Eghbalzadeh, A., Hayati, M., Rezaei, A., Javan, M., 2018. Prediction of equilibrium scour depth in uniform non-cohesive sediments downstream of an apron using computational intelligence. *Eur. J. Environ. Civ. Eng.*, 22(1): 28-41. <https://doi.org/10.1080/19648189.2016.1179677>
- Farooq, R., Ghumman, A.R., 2019. Impact Assessment of Pier Shape and Modifications on Scouring around Bridge Pier. *Water*, 11(9): 1761. <https://doi.org/10.3390/w11091761>
- Firat, M., Gungor, M., 2009. Generalized Regression Neural Networks and Feed Forward Neural Networks for prediction of scour depth around bridge piers. *Adv. Eng. Softw.*, 40(8): 731-737. <https://doi.org/10.1016/j.advengsoft.2008.12.001>
- Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. *Mach. Learn.*, 63(1): 3-42. <https://doi.org/10.1007/s10994-006-6226-1>
- Guan, D., Liu, J., Chiew, Y.-M., Hong, J.-H., Cheng, L., 2023. A comparison between artificial neural network algorithms and empirical equations applied to submerged weir scour evolution prediction. *International Journal of Sediment Research*, 38(1): 105-114. <https://doi.org/10.1016/j.ijsrc.2022.07.001>
- Guven, A., Azamathulla, H.M., 2012. Gene-expression programming for flip-bucket spillway scour. *Water Sci. Technol.*, 65(11): 1982-1987. <https://doi.org/10.2166/wst.2012.100>
- Guven, A., Gunal, M., 2008. Genetic Programming Approach for Prediction of Local Scour Downstream of Hydraulic Structures. *J. Irrig. Drain. Eng.*, 134(2): 241-249. [https://doi.org/10.1061/\(ASCE\)0733-9437\(2008\)134:2\(241\)](https://doi.org/10.1061/(ASCE)0733-9437(2008)134:2(241))
- Hamidifar, H., Omid, M.H., Nasrabadi, M., 2011. Scour Downstream of a Rough Rigid Apron. *World Applied Sciences Journal* 14(8): 1169-1178.
- Hien, L.X., Hien, L.T.T., Ho, H.V., Lee, G., 2024. Benchmarking the performance and uncertainty of machine learning models in estimating scour depth at sluice outlets. *J. Hydroinformatics*: jh2024297. <https://doi.org/10.2166/hydro.2024.297>
- Hopfinger, E.J., Kurniawan, A., Graf, W.H., Lemmin, U., 2004. Sediment erosion by Görtler vortices: the scour-hole problem. *J. Fluid Mech.*, 520: 327-342. <https://doi.org/10.1017/S0022112004001636>
- Kambekar, A.R., Deo, M.C., 2003. Estimation of pile group scour using neural networks. *Applied Ocean Research*, 25(4): 225-234. <https://doi.org/10.1016/j.apor.2003.06.001>
- Karbasi, M., Azamathulla, H.M., 2017. Prediction of scour caused by 2D horizontal jets using soft computing techniques. *Ain Shams Eng. J.*, 8(4): 559-570. <https://doi.org/10.1016/j.asej.2016.04.001>
- Ke, G. et al., 2017. LightGBM: a highly efficient gradient boosting decision tree, In *Proceedings of The 31st International Conference on Neural Information Processing Systems*, Long Beach, California, USA; pp. 3149-3157
- Kells, J.A., Balachandar, R., Hagel, K.P., 2001. Effect of grain size on local channel scour below a sluice gate. *Can. J. Civ. Eng.*, 28(3): 440-451. <https://doi.org/10.1139/01-012>
- Kennedy, J., Eberhart, R., 1995. Particle swarm optimization, In *Proceedings of ICNN'95 - International Conference on Neural Networks*, 27 Nov.-1 Dec. 1995; pp. 1942-1948 vol.4. 10.1109/ICNN.1995.488968
- Lantz, W.D., Crookston, B.M., Palermo, M., 2022. Evolution of local scour downstream of Type A PK weir in non-cohesive sediments. *Journal of Hydrology and Hydromechanics*, 70(1): 103-113. <https://doi.org/10.2478/johh-2021-0035>
- Laucelli, D., Giustolisi, O., 2011. Scour depth modelling by a multi-objective evolutionary paradigm. *Environ. Modelling Soft.*, 26(4): 498-509. <https://doi.org/10.1016/j.envsoft.2010.10.013>
- Le, H.T.T., Nguyen, C.V., Le, D.-H., 2022. Numerical study of sediment scour at meander flume outlet of boxed culvert diversion work. *PLOS ONE*, 17(9): e0275347. <https://doi.org/10.1371/journal.pone.0275347>
- Le, X.-H., Huynh, T.T., Song, M., Lee, G., 2024. Quantifying Predictive Uncertainty and Feature Selection in River Bed Load Estimation: A Multi-Model Machine Learning Approach with Particle Swarm Optimization. *Water*, 16(14): 1945. <https://doi.org/10.3390/w16141945>
- Le, X.-H., Le, T.T.H., 2024. Predicting maximum scour depth at sluice outlet: a comparative study of machine learning models and empirical equations. *Environ. Res. Commun.*, 6(1): 015010. <https://doi.org/10.1088/2515-7620/ad1f94>
- Lee, T.L., Jeng, D.S., Zhang, G.H., Hong, J.H., 2007. Neural Network Modeling for Estimation of Scour Depth Around Bridge Piers. *J. Hydrodyn.*, 19(3): 378-386. [https://doi.org/10.1016/S1001-6058\(07\)60073-0](https://doi.org/10.1016/S1001-6058(07)60073-0)
- Lim, S.-Y., Yu, G., 2002. Scouring Downstream of Sluice Gate, In *Proceedings of First International Conference on Scour of Foundations (ICSF-1)*, Texas A&M University, College Station, Texas, USA, November 17-20, 2002; pp. 395-409.
- Manes, C., Brocchini, M., 2015. Local scour around structures and the phenomenology of turbulence. *J. Fluid Mech.*, 779: 309-324. <https://doi.org/10.1017/jfm.2015.389>

- Melville, B.W., 2014. Scour at Various Hydraulic Structures: Sluice Gates, Submerged Bridges and Low Weirs. *Australas. J. Water Resour.*, 18(2): 101-117. <https://doi.org/10.1080/13241583.2014.11465444>
- Mutlu Sumer, B., 2007. Mathematical modelling of scour: A review. *Journal of Hydraulic Research*, 45(6): 723-735. <https://doi.org/10.1080/00221686.2007.9521811>
- Najafzadeh, M., Barani, G.A., 2011. Comparison of group method of data handling based genetic programming and back propagation systems to predict scour depth around bridge piers. *Scientia Iranica*, 18(6): 1207-1213. <https://doi.org/10.1016/j.scient.2011.11.017>
- Najafzadeh, M., Lim, S.Y., 2015. Application of improved neuro-fuzzy GMDH to predict scour depth at sluice gates. *Earth Sci. Inform.*, 8(1): 187-196. <https://doi.org/10.1007/s12145-014-0144-8>
- Najafzadeh, M., Rezaie Balf, M., Rashedi, E., 2016. Prediction of maximum scour depth around piers with debris accumulation using EPR, MT, and GEP models. *J. Hydroinformatics*, 18(5): 867-884. <https://doi.org/10.2166/hydro.2016.212>
- Palermo, M., Pagliara, S., Roy, D., 2021. Effect of debris accumulation on scour evolution at bridge pier in bank proximity. *Journal of Hydrology and Hydromechanics*, 69(1): 108-118. <https://doi.org/10.2478/johh-2020-0041>
- Parsaie, A., Haghiabi, A.H., Moradinejad, A., 2019. Prediction of Scour Depth below River Pipeline using Support Vector Machine. *KSCE J. Civ. Eng.*, 23(6): 2503-2513. <https://doi.org/10.1007/s12205-019-1327-0>
- Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A., 2018. CatBoost: unbiased boosting with categorical features. *ArXiv*. <https://doi.org/10.48550/arXiv.1706.09516>
- Qaderi, K., Javadi, F., Madadi, M.R., Ahmadi, M.M., 2021. A comparative study of solo and hybrid data driven models for predicting bridge pier scour depth. *Mar. Georesour. Geotechnol.*, 39(5): 589-599. <https://doi.org/10.1080/1064119X.2020.1735589>
- Sarathi, P., Faruque, M.A.A., Balachandar, R., 2008. Influence of tailwater depth, sediment size and densimetric Froude number on scour by submerged square wall jets. *Journal of Hydraulic Research*, 46(2): 158-175. <https://doi.org/10.1080/00221686.2008.9521853>
- Sarkar, A., Dey, S., 2005. Scour downstream of aprons caused by sluices. *Proceedings of the Institution of Civil Engineers - Water Management*, 158(2): 55-64. <https://doi.org/10.1680/wama.2005.158.2.55>
- Seyedian, S.M., Kisi, O., 2024. Uncertainty analysis of discharge coefficient predicted for rectangular side weir using machine learning methods. *Journal of Hydrology and Hydromechanics*, 72(1): 113-130. <https://doi.org/10.2478/johh-2023-0043>
- Sharafati, A., Haghbin, M., Motta, D., Yaseen, Z.M., 2021. The Application of Soft Computing Models and Empirical Formulations for Hydraulic Structure Scouring Depth Simulation: A Comprehensive Review, Assessment and Possible Future Research Direction. *Arch. Comput. Methods Eng.*, 28(2): 423-447. <https://doi.org/10.1007/s11831-019-09382-4>
- Verma, D.V.S., Goel, A., 2005. Scour Downstream of a Sluice Gate. *ISH J. Hydraul. Eng.*, 11(3): 57-65. <https://doi.org/10.1080/09715010.2005.10514801>
- Virtanen, P. et al., 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3): 261-272. <https://doi.org/10.1038/s41592-019-0686-2>
- Watanabe, S., 2023. Tree-Structured Parzen Estimator: Understanding Its Algorithm Components and Their Roles for Better Empirical Performance. *ArXiv*. <https://doi.org/10.48550/arXiv.2304.11127>
- Xie, C., Lim, S.-Y., 2015. Effects of Jet Flipping on Local Scour Downstream of a Sluice Gate. *J. Hydraul. Eng.*, 141(4): 04014088. [https://doi.org/10.1061/\(ASCE\)HY.1943-7900.0000983](https://doi.org/10.1061/(ASCE)HY.1943-7900.0000983)

Received 28 August 2024
Accepted 8 December 2024