

A Study for Local Scour Prediction using Machine Learning Based on Distribution Prediction

Jaemyeong Choi¹, Soonchul Kwon^{1,*}

¹Civil Engineering, Pusan National University, Republic of Korea

Abstract: Scour occurs due to the strong eddies and increased flow velocity generated by structures around bridge piers, which disrupt the river flow. Scour adversely affects bridge piers by altering the riverbed cross-section around the river. Recent advances in scour depth prediction have included empirical formulas based on field data and experimental results, numerical analyses that directly calculate fluid dynamics, and machine learning models that learn patterns from data. This study developed a bridge pier scour depth prediction model using the NGBoost model, which quantifies uncertainty based on probability distribution predictions rather than single predicted values through machine learning techniques. As a result of comparative analysis of scour prediction with various empirical formulas, the NGBoost model showed superior predictive power, which suggests that NGBoost is a model that can more accurately predict scour occurrence based on learning the hydraulic phenomena occurring around effective piers with big data.

Keywords: Scour, Machine Learning, Prediction, Big Data, Risk Assessment, Climate Change

I. Introduction

Stream scour consistently has a detrimental effect on bridge piers. When stream flow encounters a bridge, it accelerates around the pier, causing scour in the riverbed (Fig. 1). Pier scour is a complex hydrodynamic problem caused by fluid-structure-like interactions around the bridge foundation. It is influenced not only by the structural characteristics of the bridge but also by environmental factors [1,2]. In particular, local scour around structures such as piers is dominated by the strong vortex structures generated when the flow is disturbed by the structure. As the main flow approaches the pier, the pressure gradient formed across the pier causes a portion of the flow to shift into a strong downflow toward the riverbed. This downflow generates a strong horseshoe-shaped vortex at the pier foundation, transporting sediment in the riverbed and forming a wake flow behind the pier.

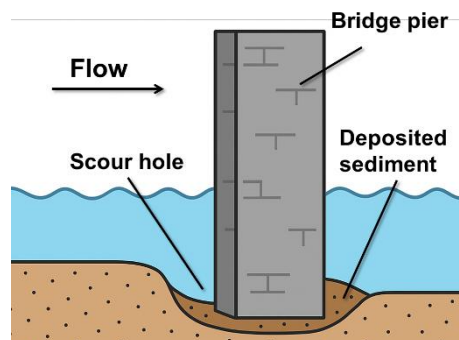


Figure 1: Illustration of local scour around pier

Horseshoe vortices, a major cause of scour, exert high shear stress on the riverbed, eroding bed materials. Scour progresses and intensifies when the outflow of sediment from around a pier exceeds the inflow of sediment [3,4]. This type of scour is a factor that threatens the structural stability of the bridge. Scour is the leading cause of bridge failures in the United States, with an average of 22 bridges per year failing or being closed due to severe deformation. Of the 503 bridge failures that occurred in the United States between 1989 and 2000, 53% were attributed to hydraulic factors such as flooding and scour [5]. Bridge damage and collapse can result in significant social and economic impacts due to structural reconstruction and maintenance costs, and in extreme cases, loss of life [6]. Research on scour depth prediction has evolved through various approaches. Major prediction techniques include: 1) empirical formulas based on field observations and experimental data, 2) numerical analysis that directly solves the fluid dynamics equations, and 3) machine learning models that learn patterns from large amounts of data. Empirical formulas formalize the relationships among key factors affecting scour depth through statistical analysis of hydraulic experiments or field observations.

However, because most formulas are developed from limited data in controlled laboratory environments, they struggle to accurately reflect the complex hydraulic and geomorphological characteristics of real rivers. In particular, complex bridge pier geometries, unsteady flow during floods, and the behavior of cohesive sediments are typical factors that empirical formulas fail to account for. Furthermore, when multiple empirical formulas are applied to the same conditions, significant deviations in predicted results often occur, raising concerns about reliability [7].

To further understand the physical mechanism of the scour phenomenon, a numerical analysis, i.e., a computational fluid dynamics-based approach, was used to directly solve the governing fluid equations on a computer to simulate in detail the three-dimensional flow structure around the pier and the resulting pseudo-transport process.

NGBoost (Natural Gradient Boosting) is a boosting technique. It is a specialized framework for performing probabilistic predictions that quantify prediction uncertainty, going beyond the single value predictions of existing models. NGBoost predicts probability distributions rather than single predicted values, thereby providing confidence intervals for the results, supporting more safety-oriented decision-making. In this study, we utilize the NGBoost model, which has not been utilized in the field of scour depth prediction, to predict scour depth, including prediction intervals, through distributional prediction. We compare this prediction with existing empirical models.

II. Experimental Section

NGBoost is a boosting technique. It goes beyond predicting a single value like conventional models and specializes in probabilistic predictions that can quantify the uncertainty of predictions. While the goal of conventional boosting models is to find predictions that minimize error, NGBoost seeks a probability distribution that maximizes the probability of actual data. To achieve this, NGBoost assumes that data follows a specific probability distribution and introduces the concept that the model is trained to predict the mean and standard deviation of that distribution. Conventional gradient boosting uses a general gradient to reduce the loss function. However, the space of probability distributions is not flat like Euclidean space but rather curved, making the general gradient a suboptimal path. NGBoost utilizes natural gradients to address this issue. Natural gradients consider the curvature of the probability distribution space and suggest directions for optimizing model parameters most efficiently and reliably. This helps the model learn the optimal probability distribution more quickly and reliably.

This study used the USGS Bridge Pier Scour database. The database contains 569 laboratory data sets and 1,858 field data sets, of which 552 laboratory data sets were used. The variable ranges of the data used in the study are shown in Table 1.

Table 1: Variables range data used in the study

	b (ft)	V (ft/s)	V_c (ft/s)	y (ft)	d_{50} (mm)	$y_s f(t)$	Data
Min	0.05	0.49	0.73	0.07	0.22	0.01	552
Max	3.00	7.08	4.18	6.23	7.80	4.63	
Mean	0.35	1.68	1.43	0.88	1.19	0.44	

III. Results

To determine the influence of various factors on the occurrence of scour, the correlation between each parameter was analyzed using the Pearson correlation coefficient, and the results are presented in Fig. 2. A strong positive correlation ($r = 0.86$) was found between the pier width (b) and the scour depth (y_s). This is because as the pier width increases, the flow cross-section narrows, increasing the velocity and consequently promoting sediment movement. This correlation may be overestimated when analyzing other data. Water depth (y) and scour depth (y_s) showed a positive correlation ($r = 0.53$). It is interpreted that the deeper the water depth, the greater the hydraulic force acting on the pier, which in turn intensifies the scour. On the other hand, the correlation with the velocity (V) ($r = 0.14$) was relatively.

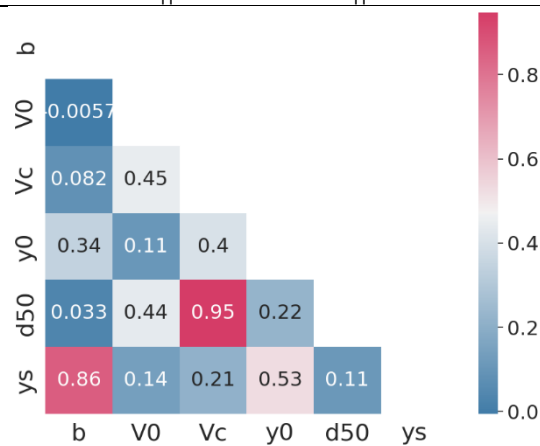


Figure 2: Input variables correlation analysis

On the other hand, the correlation with the velocity (V) ($r = 0.14$) was relatively weak, which is thought to be because scour is more sensitive to the relative velocity compared to the critical velocity than to the absolute velocity. The critical velocity (V_c) and scour depth (y_s) showed a low positive correlation ($r = 0.21$).

In theory, a high critical velocity should suppress sediment transport and reduce scour, but the positive correlation appears to be due to the tendency for large bridges to be located in sections with high V_c values.

The low correlation ($r = 0.11$) between particle diameter (d_{50}) and scour depth (y_s) suggests that scour is determined by a complex interaction with other factors, such as flow velocity and water depth, rather than solely by particle size. Meanwhile, d_{50} and V_c showed a very high correlation ($r = 0.95$), which reflects the physical relationship in which the critical velocity is determined by the difference in particle diameter and density.

This reflects the fundamental characteristics of sediment transport dynamics. In summary, the correlation analysis results confirmed that scour depth is determined by the interaction of geometric factors, flow characteristics, and material properties. In particular, y_s - b (0.86) and d_{50} - V_c (0.95) were analyzed as hydraulic and topographic governing factors of scour. Based on the correlation analysis between these input-dependent variables, this study derived a total of five dimensionless variables using the Buckingham- π theorem.

Table 2: NGBoost performance evaluation

Model	Data	RMSE	R^2	SI	Bias	Se	MAPE(%)
NGBoost	Train	0.21	0.82	0.14	-0.0010	0.21	14.38
	Validation	0.25	0.75	0.18	-0.0047	0.25	17.39
	Test	0.23	0.75	0.16	0.0184	0.23	17.17

Table 2 shows the results of the performance evaluation indices of the NGBoost model. The NGBoost model exhibits high accuracy (high R^2 , low root mean square error) and low bias ($B \approx 0$) in predicting pier scour depth. It also demonstrates excellent performance in quantifying the uncertainty (SE) of the predictions in a probabilistic and stable manner. The high coefficient of determination (R^2) demonstrates excellent fit, demonstrating that the model has learned the physical relationships in the data well in predicting pier scour depth. However, some nonlinear factors (e.g., flow disturbances, unsteady flow rates) remain as residuals. The root mean square error (RMSE), which represents the overall prediction accuracy of the model, is interpreted in units of scour depth, and was found to have little bias in the training and validation sets.

For the training data, the center of the residuals was approximately 0, the distribution was very similar to a normal distribution, and B was approximately 0 (-0.001), confirming that there was almost no bias between the predicted and measured values. The fact that RMSE and Se were identical (≈ 0.21) indicates that the model's residual distribution was stable and had a consistent error level. Unlike the training data, which was approximately normal, the error distribution in the validation data was wider than that in the training data and showed an asymmetric shape with a slightly longer right tail. The fact that RMSE and Se are identical (≈ 0.21) indicates that the model's residual distribution is stable and has a consistent error level. Unlike the training data, which is approximately normally distributed, the error distribution on the validation data is wider and slightly asymmetrical, with a longer right tail. This suggests the possibility of some over prediction, and the RMSE, at 0.25, is slightly higher than that on the training data.

However, since most of the residuals still exist within the $B \pm 1.96Se$ range, the generalization performance of the model is good, and the predictions are overall balanced, so a slight increase in variance compared to training is judged to be a realistic generalization error level. As a result of analyzing the comprehensive performance index, in the case of NGBoost, the RMSE was 0.21 and 0.82 on the training data, and the RMSE was 0.25 and 0.75 on the validation data, showing relatively small performance degradation, suggesting that it is not overly specialized in the training data and shows good generalization performance, and that the probabilistic prediction performance is stable. Overall, NGBoost is not limited to point predictions, but provides distribution predictions through learning conditional distributions, which can support more safety-oriented decision-making by providing confidence intervals for the results. Based on the model's error distribution estimates and performance evaluation metrics, the NGBoost model exhibits high accuracy, low root-mean-square error (RMSE), and a bias ($B \approx 0$) in predicting pier scour depth. The high coefficient of determination (R^2) demonstrates excellent fit, demonstrating a good understanding of the physical relationships in the data for predicting pier scour depth.

IV. Conclusion

Data-driven machine learning techniques are actively being adopted to improve the accuracy and efficiency of scour depth prediction. While many machine learning studies on scour depth prediction have achieved accurate predictions, interpretable machine learning research is scarce. Furthermore, NGBoost, a boosting model specialized in probabilistic prediction that quantifies prediction uncertainty beyond the single value prediction of existing models, has not been used in previous studies. Therefore, this study developed a scour depth prediction model for bridge piers using the NGBoost model, which can quantify uncertainty based on probability distribution predictions using machine learning techniques. This suggests that NGBoost is a model capable of more accurately predicting scour occurrence based on learning effective hydraulic relationships. These results demonstrate the NGBoost model's potential for engineering utility and practical application in scour analysis and the development of preventive measures.

V. Acknowledgements

“This work is financially supported by Korea Ministry of Environment (MOE) as Graduate School specialized in Climate Change.”

References

- [1] C. Wang, X. Yuo, F. Liang, A review of bridge scour: mechanism, estimation, monitoring and counter measures, *Nature Hazards*, 87, 2017, 1881-1906.
- [2] J.V. Klinga, A. Alipour, Assessment of structural integrity of bridges under extreme scour conditions, *Engineering Structures*, 82(1), 2015, 55-71.
- [3] M. A. Rafiqii, M.A. Lone, M. A. Tantray, Exploring the challenges and solutions for river scour in dynamic environments: a comprehensive review, *World Journal of Engineering*, 21(6), 2024, 1045-1063.
- [4] M.R. Bharadwaj, L.K. Gupta, M. Pandey, M. Valyrakis, Countermeasures for local scour around the bridge pier: a review, *Acta Geophysica*, 73, 2025, 701-728.
- [5] K. Wardhana, F.C. Hadipriono, Analysis of recent bridge failures in the United States, *Journal of Performance of Constructed Facilities*, ASCE, 17(3), 2009, 144-150.
- [6] L. Brandimarte, P. Paron, G.D. Baldassarre, Bridge Pier Scour: A review of processes, measurements and estimates, *Environmental Engineering and Management Journal*, 11(5), 975-989.
- [7] D.M. Sheppard, W. Miller, Live-bed local pier scour and experiments, *Journal of Hydraulic Engineering*, 132(7), 2006, 635-642.
- [8] D.G. Park, Policy measures to enhance the infiltration performance of permeable pavers, 2022, The Seoul Institute, South Korea.