

Evaluation of nonparametric tree models for predicting the scour depth of bridge piers

R. Dalvand*, M. Komasi

Department of Civil Engineering, Faculty of Engineering, Ayatollah Ozma Borujerdi University, Borujerd, Iran.

E-mail: rezvan.dalvand1@iran.ir

Received: 04.06.2019. Accepted: 03.07.2019

One of the main causes of damage to bridges, especially during flood event is the scour around the bridge. Determination of the depth of scour around the bridge piers plays a very important role in designing the bridges against this destructive factor. The complexity of the bridge scour and the effects of different parameters on its estimation more clearly reveal the necessity of using a nonlinear and comprehensive model in this field. In this present study, decision tree models, as nonparametric models, are used to estimate the scour depth. Furthermore, the statistics of different bridges and four tree methods are used. The data used to train and test decision trees including flow the velocity of upstream, the median grain size, flow depth, and the pier width, the skew of the pier to approach flow, the length of the pier, the grain size of bed material for which 84 percent is finer, a multiplying factor, input variables, and the depth of scour as output in the model. 75% of the available data is used for model training and the remaining 25% for testing. The results show that among the four models (CART, C5, QUEST, CHAID) examined, C5 model, considering the comparison of the root mean square error parameters and the coefficient of determination, is more accurate in computing the scour depth of the bridge, the amount coefficients of determination in this model is in training and testing steps are 0.92 and 0.76, as well as the mean square error values of the error is 0.56 and 0.72 respectively. Furthermore, the results reveal the QUEST model does not have a proper accuracy in scour depth estimation. Furthermore, the analysis of the models shows flow depth, the flow velocity in the upstream have the greatest effect on the scour depth.

Keywords: Scouring; sensitivity analysis; tree models

Introduction

Scouring at the base of bridge piers is the major cause of bridge collapses worldwide. Therefore, computing the scour risk of bridge foundations is the key to a correct management and allocation of resources for maintenance and scour mitigation works (Tubaldi et al. 2017). Scouring is a natural phenomenon caused by the flow of rivers or floods and the result is erosion by the flow of water, causing the transfer and erosion of materials from the floor, the beaches and around bridge piers. In many cases, it destroys the bridge and hydraulic structures. By controlling and protecting the bridges against scouring and proved proper scouring forecasting, it is possible to prevent these damages. Scour depth calculation methods are divided into three methods: experimental equation, laboratory, and data analysis. Heretofore, there are different experimental equations in the field of calculation of scouring around bridge piers. Among the equations presented in this field are Sutherland (1988) and Melville, Lim and Cheng (1998) and Mia and Nago (2003). One of the most famous equations provided by the University of Colorado, known as the CSU equation and also the HEC-18 which calculates the maximum scour depth around the bridge piers in two modes of clear water and a moving bed (Arneson et al. 2012). Another method is to calculate the scour depth of the bridge piers using a laboratory method, in this field, as a sample of Hosseini et al. (2011) in this research, the application of submerged plates in controlling scour around the bases of rectangular bridge piers with circular nose was investigated. In a study titled "Operation of collar geometry", they found locally scoured at the base of the bridge piers which use of both square collar and circular collar, especially under the substrate, reduces scour depth and also the square type versus the circular type, has a better performance in reducing the scour depth (Sharyati et al., 2009). The complexity of the scouring phenomenon around the bridge piers is calculated by some error due to the three-dimensional nature of the flow and the formation of multiple vortices in the scouring hole. In another research in this field, the study of scour depth at the sloping base of bridges along the direct path has been studied by laboratory studies; this study aims to examine the effect of pier inclination angle on the scouring process by conducting tests along a straight channel in a Plexiglas laboratory flume. To this end, cylindrical piers with four different inclination angles were placed in the flume along a straight channel and tests were

conducted at four different flow rates under clear water conditions. The obtained results showed that the maximum and minimum scouring depths occurred at pier angles of 0 and 15 respectively. Similar to the scoured depth, dimensions of the scoured hole decreased with increasing pier angle (i.e., changing the pier orientation with respect to the flow direction). (Karimi et al. 2017 (Accordingly, researchers use data analysis methods beside with experimental equations and laboratory methods. In the research, groundwater level prediction has been based on classification and regression tree, which is based on the response relationship between influential factors such as rainfall and reservoir level and the change of groundwater level, the influential factors of groundwater level were selected. Then the classification and regression tree (CART) model was constructed by the subset and used to predict the groundwater level and for comparison, this model has been used to support vector machine (SVM). It is indicated that CART model has not only better fitting and generalization ability, but also strong advantages in the analysis of landslide groundwater dynamic characteristics and the screening of important variables. It is an effective method for prediction of ground water level in landslides.)Zhao et al. 2016(Also, Soleimanpour et al. (2017) in a research ,the decision tree was used to determine the most effective undesirable factors in drinking water in Kazeroun plain located west of Fars province. The results of this research were obtained using the CART decision tree. The results showed that the two factors of soluble salts and calcium have a greater effect on the poor quality of drinking water in this plain. Regarding the history of research and research in this field, the study of CART, C5, QUEST and CHAID models has not yet been investigated. In this study, we tried to investigate all decision tree models for measuring the scour depth of bridges, and a comparison is made between these methods. This paper is organized as follows: materials and methods are presented in Section II. Conclusion and analyze is presented in Section III. Section IV explains the Summery and Conclusion.

Materials and Methods

The local scour around the piers of the bridge is one of the main causes of the bridge failure. The local scour of river water has a catastrophic effect on the engineering structures. The damage to the hydraulic and civil structure due to local scour is a global problem and has been studied experimentally and numerically by many researchers for several decades (Zaid et al., 2019). Scouring, having a variety of types, is due to reduction cross section, natural scouring, local scouring, etc. In this study, data from Federal Highway Administration of the United States (Mueller and Wagner, 2005) have been used to construct models and validate decision trees. This data is taken from the field and is related to the scour depth around the piers of bridge in different states. This information is the velocity of upstream (V), the median grain size (D50), flow depth (Y) and the pier width. (b), the skew of the pier to approach flow (α), the length of the pier. (h), the grain size of bed material for which 84 percent is finer (D84), a multiplying factor (k), input variables and the depth of scour (ds). The data analysis used in this study is shown in Table 1.

Table 1. Specifications of the study data (Mueller and Wanger, 2005).

Data type	Max	Min	Variance	Average	Mode
K	3	1	0.8	1.8	1
b (ft)	18	0.9	16.9	5.1	0.9
L (ft)	90	8	128.6	31.3	24
α (deg)	85	0	102.1	5.4	0
Y(ft)	65.7	0.5	137.1	13.9	1.9
v (ft per s)	12.3	0.2	6.2	4.3	3.7
D (50 Mm)	108	0.1	333	8.1	0.9
D (84 Mm)	0.233	1	9.1331	1.19	5.3
ds (ft)	0.13	5	7.9	5.3	0.1

Non-parametric decision tree models

Decision tree is a simple classification algorithm and has been widely used in knowledge discovery and pattern recognition fields, which can be used to deal with the multi-classification tasks (Guan et al., 2017). Decision trees are a new generation of data analysis technicians which today is used in various sciences to extract laws and relationships governing a phenomenon. The decision tree forms an inverted tree with its root at the highest point and the leaves at the bottom of it. The CART tree classification regression algorithm is introduced by Breiman et al. in 1984. It was a binary order tree that divides the problem into small parts. (Fürnkranz et al., 2012) This method created its branches in binary form and only based on an independent variable, it is based on the condition defined in each node and the information on that node is divided into two parts. In this algorithm, for selecting the test variable, the least squared deviation criterion according to (1) is used.

$$ss(t) = \sum_{i=1}^{N_t} (y_i(t) - \bar{y}(t))^2 \quad (1)$$

where $ss(t)$ the least is squared deviation criterion, N_t the number of records in the node t , $y_i(t)$ is the value of the target variable in the node t , and $\bar{y}(t)$ is the mean of the target variable in the node t . Generally, an input variable is used as the best attribute to create a branch in node t , the function of the following is maximized.

$$Q(X,t)=SS(t)-[SS(t_R)+SS(t_L)] \quad (2)$$

where $SS(t_R)$ and $SS(t_L)$ are in the right and left branches of the node respectively (Jabari and Samadi, 1392).

Another type of tree model is the C5 algorithm, which is able to categorize continuous data and noise data. This algorithm is used in cases where the data has a discrete domain act the same method as before and for data that has a continuous domain for all selectable states, the threshold limit is considered and for that threshold limit, it measures the amount of information efficiency and the threshold that has the highest efficiency will be considered as the decision making indicator for that node (Hamedani, 2012).

Another type of tree model is the CHAID method. In order to construct a decision tree, CHAID divides the data into similar subsets in turn as far as each subset contains a specified number of samples. In some cases, this algorithm can produce a tree that acts non-binary; in fact, it uses the separation method instead of binary separation. In this way, the root node can be divided by more than two times. This algorithm uses the chi-square test to decide on each branch to determine branch nodes. Then, the branches of the tree are made to fulfill the criterion of stopping or reaching the level of complexity demanded, pruned. In other words, CHAID first finds the differences of each sample with other samples and produces the desired tree. Pruning the tree through the find is the same difference. This decision tree combines statistical methods with other clustering and classification methods and obtains a criterion for the feature selection for each tree inner node (Chattamvelli, 2011). Another type is the QUEST model, which resembles the CART tree of a binary single-variable tree which uses the linear separation criterion. The growth process of the QUEST decision tree involves selecting a separator predictor variable, selecting a separation point for the selected predictor variable. In QUEST algorithm, only single-variable divisions are considered (Lim et al., 2000). The most important feature that distinguishes QUEST method from the other methods, it is possible to prune the tree after its complete construction and is usually implemented in this way which is considered a threshold limit. If the probability of occurrence is a leaf of the tree is less than the threshold of its adjacent leaves then it removes the leaf Or if necessary, combines with adjacent leaves. The purpose of this is to reduce the height of the tree to prevent over-learning and eliminate noise data (Hamedani, 2012).

Theory and Analysis

In order to create a model, the data are randomly divided into two parts: training and testing dataset and divided proportionally 75% and 25% respectively to better compare the performance of the proposed models from: Root Mean Squared Error (RMSE), Coefficient of Determination (R^2) have been used. RMSE is the difference between the predicted value of the model or the statistical estimator and the real value. RMSE is a good tool to compare prediction errors by a dataset and is used to compare multiple datasets whatever the amount is lower the model has a lower error rate (Zahiri, 2015).

$$RMSE = \sqrt{\frac{1}{N} \sum (d_{so} - d_{sc})^2} \quad (3)$$

$$R^2 = 1 - \frac{\sum (d_{so} - d_{sc})^2}{\sum (d_{so} - \bar{d}_{so})^2} \quad (4)$$

In eq (3) to (4), RMSE: Root mean square error, N: Number of data, d_{so} : Observation scour depth, d_{sc} : computational scour depth, R^2 : Coefficient of Determination, d_{so} : observation scour depth; \bar{d}_{sc} : mean depth of computing scour, Ratio, $\sum d_{so}$: total depth of observation scour, $\sum d_{sc}$: total sum of scouring depths are computational. After creating the model using decision trees, CART, QUEST, CHAID, C5 a set of rules will be identified and displayed. Since the CART and C5 models are not able to estimate values beyond the scope described in the training, the training set should be selected in such a way the upper and lower bounds of each of the variables in the training set.

CART tree model results

After investigating the variables affecting the scouring phenomenon in this model, the velocity of upstream (V), the median grain size (D50), flow depth (Y) and the pier width (b), the skew of the pier to approach flow (α), the length of the pier. (h), the grain size of bed material for which 84 percent is finer (D84), a multiplying factor (k), input variables and the depth of scour (ds) was introduced as the output in the model. The proposed model of the CART decision tree will be in Figure 1 after the construction and pruning of the additional branches. As seen in this figure, in the first node, flow depth (Y), Selected as the best trait to create a branch, this indicates the effect of the, flow depth on the scour. After depth of flow in upstream, the remaining variables were used. As shown in Figure 2, the CART algorithm consists of several nodes and the end of each node as output indicates the predicted scour depth for the node's information. Each node contains the number of data in that node and the average scour depth predicted at the end of each specified node for example, in the first node, its number is zero the total number of training data is 225 This node contains 100% of the training data and the average scour depth predicted is 65.3 feet in the following, all nodes are sorted in the same way, in this model, nodes are considered as 26 pcs.

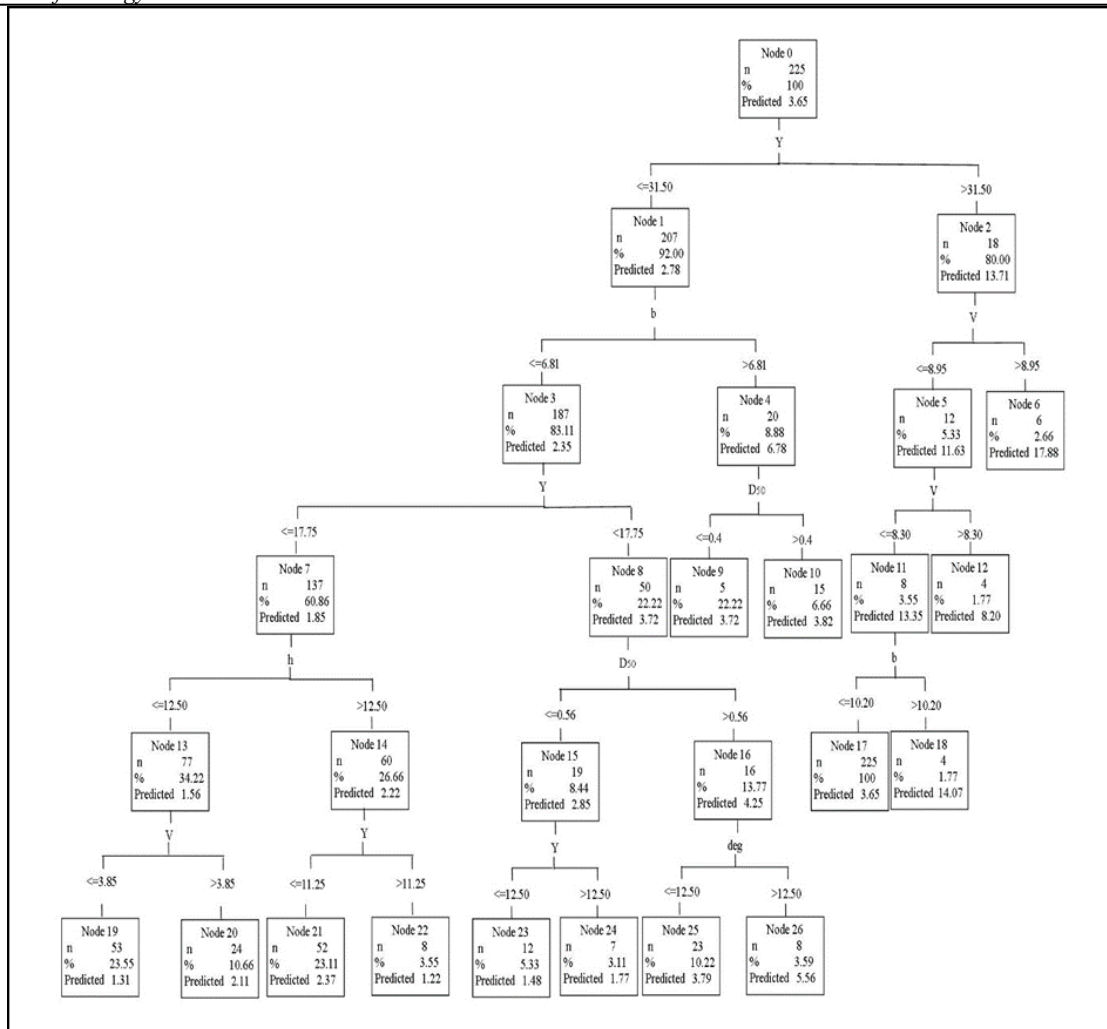


Figure 1. CART Model Tree View to Estimate Scour Depth.

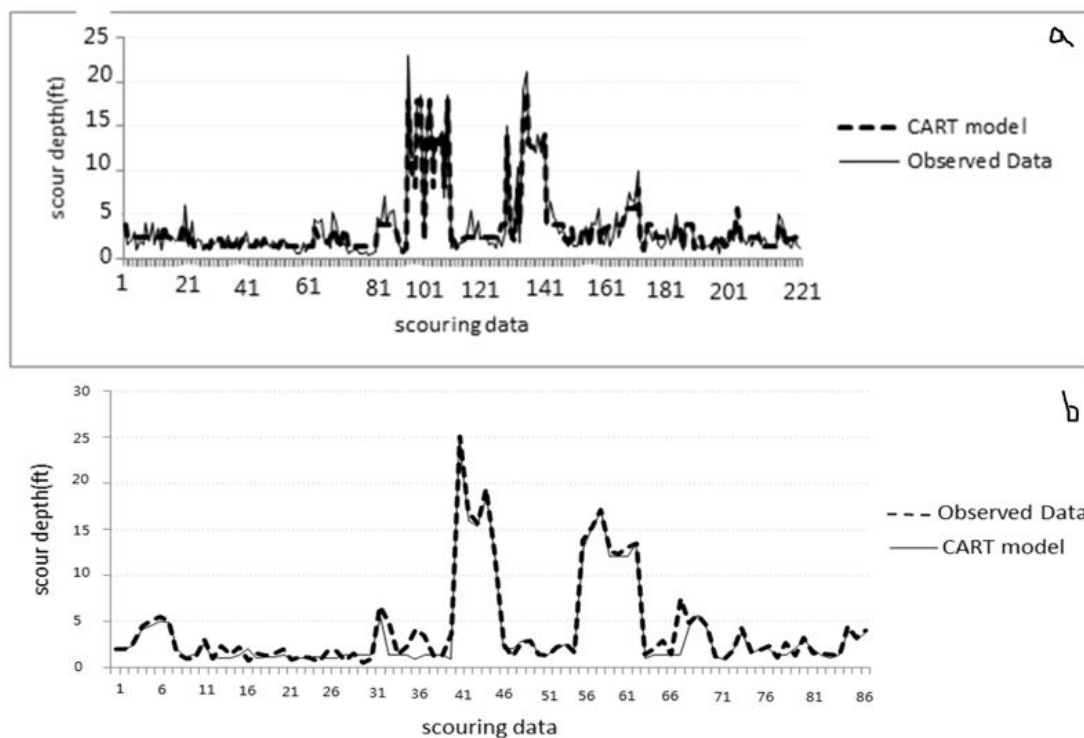


Figure 2. (a) Observational data and CART model at training stage; (b) Observational data and CART model at testing stage; the results of the observational data are plotted against the computational data.

In Figures 2 (a) and (b), the results of the observational data are plotted against the computational data. In many cases, the results of computational quantities are matches with observational values; this indicates that the CART model with a coefficient of 0.88 at the testing stage and 0.51 in the training stage has a good performance.

C5 tree model results

After examining the variables affecting scouring is in the first node of this model, the flow depth factor (Y) selected as the best trait to create a branch, which shows effect of flow depth on scouring. After creating the C5 model, this model was used to estimate scour depth using training and testing data that results are presented in Figure 3.

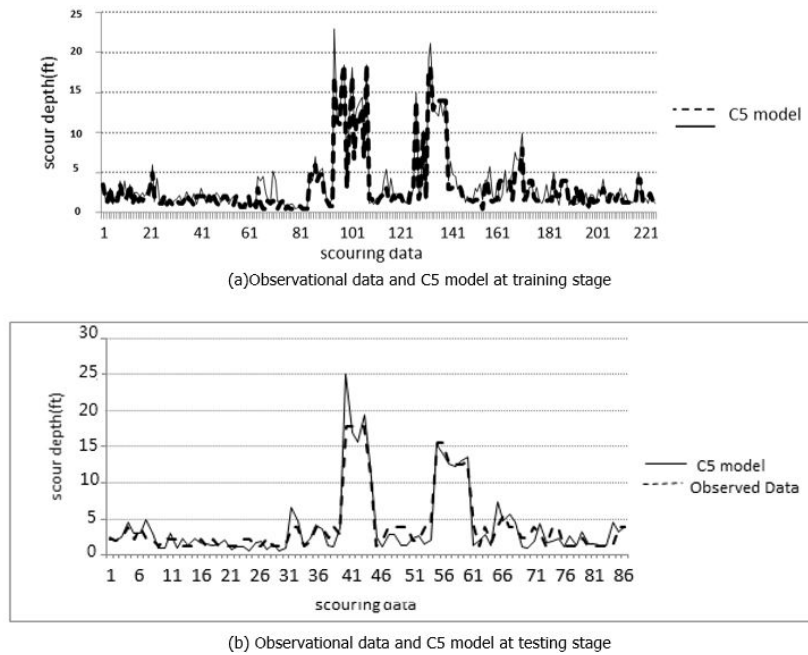


Figure 3. (a) and (b), the results obtained from the C5 model.

As shown in Figures 3 (a) and (b), the results obtained from the C5 model are well matched with observational values, Figure 3(a) is related to the training step and Figure 3(b) is related to the testing step, in both phases, the C5 has a good performance.

CHIAD tree model results

After entering the scouring parameters as input and scour depth as the modeling output in the software, in CHIAD model, the length of the pier (h) is selected as the best trait to create the branch. This indicates the effect of the length of the pier (h) on the scour depth in the CHAID model (Figure 4).

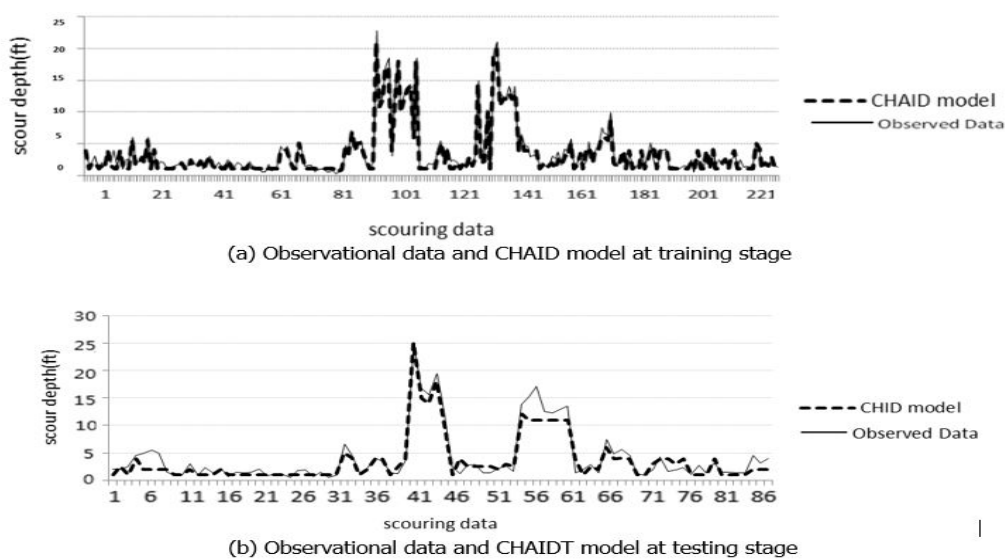


Figure 4. Results of CHAIDT model at testing stage.

QUEST tree model results

According to modeling carried out in previous models, modeling was done in the QUEST model and eight input parameters were given to the software, as well as the output, the depth of scour was considered, after analyzing the data and results, the

pier width (b) was selected as the best trait to create the branch, but this model does not offer interesting results for the calculation of scour depth, in the following the results of this model are presented in Tables 2 and 3.

Table 2. Comparison of the efficiency of tree models estimating the scour depth in the training step.

Tree model	CART	C5	QUEST	CHAID
DR	1.05	0.82	0.64	0.65
R ²	0.88	0.92	0.1	0.85
RMSE(ft)	0.59	0.56	4.33	0.89

Table 3. Comparison of the efficiency of tree models estimating the scour depth in the testing step.

Tree model	CART	C5	QUEST	CHAID
DR	0.93	0.89	1.84	0.59
R ²	0.51	0.76	0.12	0.76
RMSE(ft)	0.61	0.72	5.2	0.44

According to the analyzes made in Tables 2 and 3, if DR equals one, it indicates that the model results and observational values are consistent that's in the CART model of these values are close to number one in the training and testing phases, which indicates that the results of this model and the observed values are in good agreement. If DR>1, the model will estimate scour more than the actual value, as shown in Tables 2 and 3, the QUEST model in the testing step calculated this parameter much more than a one which indicates that this model measures the scour values much more than its actual value. Otherwise, if DR<1, the scour is less than the real value in these models, the computational scour values in the C5 and CHAID models in both training and testing and QUEST model training stages calculated the value of this parameter less than 1, indicating that these models have a value for scouring Less than real values. It is also observed that the CART and C5 models have a low RMSE in the training step, and the CHAID model at the testing stage has lower values of this parameter. The Coefficient of Determination (R²), because of the variance ratio or the oscillation of a variable (which can be predicted by another variable) is very useful. . This estimate allows a certain amount to be predicted by a given graph or model. The range of this parameter is between zero and one, the more the parameter is closer to one, and the more data are close to the fit line. In both methods, CART and C5 have values close to number one, which indicates that these two models have better correlation with observational values than other models. The QUEST model, as mentioned earlier, has not yielded acceptable results in this regard.

Sensitivity analysis of models

The β parameter is used to measure the sensitivity of each model to input variables. Application of the coefficient β is when there is more than one independent variable in the model, in this case, the value of β helps to compare the relative contribution of each variable to the prediction of the dependent variable, and in other words, it can be determined which variables have the greatest effect on the dependent variable. In this analysis of dependent variable, the scour depth calculated by models and independent variables is considered by other parameters, and this coefficient is calculated by equation (4).

$$\beta = \frac{\text{cov}(x_i, y_i)}{\text{var}(y_i)} \quad (4)$$

In this case: x_i : values of independent parameters, y_i : the values of the parameter are dependent.

Table 4. Percentage of sensitivity (β) effective variables on scour depth in different tree models.

Model	QUEST	CHAID	C5	CART
K	5.2	4.4	5.4	0.9
b	-4.1	-2	0.5	1.4
h	-11.2	-13.5	4.2	2.9
α	8.5	-17.7	-13.4	10.8
Y	41.5	28.5	40.3	45.9
V	-10.7	19.1	40.7	36.8
D ₅₀	-2.4	-26.6	-26.5	-25.7
D ₈₄	9.3	7.1	5.5	4.3

The coefficient β is measured relative to the value of one, the coefficient β one, greater than one and smaller than one, indicates the coherence, severity, and tenderness of the sensitivity of the parameters with the calculated scour depth of the models. According to Table 4, in four models presented for estimation of scour depth, flow depth (Y) and velocity of upstream

(V) have the greatest impact on scour depth, after these two parameters, a multiplying factor (k) has the most effect on the scour depth. The sensitivity coefficient of some of the parameters with a negative sign indicates the low effect of these parameters on the scour depth.

Conclusion

In the present study, four models CART, C5, QUEST and CHAID are examined. According to the results obtained through the data QUEST and the CART model has less nodes than other models, then both of them can be used in practice. After the construction of the C5 model for estimation of scour depth, using training and testing data, the results from the C5 and CHAID models are well suited for observational values. In order to compare the performance of the proposed models, statistical analyses were applied: the root mean square error parameters and the coefficient of determination. According to the parameters, the C5 model and the CART model have acceptable and reliable responses among the four examined models, and after these two models, CHAID responses are satisfying. The sensitivity analysis was performed using β coefficient that used to measure the sensitivity of each model to input variables, after obtaining this coefficient for different parameters, four models presented for estimation of scour depth were observed: flow depth (Y) and velocity of upstream (V) have the greatest impact on scour depth, after these two parameters, a multiplying factor (k) has the most effect on the scour depth. After examining the models, it was observed that the C5 and CART model with coefficients 0.92 and 0.88 at training stage and 0.76 and 0.51 at the testing stage were more accurate than the four models and the CHAID model is in the category after these two models with a coefficient of determination 0.85 and 0.76 respectively in the training and testing phase and the QUEST model has not been accurate in terms of data type.

References

- Arneson, L. A., Zevenbergen, L. W., Lagasse, P. F., & Clopper, P. E. (2012). Evaluating scour at bridges.
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees. Wadsworth & Brooks. Cole Statistics/Probability Series.
- Chattamvelli, R. (2011). Data mining algorithms. Alpha Science International, Ltd.
- Fürnkranz, J., Gamberger, D., & Lavrač, N. (2012). Foundations of rule learning. Springer Science & Business Media.
- Guan, X., Liang, J., Qian, Y., & Pang, J. (2017). A multi-view OVA model based on decision tree for multi-classification tasks. Knowledge-Based Systems, 138, 208-219.
- Hamedani, T. (2012). Decision tree algorithms" Mashhad Ferdowsi University, Iran.
- Hossieni, H., Dalir, A., Farsadizadeh, D., Aronaghi, H., & Ghorbani, M. (2011). Application of submerged plates to control the scouring around the base of a rectangular bridge with a rounded nose. Civil Engineering and Surveying Journal, 9(10), 58-67.
- Jabari, A., & Samadi, M. (2013). Application of M5 algorithm and prediction of scour depth in downstream overflows. Iran Water Resources Management Co.
- Karimi, N., Heidarnejad, M., & Masjedi, A. (2017). Scour depth at inclined bridge piers along a straight path: A laboratory study. Engineering Science and Technology, An International Journal, 20(4), 1302-1307.
- Lim, S. Y., & Cheng, N. S. (1998). Prediction of live-bed scour at bridge abutments. Journal of Hydraulic Engineering, 124(6), 635-638.
- Lim, T. S., Loh, W. Y., & Shih, Y. S. (2000). A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. Machine Learning, 40(3), 203-228.
- Tubaldi, E., Macorini, L., Izzuddin, B. A., Manes, C., & Laio, F. (2017). A framework for probabilistic assessment of clear-water scour around bridge piers. Structural Safety, 69, 11-22.
- Melville, B. W., & Sutherland, A. J. (1988). Design method for local scour at bridge piers. Journal of Hydraulic Engineering, 114(10), 1210-1226.
- Mia, M. F., & Nago, H. (2003). Design method of time-dependent local scour at circular bridge pier. Journal of Hydraulic Engineering, 129(6), 420-427.
- Mueller, D. S., & Wagner, C. R. (2005). Field observations and evaluations of streambed scour at bridges (No. FHWA-RD-03-052).
- Sharyati, H., khodashenas, R., Esmaili, K. (2009). Investigate the operation of the coupling geometry in local scouring at the bridge base. Hydraulic Scientific Journal, 11(5), 14-27.
- Soleimanpour, S. M., Mesbah, S. H., & Hedayati, B. (2018). Application of CART decision tree data mining to determine the most effective drinking water quality factors (case study: Kazeroon plain, Fars province). Iranian Journal of Health and Environment, 11(1), 1-14.
- Zahiri, J. (2015). Riprap Design for Bridge Piers Using Nonparametric Models.
- Zaid, M., Yazdanfar, Z., Chowdhury, H., & Alam, F. (2019). A review on the methods used to reduce the scouring effect of bridge pier. Energy Procedia, 160, 45-50.
- Zhao, Y., Li, Y., Zhang, L., & Wang, Q. (2016). Groundwater level prediction of landslide based on classification and regression tree. Geodesy and Geodynamics, 7(5), 348-355.

Citation: Dalvand, R., Komasi, M. (2019). Evaluation of nonparametric tree models for predicting the scour depth of bridge piers. Ukrainian Journal of Ecology, 9(3), 1-7.



This work is licensed under a Creative Commons Attribution 4.0. License