

Estimation of scour depth around circular piers: applications of model tree

Amir Etemad-Shahidi, Lisham Bonakdar and D.-S. Jeng

ABSTRACT

Scour around bridge piers is one of the main causes of bridge failures and is of great importance for hydraulic engineers and scientists. Prediction of the scour depth around piers is complicated, and accurate results are rarely achieved by the existing models. Recently, data mining approaches such as artificial neural networks and fuzzy inference systems have been applied successfully to predict scour depth around hydraulic structures. In this study, an alternative robust data mining approach was used for the predictions of the scour depth around piers, and the results were compared with those of three empirical approaches. Performances of developed models were tested by experimental data sets collected in laboratory experiments and field measurements, together with existing empirical approaches. Statistical measures indicate that the proposed M5' model provides a better prediction of scour depth than the empirical approaches.

Key words | bridge pier, data mining approaches, M5' model tree, scour, soft computing, unidirectional flow

Amir Etemad-Shahidi (corresponding author)

D.-S. Jeng

Griffith School of Engineering,
Griffith University Gold Coast Campus,
QLD 4222,
Australia
E-mail: a.etemadshahidi@griffith.edu.au

Lisham Bonakdar

Leichtweiss-Institute for Hydraulic Engineering and
Water Resources,
Technische Universität Braunschweig,
Beethovenstr. 51a,
38106 Braunschweig,
Germany

NOTATION

CC	correlation coefficient
D	pier diameter
d_{50}	median sediment diameter
Fr	Froude number
g	gravitational acceleration
I_a	index of agreement
n	number of measurements
Re	pier Reynolds number
S	equilibrium scour depth
sd	standard deviation
SDR	standard deviation reduction
SI	scatter index
S/Y	dimensionless scour depth
U	flow velocity
U_c	critical flow velocity
x	measured value
Y	the flow depth
Y/D	relative water depth
y	predicted value

ρ	fluid density
μ	fluid dynamic viscosity

INTRODUCTION

Local scour around piers is one of the common reasons of bridge failures during floods. Numerous bridge damages due to extreme scour around their piers have been reported recently (FDOT 2010). These damages result in huge economic loss and even human loss (Toth & Brandimarte 2011). Several bridges have been damaged due to storm and flood-induced scour around the world in both developed and developing countries (e.g., Blodgett 1978).

An accurate estimation of the maximum scour depth around bridges is vital in the design of bridge piers in term of safety and economics (e.g., Muzzammil 2010; Muzzammil & Alam 2011; Khan *et al.* 2012). Numerous studies have been conducted in the recent decades to develop a robust method for

estimation of the equilibrium scour depth due to the current (e.g., Melville 1997; Bateni *et al.* 2007; Azmathulla *et al.* 2010; Ghaemi *et al.* 2013; Etemad-Shahidi & Rohani 2014).

There have been numerous small-scale laboratory experiments, mainly on cylindrical piers, using dimensional analysis of different formulae available in the literature. In these formulae, both the scour depth and influential parameters such as flow velocity and depth are given by non-dimensional variables. For example, Shen (1971) suggested a formula based on the pier Reynolds number while Breusers *et al.* (1997) used only the relative water depth in their equation. On the other hand, the HEC-18 equation (USDOT 2001) considered the Froude number and relative water depth as the governing parameters. In another approach, Melville (1997) considered the relative sediment size, relative approach velocity, and relative pier diameter in their equation. However, these semi-empirical methods show a large difference in the estimation of the scour depth (e.g., Breusers & Raudkivi 1991; Bateni *et al.* 2007). This discrepancy comes from the complexity of the problem, limited number of considered variables (Ettema *et al.* 1998), and the scaling effects (Lee & Sturm 2009), which is more vital in the prototype cases (Gulbahar 2009). Gaudio *et al.* (2013) showed that some of the semi-empirical scour formulae are very sensitive to different input parameters and a small error in an input parameter might significantly change the scour depth. However, they did not provide or suggest the most accurate formula.

Nowadays, traditional statistical analysis is replaced by artificial intelligence (AI)-based approaches which have been applied in different fields of engineering (Muzzammil & Ayyub 2010). Researchers have recently invoked data mining approaches to resolve the above-mentioned issues. Recently, these approaches have been used for tackling various complex problems in hydraulic engineering (e.g., Bhattacharya & Solomatine 2005; Zanganeh *et al.* 2009; Ayoubloo *et al.* 2010; Azamathulla & Ghani 2010; Farhoudi *et al.* 2010; Zanganeh *et al.* 2011; Azamathulla 2012; Etemad-Shahidi & Taghipour 2012; Pal *et al.* 2013). Artificial neural networks (ANN) are the most commonly used method in this category. ANNs have been invoked to estimate scour around culverts (Liriano & Day 2001), downstream of a ski-jump bucket (Azmathulla *et al.* 2005), scour below pipelines (Kazeminezhad *et al.* 2010), scour around pile groups (Ghazanfari *et al.*

2011), local scour depth at bridge piers (Toth & Brandimarte 2011), and scour depth around spur dikes (Karami *et al.* 2012). Bateni *et al.* (2007) applied ANNs and adaptive neuro-fuzzy inference systems (ANFIS) to estimate scour depth. They found that ANN outperforms ANFIS and previous empirical approaches and could be a suitable procedure to predict scour depth.

In summary, there have been several attempts to apply data mining methods for the prediction of scour depth around bridge piers (e.g., Bateni *et al.* 2007; Toth & Brandimarte 2011; Azamathulla 2012; Khan *et al.* 2012; Pal *et al.* 2013; Akib *et al.* 2014). However, the previous models did not provide a transparent and compact relationship between the governing parameters that can give us insight about the physics of the process. In addition, most of the previously developed models were based on small-scale laboratory experiments rather than field measurements to evaluate their performance in prototype situation. An alternative data mining approach called M5' (Wang & Witten 1997) has been recently applied to provide compact and physically sound formulae in engineering problems. The main advantages of the model trees are that they are easily applied and yield comprehensible, compact, and transparent formulae (e.g., Bonakdar & Etemad-Shahidi 2011; Etemad-Shahidi & Jafari 2014). This method has been successfully used in modeling sediment transport (Bhattacharya *et al.* 2007), wind estimating from waves (Daga & Deo 2009), wave height predictions (Etemad-Shahidi & Mahjoobi 2009), land cover classification (Pal 2006), evapotranspiration (Pal & Deswal 2009), and design of rubble-mound breakwaters (Etemad-Shahidi & Bonakdar 2009; Etemad-Shahidi & Bali 2011; Jafari & Etemad-Shahidi 2012). The aim of this study is to explore how much this method will lead to an improvement in the scour depth prediction, particularly in terms of accuracy and efficiency. To achieve this goal, different M5' models are developed, and the results are compared with those of existing formulae and against the available laboratory experimental data.

PREVIOUS APPROACHES AND THE USED DATA SET

Previous approaches

Scour depth around piers is governed by variables characterizing the flow, fluid, sediments, and pier geometry, which

can be expressed as (Ettema *et al.* 1998)

$$S = f(\rho, \mu, U, Y, g, d_{50}, U_c, D) \quad (1)$$

where S is the scour depth, ρ is the fluid density, μ is the fluid dynamic viscosity, U is the approach flow velocity, Y is the flow depth, g is the gravity, d_{50} is the median sediment diameter, U_c is the critical velocity for initiation of sediment motion, and D is the pier diameter. The formulae obtained from small-scale laboratory experiments commonly invoke dimensional analysis for the estimation of scour depth. One of the commonly used functional relationships between dimensionless numbers is as follows (Ataie-Ashtiani & Beheshti 2006):

$$S/Y = f(Y/D, Fr, U/U_c, D/d_{50}, Re) \quad (2)$$

where Fr is the Froude number of approach flow $U/(gY)^{1/2}$ and Re is the pier Reynolds number ($\rho UD/\mu$). Using the following functional relationship, several semi-empirical formulae have been suggested previously for scour depth prediction and three of them which use different dimensionless numbers are mentioned in Table 1. As pioneers of this field, Shen *et al.* (1969) used selected laboratory data from Chabert & Engeldinger (1956) and Shen *et al.* (1966) studies and stated that scour depth around circular piles depends on the pier Reynolds number. Using the same data set, HEC-18

formula was developed and then modified and became USDOT (2001). In this formula Re is ignored and dimensionless scour depth is mainly a function of Froude number and relative water depth. On the other hand, Melville (1997) used a more extensive laboratory data set and by physical argument and push curve fitting stated that the dimensionless scour depth around circular piles depends on relative water depth, relative velocity (U/U_c), and relative size of the sediments (D/d_{50}).

Johnson (1995) applied seven equations to field data in both live and clear conditions. Her results showed that Shen's (1971) formula performs better in shallow conditions while the USDOT formula is better for $Y/D > 1.5$. She also found that there is a significant difference between the results of different formulae and most of the semi-empirical equations overestimate the scour depth. Gulbahar (2009) compared the performances of different equations using field data in different hydrological conditions. This study showed that there is no unique best formula and the skills of different methods vary in different conditions.

Recently, soft computing methods have been widely applied to handle complicated hydraulic engineering problems (e.g., Zanganeh *et al.* 2009; Yasa & Etemad-Shahidi 2013). For example, Bateni *et al.* (2007) developed ANN and ANFIS models for predicating the scour depth and its temporal evolution. They compared their results with those of previous empirical approaches and reported that

Table 1 | Three semi-empirical formulae using different dimensionless numbers

Model	Formula	Notes
USDOT (2001)	$S/Y = K Kw (Y/D)^{-0.65} (Fr)^{0.43}$	$S_{\max} = 3.0 D$ for $Fr > 0.8$ $S_{\max} = 2.4 D$ for $Fr < 0.8$ $K = f$ (nose shape, current angle of attack, mode of sediment transport, armoring by bed material) $Kw =$ correction factor when $(Y/D) < 0.8$; $(D/d_{50}) > 50$ & $Fr < 1$ $Kw = 2.58 (Y/D)^{0.34} (Fr)^{0.65}$ for $U/U_c < Kw = (Y/D)^{0.13} (Fr)^{0.25}$ for $U/U_c > 1$
Breusers <i>et al.</i> (1977)	$S/D = 2K_V \tanh(Y/D)$	$K_V = 1$ for $U/U_c > 1$ $K_V = (1 - 2U/U_c)$ for $0.5 > U/U_c > 1$ $K_V = 0$ for $0.5 < U/U_c$
Melville (1997)	$S/D = K$	$K = f$ (nose shape, relative water depth, current angle of attack, relative velocity, relative sediment size)
Conventional nonlinear regression	$S/Y = 1.46 (Y/D)^{-0.56} (Fr)^{0.57} (U/U_c)^{0.12}$	

a multi-layer perception model outperforms the ANFIS and other regression models in predicting the scour depth. They attributed the superiority of ANN to its ability in solving complex problems. Azmathulla *et al.* (2010) used genetic programming to predict the scour depth. They also compared their results with those of USDOT (2001) and showed that their model outperforms both ANN and regression equations. Recently, Pal *et al.* (2012) used field data of Mueller & Wagner (2005) to develop a model for scour depth prediction using M5 and showed that their formula outperforms those of previous ones. However, they did not provide a dimensionally homogeneous formula.

Data set

To have a wider range of parameters, 14 data sets, i.e., Chabert & Engeldinger (1956), Hancu (1971), Ettema (1980), Jain & Fischer (1980), Chee (1982), Chiew (1984), Yanmaz & Altinbilek (1991), Kothyari *et al.* (1992), Graf (1995), Melville (1997), Melville & Chiew (1999), Oliveto & Hager (2002), Sheppard & Miller (2006), and unpublished data from the University of Auckland were used to predict the equilibrium scour depth in this study. The whole data set consists of 283 laboratory experimental data which were used for developing the models and evaluating the existing formulae. The

distribution and the statistics of the governing dimensionless parameters are shown in Figures A1–A5 (Appendix A, available online at www.iwaponline.com/jh/017/051.pdf). As shown in Appendix A, the flow conditions are mostly subcritical with 75% clear water conditions and 25% live bed tests.

The above-mentioned data sets were first used to evaluate the performances of the existing formulae. As mentioned before, semi-empirical approaches reported in the literature have different forms with different dimensionless numbers. Among these, three different formulae which have been more commonly used in engineering applications, i.e., Breusers *et al.* (1977) (which considers Y/D and is hyperbolic), Melville (1997) (which considers U/U_c and D/d_{50}), and USDOT (2001) (which considers Fr and Y/D) were selected for the evaluations. Figures 1–3 show that the scatters between the measured and predicted scour depths estimated by these approaches are large. It is worth noting that the existing models predict more or less constant scour depths for the measured values greater than 0.25 m. In addition, Breusers *et al.*'s (1977) formula tends to underpredict scour depths. This is mainly because in this formula scour depth is zero for $U/U_c < 0.5$.

The following statistical parameters were used for the quantitative evaluation of the models skills: index of

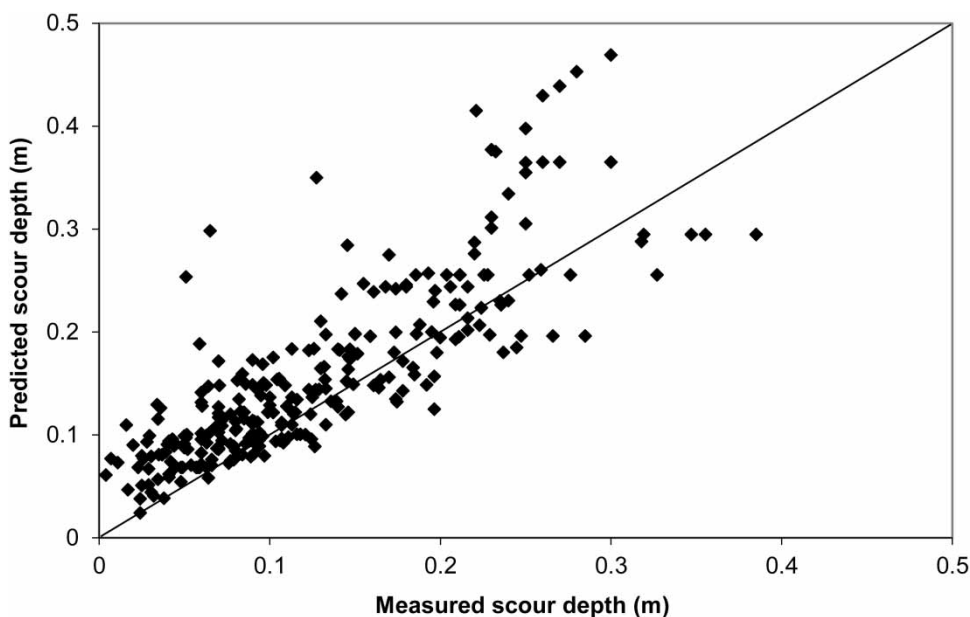


Figure 1 | Comparison between the measured and predicted scour depths using $K = 2.2$ and Kw in the empirical formula of USDOT (2001), all laboratory data.

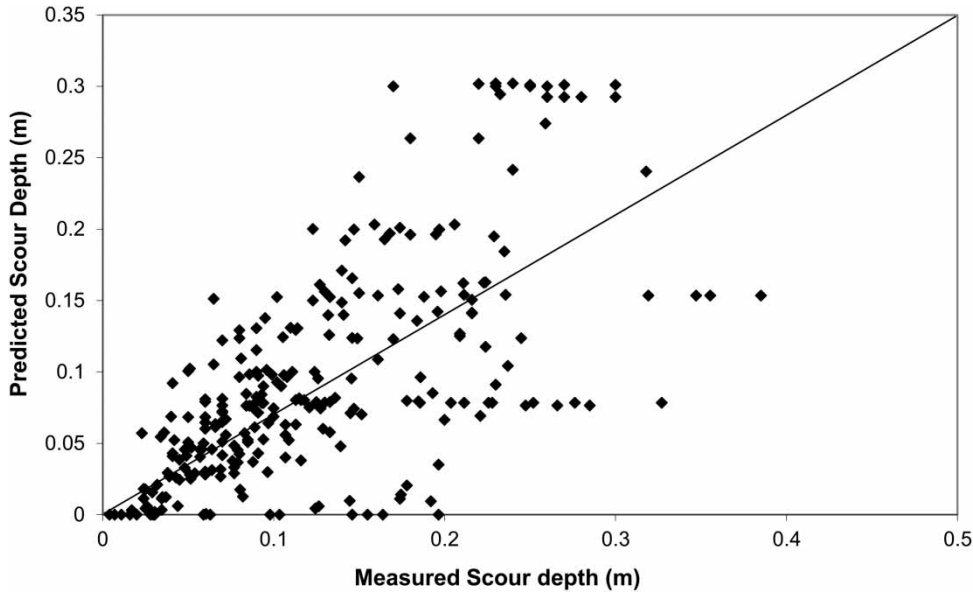


Figure 2 | Comparison between the measured and predicted scour depths using the empirical formula of Breusers et al. (1977), all laboratory data.

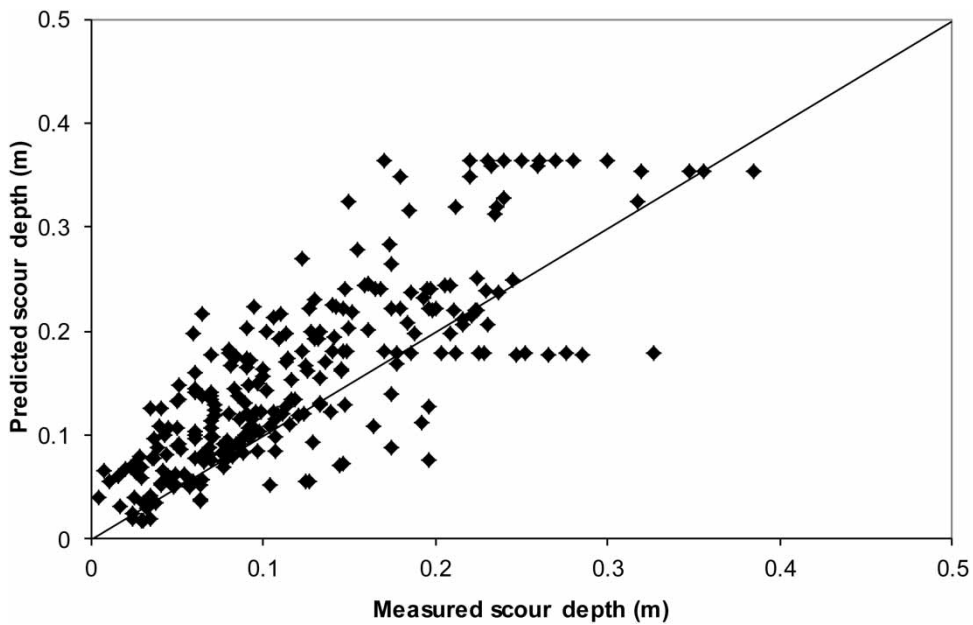


Figure 3 | Comparison between the measured and predicted scour depths using the empirical formula of Melville (1997), all laboratory data.

agreement (I_a), scatter index (SI), and ‘Bias’

$$I_a = 1 - \frac{\sum_{i=1}^n (x_i - y_i)^2}{\sum (|x_i - \bar{x}| + |y_i - \bar{y}|)^2} \quad (3)$$

$$SI = \frac{\sqrt{1/n \sum (y_i - x_i)^2}}{\bar{x}} \quad (4)$$

$$\text{Bias} = \bar{y} - \bar{x} \quad (5)$$

where x_i and y_i denote the predicted and the measured values, respectively, and n is the number of measurements. \bar{x} and \bar{y} are the corresponding mean values of the predicted and measured parameters. The error measures of these formulae are also shown in Table 2. This table shows that the Melville approach yields more accurate results while Breusers et al.'s (1977) formula is the least reliable one and generally underpredicts the scour depths, which is not safe for design purposes.

DECISION TREE AND M5' ALGORITHM

A decision tree is one of the most recent data mining methods that can be applied for classifications and predictions. In general, decision trees can be divided into two main types: classification trees and regression trees. The first type classifies instances or data records based on some attributes (input parameters) and is used when the model's output includes non-numeric values while a regression tree is applied when the model's output includes numeric values. A decision tree is similar to an inverse tree with a root node at the top and some leaves at the bottom. In general, decision trees represent a disjunction of conjunctions of constraints on the values of input parameters. Unlike other soft computing methods such as ANNs, decision trees represent rules or formulae. In fact, each path from the tree root to a leaf corresponds to a conjunction of attribute tests and the tree itself to a disjunction of these conjunctions. Decision trees classify instances by sorting them down the tree from the root node to some leaf node. Each node in the tree specifies a test of some attribute of the instance, and each branch descending from that node

corresponds to one of the possible values for this attribute (Hand et al. 2001; Kantardzic 2003).

Model trees, which are a type of decision tree with linear regression functions at the leaves, form the basis of a modern technique for predicting continuous numeric values. Structurally, a model tree takes the form of a decision tree with linear regression functions instead of terminal class values at its leaves. The M5 model tree is a numerical prediction algorithm, and the nodes of the tree are chosen over the attribute (input parameters) that maximizes the expected error reduction as a function of the standard deviation of the output parameter (Zhang & Tsai 2007). The M5 model tree was first introduced by Quinlan (1992) and was expanded in a method called M5' by Wang & Witten (1997). Model trees have a large number of advantages, making them a suitable regression method for performance analysis. The prediction accuracy of model trees is comparable to that of techniques such as ANNs (Etemad-Shahidi & Mahjoobi 2009) and is known to be higher than that of CART (Classification And Regression Tree) method (Ould-Ahmed-Vall et al. 2007). The advantage of a model tree is that it can efficiently handle large data sets with a high number of attributes and high dimensions.

In this study, first, M5' model trees algorithm constructs a tree by recursively splitting the instance space. Figure 4 illustrates a tree structure of training procedure corresponding to a given 2-D input parameter domain of x_1 and x_2 . The splitting condition is used to minimize the intra-subset variability in the values down from the root through the branch to the node. The variability is measured by the standard deviation of the values that reach that node from the root through the branch, with calculating the expected reduction in error as a result of testing each attribute at that node. In this way, the attribute (input parameter) that maximizes the expected error reduction is chosen. The splitting process would stop if either the output values of all the instances that reach the node vary slightly or only a few instances (data records) remain. The standard deviation reduction (SDR) is calculated as (Quinlan 1992)

$$\text{SDR} = \text{sd}(T) - \sum_i \frac{|T_i|}{|T|} \times \text{sd}(T_i) \quad (6)$$

where T is the set of examples that reach the node, T_i are the sets that result from splitting the node according to the

Table 2 | Statistical measures of different models for predicting S/Y , laboratory data

Model	I_a	Bias	SI (%)
USDOT (2001), all data	0.92	0.28	61
Breusers et al. (1977), all data	0.64	-0.45	92.8
Melville (1997), all data	0.91	0.245	61
Conventional nonlinear regression	0.85	-0.226	67
M1, all data	0.97	0.042	37
M2, all data	0.95	0.003	49
M2, testing data	0.95	0.004	49

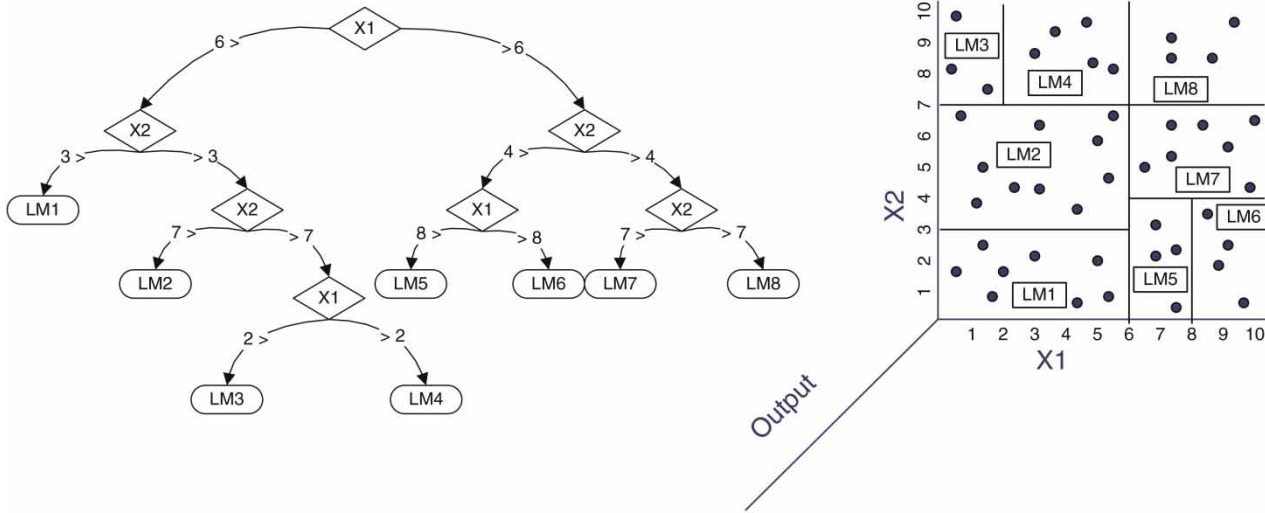


Figure 4 | Example of a M5 model tree (Models 1–7 are linear regression models) (Bonakdar & Etemad-Shahidi 2011).

chosen attribute, and *sd* is the standard deviation (Wang & Witten 1997). After the tree has been grown, a linear multiple regression model is built for every inner node, using the data associated with that node and all the attributes that participate in tests in the sub-tree rooted at that node. Then, linear regression models are simplified by dropping attributes if it results in a lower expected error on future data.

In the second stage, all sub-trees are considered for pruning. Pruning occurs if the estimated error for the linear model at the root of a sub-tree is smaller or equal to the expected error for the sub-tree. After pruning, there is a possibility that the pruned tree might have discontinuities between nearby leaves. Therefore, to compensate discontinuities among adjacent linear models in the leaves of the tree a regularization process is made, which is called smoothing process. In this process, the estimated value of the leaf model is filtered along the path back to the root. At each node, that value is combined with the value predicted by the linear model for that node as follows:

$$P' = \frac{np + kq}{n + k} \tag{7}$$

where *P'* is the prediction passed up to the next higher node, *p* is the prediction passed to this node from below, *q* is the value predicted by the model at this node, *n* is the number of training instances that reach the node below, and *k* is a

constant (Wang & Witten 1997). This process usually improves the prediction, especially for models based on training sets containing a small number of instances (Zhang & Tsai 2007). M5 has been used successfully in prediction of scour around pipelines and pile groups (Etemad-Shahidi & Ghaemi 2011; Yasa & Etemad-Shahidi 2013). The software used in this study was WEKA developed by University of Waikato, New Zealand. After uploading the data set, the required classifier (trees in this case) needs to be selected. In trees classifier, different algorithms are available and M5' was the one chosen in this study. Then, the user can determine the minimum number of instances in each leaf and the percentage of the data set to be used for training the model. The developed model can be validated either by a new set of data or using the so-called cross-validation method.

MODELING, RESULTS, AND DISCUSSION

The success of data mining methods such as M5' depends on the quality and quantity of the used data. In this study, 283 data records from 14 different data sets were used for developing the models. Models based on dimensionless variables have a wider domain of applicability and can be applied to the prototype cases. Hence, the governing input parameters considered in the modeling were the dimensionless ones

mentioned in Equation (2). This ensures the generalization ability of the results. First, a conventional nonlinear multivariate regression model was developed using the data set as a base prediction model, and a single formula was derived (Table 1). Then, the data set was randomly divided into two parts: 70% of them were used for training and the rest were used for testing the M5 model. However, the ranges of parameters used for training were checked to cover those used for testing to guarantee a proper modeling. The ranges of parameters used for the training and testing phases are shown in Table 3. As seen, the used ranges for training are wide and cover both clear water and live bed conditions. The first developed model (hereafter called M1) was based on all the dimensionless parameters of Equation (2). The

comparison between the measured and predicted scour depth using this linear model is presented in Figure 5. As seen, the scatter is less compared to those of previous figures, but the model slightly underestimates high values of scour depth. This could be due to the lack of data records in this range. The error statistics of all models including the existing ones, nonlinear regression model, and developed model trees are given in Table 2, showing the high performance of the M1. In brief, the developed linear model yields accurate results. Nevertheless, the tree and formulae (not shown) made by this model were complex. In total, 11 formulae were generated for different ranges of Y/D , Fr , and Re . The given formulae were mostly linear combinations of Fr , Y/D , and U/U_c and the other variables were either neglected or had small coefficients.

To develop a simpler and more transparent model, a logarithmic transformation was applied to the input and output parameters (Bhattacharya et al. 2007; Etemad-Shahidi et al. 2011). Then, based on the results of the linear, conventional nonlinear regression and previous empirical models, the input parameters were excluded one by one. In this approach, a simple nonlinear but still accurate model was obtained. The inputs of this nonlinear model (hereafter called M2) were U/U_c , Fr , and Y/D . The comparison

Table 3 | Ranges of parameters used for training and testing the models

Parameter	Training data	Testing data
U/U_c	0.32–6.7	0.33–5.3
Y/D	0.05–21.05	0.05–21.05
D/d_{50}	3.65–904.76	13.33–904.76
Fr	0.08–2.14	0.11–1.06
Re	3,408–328,320	6,612–228,000
S/Y	0.02–8.17	0.02–6.65

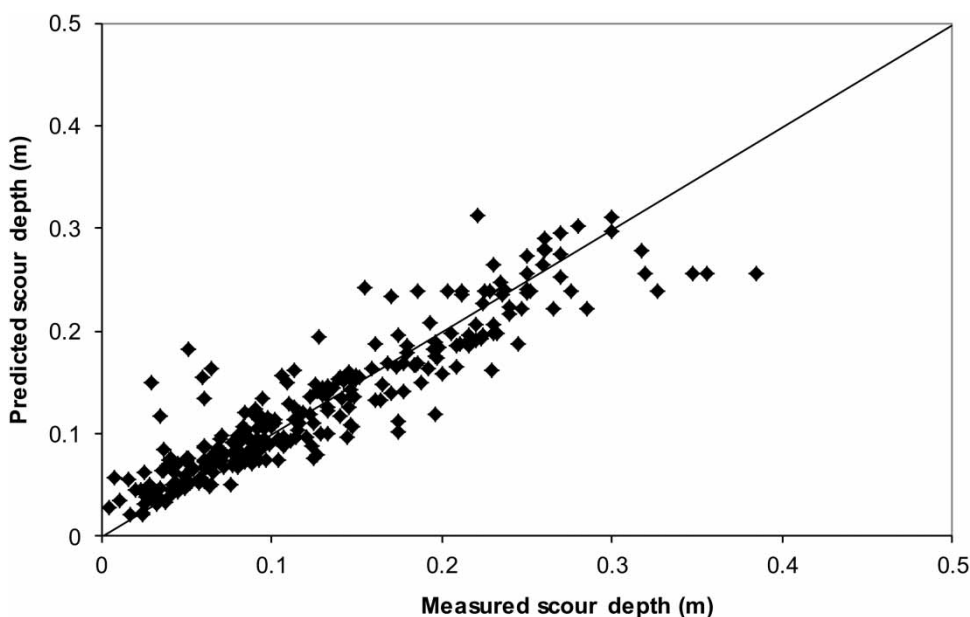


Figure 5 | Comparison between the measured and predicted scour depth by M5' model tree (linear model, M1), all laboratory data.

between the measured and predicted scour depths using this model is presented in Figure 6. As seen, the data points still fall close to the 1:1 line, and the scatter is comparable to that of M1. As shown in the figure, M2 provides better predictions in the region of small scour depth than that of large scour depths. The error statistics of this model for testing data and all data, listed in Table 2, show the skill of this model. It can be concluded that the proper selection and transformation of the input parameters will improve the accuracy and reduce the model's complexity. Compared to M1 with 11 complex formulae, this model yielded only three simple and physically sound ones, i.e.,

$$\text{If } U/U_c < 1 \text{ then } S/Y = 1.84 (Y/D)^{-0.60} (\text{Fr})^{-0.46} (U/U_c)^{0.56} \quad (8a)$$

$$\text{If } U/U_c > 1 \text{ and } Y/D > 2.4 \text{ then } S/Y = 1.88 (Y/D)^{-0.97} (\text{Fr})^{0.13} \quad (8b)$$

$$\text{If } U/U_c > 1 \text{ and } Y/D < 2.4 \text{ then } S/Y = 1.40 (Y/D)^{-0.54} (\text{Fr})^{0.20} \quad (8c)$$

It is apparent from Equation (8) that Y/D , Fr , and U/U_c are the most important dimensionless parameters on the

relative scour depth around piers, while the influences of other parameters such as Reynolds number are marginal. The form of the developed model is similar to those of USDOT (2001) and derived nonlinear regression model. However, it reveals the interaction between hydraulics and sediment transport by considering the critical velocity and relative width of the pier. It is interesting to note that the model tree can distinguish between clear and live bed conditions automatically and show that the scour depth becomes independent of U/U_c in live bed condition which is in line with the findings of Melville (1997). In addition, M5 successfully yields a different formula for wide piers, and the splitting value is very close to the one used for wide piers (Johnson 1995; Jones & Sheppard 2000).

In terms of dimensional parameters, Equation (8b) implies that in live bed conditions, the scour depth is linearly related to the pier diameter and is independent of water depth in relatively deep waters. On the other hand, Equation (8c) shows that in the case of relatively shallow water and live bed condition, the scour depth depends on the water depth as well. Both these results are in line with the existing knowledge of physics of the scour process.

In summary, it can be inferred that the nonlinear M5' model has succeeded in capturing the relationship among the scour governing parameters. Another advantage of M5'

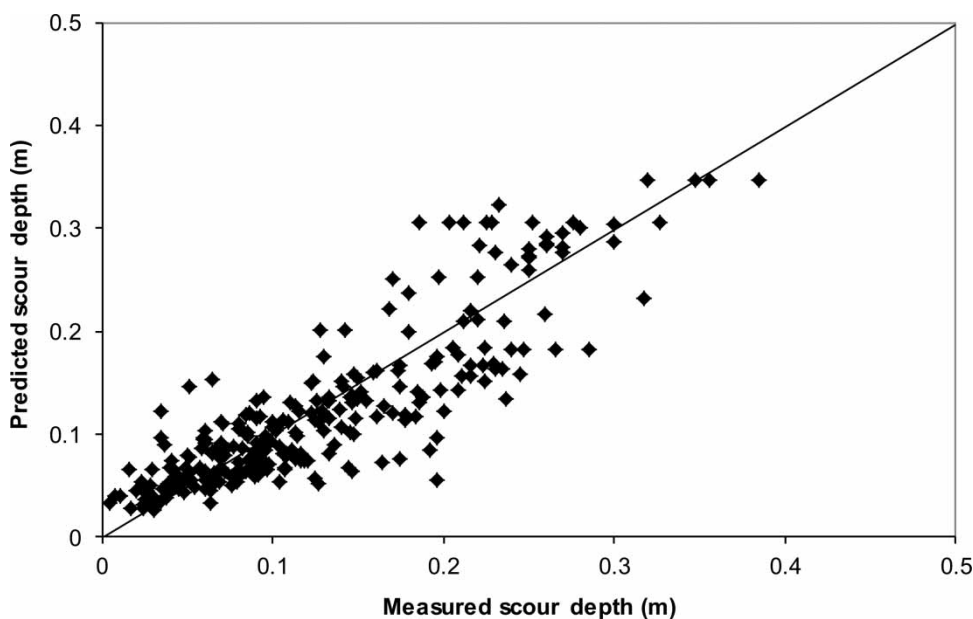


Figure 6 | Comparison between the measured and predicted scour depth by M5' model tree (M2), all laboratory data.

was that it yielded a physically sound and simple equation relating the input variables to the output. This is not the case with traditional data mining methods such as ANN. The performance of Equation (8) was superior to those of other methods while that of Melville (1997) outperformed other existing formulae. Among other data mining approaches, group method of data handling (GMDH) can also be used to provide formulae for scour depth around piers. GMDH, which is based on the principles of heuristic self-organizing, can be improved by a GMDH-back propagation method (GMDH-BP) or other evolutionary algorithm. However, the formulae developed by this method are very complex (e.g., Najafzadeh et al. 2013) and hard to be physically justified. The application of GMDH-BP requires accurate determination of several parameters, such as topology of network, weightings, and operations; while using M5 the only parameter that needs to be determined is the minimum number of data sets in each leaf. In addition, the execution of heuristics models generally is computationally expensive while executing a M5 model usually takes a couple of seconds.

APPLICATION TO THE FIELD MEASUREMENTS

Field data were also used to evaluate the performance of different models. The field data were obtained from the study of Sheppard et al. (2011). This data set contains 791 good quality field equilibrium local scour data points. A total of 71 field data sets were selected and used to evaluate the performance of different formulae. All these data were for single, circular piers founded in non-cohesive sediments. The error statistics of different models are given in Table 4. As seen, even in this case, the developed model outperforms other formulae in

predicting the scour depth. Compared to Table 2, the 'Bias' of M2 has increased significantly. This is mainly because the maturity of the scour depth is not known in the field during measurements which results in a larger 'Bias' for most of the models. In addition, the conditions in the field are not ideal, and therefore the measurements could be less accurate compared to those of laboratory experiments. This is in line with the findings of Landers et al. (1999). They evaluated formulae developed in the laboratory by use of transformed data and smoothing techniques to assess general trends in the data. They found only minimal agreement between the field data and laboratory-based relationships. Similar results were obtained by Pal et al. (2012), and they also found that the existing formulae may not be suitable for application in the field.

One of the limitations of the present model is that its application is limited to the range of used parameters and cannot be directly used to analyze complexities such as pier geometry and armoring by bed materials. In addition, most of the data points used for developing the formulae were obtained from experiments in the clear water critical conditions, and therefore Equation (8a) is statistically more significant than the others.

SUMMARY AND CONCLUSION

In this study, 14 different laboratory data sets with a wide range of variables were used to develop a model for prediction of the current-induced scour depth around circular piers. Since the selection of input variables is very important for the model's accuracy, all governing dimensionless parameters were first used as the inputs of the model and an accurate but complex model was developed. Then, to establish a simpler model, an appropriate transformation of governing parameters was used. In this way, a simple model was obtained for estimation of relative scour depth based on the Froude number, the relative water depth, and relative flow velocity. The obtained formulae were transparent and compact and also revealed the physics of phenomena by distinguishing between different regimes, goals which are rarely achieved by other data mining methods. Drawing out the physics and knowledge from data mining models is as important as their accuracy.

Table 4 | Statistical measures of different models for predicting S/Y , field data

Model	I_a	Bias	SI (%)
USDOT (2001)	0.66	0.44	245
Breusers et al. (1977)	0.71	0.40	73.1
Melville (1997)	0.47	0.81	128
M2	0.88	0.13	46.0

Using the statistical measures, it was shown that the obtained model is superior to the existing empirical approaches using both laboratory and field measurements.

The used approach is very promising considering the time savings in both the development and run-time of the model tree compared with those of other AI-based approaches such as ANN, SVM, GMDH-BP, and especially genetic programming. The appropriate transformation of the governing parameters combined with using rule-based models such as M5 provide an alternative and quick solution to provide compact and transparent design formulae with reasonable accuracy.

ACKNOWLEDGEMENT

We would like to thank the University of Waikato, New Zealand for providing WEKA software (<http://www.cs.waikato.ac.nz/~ml/weka/>).

REFERENCES

- Akib, S., Mohammadhassani, M. & Jahangirzadeh, A. 2014 Application of ANFIS and LR in prediction of scour depth in bridges. *Comp. Fluids* **91** (3), 77–86.
- Ataie-Ashtiani, B. & Beheshti, A. 2006 Experimental investigation of clear-water local scour at pile groups. *J. Hydraul. Eng. ASCE* **138** (3), 177–185.
- Ayoubloo, M. K., Etemad-Shahidi, A. & Mahjoobi, J. 2010 Evaluation of regular wave scour around a circular pile using data mining approaches. *Appl. Ocean Res.* **32**, 34–39.
- Azmathulla, H. M. 2012 Gene-expression programming to predict scour at a bridge abutment. *J. Hydroinf.* **14** (2), 324–331.
- Azmathulla, H. M. & Ghani, A. 2010 Genetic programming to predict river pipeline scour. *J. Pipeline Syst. Eng. Pract.* **1** (3), 127–132.
- Azmathulla, H., Deo, M. C. & Deolalikar, P. B. 2005 Neural networks for estimation of scour downstream of a ski-jump bucket. *J. Hydraul. Eng. ASCE* **131** (10), 898–908.
- Azmathulla, H., Ghani, A., Zakaria, N. & Guven, A. 2010 Genetic programming to predict bridge pier scour. *J. Hydraul. Eng. ASCE* **136** (3), 1–6.
- 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**, 401–414.
- Bhattacharya, B. & Solomatine, D. P. 2005 Neural networks and M5 model trees in modelling water level-discharge relationship. *Neurocomputing* **63**, 381–396.
- Bhattacharya, B., Price, R. K. & Solomatine, D. P. 2007 Machine learning approach to modeling sediment transport. *J. Hydraul. Eng. ASCE* **133** (4), 440–450.
- Blodgett, B. A. 1978 *Countermeasures for Hydraulic Problems at Bridges. Vol. 1*, Federal Washington DC, HWY Administration US Department of Transportation, pp. 10–23.
- Bonakdar, L. & Etemad-Shahidi, A. 2011 Predicting wave run-up on rubble-mound structures using M5 model tree. *Ocean Eng.* **38**, 111–118.
- Breusers, H. N. C. & Raudkivi, A. J. 1991 *Scouring*. A. A. Balkema, Rotterdam.
- Breusers, H., Nicollet, G. & Shen, H. 1977 Local scour around cylindrical piers. *J. Hydraul. Eng. ASCE* **15** (3), 211–252.
- Chabert, J. & Engeldinger, P. 1956 *Etude Des Affouillements Autour Des Piles Des Ponts*. Laboratoire de Hydraulique, Chatou, France (in French).
- Chee, R. K. W. 1982 *Live-bed Scour at Bridge Piers*. Report No. 290, School of Engineering, The University of Auckland, New Zealand.
- Chiew, Y. M. 1984 *Local Scour at Bridge Piers*. Report No. 355, School of Engineering, The University of Auckland, Auckland, New Zealand.
- Daga, M. & Deo, M. C. 2009 Alternative data-driven methods to estimate wind from waves by inverse modeling. *Nat. Hazard.* **49**, 293–310.
- Etemad-Shahidi, A. & Bali, B. 2011 Stability of rubble-mound breakwater using H50 wave height parameter. *Coast. Eng.* **59**, 38–45.
- Etemad-Shahidi, A. & Bonakdar, L. 2009 Design of rubble-mound breakwaters using M5' machine learning method. *Appl. Ocean Res.* **31**, 197–201.
- Etemad-Shahidi, A. & Ghaemi, N. 2011 Model tree approach for prediction of pile groups scour due to waves. *Ocean Eng.* **38**, 522–527.
- Etemad-Shahidi, A. & Jafari, E. 2014 New formulas for prediction of wave overtopping at inclined structures with smooth impermeable surface. *Ocean Eng.* **84**, 124–132.
- Etemad-Shahidi, A. & Mahjoobi, J. 2009 Comparison between M5' model tree and neural networks for prediction of significant wave height. *Ocean Eng.* **36**, 1175–1181.
- Etemad-Shahidi, A. & Rohani, M. S. 2014 Prediction of scour at abutments using piecewise regression. *Proc. ICE Water Manage.* **167**, 79–87.
- Etemad-Shahidi, A. & Taghipour, M. 2012 Predicting longitudinal dispersion coefficient in natural streams using M5' model tree. *ASCE J. Hydraul. Eng.* **138**, 542–554.
- Etemad-Shahidi, A., Yasa, R. & Kazeminezhad, M. K. 2011 Prediction of wave-induced scour depth under submarine pipelines using machine learning approach. *Appl. Ocean Res.* **33**, 54–59.
- Ettema, R. 1980 *Scour at Bridge Piers*. Report No. 216, University of Auckland, Auckland, New Zealand.
- Ettema, R., Melville, B. W. & Barkdoll, B. 1998 Scale effect in pier-scour experiments. *J. Hydraul. Eng.* **124**, 639–642.

- Farhoudi, J., Hosseini, S. M. & Sedghi-Asl, M. 2010 [Application of neuro-fuzzy model to estimate the characteristics of local scour downstream of stilling basins](#). *J. Hydroinf.* **12** (2), 201–211.
- FDOT 2010 *Bridge Scour Manual*. Hydraulic Engineering Circular, Florida Department of Transportation, Florida.
- Gaudio, R., Tafarajnoruz, A. & De Bartolo, S. 2013 [Sensitivity analysis of bridge pier scour depth predictive formulae](#). *J. Hydroinf.* **15** (3), 939–951.
- Ghaemi, N., Etemad-Shahidi, A. & Ataie-Ashtiani, B. 2013 [Estimation of current-induced pile groups scour using a rule based method](#). *J. Hydroinf.* **15**, 516–528.
- Ghazanfari, S., Etemad-Shahidi, A., Kazeminezhad, M. H. & Mansoori, A. R. 2011 [Prediction of pile groups scour in waves using support vector machines and ANN](#). *J. Hydroinf.* **13**, 609–620.
- Graf, W. H. 1995 *Load Scour Around Piers*. Annual Report, Laboratoire de Recherches, Hydriques, Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland, pp. B.33.1–B.33.8.
- Gulbahar, N. 2009 [Case study: comparing performance of pier scour equations using field data for the Gediz bridge, Manisa, Turkey](#). *Can. J. Civil Eng.* **36**, 801–812.
- Hancu, S. 1971 [Sur le calcul des affouillements locaux dams la zone des piles des ponts](#). In *Proceedings of the 14th IAHR Congress*, Vol. 3, Paris, France, pp. 299–313.
- Hand, D., Heikki, M. & Padhraic, S. 2001 *Principles of Data Mining*. The MIT Press, Cambridge, MA, p. 546.
- Jafari, E. & Etemad-Shahidi, A. 2012 [Derivation of a new model for prediction of wave overtopping at rubble-mound structures](#). *ASCE J. Waterway Port Coast. Eng.* **138**, 42–52.
- Jain, S. C. & Fischer, E. E. 1980 [Scour around bridge piers at high flow velocities](#). *J. Hydraul. Eng. ASCE* **106** (11), 1827–1842.
- Johnson, P. 1995 [Comparison of pier-scour equations using field data](#). *J. Hydraul. Eng. ASCE* San Francisco, CA, **121** (8), 626–629.
- Jones, J. & Sheppard, D. 2000 [Scour at wide bridge piers](#). In *Joint Conference on Water Resource Engineering and Water Resources Planning and Management 2000*, July 30–August 2, ASCE, Minneapolis, Minnesota, pp. 1–10.
- Kantardzic, M. 2003 *Data Mining: Concepts, Models, Methods, and Algorithms*. Wiley, New York, p. 343.
- Karami, H., Ardeshtir, A., Saneie, M. & Salamatian, S. A. 2012 [Prediction of time variation of scour depth around spur dikes using neural networks](#). *J. Hydroinf.* **14** (1), 180–191.
- Kazeminezhad, M. H., Etemad-Shahidi, A. & Yeganeh-Bakhtiary, A. 2010 [An alternative approach for investigation of wave-induced scour around pipelines](#). *J. Hydroinf.* **12**, 51–65.
- Khan, M., Azamathulla, H. M. & Tufail, M. 2012 [Gene-expression programming to predict pier scour depth using laboratory data](#). *J. Hydroinf.* **14** (3), 628–645.
- Kothiyari, U. C., Grade, R. J. & Ranga Raju, K. G. 1992 [Temporal variation of scour around circular bridge piers](#). *J. Hydraul. Eng. ASCE* **118** (8), 1091–1106.
- Landers, M., Mueller, D. & Richardson, E. 1999 [U.S. geological surveys of field measurements of pier scour](#). In: *Stream Stability and Scour at Highway Bridges* (E. V. Richardson & P. F. Lagasse, eds), ASCE, San Francisco, CA, pp. 585–607.
- Lee, S. O. & Sturm, T. W. 2009 [Effects of sediment size scaling on physical modeling of bridge pier scour](#). *J. Hydraul. Eng. ASCE* **135**, 793–802.
- Liriano, S. L. & Day, R. A. 2001 [Prediction of scour depth at culvert outlets using neural networks](#). *J. Hydroinf.* **3**, 231–238.
- Melville, B. W. 1997 [Pier and abutment scour: integrated approach](#). *J. Hydraul. Eng. ASCE* **123**, 125–136.
- Melville, B. W. & Chiew, Y. M. 1999 [Time scale for local scour depth at bridge piers](#). *J. Hydraul. Eng. ASCE* **125** (1), 59–65.
- Mueller, D. S. & Wagner, C. R. 2005 [Field observations and evaluations of streambed scour at bridges](#). Federal Highway Administration Report FHWA-RD-03-052, Washington, DC.
- Muzzammil, M. 2010 [ANFIS approach to the scour depth prediction at a bridge abutment](#). *J. Hydroinf.* **12** (4), 474–485.
- Muzzammil, M. & Alam, J. 2011 [ANFIS-based approach to scour depth prediction at abutments in armored beds](#). *J. Hydroinf.* **13** (4), 699–713.
- Muzzammil, M. & Ayyub, M. 2010 [ANFIS-based approach for scour depth prediction at piers in non-uniform sediments](#). *J. Hydroinf.* **12** (3), 303–317.
- Najafzadeh, M., Barani, G. A. & Azamathulla, H. 2013 [GMDH to prediction of scour depth around vertical piers in cohesive soils](#). *Appl. Ocean. Res.* **40**, 35–41.
- Oliveto, G. & Hager, W. H. 2002 [Temporal evolution of clear-water pier and abutment scour](#). *J. Hydraul. Eng. ASCE* **128** (9), 811–820.
- Ould-Ahmed-Vall, E., Woodlee, J., Yount, C., Doshi, K. A. & Abraham, S. 2007 [Using model trees for computer architecture performance analysis of software applications](#). In *IEEE International Symposium on Performance Analysis of Systems and Software*, 25–27 April, San Jose, CA.
- Pal, M. 2006 [M5 model tree for land cover classification](#). *Int. J. Remote Sens.* **27** (4), 825–831.
- Pal, M. & Deswal, S. 2009 [M5 model tree based modelling of reference evapotranspiration](#). *Hydrol. Process.* **30** (10), 1437–1443.
- Pal, M., Singh, N. K. & Tiwari, N. K. 2012 [M5 Model tree for pier scour prediction using field dataset](#). *KSCE J. Civil Eng.* **16** (6), 1079–1084.
- Pal, M., Singh, N. K. & Tiwari, N. K. 2013 [Kernel methods for pier scour modeling using field data](#). *J. Hydroinf.* **16** (4), 784–796.
- Quinlan, J. R. 1992 [Learning with continuous classes](#). In *Proceedings of 5th Australian Joint Conference on Artificial Intelligence*, World Scientific, Singapore, pp. 343–348.
- Shen, H. W. (ed.) 1971 [Scour near piers](#). In *River Mechanics*. Colorado State University, Ft. Collins, CO.
- Shen, H., Schneider, V. & Karaki, S. 1966 *Mechanics of Local Scour*. U.S. Department of Commerce, National Bureau of Standards, Institute for Applied Technology, Boulder, CO.
- Shen, H., Schneider, V. & Karaki, S. 1969 [Local scour around bridge piers](#). *J. Hydraul. Div. ASCE* **95**, 1919–1940.
- Sheppard, D. & Miller, W. 2006 [Live-bed local scour experiments](#). *J. Hydraul. Eng. ASCE* **132** (7), 635–642.

- Sheppard, D. M., Demir, H. & Melville, B. 2011 *Scour at wide piers and long skewed piers*. National Cooperative Highway Research Program Report 682, p. 55.
- Toth, E. & Brandimarte, L. 2011 *Prediction of local scour depth at bridge piers under clear-water and live-bed conditions: comparison of literature formulae and artificial neural networks*. *J. Hydroinf.* **13** (4), 812–824.
- USDOT 2001 *Evaluating Scour at Bridges*. Hydraulic Engineering Circular, Federation of Hwy, Administration, US Department of Transportation, McLean, VA.
- Wang, Y. & Witten, I. H. 1997 Induction of model trees for predicting continuous classes. In *Proceedings of the Poster Papers of the European Conference on Machine Learning*, University of Economics, Faculty of Informatics and Statistics, Prague.
- Yanmaz, A. M. & Altinbilek, H. D. 1991 *Study of time-dependent local scour bridge piers*. *J. Hydraul. Eng. ASCE* **117** (10), 1247–1268.
- Yasa, R. & Etemad-Shahidi, A. 2013 *Classification and regression trees approach for predicting current-induced scour depth under pipelines*. *J. Offshore Mech. Arctic Eng.* **136**, 1–8.
- Zanganeh, M., Mousavi, S. J. & Etemad-Shahidi, A. 2009 *A hybrid genetic algorithm–adaptive network-based fuzzy inference system in prediction of wave parameters*. *Eng. Appl. Artif. Intell.* **22**, 1194–1202.
- Zanganeh, M., Yeganeh-Bakhtiary, A. & Bakhtyar, R. 2011 *Combined particle swarm optimization and fuzzy inference system model for estimation of current-induced scour beneath marine pipelines*. *J. Hydroinf.* **13** (3), 558–573.
- Zhang, D. & Tsai, J. J. P. 2007 *Advances in Machine Learning Applications in Software Engineering*. Idea Group Publishing, London.

First received 30 March 2014; accepted in revised form 5 October 2014. Available online 31 October 2014