PREDICTION OF BRIDGE PIER SCOUR USING GENETIC PROGRAMMING

Chuan-Yi Wang1, Han-Peng Shih2, Jian-Hao Hong3, and Rajkumar V. Raikar4

Key words: genetic programming, pier scour, clear-water conditions, traditional equations.

ABSTRACT

This paper presents the use of genetic programming (GP) as a tool to predict pier scour depths based on clear-water conditions of laboratory measurements by past researchers. Four main dimensionless parameters—pier width, approaching flow depth, threshold flow velocity, and channel open-ratio—are considered for predicting the scour depth. The performance of the GP equation is verified by comparing the results with those obtained by empirical equations. It is found that the scour depth at bridge piers can be efficiently predicted using the GP model. The advantage of the GP model is confirmed by comparing the GP results of scour depths with the large-scale model studies and field data.

I. INTRODUCTION

Pier scour has attracted significant research interest for more than a century now, and numerous studies on this subject have been published. Much of this research deals with laboratory model studies of local scour. Appropriately predicting scour depth at bridge piers is a concern to bridge engineers because the underestimation of scour depth will cause the structure to be at high risk from damage while overestimation of scour depth makes the design uneconomical. In this context, several reviews have summarized equations for pier scour depths [9, 10, 13, 14, 22, 30, 31]. However, these equations are often suitable only for conditions similar to those under which they were developed.

Soft computing tools are also gaining importance in many fields as they differ from conventional hard computing in many ways, such as their robustness, and low solution cost, and tolerance to imprecision. Alternative soft computing tools have been widely applied in solving scour problems. Predictive approaches, such as artificial neural networks (ANNs), adaptive neuro-fuzzy inference systems (ANFIS), genetic programming (GP), and linear genetic programming (LGP), have recently been shown to effectively predict scour around hydraulic structures. The American Society of Civil Engineers Task Committee [1, 2] reports the application of ANNs in different fields of hydrology. Deo et al. [12] used GP to predict scour depth downstream of spillways. Azamathulla et al. [3, 4, 7] used neural networks and GP to determine scour depth downstream of ski-jump buckets. Guven et al. [18] applied LGP for predicting scour depth at circular piles. ANFIS and genetic expression programming were used by Azamathulla et al. [6] to estimate scour below flip buckets. To predict scour depth at bridge piers, Azamathulla et al. [5] used the ANN and GP, and they reported that the performance of GP was more effective than that of regression-based models and ANNs. For scour below a submerged pipeline, Azamathulla et al. [8] employed the LGP model. Najafzadeh and Barani [33] compared the group method of data handling-based GP and the back-propagation system to predict scour depth around bridge piers.

In this context, the present study emphasizes the use of GP to establish a relationship between the estimation of the maximum scour depth at bridge piers under clear-water scour conditions and results of uniform sediment particles obtained using laboratory and field data. Further, a comparison between the GP results and traditional equations and unsteady models is presented. The applicability of the GP model to large-scale models and field data is also verified.

II. ANALYSIS OF PIER SCOUR DEPTH PARAMETERS

The scour process at bridge piers and the maximum pier scour depth at piers $d_s$, are affected by a large number of interdependent variables, namely, the characteristics of flow, fluid, sediment, pier, channel, and time [11, 13, 30]. The general relation between $d_s$ and its dependent parameters can be written as [35]
where $V = \text{average approach flow velocity}; \ y = \text{approach flow depth}; \ \rho = \text{mass density of water}; \ \rho_s = \text{mass density of sediment}; \ v = \text{kinematic viscosity of water}; \ D = \text{pier width or diameter}; \ d = \text{median sediment size}; \ \sigma = \text{geometric standard deviation of particle size distribution}; \ t_e = \text{time to develop equilibrium scour depth}; \ \alpha = \text{opening ratio } \approx (B - D)/B; \ \text{and } B = \text{channel width.}

Following the previous reviews [13], Eq. (2) can be expressed as

$$d_s = f_t(V, y, \rho, \rho_s, g, V_D, d, \sigma, t_e, \alpha) \tag{1}$$

where $\Delta = (\sigma - d)/d$ is excess pier Reynolds number. However, the influence of the Reynolds number $Re$ is insignificant for a turbulent flow over rough beds [31]. Considering uniform sediments, the pier scour depth can be written as

$$d_s = f_t\left\{\frac{V^2 - V_D^2}{\Delta d}, \frac{V}{\Delta d}, \frac{d}{d}, \sigma, t_e, \alpha\right\} \tag{2}$$

$$d_s = f_t\left\{\frac{V^2 - V_D^2}{\Delta d}, \frac{V}{\Delta d}, \frac{d}{d}, \sigma, t_e, \alpha\right\} \tag{3}$$

This study uses GP that has a structure consisting of nonlinear functions and a parameter identification process based on techniques that search for global maxima in the space of feasible parameter values. This study focuses on the use of GP to establish a relationship between the estimation of the maximum scour depth at bridge piers under clear-water scour conditions and results of uniform sediment particles obtained using the experimental data of past researchers.

Several traditional equations have been selected to predict the pier scour depth, e.g., those of Shen et al. [36], Hancu [21], Gao et al. [17], Melville [28], Oliveto and Hager [34], and Kothyari et al. [25] (Table 1). The results of these equations are compared with those obtained by GP.

### III. DATABASE USED

The experimental results of the laboratory study were used in training and testing sets of the proposed GP model. The datasets that were used were collected from the studies of Chabert and Engeldinger [11], Verstappen [38], Walker [39], Ettema [16], Kothyari [23], and Melville and Chiew [29]. Table 2 (a) presents the ranges of various parameters available, such as pier width or diameter (D), approaching flow depth (y), average approaching flow velocity (V), median sediment size (d or d₅₀), channel open ratio (α₀), equilibrium pier scour depth (dₑ), dimensionless pier width (Ddₑ), dimensionless approaching flow velocity (y/dₑ), and dimensionless threshold flow velocity ((V² - V_D²)/Vₕ), with median sediment size (d) and dimensionless pier scour depth (d_s/D). From the total 130 test sets, 105 sets were

### Table 1. Pier scour equations used in present study.

<table>
<thead>
<tr>
<th>Investigator</th>
<th>Proposed equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shen et al. [36]</td>
<td>$d_s = 0.000223 \left(\frac{V}{D} \right)^{0.695}$</td>
</tr>
<tr>
<td>Hancu [21]</td>
<td>$d_s = 2.42\left(\frac{2V}{V_{th}} - 1\right)\left(\frac{V_{th}}{gD}\right)^{0.5}$ for $0.05 \leq \frac{V_{th}}{gD} \leq 0.6$</td>
</tr>
<tr>
<td>Kothyari et al. [24]</td>
<td>$d_s = 0.66 \left(\frac{D}{d}\right)^{2.23} \left(\frac{y}{d}\right)^{0.18} \left(\frac{V^2 - V_{c1}^2}{\Delta d}\right)^{0.5}$</td>
</tr>
<tr>
<td>Gao et al. [17]</td>
<td>$d_s = 0.46K_sD^{0.6} \left(\frac{V - V_{c1}}{V_{c1} - V_{c2}}\right)^7$ where $K_s = \text{shape and alignment factor}; \ V_{c1} = \text{critical velocity of the approaching flow,}$ $17.6\Delta d + 6.05 \times 10^{5}\left(\frac{10 + y}{d^0.5}\right)^{0.3}$; $V_{c2} = \text{incipient velocity for local scour at a pier,}$ $0.645(d/D)^{0.05}V_{c1}; \ \text{and } \eta = (V_{c1}/V)^{0.85 + 2.22\log \phi}.$ For clear-water scour ($V &lt; V_{c1}, \eta &lt; 1$) and for live-bed scour ($V &gt; V_{c1}, \eta &gt; 1$)</td>
</tr>
<tr>
<td>Melville [28]</td>
<td>$d_s = K_{s0}K_sK_sK_sK_sK_s$ where $K_{s0} = \text{flow depth - foundation size factor}; \ K_s = \text{flow intensity factor}; \ K_s = \text{sediment size factor}; \ K_s = \text{shape factor (pier or abutment)}; \ K_s = \text{pier or abutment alignment factor}; \ \text{and } K_s = \text{channel geometry factor}$</td>
</tr>
<tr>
<td>Oliveto and Hager [34]</td>
<td>$d_s = 0.068K_s\sigma_{ds}^2F_d^{1.2} \log T$ where $d_s = \text{temporal variation of scour depth under steady flow conditions}; \ \eta = \text{reference length obtained as } (yD^3)^{1/3}; \ F_d = \text{dissimetric Froude number, } V/(\Delta d_d)^{1/3}; \ \text{R} = \text{relative time, } t/\eta; \ \text{and } \eta = \text{time scale, } \eta = \sigma_{ds}/(\Delta d_d)^{1/3}$</td>
</tr>
<tr>
<td>Kothyari et al. [25]</td>
<td>$d_s = 0.272\sigma_{ds}^2(F_d - F_{ds})^{2/3} \log T$ where $F_d = \text{dissimetric Froude number for inception of scour at pier given by } {F_d - 1.26\beta^2(F_d - 1)^{0.16}\sigma_{ds}^2; \ F_{ds} = \text{dissimetric Froude number for inception of sediment in approaching flow, } V/(\Delta d_d)^{1/3}; \ R_h = \text{hydraulic radius}; \ \beta = \text{element obstruction } = DI/B$</td>
</tr>
</tbody>
</table>
Table 2. Ranges of database used for training and verification.

<table>
<thead>
<tr>
<th>Variables</th>
<th>(a) Data used in training and testing sets for GP</th>
<th>(b) Data of Dey [15], Sheppard et al. [37] and Raikar [35] used for verification</th>
<th>(c) Data of Federal Highway Administration (FHWA) used for verification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pier width or diameter (D)</td>
<td>0.0285-0.24 (m)</td>
<td>0.032-0.91 (m)</td>
<td>0.51-4.57 (m)</td>
</tr>
<tr>
<td>Approaching flow depth (y)</td>
<td>0.02-0.7 (m)</td>
<td>0.035-1.9 (m)</td>
<td>0.073-5.33 (m)</td>
</tr>
<tr>
<td>Average approaching flow velocity (V)</td>
<td>0.171-1.27 (m/s)</td>
<td>0.172-1.102 (m/s)</td>
<td>0.17-3.505 (m/s)</td>
</tr>
<tr>
<td>Median sediment size (d or d₅₀)</td>
<td>0.0002-0.0078 (m)</td>
<td>0.0022-0.01425 (m)</td>
<td>0.00075-0.09 (m)</td>
</tr>
<tr>
<td>Channel open-ratio (α)</td>
<td>0.81-0.98</td>
<td>0.85-0.98</td>
<td>0.88-0.926</td>
</tr>
<tr>
<td>Equilibrium pier scour depth (dₑ)</td>
<td>0.11-0.251 (m)</td>
<td>0.0193-1.27 (m)</td>
<td>0.024-1.68 (m)</td>
</tr>
<tr>
<td>Dimensionless pier width (D/d)</td>
<td>3.65-760</td>
<td>2.25-4136</td>
<td>10.53-237</td>
</tr>
<tr>
<td>Dimensionless approaching flow depth (y/d)</td>
<td>9.35-2500</td>
<td>5.61-8227</td>
<td>13.11-451</td>
</tr>
<tr>
<td>Dimensionless threshold flow velocity ((V²- Vc²)/(Δgd))</td>
<td>0.12-30.02</td>
<td>1.55-42.78</td>
<td>0.41-21.32</td>
</tr>
<tr>
<td>Dimensionless pier scour depth (dₑ/D)</td>
<td>0.22-2.47</td>
<td>0.6-2.12</td>
<td>0.33-1.71</td>
</tr>
</tbody>
</table>

selected randomly for the training, and the remaining 25 sets were used for validating the proposed GP model.

From preceding research, it is known that the use of grouped, dimensionless parameters gave better predictions than the use of dimensional parameters [3, 5, 18]. The functional relationships used in the present models are given in Eq. (3).

IV. GP AND MODEL DEVELOPMENT

GP is an extension of John Holland’s genetic algorithms (GAs) proposed by Koza [26]. The major difference between GP and GAs is that the variable parse tree structure of GP replaces the fixed gene structure of GAs. The parse tree structures undergoing adaptation are hierarchical computer programs of dynamically varying sizes and shapes in GP. Therefore, GP typically incorporates a domain specific syntax that governs acceptable (or meaningful) arrangements of information on the chromosome and makes use of genetic operators that preserve the syntax of its tree-structure chromosomes during reproduction. The search space in GP is the space of all computer programs that includes functions and terminals appropriate to the problem domain. The function set consists of all kinds of functions and the terminal set consists of all kinds of terminals defined by the developers.

Based on the natural selection obtained by way of the evolutionary process, GP produces an optimal function set (formula). It is important to mention that GP requires the typical functional relationship to be given by the user, which may be of nonlinear form. The use of this flexible coding system allows the algorithm to perform structural optimization. This can be useful in solving many engineering problems.

In the development of the GP model, the terminal set, functional set, fitness function, algorithm control parameters, and termination criterion are defined [26]. The first three components determine the algorithm search space, whereas the last two components affect the quality and speed of search.

The function set generally consists of eight basic arithmetic operators (+, −, ×, ÷, √, log, exp, power) and constants. The terminal set includes four fundamental groups of hydraulic parameters of dₑ, as expressed in Eq. (4) [24]. The terminal set encompasses the nondimensional relationships obtained from the variables influencing dₑ, as expressed by Eq. (3). It is relevant to mention that these four nondimensional variables have a significant effect on the pier scour as reported by previous researchers.

\[
\frac{d_e}{D} = f \left( \frac{D}{d}, \frac{y}{d}, \frac{V^2 - V_c^2}{\Delta g d}, \alpha \right) = f \left( X_1, X_2, X_3, X_4 \right) \tag{4}
\]

In this study, the operations of crossover and mutation were selected as 0.4-0.8 and 0.01-0.1, respectively. The population size considered was 500-1000 members. The total number of generations was 1000-8000, and the maximum depth of the parse tree structure was allowed during 15-20 generations. The restriction in the maximum depth of the parse tree structure is aimed at achieving a balanced accuracy of the solutions and the parsimony problem in GP. The parsimony problem indicates the diverging growth of population size without an associated increase in fitness during the process of obtaining best-fit optima. However, this increase in population size will not assist in improving generalization ability [40]. The fitness function was the sum of absolute differences (\( SAD = \sum |P_i - m_i| \)) between the measured values and the estimated values present in the database. The
Fig. 1. Comparison between predicted values of $d_s/D$ by GP and measured values for (a) train sets and (b) test sets.

The optimum result found through the GP development and program was obtained when the population size was 600 members with a total of 5000 generations having crossover 0.6 and mutation 0.04. The prediction of the proposed GP formula versus the actual experimental values for the training and testing sets is shown in Figs. 1(a) and (b). It is a common result that the predictions of training sets are slightly better than the results for the testing sets.

These figures show that the proposed GP formula can learn very well the nonlinear relationship between parameters and also provide high generalization capacity. The generated prediction formula of GP is as follows:

$$d_s/D = \left[ \left( \frac{X_1}{X_4} \right) \left( \frac{X_3}{X_4} \right) \right]^{0.110} \left( \frac{X_1}{X_4} \right)^{0.200} \frac{X_1}{4.94} \left( \frac{X_1}{X_4} \right)^{0.156}$$  \hspace{1cm} (5)

The comparison of $d_s/D$ predicted using Eq. (5) with that predicted using empirical equations proposed by various researchers (Table 1) is presented in Figs. 2 and 3 and in Table 3. For this comparison, the experimental data of Chabert and Engeldinger [11], Verstappen [38], Walker [39], Ettema [16], Kothyari [23], and Melville and Chiew [29] given in Table 2 (a) are used. The performance of these formulas is validated in terms of the common statistical measures of the root mean square error (RMSE), mean absolute percentage error (MAPE), and correlation coefficient ($R$):

$$RMSE = \sqrt{\frac{1}{N} \sum (m_i - p_i)^2}$$  \hspace{1cm} (6)

$$MAPE = \frac{1}{N} \sum \left( \frac{m_i - p_i}{m_i} \right) \times 100$$  \hspace{1cm} (7)

$$R = \frac{\sum (m_i - \bar{m})(p_i - \bar{p})}{\sqrt{\sum (m_i - \bar{m})^2 \sum (p_i - \bar{p})^2}}$$  \hspace{1cm} (8)

where $N$ = number of total data items; $m_i$ = measured value; $p_i$ = predicted value; and $\bar{m}$ and $\bar{p}$ = means of measured and predicted values, respectively.

V. COMPARATIVE ANALYSIS

Table 3 (a) lists the statistics results such as the RMSE, MAPE, and $R$ of these formulas including all data ranges. The results indicate that the GP model (Eq. 5) has a superior performance to the empirical pier scour equations furnished in Table 1 for all the experimental data considered. The values of RMSE, MAPE, and $R$ for the proposed GP formula considering all data [Table 2 (a)] are 0.28, 19.8%, and 0.84, respectively, which are better than those of other equations in this study [Table 3 (a)]. The equations of Shen et al. [36], Hancu [21], and Oliveto and Hager [34] resulted in larger errors than did the other equations [Table 3 (a)].

A comparison between the proposed GP equation (Eq. (5)) and all other pier scour equations (Table 1) for different ranges of $D/d$, $y/D$, and $V/V_c$ was carried out, and the results are presented in Table 4. It is possible to determine which equations are most useful under various conditions. In this table, $V_c$ was determined using the semilogarithmic average velocity equation [28]. An excellent prediction performance of the GP can be observed. The following discussion is based on the results furnished in Table 4. It is pertinent to mention that the performance of a particular equation depends on the data range used in the analysis [Table 2 (a)] and the data limit of the equation.

For all ranges of $D/d$, $y/D$, and $V/V_c$, the proposed GP performance gives the best results that are quantitatively reflected in all statistical parameters, i.e., RMSE, MAPE, and $R$. Referring to Table 4, GP outperforms in high-value
predictions for the conditions of $D/d > 100$, $D/d \leq 50$, $y/D > 2$, $y/D \leq 1$, $V/V_c \leq 0.6$ and $0.9 < V/V_c \leq 1$, compared to all other traditional equations. It should be noted that GP is more effective at extreme ranges of $D/d$, $y/D$, and $V/V_c$.

Comparison of the various empirical equations (Table 1) is considered with reference to dimensionless pier width $D/d$. It can be confirmed that none of them give acceptable results, as reflected in higher RMSE and MAPE and lower $R$ for $D/d \leq 100$. At $D/d > 100$, only the equation by Kothyari et al. [24] gives the good results. However, for dimensionless approaching flow depth $y/D < 1$, the equations of Kothyari et al. [24] and Oliveto and Hager performed well, as reflected
Table 3. Statistical parameters of pier local scour equations.

(a) Verification and validation with all data used

<table>
<thead>
<tr>
<th>Equation</th>
<th>RMSE</th>
<th>MAPE (%)</th>
<th>Correlation coefficient R</th>
</tr>
</thead>
<tbody>
<tr>
<td>GP (Train) Eq. (5)</td>
<td>0.27</td>
<td>19.6</td>
<td>0.93</td>
</tr>
<tr>
<td>GP (Test) Eq. (5)</td>
<td>0.34</td>
<td>20.6</td>
<td>0.86</td>
</tr>
<tr>
<td>GP (All data) Eq. (5)</td>
<td>0.28</td>
<td>19.8</td>
<td>0.84</td>
</tr>
<tr>
<td>Shen et al. [36]</td>
<td>0.73</td>
<td>58.3</td>
<td>0.50</td>
</tr>
<tr>
<td>Hancu [21]</td>
<td>0.88</td>
<td>55.3</td>
<td>0.29</td>
</tr>
<tr>
<td>Kothyari et al. [24]</td>
<td>0.47</td>
<td>28.0</td>
<td>0.59</td>
</tr>
<tr>
<td>Gao et al. [17]</td>
<td>0.61</td>
<td>65.5</td>
<td>0.48</td>
</tr>
<tr>
<td>Melville [28]</td>
<td>0.58</td>
<td>60.3</td>
<td>0.68</td>
</tr>
<tr>
<td>Oliveto and Hager [34]</td>
<td>0.74</td>
<td>34.8</td>
<td>0.64</td>
</tr>
<tr>
<td>Kothyari et al. [25]</td>
<td>0.64</td>
<td>37.8</td>
<td>0.40</td>
</tr>
</tbody>
</table>

(b) Validation with data Dey [15], Sheppard et al. [37] and Raikar [35]

<table>
<thead>
<tr>
<th>Data</th>
<th>Equation</th>
<th>RMSE</th>
<th>MAPE (%)</th>
<th>Correlation coefficient R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dey [15]</td>
<td>GP Eq. (5)</td>
<td>0.25</td>
<td>21.6</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>Dey et al. [15]</td>
<td>0.29</td>
<td>22.5</td>
<td>0.90</td>
</tr>
<tr>
<td>Sheppard et al. [37]</td>
<td>GP Eq. (5)</td>
<td>0.36</td>
<td>30.2</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>Sheppard et al. [37]</td>
<td>0.29</td>
<td>21.1</td>
<td>0.84</td>
</tr>
<tr>
<td>Raikar [35]</td>
<td>GP Eq. (5)</td>
<td>0.19</td>
<td>12.7</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>Raikar [35]</td>
<td>0.17</td>
<td>10.5</td>
<td>0.95</td>
</tr>
</tbody>
</table>

(c) Data of Federal Highway Administration (FHWA) used for verification

<table>
<thead>
<tr>
<th>Data</th>
<th>Equation</th>
<th>RMSE</th>
<th>MAPE (%)</th>
<th>Correlation coefficient R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lander and Muller [27]</td>
<td>GP Eq. (5)</td>
<td>0.81</td>
<td>83.0</td>
<td>0.75</td>
</tr>
<tr>
<td>Molinas [32]</td>
<td>GP Eq. (5)</td>
<td>0.51</td>
<td>56.8</td>
<td>0.70</td>
</tr>
</tbody>
</table>

Note: $F_d < F_{d50}$ and $d_{50} < 0.8$ mm have been excluded for Oliveto and Hager [34] and Kothyari et al. [25] equations.

in lower RMSE and MAPE. However, at $1 < y/D$, all of them give reasonable results because only one performs well for RMSE, MAPE and $R$ as seen in Table 4.

With regard to the performance considering dimensionless flow velocity $V/V_c$ (Table 4), the Kothyari et al. [24] equation gives good accuracy and correlation for $V/V_c \leq 0.6$, whereas the other equations do not show good results for this range of $V/V_c$. In addition, the Melville equation has lower errors and higher $R$ than the other six equations at $0.6 < V/V_c \leq 0.9$. For $0.9 < V/V_c \leq 1$, the error for these equations tends to be high, except for the equation of Kothyari et al. [24]. There is considerable variation in the prediction of scour depth at this high value of $V/V_c$, because this range of $V/V_c$ approaches live-bed scour; therefore, the pier scour depth was difficult to measure for this range. As mentioned earlier, the prediction of a particular equation depends on the data range used in deriving the equation.

The comparison of GP performance with other empirical equations presented in Figs. 2 and 3, illustrate that the pier scour equations proposed by Gao et al. [17] and Melville [28] overestimate scour depth [Figs. 2(d) and 3(a)] because these formulas are based on high safety factors and envelop curves to data. Therefore, the correlation coefficient $R$ for these two equations is lower in some selected ranges of $D/d$, $y/D$, and $V/V_c$, indicating poor performance. However, the Shen et al. [36] equation over predicts the scour depth to some extent [Fig. 2(a)] but still performs well under the conditions of $D/d > 100$, $0 < y/D \leq 1$, and $0.6 < V/V_c \leq 0.9$. Contrary to this, the equations of Hancu [21], Kothyari et al. [24], and Kothyari et al. [25] underpredict the scour depth [Figs. 2(b), 2(c), and 3(c)]. However, the Oliveto and Hager [34] equation has good predictions under the conditions of $50 < D/d \leq 100$ and $0.6 < V/V_c \leq 0.9$ as indicated by the significant value of $R$ and also evidenced by Fig. 3(b). Furthermore, the Kothyari et al. [24] equation has an advantage over the other equations, as it is based on a larger data range, which was used for regression analysis, but with minimal $R$ in some selected ranges of $D/d$, $y/D$, and $V/V_c$. The correlation coefficient $R$ is lower, showing that there is a wide variation in the prediction of scour depth. In addition, the other equations perform well only in higher or lower selected ranges of $D/d$, $y/D$, and $V/V_c$. The robustness of GP is further evaluated with the experi-
mental data of Dey [15], Sheppard et al. [37], Raikar [35], Landers and Muller [27], and Molinas [32], which were not used in developing the GP model. Tables 2 (b) and (c) furnish details of the experimental data. The RMSE, MAPE and \( R \) values using the GP formula for these data are listed in Tables 3 (b) and (c). The statistical parameters show that the prediction performance of GP is satisfactory. Hence, the GP model can be used with a wide range of data because the data of Sheppard et al. [37] consist of large-scale pier models with a size of 110-910 mm and a flow depth of 170-1700 mm. The GP formula still has acceptable results reflected in RMSE, MAPE, and \( R \) even when the few extreme ranges of data are used for predictions. In addition, the GP model is found to be suitable for the gravel-bed data (gravel bed conditions) of Raikar. Fig. 4 shows the comparison of scour depth as predicted by the GP formula and respective sources.

For sediment particles smaller than 2 mm (sand bed conditions), the prediction is reasonable, as seen from the data of Dey and Sheppard et al. as evidenced from Figs. 4(a) and 4(b). However, for \( d_{50} > 2 \) mm (gravel bed conditions), the results show some overprediction. This might be due to two facts. First, in GP model building, the large particles were not included in the data range \( (d_{50} \leq 4.02 \text{ mm}) \) as they were in Molinas’s laboratory data and Raikar’s gravel-bed data. This could be the prominent reason for the current GP equation to overpredicting the scour depth for the measured \( d_{50}/D < 1 \) [Figs. 4(c) and 4(d)]. Second, the scour depth had not reached the equilibrium condition at the time of the field measurements owing to a short flood duration relative to the equilibrium time, or perhaps the river bed aggraded during the recession of the flood. As a safety concern, the overprediction provides a greater safety factor for practical engineering.

It is noteworthy to compare the performances of GP and the ANN. Azamathulla et al. [7] presented that the performance of GP was found to be better than that of the ANN in predicting ski-jump bucket spillway scour. The advantage of GP has been reported in the work of Azamathulla et al. [5] for predicting pier scour depth. The performances of ANN and ANFIS have also been compared. Azamathulla et al. [8], Guven and Gunal [19], and Guven et al. [18, 20] mentioned that the ANFIS models specialized in training data and have a poor generalization capacity on testing data. This issue is known as an overgeneralization problem, which is a common issue in neural network techniques.

In addition, it is interesting to compare the performances of LGP and ANFIS. Guven et al. [18], and Azamathulla et al. [8] reported that the LGP models are more flexible than the ANFIS models that were considered, with more factors being incorporated in the former. In addition, the proposed LGP models were much more practical and robust than the ANFIS models.
models. It is noted that GP has the same high generalization capacity and flexibility as LGP, indicating that GP can be a better predictor of scour depth than neural network techniques.

VI. CONCLUSIONS

This study provides an efficient approach to develop an equation for the prediction of pier scour depth using GP. The following conclusions are drawn from this study:

1. The proposed GP formula as given in Eq. (5) for the prediction of pier scour depth was developed based on limited experimental data and was proved to agree better with experimental results than did the other empirical equations considered in this study.

2. The proposed GP formula has a higher and more stable accuracy in all ranges of dimensionless pier width \((D/d)\), dimensionless approaching flow depth \((y/D)\), and dimensionless flow velocity \((V/V_c)\) than the other empirical equations considered in this study, indicating the usefulness of the method under a wide range of experimental conditions with clear-water scour at bridge piers with uniform sediments. The other equations work well only in some selected ranges of these conditions.

3. The comparisons between the GP formula in Eq. (5) and the other predictors (Dey [15], Sheppard et al. [37], Raikar [35], Landers and Muller [27], and Molinas [32]), whose data were not used in developing the GP model, show that GP has a wide range of capacities and practical applications. However, the GP model overpredicts for large particles, as this data range is not included in model building. Regarding the field data in this case, the equilibrium scour depth might not have been reached or the river bed may have aggraded.

4. The present GP model results were compared with those of empirical equations, some of which are dimensional equations (Shen et al. [36] and Gao et al. [17]), non-dimensional equations (Hancu [21], Kothyari et al. [24], and Melville [28]), and unsteady models (Oliveto and Hager [34] and Kothyari et al. [25]). The advantage of the

Fig. 4. Comparison of scour depth predicted by GP and computed from: (a) Dey [15], (b) Sheppard et al. [37], (c) Raikar [35], (d) Landers and Muller [27], and Molinas [32].
model results (smaller errors and greater $R$) indicates that the present model can be applied to varied conditions.

odi et al., Jothiprakash, V., and Deo, M. C., “Genetic programming to


13. Dey, S., “Local scour at piers, part I: a review of development of re-

10. Breusers, H. N. C. and Raudkivi, A. J.,

(5) GP can create randomly formed functions and fit the ex-

perimental results for irrelevant attribute data and even small datasets, unlike empirical formulas, which are gen-

erally based on predefined functions. In addition, there is

no restriction in the complexity and structure of the ran-

domly formed functions in GP. In addition, GP can gen-

erate a transparent mathematical structure that can be used in

general.

ACKNOWLEDGMENTS

The authors would like to thank the National Science Council of the Republic of China, Taiwan, for financially

supporting this research.

REFERENCES


8. Azamathulla, H. M., Guven, A., and Demir, Y. K., “Linear genetic pro-


12. Deo, O., Jothisprakash, V., and Deo, M. C., “Genetic programming to

predict spillway scour,” International Journal of Tomography and Sta-


13. Dey, S., “Local scour at piers, part I: a review of development of re-


15. Dey, S., “Time-variation of scour in the vicinity of circular piers,” Pro-


16. Ettema, R., Scour at Bridge Piers, Report No. 216, School of Engineer-
ing, University of Auckland, Auckland, New Zealand (1980).


19. Guven, A. and Gunal, M., “Prediction of scour downstream of grade-

control structures using neural networks,” Journal of Hydraulic Engi-


