Scientia Iranica A (2023) 30(6), 1925–1941

Sharif University of Technology
Scientia Iranica
Transactions A: Civil Engineering
http://scientiairanica.sharif.edu

Multi-gene GP and GA-FIS models to deal with scaling problem in the ANFIS model for estimating roughness coefficient in erodible channels

M. Zanganeh*

Department of Civil Engineering, Faculty of Engineering, Environmental Hazard Institute, Golestan University, Golestan, Iran.

Received 7 April 2022; received in revised form 21 June 2022; accepted 18 January 2023

KEYWORDS
Estimation;
Roughness;
GP;
GA-FIS;
Erodible channel.

Abstract. Estimation of roughness coefficient is important for a reliable hydraulic design in erodible channels. In this paper, the capability of multi-gene Genetic Programming (GP), a combined Genetic Algorithm and Fuzzy Inference System (GA-FIS) model, and Multi Regression (MR) methods are employed to estimate the roughness coefficient. These methods try to extract either an explicit or implicit relationship between roughness coefficient and input variables. In addition, traditional GP, widely used by researchers, and conventional empirical formulas are implemented to evaluate the performance of the models. Results show that the employed methods are more accurate than empirical methods. In addition, the effects of some other parameters, such as non-dimensional water depth and shear Reynolds number, are highlighted over the roughness coefficient while previously ignored in the empirical methods. Also, findings prove that the GA is a helpful tool to optimize the Fuzzy Inference System (FIS) compared with gradient-based models like ANFIS, while the scale of input variables is not in the same order. Values of R^2 for multi-gene GP and GA-FIS are 0.8504 and 0.8842, respectively, while R^2 value for the most accurate empirical method is 0.6286.

1. Introduction

Estimation of roughness coefficient is necessary for a cost-effective hydraulic design of erodible channels, which are usually used to transfer water to shrimp farms, irrigation systems, etc. Accurate estimation of this parameter is essential for numerical modeling of fluid flow in open channels. In that regard, many experimental attempts have been made to achieve some representative empirical formulas for estimating the roughness. These formulas are commonly based on fitting a function between the roughness coefficient and its influential variables [1]. Evaluation of these techniques in various conditions convinces their deficiency to estimate the coefficient in different states. For example, in an empirical formula, variable x is the most influential variable, while in another formula, this variable is replaced by variables like y. This might go back to different hydraulic conditions, leading to user confusion. Inaccurate estimation of the coefficient may lead to either the non-economical design of channels or inefficient dimensions. In other words, numerous researchers have introduced various formulas to estimate the roughness coefficient. Ackers and White (1973) [2], Simons and Richardson (1996) [3], Hammond et al. (1984) [4], and Colosimo et al. (1986) [5] introduced the following relationship for the coefficient estimation:

$$k_s = \alpha_x d_x, \tag{1}$$

*. *Tel./Fax: +98 17-3422-6490*
E-mail address: m.zanganeh@gu.ac.ir (M. Zanganeh)

doi: 10.24200/sci.2023.60241.6679

where α_x is a constant coefficient and d_x is the size of the sediment, where x percent of sediment particles are finer than their size. In another research, Wislon (1989) [6] and Yalin (1992) [7] took Shields parameter into account for estimating k_s as introduced in the following expression:

$$\theta = \frac{U_f^2}{\sqrt{g\,(S-1)\,d_{50}}}, \tag{2}$$

in which g is gravitational acceleration and d_{50} is the size of the sediment particle where 50% of sediment particles are smaller than that. S is specific gravity (ρ_s/ρ_w) and U_f is shear velocity expressed as follows:

$$U_f = \sqrt{g r_b I}, \tag{3}$$

in which r_b is the hydraulic radius and I is channel slope. Sumer et al. (1996) [1] considered w/U_f and w_* as two variables in the estimation of k_s, where w is the sediment falling velocity and w_* is calculated in the following:

$$w_* = w/\sqrt{g\,(S-1)}d_{50}. \tag{4}$$

Although several formulas have been introduced so far for estimating the roughness coefficient, some influential input variables are not considered in most of them. Therefore, it seems that some complementary works are needed to extract new formulations that not only represent a phenomenon in a more generalized way, but also investigate influential variables in the coefficient. In this paper, soft computing-based approaches, including Genetic Programming (GP) and a combined Genetic Algorithm and Fuzzy Inference System (GA-FIS), are employed because of their potential to work in an area where values of input variables are not in the same order.

In recent years, soft computing-based approaches such as Artificial Neural Networks (ANNs) and Fuzzy Inference Systems (FISs), GP, Support Vector Machines (SVMs), and so on have been employed to predict complex phenomena or estimate functions representing a complex physical process ([8–18]).

Azamathulla and Ahmad (2013) used Gene Expression Programming (GEP) to estimate the Manning's roughness coefficient. Results showed that computed discharge using estimated value of roughness coefficient by GEP was in good agreement (±10%) with the experimental results compared to the conventional formulae [19]. Roushangar et al. (2020) applied five different experimental data sets to train the model for predicting roughness coefficient in alluvial channels and then, they confirmed the accuracy of the bedform characteristics in predictor models. In addition, the sensitivity analysis revealed the Reynolds number and the relative discharge effectiveness in the process. Besides, the dune geometry evaluation showed that the densimetric Froude number was the most significant variable. Their paper employed an ANN and Multi-Layer Perceptron (MLP) with firefly algorithm (MLP-FFA) as prediction models [20]. Zanganeh and Rastegar (2020) applied ANN and ANFIS models to estimate the roughness coefficient value in an erodible channel. In these models, they used both normalized and real data to estimate roughness coefficients and found the effectiveness of other parameters like Reynolds number in roughness coefficient [21].

This paper aims to apply the GP features as explicit and the combined GA-FIS as implicit function approximators to find relationships between influential input variables and the roughness coefficient as the output variable. Many researchers have previously used these approaches to predict hydraulic phenomena. Deo et al. (2008) implemented the GP to estimate the equilibrium depth of scour downstream of spillways [22], whereas Azamathulla et al. (2008) applied the GP to estimate scour depth downstream of ski-jump buckets spillway [23]. Guven and Kişi (2011) used Linear Genetic Programming (LGP) for estimating suspended sediment yield in rivers and demonstrated better performance of LGP method than the GEP method [24]. Azamathulla et al. (2009) utilized the ANN and GP to predict scour depth at bridge piers and reported that the GP outperformed regression-based models and ANNs [25]. Azamathulla et al. (2011) employed the LGP model for scouring below a submerged pipeline [26]. Najafzadeh and Barani (2011) compared the group method of data handling-based GP and the back-propagation system to estimate scour depth around bridge piers [27]. Koç et al. (2016) investigated the GP capability to assess the stability of rubble-mound breakwaters [28]. Heřmanovský et al. (2017) applied the GP model to derive a hydrological model to estimate runoff in ungauged catchments by regionalization [29]. Assimi et al. (2017) introduced a GP for simultaneous optimization of sizing and topology of truss structures [30]. Zanganeh (2017) recently used combined GA-FIS models to predict the values of wave parameters. In the models, the GA is used to improve FIS-based wave predictor models with simultaneous optimization of clustering and fuzzy antecedent and consequent parameters. Results showed the model performance in estimating wave parameters, including significant wave height and peak spectral period [31]. More recently, Zanganeh (2020) evaluated the performance of combined PSO-FIS models including PSO-ANFIS, PSO-FIS-PSO, and PSO-FIS models for wave prediction [32].

After a brief review of the previous works in the field of water engineering in Section 1, this study reviews the features of the studied case and data to be used for model development in Section 2. Section 3 describes traditional GP and multi-gene GP models,

while Section 4 outlines the combined model of FIS and GA to estimate the roughness coefficient. In Section 5, empirical methods and a regression-based model are outlined to estimate the roughness coefficient in erodible channels, while Section 6 contains the models developed to calculate the coefficient. Finally, in Section 7, the models are evaluated to estimate the roughness coefficient.

2. Studied case and data selection

As mentioned before, the aim of the paper is to estimate the roughness coefficient expressed in terms of Nikuradse's sand roughness. This parameter is calculated by the following relationship [1]:

$$\frac{U}{U_f} = 2.46ln\left(\frac{14.8r_b}{k_s}\right), \tag{5}$$

where U is the mean flow velocity calculated as $U = Q/(Bh)$, Q flow discharge; B flume width, h flow depth, and U_f bed shear velocity.

Data sets are the most important issues in developing any soft computing-based model to predict an event or fit a function to estimate a variable. In this paper, the data sets gathered by Sumer et al. (1996) [1] at the Institute of Hydrodynamics and Hydraulic Engineering, Technical University of Denmark (ISVA) are used to develop the models. These data sets are related to experiments in a tilting flume that is 10 m long, 3 m high, and 3 m wide. In the flume, sediment and water are recirculated, while a rigidly placed lid is used to avoid surface waves. However, some experiments have been coordinated in a free surface flow condition. The flume is made of glass for the sake of visual tracking of sediments. Four kinds of sediments were used in the experiment. In addition to the roughness coefficient, other parameters like velocity profile and sediment concentration profile were measured. The list of data sets, including all directly measured and calculated variables by the author, are reported in Table A.1 of Appendix A. These experimental data represent both the suspension and the no-suspension modes of sediment transport in open channels with a unidirectional flow. In this data set, not only sheet-flow layer effects as a condensed layer near the bed but also the mechanism like turbulent bursting condition are considered. Sheet-flow regime occurs at very high velocities by which ripples are washed away, and the bed becomes plane again, as shown in Figure 1. Sumer et al. (1996) attempted to find a function to estimate the roughness coefficient as follows [1]:

$$\frac{k_s}{d_{50}} = f\left(\theta, \frac{w}{U_f}, w_*\right). \tag{6}$$

Due to the complexity of the flow field and sediment transport in erodible channels and ignoring many important parameters on channel roughness in the previous works, in this paper, new models are attempted to be developed for estimation of the parameter. To this end, a function representing the relationship among parameters is defined as follows:

$$Z(k_s, d_{50}, h, w, \rho_s, \rho_f, U_f, g, \mu_f) = 0, \tag{7}$$

in which Z is a function, k_s is roughness coefficient, d_{50} is average particle size, h is water depth, w is fall velocity, ρ_s is sediment density, ρ_f is the fluid density, U_f is shear velocity, g is gravitational acceleration, and μ_f is dynamic viscosity.

Using Buckingham theorem, the non-dimensional form of the equation to estimate the roughness coefficient can be extracted as follows:

$$\frac{\mathrm{k_s}}{d_{50}} = f\left(h/d_{50}, R_*, \theta, w/U_f\right), \tag{8}$$

in which f is a none-dimensional function that can be either an implicit function like the FIS-based model or an explicit function like the GP. R_* is shear Reynolds number and θ is Shield's parameter.

In this paper, the GP, GA-FIS, and regression-based models are developed to estimate the roughness coefficient. To this end, the selected data sets from Sumer et al. (1996) [1] are categorized as the training, validation, and testing data sets (Table A.1. in Appendix section). These three subsets have been chosen randomly for models with sufficient generalization capability. From 158 data points gathered by Sumer et al. [1], 100 data points are chosen randomly as training data points, 18 data points as validation data

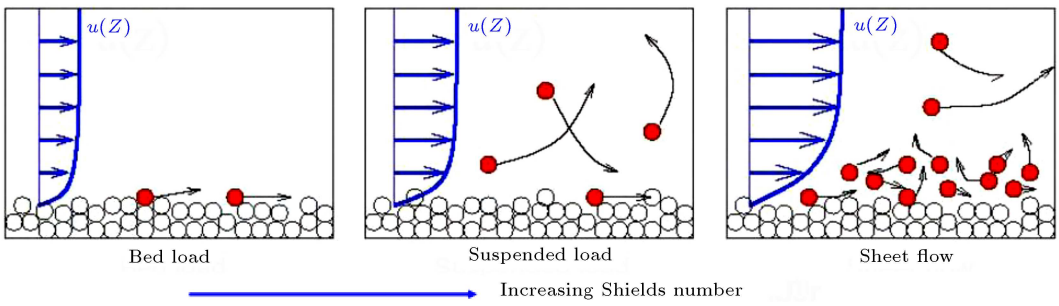

Figure 1. Different sediment transport mechanisms.

Table 1. Statistical characteristics of the data sets used for developing models to estimate roughness coefficient in erodible channels.

Characteristics	h/d_{50}	R_*	θ	w_*	w/U_f	k_s/d_{50}
Mimimum	25.00	5434	0.38	0.262	0.174	1.430
Maximum	808.46	436200	5.67	1.330	1.700	21.700
Average	164.51	177564	2.04	1.007	0.734	6.720
Range	783.46	430766	5.290	1.068	1.526	20.270

points, and the remaining 40 data points as the testing data. Table 1 outlines the statistical characteristics of the data sets used to develop the estimator models. This table reports the maximum, minimum, average, and range of the training, validation, and testing data for each input variable. As is apparent from the table, it covers a wide range of data representing various sediment transport conditions in erodible open channels. This makes the data set informative enough to be used for developing soft computing models. A debatable point about the data set goes back to having different ranges for input parameters. For example, the shear Reynolds number has a range of 5434 to 436200, while the range for Sheilds number varies from 0.38 to 5.67. This gives at least a scale of about 1 to 1147894 for Sheilds number to shear Reynolds number. This data set feature called "scaling problem" makes the training space noisy, especially for the gradient-based models like the ANFIS. A remedy to tackle the problem is normalizing the variables and bringing them in the same order. Another solution is employing an evolutionary-based model like GA-FIS model. The GA-FIS model is like the ANFIS model in which the GA is replaced by the gradient-based algorithm to tune antecedent and consequent parameters of fuzzy IF-THEN rules. Simultaneously, the GA is used to extract the structure of fuzzy IF-THEN rules.

3. Genetic Programming (GP)

The three-based GP machine learning method extended by Koza (1992) is a domain-independent problem-solving approach [33]. In this method, computer programs are evolved by influential variables of a phenomenon to extract a function for predicting the phenomenon based on Darwin's theory. The number and structure of the trees are evolved automatically during a run subjected to the user-defined constraints using training and validation (to prevent over-fitting problem) data sets by either assigning arithmetic operations $(+, -, /, *)$ or calling function such as sin, cos, log, ln, sqrt, power (Figure 2). This tool works the same as the GA by randomly generating a population of tree structures and then, mutating and crossing over the best-performing trees to develop a new population. This process is iterated until the population contain-sprograms to solve the task well.

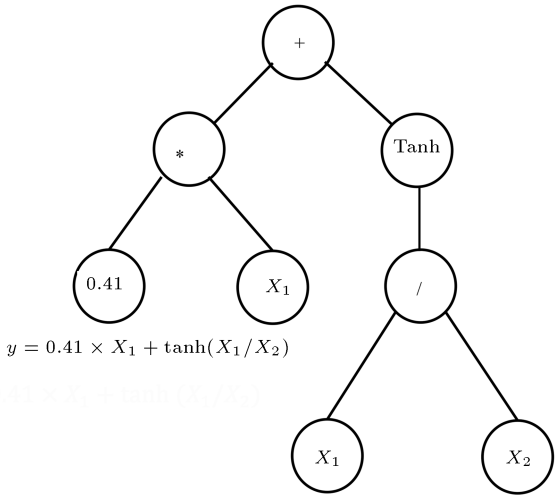

Figure 2. Basic tree program used in the GP.

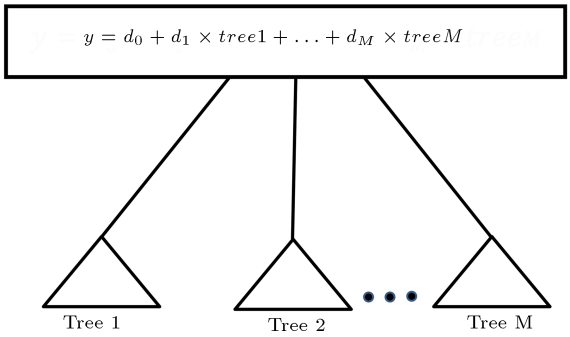

Figure 3. Multi-gene GP model combination.

The GP can be used in either a single gene or multi-gene forms depending on the complexity of the phenomenon to be estimated. In the multi-gene symbolic regression form, each estimation of output variable y is composed of the weighted output of each tree/gene in the multi-gene individual plus a bias term, as depicted in Figure 3. Each tree is a function of N input variables like $x_1, \cdots, x_N$. A multi-gene regression model mathematically can be expressed as follows:

$$y = d_0 + d_1 \times tree1 + \cdots + d_M \times treeM, \quad (9)$$

in which d_0 is bias term; $d_1, \cdots, d_M$ are the gene weights; and M is the number of genes comprising the current individual. The weights (i.e., regression coefficients) are determined by a least squares method for each gene. The single form of the GP is obtained,

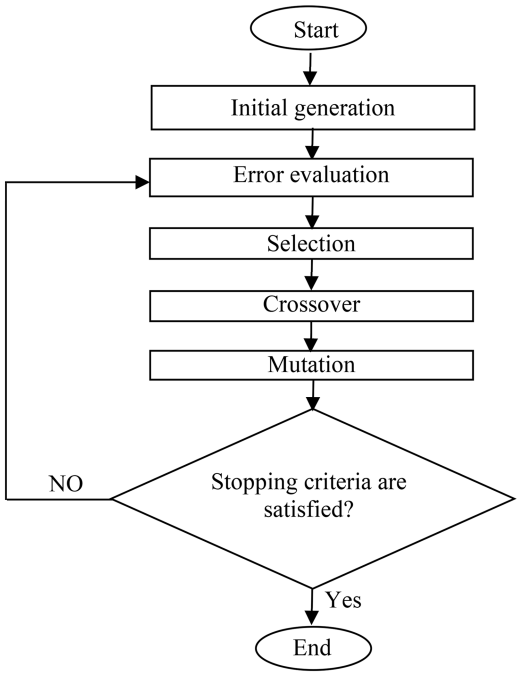

Figure 4. The flow diagram of the GP model.

while the number of genes is set to 1. Testing data as other data sets representing the system to be modeled can be used to evaluate the performance of the developed model. Note that like other soft-computing-based models, the testing data are not directly used for evolving the models. This gives an indication how well the models perform versus new data. Figure 4 shows the flow diagram of the GP model as a function estimator model. In this paper, the GPTIPS model developed by Searson et al. (2010) as a multi-gene symbolic regression function is employed in order to estimate the roughness coefficient in an erodible channel. The following expressions outline GP operators in the GPTIPS [34]:

3.1. Initial generation

The creation of individuals in the initial generation is straightforward. An individual containing a random number of genes between one and the maximum number of genes ($G_{\max}$) is generated using the "standard" algorithm for constructing symbolic expressions. To maximize the population diversity at the beginning of the run, checks are made so that duplicated genes do not appear in newly created individuals. However, such a restriction is imposed on individuals in subsequent generations. This is mainly due to the additional computational cost considerations in isolating duplicate and functionally similar genes. $D_{\max}$ as the maximum tree depth is another factor controlling the complexity of the evolved model.

3.2. Selection

Genetic operators need parent individuals to produce their children. In the GP, one of the following sampling methods selects these parents:

- Tournament: This method chooses each parent by randomly drawing a number of individuals from the population and selecting only the best of them;
- Lexictour: This method implements lexicographic parsimony pressure. Like in 'tournament', a random number of individuals are chosen from the population and the best of them is chosen. The main difference is that if two individuals are equally fit, the shortest one (the tree with fewer nodes) is chosen as the best. This technique has been shown to control bloat in different types of problems effectively.

3.3. Crossover

Individuals that have been selected for recombination should be able to acquire new genes and swap complete genes with other individuals. Both functions are performed in the multi-gene algorithm with the two-point high-level crossover operator. As two-parent individuals have been selected, two gene crossover points are selected within each parent. Then, the genes enclosed by the crossover points are swapped between parents to form two new offspring shown in Figure 5. In the two-point high-level crossover operator, if the ith gene in an individual is labeled G_i, then the crossover is performed as in the following example. The first parent individual contains the genes (G_1 G_2 G_3), and the second contains the genes (G_4 G_5 G_6 G_7) where $G_{\max} = 5$. Two randomly selected crossover points are created for each individual. The genes enclosed by the crossover points are denoted by $< \cdots >$.

$$(G_1 < G_2 > G3)\ (G4 < G_5\ G_6\ G_7 >). \tag{10}$$

The genes enclosed by the crossover points are then exchanged, resulting in the two new individuals below.

$$(G_1\ G_5\ G_6\ G_7\ G_3)\ (G_4\ G_2). \tag{11}$$

3.4. Mutation

By this operator, the GP algorithm would reshuffle the genes created in the initial generation. The mutation operator is identical to the GA model in terms of function. As reported in Figure 6, a single gene is randomly selected from the parent. Then, the mutated gene replaces the original gene in the offspring.

4. Combined GA-FIS model (GA-FIS)

To evaluate the GP model, to estimate the roughness coefficient, a GA-FIS model also is used in which both fuzzy IF-THEN rule structures and fuzzy IF-THEN rule parameters, including fuzzy antecedent and consequent parameters, are tuned simultaneously by a GA. Three reasons motivated the author to apply this model rather than the common ANFIS model.

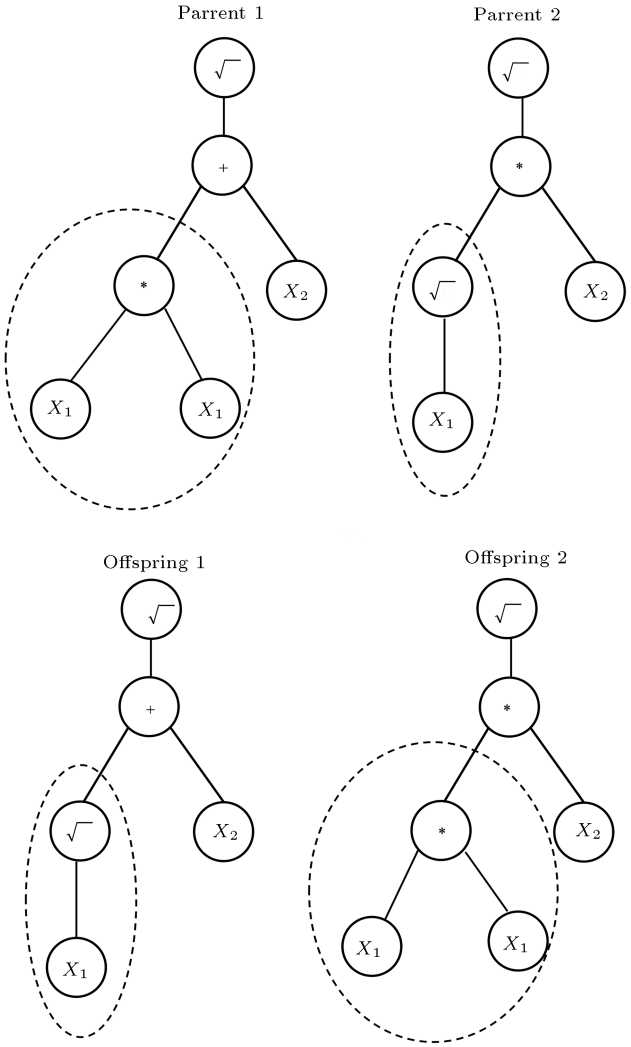

Figure 5. Crossover in the GP trees.

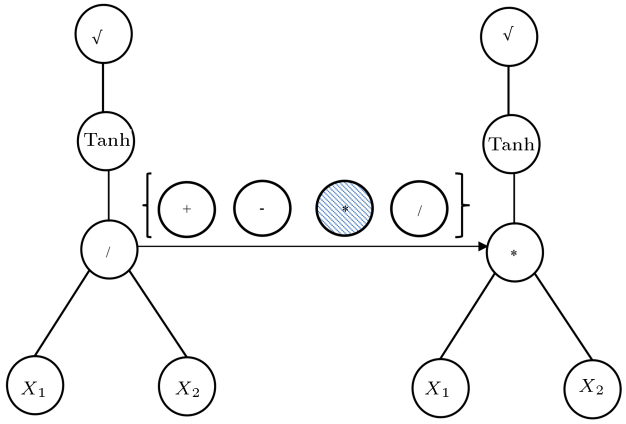

Figure 6. Mutation in the GP trees.

The first reason is ANFIS deficiency for automatic extraction of fuzzy IF-THEN rules, while the second deficiency of the ANFIS model rooted in its gradient-based essence. Gradient-based methods like the ANFIS have a problem with data sets whose input variables are not in the same order (scaling problem). This makes some training processes difficult and makes the final answer susceptible to trapping in a local optimum [35].

Fuzzy IF-THEN rules are chosen with respect to having the lowest similarities among rules such as the process performed at GENFIS2 command in the MATLAB. This process is based on subtractive clustering method [36]. The rules associated with the FIS models to estimate roughness coefficient are defined as follows: Rule 1: IF h/d_{50} is A_1&R_* is B_1&θ is C_1&w/U_f is D_1 THEN.

$$k_s/d_{50} = p_1 + q_1(h/d_{50}) + r_1 R_* + s_1\theta + t_1(w/U_f). \quad (12)$$

Rule 2: IF h/d_{50} is A_2&R_* is B_2&θ is C_2&w/U_f is D_2 THEN.

$$k_s/d_{50} = p_2 + q_2(h/d_{50}) + r_2 R_* + s_2\theta + t_2(w/U_f)$$

$$\ldots \quad (13)$$

Rule 2: IF h/d_{50} is A_i&R_* is B_i&θ is C_i&w/U_f is D_i THEN.

$$k_s/d_{50} = p_i + q_i\,(h/d_{50}) + r_i R_* + s_i\theta + t_i\,(w/U_f)\,, \quad (14)$$

where A_i, B_i, C_i, and D_i are fuzzy sets related to h/d_{50}, R_*, θ, and w/U_f, respectively, and p_i, q_i, r_i, s_i, and t_i are their associated consequent parameters in fuzzy rules. Two aspects of the GA application can be considered to optimize an FIS. The first aspect is related to employing the GA to optimize subtractive clustering parameters, while the ANFIS model introduced by Jang (1993) is implemented to tune the antecedent and consequent parameters of fuzzy IF-THEN rules extracted by the GA. This model is called GA-ANFIS model [37]. In this type of model, controlling the ANFIS parameters like epoch number is difficult. In the second aspect, only one GA model is used for simultaneous optimization of subtractive clustering parameters and fuzzy antecedent and consequent parameters. This model is called GA-FIS model, while Figure 7 shows its flow diagram. As is shown in the diagram in this model, only one GA is used to optimize FISs in which the supervision of the GA parameters is met easily. The objective function of the GA optimizer models is the minimization of the Root Mean Square Error (RMSE). Formulation of the objective function for approximating the roughness coefficient as the output is outlined in Figure 8, where ra_n is the clustering radius for the nth premise or consequent variable ($n = 1, \cdots, D$) and $ra_{\min_n}$ is the minimum clustering radius for the nth variable. This minimum value is used to prevent zero values for firing strength. *MaxNumrule* is the maximum number of fuzzy rules determined based on the prediction errors for training and validation data, i.e., $RMSE_{trn}$ and $RMSE_{Validation}$. Note that in these forms of

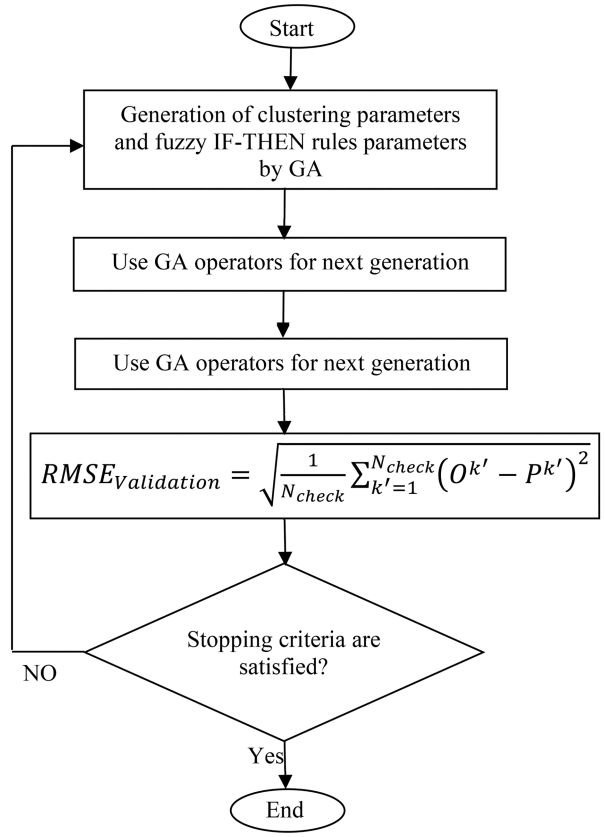

Figure 7. Flow diagram of the GA-FIS model.

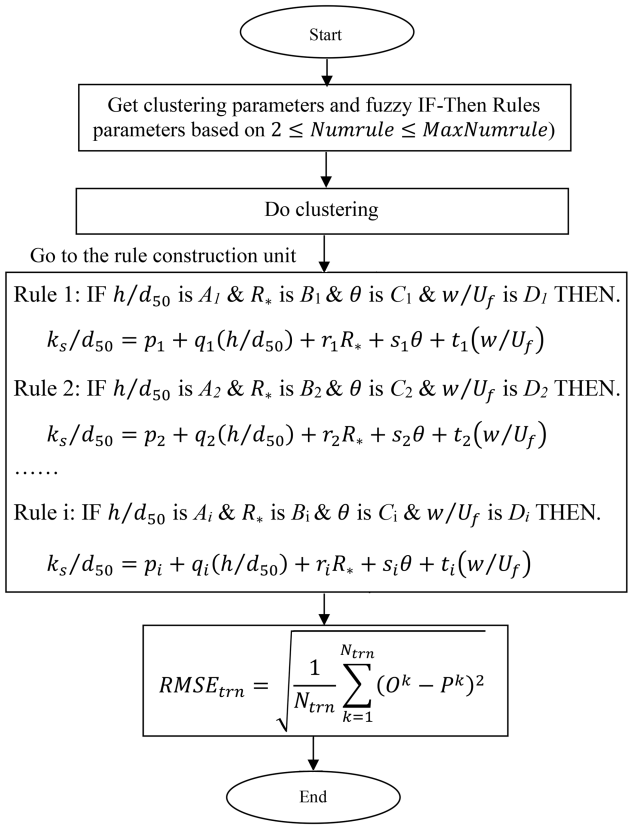

Figure 8. Objective function of the GA model.

the GA application, the numbers of the antecedent and consequent parameters are related to subtractive clustering parameters. Therefore, the number of the decision variables varies in the process of running the GA model, although *MaxNumrule* restricts the number of rules.

5. The traditional models

To verify the function of GP and GA-FIS models to estimate roughness coefficient versus traditional methods, these models are compared with the Multiple Regression (MR) techniques and some empirical formulas widely used by many researchers. The following subsections outline these models.

5.1. Multiple Linear Regression (MLR) model

In this approach, a linear relationship fits effective input and output variables through the training process as follows:

$$k_s/d_{50}\beta_0+\beta_1(h/d_{50})+\beta_2R_*+\beta_3\theta+\beta_4(w/U_f), \quad (15)$$

in which k_s/d_{50} is the output variable; β_0, β_1, β_2, β_3, and β_4 are constant parameters for the linear relation tuned by the Least Square Error (LSE) method.

5.2. Multiple Nonlinear Regression with Power function model (MNLRP)

In addition to the traditional MLR, the MNLRP can estimate an event by fitting a nonlinear relationship to input and output variables. The form of the nonlinear relation can be as follows:

$$k_s/d_{50} = \alpha_0(h/d_{50})^{\alpha_1}(R_*)^{\alpha_2}(\theta)^{\alpha_3}(w/U_f)^{\alpha_4}, \quad (16)$$

in which α_0, α_1, α_2, α_3, α_4 are constant parameters for the nonlinear relation, which are also tuned by the LSE method. The LSE method is used to tune constant parameters after taking the logarithm from both sides of the equation.

5.3. Empirical formulas

As mentioned before, for estimation of roughness coefficient, some empirical formulas have been presented so far. Wilson (1988 and 1989) presented the following expression for the roughness coefficient:

$$k_s/d_{50} = 5\theta. \quad (17)$$

Also, Yalin (1992) introduced the following relationship for estimating the coefficient:

$$\begin{cases} \frac{k_s}{d_{50}} = 5\theta + (\theta - 4)^2 \\ \quad (0.043\theta^3 - 0.289\theta^2 - 0.0203\theta + 0.125) & 1 < \theta < 4 \\ \frac{k_s}{d_{50}} = 2 & \theta < 1 \end{cases} \quad (18)$$

Sumer et al. (1996) presented the following expression for estimation of the coefficient:

$$\begin{cases} \frac{k_s}{d_{50}} = 2 + 0.6\theta^{0.25} & \frac{w}{U_f} > 0.8-1 \\ \frac{k_s}{d_{50}} = 4.5+1.8\exp\left(0.6w_*^4\right)\theta^{0.25} & \frac{w}{U_f} < 0.8-1 \end{cases} \quad (19)$$

The performance of the above-mentioned relationships is dependent on parameters representing a different sediment transport mechanism. If $w/(U_f \leq 0.8-1.0$, sediment particles are in the suspended mode, while for $w/(U_f > 0.8-1.0$, sediments are settled at the bottom of the channel. In this paper, soft computing-based models are developed by influential variables, which are identical to the empirical formula. However, additional variables like shear Reynolds number and non-dimensional water depth can be taken to increase the performance of the data-driven models.

6. Development of the models to estimate the roughness coefficient

6.1. The GP models

The GP models are developed considering two aspects to estimate the roughness coefficient. The first one goes back to the traditional single-gene GP, and the second is the multi-gene aspect. To this end, the GP parameters are set by a trial and error process reported as follows: Population size = 500, Number of generations = 1000, Tournament size = 15 (with lexicographic selection pressure), $D_{\max} = 6$, $G_{\max} = 8$, Elitism = 0.01% of population, and function node set = 'times', 'minus', 'plus', 'rdivide', 'square', 'tanh"exp', 'log', 'mult 3', 'add 3', 'sqrt', 'cube', 'negexp', 'neg', 'abs'. In addition, the default GPTIPS multi-gene symbolic regression function was used in order to minimize the RMSE of the training data. The following recombination operator event probabilities are used: Crossover events = 0.85 and mutation events = 0.1. These settings are not considered 'optimal' in any sense, but are based on experience with modeling other data sets of similar size. Run took approximately 15 minutes on a Core™ i5 PC running at 2.50 GHz with 4.0 GB of RAM.

The training processes for both traditional and multi-gene GP models are shown in Figure 9. Small error noises and the decreasing trend of RMSEs in the figure ensure suitable selection of input variables. The following expression report trees were obtained by multi-gene GP.

$$\begin{aligned} k_s/d_{50} = {} & 9.477\times10^{-6}R_*\theta \\ & -3.547\ln\left(h/d_{50}+7.452\right)-3.547\log\left(w/U_f\right) \\ & -\ln\left(2R_*+h/d_{50}\right)3.457+0.6629\sqrt{e^\theta} \\ & +\frac{9.477\times10^{-6}R_*}{w/U_f}+0.1747(w/U_f)^2 \end{aligned}$$

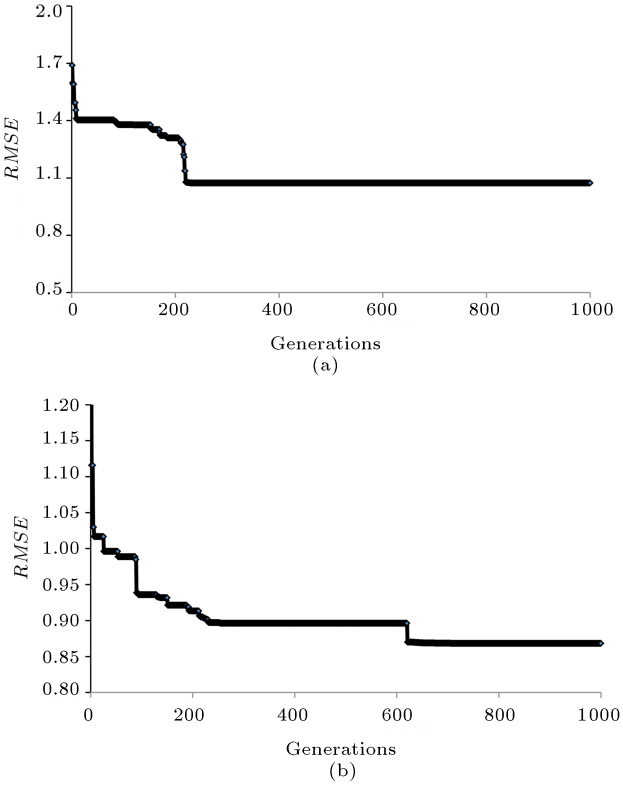

Figure 9. The training processes to estimate roughness coefficient: (a) The traditional GP model and (b) in the multi-gene GP model.

$$+0.6629(w/U_f)^3+55.34. \quad (20)$$

This expression is a superposition of the following genes:

$$\begin{aligned} \textit{Gene } 1\text{:}\ & 55.34-3.457\times\ln(h/d_{50}+7.542)^2 \\ & -3.547(\ln(w)/(U_f) \\ & -3.547\ln(2R_*+h/d_{50})), \end{aligned} \quad (21)$$

$$\textit{Gene } 2\text{:}\ 0.6629\sqrt{e^\theta}+0.1747\times(w/U_f)^2 + 0.6629(w/U_f)^3, \quad (22)$$

$$\textit{Gene } 3\text{:}\ \frac{9.477\times10^{-6}R_*\times(w/U_f+1)}{w/U_f}. \quad (23)$$

Traditional GP has obtained the following expression to estimate the roughness coefficient as follows:

$$\begin{aligned} \frac{k_s}{d_{50}} = {} & 3.46-\frac{R_*\sqrt{h/d_{50}}\,(w/U_f-\theta)\times7.662\times10^{-7}}{w/U_f} \\ & -R_*\left(w/U_f+5.236\right)(w/U_f-\theta) \\ & \times1.149\times10^{-6}. \end{aligned} \quad (24)$$

The error of multi-gene GP for estimating roughness coefficient is lower than the error of traditional GP. As

Table 2. Statistical characteristics of the data sets used for developing models to estimate roughness coefficient in erodible channels.

Model	Traditional GP	Multi-gene GP
Training error	1.0733	0.72681
Validation error	1.6204	0.9385

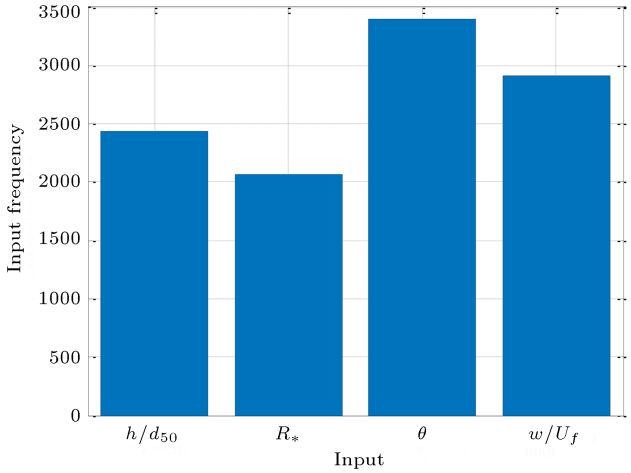

Figure 10. The frequency of input variables in the training process in multi-gene GP model.

reported in Table 2, the errors obtained by multi-gene GP for validation and training data sets are 0.9385 and 0.72681, respectively, while these errors in traditional GP for validation and training data are 1.6204 and 1.0733, respectively. This proves that the multi-gene GP model can capture the complexity of this phenomenon with higher performance than the traditional GP. In addition, Figure 10 shows the effectiveness of any input variables based on their frequency during the training process. For example, the figure shows that Shield parameter as the third input variable has been repeated near 3500 times. Therefore, it is proven that this variable plays an important role in estimating the roughness coefficient. However, repetition of other input variables proves their effectiveness in obtaining the roughness coefficient.

6.2. Development of GA-FIS estimator models

The same data sets chosen for the GP models are used to develop the GA-FIS model to estimate the roughness coefficient. As mentioned above, in the GA-FIS models, only one GA is used to optimize subtractive clustering parameters and the antecedent and consequent parameters of fuzzy IF-THEN rules [38]. The GA parameters including population number, crossover fraction, mutation coefficient, and the number of elitist chromosomes used to run the GA-FIS model are set to 400, 0.8, 0.1, and 10, respectively. These parameters were selected through a trial-and-error process. Following the selection of the GA parameters, the training

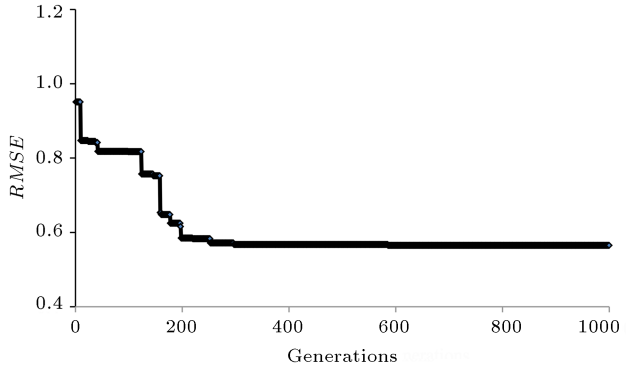

Figure 11. The training process in the GA-FIS model to estimate roughness coefficient.

process of the FIS model by the GA to estimate the roughness coefficient (GA-FIS model) is shown in Figure 11. The decreasing trend of the RMSEs for the training data sets error ensures either fair selection of input variables or the GA performance in the simultaneous tuning of subtractive clustering parameters and antecedent and consequent parameters of fuzzy IF-THEN rules. Clustering parameters in which validation and training errors are minimized simultaneously are also reported in the following expression:

$$[r_{h/d_{50}}, r_{R*}, r_{\theta}, r_{w/UU_f}, r_{k_s/d_{50}}, \gamma_{k_s/d_{50}}] =$$

$$[1.4175, 1.7952, 0.4328, 0.4376, 0.4562, 0.6427].$$

RMSEs associated with the validation and training data ses are reported in Table 3. In addition, final membership functions for input variables are reported in Figure 12. As mentioned before in this figure, each linguistic variable makes appropriate IF-THEN rules with its level (Eqs. 12 to 14). The desirable generation in which the training and validation errors are simultaneously minimized is 74, while the number of the rules appropriate by the clustering parameters is 5. As is apparent from Table 3, the errors associated with the training and validation data sets are about 0.667 and 0.863, respectively. These values prove the performance of GA-FIS model compared to the GP models based on the error estimation. However, the GP models make explicit functions easier to apply to other conditions than the implicit optimized fuzzy IF-THEN rules extracted from the GA-FIS model.

To highlight the efficiency of the GA model in

Table 3. Statistical characteristics of the data sets used for developing models to estimate roughness coefficient in erodible channels.

Model	GA-FIS	GA-ANFIS
Training error	0.667	0.864
Validation error	0.883	1.345

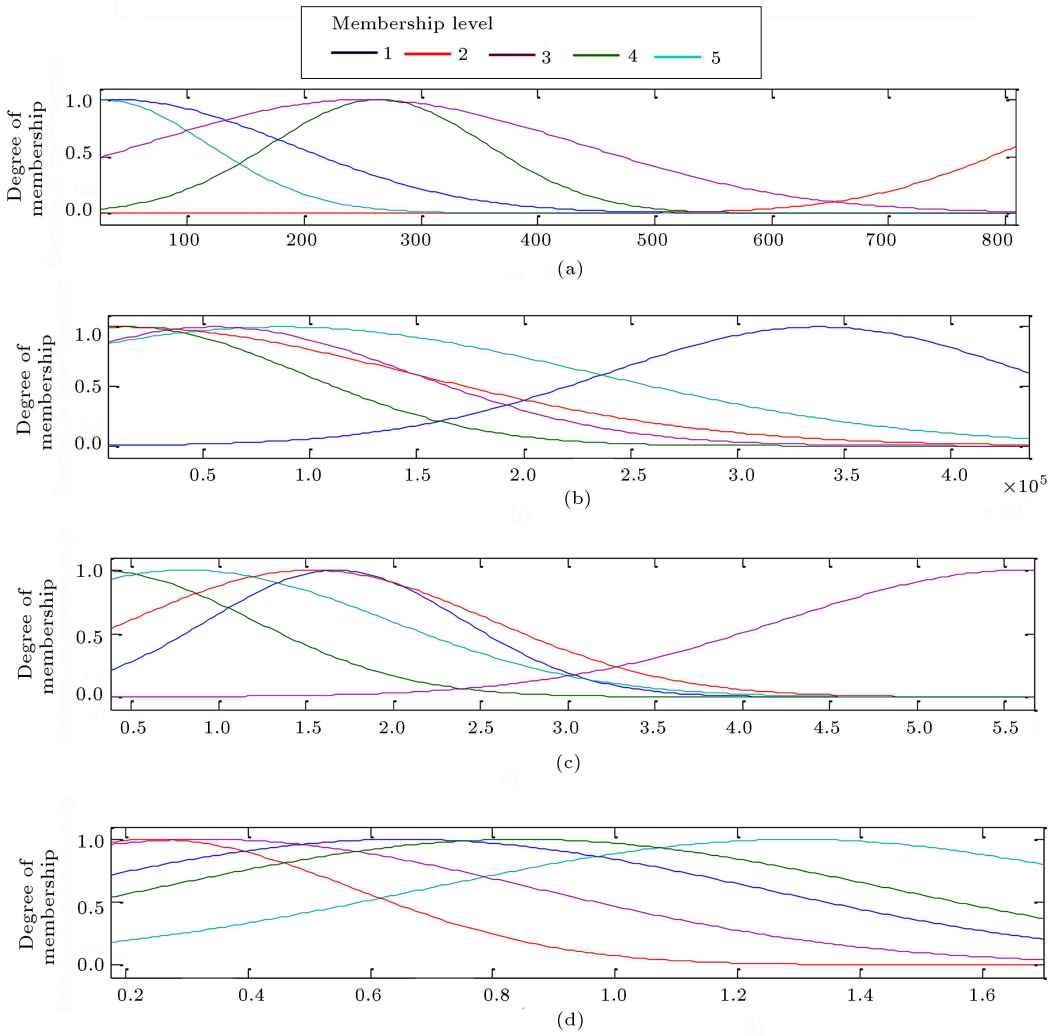

Figure 12. Final membership functions for input variables: (a) h/d_{50}, (b) R_*, (c) θ, and (d) w/U_f.

tuning fuzzy antecedent and consequent parameters in a noisy area in which scales of variables in the data set are not the same, a GA-ANFIS model is employed to estimate the roughness coefficient. In the model, the GA is only used to optimize subtractive clustering parameters, while the known ANFIS model introduced by Jang (1993) [37] is employed to tune fuzzy antecedent and consequent parameters. The GA-ANFIS model was previously employed by Zanganeh et al. (2009) to predict wave parameters. In the GA-ANFIS model, the number of training epochs for the ANFIS model is set to 500, while the GA parameters including population number, crossover fraction, mutation coefficient, and the number of elitism chromosomes are set to 100, 0.8, 10, and 0.1 population numbers, respectively. After preparing and executing the model, Figure 13 shows the training process by the GA-ANFIS model. Figure 14 shows the training process by a selected ANFIS model (in the GA-ANFIS model) randomly. As is apparent from the figure, ANFIS model cannot

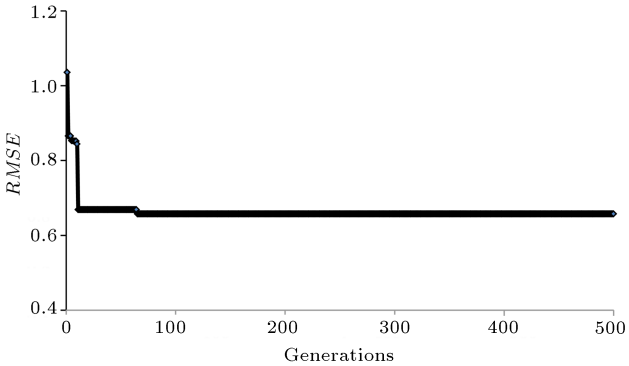

Figure 13. The training process in the GA-ANFIS model to estimate roughness coefficient.

tune the parameters in a decreasing trend. In addition, errors of GA-ANFIS models to estimate the roughness coefficient for training and validation data are 0.864 and 1.345, respectively. This shows the inefficiency of the GA-ANFIS model even in comparison with GP models, while this method does not give any explicit

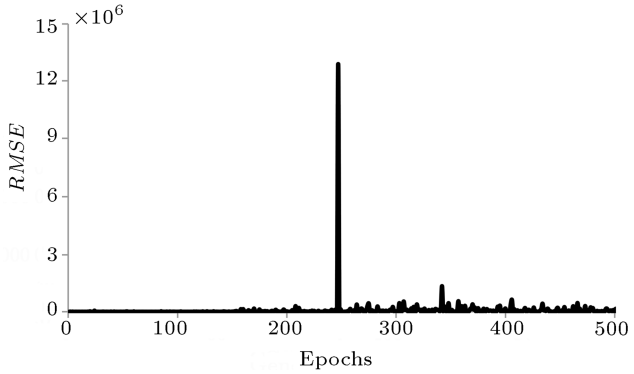

Figure 14. The training process in the ANFIS model to estimate roughness coefficient.

relationship for estimating roughness coefficient like the GA-FIS model.

6.3. Development of MLR and MNLRP models

As mentioned before, to identify linear relationships between effective input and output variables for estimating the roughness coefficient, the LSE method is used in terms of simplicity. Therefore, using the LSE method in the MATLAB, the MLR relationship for the estimation of the roughness coefficient is obtained as follows:

$$k_s/d_{50} = 2.7352 - 3.854 \times 10^{-4}\left(h/d_{50}\right) + 2.1499 \times 10^{-5} R_* - 2.252\theta - 5.7746\left(w/U_f\right). \quad (25)$$

While the MNLRP relations obtained by the LSE method to estimate the coefficient are expressed in the following:

$$k_s/d_{50} = 0.0189(h/d_{50})^{0.0348}(R_*)^{0.4436}(\theta)^{0.2601}(w/U_f)^{-0.8361}. \quad (26)$$

7. Models evaluation

Following the development of soft computing-based models, their performances are examined versus the testing data never directly used in the training process. Scatter diagrams shown in Figures 15 to 18 indicate the accuracy of GP, GA-FIS, GA-ANFIS, empirical formulas, and regression-based models by comparing the observed and estimated values. Two statistical indexes and a correlation coefficient are used to compare the models. The first index is *Bias*, expressing the mean error of the model, either over-estimating or under-estimating the observed values. The second statistical index is the RMSE. These two indexes, in addition to the correlation coefficient, are calculated by the following equations:

$$Bias = \frac{1}{N}\sum_{k=1}^{N}\left(O^k - P^k\right), \quad (27)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(O^k - P^k\right)^2}, \quad (28)$$

$$R^2 = 1 - \frac{\sum_{k=1}^{N}\left(O^k - P^k\right)^2}{\sum_{k=1}^{N}\left(\bar{O} - P^k\right)^2}, \quad (29)$$

where O^k is the observed value, P^k is the estimated value, N is number of testing data, $\bar{O}$ is the average value of the observed parameter, and *RMSE* is *Root Mean Square Error*.

Results of roughness estimator models are presented in Table 4. The table shows that the multi-gene GP model is more accurate than traditional GP and regression-based models in the estimation of roughness coefficient. In addition, the GA-FIS model estimates the roughness coefficient with acceptable accuracy, which is more accurate than the GA-ANFIS model. As apparent from Figures 15 to 18 in the estimation of roughness coefficient, the R^2 values for multi-gene GP and GA-FIS are 0.8504 and 0.8842, respectively.

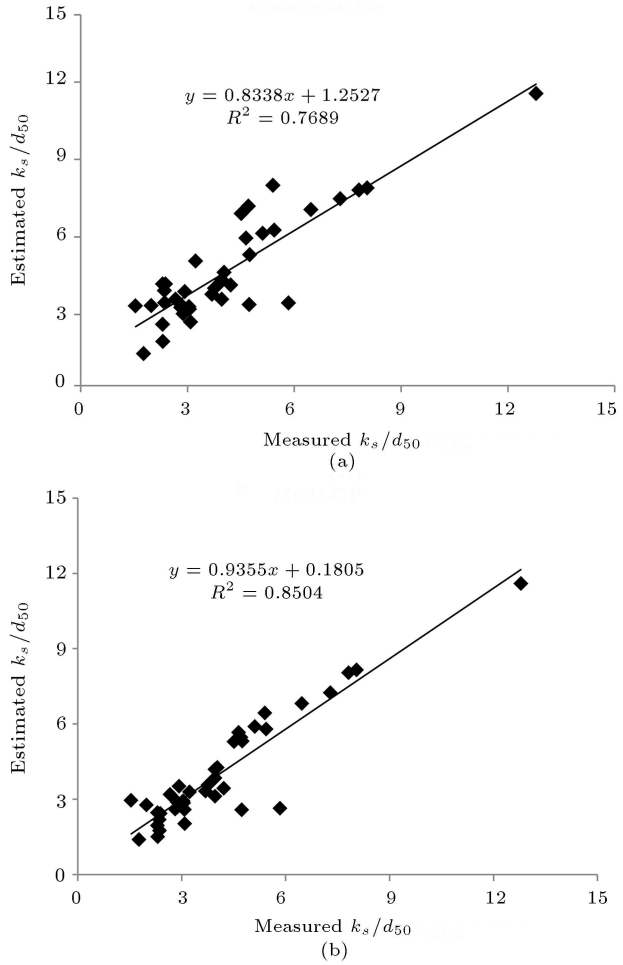

Figure 15. Measured roughness coefficients values versus estimated ones by (a) Traditional GP (1988) and (b) Multi-gene GP.

Table 4. Statistical characteristics of the models to estimate the roughness coefficient in erodible channels.

Method	RMSE	Bias
Traditional GP	0.921	–0.251
Multi-gene GP	0.821	–0.242
GA-FIS	0.775	–0.168
GA-ANFIS	1.645	0.451
MLR	2.330	1.332
MNLR	1.898	–0.523
Wilson (1988)	0.451	2.925
Yalin (1992)	1.423	3.466
Summer et al. (1996)	0.451	3.335

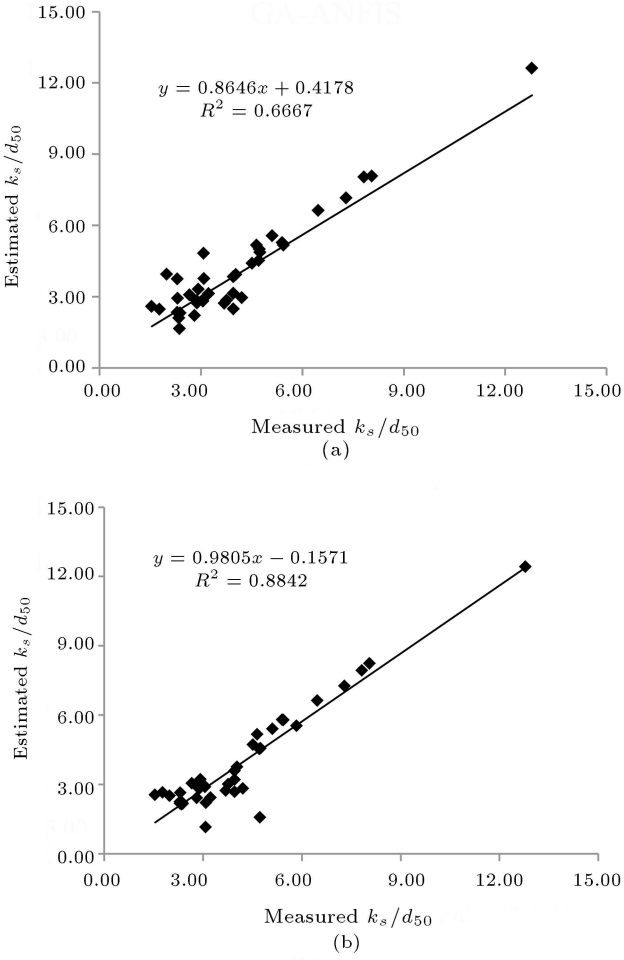

Figure 16. Measured roughness coefficients values versus estimated the ones by FIS-based model: (a) GA-ANFIS and (b) GA-FIS.

In addition, greater accuracy of the GA-FIS model than the GA-ANFIS model indicates the preference of GA over the ANFIS model in tuneing fuzzy antecedent and consequent parameters. The main reasons for that issue can be related to the GA searching skill in a noisy environment.

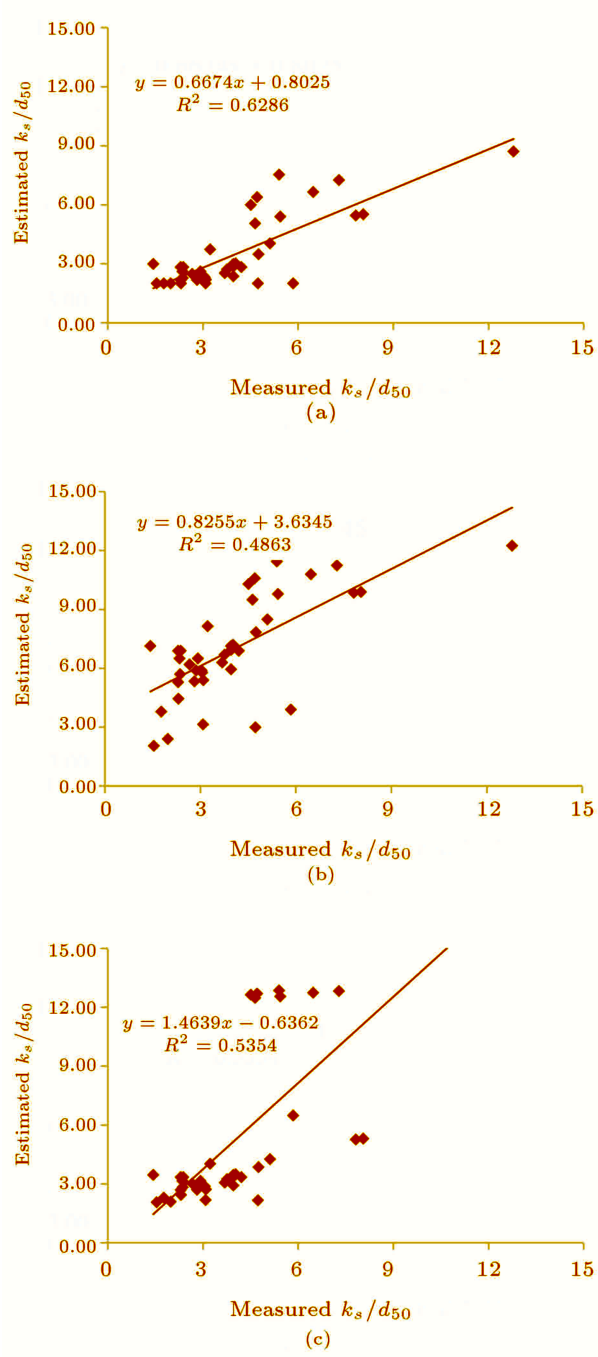

Figure 17. Measured roughness coefficients values versus the ones estimated by empirical formulas (a) Wilson (1988), (b) Yalin (1992), and (c) Sumer et al. (1996).

Zanganeh and Rastegar (2020) [21] applied ANN and ANFIS models to estimate the roughness coefficient in the erodible channel by the same data set. They used both normalized and real data sets to estimate roughness coefficients in models. The main deficiency in the application of their models was extracting an explicit relationship to estimate the roughness coefficient. On the other hand, the model application of the normalized data set may change

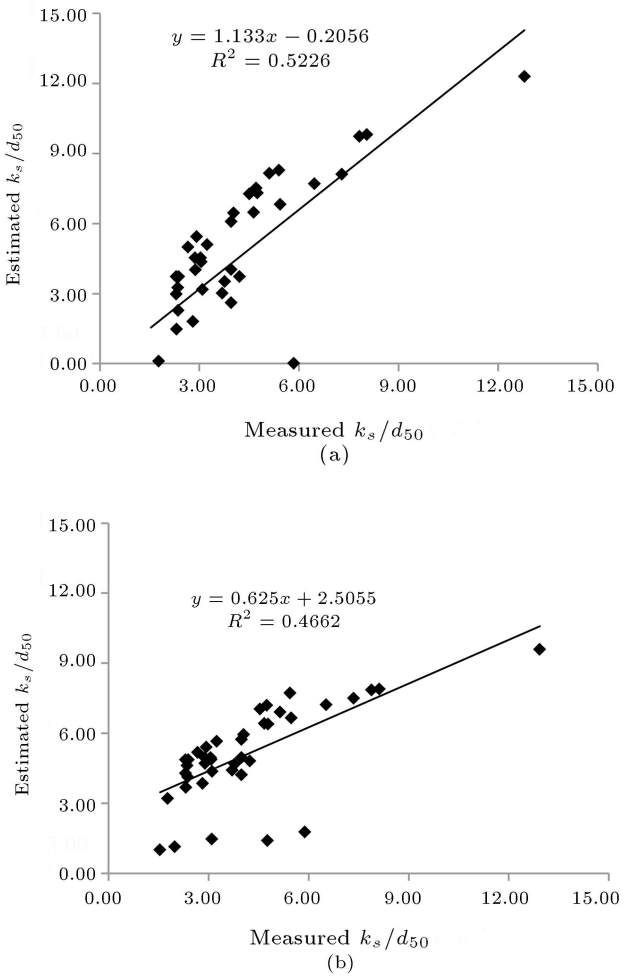

Figure 18. Measured roughness coefficients values versus the ones estimated by multi-regression methods: (a) MLR and (b) MNLR.

the physical condition of the phenomena by changing effective parameters on the roughness coefficient. The best correlation coefficient for the models was 0.8658, while the correlation coefficient was 0.8842 for the GA-FIS model. In addition, a useful relationship extracted by the multi-GP model is applicable to the same physical conditions.

8. Summary and conclusion

Scaling problems in roughness coefficient estimation in erodible open channels may affect hydraulic conditions of channels. To deal with this problem, in this paper, so-called multi-gene Genetic Programming (GP), combination of Genetic Algorithm and Fuzzy Inference System (GA-FIS) model, and Multiple Regression (MR) methods are employed to extract either an explicit or implicit relationship between the roughness coefficient and input variables involved in the coefficient value. In addition, traditional GP and conventional empirical formulas are applied to evaluate the models. Results show that the employed methods are more accurate than empirical methods. At the same time, other parameters like non-dimensional water depth and shear Reynolds number are recognized as variables affecting the roughness coefficient. Results show the GA performance to optimize an Fuzzy Inference System (FIS) compared with gradient-based models like the ANFIS when the scale of input variables is not in the same order (scaling problem). Correlation coefficients for multi-gene GP and GA-FIS are 0.8504 and 0.8842, respectively, while this value for the most accurate empirical method [7] is 0.6286.

Other implications and obtained results can be reported as follows:

- The multi-gene GP model extracts a suitable relationship for estimating the roughness coefficient in an erodible channel that can work in different hydraulic conditions;
- The GA-ANFIS model does not have reasonable accuracy despite errors in training and validation data sets in the testing data. This proves that decreasing the training trend is more important than training and validation data values;
- Empirical formula introduced by Sumer et al. (1996) [1] has difficulty with the application given the criteria to differentiate various hydraulic conditions.

Like other methods, the applied models in this paper are subject to some limitations. The main limitation of the model application is limited hydraulic conditions while this model is not properly applicable in bending channels. On the other hand, the limited hydraulic condition may affect the model generalization capability and it is necessary to study this model further in future works.

Acknowledgment

The author would like to sincerely appreciate the Deputy of Research at Golestan University (GU) for their support over the course of finalizing this research.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Sumer, B.M., Kozakiewicz, A., Fredsøe, J., et al. "Velocity and concentration profiles in sheet-flow layer of movable bed", *Journal of Hydraulic Engineering*, **122**(10), pp. 549–558 (1996).
2. Ackers, P. and White, W.R. "Sediment transport: new approach and analysis", *Journal of the Hydraulics*

Division, **99**(hy11), pp. 2041–2060 (1973).

3. Simons, D.B. and Richardson, E.V. "Resistance to flow in alluvial channels", US Government Printing Office (1996).

4. Hammond, F.D., Heathershaw, A.D., and Langhorne, D.N. "A comparison between shields threshold criterion and the Henderson movement of loosely packed gravel in a tidal channel", *Sedimentology*, **31**, pp. 51–62 (1984).

5. Colosimo, C., Copertino, V.A., and Veltri, M. "Average velocity estimation in gravel-bed rivers", In *Proc., 5th IAHR-APD Congress*, pp. 1–15 (1986).

6. Wilson, K.C. "Mobile-bed friction at high shear stress", *Journal of Hydraulic Engineering*, **115**(6), pp. 825–830 (1989).

7. Yalin, M.S., *River Mechanics*, Pergamon, First Edition (1992).

8. Shayya, W.H. and Sablani, S.S. "An artificial neural network for non-iterative calculation of the friction factor in pipeline flow", *J. of Computers and Electronics in Agriculture*, **21**(3), pp. 219–228 (1998).

9. Abdeen, M.A.M. "Artificial neutral network model for predicting the impact changing water structures' locations on the hydraulic performance of branched open channel system", *J. of Mechanics and Mechanical Engineering*, **7**(2), pp. 179–192 (2004).

10. Yang, H.C. and Chang, F.J. "Modelling combined open channel flow by artificial neural networks", *Hydrological Processes: An International Journal*, **19**(18), pp. 3747–3762 (2005).

11. Yuhong, Z. and Wenxin, H. "Application of artificial neural network to predict the friction factor of open channel flow", *J. of Communications in Nonlinear Science and Numerical Simulation*, **14**(5), pp. 2373–2378 (2009).

12. Zahiri, A. and Dehghani, A.A. "Flow discharge determination in straight compound channels using ANN", *J. of World Academy of Science, Engineering and Technology*, **58**, pp. 12–15 (2009).

13. Bateni, S.M., Borghei, S.M., and Jeng, D.S. "Neural network and neuro-fuzzy assessments for scour depth around bridge piers", *J. of Engineering Applications of Artificial Intelligence*, **20**(3), pp. 401–414 (2007).

14. Begum, S.A., Fujail, A.M., and Barbhuiya, A.K. "Artificial neural network to predict equilibrium local scour depth around semicircular bridge abutments", *6th SASTech*, Malaysia, Kuala Lumpur (2012).

15. Ghazanfari-Hashemi, S., Etemad-Shahidi, A., Kazeminezhad, M.H., et al. "Prediction of pile group scour in waves using support vector machines and ANN", *J. of Hydroinformatics*, **13**(4), pp. 609–620 (2011).

16. Kazeminezhad, M.H., Etemad-Shahidi, A., and Bakhtiary, A.Y. "An alternative approach for investigation of the wave-induced scour around pipelines", *J. of Hydroinformatics*, **12**(1), pp. 51–65 (2010).

17. Zanganeh, M., Yeganeh-Bakhtiary, A., and Bakhtyar, R. "Combined particle swarm optimization and fuzzy inference system model for estimation of current-induced scour beneath marine pipelines", *J. of Hydroinformatics*, **13**(3), pp. 558–573 (2011).

18. Zanganeh, M., Yeganeh-Bakhtiary, A., and Yamashita, T. "ANFIS and ANN models for the estimation of wind and wave-induced current velocities at Joeutsu-Ogata coast", *J. of Hydroinformatics*, **18**(2), pp. 371–391 (2016).

19. Azamathulla, H.M. and Ahmad, Z. "An expert system for predicting Manning's roughness coefficient in open channels by using gene expression programming", *J. of Neural Computing and Applications*, **23**(5), pp. 1343–1349 (2013).

20. Roushangar, K., Saghebian, S.M., Kirca, V.S., et al. "Prediction of form roughness coefficient in alluvial channels using efficient hybrid approaches", *J. of Soft Computing*, **24**(24), pp. 18531–18543 (2020).

21. Zanganeh, M. and Rastegar, A. "Estimation of roughness coefficient in erodible channels by ANNs and the ANFIS methods", *Amirkabir Journal of Civil Engineering*, **52**(2), pp. 495–512 (2020).

22. Deo, O., Jothiprakash, V., and Deo, M.C. "Genetic programming to predict spillway scour", *Int. J. of Tomography and Statistics*, **8**, pp. 32–46 (2008).

23. Azamathulla, H.M., Ghani, A.A., Zakaria, N.A., et al. "Genetic programming to predict ski-jump bucket spillway scour", *J. of Hydrodynamics, Ser. B.*, **20**(4), pp. 477–484 (2008).

24. Guven, A. and Kişi, Ö. "Estimation of suspended sediment yield in natural rivers using machine-coded linear genetic programming", *J. of Water Resources Management*, **25**(2), pp. 691–704 (2011).

25. Azamathulla, H.M., Ghani, A.A., Zakaria, N.A., et al. "Genetic programming to predict bridge pier scour", *J. of Hydraulic Engineering*, **136**(3), pp. 165–169 (2009).

26. Azamathulla, H.M., Guven, A., and Demir, Y.K. "Linear genetic programming to scour below submerged pipeline", *Ocean Engineering*, **38**(8), pp. 995–1000 (2011).

27. Najafzadeh, M. and Barani, G.A. "Comparison of group method of data handling based genetic programming and back propagation systems to predict scour depth around bridge piers", *Scientia Iranica*, **18**(6), pp. 1207–1213 (2011).

28. Koç, M.L., Balas, C.E., and Koç, D.İ. "Stability assessment of rubble-mound breakwaters using genetic programming", *J. of Ocean Engineering*, **111**, pp. 8–12 (2016).

29. Heřmanovský, M., Havlíček, V., Hanel, M., et al. "Regionalization of runoff models derived by genetic programming", *Journal of Hydrology*, **547**, pp. 544–556 (2017).
30. Assimi, H., Jamali, A., and Nariman-zadeh, N. "Sizing and topology optimization of truss structures using genetic programming", *J. of Swarm and Evolutionary Computation*, **37**, pp. 90–103 (2017).
31. Zanganeh, M. "Simultaneous optimization of clustering and fuzzy IF-THEN rules parameters by the genetic algorithm in fuzzy inference system-based wave predictor models", *J. of Hydroinformatics*, **19**(3), pp. 385–404 (2017).
32. Zanganeh, M. "Improvement of the ANFIS-based wave predictor models by the particle swarm optimization", *J. of Ocean Engineering and Science*, **5**(1), pp. 84–99 (2020).
33. Koza, J.R. "Genetic programming: on the programming of computers by means of natural selection", **1**, MIT press (1992).
34. Searson, D.P., Leahy, D.E., and Willis, M.J. "GPTIPS: an open source genetic programming toolbox for multigene symbolic regression", In *Proceedings of the International Multiconference of Engineers and Computer Scientists*, **1**, pp. 77–80 (2010).
35. Zanaganeh, M., Mousavi, S.J., and Shahidi, A.F.E. "A hybrid genetic algorithm-adaptive network-based fuzzy inference system in prediction of wave parameters", *Engineering Applications of Artificial Intelligence*, **22**(8), pp. 1194–1202 (2009).
36. Chiu, S.L. "Fuzzy model identification based on cluster estimation", *J. of Intelligent Fuzzy Systems*, **2**, pp. 267–278 (1994).
37. Jang, J.S.R. "ANFIS adaptive-network-based fuzzy inference systems", *IEEE Trans System Man Cybern*, **23**(3), pp. 665–685 (1993).
38. Chipperfield, A., Fleming, P., Pohlheim, H., et al. Genetic algorithm toolbox for use with MATLAB (1994).

Appendix A

Table A.1 shows the data used in this research. In this table, all the used input and output data are shown.

Table A.1. Gathered data sets belonging to Sumer et al. (1996) [1].

No.	h/d_{50}	R^*	θ	w/U_f	k_s/d_{50}
1	28.37	366900	1.88	0.97	7.47
2	28.83	375000	1.97	0.95	8.65
3	28.08	187200	1.43	1.01	1.43
4	188.33	10260	0.38	1.17	1.72
5	33.83	436200	2.66	0.82	14.29
6	30.00	327900	1.50	1.09	3.72
7	188.33	11160	0.45	1.08	3.35
8	808.46	8918	2.24	0.17	4.27
9	25.00	144300	0.86	1.32	2.48
10	224.50	39540	5.67	0.24	17.42
11	33.33	210000	0.67	1.70	2.46
12	797.69	5434	0.83	0.29	3.42
13	31.70	386100	2.08	0.92	10.59
14	29.83	298800	1.25	1.19	3.09
15	30.17	393600	2.16	0.91	8.84
16	30.57	385500	2.08	0.93	8.50
17	30.83	401700	2.26	0.89	10.11
18	31.33	400200	2.24	0.89	10.42
19	32.07	411300	2.36	0.87	12.15
20	32.13	419100	2.46	0.85	12.63
21	32.63	404100	2.28	0.88	11.31
22	32.53	404400	2.29	0.88	10.22
23	33.60	424200	2.52	0.84	13.17
24	26.54	195780	1.59	0.97	5.07
25	26.27	196300	1.60	0.97	4.17
26	26.46	202540	1.70	0.94	4.27
27	26.92	208780	1.81	0.91	4.67
28	28.85	197600	1.62	0.96	3.25
29	28.27	226200	2.12	0.84	6.81
30	27.69	221520	2.03	0.86	5.32
31	28.69	233220	2.25	0.81	7.29
32	28.46	236860	2.32	0.80	7.06
33	29.00	237380	2.34	0.80	7.86
34	29.62	244920	2.49	0.77	9.29
35	29.81	250120	2.59	0.76	10.10
36	29.46	249080	2.57	0.76	9.13
37	29.23	218400	1.98	0.87	4.97
38	30.38	255320	2.70	0.74	10.87
39	30.15	249340	2.58	0.76	9.46
40	30.31	258960	2.78	0.73	10.62
41	30.19	258440	2.77	0.73	10.04
42	31.15	264940	2.91	0.72	11.84
43	31.54	263900	2.88	0.72	11.18
44	32.31	269360	3.00	0.70	11.79
45	32.50	276640	3.17	0.69	12.17
46	33.46	286780	3.41	0.662	14.36
47	34.88	296920	3.65	0.639	17.76
48	34.62	297440	3.66	0.638	16.06
49	34.81	290160	3.49	0.654	15.07
50	34.23	302900	3.8	0.627	14.97
51	35.35	302900	3.8	0.627	17.17

Table A.1. Gathered data sets belonging to Sumer et al. (1996) [1] (continued).

No.	h/d_{50}	R^*	θ	w/U_f	k_s/d_{50}
52	35.65	320580	4.26	0.592	20.58
53	35.81	309660	3.97	0.613	18.49
54	35.50	311220	4.01	0.610	17.04
55	39.23	192400	1.55	0.986	2.76
56	35.62	328120	4.46	0.578	20.36
57	39.62	200200	1.64	0.948	3.12
58	36.27	333580	4.61	0.569	21.07
59	40.00	213200	1.88	0.890	4
60	41.15	213200	1.89	0.890	3.76
61	40.38	228800	2.18	0.830	5.1
62	41.15	240500	2.4	0.789	5.4
63	42.31	247000	2.54	0.768	6
64	190.00	12840	0.6	0.342	4.73
65	188.33	16140	0.95	0.322	6.98
66	186.00	16740	1.02	0.333	6.03
67	190.83	15960	0.93	0.337	4.35
68	185.83	17460	1.1	0.337	5.3
69	188.33	17340	1.09	0.338	4.31
70	187.50	19500	1.38	0.327	6.01
71	190.00	18240	1.21	0.336	4.13
72	188.00	20880	1.58	0.326	5.65
73	191.67	20820	1.58	0.322	5.05
74	198.00	22260	1.8	0.300	7.59
75	180.83	24240	2.13	0.330	5.88
76	189.17	23220	1.95	0.324	5.37
77	191.67	23340	1.97	0.317	5.29
78	193.33	25440	2.35	0.312	5.87
79	192.50	24900	2.25	0.319	4.56
80	198.17	24960	2.26	0.317	4.3
81	195.17	27300	2.71	0.304	6.52
82	196.67	27420	2.73	0.304	5.68
83	204.67	28680	2.98	0.282	8.65
84	201.67	29640	3.19	0.288	7.9
85	201.00	30780	3.44	0.283	9.38
86	200.00	29520	3.16	0.295	6.41
87	201.83	30900	3.46	0.287	7.84
88	205.00	31680	3.64	0.284	7.18
89	207.17	32940	3.94	0.277	8.57
90	214.50	34440	4.31	0.266	9.24
91	213.50	35880	4.68	0.261	10.92
92	218.50	36840	4.92	0.252	12.49
93	220.33	37680	5.16	0.248	13.38
94	231.50	35100	4.48	0.250	9.54
95	803.85	5590	0.88	0.279	3.6

Table A.1. Gathered data sets belonging to Sumer et al. (1996) [1] (continued).

No.	h/d_{50}	R^*	θ	w/U_f	k_s/d_{50}
96	793.85	6253	1.1	0.249	4.88
97	795.38	6175	1.07	0.253	4.39
98	790.77	6682	1.26	0.233	5.68
99	781.54	6916	1.35	0.226	5.78
100	784.62	7046	1.4	0.221	4.77
101	778.46	7072	1.41	0.221	4.39
102	779.23	7514	1.59	0.208	5.9
103	775.38	7605	1.62	0.205	4.58
104	776.15	7735	1.69	0.202	5.22
105	775.38	7969	1.78	0.196	4.82
106	776.92	8034	1.81	0.194	4.33
107	793.08	8177	1.88	0.191	5.56
108	792.31	8463	2.02	0.184	5.61
109	804.62	8801	2.18	0.177	6.51
110	798.46	8515	2.04	0.183	3.69
111	800.00	8749	2.15	0.178	4.04
112	26.67	270000	1.02	1.322	2.39
113	27.30	282000	1.11	1.266	2.83
114	28.33	270900	1.03	1.318	2.88
115	33.33	240000	0.87	1.488	1.94
116	28.50	270900	1.03	1.318	2.72
117	28.67	274800	1.06	1.299	2.78
118	27.07	291600	1.19	1.224	2.59
119	28.67	282300	1.12	1.265	2.88
120	27.77	315000	1.39	1.133	3.96
121	28.93	287700	1.16	1.241	3.06
122	33.33	225000	0.76	1.587	1.77
123	27.70	334800	1.57	1.066	4.75
124	29.13	290400	1.18	1.229	3.04
125	28.50	376500	1.98	0.948	8.05
126	28.67	375000	1.97	0.952	7.82
127	29.63	290400	1.18	1.229	2.87
128	33.33	243000	0.89	1.469	2.31
129	29.47	304800	1.3	1.171	2.92
130	29.53	297600	1.24	1.200	2.66
131	30.17	321000	1.44	1.112	4.03
132	33.33	267000	1.08	1.337	3.09
133	30.80	348900	1.7	1.023	5.11
134	33.33	264000	1.06	1.352	2.3
135	32.57	418200	2.45	0.854	12.79
136	25.00	169780	1.19	1.118	3.96
137	25.00	174200	1.26	1.090	3.69
138	25.23	179920	1.34	1.055	3.76
139	27.69	182000	1.38	1.043	4.21

Table A.1. Gathered data sets belonging to Sumer et al. (1996) [1] (continued).

No.	h/d_{50}	R^*	θ	w/U_f	k_s/d_{50}
140	25.58	185640	1.43	1.022	3.96
141	26.92	161200	1.07	1.177	2.81
142	27.04	213980	1.9	0.887	4.64
143	28.23	228280	2.16	0.831	6.47
144	28.69	233220	2.25	0.814	7.29
145	28.85	218400	1.96	0.869	5.44
146	38.46	166400	1.14	1.141	2.36
147	38.85	176800	1.3	1.074	2.35
148	38.85	182000	1.38	1.043	2.3
149	39.23	182000	1.38	1.043	2.38
150	40.38	197600	1.63	0.961	3.23
151	40.38	223600	2.06	0.849	4.51
152	41.92	226200	2.12	0.839	4.71
153	40.77	235300	2.29	0.807	5.4
154	186.67	10680	0.41	1.124	1.54
155	190.00	12840	0.6	0.935	4.73
156	188.33	11460	0.48	1.047	1.98
157	184.33	14640	0.78	0.820	5.84
158	188.33	13200	0.63	0.909	3.08

Biography

Morteza Zanganeh has more than 11 years of academic experience and his major field of research is Hydraulic Engineering and Hydroinformatics. He has conducted much research on the modeling of waves and their interaction with hydraulic processes including brine discharge, pump station, and so on using numerical modeling and artificial intelligence models. This has led to several peer-reviewed publications (in Applied Ocean Research, Engineering Application of Artificial Intelligence, ASME, Ocean Engineering and Science, and Journal of Hydroinformatics). He has been a faculty member at Golestan University in Iran for 10 years.