ORIGINAL PAPER

Generalized structure of group method of data handling to prognosticate scour around various cross-vane structures

Ebrahim Shahbazbeygi[1] · Fariborz Yosefvand[1] · Behrouz Yaghoubi[1] · Saeid Shabanlou[1] · Ahmad Rajabi[1]

Abstract

In this study, for the first time, the scour pattern around cross-vane structures with I, U, and J shapes in bending channels is simulated by a new artificial intelligence method called "generalized structures group method of data handling" (GSGMDH). Compared to the group method of data handling (GMDH), the GSGMDH method has more flexibility and exactness so that nodes can derive inputs from non-adjacent layers. Initially, using the input parameters, six different models are defined for each of the GMDH and GSGMDH methods. Also, the superior model forecasts objective function values with acceptable accuracy. For example, the correlation coefficient (R), the scatter index (SI), and the Nash-Sutcliffe efficiency coefficient (NSC) for the GSGMDH superior model in the training mode are approximated to be 0.913, 0.214, and 0.800, respectively. Based on the sensitivity analysis, the shape factor of cross-vane structures, the ratio of the difference between the upstream and downstream flow depths to the height of the structure, and the densimetric Froude number are introduced as the most effective input parameters. Subsequently, for the superior model, a relationship is proposed for the estimation and calculation of the scour depth. Finally, a partial derivative sensitivity analysis (PDSA) is conducted for the superior model.

Keywords Cross-vane structures · Scour · Modeling · Generalized structures group method of data handling · Densimetric Froude number

Introduction

In general, waterways have played a key duty in the development of civilizations. For example, major cities and industrial hubs have been historically created adjacent to the Yangtze, Nile, Mississippi, and Amazon rivers. The protection of the banks and riverbeds against erosion and scour is one of the most important issues in river engineering. There are several techniques used to protect riverbeds and banks against scour. The utilization of stone cross-section structures is considered a viable solution to protect the bed of canals and rivers. Scour is a three-dimensional phenomenon in which erodible materials such as gravel and sand are removed from around the structures. Commonly, scour is one of the most important reasons for the

failure of infrastructures such as piers, road bridge piles, and abutments. Due to the complete consistency of stone with the structure of rivers, structures with stone cross-section are more popular than other riverbed protective structures. Using them, the fish pathway is fully maintained in all parts of the flow. Additionally, the construction reduction and saving costs compared with common structures are one of the most important advantages of cross-section structures. Thus, thanks to the significance of such protective constructions, numerous laboratory and theoretical investigations have been done on the flow pattern in their vicinity. For instance, some criteria for designing stone cross-vane structures with different forms such as U, W, and J shapes were suggested by Rosgen (2001). Using the shear stress concept, he put forward several equations for calculating the minimum dimension of aggregates. Also, Scurlock et al. (2011a) did a laboratory-analytical survey to assess energy drop occurring on U-shaped cross-vane structures. To simulate energy drop values, they utilized the HEC-RAS numerical model and verified the results of this model with their experimental model data. Besides, the scour pattern around these constructions with U, W, and A forms in straight channels was measured by Scurlock et al. (2011b). They proposed several

---

Responsible Editor: Broder J. Merkel

✉ Fariborz Yosefvand
  fariborzyosefvand@gmail.com

[1] Department of Water Engineering, Kermanshah Branch, Islamic Azad University, Kermanshah, Iran

formulae to predict the scour depth. Scurlock et al. 2011a, b through a laboratory investigation measured the scour values near the stone weirs sequentially installed at the bend locations of an open conduit. They evaluated the results and concluded several equations for computing maximum velocity variations at the inner bank, outer bank, and centerline of the conduit. Moreover, Pagliara et al. (2013a) measured scour hole dimensions near J-Hook vanes located in an open rectangular channel in clear-water conditions. They proposed several equations with reasonable accuracy to calculate the maximum scour depth. Subsequently, Pagliara and Kurdistani (2013) performed an experiment to measure scour values downstream of these constructions located in an open r channel. Through the conduction of dimensional analyses, they introduced the variables affecting the scour parameter and indicated that the sedimentary bed morphology changes by hydraulic conditions. Pagliara et al. (2013b) measured the dimensions of scour hole in the vicinity of W-shaped constructions through a laboratory program. Through a dimensional analysis, they suggested several relationships for estimating the maximum length and depth of the scour hole. Furthermore, the authors analyzed the obtained results by the above relationships with the laboratory measurements to exhibit their accuracy. Kurdistani and Pagliara (2017) through a laboratory investigation compared the scour values around different protective constructions such as Log-Deflectors and Log-Vanes. Using the Buckingham theorem, they manage to conduct a dimensional analysis on the variables affecting the scour values. In addition, Pagliara et al. (2015a) performed an experiment on the scour pattern around Log-Vanes made of wood. These authors scaled the scour hole depth in various geometric and hydraulic circumstances and stated that the tailwater depth was a significant hydraulic factor affecting the formation of the scour hole dimension. It is worth mentioning that Pagliara and Kurdistani (2015) assessed the scour hole values at the downstream of J-Hook constructions located in bending channels in clear-water conditions. They expressed that different types of morphologies are formed downstream of these structures according to flow hydraulic conditions. Besides, Pagliara et al. (2015b) carried out an experimental program to calculate scour values around log-deflectors installed in horizontal straight channels in clear-water conditions. They evaluated such protective structures for various geometric and hydraulic conditions. Furthermore, Pagliara et al. (2016) evaluated the scour resulted from the existence of rock sills in open horizontal and bending channels. They conducted several experiments in a horizontal channel in clear-water conditions. They proposed several useful empirical formulae for assessing the bed morphology. Furthermore, Pagliara and Kurdistani (2017) experimentally investigated scour values around log-frame deflectors. They measured the scour values near these structures for various layouts of protective structures. Subsequently, Kurdistani and Pagliara (2017) experimentally evaluated the scour pattern near U-shaped constructions

installed in bending channels. By analyzing the experimental data, they divided the scour occurring in the vicinity of such structures into three different categories. Khosronejad et al. (2018) utilized a numerical model to simulate the flow pattern and the morphology of sedimentary beds in bending channels with rock vanes. To simulate the flow field, they used a model regarding the Reynolds-averaged Navier-Stokes principal in unsteady conditions. The authors validated the numerical results with the observational measurements.

Recently, various artificial intelligence (AI) algorithms have been employed to model various phenomena. Besides, such numerical models have noticeable abilities in estimating and simulating scour hole dimensions near different constructions such as piers, pipes, abutments, and submerged vanes (Azamathulla et al. 2009; Amini et al. 2011; Ebtehaj et al. 2015; Khoshbin et al. 2016; Shabanlou et al. 2018; Azimi et al. 2017, 2018, 2019a, b; Azimi and Shiri 2020). Azimi et al. (2020) simulated the scour depth around cross-vane structures through the outlier robust extreme learning machine (ORELM). The authors suggested a set of matrices to estimate the subject value.

Reviewing the literature reveals that due to the importance of the protection of banks and beds of rivers, different experimental and analytical investigations have been conducted on the scour modeling near cross-vane constructions with various forms. However, the scour depth around these constructions with I, J, and U forms installed in bending channels has not been investigated using the artificial intelligence approaches and soft computing techniques so far. So, in the current paper, the scour depth around cross-vane constructions with I, J, and U forms in bending channels is innovationally simulated by a modern artificial intelligence technique entitled "generalized structures group method of data handling (GSGMDH)." Compared to the group method of data handling (GMDH) method, this new technique has more flexibility and accuracy so that the middle nodes can receive inputs from non-adjacent nodes. By simulating the scour depth, the superior model is introduced and the most influencing inputs are detected. Eventually, a formula is presented for approximating the scour depth.

Method

The numerical models used in this study are described. Initially, the group method of data handling (GMDH) is expressed and then the new generalized structures group method of data handling (GSGMDH) is thoroughly assessed.

Group method of data handling

The GMDH model with one output and a set of inputs and is a subset of the base function elements expressed as follows:

$$Y(x_1, \ldots, x_n) = a_a \sum_{i=1}^{m} \left(a_i - f_{ij} \right) \tag{1}$$

where f denotes the initial function depending on different functions of the inputs, a represent factors, and m is the number of base function elements.

To solve the problem, the GMDH has several solutions called "partial models" or "partial derivatives." The factors of this model are reckoned using the "least square" methodology. The GMSH algorithm enhances several partial models and finds a topology of GMDH with a desirable complication shown with at least one external criterion among them. The process is called the self-organizing model. It can be said that the GMDH method is considered a self-organizing methodology that gradually creates more sophisticated issues during the evaluation of the performance of output and input values. The GMDH algorithm is used to evaluate the environmental behaviors of highly complex nonlinear problems. The major idea of GMDH is to establish an analytic method based on a progressive network according to binomial transition functions. Its structure is in the form of a progressive multilayer network which is similar to artificial neural networks, except that, unlike neural networks, the number of neurons and layers are unknown and they are determined during the training process. Ivakhnenko (1976), using the Kolmogorov-Gabor polynomial, developed the GMDH theory and expressed the relationship between input and output parameters of each system using the Voltra series that is identical to the Kolmogorov-Gabor discrete polynomial.

Mathematical basis of GMDH algorithm

Suppose there are some m variables such as x1, x2, ..., xm and a parameter y as the objective parameter. There is also data for each of the xi and the objective variable y for each of training and test modes. It means that each factor is a vector containing different numbers associated with the parameter. The initial data required to be collected for the structure of the GMDH are a combination of n values represented by the following matrix:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \\ \cdots \\ \vdots \\ y_n \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & & \text{Training dataset} & \\ x_{k1} & x_{k2} & \cdots & x_{km} \\ \cdots & \cdots & \cdots & \cdots \\ \vdots & & \text{Testing dataset} & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}$$

In this matrix, the total number of data is represented by n, so that k (k<n) samples are used for learning data and n-k samples for testing data. To start the algorithm, we encounter two problems: (1) Identifying a relationship that produces the output variable based on the input variables. (2) Prediction of y for known parameters of xi, that is, model identification and the relationship between variables (simulated) is needed to model values of the target function from that model (prediction). The principal of the GMDH is the procedure of creating a high-order polynomial known as the Volterra functional series provided as follows:

$$y = a_0 + \sum_{i=1}^{m} a_i x_i + \sum_{i=1}^{m} \sum_{j=1}^{m} a_{ij} x_i x_j + \sum_{i=1}^{m} \sum_{j=1}^{m} \sum_{k=1}^{m} a_{ijk} x_i x_j x_k$$
$$+ \cdots \tag{2}$$

To this end, in the GMDH algorithm, initially, the Volterra functional series is decomposed into multivariate quadratic polynomials:

$$y(x_i, x_j) = a_0 + a_1 x_i + a_2 x_j + a_3 x_i^2 + a_4 x_j^2 + a_5 x_i x_j \tag{3}$$

In this equation, the Volterra series is transformed into some chain recursive formulas, so that by the algebraic placement each of the recursive equations in that relationship (Volterra series) is created once again. So, the following equation:

$$y_i = f(x_{i1}, x_{i2}, x_{i3}, \ldots, x_{im}); i = 1, 2, 3, \ldots, n \tag{4}$$

is estimated by the function $\widehat{f}$:

$$y_i = \widehat{f}(x_{i1}, x_{i2}, x_{i3}, \ldots, x_{im}); i = 1, 2, 3, \ldots, n \tag{5}$$

If the function $\widehat{f}$ is expressed as Eq. 2, the function $\widehat{f}$ is written as below:

$$\begin{array}{llll} \widehat{y}_k = G(u_i, u_j) & i, j = 1, 2 \, (i \neq j) & & k = 1 \\ u_k = G(s_i, s_j) & i, j = 1, 2, K, F_1 \, (i \neq j) & F_1 \leq C_{F_2}^2 & k = 2 \\ s_k = G(p_i, p_j) & i, j = 1, 2, K, F_2 \, (i \neq j) & F_2 \leq C_{F_3}^2 & k = 3 \\ \vdots & \vdots & \vdots & \vdots \\ z_k = G(w_i, w_j) & i, j = 1, 2, K, F_l \, (i \neq j) & F_l \leq C_m^2 & k = F_{l-1} \\ w_k = G(x_i, x_j) & i, j = 1, 2, K, m \, (i \neq j) & & k = F_l \end{array} \tag{6}$$

The order of these relationships from top to bottom is a display of the decomposition steps of Eq. 4 into quadratic polynomials. On the other hand, the order of these relationships from bottom to top indicates the completion of Eq. 4 using returning equations. Indeed, the aim of the algorithm is to detect the unknown factor "a" in the Volterra function series. The resulted partial models own a structure identical to the below function:

$$\widehat{f}(x_i, x_j) = v_0 + v_1 x_i + v_2 x_j + v_3 x_i^2 + v_4 x_j^2 + v_5 x_i x_j \tag{7}$$

Given that the purpose we pursue in this algorithm is nothing but the initial system modeling, thus through the

combination of partial models and iteration of this process, the major model of the problem is calculated as below:

$$\widehat{f} = v_0 + \sum_{i=1}^{m} v_i x_i + \sum_{i=1}^{m} \sum_{j=1}^{m} v_{ij} x_i x_j + \sum_{i=1}^{m} \sum_{j=1}^{m} \sum_{k=1}^{m} v_{ijk} x_i x_j x_k + \dots \quad (8)$$

After the analysis of the problem into $C_m{}^2$ partial systems, a function with two inputs is calculated for each of them. Thus, produced partial systems are combined in pairs resulting C_m^2 $(C_m^2-1)/2$ problems or new partial models with at least three or four inputs. However, the number of variables dependent on the model or in other words the number of problem inputs is not significant and only the actual precision of the problem estimation by produced systems is important. So, considering this rule to reduce redundant computations and increase the modeling performance and precision, several produced partial functions with acceptable precision and approximation over other functions are chosen and the rest are removed.

Secondly, partial functions are chosen and then the ideal function created in the earlier stage is combined in pairs similar to the first step and new partial functions with at least five and at most six inputs are generated.

In the same way, in the next steps, the combination procedure is repeated by choosing and eliminating a number of created partial functions to finally reach a perfect result.

The main aim of these steps is to achieve a function that almost all parameters are visible and have the lowest number of models (the provided model should be simple). The other purpose considered in performing iterative amalgamation is to reach a model with the lowest error compared to other models calculated in the previous steps.

To this end, the Corrected Akaike Information Criterion (AIC_C) is used which in addition to considering the means square error (MSE) for evaluating the model accuracy considers the number of parameters that exist in the model to provide the simplest model.

$$AIC_C = n \times \log\left(\sum_{i=1}^{n} \left(y_{pre} - y_{obs}\right)^2\right) + 2z + \frac{2z(2z+1)}{n-z-1} \quad (9)$$

Here, n is the number of observations, y_{pre} and y_{obs} are the predicted and observed values, respectively, and z is the number of tuned factors through the modeling by GMDH.

In each layer of the GMDH network, to estimate partial equations, binary choices are done among inputs of that layer of the neural network to create partial equations by means of partial equations. For example, the number of binary selections in the first layer is:

$$C_m^2 = \binom{m}{2} = \frac{m(m-1)}{2} \quad (10)$$

Given that the aim of the utilization of this algorithm is to model and estimate y, so the main equation (Eq. 4) is obtained by the amalgamation of partial systems and iterating this process.

Generalized structure of GMDH

Despite a good capability of convectional GMDH for modeling complex issues, GMDH owns some disadvantages like other techniques. Such disadvantages have a significant impact on results provided by this model so that sometimes significantly decreases the modeling accuracy. Some of the main important disadvantages of this technique are as follows:

1- The structure of the defined polynomial (Eq. 7). This is a second-order polynomial so that there are only two neurons in each polynomial.
2- Input neurons in each layer are only chosen from adjacent layers.

Therefore, for problems of great complexity, the use of second-order polynomials may not yield satisfactory results. Besides, considering two inputs for each neuron leads to an increase in the number of neurons in order to achieve an acceptable model. Using adjacent neurons increase the number of produced polynomials. Thus, the existence of such problems has a noticeable influence on the accuracy and simplicity of provided models. Hence, a model entitled "generalized structure of GMDH (GSGMDH)" is presented in this paper. GSGMDH has been coded by using MATLAB software. The presented model changes the global structure of convectional GMDH so that simultaneously evaluates all possible modes in achieving the best and simplest available model using second and third polynomials and also using two and three inputs in each neuron and finally chooses the best model by means of AICc.

In fact, four states can occur: (1) second-order polynomial with two inputs, (2) second-order polynomial with three inputs, (3) third-order polynomial with two inputs, and (4) third-order polynomial with three inputs. Among these states, the first case is exactly the same relationship provided for convectional GMDH (Eq. 7). Thus, the general form of the polynomial defined in this study is expressed as follows:

$$\begin{aligned} y = {} & v0 + v1*xik + v2*xiq + v3*xip + v4*xiq*xik \\ & + v5*xip*xik + v6*xip*xiq + v7*xik*xik \\ & + v8*xiq*xiq + v9*xip*xip + v10*xip*xiq*xik \\ & + v11*xiq*xik*xik + v12*xiq*xiq*xik \\ & + v13*xip*xik*xik + v14*xip*xiq*xiq \\ & + v15*xip*xip*xik + v16*xip*xip*xiq \\ & + v17*xik*xik*xik + v18*xiq*xiq*xiq \\ & + v19*xip*xip*xip \end{aligned} \quad (11)$$

Furthermore, GSGMDH has the ability to choose neuron inputs from a non-adjacent layer. In Fig. 1, a GSGMDH
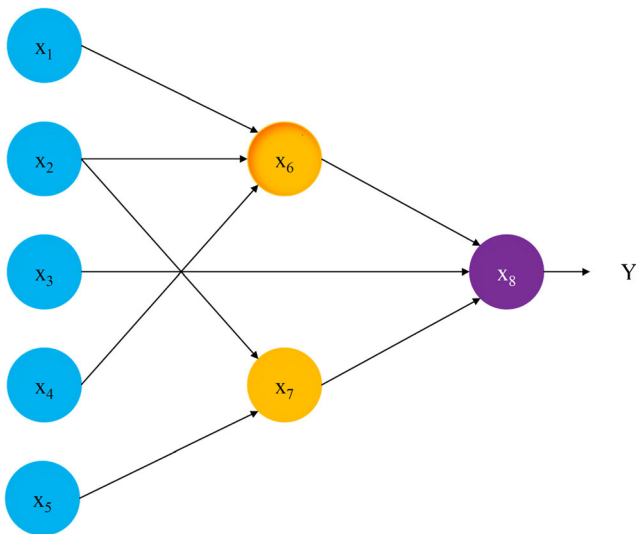
Fig. 1 An example of GSGMDH for a model with five inputs

modeling sample in which the neuron input is from the non-adjacent layer is illustrated. It is seen that Neuron x_6 has three inputs and the Neuron x_7 has two inputs. Moreover, Neuron x_8 as the output of the problem takes three inputs including x_3, x_6, and x_7. It is clear that x_3 is taken from a non-adjacent layer. The flowchart of GSGMDH is presented in Fig. 2.

Goodness of fit

In the present paper, for assessment of the introduced models, the correlation coefficient (R), variance accounted for (VAF), Root Mean Square Error (RMSE), the scatter index (SI), Mean Absolute Error (MAE), and the Nash-Sutcliffe efficiency co-efficient (NSC) are employed as follows (Azimi and Shiri 2020a, b):

$$R = \frac{\sum_{i=1}^{n}\left(F_i - \overline{F}\right)\left(O_i - \overline{O}\right)}{\sqrt{\sum_{i=1}^{n}\left(F_i - \overline{F}\right)^2 \sum_{i=1}^{n}\left(O_i - \overline{O}\right)^2}} \tag{12}$$

$$VAF = \left(1 - \frac{var(F_i - O_i)}{var(F_i)}\right) \times 100 \tag{13}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(F_i - O_i)^2} \tag{14}$$

$$SI = \frac{RMSE}{\overline{O}} \tag{15}$$

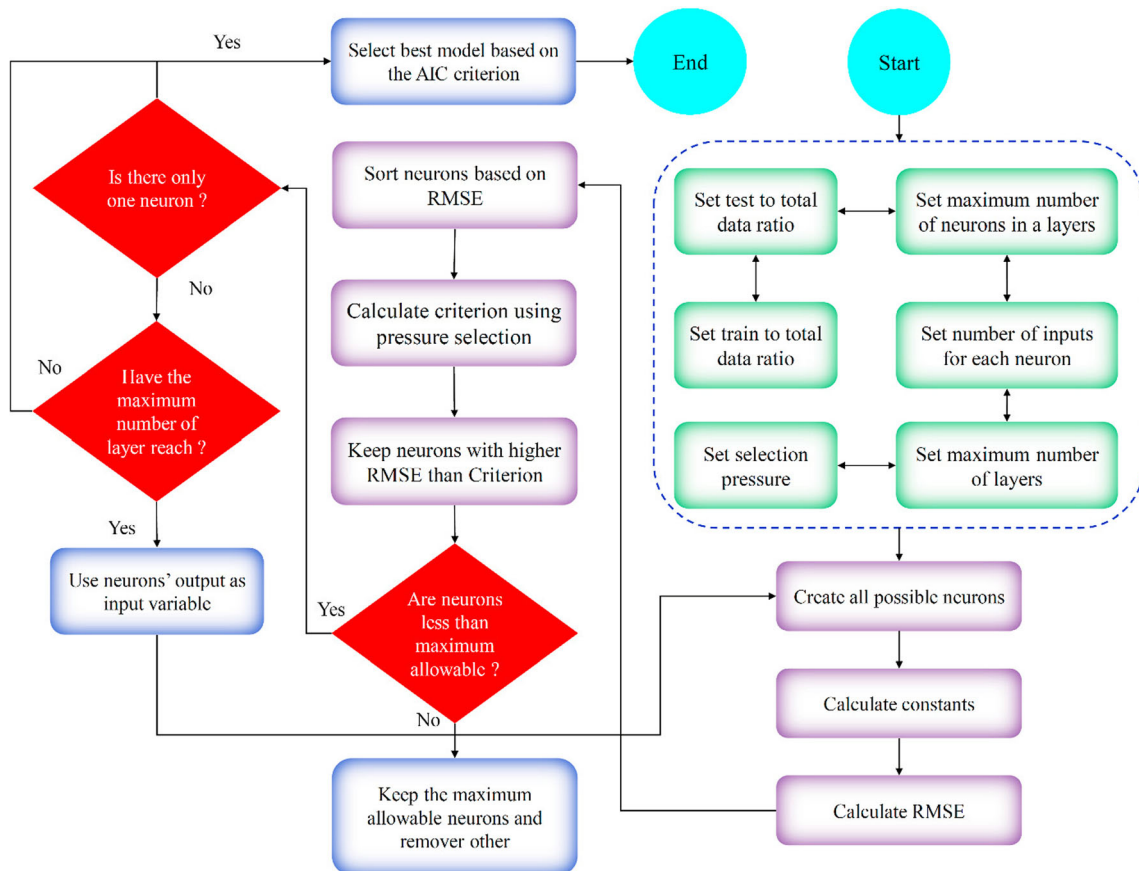$$MAE = \frac{1}{n}\sum_{i=1}^{n}|F_i - O_i| \tag{16}$$

Fig. 2 Flowchart of proposed GSGMDH

$$NSC = 1 - \frac{\sum\limits_{i=1}^{n} (O_i - F_i)^2}{\sum\limits_{i=1}^{n} \left(O_i - \overline{O}\right)^2} \qquad (17)$$

where O_i represents observational values, F_i is predicted values by AI models, $\overline{O}$ is the average of experimental values, and n is the number of experimental values. In the following parts, initially, different activation functions are evaluated. Then, the superior model and the most significant inputs are ascertained through a sensitivity analysis. Moreover, the GSGMDH superior is compared with the GMDH model. Besides, an uncertainty analysis and a partial derivative sensitivity analysis (PDSA) are accomplished. Finally, a formula is proposed for estimating the scour depth near the cross-vane constructions with various forms placed in bending channels.

Experimental models

In order to validate the AI models, the experimental values conducted by Pagliara and Kurdistani (2015), Pagliara et al. (2016), and Kurdistani and Pagliara (2017) are used. The experimental models of cross-vane constructions with J, U, and I forms were used inside a bending channel with a rectangular cross-section with a length of 15 m, a height of 0.5 m, and a width (B) of 0.5 m. They measured the scour depth (Zm) near a structure with a height of h_{st} and a width of b for densimetric Froude numbers equal to F_d. Additionally, the difference between the flow depths near the structure is Δy. The schematic of these experimental models is illustrated in Fig. 3.

Scour near cross-vane constructions

Pagliara and Kurdistani (2015), Pagliara et al. (2016), and Kurdistani and Pagliara (2017) showed that the variables governing the scour near the structures are scour depth (Z_m), height of the structure (h_{st}), tailwater depth (h_{tw}), length of the cross-vane construction (l), channel width(B), the difference between flow depth at the downstream and the upstream of the structure (Δy), discharge (Q), viscosity of water and sediment (ρ_s, ρ), gravity (g), the average diameter of sediments (d_{50}), and the radius of the channel bend (R) as follows:

$$f(Z_m, h_{st}, h_{tw}, l, B, \Delta y, Q, \rho_s, \rho, g, d_{50}, R) = 0 \qquad (18)$$

Through a dimensional analysis, they exhibited that the scour near the cross-vane constructions is a function of the below variables:

$$Z_m/h_{st} = f(l/B, h_{tw}/h_{st}, F_d, \Delta y/h_{st}, B/R) \qquad (19)$$

In the studies accomplished by Pagliara and Kurdistani (2015), Pagliara et al. (2016), and Kurdistani and Pagliara (2017), the parameter l has not been reported. Furthermore,

the shape factor of cross-vane constructions with J, I, and U shapes is also denoted byφ. Thus, Eq. 17 is rewritten as follows:

$$Z_m/h_{st} = f(h_{tw}/h_{st}, F_d, \Delta y/h_{st}, B/R, \phi) \qquad (20)$$

where F_d is the densimetric Froude number. So, the variables of Eq. 20 are assumed as the inputs. It means that six numerical models are developed through different combinations to detect the influencing factors. The developed models are depicted in Fig. 4. It is worth mentioning that 70% of the data are applied for learning the GMDH and GSGMDH models and 30% of this data are employed for testing these models.

Result and discussion

GMDH models

All GMDH numerical models are examined. As discussed in the previous section, six different artificial intelligence models are defined using the input parameters to detect the superior GMDH and the most influencing input. All estimated indices for the GMDH models are illustrated in Fig. 5. GMDH1 simulates the scour parameter using all inputs($h_{tw}/h_{st}, F_d, \Delta y/h_{st}, B/R, \phi$). This model is more accurate than other GMDHs. As an example, the VAF, MAE, and NSC for GMSH1 in the training mode are respectively computed to be 74.451, 0.233, and 0.646. Furthermore, the RMSE and SI for this model in the testing stage are respectively reckoned to be 0.277 and 0.252. For GMDH2, the SI, RMSE, and MAE in the testing stage are respectively estimated to be 0.541, 0.594, and 0.511. Besides, the NSC and VAF in the training mode for this model are -0.911 and -11.517. On the other hand, for simulating the scour parameter by this model, the shape factor (ϕ) is eliminated and this model is a function of $h_{tw}/h_{st}, F_d, \Delta y/h_{st}, B/R$. Also, in the training mode, GMDH3 estimates MAE and RMSE equal to 0.582 and 0.439, respectively. Besides, this artificial intelligence model calculates VAF, SI, and NSC in the testing mode to be 9.343, 0.574, and -0.384, respectively. GMDH3 forecasts the target values regarding the parameters $h_{tw}/h_{st}, F_d, \Delta y/h_{st}, \phi$and the effect of the parameter B/R is neglected, while for GMDH4 in the training mode, the VAF and RMSE are 31.384 and 0.502. Additionally, MAE and NSC in the testing situation are estimated to be 0.543 and -0.594. For simulating the scour pattern by GMSH4, the dimensionless parameters $h_{tw}/h_{st}, F_d, B/R, \phi$ are implemented and the effect of the $\Delta y/h_{st}$ is relinquished. For GMDH5, the RMSE, VAF, and SI in the testing stage are surmised 0.572, 6.507, and 0.529. Furthermore, the MAE and NSC in the testing situation are respectively yielded equal to 0.516 and -0.941. To approximate the scour parameter by the GMDH5 model, the dimensionless parameters $h_{tw}/h_{st}, \Delta y/h_{st}, B/R, \phi$ are employed and the impact of the parameter F_d is neglected.
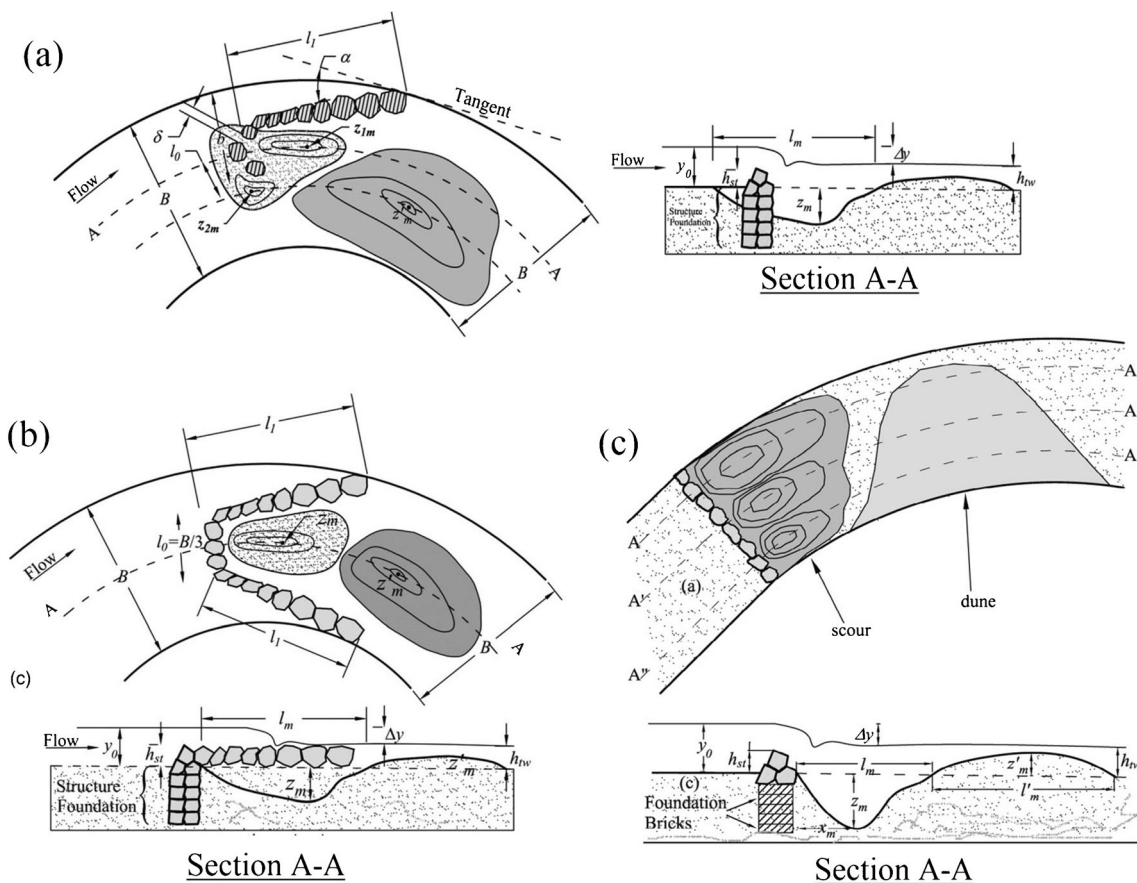
Fig. 3 Schematic of laboratory models of cross-vane constructions in bending channels (a) J-shaped, (b) U-shaped, (c) I-shaped (Pagliara and Kurdistani 2015; Pagliara et al. 2016; Kurdistani and Pagliara 2017)

Besides, GMDH6 simulates the target value using F_d, $\Delta y/h_{st}$, B/R, ϕ and for this model the effects of the parameter h_{tw}/h_{st} are removed. For this model, SI and MAE in the testing stage are respectively calculated to be 0.561 and 0.612.

In Figs. 6 and 7, the comparison of the scour values simulated by the GMDH models with the experimental measurements and their scatter diagrams is shown. GMDH1 owns the

highest precision and the lowest error among these GMDHs. For GMDH1, the correlation coefficient (R) in the training and testing stage is obtained equal to 0.863 and 0.855, respectively. Also, the R values for GMDH2, GMDH3, and GMDH4 in the testing situation are computed 0.191, 0.425, and 0.409. In contrast, GMDH5 and GMDH6 estimate the R values in the training mode equal to 0.371 and 0.423. Therefore, GMDH2 has the lowest correlation and the highest error. Furthermore, GMDH1 is the best model estimating the objective function using all input parameters. In addition, the shape factor (ϕ) and also the densimetric Froude number are known as the most important inputs.
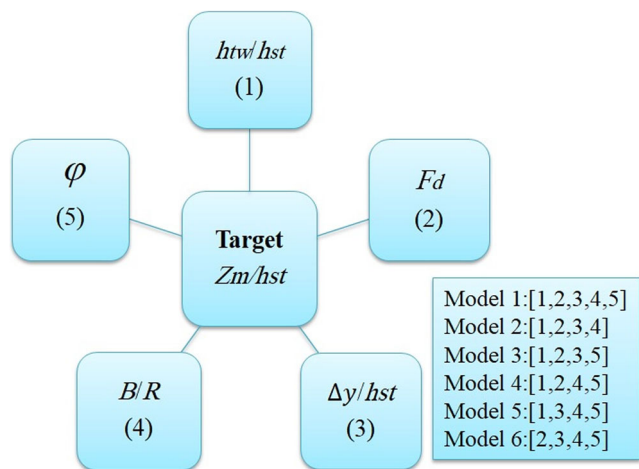
GSGMDH models

The accuracy of the GSGMDH models is assessed. The computed indices for these GSGMDH models are depicted in Fig. 8. Among all GSGMDHs, the GSGMDH1 model owns the highest accuracy and the lowest error. Additionally, the model calculated the RMSE, MAE, and SI in the training stage as 0.232, 0.166, and 0.214, respectively. For GSGMDH1, the VAF value in the training mode is improved by about 12% compared to the GMDH1 model. Moreover, the NSC value
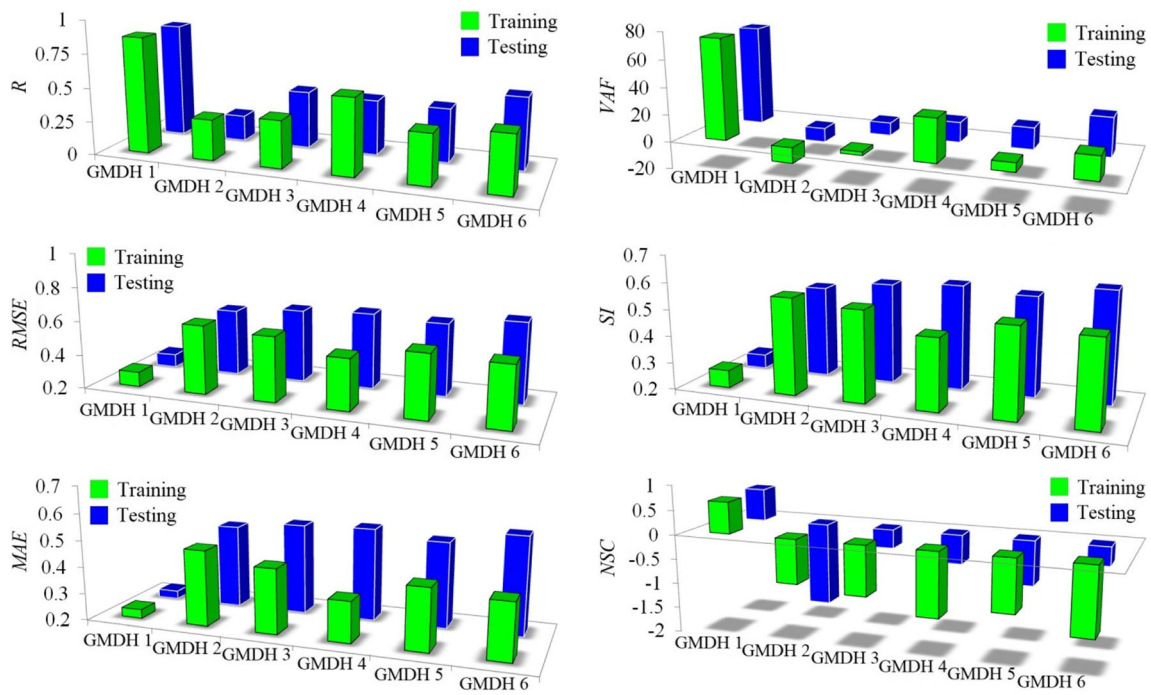

Fig. 4 Combinations of input parameters for developing different numerical models

Fig. 5 Calculated indices for GMDH models
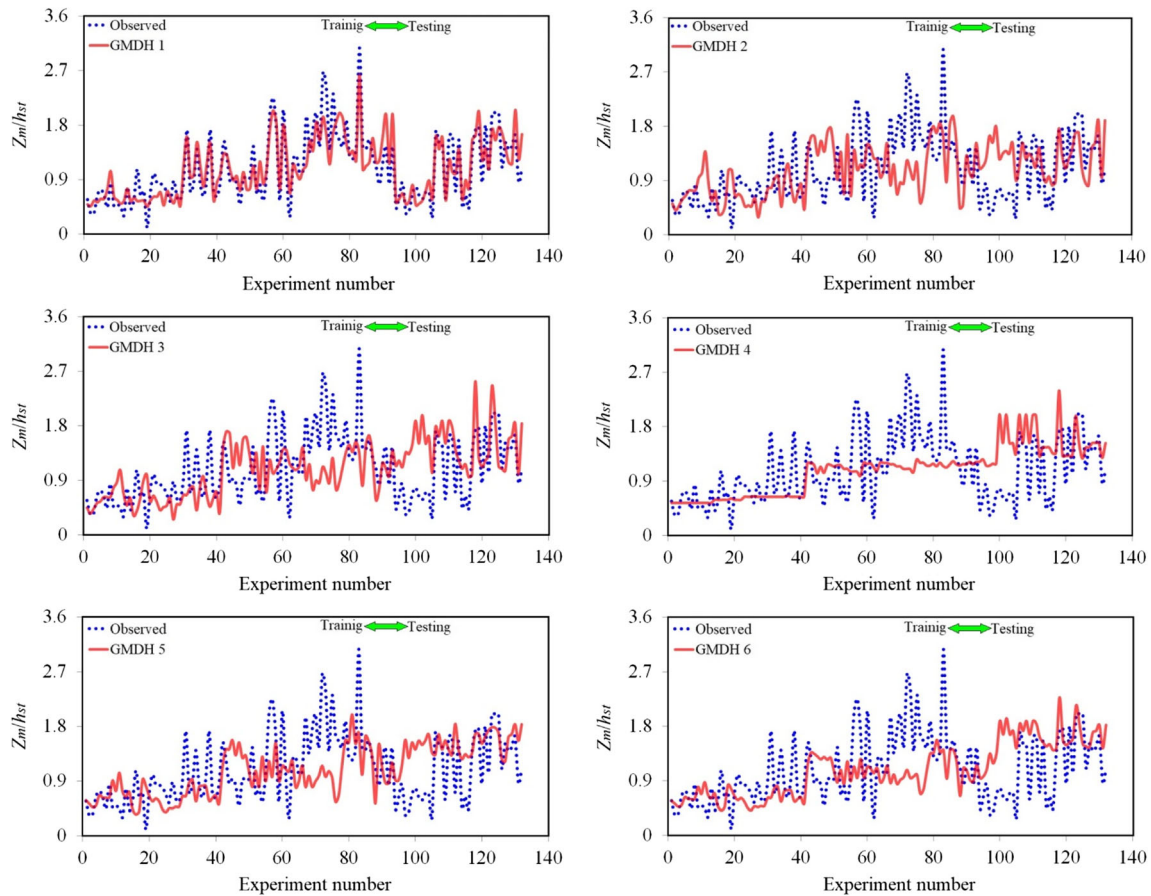

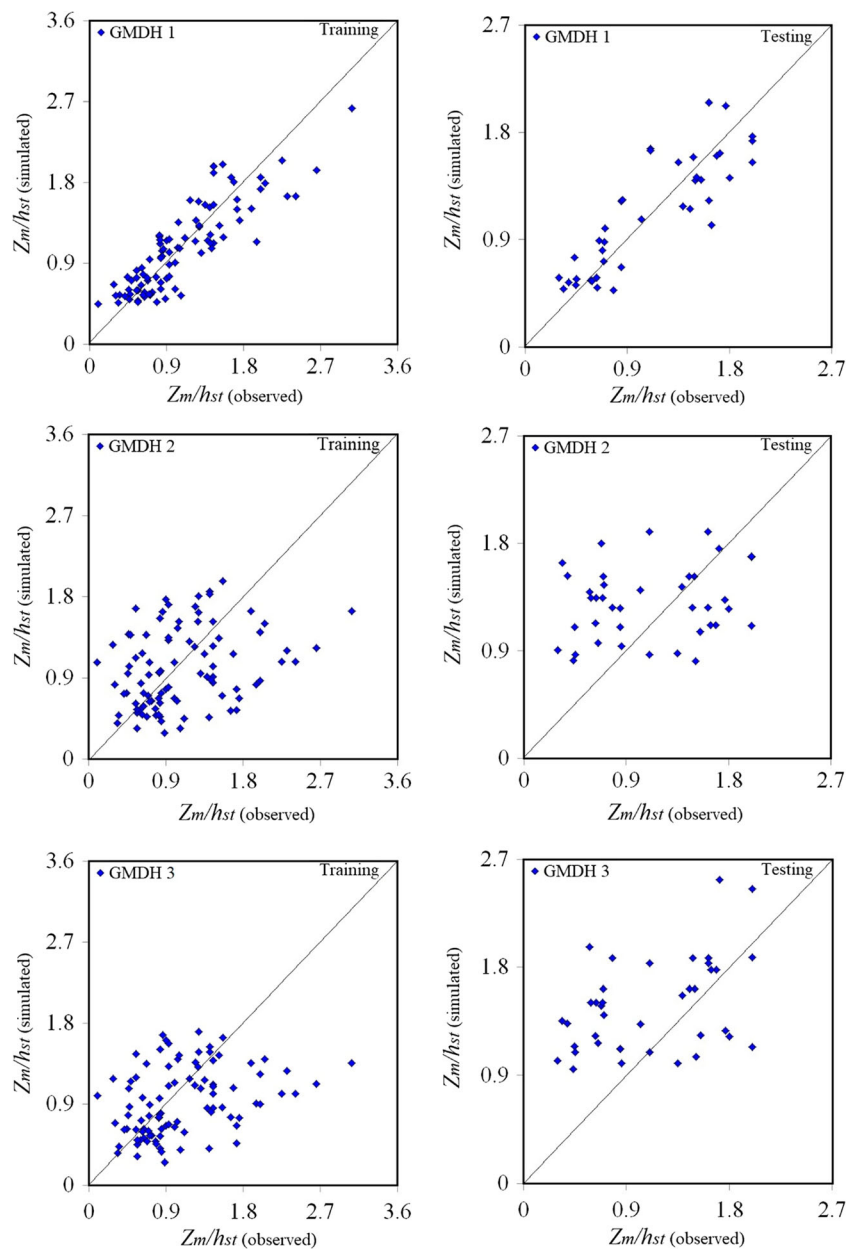**Fig. 6** Comparison of scour values simulated by GMDH models with observed

Fig. 7 Scatter plots for GMDH models

for this model in the testing mode is estimated to be 0.842. In the training mode, GSGMDH1 computes NSC and MAE equal to 0.397 and 0.263. The VAF in the testing mode of the GSGMDH model significantly increases compared to the GMDH2 model. Furthermore, for GSGMDH3 in the testing mode, the values of NSC, SI, and MAE are estimated to be 0.829, 0.186, and 0.155. The VAF index in the training mode of the GSGMDH 3 model is improved by about 41% compared to the GMDH3 model. Among all GSGMDHs, the GSGMDH4 model owns the maximum error and the lowest accuracy. In other words, the RMSE and MAE in the testing stage of this model are considered 0.335 and 0.279. However, the VAF index in the testing mode of this model is almost

twice the GMDH4 model. Additionally, the NSC, MAE, and SI for the GSGMDH5 in the testing situation are 0.542, 0.231, and 0.298, respectively. Compared to the GMDH5 model, this model has a better performance so that the VAF index in the training mode increases about 11 times the GMDH5 model. For GSGMDH6, the SI, MAE, and RMSE in the testing situation are respectively estimated to be 0.235, 0.201, and 0.295. Furthermore, the VAF index in the training and testing phases for the GSGMDH6 model is approximated to be 78.427 and 76.429, respectively.

Thus, according to the results of the GSGMDH models, the dimensionless parameters $\Delta y/h_{st}$ and ϕ are the most influencing inputs of the GSGMDH.
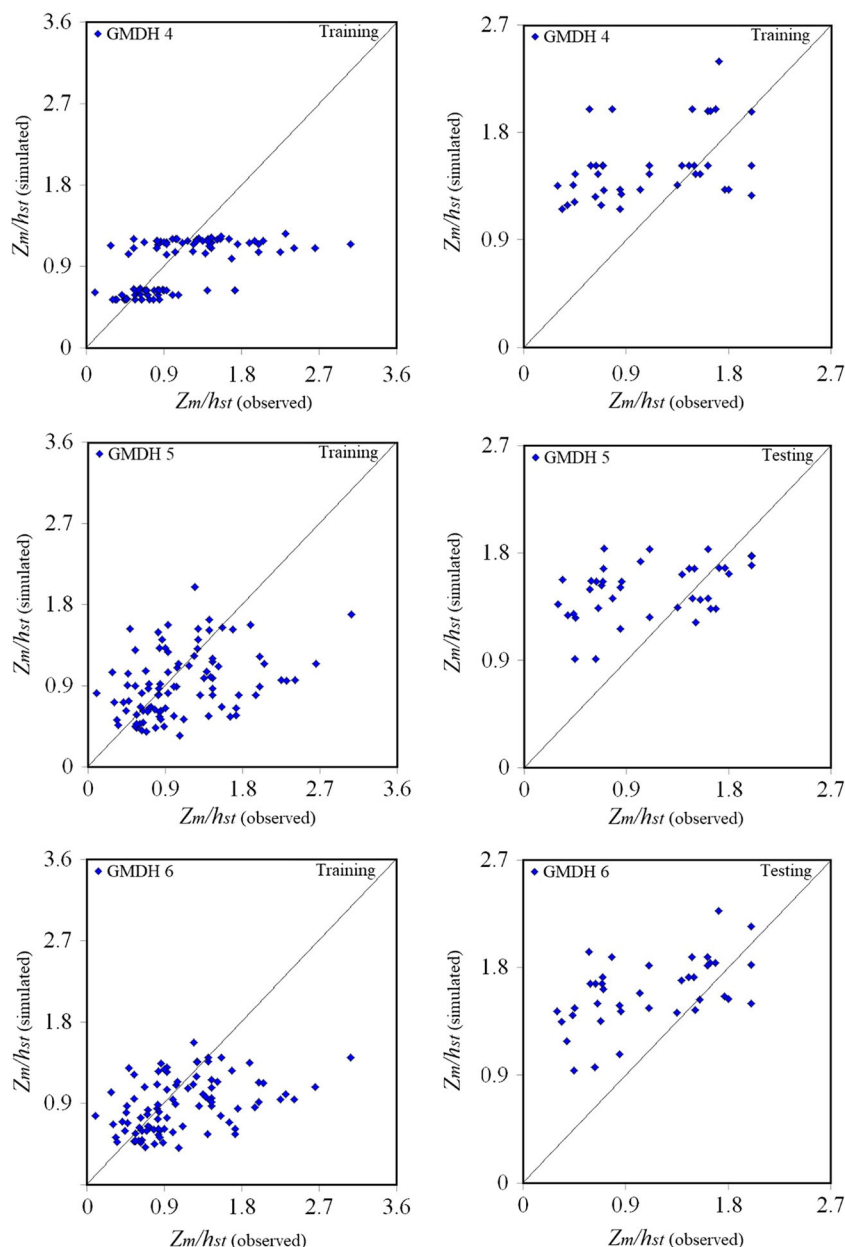
Fig. 7 (continued)

In Figs. 9 and 10, a comparison of the predicted and observed scour parameters and also the scatter plots of all GSGMDH models are depicted. the R values for the GSGMDH 1 in the training and testing modes are computed 0.913 and 0.930, respectively. Additionally, for the testing mode of GSGMDH2 and GSGMDH4, the R index values are estimated 0.821 and 0.924, respectively. Besides, the R values for GSGMDH4, GSGMDH5, and GSGMDH6 in the training stage are respectively obtained as 0.790, 0.860, and 0.886.

Thus, all GSGMDH models have a better proficiency in comparison with the GMDH models and predict the objective function values with higher accuracy.

Moreover, among all numerical models, the GSGMDH1 is determined as the best model. This model surmises the scour parameters using all inputs.

Based on the analysis of the artificial intelligence models, the parameters ϕ, $\Delta y/h_{st}$ and the densimetric Froude number are considered the most effective inputs.

Uncertainty analysis

The performance of the best models is assessed through an uncertainty analysis. Generally, the uncertainty analysis is a beneficial aid for assessing the performance of GMDHs and GSGMDHs (Karbasi and Azamathulla
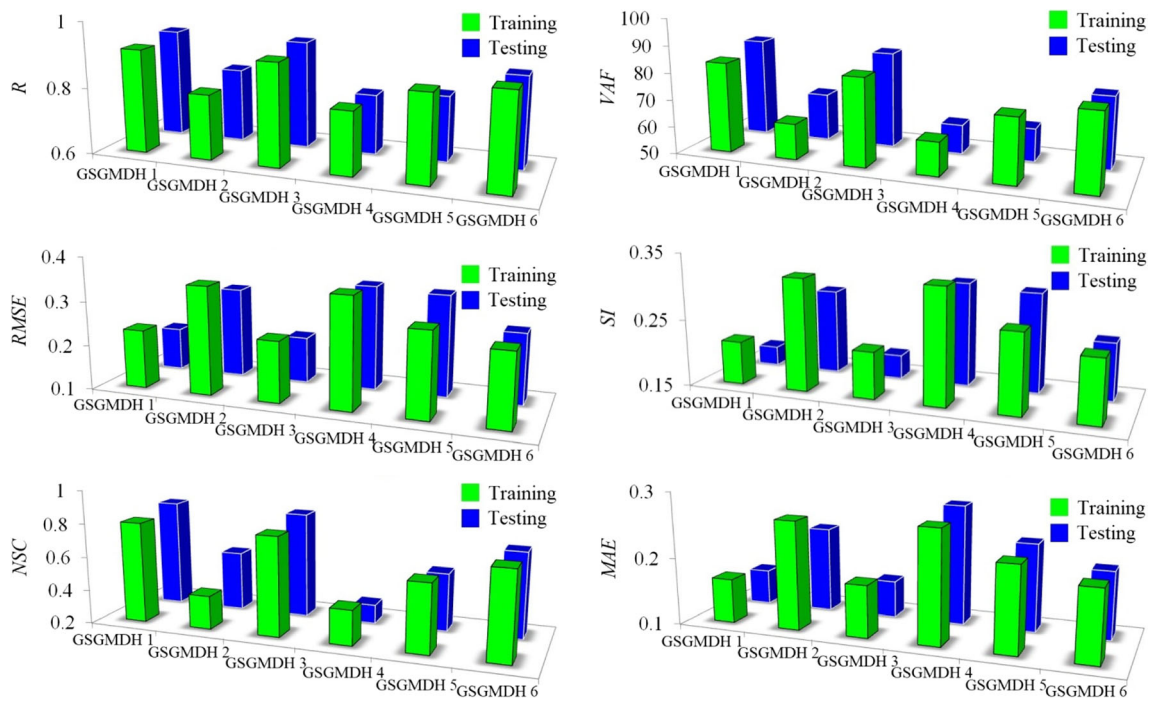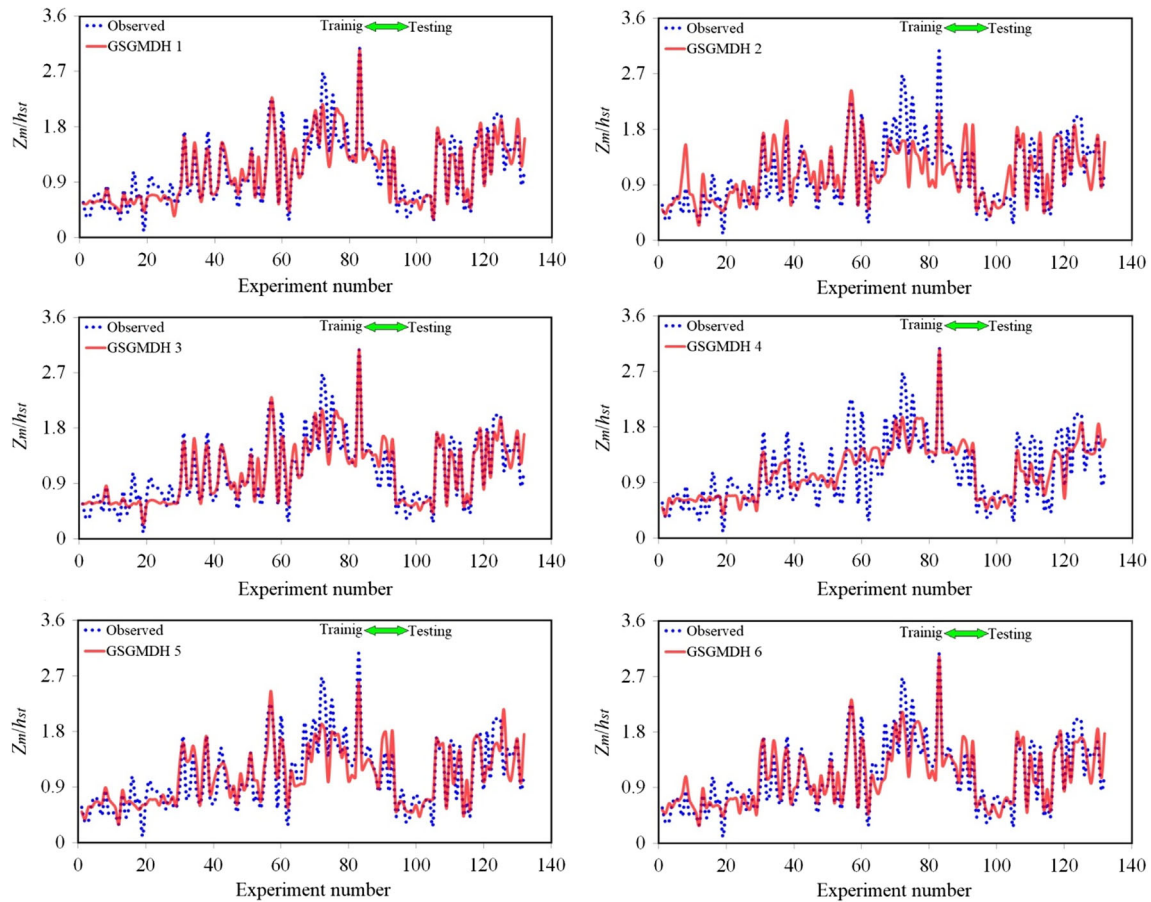
Fig. 8 Computed indices for GSGMDH models

Fig. 9 Comparison of scour values simulated by GSGMDH models with observed parameters
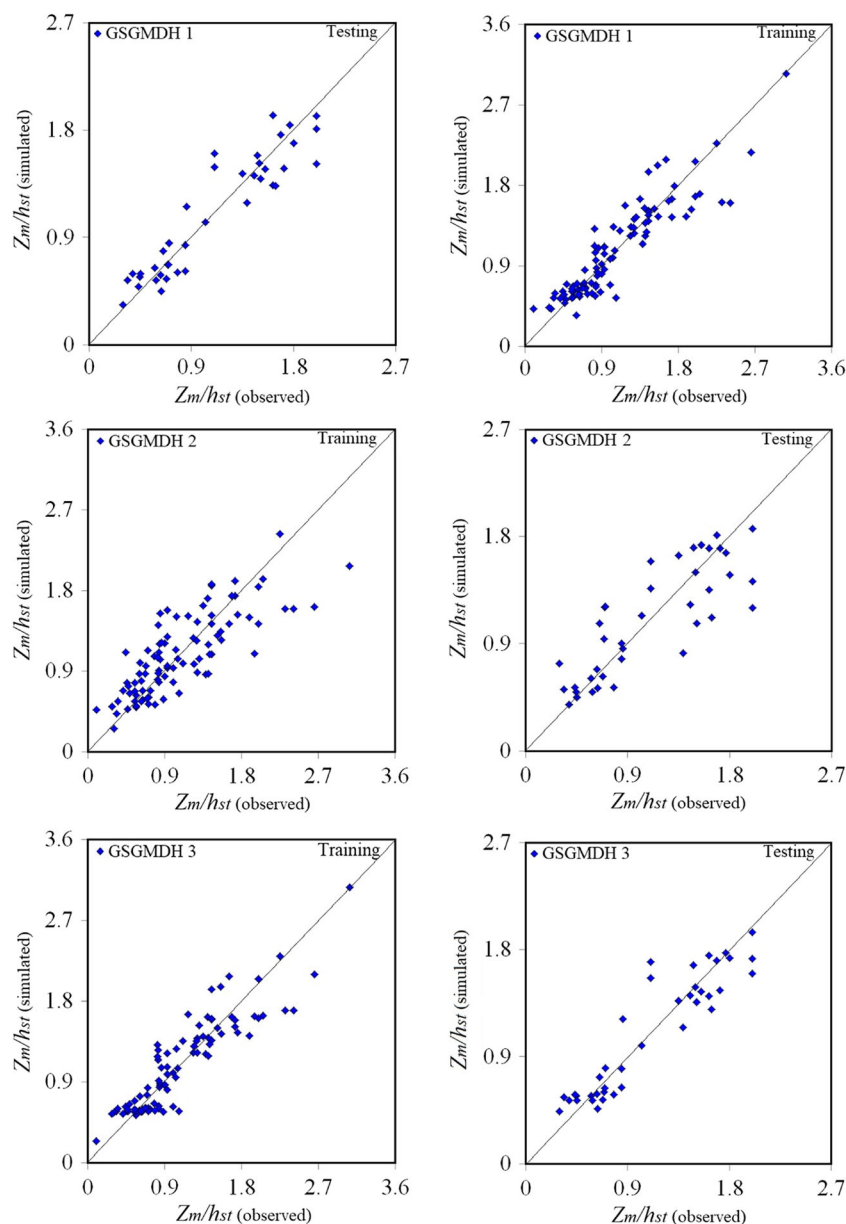
Fig. 10 Scatter plots for GSGMDH models

2016; Azimi et al. 2018, 2019a, b). In other words, this analysis is performed to measure the error computed by these models and assess their proficiency. The error predicted by a model equal to values predicted by the model (P_i) minus observed parameters (O_i)($e_i = P_i - O_i$). Besides, the average of the calculated error is surmised as $\overline{e} = \sum_{i=1}^{n} e_i$. Furthermore, the standard deviation of calculated error is in the form of $S_e = \sqrt{\sum_{i=1}^{n}(e_i - \overline{e})^2 / n - 1}$. Also, the negative sign of $\overline{e}$ shows an underestimated proficiency of this model, whereas its positive sign reveals the overestimated performance. In addition, by using the $\overline{e}$ and S_e, a confidence bound is created near the error values predicted by the Wilson score model without the

continuity correction. Then, the utilization of $\pm 1.96 S_e$, approximately results in a 95% confidence bound (Azimi and Shiri 2021). These parameters of the uncertainty analysis are listed in Table 1. In this table, the width of the uncertainty bound and the 95% prediction error interval are denoted by WUB and 95% PEI, respectively. Both the GMDH1 and GSGMDH1 models own an underestimated performance. Furthermore, the width of the uncertainty bound for the GMDH1 and GSGMDH1 models are respectively calculated − 0.049 and − 0.038. Moreover, 95% PEI for the GMDH1 model is calculated from − 0.049 to 0.049. Also, 95%PEI for the GSGMDH1 model is approximated from − 0.038 to 0.038.
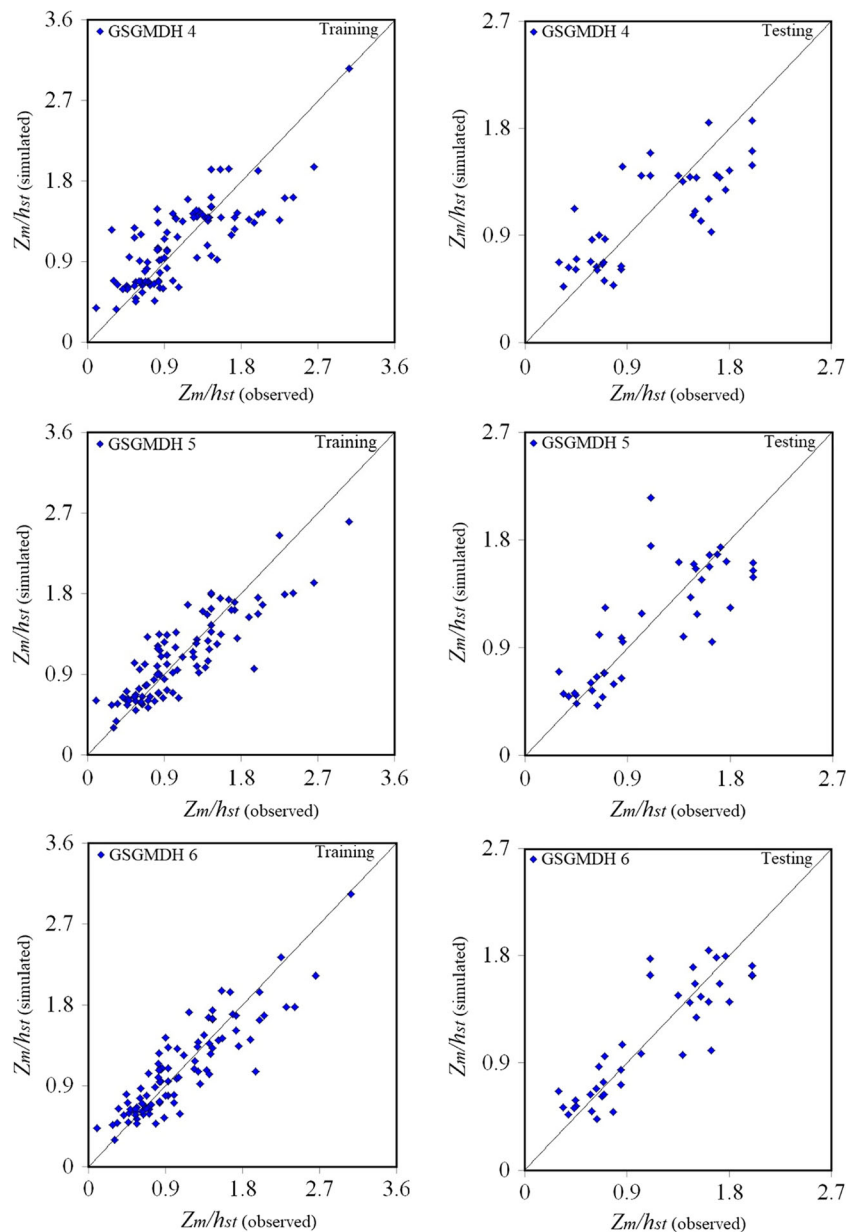
Fig. 10 (continued)

Superior model

The GSGMDH1 is introduced as the superior GSGMDH in simulating the scour depth near cross-vane construction with different forms. For this model, a formula is presented for predicting the scour depth near these protective structures as follows:

$$Z_m/h_{st} = 0.296 + -0.171*x6 + 3.056*B/R + 0.1457*x7 \quad (20)$$
$$-3.393*B/R*x6 + 2.016*x7*x6 - 0.052*x7*B/R$$
$$-0.140*x6^2 - 9.462*B/R^2 - 0.928*x7^2 + 0.188*x7*B/R*x6$$
$$+0.502*B/R*x6^2 + 4.721*B/R^2*x6 - 0.219*x7*x6^2$$
$$-2.686*x7*B/R^2 - 0.647*x7^2*x6 + 0.443*x7^2*B/R$$
$$+0.117*x6^3 + 10.864*B/R^3 + 0.506*x7^3$$

$$x6 = 0.491 - 0.845*\varphi - 2.979*\Delta y/h_{st} + 0.276*F_d$$
$$+3.635*\Delta y/h_{st}*\varphi - 0.005*F_d*\varphi + 2.347*F_d*\Delta y/h_{st}$$
$$+0.481*\varphi^2 + 7.610*(\Delta y/h_{st})^2 - 0.121*F_d^2 - 0.250$$
$$*F_d*\Delta y/h_{st}*\varphi - 0.365*(\Delta y/h_{st})^2*\varphi^2 - 5.563*(\Delta y/h_{st})^2$$
$$*\varphi - 0.045*F_d*\varphi^2 - 2.770*F_d*(\Delta y/h_{st})^2 + 0.0496*F_d^2*\varphi$$
$$-0.065*F_d^2*\Delta y/h_{st} - 0.027*\varphi^3 + 1.406*(\Delta y/h_{st})^3 + 0.004*F_d^3$$

$$(1-20)$$

$$x7 = 1.144 - 4.572*\varphi + 0.348*F_d - 1.235*h_{tw}/h_{st} + 0.836*F_d*\varphi \quad (2-20)$$
$$-2.468*h_{tw}/h_{st}*\varphi + 0.571*h_{tw}/h_{st}*F_d + 8.840*\varphi^2 - 0.278$$
$$*F_d^2 + 0.135*(h_{tw}/h_{st})^2 - 0.194*h_{tw}/h_{st}*F_d*\varphi - 0.547$$
$$*F_d*\varphi^2 + 0.142*F_d^2*\varphi + 1.379*h_{tw}/h_{st}*\varphi^2 + 0.027$$
$$*h_{tw}/h_{st}*F_d^2 + 0.056*(h_{tw}/h_{st})^2*\varphi - 0.046*(h_{tw}/h_{st})^2*F_d$$
$$-3.188*\varphi^3 - 0.003*F_d^3 - 0.040*$$

**Table 1** Uncertainty analysis results for GMDH and GSGMDH

Model	Number of samples	$\bar{e}$	S_e	WUB	95% PEI
GMDH 1	132	$-7.317E{-}17$	0.285	-0.049	-0.049 to 0.049
GSGMDH 1	132	$-1.200E{-}05$	0.223	-0.038	-0.038 to 0.038

Partial derivative sensitivity analysis

The PDSA is performed for GSGMDH1. Generally, PSDA is performed for identifying the influence of inputs on the objective parameter. In other words, PSDA is a method for identifying the variation pattern of the objective parameter according to input parameters. The positive PDSA shows that the scour parameter is enhancing, whereas the negative sign identifies that the objective function is reducing. The relative derivative of each input parameter is calculated regarding the target value (Azimi et al. 2017). For example, the PDSA values are negative for most of the h_{tw}/h_{st} values. The impact of this input variable is decreasing, while the PDSA results for the input parameters F_d, $\Delta y/h_{st}$, and B/R are positive.

Fig. 11 PSDA results for superior model
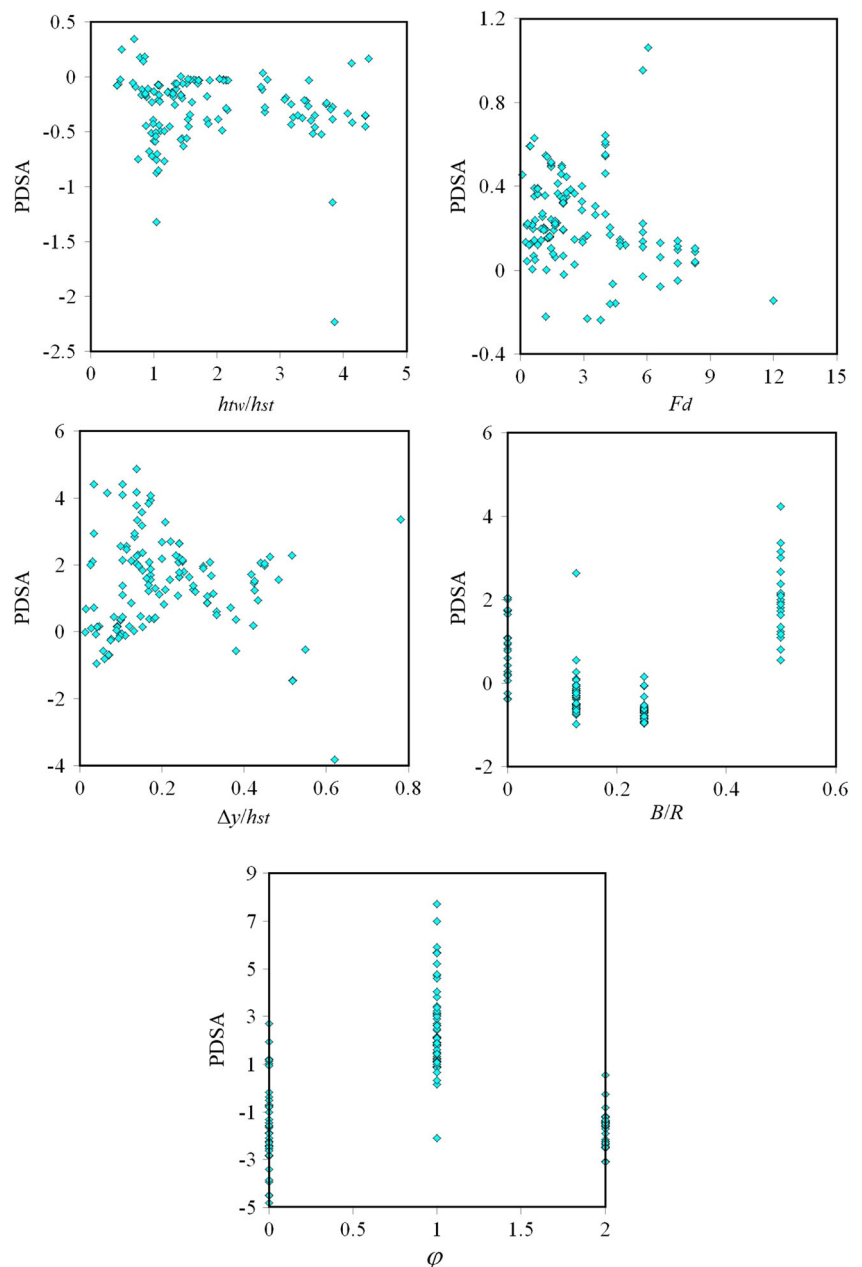
Conversely, a part of the PSDA results for the shape factor (φ) has a positive sign and the other part has a negative sign (Fig. 11).

Conclusion

Thanks to the crucial role of rivers in maintaining the ecosystem of different areas and developing cities, the preservation of their bed and banks is quite significant. In general, different approaches such as submerged vanes, log-deflectors, and cross-vane structures are used to protect riverbeds. Thanks to the ease of protection, cost-saving, and consistency with the ecosystem, cross-vane structures are more popular. In the current paper, a modern AI method entitled "generalized structure group method of data handling (GSGMDH)" was employed for predicting the scour depth near cross-vane constructions with I, U, and J shapes installed inside bending channels for the first time. Compared to the classical GMDH method, this model has higher accuracy and greater flexibility for estimating the scour depth, because, in this technique, middle nodes can take inputs from a non-adjacent layer. To simulate the scour depth, through the inputs, six models were developed for the GMDH and GSGMDH methodologies. The analysis of the models exhibited that the proficiency of the GSGMDH was better. For instance, the RMSE, R, and NSC for the GMDH in the testing stage are calculated to be 0.277, 0.855, and 0.655, while for the GSGMDH are computed equal to 0.196, 0.930, and 0.842, respectively. The sensitivity analysis indicated that the shape factor (ϕ) of cross-vane structures, the ratio of the difference between the flow depths downstream and upstream of the structure to the structure height ($\Delta y/h_{st}$), and the densimetric Froude number (F_d) are the most effective input parameters. In addition, the GMDH and GSGDMH models had underestimated performance. Subsequently, an equation for calculating the scour depth near cross-vane construction with various forms was proposed. Finally, the influence of the input parameters on the objective function was investigated by performing a PDSA.

The obtained results exhibited that the artificial intelligence (AI) approaches had a reasonable ability to predict the scour hole in the vicinity of the cross-vane structures in the curved flumes. It is worth noting that the suggested formula can be utilized in practical application particularly in the field of river engineering to approximate the scour hole depth around this type of structures. Even though the current study provides a good understanding of the simulation of scouring around cross-vane structures with different shapes by using the artificial intelligence (AI) models, further experimental and numerical investigations can be performed in the future.

Declarations

Conflict of interest   The authors declare no competing interests.

References

Amini A, Melville BW, Ali TM, Ghazali AH (2011) Clear-water local scour around pile groups in shallow-water flow. J Hydraul Eng 138(2):177–185. https://doi.org/10.1061/(ASCE)HY.1943-7900.0000488

Azamathulla HM, Ghani AA, Zakaria NA, Guven A (2009) Genetic programming to predict bridge pier scour. J Hydraul Eng 136(3):165–169. https://doi.org/10.1061/(ASCE)HY.1943-7900.0000133

Azimi H, Shiri H (2020a) Ice-Seabed interaction analysis in sand using a gene expression programming-based approach. Appl Ocean Res 98:102120. https://doi.org/10.1016/j.apor.2020.102120

Azimi H, Shiri H (2020b) Dimensionless groups of parameters governing the ice-seabed interaction process. Journal of Offshore Mechanics and Arctic Engineering 142(5):051601. https://doi.org/10.1115/1.4046564

Azimi H, Shiri H (2021) Sensitivity analysis of parameters influencing the ice–seabed interaction in sand by using extreme learning machine. Nat Hazards 105(3):1–29. https://doi.org/10.1007/s11069-021-04544-9

Azimi H, Bonakdari H, Ebtehaj I, Talesh SHA, Michelson DG, Jamali A (2017) Evolutionary Pareto optimization of an ANFIS network for modeling scour at pile groups in clear water condition. Fuzzy Sets Syst 319:50–69. https://doi.org/10.1016/j.fss.2016.10.010

Azimi H, Bonakdari H, Ebtehaj I, Gharabaghi B, Khoshbin F (2018) Evolutionary design of generalized group method of data handling-type neural network for estimating the hydraulic jump roller length. ActaMechanica 229(3):1197–1214. https://doi.org/10.1007/s00707-017-2043-9

Azimi H, Bonakdari H, Ebtehaj I (2019a) Gene expression programming-based approach for predicting the roller length of a hydraulic jump on a rough bed. ISH Journal of Hydraulic Engineering:1–11. https://doi.org/10.1080/09715010.2019.1579058

Azimi H, Bonakdari H, Ebtehaj I, Shabanlou S, Talesh SHA, Jamali A (2019b) A pareto design of evolutionary hybrid optimization of ANFIS model in prediction abutment scour depth. Sādhanā 44(7):169. https://doi.org/10.1007/s12046-019-1153-6

Azimi AH, Shabanlou S, Yosefvand F, Rajabi A, Yaghoubi B (2020) Estimation of scour depth around cross-vane structures using a novel non-tuned high-accuracy machine learning approach. Sādhanā 45(1):1–10. https://doi.org/10.1007/s12046-020-01390-6

Ebtehaj I, Bonakdari H, Khoshbin F, Azimi H (2015) Pareto genetic design of group method of data handling type neural network for prediction discharge coefficient in rectangular side orifices. Flow Meas Instrum 41:67–74. https://doi.org/10.1016/j.flowmeasinst.2014.10.016

Ivakhnenko AG (1976) The group method of data handling in prediction problems. Soviet Automatic Control 9(6):21–30

Karbasi M, Azamathulla HM (2016) GEP to predict characteristics of a hydraulic jump over a rough bed. KSCE J Civ Eng 20(7):3006–3011. https://doi.org/10.1007/s12205-016-0821-x

Khoshbin F, Bonakdari H, Ashraf Talesh SH, Ebtehaj I, Zaji AH, Azimi H (2016) Adaptive neuro-fuzzy inference system multi-objective optimization using the genetic algorithm/singular value decomposition method for modelling the discharge coefficient in rectangular sharp-crested side weirs. Eng Optim 48(6):933–948. https://doi.org/10.1080/0305215X.2015.1071807

Khosronejad A, Kozarek JL, Diplas P, Hill C, Jha R, Chatanantavet P, Heydari N, Sotiropoulos F (2018) Simulation-based optimization of in-stream structures design: rock vanes. Environ Fluid Mech 18(3):695–738. https://doi.org/10.1007/s10652-018-9579-7

Kurdistani SM, Pagliara S (2017) Experimental study on cross-vane scour morphology in curved horizontal channels. J Irrig Drain Eng 143(7):04017013. https://doi.org/10.1061/(ASCE)IR.1943-4774.0001183

Pagliara S, Kurdistani SM (2013) Scour downstream of cross-vane structures. J Hydro Environ Res 7(4):236–242. https://doi.org/10.1016/j.jher.2013.02.002

Pagliara S, Kurdistani SM (2015) Clear water scour at J-Hook Vanes in channel bends for stream restorations. Ecol Eng 83:386–393. https://doi.org/10.1016/j.ecoleng.2015.07.003

Pagliara S, Kurdistani SM (2017) Flume experiments on scour downstream of wood stream restoration structures. Geomorphology 279:141–149. https://doi.org/10.1016/j.geomorph.2016.10.013

Pagliara S, Kurdistani SM, Santucci I (2013a) Scour downstream of J-Hook vanes in straight horizontal channels. ActaGeophysica 61(5):1211–1228. https://doi.org/10.2478/s11600-013-0143-z

Pagliara S, Kurdistani SM, Cammarata L (2013b) Scour of clear water rock W-weirs in straight rivers. J Hydraul Eng 140(4):06014002. https://doi.org/10.1061/(ASCE)HY.1943-7900.0000842

Pagliara S, SagvandHassanabadi L, Mahmoudi Kurdistani S (2015a) Log-vane scour in clear water condition. River Res Appl 31(9):1176–1182. https://doi.org/10.1002/rra.2799

Pagliara S, Hassanabadi L, Kurdistani SM (2015b) Clear water scour downstream of log deflectors in horizontal channels. J Irrig Drain Eng 141(9):04015007. https://doi.org/10.1061/(ASCE)IR.1943-4774.0000869

Pagliara S, Kurdistani SM, Palermo M, Simoni D (2016) Scour due to rock sills in straight and curved horizontal channels. J Hydro Environ Res 10:12–20. https://doi.org/10.1016/j.jher.2015.07.002

Rosgen DL (2001) The cross-vane, w-weir and j-hook vane structures... their description, design and application for stream stabilization and river restoration. In Wetlands Engineering & River Restoration 2001:1–22. https://doi.org/10.1061/40581(2001)72

Scurlock S, Thornton CI, Abt SR (2011a) One-dimensional modeling techniques for energy dissipation in U-weir grade-control structures. In: Reston VA (ed) ASCE copyright Proceedings of the 2011 World Environmental and Water Resources Congress; May 22. 26, 2011, Palm Springs, California| d 20110000. American Society of Civil Engineers. https://doi.org/10.1061/41173(414)259

Scurlock SM, Thornton CI, Abt SR (2011b) Equilibrium scour downstream of three-dimensional grade-control structures. J Hydraul Eng 138(2):167–176. https://doi.org/10.1061/(ASCE)HY.1943-7900.0000493

Shabanlou S, Azimi H, Ebtehaj I, Bonakdari H (2018) Determining the scour dimensions around submerged vanes in a 180 bend with the gene expression programming technique. J Mar Sci Appl 17(2):233–240. https://doi.org/10.1007/s11804-018-0025-5