Investigating the capability of two hybrid intelligence methods to predict bedform dimensions of alluvial channels

Kourosh Qaderi, Mohammad Reza Maddahi, Majid Rahimpour and Mojtaba Masoumi Shahr-babak

ABSTRACT

Dimensions of river bedforms have an effect on total roughness. The complexity of bedform development causes empirical methods to differentiate from each other in predicting bedform dimensions. In this paper, two novel hybrid intelligence models based on a combination of the group method of data handling (GMDH) with the harmony search (HS) algorithm and shuffled complex evolution (SCE) have been developed for predicting bedform dimensions. A data set of 446 field and laboratory measurements were used to evaluate the ability of the developed models. The results were compared to conventional GMDH models with two kinds of transfer functions and an empirical formula. Also, five different combinations of dimensionless parameters as input variables were examined for predicting bedform dimensions. Results reveal that GMDH-HS and GMDH-SCE have good performance in predicting bedform dimensions, and all artificial intelligence methods were dramatically different to the empirical formula of van Rijn showing that using these methods is a key to solving complexity in predicting bedform dimensions. Also, comparing different combinations of dimensionless parameters reveals that there is no significant difference between the accuracy of each combination in predicting bedform dimensions.

Key words | bedform dimensions, group method of data handling, harmony search, river, shuffled complex evolution, van Rijn method

Kourosh Qaderi
Mohammad Reza Maddahi (corresponding author)
Majid Rahimpour
Department of Water Engineering,
Shahid Bahonar University of Kerman,
Kerman,
Iran
E-mail: *mreza_maddahi@yahoo.com*

Mojtaba Masoumi Shahr-babak
Department of Civil Engineering,
Shahid Bahonar University of Kerman,
Kerman,
Iran

INTRODUCTION

An alluvial channel is a water channel made up of loose sediments called alluvium. The sediments move as bedload and suspended load, and produce bedforms. The principles of creating different bedforms in different alluvial channels (sand-bed, gravel-bed, etc.) have been introduced in different references, e.g. van Rijn (1984). So, alluvial channels have mobile beds which is the most important parameter in creating bedforms. In fixed bed rivers like concrete channels, no sediment moves in the bed and no bedform is created. River bedforms play a significant role in friction and sediment transport. The accurate prediction of the geometric characteristics of bedforms is an essential component for estimating the flow resistance and the consequent flow conditions. In the past half-century, a number of empirical methods have been presented to relate bedform dimensions to flow characteristics. However, the presence of different kinds of variables has made the study of this problem very complicated. The method of van Rijn (1984) is one of the most widely used and applicable methods for prediction of bedform dimensions. Van Rijn's method is based on a regression analysis of 84 flume and 22 field data points. This model reveals a large variability in the prediction of

doi: 10.2166/ws.2017.230

bedform dimensions in the different conditions of natural channels (Julien & Klaassen 1995), showing underprediction of bedform dimensions under different morphodynamic conditions (Julien 1992). For this reason, other researchers have tried to achieve better methods to predict bedform dimensions (Yalin 1992; Julien & Klaassen 1995; Karim 1995; Talebbeydokhti *et al.* 2006). These methods are derived based on different river or flume data which are not appropriate for all conditions such as van Rijn's method. This shows that attempts to achieve a new relationship using a different data series will have similar results and the new predictor would not have a global use because of the complex behavior of bedform formation during different flow conditions.

Due to the uncertainty and complexity of the variables in prediction of bedform dimensions, it is difficult to present a unique empirical method for prediction of bedform dimensions. During recent years, many system identification techniques have been developed in various fields of engineering to predict the unknown behavior of complex systems based on a given input-output data set. One of these approaches is based on artificial intelligence (AI) methods. AI methods can be categorized into two major groups. The first includes data driven methods (DDM) such as artificial neural network (ANN), support vector machine (SVM), group method of data handling (GMDH), etc. The second group contains knowledge-based methods such as genetic algorithm (GA), shuffled complex evolution (SCE), harmony search (HS), etc. In recent decades, AI methods have become increasingly popular in hydrology and water resources research and engineering.

GMDH is an AI technique which belongs to the self-organizing modeling approach. In this method, the number of neurons, the number of layers and the behavior of each neuron is adjusted during the process of training, therefore, the prediction of system modeling in GMDH is more complex than in other artificial methods. GMDH has been reported to have good results in predicting debris flow (Zhang *et al.* 2013), significant wave height (Shahabi *et al.* 2016) and the discharge coefficient of rectangular sharp-crested side weirs (Ebtehaj *et al.* 2015).

Gaining better predictions by using AI methods rather than empirical models, has not been sufficient in recent decades, so scientists have tried to optimize the predictions. For this purpose, they have combined different DDM techniques with different optimization methods called hybrid intelligence systems. Many successful applications of these hybrid intelligence systems have been reported. For example, GA-ANN for flood forecasting (Wu & Chau 2006), GMDH-LSSVM, which is the combination of GMDH with least squares support vector machines (LSSVM) for river flow forecasting (Samsudin *et al.* 2011), GA-SVM for prediction of pollution in reservoirs (Su *et al.* 2015) and GMDH-HS for uplift capacity prediction of suction caissons (Masoumi Shahr-Babak *et al.* 2016).

According to the above-mentioned applications of AI and hybrid methods, it seems that using these methods can provide an accurate prediction of bedform dimensions which is an important factor in most river engineering problems. According to the literature, only Javadi *et al.* (2015) have used AI methods including ANN and SVM to predict bedform dimensions using 257 datapoints from the Rhine and Meuse rivers. Using data from just two rivers and no flume data plus a small range of variables increases doubts about using these methods with other natural or flume data sets. To solve this problem, a wider range of data sets and novel hybrid models are required.

The main objective of this research is to develop and apply two new hybrid intelligence methods called GMDH-HS and GMDH-SCE, based on a combination of GMDH with HS and SCE algorithms, to predict bedform dimensions from 447 river and flume data points with a wide range of variables. HS and SCE were used as subroutines for calibrating weights. Also, codes have been written in MATLAB to use the models not the tools in MATLAB. Finally, the results are compared to conventional methods of GMDH with two kinds of transfer functions called GMDH1 and GMDH2. The advantage of the GMDH model is that the number of neurons and layers is determined while running the model, so GMDH is faster than ANN, GP, etc. for predicting different parameters. Parameters will be defined as inputs which have the following characteristics: (1) easy to measure, (2) easy to calculate, (3) similar to the parameters extended from dimensional analysis, and (4) similar to the parameters that other researchers in the literature have used.

METHODS AND MATERIALS

Group method of data handling

GMDH is based on the principle of exploratory self-organizing which is a combination of N-Adaline (Ivakhnenko 1968). Since GMDH uses data classification both usefully and uselessly and needs fewer observational data, its structure is more precise in comparison with perceptron and needs less time for performing the calculations. A schematic diagram of this model is shown in Figure 1(a) with an additional view of the N-Adaline structure with a second order polynomial function as the active function. In Figure 1(b), *sq*, ×, X_i and Y represent squared, product, the inputs and the output, respectively. The external criterion for determining the system structure and for choosing the best neuron of each layer is defined as follows:

$$R^2 = 1 - \frac{\sum_{l=1}^{N} (y_p(l) - y_0(l))^2}{\sum_{l=1}^{N} (y_0(l) - \bar{y}_0)^2} \quad (1)$$

where R^2 is the determination coefficient, and y_0, y_p and $\bar{y}_0$ are the observation output, calculated output and average of the observation output, respectively.

In the GMDH algorithm, the data are divided into two groups of training and testing data sets. This division is based on the variance of total data from the mean value. The points with high variance are used in the testing data set to ensure that the selected models can extrapolate outside the data in the training set. Then, the data in the input matrix are taken in pairs and a quadratic polynomial with coefficients, w_i, between each pair, x_i and x_j, with the corresponding output, Y, is written. These coefficients are evaluated using a least squares estimation (LSE) method. The output of each polynomial is compared with the data points in the testing data set. The mean squared error (MSE) is used to select the polynomials which are allowed to proceed to the next layer. In the next layer, the outputs of the selected polynomials become the new input values. After repeating these steps, the lowest MSE will no longer be smaller than in the previous layer. In this situation, the GMDH run will reach the condition of termination. Then, the model will trace back the path of the polynomials that correspond to the lowest MSE in each layer. By repeating these processes, only one neuron will remain in the final layer. Each neuron performs as a nonlinear function of the inputs. In this research, two kinds of nonlinear functions (Equations (2) and (3)) have been used as an active function in each neuron: the first order polynomial and second order polynomial transfer functions, as follows (Masoumi Shahr-Babak *et al.* 2016):

$$Y = w_0 + w_1x_1 + w_2x_2 + w_3x_1x_2 + \ldots \text{ (GMDH1)} \quad (2)$$

$$Y = w_0 + w_1x_1 + w_2x_2 + w_3x_1^2 + w_4x_2^2 + w_5x_1x_2 + \ldots \text{ (GMDH2)} \quad (3)$$

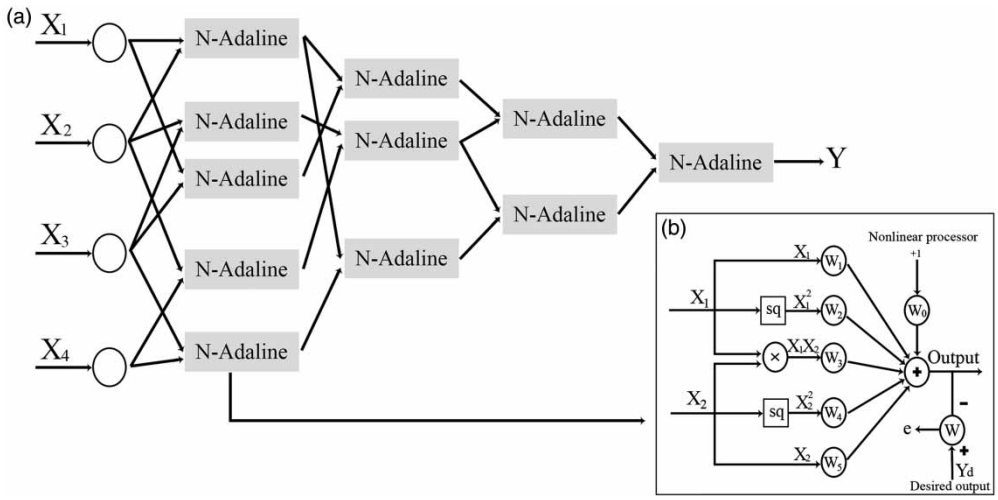

Figure 1 | A schematic diagram of the GMDH model (Masoumi Shahr-Babak *et al.* 2016).

in which w_i is the coefficient, x_i is the input and Y is the output. Also the GMDH can be combined with other evolutionary or AI models to find the coefficients of polynomials. HS and SCE algorithms are among these models.

Harmony search algorithm

The computational procedure of HS is inspired by a group of musicians searching for a musically pleasing harmony (Ayvaz 2009). This process was first adapted into engineering optimization problems by Geem *et al.* (2001). In the HS algorithm, each musician is a decision variable and the collection of notes in the musicians' memory is the value of the decision variables. Adaptation of the musical rules to the optimization problems is (1) generating a new solution vector from harmony memory (HM), (2) replacing a decision variable with a new one which is close to the current one (pitch adjusting) and (3) generating a solution vector from the possible random range (random selection). Combined utilization of these rules allows identification of the optimal or near optimal solutions. Although HS searches for the optimal solution by considering multiple solution vectors as in GA, its reproduction process is different from GA. While GA generates a new offspring from two parents in the population, HS generates it from all the existing vectors stored in HM (Ayvaz 2009).

Shuffled complex evolution algorithm

The SCE method was developed at the University of Arizona to deal with the peculiarities encountered in environmental model calibration. It combines the best features of multiple complex shuffling and competitive evolution based on the simplex search method. The use of multiple complexes and their periodic shuffling provides an effective exploration of different promising regions of attraction within the search space (Muttil & Jayawardena 2008). The method is based on a synthesis of four concepts: (1) combination of deterministic and probabilistic approaches, (2) systematic evolution of a 'complex' of points spanning the parameter space, in the direction of global improvement, (3) competitive evolution and (4) complex shuffling. The synthesis of these elements makes the SCE method effective and robust, and also flexible and efficient (Kan *et al.* 2016). A detailed presentation of the theory underlying the SCE algorithm can be found in Duan *et al.* (1993).

GMDH-HS and GMDH-SCE algorithms

Several hybrid intelligent systems have been developed based on a hybrid of DDM with optimization algorithms in various ways. Optimization algorithms are generally used to determine the structure of DDM, calibrate unknown weights or determine both of them. In this study, the HS and SCE are used to calibrate the weights of each neuron in GMDH rather than using the LSE method. Since GMDH is a self-organizing method with an unknown structure, its structure can be determined by HS and SCE algorithms. So, a hybrid integration of GMDH and HS or SCE algorithms may have a better performance by taking advantages of the characteristics of both methods together. In these algorithms, HS and SCE are employed to train and optimize the initial parameters or weights of transfer function in each neuron of the GMDH structure. The objective of the HS and SCE submodels is to determine optimal weights in order to attain the optimum structure of the GMDH model and minimum cumulative errors between the measured and predicted data sets.

Bedforms

Bedforms such as river dunes are rhythmic bed features which are developed by the interaction between water flow and sediment transport (van der Mark *et al.* 2008). River dunes are often schematized as a train of regular triangular features. Figure 2 illustrates characteristics of bedforms. The purpose of this study is to predict the length and height of dune bedforms in rivers and flumes.

Data set

In this study, the main issue is to choose the best parameters as input variables. To find the parameters using dimensional analysis, one can conclude that:

$$(\Delta, \lambda) = f\left(Fr, \mathrm{Re}, \frac{D_{50}}{h}, S_w, G_s - 1, S_p, \sigma\right) \quad (4)$$

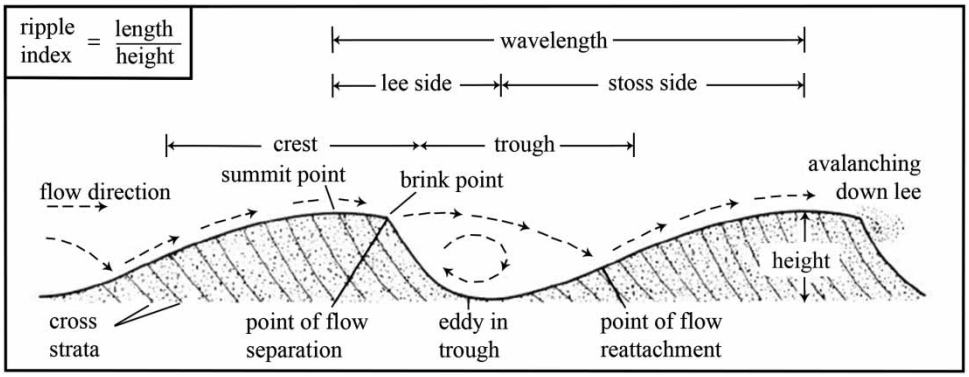

Figure 2 | Bedform characteristics.

in which Δ and λ are height and length of a bedform, respectively, Fr is the Froude number, the Re is Reynolds number, G_s is relative density of bed materials, S_w is energy slope gradient, S_p is sediment shape and σ is sediment distribution which is defined by means of geometric standard deviation as $(D_{84}/D_{16})^{0.5}$ in the literature. D_{84} and D_{16} are median bed particle diameters in which 84 and 16 percent of particles are smaller than these, respectively. Measuring these parameters is a difficult task. For this reason, in most models presented in the literature, simple parameters have been used, for example Yalin (1992) has used shear stress and critical shear stress, Fredsoe (1975) has used dimensionless bed shear stress, van Rijn (1984) has used the transport-stage parameter and D_{50}/h and Julien & Klaassen (1995) have used D_{50}/h as the input variables. The dimensionless parameters used in this study are: (1) Shear Froude number, $Fr* = u_*/\sqrt{gh}$, (2) Transport-stage parameter, $T = ((u'_*)^2 - (u_{*c})^2)/(u_{*c})^2$, (3) Particle parameter, $D_* = [(g(G_s - 1))/\upsilon^2]^{1/3} D_{50}$, (4) Shields parameter, $\theta = u_*^2/(D_{50}(G_s - 1)g)$ and (5) Suspension parameter, $z = (g(G_s - 1)D_{50}^2)/(18k\upsilon u_*)$ in which u_* is shear velocity, g is gravitational acceleration, h is flow depth, u'_* is grain shear velocity, u_{*c} is critical grain shear velocity, D_{50} is median bed particle diameter where 50 percent of particles are smaller than it, υ is kinematic viscosity and k is the von Kármán coefficient. Different combinations of these variables should be checked to see which combination yields better results for predicting bedform dimensions. Although these variables are easier to calculate, their initial parameters are sometimes hard to measure or convert to each other. For example, in most reports, the only parameters reported are h, u, s and D_{50}. To calculate the transport-stage parameter in this study, the following equations are used (van Rijn 1984):

$$u'_* = u\frac{\sqrt{g}}{C'} \tag{5}$$

in which

$$C' = 18\log\left(\frac{12R_b}{3D_{90}}\right) = 7.82\ln\left(\frac{12R_b}{3D_{90}}\right) \tag{6}$$

where u is depth-averaged flow velocity and R_b is hydraulic radius related to bed which is equal to flow depth for wide channels. Combining Equations (5) and (6) yields:

$$\frac{u}{u'_*} = \frac{\sqrt{g}}{0.4}\ln\left(\frac{h}{0.25D_{90}}\right) \tag{7}$$

On the other hand, the logarithmic distribution of the velocity profile is:

$$\frac{u}{u_*} = \frac{1}{0.4}\ln\left(\frac{h}{K_s}\right) \tag{8}$$

K_s is the reference bed level or grain roughness where water velocity at that depth is equal to zero. Scientists have presented different values for K_s in the form of aD_b in which a and b take different values. By comparing Equations (7) and (8), it is found that $u'_* = u_*/\sqrt{g}$ if $K_s = 0.25D_{90}$. u_* can also be simply calculated by use of the

logarithmic distribution of velocity profile or boundary-layer characteristics method presented by Afzalimehr & Anctil (2000).

After calculating input variables, the ratio of bedform height to bedform length and flow depth are used as output variables. The 446 data points were taken from the Klaassen study of different flumes and rivers around the world (Klaassen 1990). In this study, data sets were divided into training and testing data. The data set of 312 data points (70%) was used randomly for training and the rest (30%) was used for testing, according to specific indices including: mean, skewness, minimum and maximum of training and testing data. In addition, to select data randomly, introducing some noisy data into the training set has reduced the chances of overfitting. Also, when the number of training examples is small, their discrepancies are big, causing a serious overtraining problem. To avoid this, a large number of training data with a wide range of statistical variables is chosen. The range of statistical parameters for the training data set is shown in Table 1. SD and CV are standard deviation and coefficient of variation, respectively.

It is important that both training and testing data sets have the same statistical indices. It is observed from Table 1 that the statistical indices are approximately the same for all parameters.

In order to assess the accuracy of each model, various statistics have been used. The best known and most widely used ones are presented below. These statistics were appropriately used in the calibration phase to determine the parameters and structures.

1. Nash-Sutcliffe Efficiency Coefficient (E)
2. Root Mean Square Errors (RMSE)
3. Mean Square Relative Error (MSRE)
4. Mean Absolute Percentage Error (MAPE)
5. Relative Bias (RB)

E is used to assess the predictive power of models (Masoumi Shahr-Babak *et al.* 2016). The range of this criterion is $-\infty < E \leq 1$ in which E = 1 corresponds to a 'perfect' fit of predicted data to the observed data. The value of $0.9 \leq E < 1$ is a 'very satisfactory' prediction, whereas $0.8 \leq E < 0.9$ is 'fairly good' and $-\infty < E < 0.8$ is 'unsatisfactory' (Masoumi Shahr-Babak *et al.* 2016). RMSE reflects the performance of the prediction model. Generally, the smaller the RMSE, the better the performance. MSRE and MAPE indicate the relative absolute accuracy of the models, while RB indicates whether a model is overpredicting or underpredicting. The values of MSRE for a perfect and acceptable model is $0 \leq MSRE < 0.5$ and $0.5 \leq MSRE \leq \infty$, respectively. Ranges of MAPE for perfect and acceptable models are similar to MSRE.

The range of RB is $-\infty < RB < \infty$ (negative values indicate a general overestimation, while positive values indicate a general underestimation of the model).

RESULTS AND DISCUSSION

The combination of five dimensionless parameters which are used as inputs, is shown in Table 2.

The outputs are also Δ/λ and Δ/h. It means that there are 10 runs. The first five runs are the different combinations presented in Table 2 as inputs and one output named Δ/λ

Table 1 | Statistical parameters of the training and testing data sets

	Training					Testing				
Statistical parameters	**Min**	**Max**	**Mean**	**SD**	**CV**	**Min**	**Max**	**Mean**	**SD**	**CV**
Δ/λ	0.004	0.121	0.030	0.020	0.680	0.002	0.100	0.033	0.021	0.640
Δ/h	0.041	0.490	0.180	0.100	0.590	0.050	0.490	0.190	0.120	0.610
Fr^*	0.009	0.044	0.024	0.010	0.420	0.007	0.042	0.024	0.010	0.440
T	0.050	20.600	5.100	4.790	0.940	0.004	21.400	4.690	4.920	1.050
D_*	4.850	44.220	18.810	9.950	0.520	4.520	44.290	18.290	9.780	0.530
θ	0.310	10.190	2.270	2.190	0.960	0.290	10.180	2.070	2.050	0.990
z	0.480	43.890	12.030	9.560	0.790	0.520	48.610	11.740	10.890	0.920

Table 2 | Different combinations of dimensionless parameters

Dimensionless parameters	C1	C2	C3	C4	C5
Fr^*	x	x	x	x	
T	x	x	x	x	x
D_*	x	x	x	x	x
θ		x		x	x
z			x	x	x

and the other five runs are with the output named Δ/h. Figure 3 shows a comparison of measured bedform dimensions and the predicted values from applied methods in the training and testing period for the first combination of inputs (C1). The lines +25% and −25% present the ratio of calculated parameters to measured ones; for example, when a dot is placed between these two lines, it means that the ratio of calculated parameters to measured ones lies in the range between +0.25 and −0.25. It can be observed from Figure 3 that GMDH-HS and GMDH-SCE have better performances during the training and testing period. Regarding the integration of GMDH and HS and SCE, it is reasonable to attain a better performance by taking advantages of the self-organization of GMDH and the global optimization of the HS and SCE methods.

Although GMDH-HS and GMDH-SCE have acceptable performances in predicting bedform dimensions, in order to

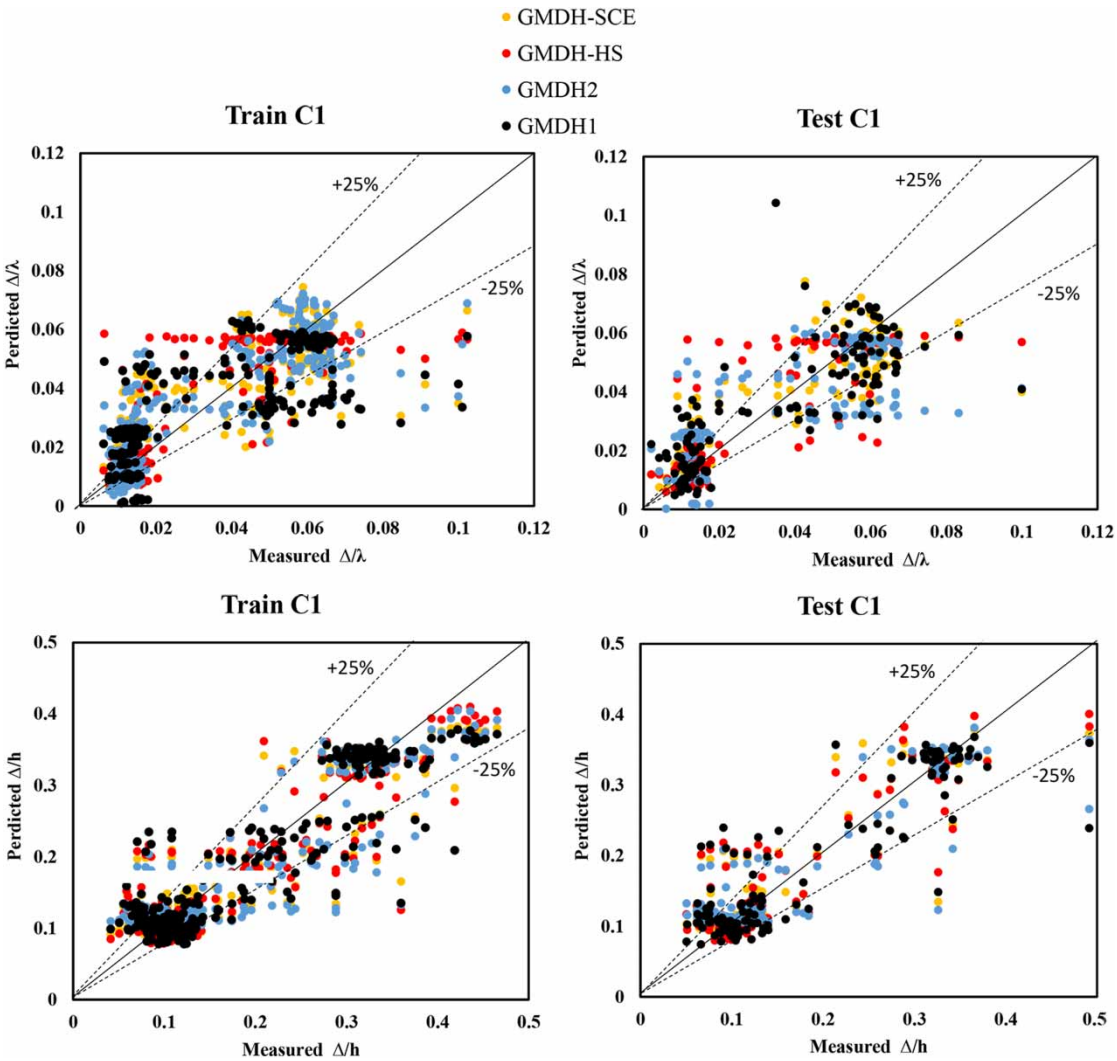

Figure 3 | Measured versus predicted bedform dimensions using applied methods for the first combination (C1).

find the best combination and method for predicting bedform dimensions, the ranking system is applied (Das & Suman 2015). The ranking system for all combinations and applied methods in Δ/λ and Δ/h prediction are presented in Tables 3 and 4, respectively. In these tables, CE is the 'coefficient of efficiency' and can be calculated as $\mathrm{CE} = 1 - \left(\sum_{i=1}^{n}(x_i - y_i)^2 / \sum_{i=1}^{n}(x_i - \bar{x})^2\right)$ in which n is the number of data series, x and y are the observed and predicted outputs respectively and $\bar{x}$ is the mean of observation outputs, R^2 is the determination coefficient, μ and σ are the mean and standard deviation of the ratio of predicted bedform dimensions to measured ones. P50 and P90 are numbers that 50 and 90 percent of relative outputs are larger than, respectively. Relative outputs is the ratio of predicted to observed outputs. For example, for training GMDH1 with C1, P50 is 1.563, which means that 50 percent of relative outputs are larger than 1.563. Lognormal and histogram values (Tables 3 and 4) are also related to the relative data with ±20% accuracy in lognormal distribution and histogram values, respectively, which are referred to the relative data, i.e. the ratio of predicted to observed outputs. Finally the indices R1, R2, R3 and R4 are ranking numbers that indicate the ranking of methods based on the best fit calculation, arithmetic calculation, cumulative probability of the ratio of predicted bedform dimensions to measured ones and prediction of bedform dimensions with accuracy, respectively. For example, in Table 3 in C1, R1 of GMDH1 was calculated to be 20. This R1 indicates that GMDH1 in C1 is 20th based on the fit calculation. Also R1, R2, R3 and R4 are defined to compare several methods. In total the best method has the minimum final rank. More details about this ranking can be found in Abu-Farsakh & Titi (2004) and Das & Suman (2015).

The ranking index (RI) is defined as the sum of four individual rank creations:

$$\mathrm{RI} = \mathrm{R1} + \mathrm{R2} + \mathrm{R3} + \mathrm{R4} \tag{9}$$

where R1, R2, R3 and R4 are based on the best fit calculation, arithmetic calculation, cumulative probability of the ratio of predicted bedform dimensions to measured ones and prediction of bedform dimensions with ±20% accuracy, respectively.

It can be observed from Table 3 that GMDH-HS has the best performance of all applied methods and is more capable of predicting Δ/λ. Also, based on the value of final rank, the fifth combination (C5) yields better answers for predicting Δ/λ.

On the other hand, GMDH-SCE has the best performance in predicting Δ/h as shown in Table 4. For this parameter, the third combination (C3) yields the best answer. So, Equations (10) and (11) are the best functions for prediction of bedform dimensions.

$$\frac{\Delta}{\lambda} = f(T, D_*, \theta, z) \tag{10}$$

$$\frac{\Delta}{h} = g(Fr^*, T, D_*, z) \tag{11}$$

In order to assess the applied methods in this research, the performance and outcome of these methods are compared with the empirical method of van Rijn (1984). Table 5 presents the results of bedform dimension prediction during training and testing periods in terms of various statistical indices. C5 and C3 are used in Table 5.

It can be observed from Table 5 that various AI methods have good performance during both training and testing periods. For prediction of Δ/λ, in the training period, GMDH-SCE obtained the best E, RMSE and MAPE statistics of 0.773, 0.011 and 31.303, respectively, while GMDH-HS and van Rijn (1984) obtained the best MSRE and RB statistics of 0.289 and −0.025, respectively. In the testing period, GMDH-HS obtained the best E, RMSE, MSRE and MAPE with the values of 0.819, 0.01, 0.542 and 34.232, respectively, while van Rijn (1984) obtained the best RB with the value of −0.027.

On the other hand, for predicting Δ/h, in the training period, GMDH-SCE obtained the best E, RMSE, MSRE and MAPE statistics of 0.88, 0.039, 0.104 and 17.279, respectively, while GMDH2 obtained the best RB statistic of −0.06. The results in the testing period are also similar to the training period and the statistics are 0.839, 0.045, 0.172 and 23.427, respectively, while van Rijn (1984) obtained the best value of RB with −0.115. It can be seen from Table 5 that all AI methods outperform the empirical method of van Rijn (1984) and also the performances of GMDH-SCE and GMDH-HS are better than other AI methods in both training and testing periods. In addition, the performance of

Table 3 | Ranking system of applied methods for predicting Δ/λ

		C1				C2				C3				C4				C5			
Period	**Index**	**GMDH1**	**GMDH2**	**GMDH-HS**	**GMDH-SCE**	**GMDH1**	**GMDH2**	**GMDH-HS**	**GMDH-SCE**	**GMDH1**	**GMDH2**	**GMDH-HS**	**GMDH-SCE**	**GMDH1**	**GMDH2**	**GMDH-HS**	**GMDH-SCE**	**GMDH1**	**GMDH2**	**GMDH-HS**	**GMDH-SCE**
Training	CE	0.537	0.701	0.767	0.691	0.565	0.724	0.750	0.727	0.685	0.778	0.729	0.780	0.610	0.712	0.725	0.770	0.600	0.746	0.728	0.774
	R^2	0.537	0.701	0.767	0.691	0.565	0.724	0.750	0.727	0.685	0.778	0.729	0.780	0.610	0.712	0.725	0.770	0.600	0.746	0.728	0.774
Testing	CE	0.517	0.627	0.725	0.717	0.534	0.693	0.735	0.718	0.682	0.741	0.723	0.738	0.641	0.678	0.720	0.728	0.590	0.663	0.819	0.736
	R^2	0.522	0.631	0.728	0.717	0.537	0.693	0.735	0.719	0.682	0.746	0.723	0.742	0.644	0.678	0.720	0.745	0.595	0.668	0.820	0.742
	SUM	2.113	2.660	2.987	2.816	2.201	2.834	2.970	2.891	2.734	3.043	2.904	3.040	2.505	2.780	2.890	3.013	2.385	2.823	3.095	3.026
R1		**20**	**16**	**6**	**13**	**19**	**11**	**7**	**9**	**15**	**2**	**8**	**3**	**17**	**14**	**10**	**5**	**18**	**12**	**1**	**4**
Training	μ	1.283	1.178	1.085	1.167	1.251	1.157	1.122	1.156	1.167	1.126	1.121	1.129	1.240	1.159	1.155	1.131	1.239	1.145	1.140	1.129
	σ	0.770	0.632	0.624	0.588	0.741	0.613	0.616	0.612	0.644	0.533	0.613	0.507	0.684	0.596	0.536	0.563	0.685	0.551	0.520	0.560
Testing	μ	1.281	1.276	1.163	1.210	1.289	1.249	1.173	1.254	1.236	1.356	1.175	1.330	1.258	1.281	1.186	1.296	1.199	1.269	1.183	1.340
	σ	1.100	1.071	0.767	0.676	1.115	0.995	0.754	1.012	0.976	1.657	0.747	1.472	1.019	0.982	0.718	1.616	0.881	1.081	0.716	1.654
	SUM	4.434	4.157	3.639	3.641	4.396	4.014	3.665	4.034	4.023	4.672	3.656	4.438	4.201	4.018	3.595	4.606	4.004	4.046	3.559	4.683
R2		**16**	**13**	**3**	**4**	**15**	**8**	**6**	**11**	**10**	**19**	**5**	**17**	**14**	**9**	**2**	**18**	**7**	**12**	**1**	**20**
Training	P50	1.564	1.261	1.140	1.275	1.525	1.206	1.111	1.257	1.223	1.199	1.194	1.203	1.441	1.240	1.181	1.138	1.388	1.176	1.191	1.133
	P90	2.507	2.314	1.854	2.232	2.476	2.248	2.073	2.311	2.342	2.097	2.064	2.040	2.417	2.111	2.152	2.276	2.591	2.154	2.076	2.173
Testing	P50	1.536	1.231	1.149	1.251	1.364	1.169	1.134	1.174	1.097	1.246	1.251	1.259	1.330	1.228	1.130	1.106	1.298	1.164	1.164	1.118
	P90	2.715	2.862	2.266	2.266	2.786	2.656	2.178	2.665	2.781	2.688	2.112	2.639	2.822	2.494	2.396	2.479	2.756	2.662	2.157	2.553
	SUM	8.322	7.668	6.409	7.024	8.151	7.279	6.496	7.407	7.443	7.230	6.621	7.141	8.010	7.073	6.859	6.999	8.033	7.156	6.588	6.977
R3		**20**	**16**	**1**	**8**	**19**	**13**	**2**	**14**	**15**	**12**	**4**	**10**	**17**	**9**	**5**	**7**	**18**	**11**	**3**	**6**
Training	Lognormal	30	40	54	40	33	45	59	42	47	54	44	55	37	49	50	54	41	54	54	55
	Histogram	33	43	57	43	36	48	62	45	50	57	47	58	40	52	53	57	44	57	57	58
Testing	Lognormal	32	43	51	42	34	53	55	49	48	56	40	51	38	47	54	54	45	52	55	55
	Histogram	35	46	54	45	37	56	58	52	51	59	43	54	41	50	57	57	48	55	58	58
	SUM	130	172	216	170	140	202	234	188	196	226	174	218	156	198	214	222	178	218	224	226
R4		**18**	**14**	**6**	**15**	**17**	**8**	**1**	**11**	**10**	**2**	**13**	**5**	**16**	**9**	**7**	**4**	**12**	**5**	**3**	**2**
RI		74	59	16	40	70	40	16	45	50	35	30	35	64	41	24	34	55	40	8	32
Final Rank		**16**	**13**	**2**	**8**	**15**	**8**	**2**	**10**	**11**	**7**	**4**	**7**	**14**	**9**	**3**	**6**	**12**	**8**	**1**	**5**

Table 4 | Ranking system of applied methods for predicting Δ/h

		C1				C2				C3				C4				C5			
Period	**Index**	**GMDH1**	**GMDH2**	**GMDH-HS**	**GMDH-SCE**	**GMDH1**	**GMDH2**	**GMDH-HS**	**GMDH-SCE**	**GMDH1**	**GMDH2**	**GMDH-HS**	**GMDH-SCE**	**GMDH1**	**GMDH2**	**GMDH-HS**	**GMDH-SCE**	**GMDH1**	**GMDH2**	**GMDH-HS**	**GMDH-SCE**
Training	CE	0.822	0.858	0.834	0.853	0.817	0.850	0.826	0.854	0.803	0.860	0.845	0.880	0.838	0.870	0.815	0.877	0.789	0.822	0.838	0.863
	R^2	0.822	0.858	0.834	0.853	0.817	0.850	0.826	0.854	0.803	0.860	0.845	0.880	0.838	0.870	0.815	0.877	0.789	0.822	0.838	0.863
Testing	CE	0.785	0.840	0.805	0.839	0.799	0.835	0.830	0.844	0.788	0.831	0.810	0.839	0.819	0.835	0.833	0.836	0.752	0.831	0.822	0.830
	R^2	0.789	0.847	0.813	0.849	0.802	0.842	0.836	0.850	0.792	0.834	0.816	0.844	0.824	0.845	0.839	0.840	0.753	0.833	0.828	0.835
	SUM	3.218	3.403	3.286	3.394	3.235	3.377	3.318	3.402	3.186	3.385	3.316	3.443	3.319	3.420	3.302	3.430	3.083	3.308	3.326	3.391
R1		**18**	**4**	**16**	**6**	**17**	**9**	**12**	**5**	**19**	**8**	**13**	**1**	**11**	**3**	**15**	**2**	**20**	**14**	**10**	**7**
Training	μ	1.090	1.075	1.098	1.075	1.106	1.079	1.091	1.082	1.101	1.061	1.076	1.063	1.092	1.073	1.097	1.069	1.110	1.098	1.092	1.078
	σ	0.379	0.349	0.333	0.329	0.402	0.348	0.319	0.356	0.379	0.340	0.351	0.317	0.366	0.343	0.362	0.329	0.450	0.425	0.418	0.373
Testing	μ	1.152	1.150	1.173	1.155	1.163	1.149	1.148	1.143	1.152	1.124	1.157	1.136	1.163	1.164	1.150	1.139	1.140	1.136	1.145	1.132
	σ	0.428	0.400	0.387	0.369	0.434	0.388	0.339	0.384	0.408	0.410	0.429	0.394	0.414	0.429	0.374	0.407	0.468	0.405	0.415	0.381
	SUM	3.049	2.974	2.991	2.928	3.105	2.964	2.897	2.965	3.040	2.935	3.013	2.910	3.035	3.009	2.983	2.944	3.168	3.064	3.070	2.964
R2		**16**	**9**	**11**	**3**	**19**	**6**	**1**	**8**	**15**	**4**	**13**	**2**	**14**	**12**	**10**	**5**	**20**	**17**	**18**	**7**
Training	P50	1.118	1.094	1.160	1.109	1.129	1.104	1.160	1.097	1.156	1.123	1.108	1.074	1.116	1.083	1.124	1.078	1.121	1.124	1.104	1.085
	P90	1.799	1.821	1.722	1.686	1.968	1.813	1.596	1.815	1.919	1.624	1.776	1.511	1.780	1.778	1.864	1.731	1.916	1.727	1.789	1.729
Testing	P50	1.133	1.170	1.240	1.162	1.167	1.177	1.215	1.180	1.177	1.140	1.153	1.142	1.163	1.169	1.169	1.147	1.223	1.214	1.176	1.144
	P90	1.983	1.996	1.974	1.928	2.153	1.934	1.953	1.961	1.967	2.032	2.094	1.934	2.024	2.085	1.978	1.983	2.132	1.890	2.021	1.939
	SUM	6.033	6.081	6.096	5.885	6.417	6.028	5.924	6.053	6.219	5.919	6.131	5.661	6.083	6.115	6.135	5.939	6.392	5.955	6.090	5.897
R3		**9**	**11**	**14**	**2**	**20**	**8**	**5**	**10**	**18**	**4**	**16**	**1**	**12**	**15**	**17**	**6**	**19**	**7**	**13**	**3**
Training	Lognormal	64	69	66	69	60	70	59	66	61	65	70	78	66	71	66	72	63	64	67	72
	Histogram	67	72	69	72	63	73	62	69	64	68	73	81	69	74	69	75	66	67	70	75
Testing	Lognormal	62	62	60	64	58	62	58	59	60	56	62	72	62	61	64	65	58	58	63	66
	Histogram	65	65	63	67	61	65	61	62	63	59	65	75	65	64	67	68	61	61	66	69
	SUM	258	268	258	272	242	270	240	256	248	248	270	306	262	270	266	280	248	250	266	282
R4		**9**	**6**	**9**	**4**	**13**	**5**	**14**	**10**	**12**	**12**	**5**	**1**	**8**	**5**	**7**	**3**	**12**	**11**	**7**	**2**
RI		52	30	50	15	69	28	32	33	64	28	47	5	45	35	49	16	71	49	48	19
Final Rank		**15**	**6**	**14**	**2**	**17**	**5**	**7**	**8**	**16**	**5**	**11**	**1**	**10**	**9**	**13**	**3**	**18**	**13**	**12**	**4**

Table 5 | The comparison of applied methods with the empirical method of van Rijn (1984)

Statistical indices		Δ/λ (C5)					Δ/h (C3)				
		GMDH1	GMDH2	GMDH-HS	GMDH-SCE	van Rijn	GMDH1	GMDH2	GMDH-HS	GMDH-SCE	van Rijn
E	Training	0.600	0.746	0.728	**0.773**	−1.159	0.803	0.859	0.844	**0.880**	−1.337
RMSE		0.015	0.012	0.012	**0.011**	0.034	0.050	0.085	0.044	**0.039**	0.171
MSRE		0.524	0.324	**0.289**	0.330	0.429	0.153	0.119	0.128	**0.104**	0.354
MAPE		47.180	33.667	31.603	**31.303**	58.488	25.050	21.098	21.931	**17.279**	53.117
RB		−0.238	−0.145	−0.140	−0.129	**−0.025**	−0.100	**−0.060**	−0.076	−0.063	−0.120
E	Testing	0.589	0.663	**0.819**	0.736	−1.352	0.788	0.831	0.809	**0.839**	−1.257
RMSE		0.015	0.013	**0.010**	0.012	0.036	0.052	0.094	0.049	**0.045**	0.169
MSRE		0.809	1.232	**0.542**	2.831	0.815	0.188	0.182	0.207	**0.172**	0.346
MAPE		48.723	45.957	**34.232**	51.537	67.371	27.658	25.911	27.292	**23.427**	51.953
RB		−0.199	−0.269	−0.18	−0.340	**−0.027**	−0.151	−0.124	−0.157	−0.135	**−0.115**

GMDH-HS during training and testing periods is shown in Figures 4 and 5.

CONCLUSION

Although empirical formulae often provide useful predictions of bedform dimensions in alluvial channels, the complexity of the interaction between flow characteristics and development of bedforms is such that these formulae cannot provide the accuracy required. In this study, two hybrid intelligence methods were developed using GMDH, HS and SCE. In the prediction of bedform dimensions, unlike empirical methods, there are no limitations in the ranges of inputs using AI techniques. For this reason, different combinations of the most frequently used dimensionless parameters in the literature were examined. Results reveal the following:

(1) The combination of T, D_*, θ and z is more accurate for predicting Δ/λ, while the combination of T, D_*, Fr^* and z has a better performance in predicting Δ/h. Although these combinations have the best performances in predicting

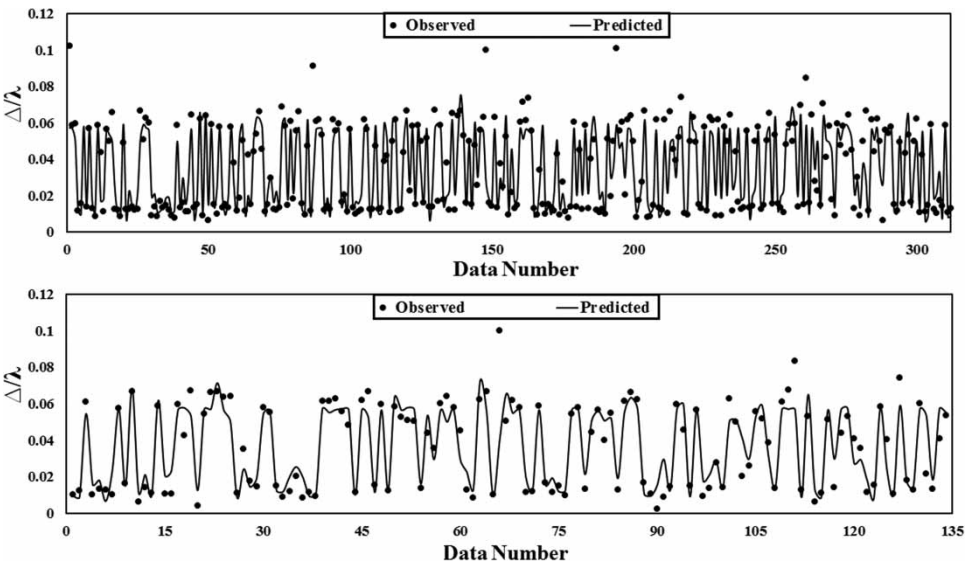

Figure 4 | GMDH-HS performance in Δ/λ prediction for training and testing data (C5 as inputs).

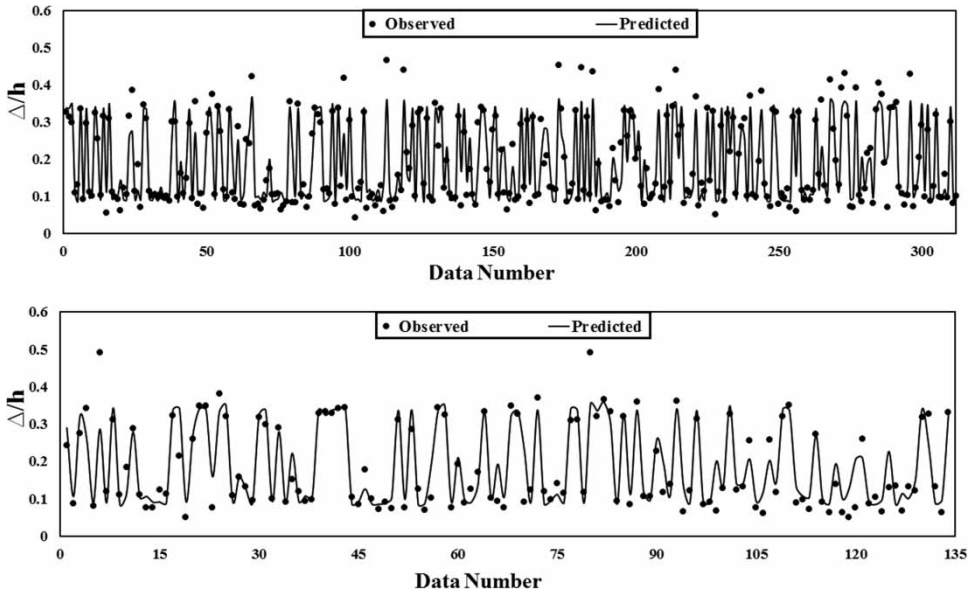

Figure 5 | GMDH-HS performance in Δ/h prediction for training and testing data (C3 as inputs).

Δ/λ and Δ/h, other combinations also have acceptable performances. So, in situations where researchers lack data, using other combinations can also yield appropriate answers.

(2) For calculation of T, the logarithmic distribution of velocity profile or the boundary-layer characteristics method can be easily used in order to calculate u'_* with regard to the effects of height of roughness caused by bedforms.

(3) The accuracy of all four AI methods in predicting the parameters Δ/λ and Δ/h in flumes and rivers with each combination of dimensionless parameters is extremely high and acceptable.

(4) The performance of GMDH-SCE for predicting bedform dimensions is better than other methods and GMDH-HS is in second place.

5) All AI methods have much better performances than the empirical method of van Rijn (1984). However, GMDH-HS and GMDH-SCE outperform all other methods for predicting bedform dimensions.

REFERENCES

Abu-Farsakh, M. Y. & Titi, H. H. 2004 Assessment of direct cone penetration test methods for predicting the ultimate capacity of friction driven piles. *Journal of Geotechnical and Geoenvironmental Engineering* **130** (9), 935–944. https://doi.org/10.1061/(ASCE)1090-0241(2004)130:9(935).

Afzalimehr, H. & Anctil, F. 2000 Accelerating shear velocity in gravel bed channels. *Journal of Hydrological Sciences* **45** (1), 113–124. http://dx.doi.org/10.1080/02626660009492309.

Ayvaz, M. T. 2009 Application of harmony search algorithm to the solution of groundwater management models. *Advances in Water Resources* **32** (6), 916–924. https://doi.org/10.1016/j.advwatres.2009.03.003.

Das, S. K. & Suman, S. 2015 Prediction of lateral load capacity of pile in clay using multivariate adaptive regression spline and functional network. *Arabian Journal of Science & Engineering* **40** (6), 1565–1578. DOI: 10.1007/s13369-015-1624-y.

Duan, Q. Y., Gupta, V. K. & Sorooshian, S. 1993 Shuffled complex evolution approach for effective and efficient global minimization. *Journal of Optimization Theory & Applications* **76** (3), 501–521. DOI: 10.1007/BF00939380.

Ebtehaj, I., Bonakdari, H., Zaji, A. H., Azimi, H. & Khoshbin, F. 2015 GMDH-type neural network approach for modeling the discharge coefficient of rectangular sharp-crested side weirs. *Engineering Science & Technology, an International Journal* **18** (4), 746–757. https://doi.org/10.1016/j.jestch.2015.04.012.

Fredsoe, J. 1975 *The Friction Factor and Height-Length Relations in Flow Over a Dune-Covered Bed.* Institute of Hydrodynamics, Technical University of Denmark, Copenhagen, Denmark, Progress report.

Geem, Z. W., Kim, J. H. & Loganathan, G. V. 2001 A new heuristic optimization algorithm: harmony search. *Simulation* **76** (2), 60–68.

Ivakhnenko, A. G. 1968 The group method of data handling, a rival of the method of stochastic approximation. *Soviet Automatic Control* **1** (3), 43–55.

Javadi, F., Ahmadi, M. M. & Qaderi, K. 2015 Estimation of river bedform dimension using artificial neural network (ANN) and support vector machine (SVM). *Journal of Agricultural Science & Technology* **17** (4), 859–868.

Julien, P. Y. 1992 *Study of Bedform Geometry in Large Rivers*. Delft Hydraulics, The Netherlands.

Julien, P. Y. & Klaassen, G. J. 1995 Sand-Dune geometry of large rivers during floods. *Journal of Hydraulic Engineering* **121** (9), 657–663. https://doi.org/10.1061/(ASCE)0733-9429(1995)121:9(657).

Kan, G., Liang, K., Li, J., Ding, L., He, X., Hu, Y. & Amo-Boateng, M. 2016 Accelerating the SCE-UA global optimization method based on multi-core CPU and many-core GPU. *Advances in Meteorology* **2016**. http://dx.doi.org/10.1155/2016/8483728.

Karim, F. 1995 Bed configuration and hydraulic resistance in alluvial–channel flows. *Journal of Hydraulic Engineering* **121** (1), 15–25. https://doi.org/10.1061/(ASCE)0733-9429(1995)121:1(15).

Klaassen, G. J. 1990 *Experiment with Graded Sediments in a Straight Flume*. Delft Hydraulics, Volume B, The Netherlands.

Masoumi Shahr-Babak, M., Khanjani, M. J. & Qaderi, K. 2016 Uplift capacity prediction of suction caisson in clay using a hybrid intelligence method (GMDH-HS). *Applied Ocean Research* **59**, 408–416. https://doi.org/10.1016/j.apor.2016.07.005.

Muttil, N. & Jayawardena, A. W. 2008 Shuffled complex evolution model calibrating algorithm: enhancing its robustness and efficiency. *Hydrological Processes* **22** (23), 4628–4638. DOI: 10.1002/hyp.7082.

Samsudin, R., Saad, P. & Shabri, A. 2011 River flow time series using least squares support vector machines. *Hydrology & Earth System Sciences* **15**, 1835–1852. DOI: 10.5194/hess-15-1835-2011.

Shahabi, S., Khanjani, M. J. & Hessami Kermani, M. R. 2016 Hybrid wavelet-GMDH model to forecast significant wave height. *Water Science & Technology: Water Supply* **16** (2), 453–459. DOI: 10.2166/ws.2015.151.

Su, J., Wang, X., Zhao, S., Chen, B., Li, C. & Yang, Z. 2015 A structurally simplified hybrid model of genetic algorithm and support vector machine for prediction of chlorophyll a in reservoirs. *Water* **7** (4), 1610–1627. DOI:10.3390/w7041610.

Talebbeydokhti, N., Hekmatzadeh, A. A. & Rakhshandehroo, G. R. 2006 Experimental modeling of dune bed form in a sand-bed channel. *Iranian Journal of Science & Technology, Transactions of Civil Engineering* **30** (4), 503–516.

van der Mark, C. F., Blom, A. & Hulscher, S. J. M. H. 2008 Quantification of variability in bedform geometry. *Journal of Geophysical Research, Earth Surface* **113** (F3). DOI: 10.1029/2007JF000940.

van Rijn, L. C. 1984 Sediment transport, part III: bedforms and alluvial roughness. *Journal of Hydraulic Engineering* **110** (2), 1733–1754. https://doi.org/10.1061/(ASCE)0733-9429(1984)110:12(1733).

Wu, C. l. & Chau, K. W. 2006 A flood forecasting neural network model with genetic algorithm. *International Journal of Environment & Pollution* **28** (3–4), 261–273. https://doi.org/10.1504/IJEP.2006.011211.

Yalin, M. S. 1992 *River mechanics*. Pergamon Press Ltd., Oxford.

Zhang, H., Liu, X., Cai, E., Huang, G. & Ding, C. 2013 Integration of dynamic rainfall data with environmental factors to forecast debris flow using an improved GMDH model. *Computers & Geosciences* **56**, 23–31. https://doi.org/10.1016/j.cageo.2013.02.003.

First received 7 June 2017; accepted in revised form 13 November 2017. Available online 27 November 2017