



OPEN

Modeling of wave run-up by applying integrated models of group method of data handling

Amin Mahdavi-Meymand¹, Mohammad Zounemat-Kermani², Wojciech Sulisz¹✉ & Rodolfo Silva³

Wave-induced inundation in coastal zones is a serious problem for residents. Accurate prediction of wave run-up height is a complex phenomenon in coastal engineering. In this study, several machine learning (ML) models are developed to simulate wave run-up height. The developed methods are based on optimization techniques employing the group method of data handling (GMDH). The invasive weed optimization (IWO), firefly algorithm (FA), teaching–learning-based optimization (TLBO), harmony search (HS), and differential evolution (DE) meta-heuristic optimization algorithms are embedded with the GMDH to yield better feasible optimization. Preliminary results indicate that the developed ML models are robust tools for modeling the wave run-up height. All ML models' accuracies are higher than empirical relations. The obtained results show that employing heuristic methods enhances the accuracy of the standard GMDH model. As such, the FA, IWO, DE, TLBO, and HS improve the *RMSE* criterion of the standard GMDH by the rate of 47.5%, 44.7%, 24.1%, 41.1%, and 34.3%, respectively. The GMDH-FA and GMDH-IWO are recommended for applications in coastal engineering.

Wave-induced inundation in coastal zones is a very complex phenomenon of fundamental importance for many disciplines. The design of coastal areas and coastal zone management strongly depend on the storm and tide-induced inundation studies. Moreover, wave-induced inundation is of fundamental importance for the erosion processes, the maintenance of beaches, and the biological processes in surges zones. Proper understanding of inundation and the prediction of flooding is crucial for hydraulic and coastal engineering and the sustainable development of coastal areas.

Coastal areas are very vulnerable zones to storm and tide-induced inundation. Wave-induced inundation may cause severe damage in these areas and the consequences of severe flooding are often fatal for the coastal zone population, which constitutes from 10 to 20% of the global population. In fact, presently about 40% of the world's population lives within 100 km of the coast. One of the key elements describing storm and tide-induced inundation in coastal areas is a wave run-up height. Wave run-up is equal to the vertical distance of the water's edge on a coastal structure or on the foreshore of the beach, which is varying with time. In Fig. 1, a schematic view of the run-up (*R*) process is shown.

In Fig. 1, *a* is the beach or structure slope, and *SWL* is the water level. Structure overtopping, dune erosion during storm conditions¹, and sediment transport in the coastal zone² are examples of some phenomena which the run-up processes may cause. In this respect, designing the protection structures in coastal areas cannot be made without considering the accurate prediction of the wave run-up parameters. Hence, the simulation of the general characteristics of the run-up is of great importance in coastal engineering.

In the past decades, numerous studies applied experimental tests^{3–5} and numerical modeling for simulating and predicting the wave run-up in coastal regions. In recent years modeling complex and nonlinear problems in science and engineering using machine learning (ML) models have received widespread attention and, in general, the capability of these approaches has been confirmed. Selected recent applications of ML in coastal engineering are discussed below.

Chang and Lin⁶ used an artificial neural network (ANN), harmonic model (HM), and global ocean tidal model (NAO.99b) to simulate tides at selected points. The comparisons of the results obtained by ANN with corresponding results obtained by applying HM and NAO.99b confirmed the superiority of the ANN model. Erdik and Savci⁷ simulated wave run-up on armored rock slopes by applying Takagi–Sugeno (TS) fuzzy models.

¹Institute of Hydro-Engineering, Polish Academy of Sciences, Warsaw, Poland. ²Water Engineering Department, Shahid Bahonar University of Kerman, Kerman, Iran. ³Instituto de Ingeniería, Universidad Nacional Autónoma de México, 04510 Mexico City, Mexico. ✉email: sulisz@ibwpan.gda.pl

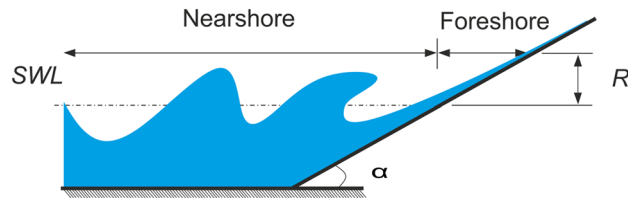


Figure 1. A schematic presentation of wave run-up on the foreshore.

The results showed that TS models provided better accuracy in simulating a run-up than the empirical methods of Van der Meer and Stam⁸. Shiri et al.⁹ used an adaptive neuro-fuzzy inference system (ANFIS), ANN, linear regression, and autoregressive methods for modeling sea-level variations. The results confirmed the superiority of ML techniques in comparison with the regression methods. Bonakdar and Etemad-Shahidi¹⁰ used the M5 model tree, Takagi–Sugeno fuzzy model, and empirical formulae to simulate wave run-up on rubble-mound structures. The study showed that the M5 results were better than the corresponding results obtained by the application of the remaining methods. Elbisy¹¹ investigated the ability of multiple additive regression trees (MART) and ANN in the simulation of wave run-up. The results showed that MART is more accurate than ANN. Abolfathi et al.¹² used the M5 decision tree algorithm for predicting wave run-up. The study showed that the M5 method provides more accurate results than a regression-based model. Pourzangbar et al.¹³ simulated non-breaking wave-induced scour depth at breakwaters with genetic programming and ANN methods. The study indicated that ML methods provide better results than empirical relations. Zhang et al.¹⁴ predicted real-time tidal levels using an integrative grey-GMDH neural network. The results revealed that the applied method can predict real-time tidal levels with high accuracy. In another study, Wang et al.¹⁵ predicted seawater levels by developing a hybrid model of ANFIS and wavelet decomposition. The results showed that the integrated method improved the accuracy of ANFIS. Gao et al.¹⁶ developed a multi-blocks fuzzy cognitive map model for predicting the steady turning motion of ships. The results confirmed the high performance of the proposed algorithm. Zeinali et al.¹⁷ applied two recurrent ANNs for the prediction of shoreline changes at Narrabeen, Australia. The results confirmed a good accuracy of the applied recurrent ANNs. Rehman et al.¹⁸ applied ANN and the response surface methodology (RSM) for wave run-up prediction. The obtained results showed that both the ANN and RSM are appropriate methods for the prediction of wave run-up. Also, the study of Yao et al.¹⁹ confirmed a good accuracy of ANN in simulating wave run-up.

Recently, the applications of meta-heuristic algorithms integrated with ML methods have been reported in many studies. Masoumi Shahr-Babak et al.²⁰ applied hybrid GMDH-HS to predict the uplift capacity of suction caisson in clay. They reported that the hybrid model can predict the suction caisson uplift capacity with acceptable accuracy. Najafzadeh et al.²¹ predicted bridge pier scour depth by developing several integrative GMDH methods. The study showed that particle swarm optimization (PSO) provides more accurate results than a genetic algorithm (GA) and gravitational search algorithm (GSA). Mahdavi-Meymand et al.²² used several meta-heuristic optimization algorithms to optimize ANFIS parameters to estimate the spillway aerator air demand in dams. They showed that hybrid models provide more accurate results than corresponding ML models. Dodangeh et al.²³ applied GA and harmony search (HS) algorithm to optimize GMDH and support vector regression parameters for flood-susceptibility prediction. The results showed that both the GA and HS algorithms improved the accuracy of GMDH and SVR. Sharafati et al.²⁴ confirmed the good accuracy of ANFIS-BBO in the prediction of long contraction scour depth. Sharafati et al.²⁵ used teaching–learning-based optimization (TLBO), biogeography-based optimization (BBO), and invasive weed optimization (IWO) algorithms to optimize ANFIS parameters for the prediction of scour depth downstream of weirs. The results showed that ANFIS-IWO is a reliable technique for the prediction of scour depth.

Qaderi et al.²⁶ developed a shuffled complex evolutionary (SCE) algorithm integrated with GMDH to predict bridge pier scour depth. The results showed a good performance of GMDH-SCE. Alizadeh et al.²⁷ compared the performance of PSO, GA, and imperialist competitive algorithm (ICA) integrated with a support vector machine (SVM) for the estimation of drilling fluid density. The results showed the high performance of PSO in estimating drilling fluid density. Milan et al.²⁸ applied three optimization algorithms, comprising particle swarm optimization (PSO), gray wolf optimization (GWO), and Harris hawk optimization (HHO) integrated with ANFIS for predicting optimal groundwater exploitation. The results indicated that all optimization algorithms increase the ANFIS accuracy. Haghbin et al.²⁹ developed support vector regression (SVR) integrated with IWO for predicting channel sinuosity. The results showed that IWO significantly increases the accuracy of SVR.

In this study, invasive weed optimization (IWO), firefly algorithm (FA), teaching–learning-based optimization (TLBO), harmony search (HS), and differential evolution (DE) optimization techniques are embedded with GMDH to predict a two percent wave run-up height in coastal regions. It is worth noting that the applications of GMDH integrated with novel meta-heuristic models are scarce. Indeed, to the best knowledge of the authors, no study has ever developed and presented the application of hybrid models of GMDH-IWO for modeling problems in any field of research.

Methods

In this section, a brief review of empirical relations for the prediction of wave run-up as well as the description of GMDH and optimization algorithms are presented.

Empirical relations. In coastal engineering, several different wave run-up values can be investigated, such as the mean value (R), the 33 percent value of the wave run-up i.e. the significant run-up (R_s), the two percent value of the wave run-up ($R_{2\%}$), etc. The mean run-up (R) can be calculated as the average run-up of all observed waves. However, the mean value is of limited interest for engineers and scientists. In this respect, other parameters might be used e.g. R_s , $R_{2\%}$ and $R_{10\%}$. Note that $R_{i\%}$ refers to the run-up level reached and exceeded by $i\%$ of the incoming waves^{30–36}. Numerous studies have been conducted on the regular wave run-up on smooth and rough beaches. Miche³¹ presented an equation to predict wave run-up for non-breaking waves:

$$\frac{R_{\max}}{H} = \sqrt{\frac{\pi}{2\alpha}} \quad (1)$$

where R_{\max} is the maximum vertical run-up, H is the wave height, and α is the beach slope.

Hunt³⁷ presented empirical relations to predict wave run-up on impermeable slopes based on breaking wave shape in the surf zone. The equation for standing waves on a steep slope was proposed in the following form:

$$\frac{R_{\max}}{H} \approx 3 \quad (2)$$

where R_{\max} may be calculated from:

$$R_{\max} = \frac{\tan \alpha}{\sqrt{H/L_0}} = \xi_0 \quad (3)$$

in which ξ_0 is the surf similarity parameter or Iribarren number³⁸, L_0 is the deep-water wavelength

$$L_0 = \frac{gT^2}{2\pi} \quad (4)$$

g is the gravitational acceleration and T is the wave period.

One of the first equations presented for estimating irregular wave run-up on mild uniform slopes ($\tan \alpha \leq 1/3$) was proposed by Wassing³⁹:

$$R_{u2\%} = 8H_{1/3} \tan \alpha \quad (5)$$

Ahrens⁴⁰ studied irregular wave run-up on smooth-impermeable slopes ($1/4 \leq \tan \alpha \leq 1.1$) and suggested two equations for breaking and non-breaking waves. Coastal Engineering Manual⁴¹ analyzed Ahrens⁴⁰ data and proposed two relations for predicting irregular wave run-up. Mase⁴² performed laboratory experiments on the irregular run-up on mild slopes ($2^\circ \leq \alpha \leq 11.4^\circ$) and proposed an equation for estimating run-up due to breaking waves.

Van der Meer and Stam⁸ suggested the following relations for estimating wave run-up on smooth slopes for irregular waves:

$$\frac{R_{u2\%}}{H_s} = 1.5\gamma\xi_{\max} \quad (6)$$

where H_s is the significant wave height, γ is the reduction factor that depends on various parameters such as roughness, shallow water conditions, oblique wave attack, berms, and ξ_{\max} is known as the Iribarren number corresponding to the maximum wave period.

In another study, Van der Meer and Stam⁸ applied the regression model proposed by Van der Meer³ and derived the following relations for predicting irregular wave run-up on permeable and impermeable slopes:

$$\begin{aligned} \frac{R_{u2\%}}{H_s} &= 0.96\xi_m & 1 \leq \xi_m \leq 1.5 \\ \frac{R_{u2\%}}{H_s} &= 1.17\xi_m^{0.46} & 1.5 < \xi_m \leq 3.1 \\ \frac{R_{u2\%}}{H_s} &= 1.97\xi_m^{0.46} & 3.1 < \xi_m \leq 7.5 \end{aligned} \quad (7)$$

where ξ_m is the Iribarren number corresponding to the average wave period.

Schimmels et al.⁵ proposed the following relations:

$$\frac{R_{u2\%}}{H_{mo}} = \gamma_p [1.65\xi_{m-1,0}] \quad \text{with a maximum of} \quad \frac{R_{u2\%}}{H_{mo}} = A\gamma_p \left[4 - \frac{1.5}{\sqrt{\xi_{m-1,0}}} \right] \quad (8)$$

where H_{mo} is the wave height, γ_p is the porosity coefficient, $\xi_{m-1,0}$ is the Iribarren number.

In this study, the formulas derived by Van der Meer and Stam⁸, and Schimmels et al.⁵ are applied to determine the two percent wave run-up height.

Group method of data handling. The group method of data handling (GMDH) is a machine learning model belonging to artificial neural networks (ANNs), which was introduced by Ivakhnenko⁴³ for modeling complex systems. This method has been successfully applied in different fields of science and engineering. Similar to ANNs, the GMDH consists of neurons connected in different layers. In the GMDH network, the neurons in the next layer are produced as a combination of two neurons from the previous layer. Then, the output of each neuron is calculated by quadratic polynomial expressions, and the most effective neurons are selected to be con-

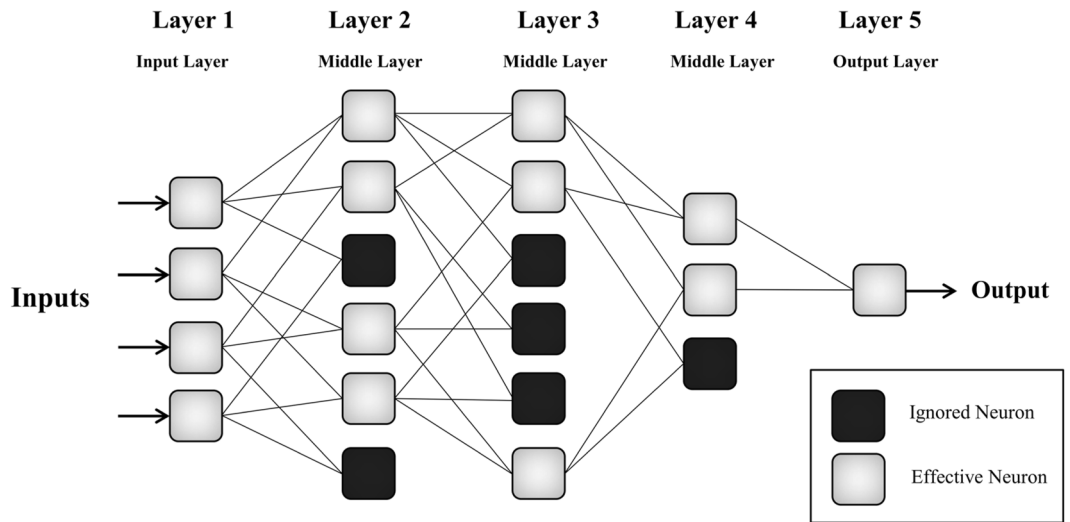


Figure 2. A schematic plot of GMDH structure with three hidden layers.

connected to neurons in the next layer. In other words, the GMDH algorithm generates the structure of the network through successive generations of quadratic regression polynomials with two input variables or neurons. In Fig. 2, a schematic plan of a five-layer GMDH with effective and ignored neurons is shown.

The GMDH network is created and trained layer-by-layer and neuron-by-neuron. This strategy allows users to have access to the neurons’ information during the network training process. As can be seen in Fig. 2, in the second, third, and fourth layers, there are some ignored neurons. The ignoring of neurons and the selection of appropriate ones prevent the immense growth of the network. The number of ignored neurons in each middle layer is affected by the error evaluation criteria e.g. *MSE* criterion:

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i^m - y_i^o)^2 \tag{9}$$

where N is the number of data, y^p is the predicted output of neurons in each layer, and y^m is the observed value. In the GMDH structure, contrary to conventional matrix structure, a number of mathematical equations are applied to speed up calculation process⁴⁴. The output of each neuron is calculated according to Volterra–Kolmogorov–Gabor (VKG) polynomial. The second-order polynomial is incorporated in the structure of GMDH⁴³ and is used in this study as the transfer functions in each neuron. The second-order polynomial may be written in the following form:

$$y = c_0 + c_1x_1 + c_2x_2 + c_3x_1^2 + c_4x_2^2 + c_5x_1x_2 \tag{10}$$

where y is the output, (x_1, x_2) is the input vector, and c is the weighting coefficient. The intricacy of the neurons will be increased layer by layer, which causes that the final network is becoming complex⁶⁷. The weighting coefficients are calculated using regression techniques:

$$c = (A^T A)^{-1} A^T Y \tag{11}$$

where c represents the weighting coefficient vector, A denotes the following matrix:

$$A = \begin{bmatrix} 1 & x_1^1 & x_2^1 & x_1^1 x_2^1 & (x_1^1)^2 & (x_2^1)^2 \\ 1 & x_1^2 & x_2^2 & x_1^2 x_2^2 & (x_1^2)^2 & (x_2^2)^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_1^m & x_2^m & x_1^m x_2^m & (x_1^m)^2 & (x_2^m)^2 \end{bmatrix} \tag{12}$$

and Y is the matrix of outputs:

$$Y = \{y_1, y_2, \dots, y_m\}^T \tag{13}$$

in which m is the number of samples.

In this study, the number of layers is equal to 5. The maximum number of neurons in the next layer (N_{np}^{i+1}) can be calculated by applying the following equation:

$$N_{np}^{i+1} = \binom{N_{np}^i}{2} = \frac{N_{np}^i!}{2! \times (N_{np}^i - 2)!} \quad (14)$$

where N_{np}^i is the number of neurons in the i layer. Substituting N_{np}^i equal to eight i.e. the same as the number of input variables in Eq. (14), N_{np}^{i+1} results in 28. In this study, the maximum number of neurons in each layer was determined to be 10.

In summary, the following seven steps are taken to build the GMDH network (1) determining the GMDH structure—in this study 5 layers are considered for the network with a maximum of 10 neurons in each layer; (2) standardization of the data; (3) entering the data to the neurons of the next layer; (4) allocating the polynomial-based fit to each neuron in layers based on the values of two neurons of the previous layer; (5) calculating weights for a polynomial—Eq. 11; (6) calculating the output of the neurons and selected appropriateness of them—Eq. 9; (7) move to the next layer and repeat the steps of 3 to 6 to create the entire GMDH network.

Hybrid technique. Instead of using the least-squares technique of GMDH, meta-heuristic optimization algorithms can be embedded with the GMDH model. This technique can be applied to optimize either the weighting coefficients in Eq. 10 or the structure of the network. In this study, meta-heuristic optimization algorithms are used to optimize the weighting coefficients. The main difference between the general execution procedure of the standard GMDH and hybrid GMDH is the calculation process of the fifth step of the GMDH network mentioned in the previous section. In hybrid GMDHs, the meta-heuristic optimization approaches will be employed to optimize the weights of the polynomial. At the first step, a number of candidate solutions are distributed in the search space. Each member of this population represents a solution to Eq. 10. The fitness of members is calculated by applying *RMSE*, and the population is ranked. At the next step, the new values of the members will be calculated by meta-heuristic algorithms. Henceforth, the fitness of the population will be calculated and the members will be ranked. This process will be repeated until the final iteration. In the end, the best member represents the optimized values of Eq. 10. In Fig. 3, the general flowchart for setting up the hybrid GMDH models is shown.

The applied algorithms comprise new swarm intelligent-based models including invasive weed optimization, IWO, firefly algorithm, FA, and teaching–learning-based optimization, TLBO, as well as evolutionary-based models including harmony search, HS, and differential evolution, DE. The integration of the GMDH model and IWO has been developed and executed for the first time in this study. In Table 1, descriptions of the applied algorithms are presented, and a brief description of these algorithms is provided in the following sections.

Invasive weed optimization. The invasive weed optimization (IWO) algorithm is inspired by weed colonization and was first proposed by Mehrabian and Lucas⁵⁴. In this algorithm, a population of initial solutions (weeds) spreads randomly in the entire search space. The fitness of weeds is evaluated and produces a number of seeds—the population with better fitness produces more seeds. Produced seeds are randomly distributed in search space by normally distributed random numbers with a mean equal to zero, but with a varying variance. The IWO applies the standard deviation (σ) of the random function which is defined between the ranges of the pre-defined initial value ($\sigma_{initial}$) to a final value (σ_{final}) and is calculated in each step from the following equation⁵⁴:

$$\sigma_{iter} = \frac{(iter_{max} - iter)^n}{(iter_{max})^n} (\sigma_{initial} - \sigma_{final}) + \sigma_{final} \quad (15)$$

where σ_{iter} is the standard deviation of the present iteration, $iter_{max}$ is the maximum number of iterations, and n is the nonlinear modulation index usually set as 2⁵⁵.

After some iterations, the number of produced plants reaches a maximum value. At this stage, competitive exclusion eliminates undesirable plants based on the fitness function. Consequently, those with better fitness would survive and are allowed to replicate. This process might be continued either reaching the maximum epoch or achieving the exact solution.

Firefly algorithm. This algorithm was inspired by the flashing and illuminating behavior of the fireflies and was proposed by Yang⁵⁶. This algorithm is based on three general rules, (1) all fireflies in the search space are considered to be unisex so they can be attracted to the others; (2) the attractiveness of a firefly is proportional to its light intensity. Hence, the brighter fireflies attract the less bright ones. In the case of fireflies of similar brightness, their movements are assumed to be random; (3) the brightness of a firefly is determined by the objective function. In FA, the new position of agents (X_i^{t+1}) is calculated according to the following equation:

$$X_i^{t+1} = X_i^t + \beta_0 e^{-\gamma r_{ij}^2} (X_j^t - X_i^t) + \alpha \varepsilon_i \quad (16)$$

where a is a randomized parameter (mutation coefficient), ε_i is the random vector, β_0 is the attractiveness at distance $r=0$, and γ is the absorption coefficient of light. More information about the FA algorithm can be found in the original work of Yang⁵⁶.

Differential evolution. The differential evolution (DE) is a stochastic and population-based algorithm proposed by Storn and Price⁵⁷ applied in global optimization problems. The main structure of DE is based on extracting individual differences from a current population to build a new population. In other words, it is one of the optimization algorithms which has both evolutionary and swarm intelligence features⁵⁸. Mutation, crossover,

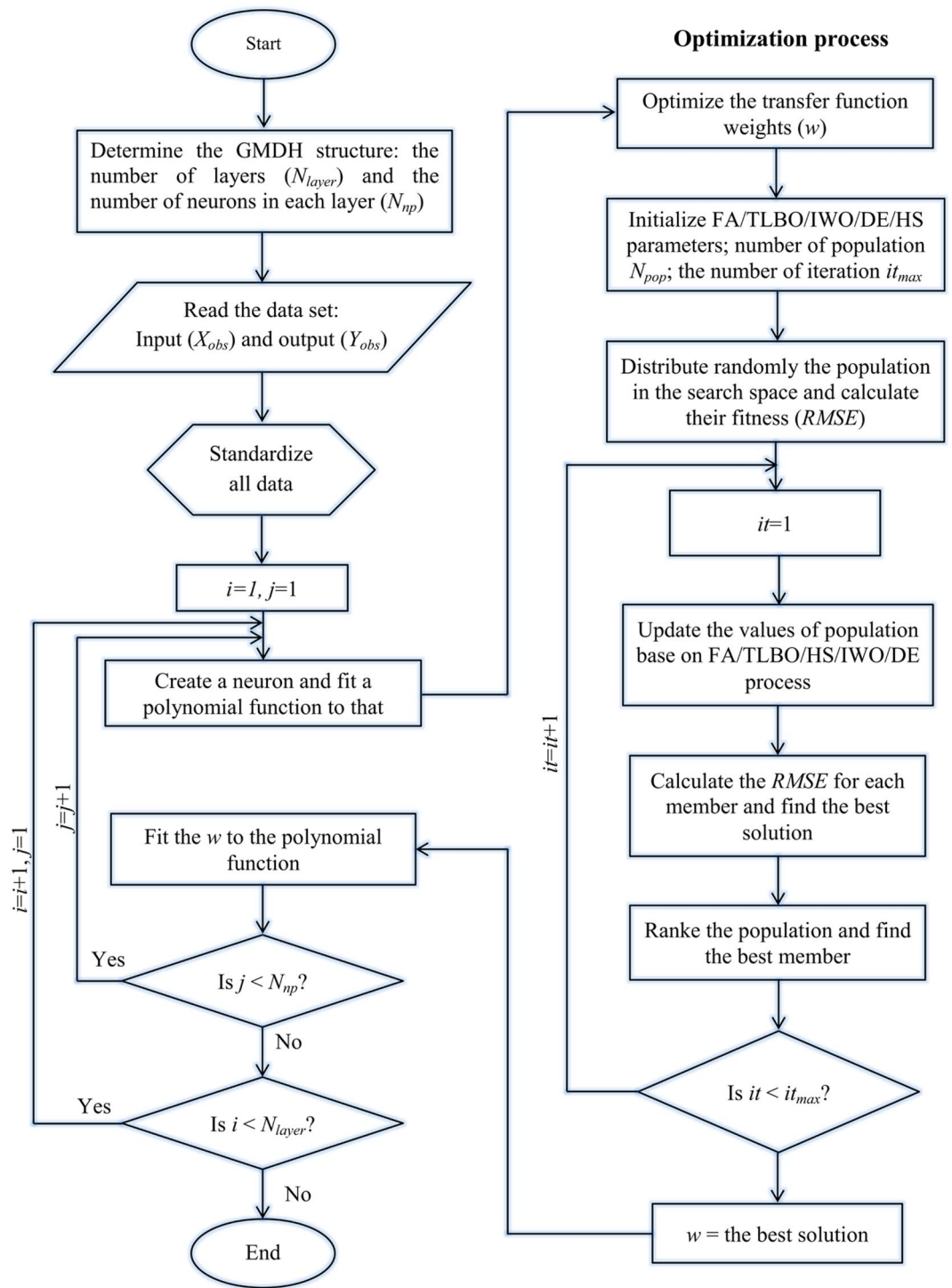


Figure 3. The general flowchart of the constructed hybrid GMDH models.

and selection are evolutionary operators, while distance and direction of the population can be considered as swarm intelligence features.

Assuming three agents of the population as X_{i1} , X_{i2} , and X_{i3} , the mutation scheme can be expressed as a trial vector (v_i^t) which will be developed for each member of the population by applying the following equation:

$$u_i^t = X_{i1}^t + F \times (X_{i2}^t - X_{i3}^t) \tag{17}$$

where t is the iteration index and F is the scale factor that controls the amount of differential variation.

In the crossover scheme, offspring (u_{i1}^t) will be generated as:

Category	Algorithms	Inspired by	Reason for the selection
Swarm intelligence	IWO	Plants and animals	High performance in combination with ANFIS and ANN models ^{45,46}
	FA	Insects movement	A robust and efficient algorithm in optimization problems ^{47,48}
	TLBO	Teaching–learning process	A simple and reliable algorithm that does not need any initial parameters ^{49,50}
Evolutionary	HS	Harmony in music	Successful application of hybrid GMDH-HS in modeling nonlinear problems ^{20,51}
	DE	Biological insights	Popular, simple, effective ⁵² and successful applications of GMDH-DE ⁵³

Table 1. Presentation of the meta-heuristic algorithms.

$$v_{ij}^t = \begin{cases} u_{ij}^t & \text{if } \text{rand}_i^j(0, 1) \in C_r \text{ or } j = j_i^{\text{rand}} \\ X_{ij}^t & \text{otherwise} \end{cases} \tag{18}$$

where C_r is the crossover controller assuming values between 0 to 1 and j_i^{rand} is a random integer number between 1 and the dimension of the problem, D . In the end, the selection operator is applied and the new position is calculated from:

$$X_i^{t+1} = \begin{cases} v_i^t & \text{if } f(v_i^t) \leq f(X_i^t) \\ X_i^t & \text{otherwise} \end{cases} \tag{19}$$

Teaching–learning-based optimization. The teaching–learning-based optimization (TLBO) is inspired by the philosophy of the teaching and learning process and was proposed by Rao et al.⁴⁹ The algorithm is based on the evaluation of the influence of a teacher on the performance of students. The TLBO consists of two main phases: Teaching and Learning Phases. Among all the designated populations in the search space, the best solution (base on fitness) is assigned to the teacher, and the learners would learn and update their knowledge from the teacher according to the teaching operation:

$$x_i^{\text{new}} = x_i^{\text{old}} + \text{rand}_i \times (x_{\text{best}} - T_F \times x_{\text{mean}}) \tag{20}$$

where x_i^{new} is the new positions of the i th learner, x_i^{old} is the old positions of the i th learner, rand_i is a random number between 0 to 1, x_{best} is the position of the teacher, x_{mean} is the mean individual position of the current class, and T_F is the teaching factor that is applied to change the mean value.

After this stage, learners increase their knowledge by interacting randomly with other learners in class as the equation below:

$$x_i^{\text{new}} = \begin{cases} x_i^{\text{old}} = \text{rand}_i \times (x_i - x_j), & \text{if } f(x_i) < f(x_j) \\ x_i^{\text{old}} = \text{rand}_i \times (x_j - x_i), & \text{otherwise} \end{cases} \tag{21}$$

where x_j is a member of the population that is selected randomly.

Harmony search. The harmony search (HS) is a meta-heuristic optimization method inspired by musicians which simulate the improvisation process of the group of musicians⁵⁹. To create the HS structure, some parameters, including the harmony memory consideration rate ($HMCR$), pitch adjusting rate (PAR), fret width (FW), and harmony memory size (HMS) have to be set at the initialization stage. The harmony memory is a matrix of candidate solutions. The $HMCR$ varies between 0 and 1 and controls the balance between exploration and exploitation. The PAR and FW are useful parameters in adjusting the convergence rate of the algorithm. The quantity $X_j = [x_1^j, x_2^j, \dots, x_D^j]$ represents the j th harmony vector. The HM is filled with the HMS harmony vectors:

$$HM = \begin{bmatrix} x_1^1 & x_2^1 & \dots & x_D^1 \\ x_1^2 & x_2^2 & \dots & x_D^2 \\ \dots & \dots & \dots & \dots \\ x_1^{HMS} & x_2^{HMS} & \dots & x_D^{HMS} \end{bmatrix} \tag{22}$$

At the first iteration, HM is filled with random solution vectors. In the next iterations, the HM must be updated with a new solution. If the new solution vector is better than the worst vector, then it is stored and replaces the worst vector.

Experimental data and model setup. *Datasets.* In this study, eight parameters are used as the input vector to build and train the GMDH models for estimating the wave run-up ($R_{u2\%}$). The input vector includes spectral peak period (T_p), mean wave period (T_m), significant wave height (H_s), beach slope ($\cot\alpha$), the relative size of bed material (D_{85}/D_{15}), the surf similarity parameter corresponding to the mean wave period (ξ_m), the

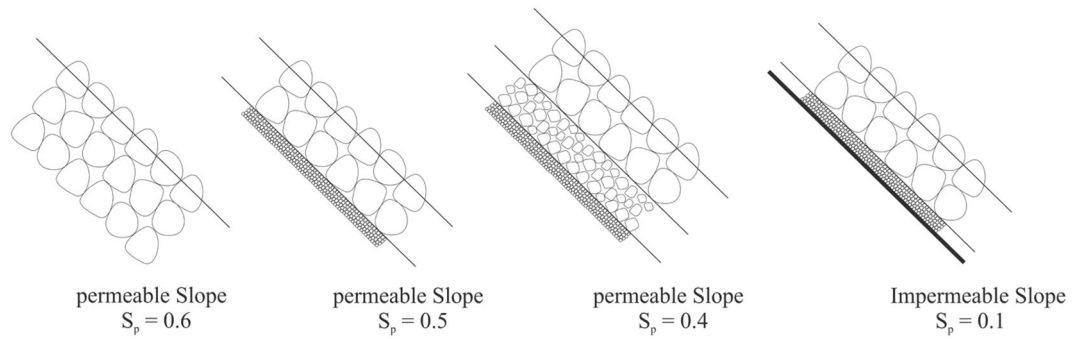


Figure 4. Bed permeability parameter for different types of bed materials.

Parameter	Type	Data range	Average	Standard deviation (SD)	Correlation coefficient ($R_{u2\%}$)
$R_{u2\%}$ (m)	Output	[0.043,1.6]	0.237	0.211	1
T_p (s)	Input	[1.33,5.1]	2.542	0.884	0.351
T_m (s)	Input	[1.24,4.4]	2.181	0.707	0.403
H_s (m)	Input	[0.461,1.18]	0.139	0.138	0.938
$\cot\alpha$	Input	[1.5,4]	2.709	0.783	0.0003
D_{85}/D_{15}	Input	1.25,2.25	1.641	0.485	0.028
S_p	Input	0.1,0.5,0.6	0.276	0.205	0.009
ξ_m	Input	[0.991,7.584]	3.202	1.367	0.0154
ξ_p	Input	[1.047,8.869]	3.740	1.681	0.0129

Table 2. Statistical summary of the data used in this study.

surf similarity parameter corresponding to the peak period (ξ_p), and the bed permeability (S_p). The S_p characterizes bed material especially particle grading and permeability. Figure 4 shows the range of S_p for different types of coastal slopes. The S_p data covers both impermeable and permeable slopes (see Table 2).

The 256 data set was extracted from Van der Meer and Stam⁸ report. All tests were conducted in a wave flume of 1 m wide, 1.2 m deep, and 50 m long. Data were measured for two types of particle grading including uniform rock and riprap, and different values of S_p . In this study, all available parameters that have a direct effect on the wave run-up height were selected as input parameters for the predictive models. Information about the statistical properties of the dataset is presented in Table 2.

Evaluation criteria. In the present study, the statistical assessment of the training data set was conducted for checking the reliability of the developed GMDH models. A similar assessment of the testing data set was conducted for the evaluation of the performance of GMDH models and empirical relations. This was done by applying different statistical parameters including root mean square error (*RMSE*), determination coefficient (R^2), mean absolute error (*MAE*), and the index of agreement (*IA*).

The *RMSE* determines the root mean square error between the observed and predicted values, while the *MAE* describes the distance between the observed and predicted values. The R^2 indicates how well the predicted values fit the regression model. The *IA* is the measure of the agreement between the predicted and observed values. These parameters are calculated as follow:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i^m - x_i^o)^2} \tag{23}$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |x_i^m - x_i^o| \tag{24}$$

$$R^2 = \frac{\left(\sum_{i=1}^N (x_i^m - \bar{x}^m)(x_i^o - \bar{x}^o) \right)^2}{\sum_{i=1}^N (x_i^m - \bar{x}^m)^2 \sum_{i=1}^N (x_i^o - \bar{x}^o)^2} \tag{25}$$

Model	Structure		Meta-heuristic parameters		
	Number of middle layers	Maximum neurons in each layer	Parameter	Value	Epoch
GMDH	5	10	–	–	300
GMDH-IWO	5	10	Maximum no of seeds	10	300
			Minimum no of seeds	1	
			Initial Standard deviation ($\sigma_{initial}$)	0.5	
			Final Standard deviation (σ_{final})	0.001	
			Search Space Range	[–10,10]	
GMDH-FA	5	10	Mutation Coefficient (a)	0.2	300
			Attraction Coefficient (β_0)	2	
			Light Absorption Coefficient (γ)	1	
			Search Space Range	[–10,10]	
GMDH-DE	5	10	Lower bound of scaling factor	0.2	300
			Upper bound of scaling factor	0.8	
			Crossover controller (C_c)	0.9	
			Search space range	[–10,10]	
GMDH-HS	5	10	Fret width damp ratio	0.995	300
			Pitch adjustment rate (PAR)	0.1	
			Harmony memory consideration rate ($HMCR$)	0.9	
			Search space range	[–10,10]	
GMDH-TLBO*	5	10	Search space range	[–10,10]	300

Table 3. Structure of models and initial parameters of meta-heuristic algorithms. *The TLBO algorithm can be applied without allocating any specific primary or adjusting parameter.

Method	Statistics			
	RMSE (m)	R^2	MAE (m)	IA
GMDH-TLBO	0.0250	0.9863	0.0184	0.9965
GMDH	0.0254	0.9853	0.0195	0.9963
GMDH-DE	0.0274	0.9829	0.0209	0.9956
GMDH-FA	0.028	0.9823	0.0188	0.9955
GMDH-IWO	0.0355	0.9759	0.0198	0.9932
GMDH-HS	0.0359	0.9730	0.0219	0.9928

Table 4. Statistical parameters for the evaluation of the performance of derived models at a training stage.

$$IA = \frac{\sum_{i=1}^N (x_i^m - x_i^o)^2}{\sum_{i=1}^N (|x_i^m - \bar{x}^o| + |x_i^o - \bar{x}^o|)^2} \quad (26)$$

where N is the number of data, x_i^m denotes the predicted value, x_i^o is the observed value, and bar stands for the average of the variables. The best fitting between models and observed data is for $RMSE$ and MAE equal to 0, and R^2 and IA equal to 1.

Models setup and structure. To begin with the model development process and implementation of data, the data set was divided randomly into a training set comprising 80% of the available data set and a testing set covering 20% of the data set. The general structure of the GMDH models is presented in Table 3. The allocated values of parameters are based on their values from the previous studies^{60–64}. For all the developed models, 300 epochs were considered. It is worth noting that this iteration number was reached based on the convergence criteria of GMDH models.

Results

In this study, the GMDH model and the five hybrid GMDH models described in the previous section were used to estimate the wave run-up. Results obtained for the training stage are presented in Table 4.

The results in Table 4 show that the applied GMDH models are capable of modeling wave run-up, which confirm low values of $RMSE$ and MAE , and the close to one values of R^2 and IA . Among all methods developed in this study GMDH-TLBO, with the lowest $RMSE$ and MAE ($RMSE = 0.0254$ m and $MAE = 0.0195$ m) and the highest R^2 and IA ($R^2 = 0.9863$ and $IA = 0.9965$) provide the best results at the training stage. However, the results

Method	Statistics				Enhancement (+) or deterioration (-) of the RMSE in performance (%)	Establishment	Execution time
	RMSE (m)	R ²	MAE (m)	IA			
GMDH-FA	0.0209	0.9908	0.0172	0.9977	+47.49	Difficult	High
GMDH-IWO	0.0220	0.9908	0.0170	0.9975	+44.74	Difficult	High
GMDH-TLBO	0.0235	0.9888	0.0180	0.997	+41.08	Difficult	High
GMDH-HS	0.0262	0.9864	0.0211	0.9964	+34.26	Difficult	High
GMDH-DE	0.0302	0.984	0.0240	0.9953	+24.12	Difficult	Medium
GMDH	0.0398	0.9674	0.028	0.9916	Base	Not Simple	Low
Schimmels et al. ⁵	0.0763	0.9549	0.0638	0.9711	-91.52	Simple	Low
Van der Meer and Stam ⁸	0.1224	0.8178	0.0747	0.927	-207.22	Simple	Low

Table 5. Statistical parameters for the evaluation of the performance of derived models at a testing stage.

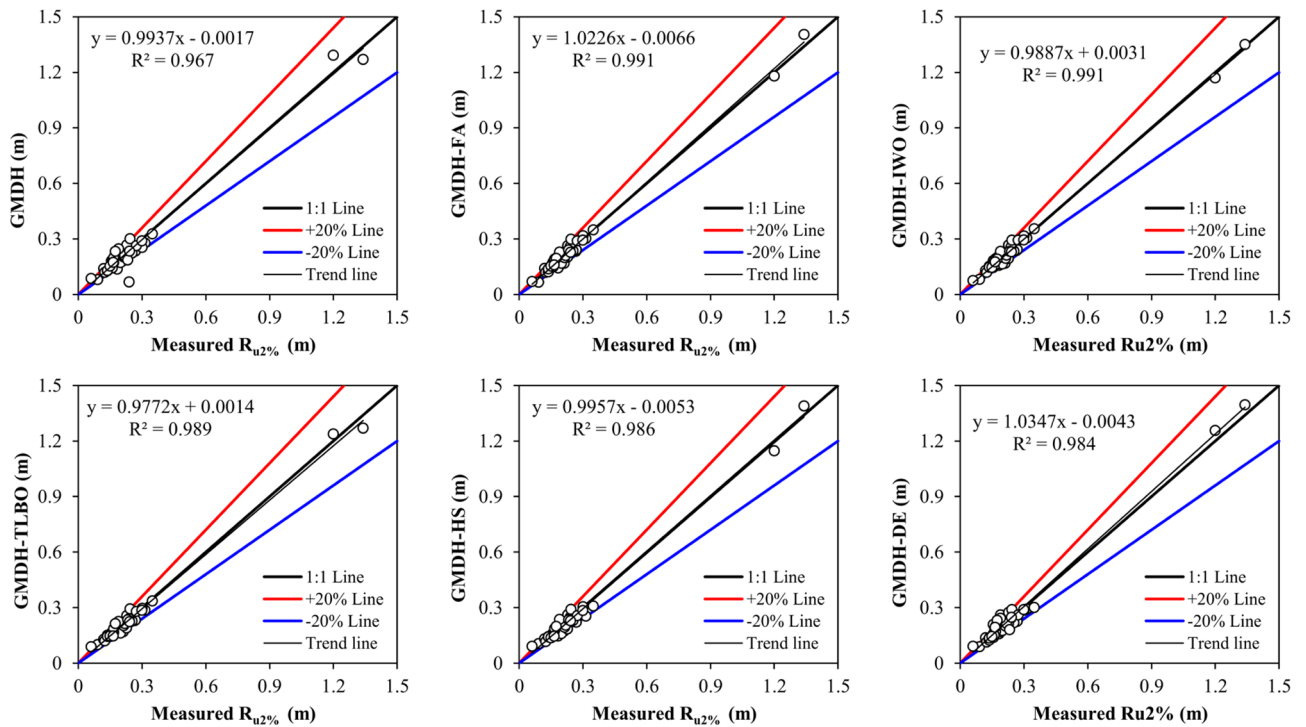


Figure 5. Scatter plots obtained for the testing stage.

show that HS, FA, DE, and IWO, decrease the performance of the standard GMDH. Nevertheless, the decision on the superiority of the derived models can be conducted by the evaluation of the results at the testing stage. In this respect, Table 5 summarizes the statistical parameters derived for the evaluation of the performance of developed models at a testing stage.

The results in Table 5 show that the standard and hybrid GMDH models provide more accurate results than the empirical relations. However, the development of the derived models is a time consuming process and their execution requires more time in comparison to the empirical models. The average values of RMSE, MAE, R², and IA for the GMDH models are 0.027 m, 0.0209 m, 0.9847, and 0.9959, respectively, whereas for empirical relations are 0.0993 m, 0.0692 m, 0.886, and 0.9490, respectively. The developed GMDH models improve the prediction of wave run-up by about 72.81% in comparison to the empirical methods. Among the developed GMDH models, the GMDH-FA with the lowest values of RMSE and MAE (RMSE=0.0209, MAE=0.0172) and the highest R² and IA (R²=0.9908, IA=0.9977) can be considered as the most precise predictive model. The application of meta-heuristic optimization algorithms improves the performance of the standard GMDH model by 39.70%. The performance of most hybrid GMDH models is lower at the training stage than the standard GMDH. The results confirm a relatively good performance of the derived models at the testing stage. The standard GMDH model may trap into the local solutions, which causes the over-fitting problem. Figure 5 presents the scatter plots for the testing stage. Scatter points of the hybrid GMDH models are closer to 1:1 line than corresponding points obtained by applying the standard GMDH technique, which indicates that the meta-heuristic algorithms increase the performance of GMDH.

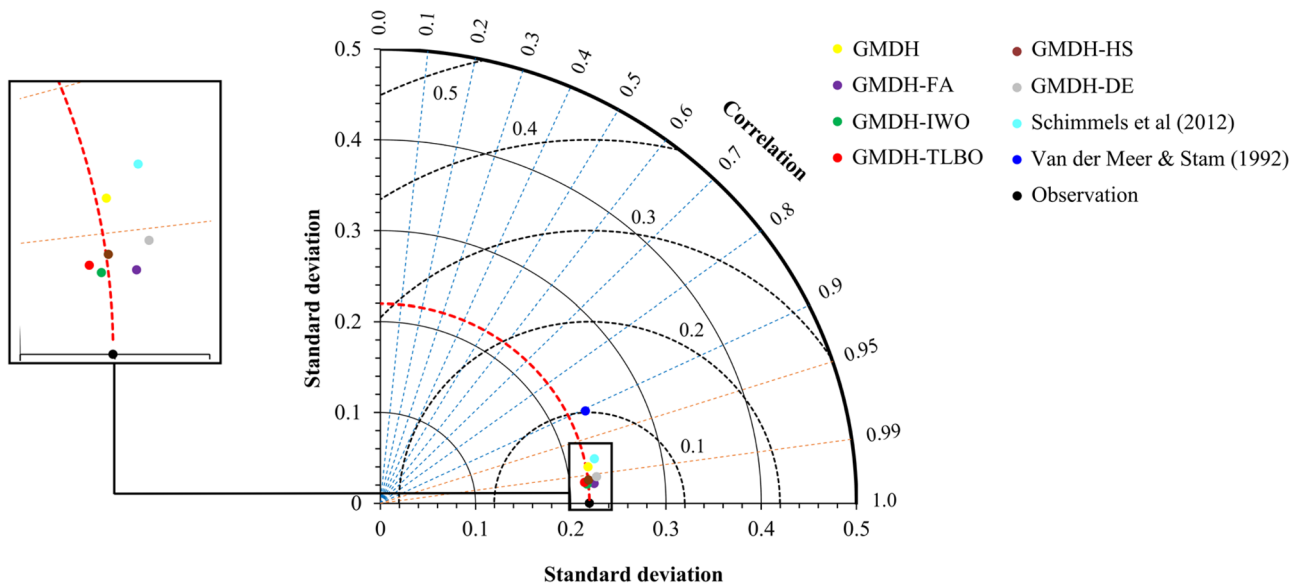


Figure 6. Taylor diagram for the tested data set.

Model	Empirical relations and standard GMDH	GMDH-IWO	GMDH-FA	GMDH-DE	GMDH-HS	GMDH-TLBO
CPU* time (min)	<1	116	405	11	43	116

Table 6. The CPU time for 300 epochs of the applied GMDH models. *Core i7; RAM: 8 GB.

Methods	p-value	Significantly different (95%)	Significantly different (99%)
Schimmels et al. ⁵ , Van der Meer and Stam ⁸	0.7251	NO	NO
GMDH, GMDH-FA, GMDH-DE, GMDH-IWO, GMDH-TLBO, GMDH-HS, Schimmels et al. ⁵ , Van der Meer and Stam ⁸	<.0006	YES	YES
GMDH, GMDH-FA, GMDH-DE, GMDH-IWO, GMDH-TLBO, GMDH-HS	0.994	NO	NO

Table 7. The results of the Kruskal–Wallis test for assessing the significant statistical differences between the applied models.

To provide further insight into the outcomes of the derived models, the results of the models are plotted in the Taylor diagram⁶⁵. The purpose of the Taylor diagram is to present on a single plot three statistical indices including the standard deviation, centered *RMSE*, and correlation. In Fig. 6, the values on the vertical and horizontal axes represent the standard deviation, the values on the dashed lines represent the correlation, and the values on the dash-curved lines represent the centered *RMSE*. Figure 6 shows that the dots representing the GMDH models are closer to the observation points than corresponding dots obtained by the application of the empirical relations. The results of the standard and hybrid GMDH models are magnified in a separate box. The plots show the superiority of the GMDH-FA over the remaining GMDH models.

Further discussion. An important factor in the analysis of the performance of machine learning methods and meta-heuristic techniques is the speed of the derived algorithms. Based on the observed CPU time presented in Table 6, it is becoming clear that the standard GMDH model is more efficient in terms of computational cost than the derived hybrid GMDH models. It is worth mentioning that the application of empirical equations in practical projects is easier than the ML models. However, nowadays, with tremendous progress that has been made in technology and minicomputers, the applications of ML- trained models to determine wave run-up, is straightforward.

In the previous section, the analyses of the performance of the GMDH models versus the empirical relations showed the superiority of the developed GMDH models. But in spite of that, the question may arise whether the differences between the results obtained by applying the developed models are statistically significant. Thus, the Kruskal–Wallis tests were carried out and the results are presented in Table 7. The non-parametric Kruskal–Wallis test is a technique often applied in statistical analyses. Mahdavi-Meymand et al.¹⁸ used the Kruskal–Wallis test to compare several machine learning (ML) techniques and empirical equations applied to predict spillways

Authors	Methods	Type of the machine learning model	Calculated R^2	Remarks
Abolfathi et al. ¹²	M5' Decision tree	Decision tree	0.970	The results show the general ability of M5' to simulate wave run-up
Bakhtyar et al. ⁶⁶	ANFIS	Hybrid intelligent systems	0.960	The comparison of results confirms the high accuracy of ANFIS in predicting wave run-up
	Empirical formulas	–	0.890	
Erdik and Savci ⁷	ANFIS	Hybrid intelligent systems	0.621	TS Fuzzy is a capable tool for modeling wave run-up
	Empirical formulas	–	0.559	
Bonakdar and Etemad-Shahidi ¹⁰	M5 model tree	Decision tree	0.920	M5 results are better than TS Fuzzy and empirical formulas
	TS Fuzzy	Hybrid intelligent systems	–	
	Empirical formulas	–	0.902	
Elbisy ¹¹	MART*	Decision tree	0.974	The MART model is more accurate than the ANN
	ANN	Neural computing	0.837	
Rehman et al. ¹⁸	ANN	Neural computing	0.995	Both ANN and RSM are robust methods for the prediction of wave run-up
	Response surface methodology (RSM)	–	0.999	
Yao et al. ¹⁹	ANN	Neural computing	0.987	The ANN performance in predicting wave run-up was confirmed
The current study	GMDH	Neural computing	0.985	The results show that the GMDH models provide more accurate results than empirical relations. Applications of optimization algorithms increases the accuracy of standard models. The Kruskal–Wallis test shows that there are no significant statistical differences between the classical and hybrid GMDH models
	GMDH-TLBO	Hybrid intelligent systems	0.986	
	GMDH-FA		0.991	
	GMDH-DE		0.984	
	GMDH-HS		0.973	
	GMDH-IWO		0.976	
	Van der Meer and Stam ⁸ Schimmels et al. ⁵	–	0.818 0.955	

Table 8. Summary of the studies conducted to predict wave run-up in coastal regions. *Multiple additive regression trees.

air demand and reported that there is no significant difference at the 99% confidence level between the applied ML approaches.

The results in Table 7 show that the probability value of the Kruskal–Wallis test for the empirical relations is 0.7251, which is higher than 0.05 and 0.01. This shows that there are no significant statistical differences between the results obtained by applying two empirical relations at both 95% and 99% confidence levels. However, significant differences exist between the results obtained by applying the GMDH models and empirical relations (p -value < 0.0006). Thus, the GMDH models may be recommended to be applied instead of empirical relations to predict wave run-up. Moreover, although the statistical criteria confirmed that meta-heuristic algorithms increase the efficiency of the GMDH, the results in Table 7 show that there is no significant statistical difference between the outcome of different GMDH models. Hence, it is recommended to use the standard GMDH in situations where computational time is an important factor for users. More insight into the results obtained in the modelling of wave run-up in coastal regions is provided in Table 8.

Based on the findings of this study following potential subjects are proposed for future studies:

Due to literature restrictions, it is recommended to take into consideration other possible future data resources to further evaluate predictive models developed in the present study.

There is not a robust formula or procedure to select the best architecture of the GMDH model. More studies should be conducted to facilitate the construction of GMDH models.

Based on the data used in this study and related calculated values of the correlation coefficient between the input and output variables, it was found that the beach slope has a limited effect on wave run-up height. Future laboratory studies should take into account a wider range of beach slopes.

It is recommended to consider other non-parametric statistical tests such as Mann–Whitney test and evaluate the results of different methods.

Conclusion

The ability to accurately estimate the maximum wave run-up is vital for the maintenance and development of coastal areas and the safety of the coastal zone population. In this study, hybrid swarm and evolutionary intelligent GMDH models as well as the standard GMDH technique were developed and applied to predict the two percent value of the wave run-up ($R_{2\%}$). The invasive weed optimization (IWO), firefly algorithm (FA), differential evolution (DE), teaching–learning-based optimization (TLBO), and harmony search (HS) optimization algorithms were used as the meta-heuristic optimization methods to train the GMDH. The results show that the GMDH models have better performance than the empirical relations. The Kruskal–Wallis tests show significant statistical differences between the results of the empirical relations and GMDH models.

The results show that the application of meta-heuristic optimization algorithms improves the performance of the standard GMDH model by 39.70% at the testing stage. However, the performance of most hybrid GMDH models is lower than the standard GMDH at the training stage. Among all developed models, the GMDH-FA

provides the best results in the testing stage with $RMSE=0.0209$ m and $IA=0.9977$. Moreover, the results show that among all methods developed in this study GMDH-TLBO, with the lowest $RMSE=0.0254$ m and $MAE=0.0195$ m, and the highest $R^2=0.9863$ and $IA=0.9965$.

The computational costs indicate that the standard GMDH model and empirical equations are significantly faster than the embedded meta-heuristic GMDH techniques. Moreover, the results show that there are no significant statistical differences between the GMDH models and meta-heuristic algorithms. Hence, the application of time-consuming models such as GMDH-FA is not recommended in situations where computational cost is a decisive factor.

Received: 29 October 2021; Accepted: 12 April 2022

Published online: 18 May 2022

References

- Ruggiero, P., Komar, P. D., Marra, J. J., McDougal, W. G. & Beach, R. A. Wave runup, extreme water levels and the erosion of properties backing beaches. *J. Coastal Res.* **17**, 407–419 (2001).
- Puleo, J. A., Beach, R. A., Holman, R. A. & Allen, J. S. Swash zone sediment suspension and transport and the importance of boregenerated turbulence. *J. Geophys. Res.* **105**(C7), 17021–17044 (2000).
- Van der Meer, J. W. Rock slopes and gravel beaches under wave attack. PhD thesis, Technical University of Delft, Netherlands (1988).
- Muttray, M., Oumeraci, H. & Ten Over, E. Wave reflection and wave run-up at rubble mound breakwaters. In *Proceedings of 30th International Conference on Coastal Engineering, San Diego, California, USA* (2006).
- Schimmels, S. *et al.* Wave run-up observations on revetments with different porosities. *Coast. Eng. Proc.* **33**, 1–14 (2012).
- Chang, H. K. & Lin, L. C. Multi-point tidal prediction using artificial neural network with tide-generating forces. *Coast. Eng.* **53**, 857–864 (2006).
- Erdik, T. & Savci, M. E. Takagi-sugeno fuzzy approach in rock armored slopes for 2% wave runup estimation. *Coast. Eng. J.* **5**, 161–177 (2008).
- Van der Meer, J. W. & Stam, J. C. Wave runup on smooth and rock slopes of coastal structures. *J. Waterw. Port Coast. Ocean Eng.* **118**, 534–550 (1992).
- Shiri, J., Makarynskyy, O., Kisi, O., Dierickx, W. & Fakheri Fard, A. Prediction of short-term operational water levels using an adaptive neuro-fuzzy inference system. *J. Waterw. Port Coast. Ocean Eng.* **137**, 344–354 (2011).
- Bonakdar, L. & Etemad-Shahidi, A. Predicting wave run-up on rubble-mound structures using M5 model tree. *Ocean Eng.* **38**, 111–118 (2011).
- Elbisy, M. S. Estimation of regular wave run-up on slopes of perforated coastal structures constructed on sloping beaches. *Ocean Eng.* **109**, 60–71 (2015).
- Abolfathi, S., Yeganeh-Bakhtiary, A., Hamze-Ziabari, S. M. & Borzooei, S. Wave run up prediction using M5 model tree algorithm. *Ocean Eng.* **112**, 76–81 (2016).
- Pourzangbar, A. *et al.* Prediction of non-breaking wave induced scour depth at the trunk section of breakwaters using Genetic Programming and Artificial Neural Networks. *Coast. Eng.* **121**, 107–118 (2017).
- Zhang, Z. G., Yin, J. C. & Liu, C. A modular real-time tidal prediction model based on grey-GMDH neural network. *Appl. Artif. Intell.* **32**, 2 (2018).
- Wang, B., Wang, B., Wu, W., Xi, C. & Wang, J. Sea-water-level prediction via combined wavelet decomposition, neuro-fuzzy and neural networks using SLA and wind information. *Acta Oceanol. Sin.* **39**, 157–167 (2020).
- Gao, X., Pan, X., Liu, X., Pedrycz, W. & Wang, Z. Modeling of the ship steady turning motion based on multiblocks of fuzzy cognitive maps. *Appl. Ocean Res.* **110**, 102604 (2021).
- Zeinali, S., Dehghani, M. & Talebbeydokhti, N. Artificial neural network for the prediction of shoreline changes in Narrabeen, Australia. *Appl. Ocean Res.* **107**, 102362 (2021).
- Rehman, K., Khan, H., Cho, Y. S. & Hong, S. H. Incident wave run-up prediction using the response surface methodology and neural networks. *Stoch. Environ. Res. Risk Assess.* **36**, 17–32 (2022).
- Yao, Y., Yang, X., Lai, S. H. & Chin, R. J. Predicting tsunami-like solitary wave run-up over fringing reefs using the multi-layer perceptron neural network. *Nat. Hazards* **107**, 601–616 (2021).
- Masoumi Shahr-Babak, M., Khanjani, M. J. & Qaderi, K. Uplift capacity prediction of suction caisson in clay using a hybrid intelligence method (GMDH-HS). *Appl. Ocean Res.* **59**, 408–416 (2016).
- Najafzadeh, M., Saberi Movahed, F. & Sarkamaryan, S. NF-GMDH based self-organized systems to predict bridge pier scour depth under debris flow effects. *Mar. Georesour. Geotechnol.* **36**, 5 (2017).
- Mahdavi-Meymand, A., Scholz, M. & Zounemat-Kermani, M. Challenging soft computing optimization approaches in modeling complex hydraulic phenomenon of aeration process. *ISH J. Hydraul. Eng.* **27**, 1–13 (2019).
- Dodangeh, E. *et al.* Novel hybrid intelligence models for flood-susceptibility prediction: meta optimization of the GMDH and SVR models with the genetic algorithm and harmony search. *J. Hydrol.* **590**, 125423 (2020).
- Sharafati, A., Haghbin, M., Torabi, M. & Yaseen, Z. M. Assessment of novel nature-inspired fuzzy models for predicting long contraction scouring and related uncertainties. *Front. Struct. Civil Eng.* **15**, 665–681 (2021).
- Sharafati, A. *et al.* Scouring depth assessment downstream of weirs using hybrid intelligence models. *Appl. Sci.* **10**(11), 3714 (2020).
- Qaderi, K., Javadi, F., Madadi, M. R. & Ahmadi, M. M. A comparative study of solo and hybrid data driven models for predicting bridge pier scour depth. *Mar. Georesour. Geotechnol.* (2020).
- Alizadeh, S. M., Alrueyemi, I., Daneshfar, R., Mohammadi-Khanaposhtani, M. & Naseri, M. An insight into the estimation of drilling fluid density at HPHT condition using PSO-, ICA-, and GA-LSSVM strategies. *Sci. Rep.* **11**, 7033 (2021).
- Milan, S. G., Roozbahani, A., Arya Azar, N. & Javadi, S. Development of adaptive neuro fuzzy inference system: evolutionary algorithms hybrid models (ANFIS-EA) for prediction of optimal groundwater exploitation. *J. Hydrol.* **598**, 126258 (2021).
- Haghbin, M., Sharafati, A. & Motta, D. Prediction of channel sinuosity in perennial rivers using Bayesian Mutual Information theory and support vector regression coupled with meta-heuristic algorithms. *Earth Sci. Inform.* **14**, 2279–2292 (2021).
- Coastal Flood Hazard Analysis and Mapping Guidelines Focused Study Report, FEMA. Wave Run-up and Overtopping (2005).
- Miche, R. Mouvements ondulatoires de la mer en profondeur constante ou décroissante: forme limite de la houle lors de son déferlement. Application aux digues maritimes, Paris (1944).
- Grantham, K.N. A model study of wave run-up on sloping structures. Technical Report, Series 3, Issue 348, Institute of Engineering Research in the University of California, Berkeley California, USA (1953).
- Saville, T. Laboratory data on wave run-up and overtopping on shore structure. Technical Memorandum, 64, Beach Erosion Board, US Army Corps of Engineers, Washington, DC, USA (1955).
- Saville, T. Wave run-up on shore structures. *J. Waterw. Div.* **952**, 1–14 (1956).

35. Saville, T. Wave run-up on composite slopes. In: *Proceedings of the 6th International Conference on Coastal Engineering*, ASCE 691–699 (1958).
36. Savage, R. P. Wave run-up in roughened and permeable slopes. *J. Waterw. Harbors Div.* **99**, 1–38 (1958).
37. Hunt, I. A. Design of seawalls and breakwaters. *J. Waterw. Harbors Div.* **85**, 123–152 (1959).
38. Iribarren, C. R. Nogales, C. Protection des ports. In: *Proceedings of the XVIIth International Navigation Congress, Section II, Communication*, 31–80 (1949).
39. Wassing, F. Model investigations of wave run-up on dikes carried out in the Netherlands in the last 20 years. In: *Proceedings of the 6th Coastal Engineering Conference, Florida, USA* (1957).
40. Ahrens, J. P. Irregular wave runup on smooth slopes. Q CETA No. 81–17, US Army Corps of Engineers, Coastal Engineering Research Center, Ft. Belvoir, VA (1981).
41. US Army Corps of Engineers. Coastal Engineering Manual, Engineer Manual 1110–2 1100, Washington (2002).
42. Mase, H. Random wave runup height on gentle slope. *J. Waterw. Port Coast. Ocean Eng.* **115**, 649–661 (1989).
43. Ivankhnenko, A. G. Polynomial theory of complex systems. *IEEE Trans. Syst. Man Cybern. SMC* **1**, 364–378 (1971).
44. Sohani, A., Sayyaadi, H. & Hoseinpoori, S. Modeling and multi-objective optimization of an M-cycle cross-flow indirect evaporative cooler using the GMDH type neural network. *Int. J. Refrig.* **69**, 186–204 (2016).
45. Tien Bui, D. et al. New hybrids of ANFIS with several optimization algorithms for flood susceptibility modeling. *Water* **10**, 1–28 (2018).
46. Bazdar, H., Fattah, H. & Ghadimi, F. Hybrid ANN with invasive weed optimization algorithm a new technique for prediction of gold and silver in Zarshuran Gold Deposit, Iran. *J. Tethys* 273–285 (2015).
47. Baykasoglu, A. & Ozsoydan, F. B. Adaptive firefly algorithm with chaos for mechanical design optimization problems. *Appl. Soft Comput.* **36**, 152–164 (2015).
48. Peng, H., Zhu, W., Deng, C. & Wu, Z. Enhancing firefly algorithm with courtship learning. *Inform. Sci.* **543**, 18–42 (2021).
49. Rao, R. V., Savsani, V. J. & Vakharia, D. P. Teaching–learning–based optimization: a novel method for constrained mechanical design optimization problems. *Comput. Aided Des.* **43**, 303–315 (2011).
50. Rao, R. V. *Teaching learning based optimization algorithm and its engineering applications* (Springer, 2015).
51. Qaderi, K., JalaliEmail, K., Etminan, S., Masoumi Shahrbabak, M. & Homaei, M. Estimating soil hydraulic conductivity using different data-driven models of ANN, GMDH and GMDH-HS. *Paddy Water Environ.* **16**, 823–833 (2018).
52. Awad, N. H., Ali, M. Z., Mallipeddi, R. & Suganthan, P. N. An efficient differential evolution algorithm for stochastic OPF based active–reactive power dispatch problem considering renewable generators. *Appl. Soft Comput.* **76**, 445–458 (2019).
53. Rayegani, F. & Onwubolu, G. C. Fused deposition modelling (FDM) process parameter prediction and optimization using group method for data handling (GMDH) and differential evolution (DE). *Int. J. Adv. Manuf. Technol.* **73**, 509–519 (2014).
54. Mehrabian, A. R. & Lucas, C. A novel numerical optimization algorithm inspired from weed colonization. *Ecol. Inform.* **1**, 355–366 (2006).
55. Ramu Naidu, Y. & Ojha, A. K. A space transformational invasive weed optimization for solving fixed-point problems. *Appl. Intell.* **48**(4), 942–952 (2018).
56. Yang, X. Firefly algorithms for multimodal optimization. In: *Proceedings of the 5th Symposium on Stochastic Algorithms. Foundations and Applications* 5792, 169–178 (2009).
57. Storn, R. & Price, K. Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. *J. Glob. Optim.* **11**, 341–359 (1997).
58. Jadon, S. S., Tiwari, R., Sharma, H. & Bansal, H. C. Hybrid artificial bee colony algorithm with differential evolution. *Appl. Soft Comput.* **58**, 11–24 (2017).
59. Geem, Z. W., Kim, J. H. & Loganathan, G. A new heuristic optimization algorithm: harmony search. *SIMULATION* **76**, 60–68 (2001).
60. Naidu, Y. R. & Ojha, A. K. A space transformational invasive weed optimization for solving fixed-point problems. *Appl. Intell.* **48**, 942–952 (2018).
61. Zhou, Y., Chen, H. & Zhou, G. Invasive weed optimization algorithm for optimization no-idle flow shop scheduling problem. *Neurocomputing* **137**, 285–292 (2014).
62. Shabani, M., Mirroshandel, S. A. & Asheri, H. Selective refining harmony search: a new optimization algorithm. *Exp. Syst. Appl.* **81**, 423–443 (2017).
63. Ugolotti, R., Nashed, Y. S. G., Mesejo, P., Ivekovic, S. & Mussi, L. Particle Swarm optimization and differential evolution for model-based object detection. *Appl. Soft Comput.* **13**, 3092–3105 (2013).
64. Zounemat-Kermani, M. & Mahdavi-Meymand, A. Hybrid meta-heuristics artificial intelligence models in simulating discharge passing the piano key weirs. *J. Hydrol.* **569**, 12–21 (2019).
65. Taylor, K. E. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.* **106**, 7183–7192 (2001).
66. Bakhtyar, R., Yeganeh Bakhtiary, A. & Ghaheri, A. Application of neuro-fuzzy approach in prediction of runup in swash zone. *Appl. Ocean Res.* **30**(1), 17–27 (2008).
67. Mahdavi-Meymand, A., Zounemat-Kermani, M. & Qaderi, K. Prediction of hydro-suction dredging depth using data-driven methods. *Front. Struct. Civ. Eng.* **15**(3), 652–664 (2021).

Acknowledgements

This study was partially developed under the framework of CEMIE-Océano. Project FSE-2014-06-249795 financed by CONACYT-SENER- Sustentabilidad Energética.

Author contributions

Conceptualization: A.M., M.Z., W.S., R.S.; Methodology: A.M.; Analysis and interpretation of results A.M., M.Z., W.S., R.S.; Writing—draft preparation: A.M.; Writing—review and editing: M.Z., W.S., R.S.; Supervision: M.Z., W.S.; Project administration: W.S.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to W.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022