Ocean Engineering 279 (2023) 114465

Contents lists available at ScienceDirect

Ocean Engineering

journal homepage: www.elsevier.com/locate/oceaneng

Genetic programming for the prediction of berm breakwaters recession

Alireza Sadat Hosseini [a], Amir Kabiri [b], Amir H. Gandomi [c,d,*], Mehdi Shafieefar [a]

[a] Faculty of Civil and Environmental Engineering, Tarbiat Modares University, Tehran, Iran
[b] School of Civil Engineering, College of Engineering, University of Tehran, Tehran, Iran
[c] Faculty of Engineering & IT, University of Technology Sydney, Sydney, Australia
[d] University Research and Innovation Center (EKIK), Óbuda University, 1034, Budapest, Hungary

ARTICLE INFO

Handling Editor: Prof. A.I. Incecik

Keywords:
Berm Breakwater
Recession (Rec)
Rebuild experiments
Cumulative experiments
Multi-objective genetic programming (MOGP)

ABSTRACT

The response of berm breakwaters to wave forces has been examined with rebuild and cumulative experiments. In rebuild experiments, the breakwaters were reconstructed after each test, whereas in cumulative experiments the structural damages were examined at the end of the experiment. This study presents a new method to investigate the berm breakwaters recession considering datasets collected of both types of experiments. Cumulative experimental results were converted to their equivalent rebuild experimental results by modifying the number of waves for the reported damage. After homogenizing the data, the datasets were divided into the train, validation, and test subsets. The data were analyzed using the Multi-Objective Genetic Programming (MOGP) approach, and a prediction model was created to evaluate the berm breakwater recession. The results obtained from the MOGP model were compared to outcomes computed using implicit formulas available in the literature showing that the MOGP model is accurate (R2 = 0.911 and RMSE = 0.111) with a relatively broader applicability range. The impact of each input parameter on the berm breakwater recession was examined using parametric and sensitivity analyses. The stability number was the most important parameter impacting the damage on the coastal structure. The results are in line with findings reported in previous studies.

1. Introduction

Berm breakwaters were first introduced four decades ago mainly to minimize the armor-stone quarry and construction equipment. Berm breakwaters have a bulkier cross-section compared to conventional and concrete block breakwaters. However, the possibility of using smaller rock materials for construction mostly outweighs the rock volume regarding the material supplies and construction costs (Juhl and Jensen, 1995). An illustration of the berm breakwater recession and the involved parameters are presented in Fig. 1.

The stability number, $H_0 = H_s/\Delta D_{n50}$, is the main contributing parameter in the design of berm breakwaters, which was first proposed by Van der Meer and Pilarczyk (van der Meer and Pilarczyk, 1984). The wave period was then added to the stability number to generalize its relation to wave parameters (van der Meer, 1988), (Moghim et al., 2011). Although it was shown that using the wave period parameter, $T_0 = T\cdot(g/D_{n50})^{0.5}$, can increase the scattering of the recession data (van der Meer and Sigurdarson, 2017), researchers (Moghim et al., 2011), (Tørum and Krogh, 2000)– (Shekari and Shafieefar, 2013) have incorporated this parameter in their post-analyses of the experimental studies and derived reliable formulations. Later (Moghim et al., 2011), It is shown that the order of wave height and wave period effectiveness on the reshaping of berm breakwaters is not the same (Moghim et al., 2011). Therefore, a modified stability number, $H_0\sqrt{T_0}$, is introduced.

Some variation of views, which have been put forward based on the experiments, stems from the difference in the experiment methods which lead to substantial differences in the data sources. Berm breakwater experiments follow two major approaches: cumulative, in which cumulative damages are recorded in the experiments, and rebuild, wherein, breakwater sections are reconstructed after each test. Since this difference in the test method and their results significantly affects the data analysis, a general dataset consisting of these two sources must be homogenized before analysis can be carried out. This is one of the main parts of this paper and will be discussed in more detail.

Thus far, researchers primarily aimed to establish experimental formulae for estimating the stability of berm breakwaters (Moghim et al., 2011), (Lykke Andersen and Burcharth, 2010), (Sigurdarson et al., 2009a; Sigurdarson et al., 2009b; Sigurdarson and van der Meer, 2013; Sigurdarson and van der Meer, 2014; Moghim and Lykke Andersen,

* Corresponding author. Faculty of Engineering & IT, University of Technology Sydney, Sydney, Australia.
E-mail addresses: a.sadat@modares.ac.ir (A. Sadat Hosseini), a.kabiri@ut.ac.ir (A. Kabiri), gandomi@uts.edu.au (A.H. Gandomi), shafiee@modares.ac.ir (M. Shafieefar).

https://doi.org/10.1016/j.oceaneng.2023.114465
Received 18 November 2022; Received in revised form 26 February 2023; Accepted 5 April 2023
Available online 20 April 2023

Nomenclature

Symbol Description Unit

H_s	Significant wave height m
Δ	Relative mass density
D_{n50}	Nominal rock diameter m
g	Gravitational acceleration = 9.81 m/s^2
T_m	Mean wave periods
H_0	Stability number (also known as N_s) = $H_s/\Delta D_{n50}$
T_{0m}	Wave period parameter = $T_m \cdot (g/D_{n50})^{0.5}$
H_0T_0	Dynamic stability parameter
$H_0\sqrt{T_0}$	Modified dynamic stability parameter
h	Water depth at the toe of the structure m
h_b	Level of the berm in a berm breakwater below SWL, Negative values: submerged berm, Positive values: emerged berm m
N	Number of waves in a test = Test Duration/T_m
B	Berm width of berm breakwater m
f_g	Rock gradation factor = D_{85}/D_{15}
G_c	Width of rock armored crest m
D_{85}	85 percent value of the sieve curve m
D_{15}	15 percent value of the sieve curve m
α_d	Berm breakwater lower slope angle (°)
R_c	Crest freeboard, level of crest relative to still water level m
Rec	Berm breakwater recession m
k	The regression line slope for predicted versus observed data
k'	The regression line slope for observed versus predicted data
D_i	Formula prediction domain when average values were substituted for all inputs except for the ith one var
S_i	The sensitivity of the formula to its ith input parameter %

2015). These research studies have led to a deeper understanding of the types of berm breakwaters (van der Meer and Sigurdarson, 2017). More recently, Lykke Andersen et al. (2014) modified an older formulation (Lykke Andersen and Burcharth, 2010) with a large application area on berm breakwaters using the modified dynamic stability number. More recently, Ehsani et al. (2020) performed an experimental study on Icelandic-type berm breakwaters and damage, presenting formulae for the eroded area of multi-layered berm breakwaters. Shafieefar et al. (2020) proposed the idea of using secondary toe berm to reduce bottom settlements and increase the geotechnical stability of berm breakwaters constructed in water depths less than 20 m. A considerable influence on recession reduction was observed when increasing the width and thickness of the toe berm.

Machine learning (ML) approaches and soft computing tools, such as Artificial Neural Network (ANN), Support Vector Machine (SVM), Model Trees (MT), Fuzzy Logic (FL), and Genetic Programming (GP), are powerful methodologies that have been used to solve many complex problems dealing with data points. These methods have had successful applications in marine engineering. For instance, an ANN model was developed by Van Gent et al. (van Gent et al., 2007) to estimate wave overtopping discharges for a wide range of coastal structures, including berm breakwaters. Formentin et al. (2017) introduced an ANN tool using an extensive experimental dataset consisting of nearly 18,000 tests to predict the main parameters describing the wave-structure interaction processes, such as mean wave overtopping discharge, wave transmission, and wave reflection coefficients. In addition, success has been achieved by using machine learning tools, such as SVM and hybrid GA with SVM models, for berm breakwaters, particularly to predict the damage level in non-reshaping berm breakwaters (Harish et al., 2014). In a similar study, Harish et al. (2015) developed an SVM tool and Particle Swarm Optimization (PSO) with SVM hybrid to predict the damage level of non-reshaping berm breakwaters. (MAST, 1997). An M5' machine learning approach is used to formulating the berm recession (Hosseini and Shafieefar, 2016). A Bayesian probabilistic model was developed by Pontiki (2019) for berm breakwaters failures probability prediction and the corresponding uncertainties in the arctic regions.

Looking at the history of research in the field of berm breakwaters, there have been invaluable experimental and analytical studies that resulted in some design formulas with their own inevitable limitations. Although efforts have been made on developing or modifying the existing formulations based (van Gent et al., 2007) on a larger database consisting of multiple data sources, the outputs were either hard to access (in form of an ANN) or too complicated (with a higher probability of user mistakes) to be utilized as a design tool.

In this study, reliable experimental data on berm recession were collected from different sources. Based on the fundamental differences in the test data due to the different experimental approaches, it was not practically acceptable to use all data at once without any modifications. That is why researchers developed different formulas based on the experimental approach they adopted. Therefore, by developing a code, data homogenization considering the effect of important parameters such as the number of waves and damage parameters was conducted. Then, Multi-Objective Genetic Programming (MOGP) technique was used on the homogenized data to find a relationship between the inputs (sea state and structural) and the target parameter (berm recession). Since various experimental outputs from different sources were used, which is the main advantage of this study, the general model allowed more comprehensive predictions of recession in berm breakwaters for design purposes. The performance of the MOGP model was examined and compared with the existing formulas in the literature. Moreover, following the development of a MOGP model, both parametric and sensitivity analyses were conducted to examine its performance. Fig. 2 summarizes the flow of work in the present study.

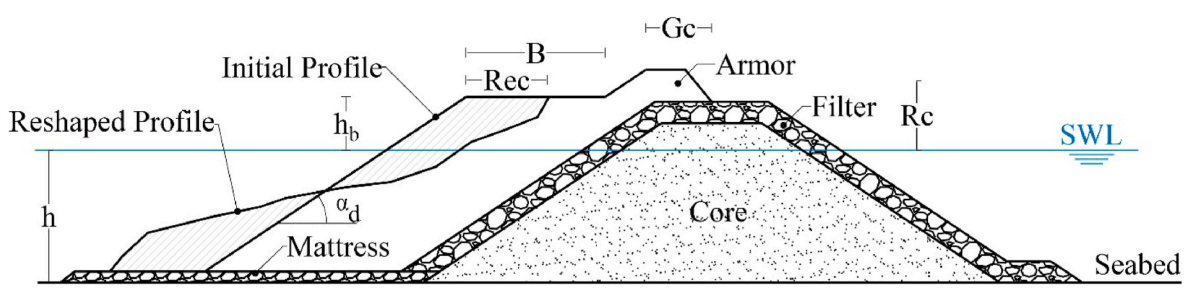

Fig. 1. Schematic illustration of the important parameters in berm breakwaters recession.

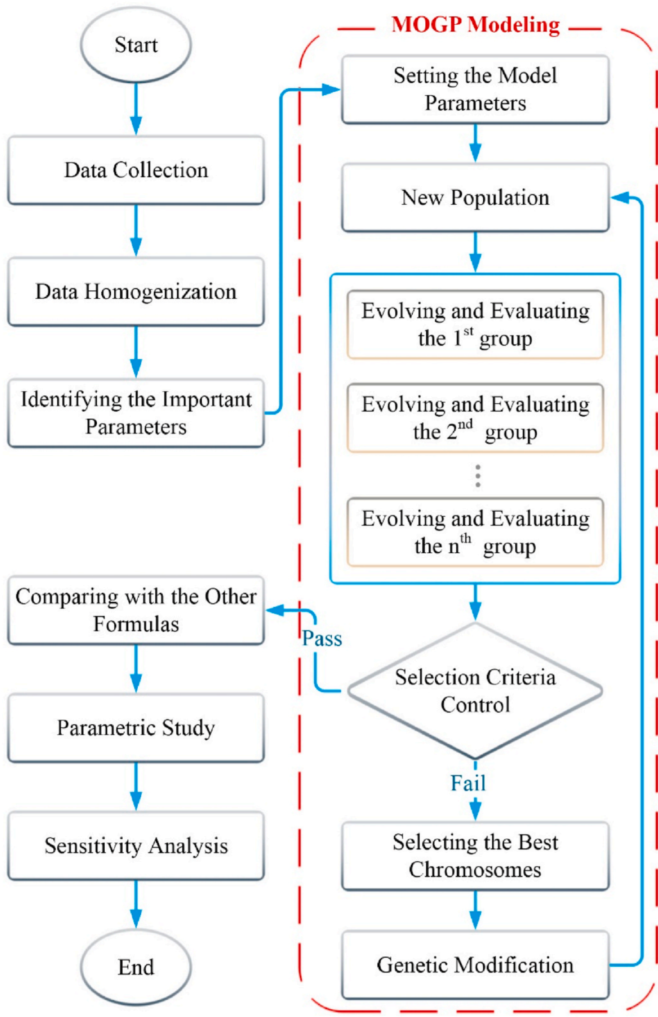

Fig. 2. Workflow of this study.

2. Methodology

2.1. Experimental data

Since the recession of berm breakwaters is a progressive damage phenomenon, normally, the experiments were categorized into two different groups, cumulative and rebuild. The former group considers the effect of a cycle of test conditions by the method of cumulative damage, in which the breakwater section experiences a test sequence without rebuilding the structure. Obviously, at each stage of the experiment, a few rocks may be detached from the breakwater berm and fall to the toe of the structure. As each wave collides with the structure, the interlocking of stable stones increases (gaining more interlocking), not only due to the separation of the unstable stones from their original location but also because the reshaped profile has a milder frontal slope and is more stable than the initial slope. Therefore, when the maximum wave in a spectrum hits the structure in the final test, the initial relative stability that the breakwater section gained from the previous tests leads to lower values of berm recession. This methodology can contain deviations from possible events in reality. In real cases, the structure can be prone to waves close to the maximum or design wave of a breakwater section right after its construction. This can cause the cumulative method to result in lower values than the probable environmental conditions.

The second group of experiments, so-called the rebuild method, considers an undamaged section of berm breakwaters being exposed to the sea states in the physical modeling of the structure. Since a spectral wave is composed of a range of small waves to the high ones, the first wave train that strikes the structure is not necessarily a destructive wave for that breakwater section, but in any case, the structure experiences the highest wave of the spectrum during the test. This test condition gives more reliable and confident values for engineering design and judgment. Therefore, generally, the results of rebuild test measurements are of higher values than those extracted from cumulative tests, which explains the difference in the recession prediction results using the formulae of Hall and Kao (1991) and Tørum et al. (2003) in comparison with those presented by Lykke Andersen and Burcharth (Lykke Andersen and Burcharth, 2010), Moghim et al. (2011) and Shekari and Shafieefar (2013).

Table 1 presents the datasets used in the analyses (van der Meer and Sigurdarson, 2017) (Sadat Hosseini and Shafieefar, 2014) (van der Meer and Sigurdarson, 2017) (Moghim et al., 2011) (Research and Association, 2007) (Koza, 1992).Overall, a total of 805 data points were considered for developing the Genetic Programming by dividing the data into Train, Validation, and Test subsets after the homogenization.

Berm breakwaters recession depends on the wave height and period, number of waves or storm duration, as well as geometrical parameters, such as rock relative density and gradation, berm elevation and width, and the initial slope ((Moghim et al., 2011), (Tørum et al., 2003), (Shekari and Shafieefar, 2013), (Ehsani et al., 2020), (Hosseini and Shafieefar, 2016)). There are a few non-dimensional parameters that have been used by researchers and presented in Eq. (1). $H_0\sqrt{T_{0p}}$ was selected as the representative of the stability against a particular sea state. h/D_{n50} is owed to the representation of the average number of rocks in-depth, h_b is divided by H_s to represent the relative berm elevation with respect to wave height. This parameter is generally used as a design parameter in berm breakwaters. The number of waves was used in a logarithmic form (Hall and Kao, 1991). The rock gradation factor and the lower front slope of the berm breakwater were also used as the input parameters. Some researchers (Shekari and Shafieefar, 2013), (Sadat Hosseini and Shafieefar, 2014) used the berm width as an input parameter in their studies which means that a recursive process is needed to find the berm recession through convergence in the formulas. However, this methodology for design is tedious and more importantly, some researchers avoided the use of this parameter because it was not found an effective parameter on berm recession (Moghim et al., 2011). Therefore, this parameter was not introduced to the MOGP model. Rec/D_{n50} represents the average number of rocks displaced from the berm breakwater.

$$\frac{Rec}{D_{n50}} = f\left(H_0\sqrt{T_{0p}}, \frac{h}{D_{n50}}, \frac{h_b}{H_s}, LnN, f_g, cot\ \alpha_d\right) \tag{1}$$

2.2. Data homogenization

Each of the datasets in Table 2 contains several test series, and each test series contains several tests. Tests are performed under different conditions and are divided into two categories: rebuild and cumulative. The results of the rebuild tests will be used directly in the MOGP model. But the results of cumulative tests must first be converted to the equivalent of their rebuild tests. The general flowchart of this process is shown in Fig. 3.

Conversion of cumulative tests begins with the detection of separate sets of tests from each other. Each cumulative collection is divided into separate sets by comparing some parameters related to the physical characteristics of the tests. These parameters are H_s, D_{n50}, D_{85}/D_{15}, Rc, B, h_b, and G_c. If the values of these parameters are the same in some tests, those tests will be classified into one test series.

In the initial classification based on physical characteristics, it is realized that in some test series only the breakwater was reconstructed, and the physical characteristics remained the same. To identify and separate these special cases, their recession value was compared. In each

Table 1
Datasets corresponding to cumulative and rebuild test conditions used in this study.

Cumulative										
Dataset	Number of tests	H_0	T_{0m}	h/D_{n50}	h_b/H_s	N/3000	B/D_{n50}	f_g	$\cot\alpha_d$	Rec/D_{n50}
Keilisnes (van der Meer and Sigurdarson, 2017)	3	2.11–2.40	29.12–38.82	27.65–28.26	−0.44 ~ −0.25	2.67–5.80	12.96	1.28	1.3	2.59–8.21
MAST II (1997) (MAST, 1997)	15	1.95–2.95	23.65–33.79	11.36	−1.37 ~ −0.46	0.67	15.91–18.18	1.8	1.25	1.45–10.91
Project 1 (3 series) (van der Meer and Sigurdarson, 2017)	32	1.76–2.94	21.87–34.98	10.17–13.36	−2.13–0	0.28–0.82	9.53–15.89	1.52–1.72	1.10–1.50	1.18–10.77
Project 2 (van der Meer and Sigurdarson, 2017)	7	2.39–3.00	22.57–37.70	7.28	−0.37 ~ −0.30	0.55–0.91	13.97	1.43	1.5	2.62–11.64
Project 3 (van der Meer and Sigurdarson, 2017)	11	2.11–2.67	33.49–34.71	10.96–12.80	−0.58 ~ −0.46	0.22–0.33	6.61–8.74	1.65	1.5	1.04–5.01
Project 4 (van der Meer and Sigurdarson, 2017)	16	1.72–2.95	20.39–32.29	10.53–14.32	−1.18 ~ −0.41	0.28–1.43	10.98–13.89	1.10–1.58	1.25	0.74–7.78
Project 5 (van der Meer and Sigurdarson, 2017)	12	2.58–2.71	21.86–28.51	17.20–19.67	−1.33 ~ −0.73	0.84–1.09	13.27–17.76	1.6	1.33	4.05–15.85
Lykke Andersen (2006) (Lykke Andersen, 2006)	446	1.70–3.00	20.22–40.12	8.76–22.33	−0.62–1.48	1	7.73–33.01	1.35–1.45	1.25	0.36–19.22
Lykke Andersen (2008) (Lykke Andersen, 2006)	11	1.71–2.96	20.37–32.16	11.56	−0.83 ~ −0.48	0.4	7.14	1.3	1.3	0.29–6.65
Rebuild										
Moghim (2009), NTNU (Moghim et al., 2011)	10	1.87–2.89	22.35–27.02	26.17	−0.75 ~ −0.49	0.67	9.5	1.1	1.3	0.83–6.90
Moghim (2009), TMU (Moghim et al., 2011)	125	1.80–3.00	24.07–37.45	14.12–26.47	−1.17 ~ −0.21	0.17–2.00	17.65–26.47	1.5	1.25	2.65–17.35
Motalebi (2010), TMU (Motalebi, 2010)	117	1.87–2.96	19.61–25.94	8.00–16.47	−1.24 ~ −0.44	1	9.52–17.65	1.82	1.25	1.04–8.59

Table 2
Breakwater and sea states specifications.

Sea state	Δ	D_{n50}	Cot(a)	T_m	H_s	P	M_{50}	N_w
1	1.63	0.0323	1.25	1.17	0.096	0.55	10t	3000
2	1.63	0.0323	1.25	1.23	0.106	0.55	10t	3000
3	1.63	0.0323	1.25	1.27	0.112	0.55	10t	3000

set, the number of waves is constant. So, it is evident that in these cases, the recession of the first test in a set will be less than the overall recession after the last test of its previous set. As a result, these cases can also be identified and distinguished from each other.

Once the separate sets were detected in a dataset, the equivalent rebuild tests of each cumulative test can be achieved using the method described by Van der Meer (van der Meer and Sigurdarson, 2017). This method is briefly explained in the following paragraph.

There are two or more sea states in each set of cumulative data (Table 2). For each sea state, the S_d - N_w (damage-Number of waves) curve will be found using the original formula, given in Eq. (2) (van der Meer, 1988):

$$\frac{H_s}{\Delta D_{n50}} = \begin{cases} 6.2P^{0.18}\left(\frac{S_d}{\sqrt{N_w}}\right)^{0.2} \xi_m^{-.05}\xi_m < \xi_{cr}\ (Plunging\ waves) \\ \frac{H_s}{\Delta D_{n50}} = 1.0P^{-0.13}\left(\frac{S_d}{\sqrt{N_w}}\right)^{0.2}\sqrt{cot\ \alpha}\ \xi_m^P \xi_m \geq \xi_{cr}\ (Surging\ waves) \end{cases} \tag{2}$$

Wherein,

$$\xi_{cr} = \left[6.2P^{0.31}\sqrt{tan\ \alpha}\right]^{\frac{1}{p+0.5}}, \xi_m = tan\ \alpha \Bigg/ \left[\frac{2\pi H_s}{gT_m^2}\right]^{0.5} \tag{3}$$

By using this formula, if S_d is specified, N_w can be calculated and vice versa. The following conditions should also be considered to using this formula (Research and Association, 2007).

- This formula only covers deep water conditions at the toe ($h > 3H_s$).
- According to equation (2), S_d is a function of the square root of N_w. But when the number of waves is less than 1000 ($N_w < 1000$) the equivalent number of waves should be used in the original formula, which is $N_{eq} = N_w^2/1000$. Therefore, there is a linear relationship between the damage parameter, S_d, and the number of waves, N_w, for $N_w < 1000$.
- Regardless of ξ_{cr} and ξ_m, only the plunging waves equation should be used if $cot\ \alpha \geq 4$.

The conversion of a cumulative test into its equivalent rebuild test begins with specifying the number of waves in the first sea state (N_{w1}). Next, the damage (S_{d1}) is calculated for N_{w1} and the first sea state.

Then, the number of waves that caused Sd1 damage in the second sea state is calculated (N_{w12}). Finally, by adding the calculated number of waves (N_{w12}) and the number of Sea state 2 waves (N_{w2}), the final damage of Sea state 2 (S_{d2}) can be calculated. The number of waves and the damage of subsequent sea states are calculated in the same way, so that a set of cumulative data are converted to several rebuild data. Fig. 4 shows an example of converting three consecutive cumulative tests to their equivalent rebuild tests. The breakwater and sea state specifications are presented in Table 2.

2.3. Model development

2.3.1. The genetic programming methodology

Based on the principle of Darwinian Natural Selection theory, Gene Programming (GP) was introduced by Koza (1992) as a useful and powerful prediction algorithm to establish meaningful relationships between the input and output parameters involved in a problem. GP creates a population of random compositions of the functions (such as algebraic functions) and terminals (the variables or constants used in a problem) (Sadat Hosseini et al., 2021) for the first generation. After preparing the population of the first generation, genetic operations such as crossover (also known as recombination) and mutation are used to create new generations. Crossover generates new offspring by combining the parental genetic information (exchanging the sub-trees

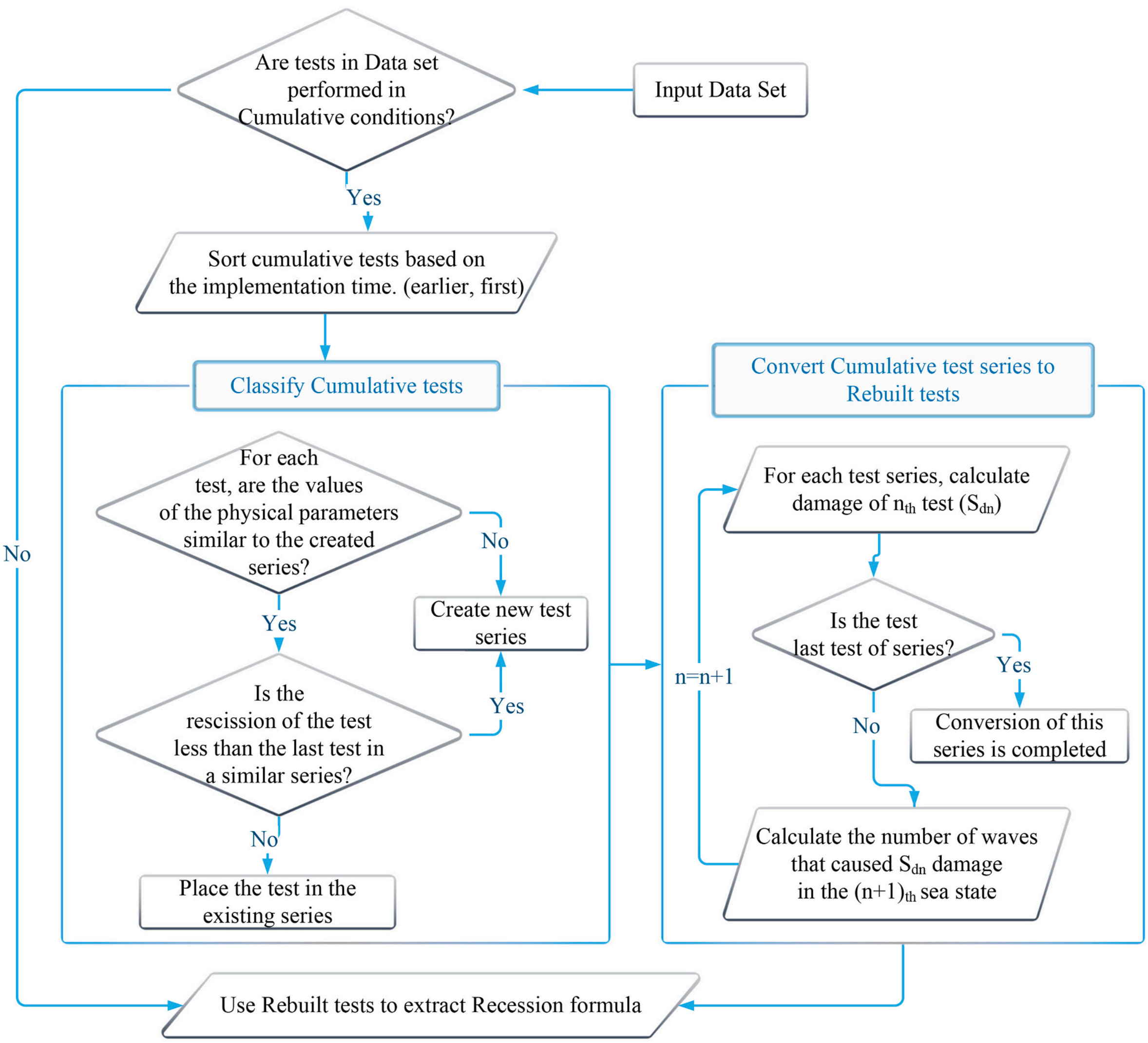

Fig. 3. Process of data homogenization.

under a selected node). Mutation increases the population variation by creating new individuals from an existing tree in the population (Fig. 5). The higher the fitness value, the greater the chance of remaining in the population of successive generations.

In the standard GP, a symbolic regression technique can be implemented. In this method, the nodes of a tree (gene) contain functions or terminals. In this case, a new language was developed by Karva (Salgotra et al., 2020) that makes it possible to read the chromosomes' information. Genes exhibiting K-expressions are intelligible computer programs in the Karva language. These extremely compact expressions simply consist of letters that represent the variables, and the constant numbers correspond to a specific problem (Ferreira, 2001). Therefore, according to Fig. 6, symbolic mathematical expressions are directly encoded by genes. This model consists of two genes that can predict the output value using linear and nonlinear functions on three inputs A, B and C. Although nonlinear functions such as exp and Ln was used in the structure of these genes, a weighted linear combination of the genes produces the overall model based on the following general formulation for Multi Gene Genetic Programming (MGGP):

$$\hat{y}(x,w,r) = w_0 + \sum_{i=1}^{i=n} w_i G_i(r,x) \tag{4}$$

Wherein, n is the number of genes, and y is the output, which is a function of inputs, x. The ith gene weight is w_i, and bias term is w_0. The vector of unknown parameters for each gene is r, and the vector of outputs is G. One of the major differences of MGGP with the standard GP is its higher accuracy and efficiency in modeling nonlinear complex problems (Gandomi et al., 2021) like the problem at hand. Having the input variables, different functions and a range of random constant values, the initial population of MGGP can be constructed with maximum diversity in order not to have duplicate genes.

The tree-based GP and MGGP generally optimize the single objective of a problem according to a fitness function which is mostly the goodness-of-fit in symbolic regression problems. This single objective approach can lead to very complex and non-robust models by incorporating less important and ineffective terms in the model. Although less complicated genes can be developed using MGGP in comparison with traditional GP, the tendency of MGGP to produce genes that have a negligible effect on the overall performance causes over-complexity problems (Searson, 2015). To restrict the model from overexpansion a multi-objective strategy is implemented into the regression genes. This method, which is called Multi-Objective Genetic Programming (MOGP) can optimize the development complexity and the goodness-of-fit simultaneously. Amongst the different techniques of MOGP, in this paper, the GPTIPS 2 toolbox (Searson, 2015) was used in MATLAB. In this method, the generated models from the MGGP algorithm are sorted based on their complexity and accuracy. The top 50% of the population survive to the next generation, while the rest are omitted. For more

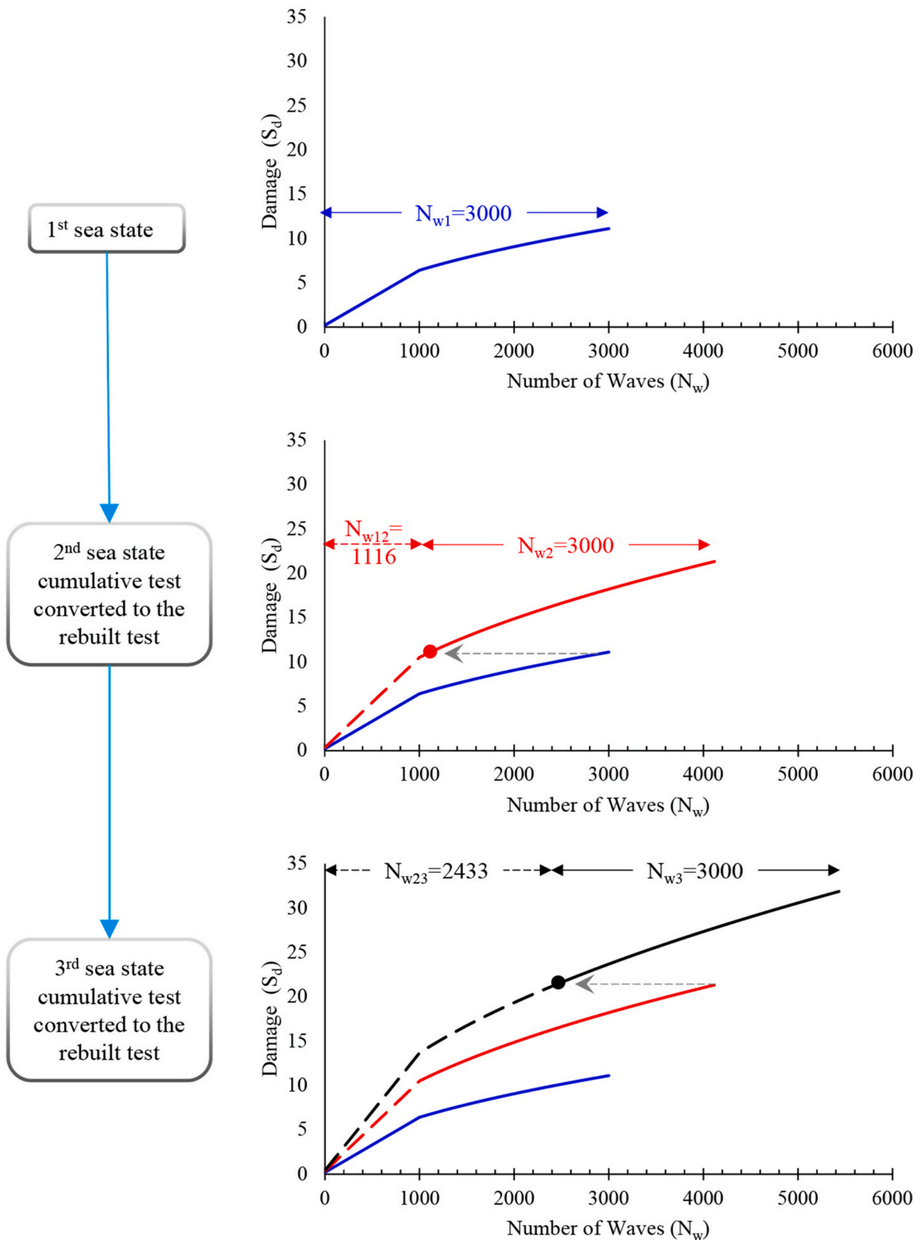

Fig. 4. Methodology of converting cumulative tests to the equivalent rebuild tests (van der Meer and Sigurdarson, 2017).

information on the details of the MOGP approach and its applications, references are made to Searson (2015) and Gandomi et al. (2021).

2.3.2. Preparation of datasets

Based on the data homogenization approach, a unified and homogeneous data package was developed. Consequently, the data points whose H_0 is smaller than 1.7 or greater than 3.0 were omitted from the dataset based on the practical range of H_0 (van der Meer and Sigurdarson, 2017). Moreover, the data points corresponding to $Rec/Dn50 > 15$ were omitted to be in the practical range of the Berm Breakwaters design.

These data are randomly divided into Training, Validation, and Test subsets for the analysis. The Training subset (75% of data) was used to develop the MOGP algorithm, while the Validation data (15% of data) were used to examine the generalization of the predictions of the model on data that was not used in the model Training. Based on the number of terminals, functions, genes, the depth of the trees, fitness termination value and other influencing parameters in modeling, and through the optimization process in MOGP, the best model on the data were obtained. The remaining 10% of data were finally used for model performance assessment against other formulae reported in the literature.

2.3.3. MOGP formulation

Six input parameters, including $H_0\sqrt{T_{0p}}$, h/D_{n50}, h_b/H_s, $N/3000$, f_g and $\cot\alpha_d$, substantially contributed to the generation of the MOGP model. The stability number, $H_0\sqrt{T_{0p}}$ is the most determining parameter in the stability and/or reshaping of berm breakwaters. The other non-dimensional parameters were selected in accordance with previous studies (Moghim et al., 2011), (Lykke Andersen and Burcharth, 2010), (Hosseini and Shafieefar, 2016) and physical logics. To devise a correct and valid MOGP model for the prediction of the target parameter, Rec/D_{n50}, several runs were conducted, the model variables were changed, and the accuracy of the models was controlled using the selection criteria explained here.

On the one hand, the model should be simple enough to be easily used for practical estimations, but on the other hand, it has to be flexible

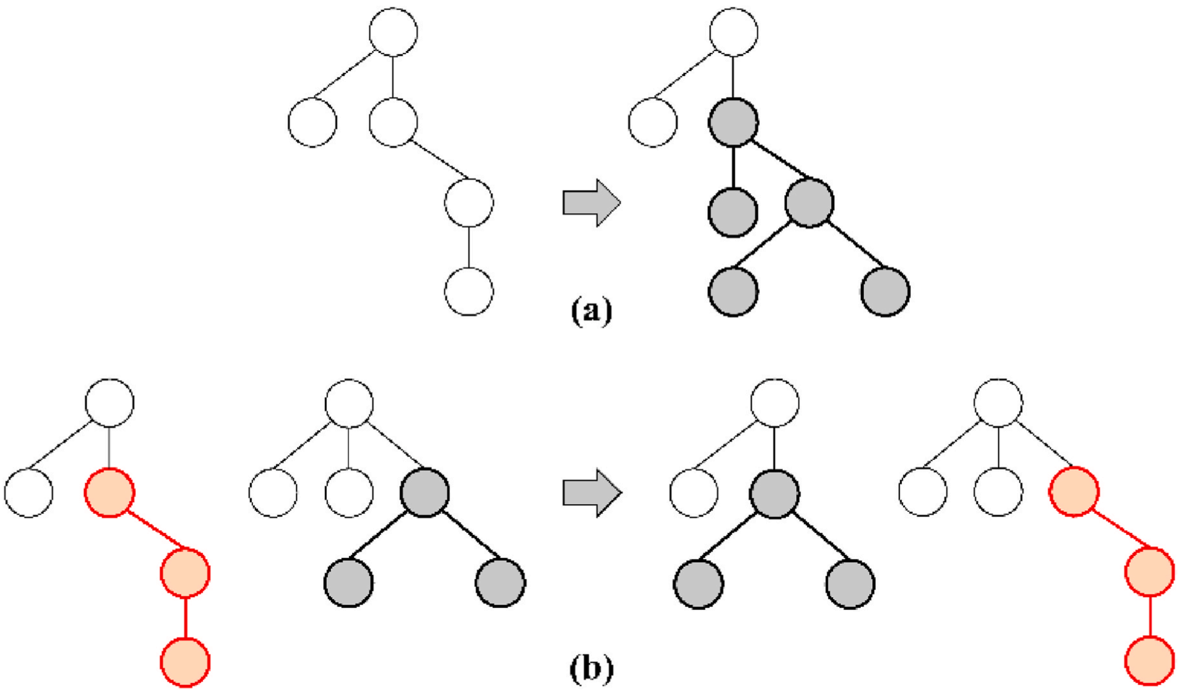

Fig. 5. Example scheme of a) Mutation and b) Crossover in a population.

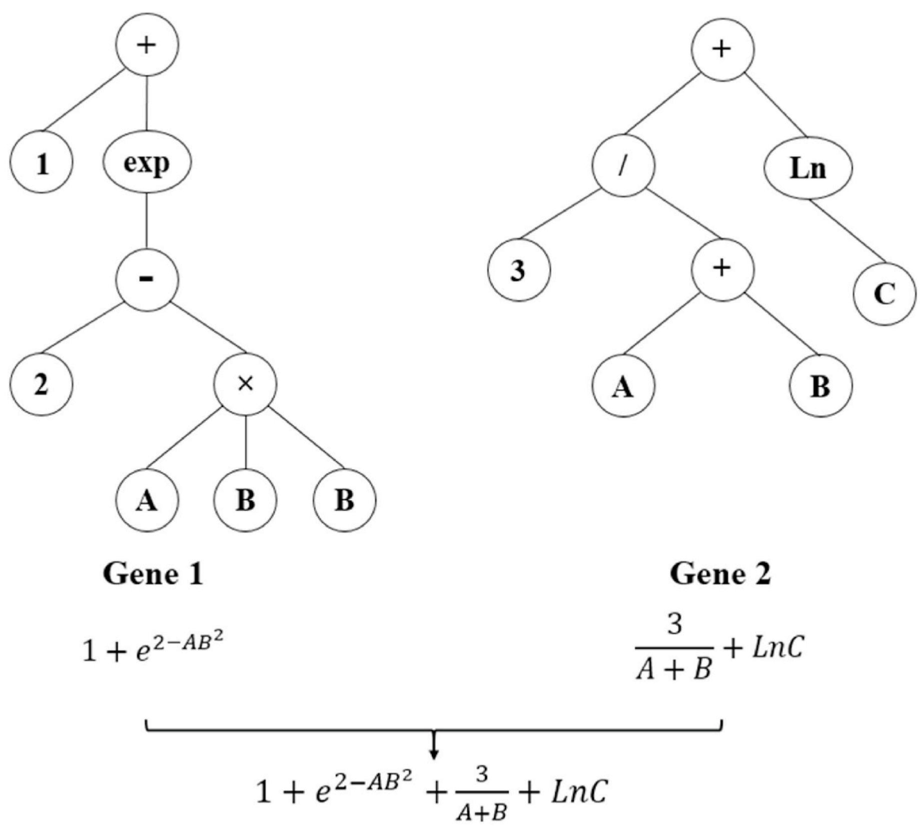

Fig. 6. Example of an MGGP model consisting of two genes and three inputs.

enough to provide good approximations of the berm recession. However, the latter was of greater importance for the authors in developing the MOGP model. The following three objectives were considered in the process of Validation control and Training performance check:

1) The model on the Training subset has the best fitness value and the least error among all derived GP models.
2) The trained model shows the best fitness value and the least error on the Validation subset.

Models were controlled regarding the first and second criteria to examine the accuracy of the predictions, using root mean squared error (RMSE) as follows:

$$RMSE = \left[\frac{1}{n}\sum (O-P)^2\right]^{0.5} \tag{5}$$

where O and P are the objective (experimental output) and predicted output by MOGP model, respectively; $\overline{O}$ and $\overline{P}$ are the average values of O and P data points, respectively; and n is the sample number. The predicted model with a specific range of parameters (Table 2) was evaluated using the correlation of coefficients:

$$R = \frac{\sum (O-\overline{O})(P-\overline{P})}{\left[\sum (O-\overline{O})^2 (P-\overline{P})^2\right]^{0.5}} \tag{6}$$

A model prediction with a higher R value, a lower RMSE, and a lower model complexity is more likely to be selected.

3) The final model is as simple as possible based on the complexity of the problem for practical use.

Selecting optimum values for the model parameters not only decreases the run time but also results in less complicated models with relatively higher accuracy. By employing a trial-and-error approach, the parameters of the MOGP model on the recession data are obtained. Population size, or the number of chromosomes, serves as the critical parameter that determines the number of evolved programs in the modeling. The complexity of a certain problem and the number of possible solutions determine the appropriate functions and population size. In the analyses, ten well-known mathematical functions and arithmetic operators such as +, -, × ,/, √, exp, Ln, power, add3, mult3, wherein add3(a,b,c) = a+b + c and mult3(a,b,c) = a × b × c, were selected based on the literature on berm breakwaters. The number of functions was limited to hinder the model overgrowing. Six sets for the population size (i.e., 500, 1000, 3000, 5000, 10000 and 15000) were considered. The complexity of the model terms is determined by the maximum number of genes allowed and the maximum tree depth.

3. Results

By employing the trial-and-error approach, the hyperparameters of the MOGP model on the recession data are as presented in Table 3.

Table 3
Hyperparameters of the MOGP model on berm breakwaters recession.

Hyperparameter	Value or setting
Population size	10000
Number of generations	5000
Maximum number of genes allowed	3
Maximum tree depth	3
Training set	0.75
Validation set	0.15
Crossover events	0.85
High-level crossover	0.20
Low-level crossover	0.80
Sub-tree mutation	0.90

Among the six sets for the population size, the best model was obtained using the population size of 10,000. Increasing the population size had a negligible effect on the model precision. The complexity of the model terms is determined by the maximum number of genes allowed and the maximum tree depth.

The best-of-run model was the one with the highest R and the lowest RMSE and complexity. The final formulation for the prediction of the recession in berm breakwaters is as follows:

Gene 1 and bias term:

$$-81.74 \times 10^{-5} \times 1.378^{x_3} {x_1}^3\, x_2 {x_3}^2 - 1.245 \tag{7}$$

Gene 2:

$$21133{x_1}^{2.094}\, x_3 x_4 \times Ln(x_5)^{7.187^{x_5}} \tag{8}$$

Gene 3:

$$0.02067{x_1}^{1.971}{x_4}^{-0.1087}\, Ln(x_2 x_4) \tag{9}$$

Wherein, $x_1 = H_0\sqrt{T_{0p}}$, $x_2 = N/3000$, $x_3 = h_b/H_s$, $x_4 = h/D_{n50}$, $x_5 = f_g$. For the range of data used in the modeling process, cotα was not as influencing as the other parameters since it could not stand out in any of the genes after 5000 generations. However, it should be noted that the driven formula is only valid for the range of parameters in the input data. The overall formula for the prediction of berm recession is the summation of the above genes:

$$\begin{aligned}\frac{Rec}{D_{n50}} =& 0.02067\left(H_0\sqrt{T_{0p}}\right)^{1.971}\left(\frac{h}{D_{n50}}\right)^{-0.1087} Ln\left(\frac{N}{3000}\times\frac{h}{D_{n50}}\right)\\ &+21133\left(H_0\sqrt{T_{0p}}\right)^{2.094}\left(\frac{h_b}{H_s}\right)\left(\frac{h}{D_{n50}}\right)\times Ln(f_g)^{7.187^{f_g}}\\ &-81.74\times 10^{-5}\times 1.378^{\left(\frac{h_b}{H_s}\right)}\left(H_0\sqrt{T_{0p}}\right)^3\left(\frac{N}{3000}\right)\left(\frac{h_b}{H_s}\right)^2 - 1.245\end{aligned} \tag{10}$$

To evaluate the acceptability of the MOGP model, the Frank and Todeschini (1994) criterion was used, in which the ratio of the amount

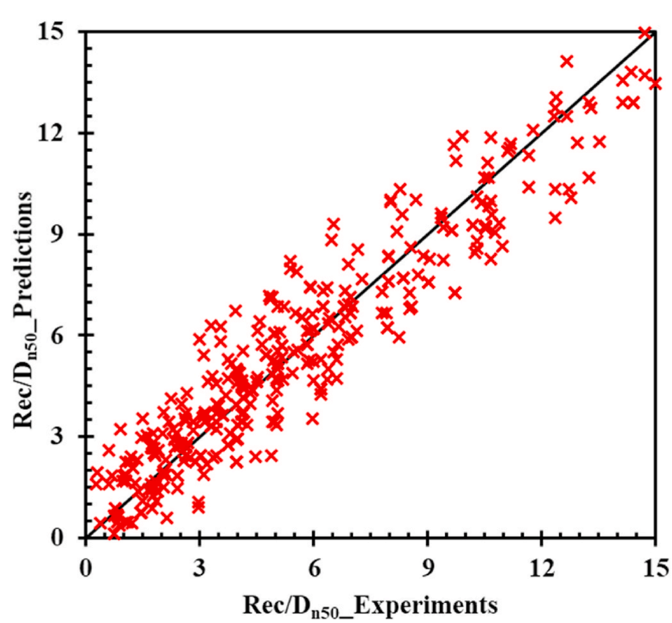

Fig. 7. Scatter diagram of the MOGP model predictions.

Table 4
Results of the statistical indices for the Training and Validation subsets in the MOGP formula.

$$k = \frac{1}{O^2} \sum O \times P \tag{11}$$

$$k' = \frac{1}{P^2} \sum O \times P \tag{12}$$

Train (75% data)		Validation (15% data)	
R^2	RMSE	R^2	RMSE
0.911	0.111	0.906	0.134

For the external validation of the proposed formula, at least one of the regression line slopes (i.e. k and k'), passing through the origin, should be close to one (i.e. $0.85 < k, k' < 1.15$) (Golbraikh and Tropsha, 2002).

of data to the number of inputs should be greater than five. This ratio for the model on the berm recession data were 60, which indicates that a valid number of data were used in the analyses. According to Smith (1986), the error value (e.g., RMSE) must be at a minimum, while the correlation of coefficients (i.e. R) must be higher than 0.8. These statistical parameters were checked for the Training, Validation and Test subsets in the MOGP model. Fig. 7 depicts the predictions of the MOGP model versus the experimental data, in which a good correlation can be seen.

Table 4 presents the calculated statistical indices (i.e., RMSE and R^2) for the proposed formula. Accordingly, it is apparent that the results of Training and Testing are very close, indicating that the models did not overfit.where O values are the output values from the experiments; and P represents the MOGP prediction values. The calculated values of the validation criteria are $k = 0.959$ and $k' = 1.019$. These values show that the MOGP formula meets the external validation criteria.

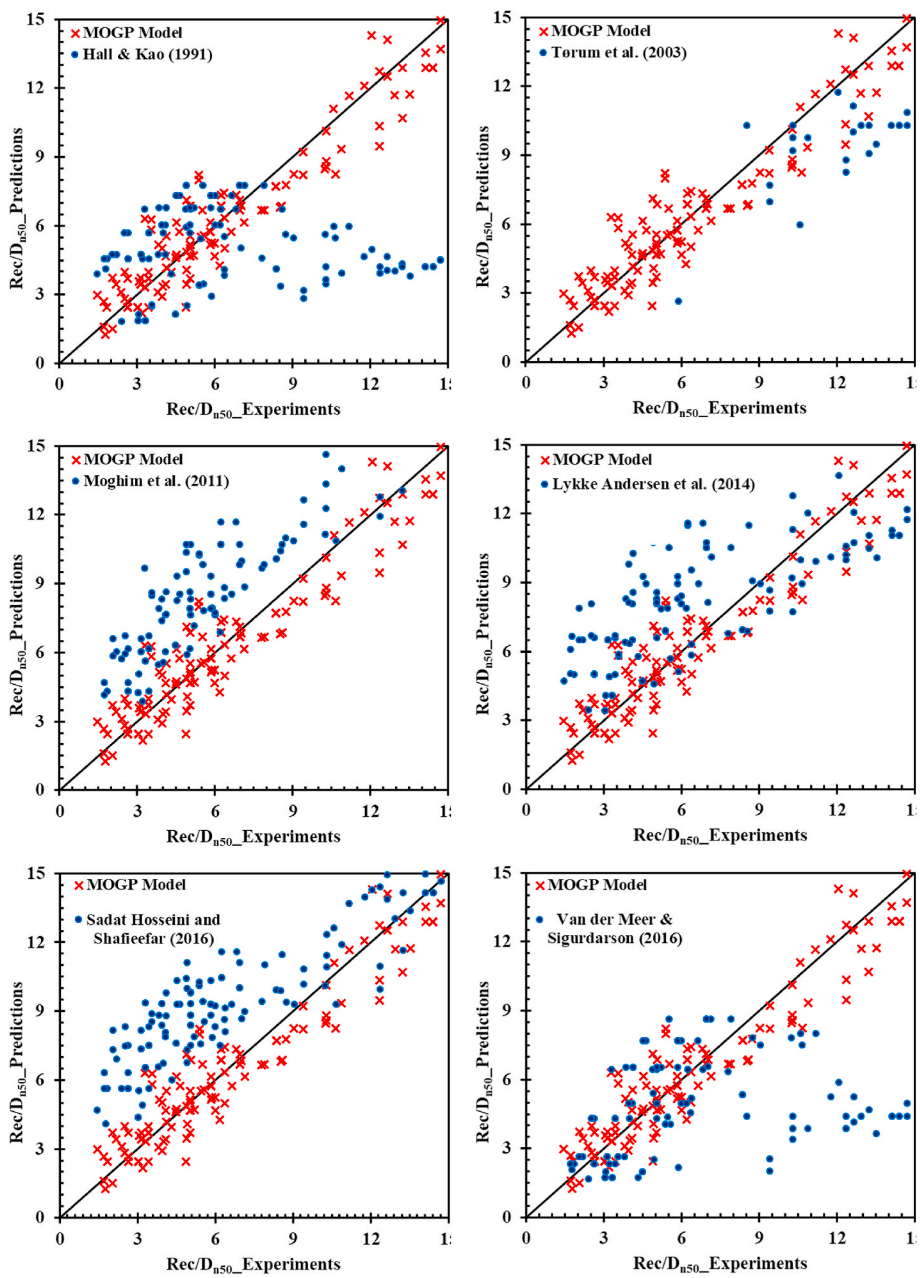

Fig. 8. Scatter diagram of predictions versus the experimental data for comparing the available empirical models with MOGP predictions on the Test dataset (10% of total data).

Table 5
Statistical indices for the external validation on the Test dataset (10% data).

Equation	R^2	RMSE	Percentage of data coverage	Range of applicability
MOGP Model (Eq. (10))	0.904	0.128	100	$5.8 < H_0\sqrt{T_0} < 28.2$ $500 < N < 17400$ $1.5 < h_b/H_s$[a] < 3.1 $7.3 < h/D_{n50} < 28.3$ $1.1 < f_g < 1.8$
Hall and Kao (Hall and Kao, 1991)	0.691	0.189	66.8	$2 < H_0 < 5$
Tørum et al. (Tørum et al., 2003)	0.676	0.204	51.6	$12.5 < h/D_{n50} < 25$ $1.3 < f_g < 1.8$
Moghim et al. (Moghim et al., 2011)	0.823	0.133	19.7	$7.7 < H_0\sqrt{T_0} < 24.4$ $500 < N < 6000$ $0.12 < h_b/H_s < 1.24$ $8 < h/D_{n50} < 16.5$ $1.2 < f_g < 1.5$
Sadat Hosseini and Shafieefar (Hosseini and Shafieefar, 2016)	0.743	0.176	32.1	$6.3 < H_0\sqrt{T_0} < 24.4$ $500 < N < 6000$ $0.1 < h_b/H_s < 6.82$ $8.0 < h/D_{n50} < 16.5$ $12 < B/D_{n50} < 29.41$
Lykke Andersen et al. (Andersen et al., 2014)	0.760	0.181	100	–
Van der Meer and Sigurdarson (van der Meer and Sigurdarson, 2017)	0.690	0.251	100	–

[a] Positive values of hb stand for emerged berm.

4. Discussion

4.1. Comparison with the literature formulas

A direct comparison was made between the predictions of the model and the available implicit formulas in the literature on the Test data (10% of the total data). Six empirical formulae were selected, and the scatter diagrams of the predictions by these models in comparison with the predictions of the MOGP model are presented in Fig. 8. The correlation of coefficients and Root Mean Square Error were calculated (Table 5). According to the table, some models showed lower performance than the reported results in the corresponding studies. This can be attributed either to the modification of the cumulative dataset carried on in this study or to some differences in the structural and hydraulic properties of the tests conducted in different labs that do not exactly match. For instance, the formula of Hall and Kao (1991) were applicable to 66.8% of the Testing data due to its range of applicability ($2 < H_0 < 5$ and $\cot\alpha d = 1{:}1.25$). This can be seen for the Tørum et al. (2003) formula and Sadat Hosseini and Shafieefar (2016) with 51.6 and 32.1 percent of data coverage, respectively. The accuracy of predictions by Moghim et al. (2011) (Shekari and Shafieefar, 2013) formula is relatively high. However, they used relatively large berm widths in their experiments, and the proposed equations have a relatively limited range of applicability compared with the MOGP model, Lykke Andersen et al. (2014) and Van der Meer and Sigurdarson (van der Meer and Sigurdarson, 2017). By comparing these scatter diagrams, another interesting point that stands out is that the formulas which were developed on rebuild data such as Moghim et al. (2011) and Sadat Hosseini and Shafieefar (2016) yield relatively overestimated values while those that were originally developed on cumulative sets such as Hall and Kao (1991) and Tørum et al. (2003) result in lower recession values comparing with that of the MOGP model. The scattering of the results tends to increase for $Rec/D_{n50}>10$ for Hall and Kao (1991) and Van der Meer and Sigurdarson (van der Meer and Sigurdarson, 2017), which is mainly due to the fact that $Rec/D_{n50}>10$ is considered impractical in their research.

4.2. Parametric study

To examine the pattern of the target parameter (non-dimensional recession) regarding the changes in the corresponding input parameters and the robustness of the prediction model, a parametric study was conducted. The general trends of model predictions against the inputs are illustrated in Fig. 9 for the presented formula. As seen in this figure, there is a relatively good agreement between the trends and the structure of the MOGP formula, and the rising effects of the stability parameter were captured. According to the literature, the stability number affects the recession values exponentially. The other interesting trend belongs to the number of waves, in which the berm recession rose dramatically up to around the corresponding value for N = 3000 waves and then the slope of the diagram decreased and consequently leveled off, as was observed in previous studies (Moghim et al., 2011), (Shekari and Shafieefar, 2013). An inverse parabolic trend can be seen for h_b/H_s and f_g. The former indicates that the highest damage occurs when the water level is equal to the berm elevation (i.e., $h_b = 0$), and the latter shows that there is a value of gradation factor that maximizes the recession. The same general trend as Tørum et al. (2003) study can be seen after $f_g = 1.5$. One advantage of the proposed model is the relatively wider domain of the input parameter (submerged to emerged berm levels) due to the larger parameter space covered by the data used here. As it is illustrated for h/D_{n50}, the deeper the water in front of the breakwater, the higher the damage probability. This is compatible with the experimental observations wherein higher waves can collide with the structure before breaking.

4.3. Sensitivity analysis

The level of influence and contribution of each contributing input parameter in the MOGP model was examined by calculating the sensitivity percentages of the output parameter by using the following equations (Sadat Hosseini et al., 2021):

$$D_i = f_{max}(x_i) - f_{min}(x_i) \tag{13}$$

$$S_i = \frac{D_i}{\sum D} \times 100 \tag{14}$$

where $f_{max}(x_i)$ and $f_{min}(x_i)$ correspond to the maximum and minimum values of the output parameter by substituting the *i*th input parameter (the target of sensitivity analysis) in the MOGP formula, while the average values of the other contributing parameters were inserted. The results of the sensitivity analysis in Fig. 10 are presented as a bar chart of the influence percentages for each input parameter in the modified data. It can be inferred that the most effective parameter is the stability parameter $H_0\sqrt{T_{0p}}$, while the berm elevation and the number of waves stand in the second and third places, respectively with around 20 percent sensitivity. The gradation parameter has by far lower effect on the recession value. The lower slope angle was omitted in the MOGP model generation process through the rule of "survival of the strongest species" as explained before. Comparatively, some previous research works did not consider some less influencing structural and sea state parameters in their presented formulations (van der Meer and Sigurdarson, 2017), (Sigurdarson et al., 2009b).

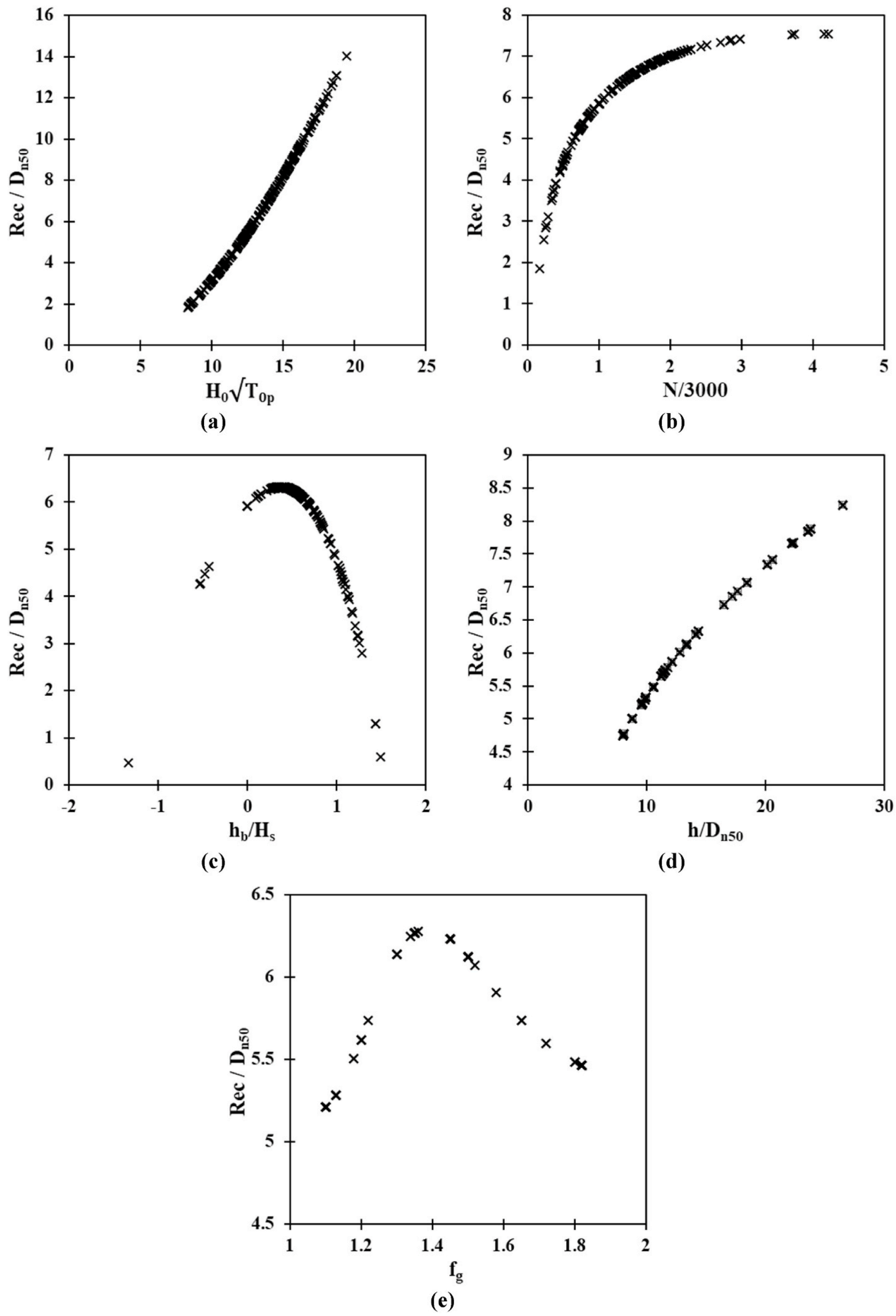

Fig. 9. Parametric study of the contributing inputs and target parameter in the MOGP model: a) $\mathrm{Ln}(H_0\sqrt{T_0})$, b) Ln(N), c) f_g.

4.4. Future research

Any new reliable datasets can be added to the existing data for enhancing the range of applicability and training of the machine learning model. The methodology of this study can be reconducted using other relationships between the damage parameter and the number of waves. Other soft computing techniques can be used to develop models after the homogenization of the data and compared with the present formulation.

5. Conclusion

An up-to-date form of Multi-Gene Genetic Programming (MGGP), namely Multi-Objective Genetic Programming (MOGP), was used to develop a formula for predicting recession in rubble mound mass armor berm breakwaters. The utilized data for the analysis consisted of two major subsets, namely rebuild and cumulative, based on the methodology of the experiments. Before developing a formula, the results of cumulative tests were converted to the equivalent of their rebuild tests by using the number of waves and the damage parameter.

The homogenized data were used as the input for MOGP modeling. The performance of the model was checked based on the correlation of coefficient and root mean square error. These two indicators were 0.911 and 0.11 for the Training set, and 0.906 and 0.134 for the Validation subset, respectively.

The accuracy of model predictions was subsequently examined for the Test dataset in comparison with the proposed formulas in the literature. Results show that the MOGP model not only covers a broader range of applicability but also can acceptably predict the recession in comparison to the other formulas.

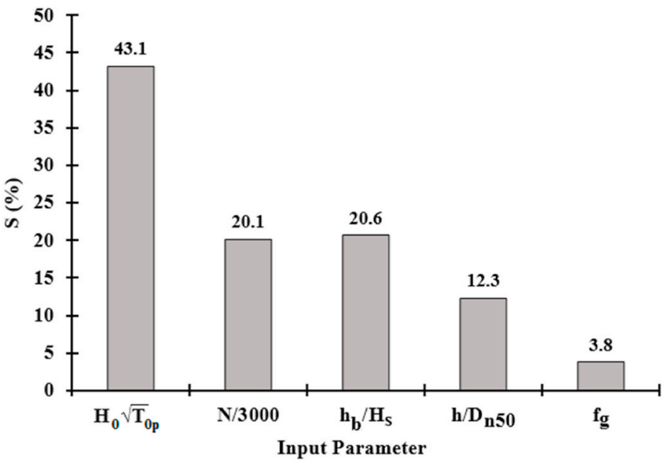

Fig. 10. Sensitivity analysis bar chart for the input parameters of the MOGP model.

A parametric study was conducted on the MOGP formula to analyze the response of the target parameter (Recession) to changes in the inputs of the proposed formula. Also, a sensitivity analysis was conducted to study the dependency level of the recession on each of the contributing input parameters. It was generally found that the modified dynamic stability parameter plays the most significant role in the datasets, which has by far the highest influence on the predictions for about 43 percent sensitivity followed by the number of waves, N, and berm elevation, h_b, in the second place (around 20%), and about 12 percent sensitivity on the water depth, h, and four percent on the rock gradation factor, fg.

Overall, each formula reported in the literature works well for its range of data. However, in engineering studies and the design of marine structures, it is difficult to use different formulas based on the applicability criteria. The database used for deriving the MOGP model corresponds to a relatively larger amount of data which can overcome this problem to some extent.

Ethical statement

The paper has been submitted with full responsibility, following due ethical procedure, and there is no duplicate publication, fraud, plagiarism. None of the authors of this paper has a financial or personal relationship with other people or organizations that could inappropriately influence or bias the content of the paper.

Funding body

The authors received no financial support for the research, authorship and publication of this article.

CRediT authorship contribution statement

Alireza Sadat Hosseini: Conceptualization, Methodology, Software, Writing – original draft. **Amir Kabiri:** Software, Data curation, Writing – original draft, Visualization. **Amir H. Gandomi:** Methodology, Supervision, Writing – review & editing. **Mehdi Shafieefar:** Supervision, Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The Python and Matlab codes are available on request. Each dataset needs a permission from its original owner.

Acknowledgements

The authors of this paper would like to officially acknowledge the considerable and invaluable efforts of the researchers who provided us with data acquired in their experimental studies. Moreover, the authors highly appreciate the provision of part of the experimental data by Prof. J. van der Meer, and his suggestions which helped to improve the paper's quality.

References

Andersen, T.L., Moghim, M.N., Burcharth, H.F., 2014. Revised recession of reshaping berm breakwaters. Coast. Eng. 2.

Ehsani, M., Moghim, M.N., Shafieefar, M., 2020. An experimental study on the hydraulic stability of Icelandic-Type berm breakwaters. Coast Eng. 156, 103599.

Ferreira, C., 2001. Gene Expression Programming: a New Adaptive Algorithm for Solving Problems. *arXiv preprint cs/0102027*.

Formentin, S.M., Zanuttigh, B., van der Meer, J.W., 2017. A neural network tool for predicting wave reflection, overtopping and transmission. Coast Eng. J. 59 (1), 1750006.

Frank, I.E., Todeschini, R., 1994. The Data Analysis Handbook. Elsevier.

Gandomi, M., Kashani, A.R., Farhadi, A., Akhani, M., Gandomi, A.H., 2021. Spectral acceleration prediction using genetic programming based approaches. Appl. Soft Comput. 106, 107326 https://doi.org/10.1016/j.asoc.2021.107326.

Golbraikh, A., Tropsha, A., 2002. Beware of q2. J. Mol. Graph. Model. 20 (4), 269–276.

Hall, K.R., Kao, J.S., 1991. A study of the stability of dynamically stable breakwaters. Can. J. Civ. Eng. 18 (6), 916–925.

Harish, N., Lokesha, N., Mandal, S., Rao, S., Patil, S.G., 2014. Parameter optimization using GA in SVM to predict damage level of non-reshaped berm breakwater. The International Journal of Ocean and Climate Systems 5 (2), 79–88.

Harish, N., Mandal, S., Rao, S., Patil, S.G., 2015. Particle Swarm Optimization based support vector machine for damage level prediction of non-reshaped berm breakwater. Appl. Soft Comput. 27, 313–321. https://doi.org/10.1016/j.asoc.2014.10.041.

Hosseini, A.S., Shafieefar, M., 2016. Prediction of seaward slope recession in berm breakwaters using M5'machine learning approach. China Ocean Eng. 30 (1), 19–32.

Juhl, J., Jensen, O.J., 1995. Features of berm breakwaters and practical experience. In: International Conference on Coastal and Port Engineering in Developing Countries. Rio de Janeiro, Brazil.

Koza, J.R., 1992. Genetic Programming: on the Programming of Computers by Means of Natural Selection, vol. 1. MIT press.

Lykke Andersen, T., 2006. Hydraulic Response of Rubble Mound Breakwaters: Scale Effects-Berm Breakwaters, vol. 27. Aalborg University report. *ISSN 0909-4296*, Dissertation.

Lykke Andersen, T., Burcharth, H.F., 2010. A new formula for front slope recession of berm breakwaters. Coast Eng. 57 (4), 359–374. https://doi.org/10.1016/j.coastaleng.2009.10.017.

MAST II, 1997. Berm Breakwater Structures, Final Report under MAST-CONTRACT MAS2-CT94-0087.

Moghim, M.N., Lykke Andersen, T., 2015. Armor stability of hardly (or partly) reshaping berm breakwaters. Coast Eng. 104, 1–12. https://doi.org/10.1016/j.coastaleng.2015.06.003.

Moghim, M.N., Shafieefar, M., Tørum, A., Chegini, V., 2011. A new formula for the sea state and structural parameters influencing the stability of homogeneous reshaping berm breakwaters. Coast Eng. 58 (8), 706–721. https://doi.org/10.1016/j.coastaleng.2011.03.006.

Motalebi, A., 2010. Experimental Study of Investigating the Influence of Armor Stone Size in Stability of Berm Breakwaters." Master Thesis. Tarbiat modares University.

Pontiki, M., 2019. Bayesian damage prediction of berm breakwaters in the arctic. In: Ports 2019: Port Planning and Development. American Society of Civil Engineers Reston, VA, pp. 12–21.

Research, C.I., Association, I., 2007. C. C. U. R. en R. (Netherlands), and C. d'etudes maritimes et fluviales, The rock manual: the use of rock in hydraulic engineering, vol. 683. Ciria.

Sadat Hosseini, A., Shafieefar, M., 2014. Estimation of reshaped profile of berm breakwaters using experimental data. Int. J. Marit. Technol. 2, 55–65.

Sadat Hosseini, A., Hajikarimi, P., Gandomi, M., Moghadas Nejad, F., Gandomi, A.H., 2021. Genetic programming to formulate viscoelastic behavior of modified asphalt binder. Construct. Build. Mater. 286, 122954 https://doi.org/10.1016/j.conbuildmat.2021.122954.

Salgotra, R., Gandomi, M., Gandomi, A.H., 2020. Time series analysis and forecast of the COVID-19 pandemic in India using genetic programming. Chaos, Solit. Fractals 138, 109945. https://doi.org/10.1016/j.chaos.2020.109945.

Searson, D.P., 2015. GPTIPS 2: an open-source software platform for symbolic data mining. In: Handbook of Genetic Programming Applications. Springer, pp. 551–573.

Shafieefar, M., Shekari, M.R., Hofland, B., 2020. Influence of toe berm geometry on stability of reshaping berm breakwaters. Coast Eng. 157, 103636.
Shekari, M.R., Shafieefar, M., 2013. An experimental study on the reshaping of berm breakwaters under irregular wave attacks. Appl. Ocean Res. 42, 16–23.
Sigurdarson, S., van der Meer, J., 2013. Front slope stability of the Icelandic-type berm breakwater. In: Coastal Structures 2011*: (In* 2 Volumes*)*. World Scientific, pp. 435–446.
Sigurdarson, S., van der Meer, J., 2014. Design of berm breakwaters: recession, overtopping and reflection. In: From Sea to Shore–Meeting the Challenges of the Sea: (Coasts, Marine Structures and Breakwaters 2013*)*, pp. 64–74.
Sigurdarson, S., van der Meer, J.W., Burcharth, H.F., Sørensen, J.D., 2009a. Optimum safety levels and design rules for the Icelandic-type berm breakwater. In: Coastal Structures 2007*: (In* 2 Volumes. World Scientific, pp. 53–64.
Sigurdarson, S., van der Meer, J.W., Tørum, A., Tomasicchio, R., 2009b. Berm recession of the Icelandic-type berm breakwater. In: Coastal Engineering 2008: *(In 5 Volumes)*. World Scientific, pp. 3311–3323.
Smith, G.N., 1986. Probability and Statistics in Civil Engineering, vol. 244. Collins professional and technical books.
Tørum, A., Krogh, S.R., 2000. Berm breakwaters. Stone quality. In: SINTEF Report No. STF22 A00207. SINTEF, Civil and Environmental Engineering.
Tørum, A., Kuhnen, F., Menze, A., 2003. On berm breakwaters. Stability, scour, overtopping. Coast. Eng. 49 (3), 209–238.
van der Meer, J.W., 1988. Rock Slopes and Gravel Beaches under Wave Attack.
van der Meer, J.W., Pilarczyk, K.W., 1984. Stability of rubble mound slopes under random wave attack. Coast Eng. 1985, 2620–2634.
van der Meer, J., Sigurdarson, S., 2017. Design and Construction of Berm Breakwaters. World Scientific.
van Gent, M.R.A., van den Boogaard, H.F.P., Pozueta, B., Medina, J.R., 2007. Neural network modelling of wave overtopping at coastal structures. Coast. Eng. 54 (8), 586–593.