Wave overtopping at berm breakwaters: Experimental study and development of prediction formula

Karthika Pillai^a, Amir Etemad-Shahidi^{b,c*}, Charles Lemckert^{d,e}

^aGriffith School of Engineering, Griffith University, QLD, 4222, Australia

E-mail: karthika.pillai@griffithuni.edu.au

^bGriffith School of Engineering, Griffith University, Gold Coast, 4222, Australia

^cSchool of Engineering, Edith Cowan University, Joondalup, 6027, Australia

E-mail: a.etemadshahidi@griffith.edu.au

^dFaculty of Arts and Design, University of Canberra, Bruce, ACT, 2603, Australia

^eGriffith School of Engineering, Griffith University, Gold Coast, 4222, Australia

E-mail: Charles.Lemckert@canberra.edu.au

ABSTRACT

Prediction of overtopping rate plays a key role in the design and safety assessment of breakwater structures. In this study, the mean wave overtopping rate at hardly and partly reshaping berm breakwaters were investigated by means of small-scale physical model tests. The test program was designed to create a database with a wider range of wave steepness, berm width, berm level and crest level for the hardly/partly reshaping conditions with an initial slope of 1:1.50. The new formula was developed covering the comprehensive test conditions (including existing tests and newly-collected ones) using scaling arguments and data mining techniques. Dimensionless governing parameters were used in the development of the formulae, and it was found that the new formula represented the influence of relative water depth on the wave overtopping rate better than the other prediction models. Accuracy [metrics](#) such as *RMSE* and *Bias* indicated that the developed formula outperformed the existing prediction models [within the range of significant overtopping rates](#). Moreover, the developed formula is simple and gives better insights into the physical significance of the variables in the overtopping process, and thus it is hoped it will be more appealing for engineering applications.

Keywords:

Wave overtopping, berm breakwater, laboratory experiments, M5` model tree, prediction formula.

1. Introduction

Berm breakwaters have been constructed since the early 1980s in order to protect harbours from rough seas. The presence of a porous berm above the still water level provides improved dissipation of wave energy compared to conventional rubble mound structures. They also offer considerable cost savings in contrast to conventional structures in areas of large water depth and high waves, as they use smaller stones in the construction of the primary armour [37]. Often berm breakwaters are constructed using split barges and excavators, and the use of large cranes become optional [46]. This opens up opportunities for cost savings and small contractors. Wave overtopping is a key factor in the design of berm breakwaters since berths in the harbours are usually located just behind the breakwater structures [34]. The overtopping rates must be less than the allowable rate under the design and operating conditions to ensure the safety of people and property on and behind the breakwater [20, 45, 46].

Overtopping is a relevant phenomenon in many ports worldwide and significant effort has been made in the prediction of it. The existing approaches used to estimate the overtopping rates at berm breakwaters can be broadly classified as empirical models and soft computing methods. The most widely used are the empirical formulae, wherein the dimensionless measured overtopping rates from physical model tests are correlated to the dimensionless structural and hydraulic parameters through regression analysis [13, 31]. The most popular empirical formulae used to estimate the overtopping rate at berm breakwaters are those of Van der Meer and Janssen [44], Lykke Andersen [25] and EurOtop [16]. More recently, Etemad-Shahidi and Jafari [13] proposed a prediction model based on the CLASH

[9] database. Moghim et al.[28] used the Lykke Andersen [25] data while Sigurdarson and Van der Meer [35, 46] used mostly their own collected data, as well as the Lykke Andersen [25] data, to develop new models. The Artificial Neural Network (ANN) method of prediction [18, 47, 48, 51] was developed as a versatile tool to estimate the overtopping rate at a wide variety of coastal structures. In general, it performs complex regression analysis in order to estimate the overtopping rate and is often more accurate than the empirical formulae. Often, these models do not result in explicit equations to be used by engineers. Moreover, engineers need to have some data mining knowledge for practical application of ANN models which may not always be the case [13]. The empirical formulae are, therefore, more transparent in representing the physics of the overtopping process and provide insight into the importance of variables in the overtopping process.

The reliability of predictors ultimately depends on the data used to develop them. Generally, an underestimation of the design overtopping rate can cause extensive damage to the port infrastructure and loss of life in the coastal region protected by the structure, whereas, an overestimation may lead to increase in the crest level (to reduce overtopping) and thus result in increased construction costs, loss of visual amenity and construction time. A detailed review of the most common prediction models applicable for berm breakwaters, the data used to develop them, and their differences in estimating the influence of variables in the predicted overtopping rate have been performed by Pillai et al. [31]. Their study revealed significant inconsistencies among the prediction models in estimating the influence of governing variables. Importantly, excessive wave overtopping is one of the most threatening causes of failure of these structures in the context of global climate change

[41]. These effects increase dramatically for structures built in shallow water [22]. Thus, a reliable prediction of wave overtopping rate is essential to determine the geometry of the structure and thus the cost of new constructions and possible retroffing of existing structures.

The aim of this study is to better understand the wave overtopping process at hardly/partly (HR/PR) reshaping berm breakwaters through physical model tests and to develop robust formula for the prediction of the wave overtopping rate. In order to achieve this goal, the existing measured overtopping rates from the physical model tests performed by Lykke Andersen [25], Bătăcui and Ciocan [1], Lykke Andersen et al. [26], Sigurdarson et al. [36] were collected and a database exclusively for berm breakwaters was developed. Most of these data represented fully reshaping (FR) conditions from the experiments of Lykke Andersen [25]. The data from the model tests conducted as a part of this study were added and the existing database was extended. In order to derive the prediction formulae, the M5' model tree [49] and the software package SPSS were used together with scaling arguments. Recently, these approaches were successfully used to develop accurate prediction models at vertical structures [15], smooth sloped structures [12] and rubble mound breakwaters [23]. The new formulae were trained on two-thirds of the data and validated using the remaining one-third. In addition, the performances of the new formulae were qualitatively and quantitatively compared with those of available prediction models.

The paper is outlined as follows. Subsequent to the introduction, Section 2 briefly reviews the available prediction formulae for estimating the wave overtopping rate at berm

breakwaters. Then, the selection of measured overtopping rates from existing small-scale physical model tests is detailed. Next, the necessity and particulars of the experimental investigation undertaken are discussed in Section 3. In Section 4, the performances of the available prediction models are evaluated. Model trees and M5' algorithms are described in Section 5. Section 6 presents the new formula for estimating the wave overtopping rate at berm breakwaters and discusses the performance in comparison with the existing prediction models. Finally, the summary and concluding remarks from the study are presented in Section 7.

2. State of art

2.1 Overtopping prediction models

In the empirical approach, the dimensionless overtopping rate (q_*) is expressed as a function of the relative crest freeboard (R_*) as:

$$q_* = a \exp\left[-(bR_*)^c\right] \quad (1)$$

or

$$q_* = a R_*^{-b} \quad (2)$$

where a , b and c are empirically fitted coefficients from physical model test data, and q_* and R_* are defined as:

$$q_* = \frac{q}{\sqrt{g H_{m0}^3}} \quad (3)$$

$$R_* = \frac{R_c}{H_{m0}} \quad (4)$$

where q is the mean overtopping discharge per meter of structure width, g is the acceleration due to gravity, H_{m0} is the estimate of significant wave height from spectral analysis and R_c is the crest freeboard which is the vertical distance from the Still Water Level (SWL) to the point where the overtopping is measured (Fig. 1).

Van der Meer and Janssen [44] proposed the following probabilistic formulae (hereafter VJ) to estimate the overtopping rate at bermed dikes which could be applicable for berm breakwaters.

$$\text{If } \xi_{op} < 2; \quad q_* \sqrt{\frac{s_{op}}{\tan \alpha}} = 0.06 \exp \left(-5.2 \frac{R_c}{H_{m0}} \frac{\sqrt{s_{op}}}{\tan \alpha} \frac{1}{\gamma_f \gamma_b \gamma_h \gamma_\beta} \right) \quad (5)$$

$$\text{If } \xi_{op} > 2; \quad q_* = 0.2 \exp \left(-2.6 \frac{R_c}{H_{m0}} \frac{1}{\gamma_f \gamma_b \gamma_h \gamma_\beta} \right) \quad (6)$$

where ξ_{op} is the surf similarity parameter based on structure slope ($\tan \alpha$) and wave steepness (s_{op}) are given as:

$$\xi_{op} = \frac{\tan \alpha}{\sqrt{s_{op}}} \quad (7)$$

$$s_{op} = \frac{H_{m0}}{L_{0p}} = \frac{2\pi H_{m0}}{g T_{0p}^2} \quad (8)$$

Here, L_{0p} is the deep water wave length, T_{0p} is the peak wave period of the spectra, γ_f , γ_b , γ_h and γ_β are the reduction factors for roughness, presence of berm, water depth and wave

obliquity, respectively. According to Van der Meer and Janssen [44], the wave overtopping rate depended mainly on the R_c of the coastal structure wherein, the presence of a berm further reduced the overtopping rate. The reduction factors are defined as:

$$\gamma_b = 1 - r_b(1 - r_{hb}) \quad (9)$$

$$r_b = \begin{cases} 1 - \frac{\tan \alpha_{eq}}{\tan \alpha} \\ \left(\frac{B/H_{m0}}{2 \cot \alpha + B/H_{m0}} \right) \end{cases} \quad \text{for horizontal berm} \quad (10)$$

$$r_{hb} = 0.5 \left(\frac{h_b}{H_{m0}} \right)^2; \quad 0 \leq r_{hb} \leq 1 \quad (11)$$

In addition, the reduction factors to account for shallow foreshore and oblique wave incidence were defined as:

$$\gamma_h = \begin{cases} 1 - 0.03 \left(4 - \frac{h}{H_{m0}} \right)^2 & \frac{h}{H_{m0}} < 4 \\ 1 & \frac{h}{H_{m0}} \geq 4 \end{cases} \quad (12)$$

$$\gamma_\beta = 1 - 0.0033 \beta^\circ \quad (13)$$

where B is the berm width, h_b is the water depth above the berm (negative if the berm is above SWL), h is the water depth at the toe of the structure, β is the mean wave direction, and α_{eq} is the equivalent slope of structure considering the berm (Fig. 2). Van der Meer and Janssen [44] suggested that a smooth impermeable berm had the maximum influence in reducing the wave overtopping when at the SWL and hence γ_b tends to be the minimum

value (0.6) when the berm is at SWL. Nevertheless, Van der Meer and Sigurdarson [46] asserted that dikes and berm breakwaters are different mainly in the design wave heights considered, former often being designed for lower wave activity (≤ 3.0 m). Also, in the case of dikes, horizontal overtopping flow can occur over the impervious crest whereas in berm breakwaters most of the flow may come as splash after hitting the permeable crest of rocks.

Lykke Andersen [25] formula (hereafter LA) was developed from physical model tests on rubble mound berm breakwaters. The estimated overtopping rate is given as:

$$q_* = 1.79 \times 10^{-5} \left(f_{H_0}^{1.34} + 9.22 \right) s_{0p}^{-2.52} \exp \left(-5.63 R_*^{0.92} - 0.61 G_*^{1.39} - 0.55 h_{b*}^{1.48} B_*^{1.39} \right) \quad (14)$$

where

$$G_* = \frac{G_c}{H_{m0}} \quad (15)$$

$$B_* = \frac{B}{H_{m0}} \quad (16)$$

$$h_{b*} = \begin{cases} \frac{3H_{m0} - h_b}{3H_{m0} + R_c}; & h_b < 3H_{m0} \\ 0; & h_b \geq 3H_{m0} \end{cases} \quad (17)$$

$$T_0 = \sqrt{\frac{g}{D_{n50}}} T_{0,1} \quad (18)$$

$$T_0^* = \frac{19.8 \exp \left(\frac{-7.08}{H_0} \right) s_{0m}^{-0.5} - 10.5}{0.05 H_0} \quad (19)$$

$$f_{H0} = \begin{cases} 19.8 \exp\left(\frac{-7.08}{H_0}\right) s_{0m}^{-0.5}; & T_0 \geq T_0^* \\ 0.05 H_0 T_0 + 10.5; & T_0 < T_0^* \end{cases} \quad (20)$$

where G^* , h_b^* and B^* are the dimensionless parameters representing crest width, berm level and berm width, respectively. G_c is the crest width, T_0 is the dimensionless wave period, D_{n50} is the equivalent cube length exceeded by 50% of a sample of armour by weight, $T_{0,l}$ is the mean wave period based on frequency domain analysis, T_0^* is the dimensionless wave period transition point, s_{0m} is the mean wave steepness, f_{H0} accounts for the influence of stability number and $H_0 T_0$ is the stability number including the mean wave period. Here, H_0 is defined as:

$$H_0 = \frac{H_{m0}}{\Delta D_{n50}} \quad (21)$$

Δ is the relative reduced mass density given as:

$$\Delta = \frac{\rho_s}{\rho_w} - 1 \quad (22)$$

where ρ_s and ρ_w are the mass densities of stone material and water, respectively. The formula was developed for initial slopes above ($\cot\alpha_d$) and below berm ($\cot\alpha_d$) (Fig. 1) of 1.25. For $\cot\alpha \neq 1.25$, the G^* and B^* to be used in Eq. (14) are as follows:

$$G^* = \frac{G_c + 0.5(R_c + h_b)(\cot(\alpha_u) - 1.25)}{H_{m0}} \quad (23)$$

$$B^* = \frac{B + 0.5(R_c + h_b)(\cot(\alpha_u) - 1.25) + 0.5(h - h_b)(\cot(\alpha_d) - 1.25)}{H_{m0}} \quad (24)$$

where α_u and α_d are the slopes above and below the berm, respectively. LA takes into account a number of variables and thus has a complex structure when compared to the other empirical formulae [31]. Van der Meer and Sigurdarson [46] noted that the complexity of the formula makes it difficult to physically justify the influence of governing parameters on the overtopping rate.

EurOtop [16] provided the formula (hereafter EO) given in Eq. 25 to estimate (probabilistic) the overtopping rate at berm breakwater, which was similar to the one suggested for bermed dike by Van der Meer [42]. A γ_f of 0.35 for non-reshaping and 0.40 for reshaping berm breakwaters were proposed, respectively.

$$q_* = \frac{0.067}{\sqrt{\tan \alpha}} \gamma_b \xi_{m-1,0} \exp \left(-4.75 \frac{R_c}{H_{m0}} \frac{1}{\xi_{m-1,0} \gamma_f \gamma_b \gamma_v \gamma_\beta} \right) \quad (25)$$

$$\text{with a maximum of } q_* = 0.2 \exp \left(-2.6 \frac{R_c}{H_{m0}} \frac{1}{\gamma_f \gamma_\beta} \right) \quad (26)$$

where $\xi_{m-1,0}$ is the surf similarity parameter based on $T_{m-1,0}$, the wave period based on frequency domain analysis and γ_v is the reduction factor for the vertical wall. Sigurdarson and Van der Meer [35] suggested that Eq. (26) is more appropriate since the berm breakwaters have steep slopes. As the formula (Eq. 26) is simple, it is widely used by design engineers.

Sigurdarson and Van der Meer [35] modified the EO (Eq. 26) using the γ_{BB} factor which incorporated the differences in overtopping due to reshaping and s_{0p} . The data used for developing this formula was obtained mostly from their project data as well as from Lykke Andersen [25]. The formula (hereafter SV12) is given as:

$$q_* = 0.2 \exp \left(-2.6 \frac{R_c}{H_{m0}} \frac{1}{\gamma_{BB} \gamma_\beta} \right) \quad (27)$$

where γ_{BB} differentiated the HR/PR ($H_0 < 2.5$) and FR ($H_0 > 2.5$) conditions as given below:

$$\gamma_{BB} = \begin{cases} 0.68 - 4.5 S_{0p} - 0.05 \frac{B}{H_{m0}} & \text{for HR and PR berm breakwater} \\ 0.70 - 9.0 S_{0p} & \text{for FR berm breakwater} \end{cases} \quad (28)$$

Eq. (27) was further modified to Eq. (29) (hereafter SV16) by Van der Meer and Sigurdarson [46] and used in EurOtop [17].

$$q_* = 0.09 \exp \left(- \left(1.5 \frac{R_c}{H_{m0}} \frac{1}{\gamma_{BB} \gamma_\beta} \right)^{1.3} \right) \quad (29)$$

Etemad-Shahidi and Jafari [13] used regression analysis to develop the following formula (hereafter EJ) using selected data from the CLASH database:

$$q_* = \exp \left(-0.97 \frac{R_c}{H_{m0} \gamma_f} - 0.4 \frac{G_c}{H_{m0}} - 1.04 \tan \alpha_{ave} - 4.975 \right) \quad (30)$$

where $\tan \alpha_{ave}$ is defined as:

$$\tan \alpha_{ave} = \frac{3H_{m0}}{((1.5H_{m0} - h_b) \cot \alpha_u + (1.5H_{m0} + h_b) \cot \alpha_d)} \quad (31)$$

The formula is simple and considers the influences of slope and berm through the $\tan \alpha_{ave}$.

In another approach, Moghim et al. [28] used the concept of wave momentum flux to derive an overtopping formula for reshaping berm breakwaters using the data from Lykke Andersen [25]. The formula is given as:

$$q_* = 3.8 \times 10^{-8} \left(\frac{G_c}{H_{m0}} \right)^{-2.85} \exp(0.45 H_0 \sqrt{T_0}) \left(\frac{R_0}{h} \right)^{-2.9} \quad (32)$$

where

$$R_0 = R_u - R_c = \sqrt{\left[\left(\frac{\tan \alpha}{0.21} h^2 \right) \left(\frac{M_f}{\rho_w g h^2} \right)_{\max} + \left(\frac{\tan \alpha (B - Rec) h_b}{5.25} \right) \right]} - R_c \quad (33)$$

$$\left(\frac{M_f}{\rho_w g h^2} \right)_{\max} = \frac{0.21 R_u^2}{\tan \alpha h^2} - 0.04 \frac{(B - Rec) h_b}{h^2} \quad (34)$$

Here, R_u represents the run-up from the SWL calculated based on the maximum wave momentum flux M_f , R_0 the run-up in excess of the R_c that can lead to overtopping and Rec is the recession of the berm. The accuracy of the formula was found to be similar to that of the LA. Since both recession and run-up values are required for estimating the overtopping rate, this formula is not considered for further analysis in this study. A summary of the empirical prediction models is given in Table 1.

2.2 Database

The majority of existing data (486 datasets) on the measured overtopping rate at berm breakwater are obtained from the physical model tests of Lykke Andersen [25]. Mass armoured berm breakwater cross sections with initial slopes of 1.25 were tested. A

significant part of the data represents FR conditions with $H_0 \geq 3.5$ [46]. Though the data represents a vast range of other tested parameters, the predominant s_{op} is approximately 0.035. In addition, the ranges of emerged h_b for HR/PR [46] cross sections are limited. Berms below the SWL are hardly constructed in the real world and a few of the HR/PR tests had the berm level more than one significant wave height beneath the SWL. According to Van der Meer and Sigurdarson [46], the database does not describe well the influence of slopes other than 1.25. Moreover, the tested cross sections were not restructured unless either the s_{op} or the initial geometry was varied. Van der Meer and Sigurdarson [46] observed that this can affect the measured overtopping rates, especially when lower wave conditions are tested on already reshaped berm breakwaters. Physical model test data was also obtained from Bătcăui and Ciocan [1] (23 datasets). HR/PR cross sections with initial slopes of 1:1.10 and 1:1.50 were tested in these experiments. In addition, the influence of a wider range of s_{op} ($0.01 \leq s_{op} \leq 0.04$) on the overtopping rate was investigated. Since berm breakwaters are usually constructed with emerged berms, experiments were conducted with emerged berm conditions. Recently, the EurOtop database was updated ([EurOtop 2016](#)) and the measured overtopping rate from research around the world was incorporated [17, 51]. Among these, 26 data records corresponding to the wave overtopping at rubble mound berm breakwaters [26, 36] were selected.

Low overtopping discharges measured during the model tests are less likely to be accurate due to the measurement errors in these tests (e.g. [12], [15], [46]). Hence, following Van Gent et al. [47] and Verhaeghe et al. [48], data with $q \geq 10^{-6} \text{ m}^3/\text{m/s}$ were selected for the current study. It is noteworthy that $q < 10^{-6} \text{ m}^3/\text{m/s}$ represents $q < 0.3 \text{ l/m/s}$

in the prototype scale. These low overtopping rates correspond only to the allowable rates for people at seawall/dike and are less significant for the design of rubble mound berm breakwaters [17]. A summary of the parameters in the existing database are given in Table 2. In general, the tested range of parameters in Bătăcui and Ciocan [1] data are well within that of Lykke Andersen [25]. Nevertheless, Bătăcui and Ciocan [1] data represent a marginally broader range of R_c and emerged h_b compared to that of Lykke Andersen [25]. The limits of tested R_c , emerged h_b and h for the EurOtop [17] database is greater than that of Lykke Andersen [25] and Bătăcui and Ciocan [1]. As seen from the table, the existing database contains a total of 535 tests.

The HR/PR structures are often constructed with a slope of 1:1.50 to reduce damage and recession [46]. However, data on the measured overtopping rate at HR/PR berm breakwaters with $cota_u$ and $cota_d = 1.50$ is limited in the existing database except for the few data from Bătăcui and Ciocan [1] and EurOtop [17] databases. Lately, Sigurdarson and Van der Meer [35] reported a clear dependency of the overtopping rate at berm breakwaters with s_{op} . A database with more data points on a wider range of s_{op} is required to understand the overtopping phenomena during low and high s_{op} conditions. In contrast, measured overtopping rates outside the s_{op} range of 0.030 – 0.035 are lesser. Hence, the available empirical formulae developed from these data may have some limitations in estimating the overtopping rate. In order to bridge this knowledge gap in the overtopping behaviour of HR/PR berm breakwaters, physical model tests were performed in this study as detailed in the following sections.

3. Laboratory experiments

The experiments were conducted in the wave flume (14 m long $\times$ 0.80m deep $\times$ 0.45m wide) in the hydraulics laboratory of Griffith University, Australia. The flume is installed with a piston wave maker at one end which consists of a rectangular steel plate driven by a software controlled electric motor. The wave maker is able to produce both regular and irregular waves and is equipped with active wave absorption system to minimize the wave reflection from the wave board. Irregular waves of JONSWAP spectrum with a peak enhancement factor (γ) of 3.3 were used for all the tests. The tests were carried out on rubble mound berm breakwater cross sections with $cot\alpha_u$ and $cot\alpha_d$ of 1.50 (Fig. 3). A permeable rock core with a D_{n50} of 0.021 m and $D_{n85}/D_{n15} = 2$ was used to construct the model. The primary armour consisted of two layers of rock with a D_{n50} of 0.032 m. A constant crest width of 0.20m was tested in all the experiments. Overtopping volumes were collected using a chute (0.2 m wide) placed at the rear end of the crest (leeward side). The chute emptied the water into the overtopping collection tank. The overtopping volume was determined by measuring the water level in the tank before and after each test.

Three capacitance wave gauges were placed near the toe for the measurement of free surface and to estimate wave parameters H_{m0} and T_p (Fig. 4). The gauges were placed at distances of $X_{l2} = 0.50$ m and $X_{l3} = 1.30$ m where X_{l2} and X_{l3} are the distances of the second and third gauges from the first, respectively. Mansard and Funke [27] method of three fixed probes were used to separate the incident and reflected waves using a Matlab code. A data logger was used to synchronously collect the wave gauge outputs (in Volts)

and these were recorded as water surface elevation (in meters) using the work module in DASYLab 10 software.

Experiments were performed by varying the SWL, h_b , B and s_{op} . Tests covered the range of s_{op} (0.01 – 0.06) by varying the H_{m0} and T_p . Then, the structure configuration was changed to a different h_b and the same range of s_{op} was tested. Later, the procedure was repeated for a different B and SWL. Wave overtopping rates were measured after approximately 1000 waves for each wave condition (H_{m0} and T_p). For the present study, the movements of the stones were visually monitored and the structure was rebuilt (if required) at the end of each test. The ranges of tested variables are summarized in Table 3. As seen from the table, broader ranges of R_c , B and h_b (for $cota = 1.50$) were achieved from the current experiments compared to the existing data (Table 2). Moreover, approximately 85% of the new data are outside of $0.030 \leq s_{op} \leq 0.035$ range.

4. Evaluation of the existing prediction models

First the performances of the available prediction models were compared using the existing and the new data. Scatter diagrams of the estimated ($q^{*estimated}$) and measured ($q^{*measured}$) overtopping rates are shown in Fig. 5. Fig. 5a shows that the VJ overestimates the $q^{*measured}$. This could be because the formula was basically developed for bermed dikes with impermeable slopes. Berm breakwaters usually have more porous and steeper slopes in the range $1.1 \leq cota \leq 1.5$ [45] when compared to dikes. Also, the influence of s_{op} on the overtopping rate was not considered when ζ_{op} was > 2 . Pillai et al. [31] identified that these

factors could result in overestimation of the $q^*_{measured}$. A similar observation was noticed for the conventional rubble mound breakwaters by Jafari and Etemad-Shahidi [23].

As can be seen from the Fig. 5b, the existing data are well predicted by the LA. However, the uncertainty in the estimation of the new data is higher than that of the existing data and it increases for relatively low values of $q^*_{measured}$ ($< 10^{-4}$). It needs to be mentioned that the Lykke Andersen [25] tests with low $q^*_{measured}$ values (for $q < 10^{-6}$ m³/m/s) are well predicted (not shown here) by the LA, probably because they were used in the development of the model. However, such low overtopping rates are not significant in the design of rubble mound breakwaters and are liable to scale effects and measurement errors (e.g. [46], [47]). In addition, the high values of $q^*_{measured}$ ($> 10^{-3}$) are somewhat overestimated by the LA.

Though the EO is simple, Fig. 5c demonstrates that considerable uncertainties exist with the wide scatter of $q^*_{estimated}$. Similar to the VJ, this could probably be due to the absence of s_{0p} in the formula or due to the assumed γ_b for reshaping and non-reshaping berm breakwaters. To a similar extent, the $q^*_{estimated}$ using the EO also provided a large scatter for rubble mound breakwaters [23]. Compared to the EO, there is less scatter observed in the $q^*_{estimated}$ using the SV12 (Fig. 5d) probably because the formula takes into account the influences of s_{0p} and reshaping through the improved reduction factor of γ_{BB} . However, considerable disagreement exists, especially for the relatively low $q^*_{measured}$ ($< 10^{-4}$). The change in the functional form of R_c/H_{m0} does not seem to improve the $q^*_{estimated}$ using the SV16 (Fig. 5f) as the underestimation is more than that of the SV12. Fig. 5e shows that the

lower $q^{*estimated}$ are somewhat overestimated by the EJ. Nevertheless, scatter is much less than those of the VJ, EO and SV.

$q^{*estimated}$ using the ANN [47, 48] and the ANN-17 [18, 51] show a predominant underestimation (Fig. 5g and Fig. 5h, respectively). Likewise, underestimation was observed for the conventional rubble mound breakwaters by Jafari and Etemad-Shahidi [23]. This is rather interesting since they generally provide more accurate estimation of overtopping rates compared to the empirical formulae. However, the ANN predictions in the range $10^{-6} \leq q^{*measured} \leq 10^{-5}$ are less reliable and indicative [46]. The uncertainty in the prediction could be due to the limited data on rubble mound berm breakwaters (82 test data from Lykke Andersen [25]) used to develop the ANN model. Though ANN-17 [18, 51] is a more robust tool (in estimating the extreme cases of overtopping rate) developed on a comprehensive database when compared to the ANN, Fig. 5h indicates considerable underestimations.

To sum up, the $q^{*estimated}$ by the available prediction models shows some disparities as exemplified by the scatter diagrams which increase for the HR/PR berm breakwaters with initial $\cot\alpha = 1.50$ (data from this study) and for extreme overtopping rates ($q^{*measured} \leq 10^{-4}$ or $q^{*measured} \geq 10^{-3}$). Low discharges are relevant for situations such as partially constructed breakwaters, lower cost defences [43] and during operational conditions, whereas, high overtopping rates are crucial for estimating the crest level of the berm breakwater and in turn affect the geometry and overall cost of construction. Thus, there exists an underlying necessity to develop more accurate formulae from a comprehensive database that includes

the HR/PR conditions better. The data mining technique using the model tree and SPSS were adopted for developing the new formulae as detailed in the next section.

5. Model tree

The availability of enhanced computing power has led to the use of machine learning techniques to identify the relationships between variables in an extensive database. Model trees (MT) are one such technique which can be used for linguistic as well as numeric variables [15]. Some of the areas in which the M5MT has been effectively used are flood forecasting [39, 40], modelling sediment yield [21], prediction of run-up and scour depth around piles due to waves [4, 11] and estimation of local scour around bridge piers [30]. In the field of coastal engineering, the M5MT method was used to predict the armour stability [10] and overtopping discharge [23] at rubble mound breakwaters. In addition, it was utilized for the wave height prediction in Lake Superior [14], to develop the wave spectrum at the Gulf of Maine: USA and Bay of Bengal: India [33], estimation of wind speed from wave measurements [8], to develop an improved overtopping prediction model for inclined coastal structures with smooth impermeable surface [12] and vertical structures [15].

MTs work on the principle that a complex problem can be solved by dividing it into a number of simpler sub problems and thereby solving them [23]. It is a variant of the decision tree where each branch node represents a choice between alternatives and each leaf node represents a classification or decision [14]. The primary difference between a regression tree and an MT is that, if, as a result of sorting, the leaf is associated with an output value, it is called a regression tree. While, if the leaf is associated with a piecewise

regression function of the input variables, it is called an MT [38]. The algorithm used to build the MT is called M5 [32] which was improved to M5' in order to deal with missing values [49]. The three main steps in the M5' algorithm includes: 1) building the tree 2) pruning and 3) smoothing. They are described in detail by many researchers such as Quinlan [32], Wang and Witten [49], Etemad-Shahidi and Mahjoobi [14], Jung et al. [24], Etemad-Shahidi et al. [15] and Bonakdar and Etemad-Shahidi [3] the excerpts of which are enumerated below.

The MT developed by M5' algorithm is similar to an inverted tree where the data is fed into the root which is at the top and the related prediction of the introduced data is obtained from leaves, which are at the bottom. The M5' uses a recursive algorithm that creates the tree by splitting the data (total) space (Fig. 6) into subsets. The splitting condition is chosen in such a way that the variability is reduced in the subsets all the way from the root to the leaves through the branches. The Standard Deviation Reduction (*SDR*) factor is used as the measure of the variability and is given as:

$$SDR = \frac{M}{|T|} sd(T) - \sum \frac{|T_i|}{|T|} \times sd(T_i) \quad (35)$$

where M is the number of tests without missing values of that variable, T is the set of tests that reach the node (before splitting), T_i is the set that results from splitting at the node for the chosen variable and sd is the standard deviation. The *SDR* is estimated for every sub-space obtained as a result of the splitting process and it is compared with the data space before splitting. The algorithm considers all the input parameters for splitting and chooses that

attribute which maximizes the expected error reduction. The splitting process is stopped when the *SDR* is less than 5%, or only a few data remain. A linear regression model is then developed for the data associated with each leaf.

The accuracy of the model developed can increase with the growth of the MT and this often results in a large tree with several leaves (linear models). However, there is a possibility of overfitting and hence, pruning of the tree becomes inevitable. During pruning, the average absolute difference between the predicted value and the actual output is estimated for each of the training data that reaches the node. Then the linear model is simplified by eliminating some terms to minimize the estimated error, multiplied by the factor $(n + v) / (n - v)$. Here, n is the number of training data points that reach the node and v is the number of input parameters in the model that represent the output value at that node. This multiplication is done to avoid underestimation while handling the new data. If the estimated error is lower than the expected one, the leaf node can be pruned. The final stage is smoothing or regularization which removes the discontinuities between the adjacent linear models (pruned leaves). Also, it is important to consider smoothing when the models are constructed from very few training data. The predicted value is computed at each subspace and it is smoothed along its path back to the root at each node. Finally, the MT produces a set of linear multi-variable equations.

6. Results and discussion (new formula)

To develop robust prediction formulae, the data from the present study (Table 3) and the existing data (Table 2) were utilized as a comprehensive database with a wider range of

tested parameters could result in a more generic prediction model. About two-thirds of the data were randomly selected for training and the remainder was used for validation of the model. The software package WEKA [19] was used to implement the M5`MT algorithm. Varying levels of influences were reported [31] for the structural parameters such as R_c , B , h_b , G_c , $\tan\alpha$ and hydraulic parameters such as H_{m0} , T_{0p} (or L_{0p}) on the overtopping rate which can be expressed as:

$$q = f\{R_c, G_c, B, h_b, h, \tan \alpha, H_{m0}, T_{0p} \text{ (or } L_{0p})\} \quad (36)$$

Dimensionless variables were used to develop the new formulae in order to eliminate the scale effects. The functional form of the new overtopping model was similar to the existing empirical formulae with q^* as the output and the exponential forms of the governing variables as inputs. Several models using different dimensionless input variables were developed. By examining the correlation coefficients, the less important parameters were eliminated and the model development was simplified. Finally, the model that outperformed others was of the form:

$$q = f\left\{\frac{R_c}{H_{m0}\xi_{0p}}, \frac{G_c}{L_{0p}}, \frac{h_b}{H_{m0}}, \frac{h}{L_{0p}}\right\} \quad (37)$$

The following formulae were developed for HR/PR ($H_0 < 2.5$) and FR ($H_0 > 2.5$) berm breakwaters [45], respectively:

$$q_* = \begin{cases} 3.11 \times 10^{-5} \left(\tanh \frac{G_c}{L_{0p}} \right)^{-1.31} \left(\tanh \frac{2\pi h}{L_{0p}} \right)^{0.76} \exp \left(-7.16 \frac{R_c}{H_{m0} \xi_{0p}} + 1.22 \frac{h_b}{H_{m0}} + 2.96 \tan \alpha_{eq} \right) & HR / PR \\ 3.78 \times 10^{-5} \left(\tanh \frac{G_c}{L_{0p}} \right)^{-1.79} \left(\tanh \frac{2\pi h}{L_{0p}} \right)^{0.98} \exp \left(-8.11 \frac{R_c}{H_{m0} \xi_{0p}} + 0.91 \frac{h_b}{H_{m0}} + 0.06 H_0 \sqrt{T_0} \right) & FR \end{cases} \quad (38)$$

where

$$\xi_{0p} = \frac{\tan \alpha_{eq}}{\sqrt{S_{0p}}} \quad (39)$$

$$\tan \alpha_{eq} = \frac{3 H_{m0}}{\left((1.5 H_{m0} + h_b) \cot \alpha_u + (1.5 H_{m0} - h_b) \cot \alpha_d + B \right)} \quad (40)$$

A disadvantage of the formulae (Eq. 38) developed by the model tree approach is the discontinuity between the individual formulae fitted for each data sub-space, even though it is reduced using the smoothing process [15]. The transition from HR/PR to FR should take place gradually and when $H_0 = 2.5$, the formulae should yield similar results. To overcome this issue and to develop a simpler model, other non-conventional functional forms were considered. Several formulae with various levels of complexity and accuracy were developed and finally a simple and accurate model was developed as follows:

$$q_* = 7.39 \times 10^{-3} \exp \left(-7.71 \frac{R_c}{H_{m0} \xi_{0p}} - 8.24 \left(\frac{G_c}{L_{0p}} \right)^{0.5} + 1.05 \frac{h_b}{H_{m0}} + 1.03 \tanh \frac{2\pi h}{L_{0p}} + 0.08 H_0 \sqrt{T_0} \right) \quad (41)$$

In Eq. (41), a positive coefficient indicates that the parameter has a direct relationship with the output (q_*). As such, the equation shows that by increasing the dimensionless crest

freeboard, the overtopping rate decreases. The use of $R_c/(H_{m0}\zeta_{0p})$ variable (rather than R_c/H_{m0}) was found to improve the model accuracy as it has been used for the overtopping prediction models of conventional rubble mound breakwaters [23] and dikes/revetments [44]. $\tan\alpha_{ave}$ takes into account the influences of B and h_b on the initial slope of the structure. The lower the $\tan\alpha_{ave}$, the lower the ζ_{0p} , the higher the $R_c/(H_{m0}\zeta_{0p})$ and thus the lower the estimated overtopping rate. A justification for this trend in the case of conventional rubble mound breakwaters was given by Jafari and Etemad-Shahidi [23], which supports our results. As per Eq. (41), a low s_{0p} leads to high ζ_{0p} , low $R_c/(H_{m0}\zeta_{0p})$, and thus a high wave overtopping. A similar increase in overtopping rate with higher s_{0p} was reported for berm breakwaters by Christensen et al. [7]. h_b/H_{m0} term is positive in the formulae indicating that an emerged berm ($-h_b/H_{m0}$) would produce less overtopping rate compared to a submerged one. This is not only in agreement with the physics of the overtopping process but also with the results of the sensitivity analysis conducted by Pillai et al. [31]. It should be noticed that the formula is applicable within the limits of h_b/H_{m0} shown in Table 4.

Eq. (41) indicates that as G_c increases, the measured overtopping rate decreases. This is justifiable since the overtopping is measured at the lee side of the crest and as the (rough) crest width increases, less overtopping water reaches the lee side [23]. Zanuttigh et al. [51] also observed that the use of wave length for scaling the crest width takes into account the wave reflection from the structure and hence the selection of parameter G_c/L_{0p} by the M5`MT over the conventional form of G_c/H_{m0} appears reasonable. The exponent in the term G_c/L_{0p} was obtained by minimizing the scatter. The functional form of $\tanh(2\pi h/L_{0p})$

represents the proper dependence of overtopping rate to the changes in h at deep water levels conforming to the findings of Cheon and Suh [5]. $H_0\sqrt{T_0}$ in the formula is the stability index [29] to compensate for the reshaped profiles, which logically should give increased overtopping when compared to non-reshaped profiles.

In summary, Eq. (41) uses a reasonable number of independent variables to estimate the overtopping rate at berm breakwaters. It takes into account all the governing variables and the influence of reshaping on the overtopping rate. It is noteworthy that the new formula considers the influence of water depth at the toe through a simple dimensionless parameter. Recent studies such as Isobe [22] and Chini and Stansby [6] showed that the water level plays a major role in the overtopping rate during the SLR due to climate change. The formula is valid within the range of parameters mentioned in Table 4. In the following sections, the performance of Eq. (41) is compared with the existing prediction models.

Fig. 7 shows the diagram of $q^{*measured}$ vs $q^{*estimated}$ using Eq. (41). As can be seen from the Fig. 7a, the data are more concentrated around the optimal line compared to those of the previous models (Fig. 5). The test data are also well predicted by the formulae (Fig. 7b), with less number of data with considerable overestimation ($q^{*estimated} > 10 \times q^{*measured}$) or underestimation ($q^{*estimated} < 10 \times q^{*measured}$). The scatter is much less for the higher $q^{*measured}$ ($> 10^{-3}$) which is important for the economic and safe design of the structure in extreme conditions. The prediction models were quantitatively compared using the accuracy metrics of *Bias*, the coefficient of determination (R^2), Root Mean Square Error (*RMSE*), and Discrepancy Ratio (*DR*); defined as follows:

$$Bias = \frac{1}{m} \sum_{i=1}^m (\log q_{*estimated} - \log q_{*measured}) \quad (42)$$

$$R^2 = \frac{\left(\sum_{i=1}^m (\log q_{*measured} - \overline{\log q_{*measured}}) (\log q_{*estimated} - \overline{\log q_{*estimated}}) \right)^2}{\sum_{i=1}^m (\log q_{*measured} - \overline{\log q_{*measured}})^2 \sum_{i=1}^n (\log q_{*estimated} - \overline{\log q_{*estimated}})^2} \quad (43)$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (\log q_{*estimated} - \log q_{*measured})^2} \quad (44)$$

$$DR = \frac{q_{*estimated}}{q_{*measured}} \quad (45)$$

where m is the number of measurements, $\overline{\log q_{*measured}}$ and $\overline{\log q_{*estimated}}$ are the averages of $\log q_{*measured}$ and $\log q_{*estimated}$, respectively.

As seen from the Table 5, most of the accuracy measures of Eq. (41) are better than those of the existing prediction models. Furthermore, the observations from the scatter diagrams are strengthened by these accuracy measures. For example, *Bias* values show that VJ mostly overestimates the $q_{*measured}$ while the ANN and ANN-17 underestimate which are in line with the explanation given earlier for Figs. 5a, 5g and 5h. Moreover, Eq. (41) has a negligible positive bias indicating a negligible amount of overestimation. R^2 which is a measure of the correlation between the data also shows that it is higher for the new formulae compared to other models. *RMSE* estimates also indicate that the Eq. (41) is more accurate than the available models. It should be mentioned that the *RMSE* of Eq. (41) for the Lykke Andersen [25] data ($q > 10^{-6} \text{ m}^3/\text{s/m}$) is similar to that of the LA. However, Eq. (41) better estimates the Bătăcui and Ciocan [1] and EurOtop [17, 51] data which represents the hardly/partly reshaping structures with initial slopes which were outside the

validated range of the LA (1:1.25). For the sake of brevity, the details are shown in the Appendix A.

Table 5 also gives information related to the DR of different models. A log DR equal to zero implies a perfect match between the $q^*_{measured}$ and the $q^*_{estimated}$. The distribution of the log DR estimations was further analyzed using histogram (Fig.8). It can be seen that the log DR for most of the models are close to zero. However, the VJ has positive skewness and approximately 70% of the data points are more than 10 times overestimated. While the ANN (ANN-17) has a distinct negative skewness and nearly 50 (55) % of the data points is more than 10-fold underestimated. The distributions of the EO, SV12 and SV16 are nearly symmetric indicating a wide scatter of the data. Almost 20% of the estimations using the EO and SV16 are more than 10 times underestimated which is similar to the results of rubble mound breakwater reported by Jafari and Etemad-Shahidi [23] (using EO). Fig. 8 shows that Eq. (41) has the maximum percentage of log DR in the range -0.50 to 0.50. Nevertheless, Table 5 shows that all the models except the VJ, ANN and ANN-17 have more than 90% of the estimations within the range of $-1 \leq \log DR \leq 1$.

The influence of the relative water depth on the overtopping rate has rarely been considered explicitly before. Fig. 9 plots the DR values of different formulae as a function of $\tanh(2\pi h/L_{op})$. A good model should have its DR independent of the input parameter. It can be seen that the performances of the VJ and EO are very sensitive to the value of relative water depth (Figs. 9a and 9c) and they overestimate (underestimate) deep (shallow) water conditions. The ANN and ANN-17 predictions indicate predominant underestimation

of the overtopping rate (Fig. 9d and 9h), that can be dangerous. Similar is the performance of the SV16 (Fig. 9g) but with more scatter of the results. The LA, SV12 and EJ (Figs. 9b, 9e and 9f) show some scatter while less scatter (and no trend) is observed when using developed formula (Fig. 9i). In other words, this figure clearly shows that the influence of the relative water depth is included in a better way in Eq. (41). The developed formula can equip the engineers to consider the influence of relative changes in the water depth as a result of SLR on the overtopping rate and in turn design a structure more resilient to its effects. A relatively new parameter that is important in the recession, which also influences the overtopping discharge, is the horizontal armour width (A_h) [46]. Fig. 10 illustrates the DR of different formulae vs A_h . It shows the scatter of the existing formulae while Fig.10i shows less scatter, indicating that this parameter is considered implicitly in the developed the formula (Eq.41).

Wave overtopping estimated by the empirical formulae has some intrinsic limitations due to the complex nature of the overtopping process and also due to the natural randomness of the variables in the model. These uncertainties in Eq. (41) are considered using a safety margin of one standard deviation following the guidelines of EurOtop [17], Etemad-Shahidi and Jafari [12] and Etemad-Shahidi et al. [15]. Based on the scatter of the result, the formulae for the deterministic design or safety assessment of berm breakwaters are given as:

$$q_* = 10^{0.35N} (7.39 \times 10^{-3}) \exp \left(-7.71 \frac{R_c}{H_{m0} \xi_{0p}} - 8.24 \left(\frac{G_c}{L_{0p}} \right)^{0.5} + 1.05 \frac{h_b}{H_{m0}} + 1.03 \tanh \frac{2\pi h}{L_{0p}} + 0.08 H_0 \sqrt{T_0} \right) \quad (46)$$

where N can be obtained for various exceedance levels (Table 6), assuming a normal distribution. For instance, when the acceptable risk is 5%, the value of N to be used is 1.65 and the improvement factors in Eq. (46) will be $7.39 \times 10^{-2.42}$.

It should be mentioned that M5`MT can be developed faster compared to the other soft computing methods such as ANN. The development of ANN requires the time-consuming optimization of the neural network topology (e.g. Etemad-Shahidi et al. [15]). In addition, in Eq. (41) physical significance of the variables is expressed in simple dimensionless forms that are easy to justify whereas the ANN remains a black-box in this regard. However, the shortcoming of the M5MT method is that they can only produce linear relationships (e.g. Etemad-Shahidi and Bonakdar [10]). This was overcome using SPSS to develop non-linear functional forms of the dependent variables. Thus, pre-processing and appropriate selection of the functional forms of the influencing variables have to be performed in order to obtain an optimum model (see Etemad-Shahidi et al. [15] for more details).

7. Summary and conclusions

Wave overtopping at hardly and partly reshaping berm breakwaters was investigated through small-scale experiments. Tests covered a broader range of wave steepness and berm level conditions when compared with the existing data. A comprehensive database was developed by combining the available and the newly-collected data, which was then used to evaluate the existing prediction models. Uncertainties, which were mostly due to the limitation of existing data, were observed when applying the existing models on the

wider range of parameters considered in this study. This confirmed that the new formula was required to cover the range of hydrodynamic conditions represented by the conducted experiments. In order to develop the new formula, the data was first divided into train and test data. The formula was developed using M5`model tree, SPSS and scaling arguments. Several models were developed using the traditional functional form of the overtopping rate as the output parameter and dimensionless influencing variables as inputs. Different functional forms of the input parameters were investigated and a final model that included all the governing parameters was obtained. The proposed formula was able to successfully justify the influences of the governing variables on the overtopping rates at berm breakwaters. Additionally, the influence of the relative water level on the overtopping rate was more skillfully represented in the new formula when compared with the other prediction models. The performance of the derived formula was compared using accuracy metrics with the results revealing that it outperformed the existing models for $q \geq 10^{-6} \text{ m}^3/\text{m/s}$. Also, the developed formula is physically sound and rather compact. Finally, the formula was modified to consider the acceptable risk level, thus making it an appropriate design tool for engineers. It is hoped that the developed formula will result in a more reliable design tool to estimate the overtopping performance of berm breakwaters. This is a very promising step in the direction of incorporating the effects of sea level rise into the functional design of berm breakwaters.

Acknowledgement

The University of Waikato is highly appreciated for providing the WEKA software. We are also grateful to Prof Thomas Lykke Andersen, Adrian Batacui and Radu Ciocan for

using their dataset. We extend our appreciation to Professor Barbara Zanuttigh for kindly providing the EurOtop [17] database. [The staff of the Hydraulics Laboratory at Griffith University](#) are acknowledged for the contribution to the physical model tests.

Appendix A - Accuracy metrics (*RMSE*, *Bias*) of the prediction models

Data [Reference] (number of data)	Lykke Andersen [25] (486)	Batacui and Ciocan [1] (23)	Current study (166)	All data (701)
Formula				
VJ	1.05, 0.91	0.55, -0.42	1.20, 1.05	1.06, 0.87
LA	0.33, -0.07	0.51, 0.28	0.61, 0.22	0.44, 0.04
EO	0.65, 0.16	1.40, -1.35	0.75, 0.28	0.71, 0.11
ANN	1.05, -0.92	1.13, -1.05	0.84, -0.61	0.99, -0.84
ANN-17	1.15, -1.11	0.86, -0.76	0.87, -0.69	1.06, -0.96
SV12	0.64, 0.00	0.65, -0.55	0.63, 0.15	0.64, 0.01
EJ	0.46, -0.06	1.04, -0.96	0.51, 0.20	0.50, -0.03
SV16	0.78, -0.04	0.67, -0.55	0.72, 0.12	0.76, -0.03
Eq. (41)	0.33, -0.03	0.36, -0.15	0.38, 0.11	0.36, 0.01

Table captions

Table 1. Summary of the wave overtopping prediction formulae for berm breakwaters.

Table 2. Ranges of parameters in the existing database.

Table 3. Ranges of parameters tested.

Table 4. Ranges of used dimensionless parameters.

Table 5. Accuracy measures of different formulae for test (all) data.

Table 6. N values for various levels of acceptable risk [50].

Table 1. Summary of the wave overtopping prediction formulae for berm breakwaters

Formula	Reference
$\text{If } \xi_{op} < 2 \quad q_* \sqrt{\frac{S_{0p}}{\tan \alpha}} = 0.06 \exp \left(-5.2 \frac{R_c}{H_{m0}} \frac{\sqrt{S_{0p}}}{\tan \alpha} \frac{1}{\gamma_f \gamma_b \gamma_h \gamma_\beta} \right)$	Van der Meer and Janssen [44]
$\text{If } \xi_{op} > 2 \quad q_* = 0.2 \exp \left(-2.6 \frac{R_c}{H_{m0}} \frac{1}{\gamma_f \gamma_b \gamma_h \gamma_\beta} \right)$	
$q_* = 1.79 \times 10^{-5} \left(f_{H0}^{1.34} + 9.22 \right) S_{0p}^{-2.52} \exp \left(-5.63 \left(\frac{R_c}{H_{m0}} \right)^{0.92} - 0.61 \left(\frac{G_c}{H_{m0}} \right)^{1.39} - 0.55 h_b^{1.48} B_*^{1.39} \right)$	Lykke Andersen [25]
$q_* = 0.2 \exp \left(-2.6 \frac{R_c}{H_{m0}} \frac{1}{\gamma_f \gamma_\beta} \right)$	EurOtop [16]
$q_* = 0.2 \exp \left(-2.6 \frac{R_c}{H_{m0}} \frac{1}{\gamma_{BB} \gamma_\beta} \right)$	Sigurdarson and Van der Meer [35]
$q_* = \exp \left(-0.97 \frac{R_c}{H_{m0} \gamma_f} - 0.4 \frac{G_c}{H_{m0}} - 1.04 \tan \alpha_{ave} - 4.975 \right)$	Etemad-Shahidi and Jafari [13]
$q_* = 0.09 \exp \left(- \left(1.5 \frac{R_c}{H_{m0}} \frac{1}{\gamma_{BB} \gamma_\beta} \right)^{1.3} \right)$	Van der Meer and Sigurdarson [46]

Table 2. Ranges of parameters in the existing database

Parameter (unit)	Range		
	Lykke Andersen [25]	Bătăcui and Ciocan [1]	EurOtop 2016 [17]
R_c (m)	0.60 – 0.17	0.19	0.12 – 0.26
B (m)	0.20 – 0.65	0.20	0.25
h_b (m)	-0.04 – 0.12	-0.07 – -0.04	-0.10 - -0.02
G_c (m)	0.09 – 0.38	0.20	0.07 – 0.13
h (m)	0.23 – 0.44	0.34	0.4 – 0.55
H_{m0} (m)	0.05 – 0.16	0.11 – 0.16	0.073 – 0.21
T_{0p} (s)	0.98 – 2.23	1.53 – 2.91	1.51 – 2.58
L_{0p} (m)	1.50 – 7.76	3.67 – 13.16	3.54 – 10.40
s_{0p}	0.01 – 0.05	0.01 – 0.04	0.01 – 0.04
ζ_{0p}	1.42 – 3.41	2.57 – 6.30	2.43 – 4.30
H_0	1.13 – 4.86	2.08 – 2.93	1.57 – 3.09
T_0	17.81 – 40.07	23.63 – 42.06	21.44 – 48.48
$H_0\sqrt{T_0}$	4.79 – 28.19	10.76 – 17.69	7.57 – 16.26
$q^{*measured}$	$6.58 \times 10^{-6} - 1.27 \times 10^{-3}$	$7.57 \times 10^{-5} - 3.12 \times 10^{-3}$	$1.32 \times 10^{-5} - 9.69 \times 10^{-4}$
Number of data	486	23	26
Total number of data	535		

Table 3. Ranges of parameters tested

Parameter (unit)	Range
R_c (m)	0.08 – 0.18
B (m)	0.25 – 0.45
h_b (m)	-0.08 – 0.05
G_c (m)	0.20
h (m)	0.38 – 0.48
H_{m0} (m)	0.07 – 0.16
T_{0p} (s)	1.17 – 1.98
L_{0p} (m)	2.13 – 6.11
s_{0p}	0.01 – 0.06
ξ_{0p}	1.55 – 3.33
H_0	1.16 – 2.63
T_0	18.61 – 31.50
$H_0\sqrt{T_0}$	5.04 – 13.22
$q^{*measured}$	$9.79 \times 10^{-6} - 1.19 \times 10^{-3}$
Total number of data	166

Table 4. Ranges of used dimensionless parameters

Parameters	Range	
	Train	Test
R_c/H_{m0}	0.63 – 2.02	0.96 – 1.79
s_{0p}	0.01 – 0.06	0.01 – 0.05
$\tan\alpha_{eq}$	0.27 – 0.66	0.31 – 0.53
ξ_{0p}	1.42 – 6.30	1.80 – 3.39
$R_c/H_{m0}\xi_{0p}$	0.20 – 0.88	0.33 – 0.71
G_c/L_{0p}	0.01 – 0.15	0.01 – 0.11
h_b/H_{m0}	-0.78 – 1.50	-0.70 – 0.45
h/L_{0p}	0.03 – 0.22	0.05 – 0.21
$\tanh 2\pi h/L_{0p}$	0.16 – 0.89	0.29 – 0.86
H_0	1.13 – 4.86	1.16 – 4.67
T_0	17.81 – 48.48	18.83 – 33.61
$H_0\sqrt{T_0}$	4.79 – 28.19	5.83 – 26.92
$q^*_{measured}$	$6.59 \times 10^{-6} - 3.12 \times 10^{-3}$	$9.11 \times 10^{-6} - 9.7 \times 10^{-4}$
Number of data	492	209

Table 5. Accuracy measures of different formulae for test (all) data

Accuracy indicator	VJ	LA	EO	ANN	ANN-17	SV12	EJ	SV16	Eq. (41)
<i>BIAS</i>	0.77 (0.87)	0.04 (0.04)	-0.06 (0.11)	-0.94 (-0.84)	-0.96 (-0.96)	0.01 (0.01)	-0.03 (-0.03)	-0.09 (-0.03)	0.05 (0.01)
<i>R</i> ²	0.12 (0.12)	0.36 (0.46)	0.11 (0.12)	0.18 (0.29)	0.29 (0.41)	0.17 (0.25)	0.12 (0.13)	0.17 (0.25)	0.49 (0.52)
<i>RMSE</i>	0.91 (1.06)	0.41 (0.44)	0.54 (0.71)	1.06 (0.99)	1.06 (1.06)	0.50 (0.64)	0.44 (0.50)	0.56 (0.76)	0.34 (0.36)
$-1 \leq \log(DR) \leq 1$ (%)	67 (56)	99 (97)	92 (83)	51 (59)	46 (47)	97 (92)	99 (97)	94 (87)	100 (99)

Table 6. N values for various levels of acceptable risk [50]

Acceptable risk (%)	N
2	2.05
5	1.65
10	1.28
33	0.44
50	0.00

Figure captions

Fig. 1. Schematic cross section of berm breakwaters [31].

Fig. 2. Equivalent slope a) considering the berm and b) without considering the berm [44].

Fig. 3. a) Schematic cross section and b) Photograph of the constructed model.

Fig. 4. a) Sketch and b) Photographs of the experiment set-up.

Fig. 5. Comparison between the $q^*_{measured}$ and the $q^*_{estimated}$ using the a) VJ b) LA c) EO d) SV12 e) EJ f) SV16 g) ANN and h) ANN-17.

Fig. 6. Graphical representation of the M5`MT algorithm [4].

Fig. 7. Comparison between the $q^*_{measured}$ and the $q^*_{estimated}$ using the Eq. (41) a) Existing and new data b) Test and train data.

Fig. 8. Histogram of $\log DR$ using the a) VJ b) LA c) EO d) ANN e) SV12 f) EJ g) SV16 h) ANN-17 and i) Eq. (41).

Fig. 9. Variations of DR as a function of the relative water depth using the a) VJ b) LA c) EO d) ANN e) SV12 f) EJ g) SV16 h) ANN-17 and i) Eq. (41).

Fig. 10. Variations of DR as a function of A_h using the a) VJ b) LA c) EO d) ANN e) SV12 f) EJ g) SV16 h) ANN-17 and i) Eq. (41).

Fig. 1. Schematic cross section of berm breakwaters [31]

Fig. 2. Equivalent slope a) considering the berm and b) without considering the [44]

All dimensions in meters

(a)

(b)

Fig. 3. a) Schematic cross section and b) Photograph of the constructed model

(b)

Fig. 4. a) Sketch and b) Photographs of the experiment set-up

(a)

(b)

(e)

(f)

Fig. 5. Comparison between the $q^*_{measured}$ and the $q^*_{estimated}$ using the a) VJ b) LA c) EO d) SV12 e) EJ f) SV16 g) ANN and h) ANN-17. (Upper dashed line represents 10 fold overestimation, lower dashed line 10 fold underestimation and the solid line a perfect agreement between $q^*_{estimated}$ and $q^*_{measured}$, respectively)

Fig. 6. Graphical representation of the M5`MT algorithm [4]

(a)

(b)

Fig. 7. Comparison between the $q^*_{measured}$ and the $q^*_{estimated}$ using the Eq. (41) a) Existing and new data b) Test and train data

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

(i)

Fig. 8. Histogram of $\log DR$ using the a) VJ b) LA c) EO d) ANN e) SV12 f) EJ g) SV16 h) ANN-17 and i) Eq. (41).

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

(i)

Fig. 9. Variations of DR as a function of the relative water depth using the a) VJ b) LA c) EO d) ANN e) SV12 f) EJ g) SV16 h) ANN-17 and i) Eq. (41) (Circles: new data; Triangles: existing data)

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

Fig. 10. Variations of DR as a function of A_h using the a) VJ b) LA c) EO d) ANN e) SV12 f) EJ g) SV16 h) ANN-17 and i) Eq. (41) (Circles: new data; Triangles: existing data)

Nomenclature

α	Structure slope angle (°).
α_{ave}	Average structure slope considering the berm (°).
α_{eq}	Equivalent slope of structure considering the berm (°).
α_d	Lower front slope (below berm) (°).
α_u	Upper front slope (above berm) (°).
β	Mean wave direction (°).
$\Delta = (\rho_s/\rho_w) - 1$	Relative reduced mass density.
γ	Peak enhancement factor.
γ_β	Reduction factor for wave direction.
γ_b, γ_{BB}	Reduction factors for the berm.
γ_f	Reduction factors for roughness or permeability.
γ_h	Reduction factor for shallow foreshore.
γ_v	Reduction factor for vertical wall.
ν	Number of input parameters in the model that represent the output value at a node.
ρ_s	Mass density of stone material (kg/m ³).
ρ_w	Mass density of water (kg/m ³).

$\xi = \frac{\tan \alpha}{\sqrt{s}}$	Surf similarity parameter.
ξ_{0p}	Surf similarity parameter based on T_{op} .
$\xi_{m-1,0}$	Surf similarity parameter based on $T_{m-1,0}$.
a, b, c	Empirical coefficients.
ANN	Artificial Neural Network by Van Gent et al. [47].
$ANN-17$	Artificial Neural Network by Formentin et al. [18] and Zanuttigh et al. [51]
B	Berm width (m).
B^*	Dimensionless berm width.
$D_{n,50} = \frac{W_{50}}{\rho_s}$	Equivalent cube length exceeded by 50% of a sample by weight (m).
DR	Discrepancy Ratio.
EJ	Etemad-Shahidi and Jafari [13] formula.
EO	EurOtop [16] formula.
f_{H0}	Factor accounting for the influence of stability number.
f_g	Gradation factor for rock material.
FR	Fully Reshaping.

$g = 9.82 \text{ m/s}^2$	Acceleration due to gravity.
G^*	Dimensionless armour crest width.
G_c	Width of the armour crest (m).
h	Water depth at toe of the structure (m).
h_b	Water depth above berm (negative if berm is above SWL) (m)
h_b^*	Dimensionless berm elevation.
H_0	Stability number.
$H_0 T_0$	Stability number including mean wave period.
H_{m0}	Significant wave height based on frequency domain analysis (m).
HR	Hardly Reshaping.
L	Wave length (m).
LA	Lykke Andersen [25] formula.
M	Number of tests without missing values of the variable.
m	Number of measurements.
M_f	Wave momentum flux.
MT	Model Tree.

N	Coefficient for various levels of acceptable risk.
n	Number of training data points that reach a node.
PR	Partly Reshaping.
q	Mean overtopping discharge per meter structure width (m ³ /m/s).
q^*	Dimensionless overtopping discharge.
$q^{*estimated}$	Estimated dimensionless overtopping discharge.
$q^{*measured}$	Measured dimensionless overtopping discharge.
r_b, r_{hb}	Reduction factors for berm (Van der Meer and Janssen, 1994).
R^*	Dimensionless crest freeboard.
R^2	Coefficient of determination.
R_0	Run-up in excess of crest freeboard (m).
R	Correlation Coefficient
R_u	Run-up from SWL (m).
R_c	Crest freeboard which is the vertical distance from SWL to the point where overtopping is measured (m).
Rec	Recession of the berm (m).

$RMSE$	Root Mean Square Error.
$s_{0m} = \frac{2\pi H_{m0}}{g T_{0m}^2}$	Deep water mean wave steepness.
$s_{0p} = \frac{2\pi H_{m0}}{g T_{0p}^2}$	Deep water peak wave steepness.
sd	Standard deviation.
SDR	Standard Deviation Reduction factor.
SLR	Sea Level Rise.
$SV12$	Sigurdarson and Van der Meer [35] formula.
$SV16$	Van der Meer and Sigurdarson [46] formula.
SWL	Still Water Level (m).
T	Tests that reach a node.
T_i	Set after splitting at the node.
$T_{0,1} = \frac{m_0}{m_1}$	Mean wave period based on frequency domain analysis (s).
T_0	Dimensionless wave period.
T_0^*	Dimensionless wave period transition point.
T_{0p}	Peak wave period (s).
T_i	Set after splitting the node.

$$T_{m-1,0} = \frac{m-1}{m_0}$$

Wave period based on frequency domain analysis (s).

VJ

Van der Meer and Janssen [44] formula.

W_{50}

Weight exceeded by 50% of a sample (kg).

X_{12}

Distance between first and second gauge (m).

X_{13}

Distance between first and third gauge (m).

References

- [1] A. Bătăcui, R. Ciocan, On berm breakwaters - An investigation into recession and overtopping, (Master thesis), Department of Civil Engineering, Aalborg University, Denmark, 2013.
- [2] B. Bhattacharya, R.K. Price, D.P. Solomatine, Machine learning approach to modeling sediment transport, *J. Hydrol. Eng.* 133(4) (2007) 440-450.
- [3] L. Bonakdar, A. Etemad-Shahidi, Predicting wave run-up on rubble-mound structures using M5 model tree, *Ocean Eng.* 38 (1) (2011) 111–118.
- [4] L. Bonakdar, H. Oumeraci, A. Etemad Shahidi, Run-up on vertical piles due to regular waves: Small-scale model tests and prediction formulae, *Coast. Eng.* 118 (2016) 1-11.
- [5] S.H. Cheon, K.D. Suh, Effect of sea level rise on nearshore significant waves and coastal structures, *Ocean Eng.* 114 (2016) 280-289.
- [6] N. Chini, P.K. Stansby, Extreme values of coastal wave overtopping accounting for climate change and sea level rise, *Coast. Eng.* 65 (2012) 27–37.
- [7] N.F. Christensen, M.S. Røge, J.B. Thomsen, T. Lykke Andersen, H.F. Burcharth, J. Q.H. Nørgaard, Overtopping on rubble mound breakwaters for low steepness waves in deep and depth limited conditions, in: *Proceedings of the 34th International Conference on Coastal Engineering*, Seoul, Korea, 2014, pp. 1–9.
- [8] M. Daga, M.C. Deo, Alternative data-driven methods to estimate wind from waves by inverse modeling, *Nat. Hazards.* 49(2) (2009) 293-310.
- [9] J. De Rouck, H. Verhaeghe, J. Geeraerts, Crest level assessment of coastal structures - General overview, *Coast. Eng.* 56 (2) (2009) 99–107.

- [10] A. Etemad-Shahidi, L. Bonakdar, Design of rubble-mound breakwaters using M5' machine learning method, *Appl. Ocean Res.* 31(3) (2009) 197-201.
- [11] A. Etemad-Shahidi, N. Ghaemi, Model tree approach for prediction of pile groups scour due to waves. *Ocean Eng.* 38(13) (2011) 1522-1527.
- [12] A. Etemad-Shahidi, E. Jafari, New formulae for prediction of wave overtopping at inclined structures with smooth impermeable surface, *Ocean Eng.* 84 (2014) 124–132.
- [13] A. Etemad-Shahidi, E. Jafari, A model for prediction of overtopping at berm breakwaters, in: *Proceedings of the 21st International Congress on Modelling and Simulation*, Goldcoast, Australia, 2015.
- [14] A. Etemad-Shahidi, J. Mahjoobi, Comparison between M5' model tree and neural networks for prediction of significant wave height in Lake Superior, *Ocean Eng.* 36(15-16) (2009) 1175-1181.
- [15] A. Etemad-Shahidi, S. Shaeri, E. Jafari, Prediction of wave overtopping at vertical structures, *Coast. Eng.* 109 (2016) 42–52.
- [16] EurOtop, Wave overtopping of sea defences and related structures: Assessment manual. N.W.H. Allsop, T. Pullen, T. Bruce, J.W. Van der Meer, H. Schüttrumpf and A.Kortenhaus, 2007. (www.overtopping-manual.com)
- [17] EurOtop, Manual on wave overtopping of sea defences and related structures: An overtopping manual largely based on European research, but for worldwide application, J.W. Van der Meer, N.W.H. Allsop, T. Bruce, J. De Rouck, A. Kortenhaus, T. Pullen, H. Schüttrumpf, P. Troch and B. Zanuttigh, 2016.

- [18] S.M. Formentin, B. Zanuttigh, J.W. Van Der Meer, A neural network tool for predicting wave reflection, overtopping and transmission. *Coast. Eng.* 59 (1) (2017).
- [19] E. Frank E, M.A. Hall, I.H. Witten, The WEKA Workbench. Online appendix for data mining: Practical machine learning tools and techniques, 4th Ed, Morgan Kaufmann, 2016.
- [20] Y. Goda, Derivation of unified wave overtopping formulas for seawalls with smooth, impermeable surfaces based on selected CLASH datasets, *Coast. Eng.* 56 (4) (2009) 385–399.
- [21] M.K. Goyal, Modeling of Sediment Yield Prediction Using M5 Model Tree Algorithm and Wavelet Regression, *Water Resour. Manag.* 28(7) (2014) 1991-2003.
- [22] M. Isobe, Impact of global warming on coastal structures in shallow water, *Ocean Eng.* 71 (2013) 51–57.
- [23] E. Jafari, A. Etemad-Shahidi, Derivation of a new model for prediction of wave overtopping at rubble mound structures, *J. Waterw. Port. Coast. Ocean Eng.* 138 (1) (2012) 42–52.
- [24] N.C. Jung, I. Popescu, P. Kelderman, D.P. Solomatine, R.K. Price, Application of model trees and other machine learning techniques for algal growth prediction in Yongdam reservoir, Republic of Korea, *J. Hydroinform.* 12(3) (2010) 262-274.
- [25] T. Lykke Andersen, Hydraulic Response of Rubble Mound Breakwaters (PhD thesis), Department of Civil Engineering, Aalborg University, Denmark, 2006.

- [26] T. Lykke Andersen, K.T. Skals, H.F. Burcharth, Comparison of homogenous and multi-layered berm breakwaters with respect to overtopping and front slope stability, in: Proceedings of the 31st International Conference on Coastal Engineering ICCE, Hamburg, 2008.
- [27] E.P.D. Mansard, E.R. Funke, The measurement of incident and reflected spectra using a least squares method, in: Proceedings of the 17th Coastal Engineering Conference Sydney, Australia 1980.
- [28] M.N. Moghim, R.F. Boroujeni, M.M.R. Tabari, Wave overtopping on reshaping berm breakwaters based on wave momentum flux, *Appl. Ocean Res.* 53 (2015) 23–30.
- [29] M.N. Moghim, M., Shafieefar, A. Torum, V. Chegini, A new formula for the sea state and structural parameters influencing the stability of homogeneous reshaping berm breakwaters. *Coast Eng.* 58 (2011) 706-721.
- [30] M. Pal, N.K. Singh, N.K. Tiwari, M5 model tree for pier scour prediction using field dataset. *KSCE J. Civ. Eng.* 16(6) (2012) 1079-1084.
- [31] K. Pillai, A. Etemad-Shahidi, C. Lemckert, Wave overtopping at berm breakwaters: Review and sensitivity analysis of prediction models. *Coast Eng.* 120 (2017) 1-21.
- [32] J.R. Quinlan, *Learning with continuous classes*. World Scientific, 1992.
- [33] S. Sakhare, M.C. Deo. Derivation of wave spectrum using data driven methods. *Mar. Struct.* 22(3) (2009) 594-609.
- [34] S. Sigurdarson, O.B. Smarason, G. Viggosson, Design considerations of berm breakwaters, in: Proceedings of the 27th International Conference on Coastal Engineering, ICCE 2000. Sydney, 2000.

- [35] S. Sigurdarson, J.W. Van der Meer, Wave overtopping at berm breakwaters in line with EurOtop, in: Proceedings of the 33rd International Conference on Coastal Engineering, ICCE 2012, Santander, Spain, 2012.
- [36] Sigurdarson, S. and Viggosson, G., 1994. Berm breakwaters in Iceland, practical experiences, in: Proceedings of the International Conference on Hydro-Technical Engineering for port and harbor construction, Japan, 1994
- [37] S. Sigurdarson, G. Viggosson, S. Benediktsson, O.B. Smarason, Berm breakwaters and quarry investigations in Iceland, in: Proceedings of the 4th International Conference on Coastal and Port Engineering in Developing Countries, Brazil, 1995.
- [38] D.P. Solomatine, K.N. Dulal, Model trees as an alternative to neural networks in rainfall-runoff modelling, *Hydrolog. Sci. J.* 48(3) (2003) 399-412.
- [39] D.P. Solomatine, M.B.L.A. Siek, Flexible and optimal M5 model trees with applications to flow predictions, in: Proceedings of the 6th International Conference on Hydroinformatics, 2004.
- [40] D.P. Solomatine, Y. Xue, M5 model trees and neural networks: Application to flood forecasting in the upper reach of the Huai river in China. *J. Hydrol. Eng.* 9(6) (2004) 491-501.
- [41] T.Q. Tuan, H. Oumeraci, A numerical model of wave overtopping on seadikes. *Coast. Eng.* 57(8) (2010) 757-772.
- [42] J.W. Van der Meer, Technical report on wave run up and wave overtopping at dikes (TAW), Technical Advisory Committee for Flood Defence in the Netherlands, Delft, 2002.

- [43] J.W. Van der Meer, T. Bruce, New physical insights and design formulae on wave overtopping at sloping and vertical structures, *J. Waterw. Port. Coast. Ocean Eng.* 140(6) (2014).
- [44] J.W. Van der Meer, J.P.F.M. Janssen, Wave run up and wave overtopping at dikes and revetments. Report No. 485, Delft Hydraulics, 1994.
- [45] J.W. Van der Meer, S. Sigurdarson, Geometrical design of berm breakwater, in: *Proceedings of the 34th International Conference on Coastal Engineering*, Seoul, Korea, 2014.
- [46] J.W. Van der Meer, S. Sigurdarson. *Design and construction of berm breakwaters*: World Scientific Publishing Co. Pte. Ltd; 2016.
- [47] M.R.A. Van Gent, H.F.P. Van den Boogaard, B. Pozueta, J.R. Medina, Neural network modelling of wave overtopping at coastal structures, *Coast. Eng.* 54 (8) (2007) 586–593.
- [48] H. Verhaeghe, J. De Rouck, J.W. Van der Meer, Combined classifier–quantifier model: a 2-phases neural model for prediction of wave overtopping at coastal structures, *Coast. Eng.* 55 (5) (2008) 357–374.
- [49] Y. Wang, I.H. Witten, Induction of model trees for predicting continuous classes, in: *Proceedings of the Poster Papers of the European Conference on Machine Learning*, Prague, 1997.
- [50] R. Yasa, A. Etemad-Shahidi, Classification and regression trees approach for predicting current-induced scour depth under pipelines, *J. Offshore Mech. Arct. Eng.* 136(1) (2014).

- [51] B. Zanuttigh, S.M. Formentina, J.W. Vandermeer, Prediction of extreme and tolerable wave overtopping discharges through an advanced neural network. *Ocean Eng.* 127 (2016).