New formulae for prediction of wave overtopping at inclined structures with smooth impermeable surface

Amir Etemad-Shahidi^{a,*} and Ebrahim Jafari^b

^{a,*} Griffith School of Engineering, Gold Coast Campus, Griffith University, QLD, 4227, Fax; +61 7 55528065, Email: a.etemadshahidi@griffith.edu.au, Corresponding author

^b School of Civil Engineering, Faculty of Engineering, University of Tehran, Email: ejafari@ut.ac.ir

Abstract

Reliable prediction of wave overtopping rate has an important role in the design and safety assessment of coastal structures. In this study, the selected data from the CLASH database were used to provide new formulas for the estimation of wave overtopping at simple sloped inclined seawalls or dikes with smooth impermeable surface. To develop the formulas, the decision tree approach together with the nonlinear regression provided by SPSS was used as a novel procedure. The conventional governing parameters were used as the input and output variables. The accuracy of the new formulas was then compared to those of previous ones and the measurements. It was shown that the presented formulas outperform the existing ones in the prediction of small and large scale overtopping rates at inclined impermeable coastal structures such as dikes and seawalls.

Key words: Wave overtopping; inclined impermeable seawalls; dikes; CLASH database; decision tree.

1. Introduction

The safety of the coastal regions against flooding largely depends on the performance of the coastal protection structures against wave attacks and storm surges. In the recent years, due to the effects of climate change, the intensity and duration of storms have increased (Verhaeghe et al., 2008). Therefore, safe design of coastal protections has become more vital. One of the most important hydraulic responses of the coastal structures is the wave overtopping rate. Reliable prediction of wave overtopping has a predominate role in determining the crest level of these structures. Excessive wave overtopping can cause failures in the structures, casualties and financial losses (Neves et al., 2008); especially when the coastal areas are protected by dikes or seawalls.

Several approaches have been proposed for the prediction of wave overtopping discharge on the inclined seawalls and dikes. The most commonly used models are the empirical formulas. In these formulas, the dimensionless overtopping rate is an exponential function of the relative crest freeboard. The summary of the previous empirical formulas for the prediction of wave overtopping rate at simple sloped smoothed structures is presented in Table 1 (see also Verhaeghe, 2005).

In these equations, q is the mean overtopping discharge per unit length of the structure, H_s is the significant wave height at the toe of structure, T_m is the mean wave period of incident waves at the toe, R_C is the crest freeboard of the structure, and h_t is the water depth on the toe of structure. The parameter of α is the slope angle of the structure, θ is the sea bed slope, and γ is the combination of different reduction factors for the effects of parameters such as surface roughness, oblique wave attack, berm, shallow water and wave walls (CEM, 2006). The parameter of S_o is the wave steepness defined as:

$$S_o = \frac{H_s}{L_o} = \frac{2\pi}{g} \frac{H_s}{T^2} \quad (1)$$

Where L_o is the wave length. S_{om} and S_{op} are derived by using the T_m and T_p in eq. 1, respectively. In the Van der Meer and Janssen (1995) and EurOtop (2007) formulas, the parameter of ξ_o is the surf similarity parameter (Iribarren number) defined as:

$$\xi_o = \frac{\tan \alpha}{\sqrt{S_o}} \quad (2)$$

These equations were obtained by fitting appropriate dimensionless numbers to the results of the laboratory model tests (Neves et al., 2008). Lately, increasing the power of computers has increased the popularity of numerical models. These models simulate the wave flow on the structure in a numerical flume. Basically, they solve governing differential equations of fluid flow in front and on the top of the structure (Verhaeghe, 2005). The main disadvantage of numerical models is the high computational cost required for an accurate simulation.

In the last decades, two artificial neural network (ANN) models have been proposed for the estimation of wave overtopping rate at all kinds of structures (Van Gent et al., 2007; Verhaeghe et al., 2008; Victor and Troch, 2012). In general, ANN models are more accurate than empirical models due to the complex regression that they make between inputs and output parameters (Kazeminezhad et al. 2010). The structure of ANN models, however, is not as transparent as that of empirical formulas. In addition, some trial and error process is required to find the optimum model (e.g. Zanganeh et al., 2009).

In summary, there are several models for the prediction of overtopping rate. However, some of them have complicated structures (such as Goda's formula, 2009, or Victor and Troch, 2012) and some of them overpredict the wave overtopping rate in some cases which cause financial losses. On the other hand, the underestimation of the overtopping rate may result in both economical and life losses due to flooding of the hinterland. In addition, the reliability of the existing formulas is not quite known. Due to the importance of dikes or

inclined seawalls (because in most cases they build to protect the populated areas in the coastal zones), in this study it has been tried to drive a simple and more accurate formula to be easily and safely used in the design of these kind of structures.

The aim of this study is to derive compact transparent and accurate formulas for the prediction of wave overtopping rate at simple sloped inclined seawalls. In order to develop the new prediction formulas, decision tree (Breiman et al., 1984) was used together with the nonlinear regression. Recently, decision tree has been successfully used in the field of ocean and coastal engineering (e.g.; Kambekar and Deo, 2010; Etemad-Shahidi et al., 2011; Jafari and Etemad-Shahidi, 2012). In this study, once the basic regimes are classified by the decision tree, nonlinear multi-variable regression is used to develop the formulas. The proposed formulas were trained and validated using the selected small scale data from the CLASH database (Van der Meer et al., 2009). Moreover, the performances of the formulas were evaluated for large scale data as an applicable example. The results were also compared to those of previous empirical formulae.

2. Description of the parameters affecting the overtopping rate

Overtopping discharge is mostly related to the wave run-up. When the rising water caused by incident waves reaches the crest of the structure and pass over it, overtopping occurs. In addition, fine droplets which are generated by wave breaking in front of structure and carried over the crest, are also considered as the wave overtopping (Verhaeghe, 2005). Several structural and hydraulic parameters affect the overtopping rate on simple sloped smoothed structures. One of the most important structural parameters is the crest freeboard (R_c). When the wave run-up height is higher than the crest freeboard height, the sea water will flow on the top of the crest and reach the lee side of the structure (Fig. 1). Besides, in general,

the number of waves that can rise over the crest is decreased by increasing the R_c . Hence, the overtopping discharge is reduced by increasing the crest height.

Significant wave height ($H_{m0,toe}$) has a high correlation with the overtopping discharge. The energy of incident waves is related to the square of the wave height. The results of previous studies have also shown that for a constant wave period, the power (energy) of incident waves increases by increasing the wave height (Moghim et al., 2011); and consequently, the wave overtopping discharge increases.

The mean period ($T_{m,toe}$) is also an important hydraulic parameter. The effect of the wave period is usually considered in the wave steepness parameter (eq. 1). By increasing the wave period, for a constant wave height, the wave length increases. According to the eq. 1, wave steepness reduces by increasing the wave length. The energy dissipation of a wave with a lower steepness is less than that of a steep wave (Basco, 1985). Hence, lower wave steepness results in a higher level of wave run-up, higher velocity of the rising water flow; and therefore, higher rate of wave overtopping. The laboratory measurements of these parameters at the toe of structure are used for the development of the new formulae.

3. Data extraction from the CLASH database

In the last decades, a number of studies have been performed to identify all aspects of the overtopping phenomena. As a result, several experimental data points are collected in different laboratories and sites all over the world. Most of these small scale data have been gathered together with the available large scale measurements within the CLASH project (De Rouck et al., 2009). An extensive database was then created based on the gathered data. It contains of more than 10500 overtopping data at various types of structures (Van der Meer et al., 2009).

In this study, data of overtopping at simple inclined coastal structures with smooth impermeable surface have been extracted from the CLASH database. The CLASH database has assigned a reliability factor RF and a structural complexity factor CF to each test. In this work, unreliable tests ($RF=4$) and the tests with very complex structure ($CF=4$) were excluded from the datasets (see appendix A for more details about the reliability of tests). Due to possible errors in the recording of zero overtopping rates; the tests with non-zero overtopping discharges were selected for further processing. In addition, the data of the structures with berms and with multiple slopes were excluded.

The extracted data include 882 small scale tests and 30 large scale measurements. Table 2 shows the characteristics of selected data for smoothed inclined seawalls. The identification of datasets in the CLASH database and the range of parameters in each dataset are also listed in the Table 2. 706 out of 882 data points have $\tan\theta = 0.001$ which represents a test on horizontal bed in most of the cases (Goda, 2009). According to Table 2, the minimum value for the relative toe depth ($h_t/H_{m0,toe}$) is 1.04. 228 tests (25.85% of all data) have $h_t/H_{m0,toe}$ less than 2 and other ones have $h_t/H_{m0,toe} > 2$. Hence, most of the wave characteristics at the toe of the structure were not recorded in the shallow water.

The amount of overtopping discharge significantly depends on the differences between wave run-up level and the crest freeboard level. In the coastal engineering manual (CEM, 2006), the value of 2.58 is given as a maximum value for $(R_u/H_{m0,toe})_{0.1\%}$. It means that if the relative crest freeboard ($R_c/H_{m0,toe}$) is greater than 2.58, only less than 0.1% of wave run-up can exceed the crest of structure. In the selected data 112 data have $R_c/H_{m0,toe} > 2.58$ and hence, the value of wave overtopping for these cases is very low. Based on the surf similarity parameters (ξ_{om}) and regarding to the values suggested by CEM (2006), 754 data have $\xi_{om} < 3.5$ and their type of wave breaking is plunging. The other remaining 128 data have $\xi_{om} > 3.5$ and break in the surging mode.

It should be mentioned that the large scale tests have been performed in large flumes in the laboratories (Verhaeghe, 2005). To avoid the scale and model effects (Verhaeghe et al., 2008), the large scale data points were excluded from the development processing of the new formulae due to their limited number compared to those of small scale tests.

4. Decision tree and nonlinear regression

The relationships between different variables in an extensive database can be discovered using the data mining techniques and statistical analysis. Through a variety of techniques, the decision tree is a new method which has been less applied to solve the engineering problems. The primary idea of this method is based on the theory that a complex problem can be solved by splitting it to smaller problems and then solving them (Bhattacharya et al., 2007, Mahjoobi and Etemad-Shahidi, 2008, Etemad-Shahidi and Ghaemi, 2011). This algorithm divides the input dataset to smaller subsets and after that, processes the subsets.

In Jafari and Etemad-Shahidi (2012), M5' model tree was used to develop a new model for over topping at rubble-mound structures. However in this study, Classification And Regression Tree (CART) method provided by Breiman et al. (1984) was used for data processing and then a non-linear regression was fitted to datasets using SPSS software. The basic idea of M5' model tree and CART is similar and the both method split the entire dataset to smaller subsets. M5' model uses a slightly complicated algorithm to divide the domain and presents only linear models. This shortcoming is overcome by using a combination of CART and non-linear regression of SPSS.

The CART algorithm is developed for analyzing the classification problems. In this method, first, the entire dataset is split to smaller subsets. The best splitting point is obtained by minimizing the variety of impurity or the diversity measured in each subset. The CART algorithm, in fact, tries to put the homogeneous data points in the same subset (Yasa and

Etemad-Shahidi, 2014). In the CART algorithm, for each split, the predictor is evaluated to find the best cut point based on an improvement score or reduction in impurity. For splitting rules and goodness of fit criteria, the Least Squared Deviation (LSD) impurity measure is used. $R(t)$ function is the weighted within node variance for the node t in the LSD measure, and it is equal to the re substitution risk estimation for the node t . $R(t)$ defines as:

$$R(t) = \frac{1}{\sum \omega_i f_i} \sum \omega_i f_i (y_i - \bar{y}(t))^2 \quad (3)$$

$$\bar{y}(t) = \frac{1}{\sum \omega_i f_i} \sum \omega_i f_i y_i \quad (4)$$

where ω_i is the weighting field value for the record i (if any), f_i is the frequency fieldvalue (if any), y_i is the target field value, and $\bar{y}(t)$ is the mean of the dependent variable at the node t . The LSD function for dividings at the node t is defined as:

$$Q(s, t) = R(t) - R(t_1) - R(t_2) \quad (5)$$

where $R(t_1)$ and $R(t_2)$ are the sum of squares of the first and second subsets. The split s is chosen to maximize the value of $Q(s, t)$ (Mahjoobi and Etemad-Shahidi, 2008; Ayoubloo et al., 2010). After the splitting process, a linear regression line is fitted to the data points in each subset. The splitting process repeats recursively until one of the stopping conditions occur (Breiman et al., 1984).

Sometimes the linear equations cannot describe the relationship between the variables properly. Nonlinear regression, unlike the traditional linear regression, can estimate an arbitrary nonlinear relationship between inputs (independent variables) and output (dependent variable). In this study, multivariate regression equations were fitted to the training set using SPSS software. In this case, an iterative estimation algorithm based on the methods proposed and implemented in NPSOL (Gill et al., 1998) was applied to estimate the regression

coefficients. This method was used to improve the linear regression equations made by CART algorithm to nonlinear ones.

5. Modeling and discussion on the results

The used overtopping database is composed of laboratory tests with different scales as well as field measurements (Verhaeghe et al., 2008). To eliminate the scale effects, dimensionless parameters were used for developing the new model. The input parameters include the relative crest freeboard ($R_c/H_{m0,toe}$), which represents the ratio between the crest level of structure and incident wave height at the toe of structure, and the surf similarity parameter (ξ_{om}). Various ranges of ξ_{om} refer to different types of wave breaking on the structure. Almost all of the phenomena in the surf zone are described by the surf similarity parameter (e.g. EurOtop, 2007; Bonakdar and Etemad-Shahidi, 2011; Etemad-Shahidi and Bonakdar, 2009). As shown in eq. 2, ξ_{om} is composed of the slope of structure and wave steepness. These two parameters (S_{om} and $\tan\alpha$) have been also used separately in some of the previous formulae. In addition to these parameters, other ones which seemed to be less effective such as the relative water depth at the toe of structure ($h_t/H_{m0,toe}$) and the sea bed slope ($\tan\theta$) were also considered as input parameters in the developing the new formulae.

The above mentioned dimensionless parameters include the most effective parameters on the wave overtopping rate which are described in the section 2. Relative overtopping discharge ($q/\sqrt{(g.H_{m0,toe}^3)}$, where q is the overtopping discharge and g is the gravity acceleration) was considered as the output parameter in the modeling. The formulae were developed based on the randomly selected 67 percent of small scale data and then tested by the remaining (33 percent) data points. Table 3 shows the range of the dimensionless parameters of the used train and test data. Different forms of functions using the governing parameters were tested. The one which outperformed other ones was as follows:

$$q^* = \frac{q}{\sqrt{g \cdot H_{m0,toe}^3}} = f\left(\frac{R_c}{H_{m0,toe}}, \xi_{om}, \frac{h_t}{H_{m0,toe}}, \tan \theta\right) \quad (6)$$

The sensitivity analysis on the effects of different input parameters and their combinations were accomplished. Results indicated that the effects of $h_t/H_{m0,toe}$ and $\tan \theta$ on the model performance can be neglected. Finally, two equations were obtained, which were separated by $R_c/H_{m0,toe}=1.62$. The models were optimized and the final formulae were:

$$\text{If } \frac{R_c}{H_{m0,toe}} \leq 1.62 \text{ then } \frac{q}{\sqrt{g \cdot H_{m0,toe}^3}} = 0.032 \cdot \exp\left(-2.6\left(\frac{R_c}{H_{m0,toe}}\right)^{1.6} (\xi_{om})^{-1.26}\right) \quad (7a)$$

$$\text{If } \frac{R_c}{H_{m0,toe}} > 1.62 \text{ then } \frac{q}{\sqrt{g \cdot H_{m0,toe}^3}} = 0.032 \times \exp\left(-5.63(\xi_{om})^{-1.26} - 3.283\left(\frac{R_c}{H_{m0,toe}} - 1.62\right)^{0.83}\right) \quad (7b)$$

Both formulae are physically sound and show that by increasing the crest freeboard, the overtopping discharge decreases. In addition, they show that by increasing the slope angle or the wave height or wave period, the overtopping discharge increases. The previous studies on sloped structures have shown that wave run-up has its maximum level for the various exceed levels (Delft Hydraulics, 1989). For example, the $(R_u/H_{m0,toe})_{5\%}$ exceeding level is equal to 1.68 (CEM, 2006). Victor and Troch (2012) also used $R_c/H_{m0,toe} < 1.7$ as the limit for low crested structures. As mentioned before, the amount of overtopping discharge remarkably depends on the differences between the crest freeboard level and the run-up level. When the $R_c/H_{m0,toe}$ is greater than 1.62, most of the run-up levels of incident waves are less than the crest freeboard level. Hence, the overtopping discharge and its dependence to the crest freeboard decrease. This fact can be concluded from the power of $R_c/H_{m0,toe}$ which is decreased in eq. 7b.

Scatter diagrams of the measured dimensionless overtopping discharges and the predicted ones, using different approaches, are shown in Fig 2. The inclined upper dashed line, solid line and lower dashed line are 10 times over estimated, perfect agreement, and 10 times under estimated lines, respectively. The test data (33% of randomly selected data) which were not used for the development of the formulae, are shown with gray diamonds in Fig 2.

Fig 2a shows that the predictions of the Owen (1982) formula are mostly overestimated. On the other hand, other formulae generally show underestimation in their predictions. Some of the previous models' predictions have a large scatter. For example, the predictions of the Owen (1982) formula in some cases are more than 100 times of the measured ones. As seen in the Fig 2, the EurOtop (2007) and Van der Meer and Janssen (1995) formulae have better performances compared to the others.

Scatter diagram of the measured and predicted overtopping rates by the new formulae is shown in Fig 3. It is clear from this figure that the predictions of the new formulae are more concentrated around the optimal line compared to those of the previous formulae. For quantitative evaluation of the formulae' accuracies, statistical indicators such as Bias, scatter index (SI), the coefficient of determination (R^2) and root mean square error ($RMSE$) were used. These parameters are defined as:

$$BIAS = \frac{1}{n} \sum_{i=1}^n (\log q_{est}^* - \log q_{meas}^*) \quad (8)$$

$$SI = \left(\frac{1}{|\bar{X}|} \sqrt{\frac{1}{n} \sum_{i=1}^n (\log q_{est}^* - \log q_{meas}^*)^2} \right) \times 100 \quad (9)$$

$$R^2 = \frac{\left(\sum_{i=1}^n (\log(q_{meas}^*) - \bar{X})(\log(q_{est}^*) - \bar{Y}) \right)^2}{\sum_{i=1}^n (\log(q_{meas}^*) - \bar{X})^2 \sum_{i=1}^n (\log(q_{est}^*) - \bar{Y})^2} \quad (10)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\log q_{meas}^* - \log q_{est}^*)^2} \quad (11)$$

Where q_{est}^* and q_{meas}^* are estimated and measured dimensionless overtopping discharge, n is the number of data, and $\bar{X}$ and $\bar{Y}$ are the average of $\log(q_{measured}^*)$ and $\log(q_{estimated}^*)$, respectively. For a quantitative judgment of models' performances, the error indicators were calculated by using the logarithm of overtopping discharges (Goda, 2009). The error measures of the previous formulae and the new ones are shown in the Table 4.

As seen, almost all the error measures of eq. 7 show the improvement in the predictions of the overtopping rate. Bias shows the average of the absolute error. Some of the Bias values are negative, indicating that their predictions are under estimated. Nevertheless, the average of absolute error for the new model is less than those of other ones. SI shows the relative scatter of the data points. As seen in Table 4, the predictions of the new formulae are less scattered than those of other ones. R^2 indicates the correlation of data. It can be seen from the table that the correlation of the new equations predictions is higher than those of Owen (1982) and Goda (2009) and close to those of Van der Meer and Janssen (1995) and EurOtop (2007). The values of $RMSE$ error also show that the developed formulae are more accurate than others.

Eq. 7 represents the average of overtopping discharge, and it can be modified for the probabilistic design or comparison with the measurements. For the deterministic design or safety, it is recommended to increase the average discharge by using the standard deviation of data. Coefficients of 2.6 in eq. 7a and 5.63 and 3.283 in eq. 7b are normally distributed with the standard deviation of 1.65, 3.09 and 1.5 respectively. Therefore, these equations can be modified for different levels of risk as follows:

$$\text{If } \frac{R_C}{H_{m0,toe}} \leq 1.62 \text{ then } \frac{q}{\sqrt{g \cdot H_{m0,toe}^3}} = 0.032 \times \exp \left((-2.6 + N \times 1.65) \left(\frac{R_C}{H_{m0,toe}} \right)^{1.6} (\xi_{om})^{-1.26} \right) \quad (12a)$$

$$\text{If } \frac{R_C}{H_{m0,toe}} > 1.62 \text{ then } \frac{q}{\sqrt{g \cdot H_{m0,toe}^3}} = 0.032 \times \exp \left((-5.63 + N \times 3.09) (\xi_{om})^{-1.26} + (-3.283 + N \times 1.5) \left(\frac{R_C}{H_{m0,toe}} - 1.62 \right)^{0.83} \right) \quad (12b)$$

The value of N can be determined for various exceedance levels by using the normal distribution curve (Table 5). As an example, the measured overtopping discharge and the predicted ones by eq. 7a and eq. 12a for 33% risk are shown in fig. 4 for $R_C/H_{m0,toe} \leq 1.62$.

Different kinds of reduction factors have been suggested previously to consider the effects of berm, surface roughness, wave walls or oblique wave attack. Since eq. 7a and eq. 7b were presented for simple slope inclined seawalls with smooth impermeable surface, the values of γ_f , γ_b and γ_v were considered to be equal to 1. However, the effect of oblique waves is considered here. The previously mentioned selected data from the CLASH database were tests with head on waves. 145 additional data points were then extracted from the CLASH database to investigate oblique waves. The range of these extracted data is shown in Table 6. In addition, Fig 5 shows the rose plot of oblique wave data in each direction. As seen, in most of the existing tests, the angle of wave attack is between 20 and 40 degrees. The effect of oblique wave attack is considered in eq. 7a as follows:

$$q^* = a \cdot \exp \left(b \left(\frac{R_C}{H_{m0,toe} \cdot \gamma_\beta} \right)^\alpha (\xi_{om})^\beta \right) \quad (13)$$

Hence, the value of γ_β was calculated as:

$$\gamma_{\beta} = \frac{\frac{R_c}{H_{m0,toe}}}{\left(\frac{\ln\left(\frac{q_{meas}^*}{a}\right)}{b(\xi_{om})^{\beta}} \right)^{\frac{1}{\alpha}}} \quad (14)$$

Several formulae have been suggested to consider the effect of oblique wave attack.

Van der Meer and Janssen (1995) proposed the following formulae for calculating γ_{β} :

$$\text{Short crested waves: } \gamma_{\beta} = 1 - 0.0033\beta \quad (15)$$

$$\text{Long crested waves: } \gamma_{\beta} = \begin{cases} 1 & 0 \leq \beta \leq 10 \\ \cos^2(\beta - 10) & 10 \leq \beta \leq 50 \\ 0.6 & \beta \geq 50 \end{cases} \quad (16)$$

Where β is the angle of wave attack. EurOtop (2007) recommended to use the eq. 15 with the minimum of $\gamma_{\beta}=0.736$. This reduction factor was considered by Goda (2009) as follows:

$$\gamma_{\beta} = 1 - 0.0096|\beta| + 0.000054\beta^2 \quad (17)$$

In all of these formulae, γ_{β} varies from 1 (for $\beta=0$) to a minimum of 0.6~0.7 for $\beta>70$. In this study, various types of equations were tested, and the best fitted equation was selected as follows:

$$\gamma_{\beta} = 1 - 0.33\sin(\beta) \quad (18)$$

Fig 6 shows the average values of measured γ_{β} as a function of the angle of wave attack. The predicted values of γ_{β} by other formulae are also shown in this figure. As seen, the predictions of eq. 18 are in good agreement with those of previous ones. In addition, the values of γ_{β} predicted by eq. 18 varies from 1 for $\beta=0$ to 0.67 for $\beta=90$, which are close to those predicted by others.

6. Application to large scale data

In this part, the developed formulae are applied to the large scale measurements. The predictions are compared to those of previous formulae and the measured overtopping rates. Scatter diagram of the measured overtopping discharges and the predicted ones are shown in Fig 7. As seen, the predictions of the new formulae are closer to the line of perfect agreement than those of previous formulae. Table 7 shows the error measures of different formulae. The *SI* values show that the predictions of the eq. 7, Owen (1982) and then EurOtop (2007) formulae are more accurate than other ones. The *Bias* values show that the predictions of all formulae are underestimated. However, except the Goda's formula, the average of absolute error of the developed formulae is less than those of others. R^2 values show that the estimated overtopping rate by the eq. 7 is highly correlated with the measured values. The calculated *RMSE* also indicates that the suggested formulae are the most accurate ones among all available formulae.

As it can be seen in Fig 7 and Table 7, generally the predictions of all formulae are underestimated. This underestimation could be caused by scale and model effects and/or measurement errors or some effects that are missing in the formulae. As mentioned in Verhaeghe et al. (2008), overtopping studies within the CLASH project have shown that overtopping measurements are affected by model and scale effects, resulting in differences between prototype and model response. On the other hand, it is quite difficult to measure overtopping in prototype and generally, measurement errors in small scale tests is less than those of large scale ones. As mentioned in the previous section, for the deterministic design or safety, it is recommended to increase the average discharge by using the eq 12a and 12b.

7. Summary and conclusions

In this study, two formulae are proposed for the prediction of wave overtopping rate at the simple slope inclined seawalls and dikes with smooth impermeable surface (for non-zero overtopping conditions). The new equations were developed using small scale data extracted from the CLASH database that covers a wide range for each effective parameter. Hence, derivation of a new formula based on this database leads to a formula which is applicable to a wider range of parameters rather than previous ones. Two third of data points were used for development of the equations and the remaining part was used for testing the model. In the modeling process, different combinations of dimensionless parameters have been tested and various models have been developed.

The formulation of the suggested model is quite compact and physically sound. Dimensionless crest freeboard and surf similarity parameter were used as the input parameters and the dimensionless overtopping discharge was used as the dependent variable. The performance of the new formulae for the prediction of wave overtopping rate was first compared to those of previous ones for small scale data and it was shown that their predictions outperform those of others. To consider different levels of acceptable risk, the new formulae were then modified to include the reliability. In addition, the effect of the oblique wave attack was also considered by a simple correction factor. Finally, the performances of different formulae were evaluated for large scale tests. It was shown that the proposed formulae are more accurate than the previous ones for large scale data sets as well as small scale ones.

Acknowledgements

The authors would like to acknowledge the organizers of EC-project CLASH for their comprehensive and freely available database on the wave overtopping phenomenon and University of Waikato for providing the WEKA software. The authors also thank Meysam Bali for his help in improving the manuscript.

Appendix A: a discussion on the reliability of used data

As mentioned in section 3, the CLASH database has assigned a reliability factor RF to each test. The reliability factor RF gives an indication of the reliability of the considered overtopping test. This reliability factor varies from $RF=1$ for very reliable test to $RF=4$ for unreliable test (Verhaeghe, 2005). As $RF=4$ means “No acceptable estimations/calculations could be made and/or measurements/analysis include faults, leading to an unreliable tests” (Verhaeghe et al., 2008), the test with $RF=4$ were excluded from the modeling process.

The tests with $RF=3$ refers to less reliable test (some estimations/calculations had to be made and/or some uncertainties about measurements/analysis exist). About 10% of all extracted data in this paper has a $RF=3$ (91 of 882 small scale data and 3 of 30 large scale data) and all of them have a $CF=1$. Table A shows the error measure of the results of different formulae by using the data with $RF=1&2$. The results using all used data ($RF=1&2&3$) are also shown in the parenthesis in the table A. As seen in the table, there is no significant difference between the results when excluding the data with $RF=3$. On the other hand, including them may increase the applicability of the formula, as other researchers such as Van Gent (2007) and Verhaeghe et al. (2008) (which were involved in the CLASH project) used these data for developing their ANN models. Therefore, in this study tests with $RF=1&2&3$ were used for analysis.

References

- Ayoubloo, MK, Etemad-Shahidi, A and Mahjoobi, J., 2010, Evaluation of regular wave scour around a circular pile using data mining approaches, *Applied Ocean Research*, 32, 34-39.
- Basco, D.R., 1985. A Qualitative Description of Wave Breaking. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, ASCE, 111 (2), 171-188.
- Bhattacharya, B., Price, R.K., Solomatine, D.P., 2007. Machine learning approach to modeling sediment transport. *Journal of Hydraulic Engineering*, ASCE, 133 (4), 440-450.
- Bonakdar, L., Etemad-Shahidi, A., 2011. Predicting wave run-up on rubble-mound structures using M5 model tree. *Ocean Engineering* 38 (1), 111-118.
- Breiman, L., Friedman, J., Olshen, R., Stone, C., 1984. *Classification and regression tree*. Belmont: Wadsworth Statistical Press.
- CEM, 2006. *Coastal Engineering Manual*. US Army Corps of Engineers, Washington, DC.
- De Rouck, J., Verhaeghe, H., Geeraerts, J., 2009. Crest level assessment of coastal structures - General overview. *Coastal Engineering* 56 (2), 99-107.
- DelftHydraulics, 1989. Slopes of loose materials: wave run-up on statistically stable rock slopes under wave attack, Report M 1983 (in Dutch), Delft Hydraulics Laboratory, The Netherlands.
- Etemad-Shahidi, A. and Bonakdar, L., 2009, Design of Rubble-Mound Breakwaters using M5 Machine Learning Method, *Applied Ocean Research*, 31, 197-201
- Etemad-Shahidi, A. and Ghaemi, N., 2011, Model tree approach for prediction of pile groups scour due to waves, *Ocean Engineering*, 38, 522-527
- Etemad-Shahidi, A., Yasa, R and Kazeminezhad, MK. , 2011, Prediction of wave-induced scour depth under submarine pipelines using machine learning approach, *Applied Ocean Research*, 33, 54-59
- EurOtop, 2007. *Wave Overtopping of Sea Defenses and Related Structures—Assessment Manual*. Gill, P.E., Murray, W., Saunders, M.A., Wright, M.H., 1998. User's Guide for NPSOL (Version 5.0): a Fortran package for nonlinear programming., Numerical Analysis Report 98-2, Department of Mathematics, University of California, San Diego.
- Goda, Y., 2009. Derivation of unified wave overtopping formulae for seawalls with smooth, impermeable surfaces based on selected CLASH datasets. *Coastal Engineering* 56 (4), 385-399.

- Jafari, E., Etemad-Shahidi, A., 2012. Derivation of a new model for prediction of wave overtopping at rubble-mound structures. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, ASCE, 138 (1), 42-52.
- Kambekar, A.R., Deo, M.C., 2010. Wave simulation and forecasting using wind time history and data-driven methods. *Ships and Offshore Structures* 5 (3), 253-266
- Kazeminezhad, M.H., Etemad-Shahidi, A. and Yeganeh-Bakhtiary, A. 2010. An alternative approach for investigation of the wave-induced scour around pipelines, *J. Hydroinformatics*, 12 (1), 51-65
- Mahjoobi, J., Etemad-Shahidi, A., 2008. An alternative approach for the prediction of significant wave heights based on classification and regression trees. *Applied Ocean Research* 30 (3), 172-177.
- 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. *Coastal Engineering* In Press.
- Neves, M.G., Reis, M.T., Losada, I.J., Hu, K.M., 2008. Wave overtopping of Povoá de Varzim breakwater: Physical and numerical simulations. *Journal of Waterway Port Coastal and Ocean Engineering*, ASCE 134 (4), 226-236.
- Owen, M.W., 1982. The Hydraulic Design of Seawall Profiles, *Proceedings of the Coastal Protection Conference*, Institution of Civil Engineers, Thomas Telford Publishing, London, UK, pp. 185-192.
- Van der Meer, J.W., Janssen, W., 1995. Wave Run-Up and Wave Overtopping at Dikes. In *Wave Forces on Inclined and Vertical Wall Structures*, Kobayashi and Demirbilek, eds., American Society of Civil Engineers, pp. 1-27.
- Van der Meer, J.W., Verhaeghe, H., Steendam, G.J., 2009. The new wave overtopping database for coastal structures. *Coastal Engineering* 56 (2), 108-120.
- 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. *Coastal Engineering* 54 (8), 586-593.
- Verhaeghe, H., 2005. Neural network prediction of wave overtopping at coastal structure, Doctorate Dissertation to Dept. civil Eng., Ghent University.
- Verhaeghe, H., De Rouck, J., van der Meer, J., 2008. Combined classifier-quantifier model: A 2-phases neural model for prediction of wave overtopping at coastal structures. *Coastal Engineering* 55 (5), 357-374.

- Victor, L and Troch, P., 2012. Wave overtopping at smooth impermeable steep slopes with low crested freeboards, *Journal of Waterway, Port, Coastal, and Ocean Engineering*, ASCE, 138 (5), 372-385.
- Yasa, R. and Etemad-Shahidi, A., 2014. Classification and regression trees approach for predicting current-induced scour depth under pipelines. *J. Offshore Mechanics and Arctic Engineering*, 136, 1-8.
- Zanganeh, M., Mousavi, J. and Etemad-Shahidi, A. 2009. A hybrid genetic algorithm-adaptive network based fuzzy inference system in prediction of wave parameters, *Engineering applications of artificial intelligence*. 22 (8), 1194-1202

Table 1. Previous empirical formulae for the prediction of wave overtopping rate at simple sloped smoothed structures

Formula	References
$\frac{q}{gH_s T_m} = a \exp\left(-b \frac{R_C}{H_s} \sqrt{\frac{S_{om}}{2\pi}} \frac{1}{\gamma_r}\right)$	Owen (1982)
<p>If $\xi_{op} < 2$ then $\frac{q}{\sqrt{g \cdot H_s^3}} \sqrt{\frac{S_{op}}{\tan \alpha}} = 0.06 \exp\left(-5.2 \frac{R_C}{H_s} \frac{\sqrt{S_{op}}}{\tan \alpha} \frac{1}{\gamma_r \cdot \gamma_b \cdot \gamma_h \cdot \gamma_\beta}\right)$</p> <p>If $\xi_{op} \geq 2$ then $\frac{q}{\sqrt{g \cdot H_s^3}} = 0.2 \exp\left(-2.6 \frac{R_C}{H_s} \frac{1}{\gamma_r \cdot \gamma_b \cdot \gamma_h \cdot \gamma_\beta}\right)$</p>	Van der Meer and Janssen (1995)
$\frac{q}{\sqrt{g \cdot H_s^3}} = \min \left\{ \frac{0.067}{\sqrt{\tan \alpha_s}} \cdot \gamma_b \cdot \xi_m \cdot \exp\left(-4.75 \frac{R_C}{\xi_m \cdot H_s \cdot \gamma_b \cdot \gamma_f \cdot \gamma_\beta \cdot \gamma_v}\right), 0.2 \exp\left(-2.6 \frac{R_C}{H_s \cdot \gamma_f \cdot \gamma_\beta}\right) \right\}$	EurOtop (2007)
$\frac{q}{\sqrt{g H_{s,toe}^3}} = \exp\left(-A - B \frac{R_C}{H_{s,toe}}\right)$	
$\begin{cases} A = A_0 \tanh\left[(0.956 + 4.44 \tan \theta) \times \left(\frac{h_t}{H_{s,toe}} + 1.242 - 2.032 \tan^{0.25} \theta\right)\right] \\ B = B_0 \tanh\left[(0.822 - 2.22 \tan \theta) \times \left(\frac{h_t}{H_{s,toe}} + 0.578 + 2.22 \tan \theta\right)\right] \end{cases}$	Goda (2009)
$\begin{cases} A_0 = 3.4 - 0.734 \cot \alpha + 0.239 \cot^2 \alpha - 0.0162 \cot^3 \alpha \\ B_0 = 2.3 - 0.5 \cot \alpha + 0.15 \cot^2 \alpha - 0.011 \cot^3 \alpha \end{cases} : 0 \leq \cot \alpha \leq 7$	

Table 2. The characteristics of the extracted data from the CLASH database for smoothed inclined seawalls. The last three rows of the table is refers to the large scale data.

Dataset ID	Number of data	$R_c / H_{m0, toe}$	ξ_{om}	$h_t / H_{m0, toe}$	Tan θ	$q^* = q / \sqrt{(g \cdot H_{m0, toe}^3)}$
DS-030	134	0.57~3.00	1.01~6.72	1.41~6.00	0.001~0.05	2.84E-5~3.79E-2
DS-035	17	0.99~2.34	2.03~3.61	2.20~5.49	0.019	4.57E-4~1.12E-2
DS-042	213	0.51~2.82	1.19~6.88	1.04~4.90	0.02~0.1	9.12E-5~4.57E-2
DS-101	29	0.50~1.47	0.66~2.35	3.50~10.3	0.001	2.00E-3~2.14E-2
DS-102	14	0.55~1.32	1.02~3.43	3.85~9.21	0.001	3.10E-3~2.85E-2
DS-103	12	0.51~0.93	1.41~4.28	3.55~6.48	0.001	1.19E-2~3.49E-2
DS-104	84	0.43~0.98	0.69~1.64	8.62~19.6	0.001	1.37E-3~4.06E-2
DS-109	16	1.19~1.48	0.83~1.06	1.39~1.73	0.034	2.51E-4~5.70E-3
DS-218	56	0.75~2.88	0.68~2.84	3.16~5.05	0.001	1.80E-5~4.25E-3
DS-220	20	1.09~2.92	1.00~3.59	4.78~6.55	0.001	5.21E-5~8.57E-3
DS-221	65	1.47~3.33	1.08~3.27	1.68~6.00	0.001~0.01	7.30E-6~1.26E-3
DS-222	31	0.97~2.75	1.02~3.81	4.68~6.86	0.001	1.69E-5~6.89E-3
DS-226	42	2.74~3.26	1.70~5.65	1.65~2.85	0.004~0.01	1.59E-5~3.81E-4
DS-227	67	0.83~3.50	1.14~4.54	1.49~5.43	0.001~0.01	5.13E-6~6.48E-3
DS-703	24	0.54~1.83	3.34~6.59	3.10~10.5	0.001	2.43E-3~4.48E-2
DS-955	16	1.08~3.22	2.40~3.69	2.40~3.41	0.017~0.03	9.39E-4~3.29E-2
DS-956	18	0.70~1.77	1.80~2.47	1.29~2.84	0.02	4.86E-3~3.47E-2
DS-959	24	0.79~3.44	0.84~3.37	3.18~7.35	0.025	5.98E-5~7.43E-3
Total small scale:	882	0.43~3.50	0.66~6.88	1.04~19.61	0.001~0.1	5.13E-6~4.57E-2
DS-110	17	0.98~3.51	0.82~2.44	3.86~7.09	0.001	4.35E-5~9.02E-3
DS-217	13	1.25~2.50	0.80~1.63	3.04~8.04	0.001	1.29E-5~4.48E-4
Total large scale:	30	0.98~3.51	0.80~2.44	3.04~8.04	0.001	1.29E-5~9.02E-3

Table 3. Characteristics of the used train and test small scale data

Parameters	Range	
	Train data	Test data
Number of data	588	294
$R_c/H_{m0,toe}$	0.44 ~ 3.50	0.43 ~ 3.43
ξ_{om}	0.66 ~ 6.88	0.69 ~ 6.83
$h_t/H_{m0,toe}$	1.04 ~ 19.61	1.04 ~ 19.23
$\tan \theta$	0.001 ~ 0.1	0.001 ~ 0.1
$q/\sqrt{(g.H_{m0,toe}^3)}$	5.1E-6 ~ 4.6E-2	7.9E-6 ~ 3.8E-2

Table 4. The error measures of different formulae for small scale test, all data

Error indicator	Owen (1982)	Van der Meer and Janssen (1995)	EurOtop (2007)	Goda (2009)	Eq. 7
<i>SI</i> (%)	26.09 (25.81)	17.37 (12.64)	17.36 (12.54)	19.73 (18.35)	13.86 (10.85)
<i>BIAS</i>	0.35 (0.43)	-0.22 (-0.10)	-0.20 (-0.07)	-0.08 (0.04)	-0.12 (0.0004)
R^2	0.45 (0.45)	0.67 (0.81)	0.67 (0.80)	0.60 (0.69)	0.73 (0.73)
<i>RMSE</i>	0.68 (0.69)	0.45 (0.34)	0.45 (0.34)	0.51 (0.49)	0.36 (0.29)

Table 5. Different values of N for various levels of acceptable risk

Acceptable risk (%)	N
0.1	3.09
2	2.05
5	1.65
10	1.28
33	0.44
50	0.00

Table 6. Range of the extracted data for inclined seawalls or dikes with oblique wave attack

Parameter	$R_c/H_{m0,toe}$	ξ_{om}	β	q^*
Range	1.06 ~ 2.63	0.97 ~ 6.51	5° ~ 80°	8.8E-6 ~ 1.0E-2

Table 7.The error measures of different formulae for large scale tests

Error measure	Owen (1982)	Van der Meer and Janssen (1995)	EurOtop (2007)	Goda (2009)	Eq. 7
<i>SI(%)</i>	12.58	15.08	13.19	19.34	10.71
<i>BIAS</i>	-0.33	-0.36	-0.39	-0.08	-0.26
<i>R²</i>	0.77	0.84	0.91	0.47	0.97
<i>RMSE</i>	0.44	0.53	0.46	0.68	0.38

Appendix A. The error measure of the results of different formulae for the tests with $RF=1&2$ (the results using the tests with $RF=1,2&3$ are shown in the parenthesis).

Error indicator	Owen (1982)	Van der Meer and Janssen (1995)	EurOtop (2007)	Goda (2009)	Eq. 7
<i>SI (%)</i>	25.34 (25.81)	12.19 (12.64)	12.03 (12.54)	17.87 (18.35)	10.50 (10.85)
<i>BIAS</i>	0.45 (0.43)	-0.09 (-0.10)	-0.06 (-0.07)	0.03 (0.04)	0.002 (0.0004)
<i>R²</i>	0.46 (0.45)	0.81 (0.81)	0.80 (0.80)	0.69 (0.69)	0.73 (0.73)
<i>RMSE</i>	0.68 (0.69)	0.33 (0.34)	0.32 (0.34)	0.48 (0.49)	0.28 (0.29)

Figure captions:

Fig 1. Wave run-up and overtopping in a low-crested structure (CEM, 2007)

Fig 2. Scatter diagrams of the measured dimensionless overtopping discharge and the predicted ones by using the formulae of (a) Owen (1982), (b) Van der Meer and Janssen (1995), (c) EurOtop (2007), (d) Goda (2009); train data (black diamonds) and test data (gray diamonds).

Fig 3. Scatter diagram of the measured and predicted overtopping rate by the new formulae; train data (black diamonds) and test data (gray diamonds).

Fig 4. The dimensionless measured overtopping discharge and the predicted ones by eq. 7a, average value (solid line), and 33% exceedance (dashed line)

Fig 5. Rose plot of oblique wave data.

Fig 6. Average values of γ_β for different angles of attack.

Fig 7. Scatter diagram of the measured overtopping discharge and the predicted ones by using the formulae of (a) Owen (1982), (b) Van der Meer and Janssen (1995), (c) EurOtop (2007), (d) Goda (2009), (e) New formulae; large scale data.

Fig 1. Wave run-up and overtopping in a low-crested structure (CEM, 2007)

Fig 2. Scatter diagrams of the measured dimensionless overtopping discharge and the predicted ones by using the formulae of (a) Owen (1982), (b) Van der Meer and Janssen (1995), (c) EurOtop (2007), (d) Goda (2009); train data (black diamonds) and test data (gray diamonds).

Fig 3. Scatter diagram of the measured and predicted overtopping rate by the new formulae;
train data (black diamonds) and test data (gray diamonds).

Fig 4. The dimensionless measured overtopping discharge and the predicted ones by eq. 7a, average value (solid line), and 33% exceedance (dashed line)

Fig 5. Rose plot of the used oblique wave data.

Fig 6. Average values of γ_β for different angles of attack.

Fig 7. Scatter diagram of the measured overtopping discharge and the predicted ones by using the formulae of (a) Owen (1982), (b) Van der Meer and Janssen (1995), (c) EurOtop (2007), (d) Goda (2009), (e) New formulae; large scale data.