

Article

Determination of Semi-Empirical Models for Mean Wave Overtopping Using an Evolutionary Polynomial Paradigm

Corrado Altomare ^{1,2,*} , Daniele B. Laucelli ³, Hajime Mase ^{4,5} and Xavi Gironella ¹

¹ Laboratori d'Enginyeria Marítima, Universitat Politècnica de Catalunya—BarcelonaTech, Calle Jordi Girona 1-3, 08034 Barcelona, Spain; xavi.gironella@upc.edu

² Department of Civil Engineering, Ghent University, Technologiepark 60, 9052 Gent, Belgium

³ Department of Civil Engineering and Architecture, Technical University of Bari, via Amendola, 126/B, 70125 Bari, Italy; d.laucelli@poliba.it

⁴ Disaster Prevention Research Institute, Kyoto University, Uji, Kyoto 611-0011, Japan; mase.hajime.q37@kyoto-u.jp

⁵ Hydro Technology Inst., Co., Ltd., Nakanoshima Daibiru 26F., Osaka 530-6126, Japan

* Correspondence: corrado.altomare@upc.edu

Received: 9 July 2020; Accepted: 27 July 2020; Published: 29 July 2020



Abstract: The present work employs the so-called Evolutionary Polynomial Regression (EPR) algorithm to build up a formula for the assessment of mean wave overtopping discharge for smooth sea dikes and vertical walls. EPR is a data-mining tool that combines and integrates numerical regression and genetic programming. This technique is here employed to dig into the relationship between the mean discharge and main hydraulic and structural parameters that characterize the problem under study. The parameters are chosen based on the existing and most used semi-empirical formulas for wave overtopping assessment. Besides the structural freeboard or local wave height, the unified models highlight the importance of local water depth and wave period in combination with foreshore slope and dike slope on the overtopping phenomena, which are combined in a unique parameter that is defined either as equivalent or imaginary slope. The obtained models aim to represent a trade-off between accuracy and parsimony. The final formula is simple but can be employed for a preliminary assessment of overtopping rates, covering the full range of dike slopes, from mild to vertical walls, and of water depths from the shoreline to deep water, including structures with emergent toes.

Keywords: Evolutionary Polynomial Regression; wave overtopping; sea dikes

1. Introduction

Wave overtopping occurs when sea waves run up coastal defenses, reach their crest, and flow over it. Coastal structures are usually built high enough to prevent significant flows that might damage infrastructures, properties, and assets and, even worse, injure people or cause casualties. Overtopping can be characterized in terms of individual discharges and volumes or flow properties, such as velocities and depth, of each wave that will eventually overtop the coastal structures. Volumes associated to extreme individual overtopping events are critical both for structural stability and the people safety, when standing on top or on the rear side of any coastal defense [1,2]. Generally most overtopping waves are rather small, but a few lead to significant volumes and high flow velocities. Individual overtopping volumes are usually analyzed in terms of probability distributions, e.g., [3–5]. Characterization of overtopping flow velocities and depths have been carried out in several works, such as [2,6–8]. Wave overtopping is a complex phenomenon, varying both in space and time, and

it is very stochastic [9,10]. Despite the importance of individual overtopping events on safety and structural stability, to consider an average discharge value is common practice worldwide for the design of coastal structures [11]. The crest level of any coastal structure is usually defined to limit the mean or average overtopping discharge below certain thresholds. The mean discharge is defined as the average discharge per meter width of the structure, q , expressed in $\text{m}^3/\text{s}/\text{m}$ or in $\text{l}/\text{s}/\text{m}$. The mean overtopping discharge depends on local wave conditions (wave height, wave period, water depth) but also on the layout and main features of the coastal defense, such as slope of the seaward face ([12]), slope of the foreshore before the defense ([13]), structural crest freeboard ([12,14]), presence of storm walls ([15]), presence of a berm ([16]), blocks or any artificial element on the surface of the structure that might increase its roughness to the wave run-up ([17]), etc. A sketch of the overtopping over a sea dike is represented in Figure 1. Both deep-water wave height and period ($H_{m0,o}$ and $T_{m-1,o}$) and local wave characteristics ($H_{m0,t}$ and $T_{m-1,t}$) are indicated. The latter ones correspond to wave height and period at the toe of the structure; h_t corresponds to the local water depth or depth at the toe of the structure. Meanwhile, R_c defines the crest freeboard, i.e., vertical distance between the still water level (SWL) and the crest of the defense. The spectral wave period is preferred to the peak period, since it was proved to show better performance for wave overtopping predictions for a different kind of wave spectra (e.g., single or double peak spectrum), as detailed in [18].

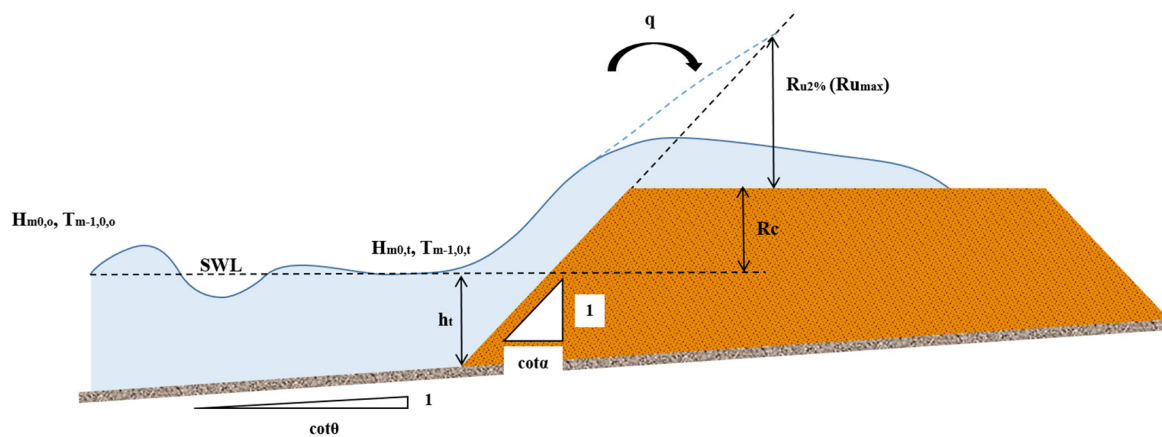


Figure 1. Sketch of wave overtopping on smooth sea dike.

Mean wave overtopping is usually assessed by means of semi-empirical models (i.e., formulas) for a preliminary design of any coastal defence. There are a large variety of formulas depending on the kind of structure and hydraulic boundary conditions. Following EurOtop [19], coastal structures can be categorized as dikes and embankment seawalls, armored rubble slopes and mounds, or vertical and steep walls. The present work will focus on smooth and impermeable sea dikes, as well as vertical and steep walls. Only simple geometries are studied aiming at characterizing the influence on the main structural characteristics, such as slope, freeboard, and toe depth on wave overtopping discharge. Therefore, complex layouts including storm walls, berms, blocks, and revetments are excluded.

For smooth sloping dikes and vertical walls, there are several formulas that differ for the structural layout (whether sea dike or vertical wall) and for different hydraulic boundary conditions (whether deep or intermediate waters, shallow waters, or impulsive conditions, see [19]). For sloping dikes, a set of formulas can be described for dikes with a slope between 1:7 and 1:2. For steeper slopes, up to vertical walls, the influence of the slope angle becomes more sensitive and is considered to fit new set of formulas [12,20,21]. In case of very or extremely shallow water conditions with mild foreshores before a sloping dike, [13,18,22] propose different models, in which overtopping discharge is proved to be function of the dimensionless freeboard, $R_c/H_{m0,t}$, and the surf similarity parameter, which is defined as the ratio between the structural slope and the square root of the wave steepness, $H_{m0,t}/L_{m-1,t}$, in which $L_{m-1,t}$ is the wavelength at the toe of the structure calculated using the deep-water formula $gT_{m-1,t}^2/2\pi$.

There is actually a gap in overtopping formulas between breaking/non-breaking waves and very shallow water conditions; see [19]. Besides, the distinction between breaking/non-breaking or impulsive/pulsating raises some concerns if we look just at an average value of the overtopping discharge, as already observed by [23], who derived a set of unified formulas for prediction of the mean rate of wave overtopping at coastal structures with smooth, impermeable surfaces. Nevertheless, the work of [23] lacked enough data in very and extremely shallow waters.

The present work aims at identifying possible models for the mean wave overtopping discharge of smooth impermeable coastal structures. The models will be derived by gathering several databases together and fitting the test results by means of a hybrid data-driven technique. The main goals are to (1) identify the main variables that determine the discharge and their correlations, going beyond the most used dimensionless freeboard and surf-similarity parameter, and (2) derive a final and unified model considered as a compromise between model parsimony and accuracy that could be employed for a preliminary assessment of mean overtopping discharge for any kind of smooth coastal defenses.

This paper is organized as follows: the state-of-art in mean wave overtopping assessment is summarized in Section 2; the scaling laws in wave overtopping are described in Section 3 together with a dimensional analysis to identify the main dimensionless groups or explanatory variable variables to be used in the analysis; Section 4 describes the experimental dataset that has been used for the analysis; in Section 5, a set of unified formulas for wave overtopping is presented; in Section 6, the new model expressions for mean overtopping assessment are discussed and a final formula is proposed; finally, the main conclusions are reported in Section 7.

2. Average Wave Overtopping Assessment in Existing Literature

Semi-empirical models are often used to assess average overtopping discharges. These models or formulas are fit on experimental databases. Hence, their range of application is restricted to the range of tested conditions and geometries of the database from which each formula is derived. This section offers a brief review of those semi-empirical formulas that might be applicable to the case of wave overtopping on smooth sea dikes and vertical walls.

2.1. Goda (2009)

Goda [23] proposed a set of unified formulas for the prediction of the mean rate of wave overtopping of smooth impermeable sea dikes. Both sloping and vertical structures were analyzed. The author employed 1254 data extracted from the database of the EU-funded CLASH project (Crest Level Assessment of coastal Structures by full scale monitoring, neural network prediction and Hazard analysis on permissible wave overtopping, contract EVK3-2001-00058).. Vertical walls with re-curved wave walls or broad crests, and the data with oblique wave incidence (angle of wave attack not being 0°), were excluded. Goda used also 198 tests carried out at Kansai University, Japan, by [24] that presented dike slopes of 1:3, 1:5, and 1:7 and foreshore slopes equal to 1:10 and 1:30. Since [24] did not provide wave conditions at the toe of the structure, the method proposed in [25] was used to calculate the wave height at the toe. In [23], the author showed that existing exponential formulas have a tendency to overestimate large overtopping rates and underestimate low overtopping rates when calibrated with the extracted dataset. He used the simple well-known exponential formula for wave overtopping:

$$\frac{q}{\sqrt{gH_{m0,t}^3}} = Q = \exp\left[-\left(A + B \frac{R_c}{H_{m0,t}}\right)\right] \tag{1}$$

where q is the mean overtopping discharge, R_c is the freeboard, $H_{m0,t}$ is the incident spectral wave height at the dike toe, and g is the gravity acceleration. The coefficients A and B are a function of the dike slope, the foreshore slope, and the dimensionless toe depth:

$$A = A_0 \tanh\left[(0.956 + 4.44 \tan \theta) \cdot \left(h_t / H_{m0,t} + 1.242 - 2.032 \tan^{0.25} \theta\right)\right] \tag{2}$$

$$B = B_0 \tanh[(0.822 - 2.22 \tan \theta) \cdot (h_t/H_{m0,t} + 0.578 + 2.22 \tan \theta)] \tag{3}$$

with:

$$A_0 = 3.4 - 0.734 \cot \alpha + 0.239 \cot^2 \alpha - 0.0162 \cot^3 \alpha \tag{4}$$

$$B_0 = 2.3 - 0.5 \cot \alpha + 0.15 \cot^2 \alpha - 0.011 \cot^3 \alpha \tag{5}$$

where θ is the foreshore slope angle with the horizontal, h_t is the still water depth at the dike toe, and α is the dike slope angle. The Equations (2)–(5) are valid in the range $0 \leq \cot \alpha \leq 7$. Both A and B coefficients increase up to a constant value if the relative toe depth increase. The datasets employed to fit Equations (2)–(5) are characterized by a dimensionless toe depth $h_t/H_{m0,t}$ bigger than 1.0, lacking data for very and extremely shallow water conditions.

2.2. Mase et al. (2013)

Mase et al. [22] proposed a set of formulae for wave run-up and overtopping discharge of sea walls in the presence of very shallow foreshores. The work includes cases of structures with an emergent toe ($h_t < 0$). An imaginary slope is defined between the sea bed at the breaking location and the point reached by the maximum run-up height; see Figure 2. The imaginary slope is employed to overcome the difficulties in schematizing complex dike and foreshore layouts. The formulae proposed by [22] are based on deep-water wave characteristics. Overtopping discharge is a function of the freeboard, the run-up, and the deep-water wave height.

$$\frac{q}{\sqrt{gH_{m0,o}^3}} = 0.018 \left(\frac{R_{max}}{H_{m0,o}} \right)^{1.5} \left(1 - \frac{R_c}{R_{max}} \right)^{6.24} \text{ for } 0 \leq R_c < Ru_{max} \tag{6}$$

with

$$\begin{aligned} Ru_{max} &= 1.54Ru_{2\%} \\ &= 1.54H_{m0,o} \left[2.99 - 2.73 \exp\left(-\frac{0.57 \tan \theta_{im}}{\sqrt{H_{m0,o}/L_{m-1,0,o}}}\right) \right] \end{aligned} \tag{7}$$

where $H_{m0,o}$ and $L_{m-1,0,o}$ are the deep-water wave height and wavelength, respectively, and θ_{im} refers to the imaginary slope as schematized in [22]. The use of deep-water conditions and calculation of wave run-up to assess the average discharge are key aspects of [22] that will be considered later on in this work.

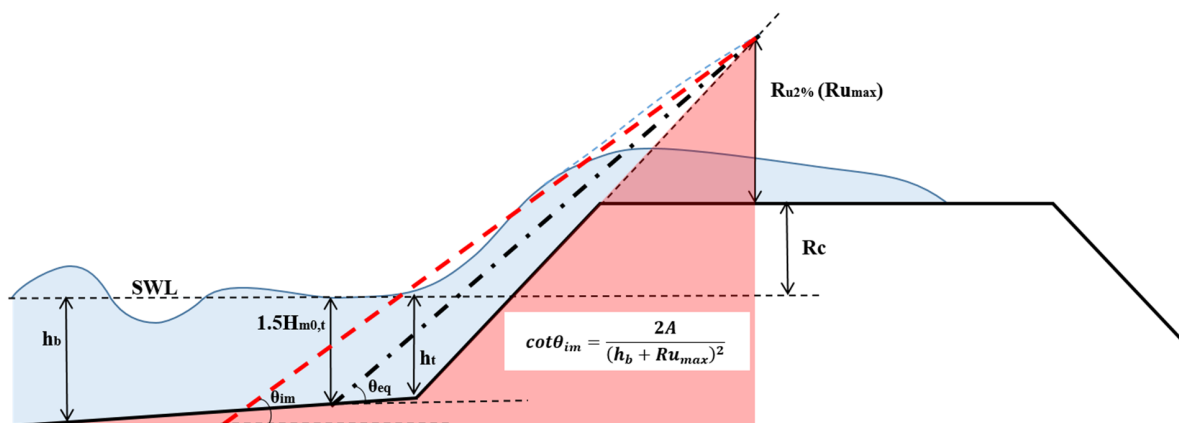


Figure 2. Scheme for definition and calculation of equivalent slope (Altomare et al., 2016) and imaginary slope (Mase et al., 2013). For the latter, the area A shadowed in red is employed.

2.3. EurOtop (2018)

EurOtop [19] proposes a set of formulas for the average overtopping discharge on sea dikes and embankment walls. In the manual, a distinction is made between gentle and steep dikes ($\cot \alpha < 2$), shallow foreshores, and breaking and non-breaking waves.

2.3.1. Breaking and Non-breaking Wave Conditions

The EurOtop formula for breaking and non-breaking waves and sea dikes with slope $\cot \alpha \geq 2$ can be expressed as follows:

$$\frac{q}{\sqrt{gH_{m0,t}^3}} = \frac{0.023}{\sqrt{\tan \alpha}} \xi_{m-1,0} \exp \left[- \left(2.7 \frac{R_c}{H_{m0,t}} \right)^{1.3} \right] \quad (8)$$

with a maximum of:

$$\frac{q}{\sqrt{gH_{m0,t}^3}} = 0.09 \exp \left[- \left(1.5 \frac{R_c}{H_{m0,t}} \right)^{1.3} \right] \quad (9)$$

The breaker parameter can be expressed as:

$$\xi_{m-1,0} = \frac{\tan \alpha}{\sqrt{s_{m-1,0}}} \quad (10)$$

being $s_{m-1,0} = H_{m0,t} / L_{m-1,0,t}$, where $L_{m-1,0,t}$ is the deep-water wavelength calculated based on the spectral period $T_{m-1,0,t}$. The influence factors presented in [19] are here omitted to take into account the effects of wave obliqueness, berm presence, and roughness, among others.

2.3.2. Very (shallow) Foreshores

The presence of very or extremely shallow foreshores induces heavy wave breaking far from the toe of the dike or seawall. The influence of the foreshore and local water depth at the dike toe are taken into account in [13], who introduced the concept of an equivalent slope in shallow water conditions. A similar formula as in [18] is used, but the surf-similarity parameter is calculated replacing the dike slope with the equivalent slope. The authors suggest using the equivalent slope when the ratio $h_t / H_{m0,t}$ is smaller than 1.5; otherwise, the dike slope only should be considered for further calculations. The new equivalent slope concept can be applied also to cases of dry toe (= negative water depth at the toe). The relative toe depth ranges between -0.25 m and 3.65 m. Mean wave overtopping can be assessed using the following equation:

$$\frac{q}{\sqrt{gH_{m0,t}^3}} = Q = 10^c \exp \left[- \frac{R_c}{H_{m0,t} (0.33 + 0.022 \xi_{m-1,0})} \right] \quad (11)$$

The c exponent in Equation (11) is considered normally distributed: the mean value of c is equal to -0.791 , and the standard deviation σ is 0.294 . The equivalent slope between the point on the foreshore with a depth of $1.5 H_{m0,t}$ and the run-up level $R_{U2\%}$ can be expressed as follows:

$$\cot \theta_{eq} = \frac{(1.5H_{m0,t} - h_t) \cdot \cot \theta + (h_t + R_{U2\%}) \cdot \cot \alpha}{(1.5H_{m0,t} + R_{U2\%})} \quad (12)$$

For the calculation, both structure and foreshore are assumed to have a straight slope without a berm, which are defined by $\cot \alpha$ and $\cot \theta$, respectively. Therefore, $\xi_{m-1,0}$ is assessed replacing the dike slope with the value of $\cot \theta_{eq}$.

2.4. Gallach (2018)

The transition from steep slopes and non-breaking waves to vertical walls was initially studied by [21] for relatively low freeboards. Gallach [12] extended the work of [21] and proposed the following formula based on three coefficients:

$$\frac{q}{\sqrt{gH_{m0,t}^3}} = a_{Gallach} \exp \left[- \left(b_{Gallach} \frac{R_c}{H_{m0,t}} \right)^{c_{Gallach}} \right] \tag{13}$$

- $a_{Gallach} = 0.0109 - 0.035(1.05 - \cot \alpha)$ and $a_{Gallach} = 0.109$ for $\cot \alpha > 1.5$;
- $b_{Gallach} = 2 + 0.56(1.5 - \cot \alpha)1.3$ and $b_{Gallach} = 2$ for $\cot \alpha > 1.5$; and
- $c_{Gallach} = 1.1$

The range of application of the formula is for slope angles $0 < \cot \alpha < 4$, the relative crest $0 < R_c/H_{m0} < 2.5$, and $0 < H_{m0,t}/h_t < 0.55$. Influence factors are omitted here.

3. Scaling Laws in Wave Overtopping

3.1. Volume Flux and Deficit in Freeboard

Ibrahim and Baldock [26] have revisited overtopping semi-empirical formulas by scaling them on the volume flux, $V_0 = H_{m0,o}L_p/2\pi$, of the incident waves. $H_{m0,o}$ is the offshore wave height and L_p is the deep-water wavelength calculated starting from the peak period T_p . The authors start from the assumption that scaling parameters for overtopping typically include the local wave height and period (or wavelength), the beach slope, and the crest freeboard. Usually, overtopping discharge is scaled on $\sqrt{(gH_{m0,t}^3)}$, which was employed based on the weir discharge equation and assuming critical flow conditions at the crest. However, the authors observed that except for cases of submerged structures, overtopping is usually characterized by subcritical flows for non-breaking wave conditions and supercritical flows in case of broken waves. Besides, the authors demonstrate that to use the deficit in freeboard, $1-R_c/Ru_{2\%}$, instead of the R_c/H_{m0} reduces substantially the scatter of data around the best fit. The deficit in freeboard has been already used in [22]; see Equation (6).

Ibrahim and Baldock [26] conclude that breaker type is not a major factor in the scaling of the overtopping for non-breaking waves; however, this parameter is employed to assess wave run-up for breaking waves or shallow foreshores, having a secondary effect on the overtopping scaling.

Therefore, to apply the method proposed by [26] to scale average discharges is required to know the deep-water wave conditions to obtain the volume flux and to choose an appropriate run-up model to assess $R_{u2\%}$.

Starting from the scaling of the overtopping volume, we can derive the scaling of the mean discharge. The overtopping volume per wave can be defined as $qT_{m-1,0}$, in which $T_{m-1,0}$ is the mean spectral wave period. By replacing L_p with $L_{m-1,0}$ ($=gT_{m-1,0}^2/2\pi$), the scaling for the mean discharge can be finally expressed as follows:

$$\frac{V_0}{T_{m-1,0,o}} = \frac{H_{m0,o}L_{m-1,0,o}}{2\pi T_{m-1,0,o}} = \frac{H_{m0,o}}{2\pi T_{m-1,0,o}} \frac{gT_{m-1,0,o}^2}{2\pi} = \frac{gH_{m0,o}T_{m-1,0,o}}{4\pi^2} \tag{14}$$

In the following analysis, we will employ both deep-water and local wave characteristics at the dike toe, the latter ones being normally employed for overtopping characterization, except in [22].

3.2. Dimensional Analysis

Correlation between mean overtopping discharge and hydraulic and structural parameters is investigated in terms of dimensionless variables. Basing on the Buckingham's π theorem, if there are n

variables in a problem containing k primary dimensions, the equation relating all the variables will have $n-m$ dimensionless groups. In mathematical terms, it is possible to write:

$$f(H_{m0,o}, T_{m-1,0,o}, H_{m0,t}, T_{m-1,0,t}, h_t, q, R_c, g, \alpha, \theta) = 0 \tag{15}$$

Hence, $n = 10$ and $k = 2$, leading to 8 dimensionless parameters. The choice of dimensionless groups is based on recent literature on overtopping:

- (1) $Q^* = 4\pi^2 q/gH_{m0,o}T_{m-1,0,o}$, based on [26], see Equation (14): deep-water wave characteristics are employed, after having verified that there is almost no difference ($R^2 = 99\%$) between the volume flux calculated employing the deep-water wave height and period with the one calculated based on local wave conditions at the toe.
- (2) Dimensionless freeboard, $R_c/H_{m0,t}$.
- (3) In [22] and [26], the overtopping is proved to be a function of the deficit of the freeboard, namely $1-R_c/Ru_{max}$, rather than the dimensionless freeboard. Notice that we use here the maximum run-up according to [22]. Considering the maximum run-up, the effects of wave periods, foreshore, and dike slope on overtopping discharge are included, as also in [13].
- (4) The slope parameter proposed in [27], $\beta = \theta T_{m-1,0,o} \sqrt{gH_{m0,o}}$, is employed to consider the influence of long waves generated by the presence of mild slopes on wave overtopping.
- (5) Relative toe depth employed as shallowness criteria, $h_t/H_{m0,o}$, is employed as proposed by [27].
- (6) The ratio between the local water depth, the dike toe, and the local wave height, $h_t/H_{m0,t}$, is considered as it was proven to have an influence on overtopping discharge [1], see Equations (2)–(3).
- (7) The combined effect of the foreshore slope and dike slope is taken into account by considering the equivalent slope defined in [13] and expressed as $\cot\theta_{eq}$. As an alternative, the imaginary slope described in [22] and expressed as $\cot\theta_{im}$ will be considered. By means of Evolutionary Polynomial Regression (EPR) application, the influence of the two conceptual slopes on the assessment of mean overtopping discharge will be evaluated and discussed.
- (8) The ratio $h_t^2/H_{m0,t}L_{m-1,0,t}$, which is defined as an impulsiveness ratio in [1].

It must be noticed that to define the wave run-up, different formulations can be employed. From one side, Equation (7) based on deep-water wave characteristics can be employed. On the other side, if local wave characteristics are used, we will refer to wave run-up formulas as described in [1], which is defined as follows:

$$Ru_{2\%} = 1.65H_{m0,t}\xi_{m-1,0} \tag{16}$$

with a maximum of

$$Ru_{2\%} = H_{m0,t} \left(4 - \frac{1.5}{\sqrt{\xi_{m-1,0}}} \right) \tag{17}$$

For steep and vertical walls ($\cot \alpha < 2$), the run-up will be expressed as:

$$Ru_{2\%} = H_{m0,t}(0.8 \cot \alpha + 1.6) \tag{18}$$

with a maximum of 3.0 and minimum of 1.8.

Equations (16)–(18) are also employed to calculate the equivalent slope $\cot\theta_{eq}$; meanwhile, Equation (7) is employed for the imaginary slope calculation, $\cot\theta_{im}$. The differences between the two methods are presented in Section 6. The most probable maximum run-up $Ru_{max} = (Ru_{max})_{99\%}$, assuming a Rayleigh distribution, is finally equal to $1.54Ru_{2\%}$.

4. Methodology

The search for a unified formula for wave overtopping assessment for smooth sea dikes and vertical walls, including cases with zero freeboard and emergent toe, is facilitated by employing

a data-driven approach, namely Evolutionary Polynomial Regression (EPR), which can be freely downloaded on <http://www.idea-rt.com/products-our-job/epr-moga-xl>.

EPR belongs to the big family of general-purpose data-driven techniques, whereas artificial neural networks (ANN) and genetic programming (GP) are probably the most well known. Based on our present understanding of the brain and nervous systems, ANN use highly simplified models composed of many processing elements ('neurons') connected by links of variable weights (parameters) to form a black-box representation of the system [28]. These models can deal with a great amount of information, while learning complex model functions from examples, i.e., by 'training' using sets of input and output data.

Genetic programming (GP) is another modeling approach that has recently increased in popularity. It is an evolutionary computing method that generates a 'transparent' and structured representation of the system being studied. The most frequently used GP method is so-called symbolic regression, which was proposed by [29]. This technique builds mathematical expressions to fit a set of data points using the evolutionary process of genetic programming.

Davidson et al. ([30,31]) introduced a new regression method for creating polynomial models based on both numerical and symbolic regression. They used GP to find the form of polynomial expressions and least squares optimization to find the values for the constants in the expressions. The incorporation of least squares optimization within symbolic regression was made possible by a rule-based component that algebraically transforms expressions into equivalent forms that are suitable for least squares optimization.

The EPR is based on the combination of numerical and symbolic regression, while eliminating the bulky, and often slow, rule-based component. The method also borrows from the idea of stepwise regression [32], using an evolutionary process based on genetic algorithms [33] rather than following a hill-climbing method of stepwise regression. More details are given in the following section.

4.1. The Evolutionary Polynomial Regression Paradigm

Evolutionary Polynomial Regression is a hybrid data-driven technique [34]. EPR is an optimization and modeling tool that combines and integrates numerical regression and genetic programming. The search for mathematical structure of models is performed through a population-based strategy that mimics the evolution of the individuals in nature.

Although the details about the main EPR paradigm are explained in the reference works ([34,35]), a concise description is reported here. EPR is built following the principle of Occam's razor, searching for the best models as a trade-off of model accuracy and parsimony. One of the general model structures in EPR can be expressed as follows:

$$\hat{Y} = \left[a_0 + \sum_{j=1}^m a_j X_1^{ES(j,1)} \cdot \dots \cdot f \left(X_k^{ES(j,k)} \right) \right] \quad (19)$$

where m is the number of additive terms, a_j are the numerical coefficients to be estimated, X_i are the candidate explanatory variables, $ES(j,z)$ is the exponent of the z^{th} input within the j^{th} term in Equation (19), and f is a function selected by the user among a set of possible alternatives (including no function selection). The exponents $ES(j,z)$ are selected from a user-defined set of candidate values (which should include 0).

It is worth noting that when an exponent is equal to zero, the relevant input X_i is basically deselected from the resulting equation (i.e., the exponent $ES(j,k) = 0$ and the relevant input $X_{kj} = 1$). This is crucial when investigating a physical phenomenon and the most important inputs to characterize it. In fact, during its search in the space of solutions, EPR can deselect a less important input by assigning a null exponent.

The EPR technique exploits a multi-objective search paradigm (EPR-MOGA) [35] in order to develop multiple Pareto optimal models. The multi-objective search in EPR-MOGA is

achieved by means of a multi-objective genetic algorithm based on the Pareto dominance criterion. First, the EPR-MOGA search options are defined by the user, such as, for example, candidate model attributes, candidate exponents for attributes, and the maximum number of parameters; based on these choices, EPR returns a set of Pareto-optimal models considering three conflicting objective functions: (1) maximization of model fitness to training data (i.e., model accuracy); (2) minimization of the number of model coefficients; and (3) minimization of the number of model inputs (i.e., both representative of model parsimony).

When the search options are defined by the user, at each generation, EPR-MOGA creates a population of models by simultaneously looking at the definition of the model structure, defining the exponents on the basis, for example, of Equation (19), and successively estimating the coefficients of polynomial terms by solving a linear inverse problem on the input data in order to guarantee a two-way relationship between each model structure and its parameters.

Then, the population of models is evaluated in the above-described Paretian dominance scheme, and only the best solutions pass to the next generation. In this way, the EPR-MOGA evolves within a few hundred generations a set of polynomial models that are optimal from the point of view of accuracy and parsimony. EPR-MOGA uses a MOGA strategy named OPTimized Multi-Objective Genetic Algorithm (OPTIMOGA), whose details are provided by [36]. The returned EPR mathematical structures are pseudo-polynomials, as they are obtained by adding a number of monomial terms, i.e., the argument of sum in Equation (19): they are linear with respect to their coefficients although not necessarily linear in their attributes, due to both exponents differing from one and the possible selection of function f in Equation (19).

Once the Pareto set of optimal models is obtained, the symbolic nature of EPR models allows a proper judgment of the optimal models looking at key aspects other than those encoded as objective functions, as for example: (1) the model structure with respect to the physical insight related to the problem; (2) similarities of mathematical structures of obtained models; (3) recurrent groups of variables in different models; (4) generalization performance of models as assessed in terms of both statistical analysis and mathematical parsimony [36–39].

4.2. Model Performance

The model expressions resulting from EPR analysis are evaluated on the basis of the coefficient of determination (CoD) and the average error (AVG) computed between the predicted and the measured data. CoD is defined as:

$$CoD = 1 - \frac{N - 1}{N} \frac{\sum_{i=1}^N (q_{predicted,i} - q_{measured,i})^2}{\sum_{i=1}^N (q_{measured,i} - mean(q_{measured,i}))^2} \quad (20)$$

where N is the number of data, and $q_{predicted,i}$ and $q_{measured,i}$ are respectively the i -th predicted and measured mean overtopping discharge. The higher the value of the CoD (with a maximum of 1), the better the model performance. The average error is defined as:

$$AVG = \sum_{i=1}^N |1 - q_{predicted} / q_{measured}| \quad (21)$$

Besides, following [13,23], the geometric mean (GEO) is introduced and defined as:

$$GEO = \exp \left[\frac{1}{N} \sum_{i=1}^N \ln x_i \right]; \text{ with } x_i = \frac{q_{predicted,i}}{q_{measured,i}} \quad (22)$$

The scatter of the data is assessed through the geometric standard deviation (*GSD*) that is calculated as the exponential value of the standard deviation of the logarithm:

$$GSD = \exp \left\{ \left[\frac{1}{N} \sum_{i=1}^N ((\ln x_i)^2 - (\ln \bar{x}_G)^2) \right]^{0.5} \right\} \quad (23)$$

Considering a quantity normally distributed, 90% of the data will be contained in the range between the mean divided by 1.64 times the *GSD* and the mean multiplied by 1.64 times the *GSD*.

5. Experimental Datasets for Fitting Overtopping Formula

Data from five different datasets have been employed for fitting the formula for mean wave overtopping. All data correspond to experimental results from wave flume facilities; thus, only long-crested waves are considered and the effects of wave obliqueness or directional spreading are neglected. Only simple structural layouts are considered, namely sloping dikes or vertical walls. The seaward face of the structure is smooth and impermeable. Defenses such as armor breakwaters with a permeable core are not considered. No storm walls or berms are considered. The overtopping is measured right after the seaward edge of the dike crest, so that the crest width is assumed to be equal to 0. All experimental tests with overtopping equal to 0 have been also excluded. The resulting data have been extracted from the following datasets:

- EurOtop [19] database: the new extended database comprises about 13,500 tests on wave overtopping and extends the original CLASH database [40]. The tests include different dike geometries and kinds of structures, 14% of which represent smooth dikes. By excluding cases with zero overtopping, oblique waves, berms, or storm walls and considering only core data, finally 1128 data have been extracted.
- Data described in [13] and corresponding to tests on sea dikes with shallow foreshores. The tests correspond to five different experimental campaigns carried out at Flanders Hydraulics Research, in Antwerp, and at the Department of Civil Engineering of Ghent University, in Ghent (both in Belgium). These experimental campaigns were conducted to study the influence of mild foreshore slopes and very or extremely shallow water conditions on wave overtopping and loading on smooth sea dikes. The geometry of the dikes resembled typical layouts from the Belgian coast.
- Data from [41] corresponding to coastal defenses in Japan built in extremely shallow foreshores or on land (emergent toe), namely seawalls with slopes of 1:3, 1:5, and 1:7 with uniform foreshore slope of either 1:10 or 1:30. A Bretschneider–Mitsuyasu spectrum was used to generate random wave trains, differently from other datasets here employed, where a JONSWAP spectrum was employed.
- Data from [42] corresponding to the experimental campaign carried out in the multi-directional wave basin, at Flanders Hydraulics Research. A 1:2 sloping dike after a 1:35 foreshore slope was employed. All data correspond to cases with mild foreshore slope ($\beta < 0.62$) in very and extremely shallow water conditions ($h_t/H_{m0,\rho} < 1$). The experimental campaign was conceived to characterize the overtopping discharge of long- and short-crested waves proceeding very obliquely with respect to the dike crest. For the present study, only perpendicular long-crested wave cases are considered.
- Data from Altomare et al. (under review) [43], where steep foreshore ($\beta > 0.62$) is used in combination with a rather steep dike (1:1) and very or extremely shallow waters ($h_t/H_{m0,\rho} < 1$). The datasets include cases with crests larger than 0, which have been excluded from the present analysis. The experiments were carried out at the Maritime Engineering Laboratory of Universitat Politècnica de Catalunya – BarcelonaTech (LIM/UPC), in Barcelona, Spain. The scope of the experimental campaign was to measure the overtopping flow properties and analyze the related safety of pedestrian standing on the promenade of a sea dike.

In total, 1679 data have been gathered. All the data are meant to cover a wide range of water depths, structural slopes, and freeboards. Table 1 reports for each employed dataset the range, $h_t/H_{m0,o}$ (shallowness criteria), foreshore and dike slopes, dimensionless freeboard, $R_c/H_{m0,t}$, and slope parameter, $\beta = \theta T_{m-1,0,o} \sqrt{(g/H_{m0,o})}$.

Table 1. Characteristics of extracted data for sea smooth sea dikes and vertical walls.

Dataset ID	Dataset Source	# of Data	$h_t/H_{m0,o}$	$\cot \alpha$	$\cot \theta$	$R_c/H_{m0,t}$	β
1	Altomare et al. (2016)	170	(−0.14)–0.86	2–6	35–50	0.41–3.50	0.29–0.62
2	EurOtop (2018)	1128	0.27–22.73	0–7	10–1000	0–6.24	0.01–3.5
3	Tamada et al. (2009)	230	(−0.27)–0.50	3, 5, 7	10–30	0.93–8.04	0.44–1.92
4	Altomare et al. (under review)	26	0.10–0.36	1	15	2.95–6.19	1.13–1.42
5	Altomare et al. (2020)	125	0.12–0.58	1.97	35	0.53–3.24	0.37–0.58

Datasets from [13,41] include cases with an emergent toe ($h_t \leq 0$), for which it is important to provide some details on the methodology followed to measure local wave characteristics. For emergent toes and more in general for extremely shallow water cases, the waves break heavily before reaching the dike. These broken waves present an unusual shape that is similar to a bore rather than to an oscillatory phenomenon around the mean water level. Their oscillatory nature can be no longer distinguished, yet the spectral momentum and, therefore, the wave height and period can be defined. The accuracy of this choice is demonstrated after showing the overtopping results in [13]. For some of the data employed for the present work, the information at the dike toe was lacking. In some cases, specifically in those from [13], the SWASH model (<http://swash.sourceforge.net/>) was employed to get the incident wave height and period at the toe. For other cases from [41], the method developed by [25] was applied. This method allows calculating the wave height at the toe even when the water depth is 0. Citing [25]: “A finite wave height is considered at the shoreline that originates from the presence of stationary wave set-up and time-varying surf-beat there.” In the case of a dry or emergent toe, the same method from [25] has been applied, replacing the dry toe ($h_t < 0$) with $h_t = 0$. In other words, the solution is forced at $h_t = 0$. A similar procedure has been found in [44].

All data are plotted in Figure 3, where a scatter plot ([45]) among the main variables is shown. Each data has been scaled to $R_c = 1$ m using Froude’s similarity law in order to compare different datasets. The selected variables are mean overtopping discharge (expressed in logarithmic scale), deep-water wave height and period, local wave height and period, foreshore and dike slope, local water depth, and maximum wave run-up. The last one has been calculated using Equation (7) and is plotted in logarithmic scale. It can be noticed that despite the large variety of data and range of each dataset, a strong relationship can be still identified between the discharge values and the local wave height. Even better correlation is noticeable in the wave run-up, where the calculation of wave run-up depends not only on the wave height but also on the wave period, foreshore, and dike slope.

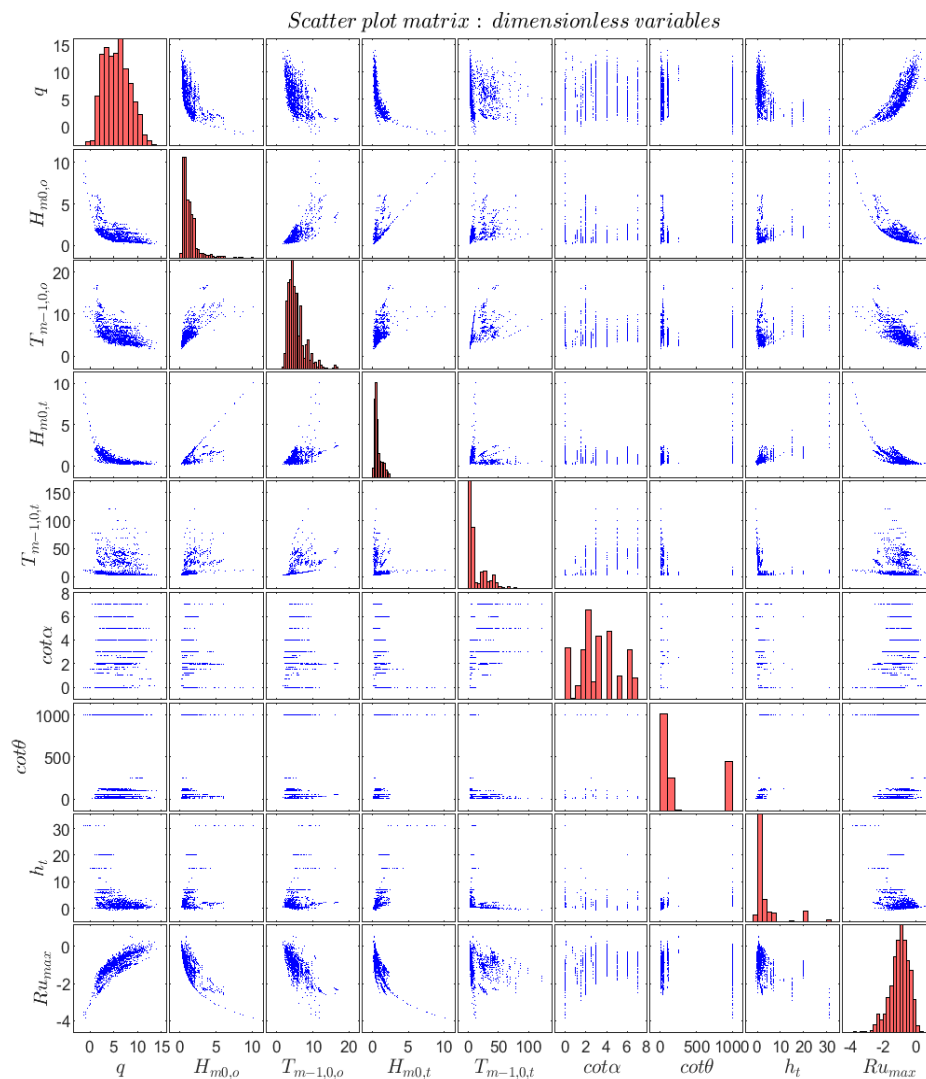


Figure 3. Scatter-plot matrix of dimensionless variables.

These dependency is shown with more detail in terms of the dimensionless variables in Figure 4, where the left plot represents the variation of dimensionless discharge $Q^* = 4\pi^2 q/gH_{m0,o}T_{m-1,o,o}$ against the dimensionless freeboard; meanwhile, in the right plot, the dependence is shown in terms of deficit in freeboard. For the former one, all data but the ones from dataset #3 are quite clustered. Dataset #3 corresponds to the data from [22], mainly with sea dikes and walls with extremely shallow waters and even an emergent toe. In the right plot of Figure 4, the maximum run-up has been calculated with Equation (7) ($Ru_{max,im}$). In this case, the data belonging to each dataset seem to follow a different trend but converge for value of the deficit in freeboard equal to 1 (i.e., freeboard equal to 0). The relationship between discharge, dimensionless freeboard, and deficit in the freeboard will be investigated further and discussed in Sections 5 and 6. It is worthwhile to notice that employing the dimensionless discharge Q^* scaled by the volume flux leads to less data scatter than employing the most used $Q = q/\sqrt{(gH_{m0,t}^3)}$, especially for low freeboards and large overtopping rates (where measurement uncertainties are less and the test repeatability is usually high): see the comparison between Figure 4 (left plot) and Figure 5.

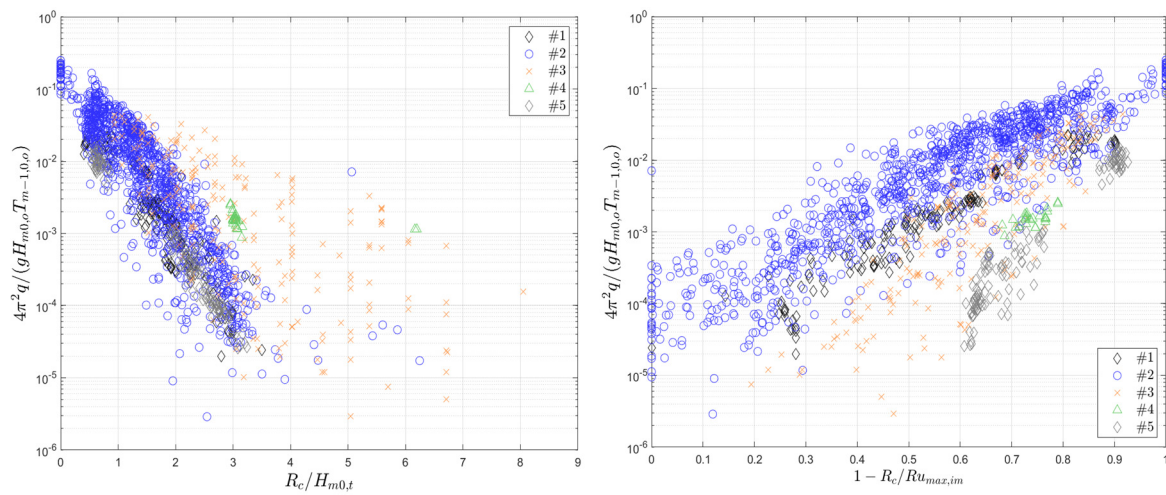


Figure 4. Variation of the dimensionless overtopping Q^* discharge versus with the dimensionless freeboard (left) and deficit in freeboard (right) for each of the five employed datasets.

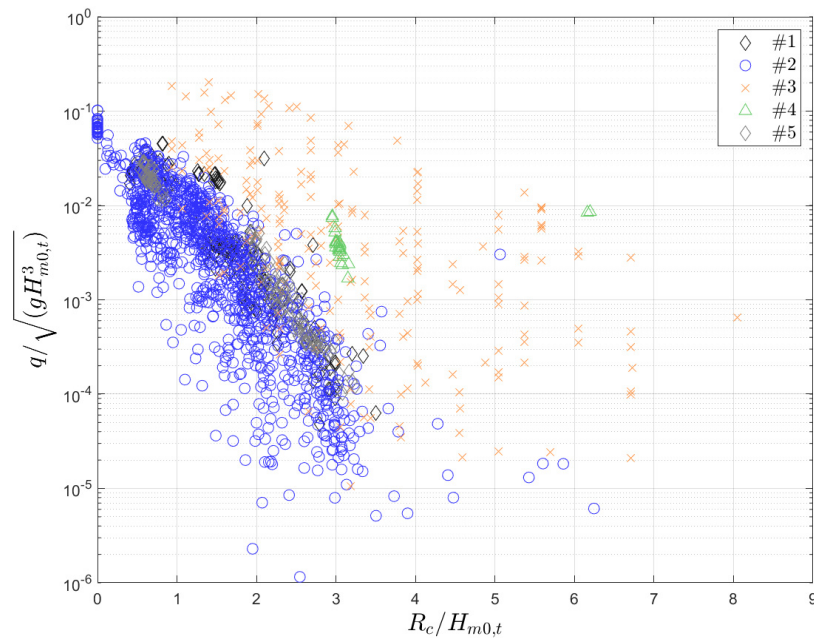


Figure 5. Variation of the dimensionless overtopping discharge $q/\sqrt{(gH_{m0}^3)}$ versus with the dimensionless freeboard.

The measured discharge values are finally compared with ones calculated using Equation (1), which is proposed as a unified formula for wave overtopping by [23]. Results are depicted in Figure 6. Three lines are added to the graphs: a solid line that corresponds to a prediction equal to the measurement and two dashed lines corresponding to overtopping predictions that are 10 times and 0.1 times the measured data. Employing Equation (1), most of the predicted data are within the range $0.1 < q_{predicted}/q_{measured} < 10$, except for data from datasets #1 and #3, which correspond to the two datasets having extremely shallow water conditions with emergent toes. The inaccuracy to predict the mean overtopping discharge of these two datasets is due to the range of applicability of Equation (1) and the range of hydraulic and geometrical parameters used to derive the coefficients expressed in Equations (2)–(5), where a clear lack of data in very or extremely shallow water conditions has been noticed in the dataset employed by [23].

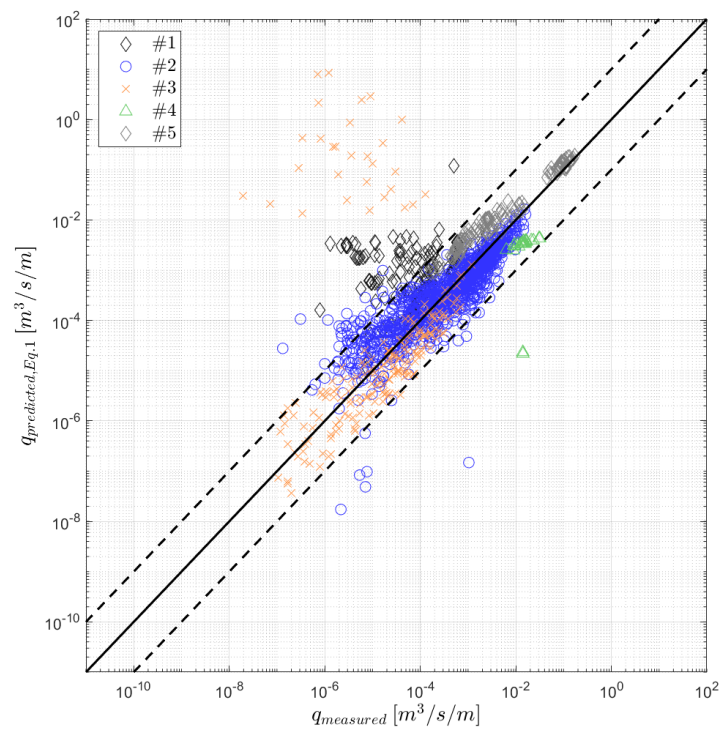


Figure 6. Comparison of Goda (2009) and measured overtopping discharge (solid line: prediction using Equation (1) equal to measured rate; dashed lines: prediction using Equation (1) equal to 10 times and 0.1 times the measured rate).

6. A Set of Unified Formulas for the Preliminary Assessment of Wave Overtopping on Smooth Sea Dikes and Vertical Walls

EPR has been already successfully applied in coastal engineering and other engineering fields ([36–39]). Here, EPR is applied to the collected data described in Section 4. The genetic structure of the formulas is assumed a priori based on the exiting literature. The set of variables employed in EPR modeling (see Section 3.2) are based on certain physical assumptions and derived by [12,13,23,26,27]. EPR modeling returns a set or family of formulas, which allows identifying recurrent and meaningful groups or explanatory variables and comparing all expressions in terms of accuracy and parsimony. Then, the so-derived formulas are validated with the experimental data.

EPR Setting and Results

The 1679 data related to hydraulic and geometrical parameters have been divided into two sets: 1343 data have been used as a training set for EPR modeling, while 336 data have been employed as a test set, used for the evaluation of the returned models and not for model construction. The test data have been randomly picked up from the whole database.

EPR has been applied to look among the explanatory variables defined in Section 3.2 and identify those influent for mean wave overtopping discharge. Since most of the studies in the literature on wave overtopping rely on the assumption of an exponential function, this hypothesis has been kept for EPR modeling. Both equivalent slope and imaginary slope have been considered as defined in [13] and [22], respectively. Accordingly, maximum run-up values have been calculated either with Equation (7), $Ru_{max,im}$, or Equations (16)–(18), $Ru_{max,eq}$. A set of model expressions returned by EPR is reported in Table 2. Looking for model parsimony, here, only expressions limited to a maximum of 3 terms (excluding the constant) and 4 inputs are reported. The values of CoD and AVG for both training and test data are reported, and the results overall are very similar to each other.

Table 2. Model expressions returned by Evolutionary Polynomial Regression (EPR). *AVG*: average error, *CoD*: coefficient of determination.

Model #	Model Expression	Training		Test	
		CoD	AVG	CoD	AVG
I	$Q^* = \exp\left(8.42\left(1 - \frac{R_c}{Ru_{max,eq}}\right)^{1.5} - 9.80\right)$	0.69	0.18	0.70	0.18
II	$Q^* = \exp\left(-4.19\sqrt{\frac{R_c}{H_{m0,t}}} - 0.37\right)$	0.65	0.20	0.66	0.20
III	$Q^* = \exp\left(8.10\left(1 - \frac{R_c}{Ru_{max,eq}}\right)^{1.5} - 0.14 \cot \theta_{eq} - 9.00\right)$	0.75	0.17	0.74	0.17
IV	$Q^* = \exp\left(-4.6\sqrt{\frac{R_c}{H_{m0,t}}} + 0.87\beta - 0.29\right)$	0.70	0.20	0.70	0.20
V	$Q^* = \exp\left(- (4.95 - 0.69\beta)\sqrt{\frac{R_c}{H_{m0,t}}} + 0.14\right)$	0.71	0.20	0.71	0.19
VI	$Q^* = \exp\left(3.68\left(1 - \frac{R_c}{Ru_{max,im}}\right) - 2.94\sqrt{\frac{R_c}{H_{m0,t}}} - 4.11\right)$	0.73	0.18	0.74	0.18
VII	$Q^* = \exp\left(-\frac{R_c}{H_{m0,t}} + 3.94\left(1 - \frac{R_c}{Ru_{max,im}}\right) - 0.08 \cot \theta_{im} - 5.80\right)$	0.76	0.16	0.74	0.16
VIII	$Q^* = \left(-2.49\sqrt{\frac{H_{m0,0}}{H_{m0,t}}} + 7.23\left(1 - \frac{R_c}{Ru_{max,im}}\right)^{1.1} - 6.65\right)$	0.75	0.17	0.74	0.17
IX	$Q^* = \exp\left(1.56\frac{H_{m0,0}}{H_{m0,t}} - 2 + 7.89\left(1 - \frac{R_c}{Ru_{max,eq}}\right)^{1.5} - 10.67\right)$	0.74	0.16	0.74	0.16
X	$Q^* = \exp\left(-1.80\sqrt{\frac{H_{m0,0}}{H_{m0,t}}} + 5.37\left(1 - \frac{R_c}{Ru_{max,im}}\right) - 0.68\frac{R_c}{H_{m0,t}} - 5.38\right)$	0.76	0.15	0.76	0.16

The dimensionless freeboard and the deficit in freeboard results are the most frequent variables in each model expression. In most of the cases, the dimensionless discharge results are functions of the square root of the dimensionless freeboard; meanwhile, the deficit in freeboard is expressed in form of power functions with exponents varying between 1 and 1.5 (the latter is the most frequent). A direct dependence of Q^* on the foreshore and dike slope has been investigated, as for example expressed in Equations (2)–(4). However, the results were poor in terms of accuracy. Instead, employing the equivalent slope or the imaginary slope improved the model accuracy (see models III and VII). Besides, it must be considered that the same run-up is calculated as a function of the equivalent or imaginary slope, hence corroborating the importance of such parameters on the expression of mean discharge. The slope parameter defined by [27] has been selected in two cases (model IV and V). In the last three models, the quantity $H_{m0,o}/H_{m0,t}$ has been selected, as derived from the ratio between the $h_t/H_{m0,t}$ and $h_t/H_{m0,o}$. Notwithstanding, models VIII, IX, and X do not add further accuracy, and the dependence of the dimensionless discharge on $H_{m0,o}/H_{m0,t}$ seems rather the result of model overfitting than something physically based.

To support the choice of a proper model expression, the model accuracy has been estimated also in terms of geometric mean (*GEO*) and standard deviation (*GSD*), which are calculated on the entire amount of data. Results are reported in Table 3. The 90% interval of the ratio $q_{predicted}/q_{estimated}$ calculated based on *GEO* and *GSD* for each model expression is reported. The model expression VII results are the most accurate, having a *GEO* almost equal to 1 (=0.99) and the lowest standard deviation. For model VII, assuming the overtopping rate as normally distributed, 90% of the predicted overtopping rate is to be located in the range between 0.20 and 4.84 times the measured overtopping discharge. Excluding models VIII, IX, and X, the second most accurate model expression is model III, which is a function of the deficit in freeboard and the equivalent slope. The accuracy of Equation (1) and Equation (6) has been also assessed, showing a larger confidence interval for both equations, with a tendency to overestimate overtopping discharge when Equation (1) is employed. The differences between equivalent and imaginary slopes will be discussed later. Yet, it is evident that a slope—a combination of foreshore slope and dike slope, and the definition of which depends also on the local water depth—is a key variable to express the mean overtopping discharge. The dependence of discharge on θ , α , and h_t was already remarked in [13,23]. The results of predicted mean discharge rates employing model III and VII are plotted against the measured ones in Figures 7 and 8, respectively.

Table 3. Geometric mean and standard deviation for each EPR model expression. The 90% interval of $q_{predicted}/q_{estimated}$ is reported. *GEO*: geometric mean, *GSD*: geometric standard deviation.

Model #	GEO	GSD	q_{pred}/q_{EXP}	
			$\mu - 1.64\sigma$	$\mu + 1.64\sigma$
I	1.00	3.40	0.18	5.58
II	1.15	3.60	0.19	6.77
III	0.99	3.08	0.20	5.01
IV	1.00	3.38	0.18	5.53
V	1.00	3.29	0.19	5.39
VI	0.99	3.14	0.19	5.10
VII	0.99	2.98	0.20	4.84
VIII	0.85	3.00	0.17	4.18
IX	0.93	3.28	0.17	5.00
X	0.89	3.11	0.17	4.54
Equation (1)	1.59	8.03	0.12	20.94
Equation (6)	0.96	24.55	0.02	38.65

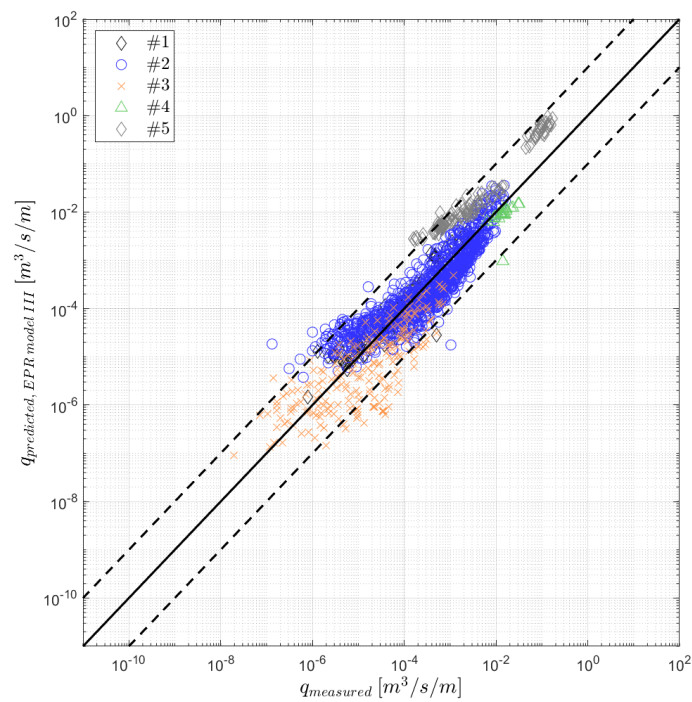


Figure 7. Comparison of overtopping prediction employing EPR model expression III and measured overtopping discharge (solid line: prediction equal to measured rate; dashed lines: prediction equal to 10 times and 0.1 times the measured rate).

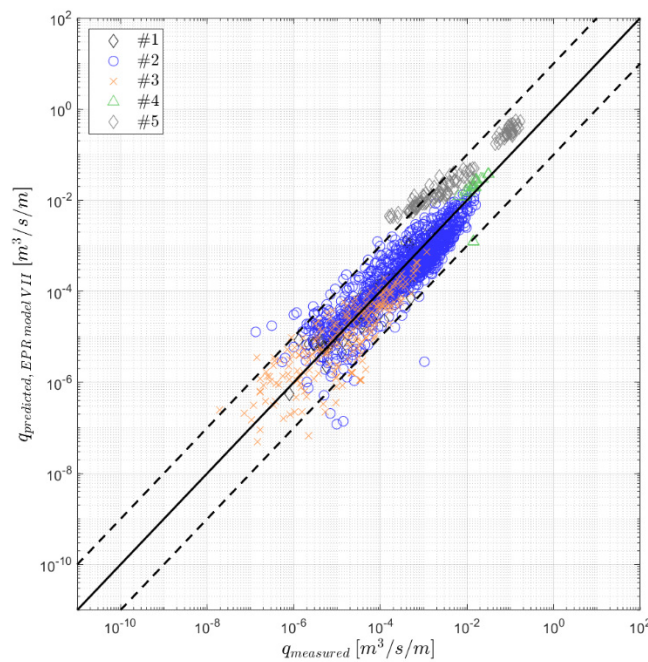


Figure 8. Comparison of overtopping prediction employing EPR model expression VII and measured overtopping discharge (solid line: prediction equal to measured rate; dashed lines: prediction equal to 10 times and 0.1 times the measured rate).

7. Discussion

The set of model expressions fitted with EPR has been presented in the previous section. The exponential structure for each model expression is confirmed to be the most suited for expressing the mean overtopping discharge as widely used in the literature. The dimensionless freeboard and deficit in freeboard resulted in the main explanatory variables for mean overtopping discharge assessment. The mean discharge has been expressed in dimensionless form by scaling it with the volume flux derived and expressed in Equation (14) in agreement with [26]. A preliminary analysis of the volume flux has shown that there is no difference employing either deep-water wave characteristics or local wave characteristics (Figure 9).

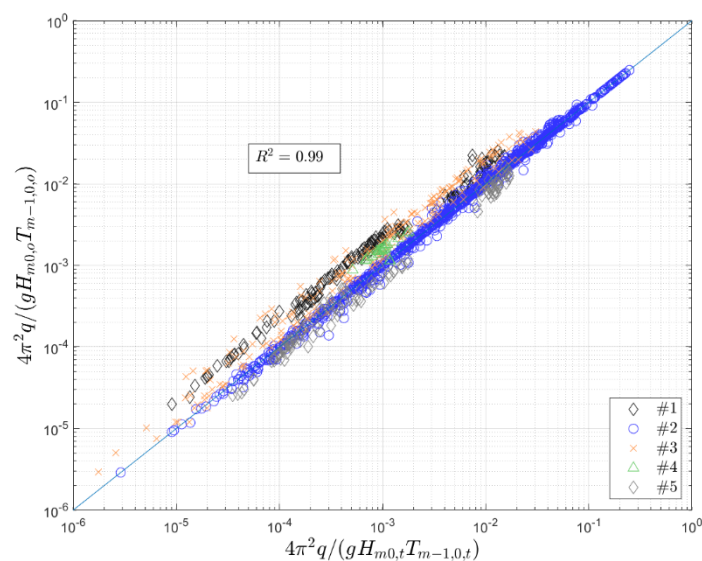


Figure 9. Comparison between the volume flux calculated employing deep-water wave characteristics and local ones.

All calculations presented in this work employ the deep-water wave characteristics for the definition of Q^* . Differently from the most used formulas from the literature (e.g., [13,19,22,23,46,47]), where the average discharge is scaled by $\sqrt{(gH_{m0,t}^3)}$, here, the scaling factor includes the wave period as a key variable to explain overtopping discharge. Assuming the extreme case of zero freeboard for different wave periods but keeping all other hydraulic and geometrical characteristics fixed, the mean discharge will be different; namely, it will be larger for longer wave periods. This result is actually corroborated by the evidence that if we assume sea storms of the same duration, which are expressed as the number of waves in one wave train, longer waves will lead to larger volumes than shorter ones. However, only breaking wave formulas, such as Equation (8), express similar relationships. For non-breaking waves, the wave period does not play a role at all. For very and shallow water conditions, the wave period is considered only within the exponent; hence, there is no influence in case of zero freeboard.

The EPR results show that the deficit in freeboard is a recurrent variable; hence, it is fundamental for mean overtopping discharge (EPR models I, III, VI, VII, VIII, IX, X), as demonstrated by [26]. The larger the deficit, the larger the overtopping. Mase et al. [22] expressed the discharge rate as a function of the deficit in freeboard. Besides, better model performance is achieved combining it with the dimensionless freeboard (models VI, VII, and X). Comparing models with only dimensionless freeboard or deficit in freeboard, better model accuracy is attained when the latter explanatory variable is employed. The main difference is that the dimensionless freeboard considers only two variables: crest freeboard R_c and local wave height $H_{m0,t}$. Employing instead the deficit in freeboard includes the wave run-up calculation and hence the influence of wave period, local water depth, foreshore slope, and dike slope. This means proper consideration of the wave transformation and eventual wave breaking to which the waves are subjected before reaching the dike toe and dike crest ([13]).

At this point, a discussion on the equivalent slope and imaginary slope concepts is mandatory. Both take into account local processes of wave transformation and breaking on the foreshore first and then on the dike slope. Both concepts are applicable to cases of structures with an emergent toe. However, for the imaginary slope, the calculation of the breaking depth is required. In this work, the formula proposed by [48] is employed to calculate the breaking depth, h_b . For the equivalent slope, if the local water depth is bigger than $1.5H_{m0,t}$, only the dike slope is assumed. In case of vertical walls, this will correspond therefore to a value equal to 0. Besides, the calculation of the equivalent slope and the imaginary slope requires an iterative process, since the wave run-up, which is necessary for the slope definition, is not known a priori and depends on the same slope. Uncertainties related to applicability range of the same run-up formulas raise questions about the accuracy of one method over another. A comparison between the calculated imaginary slope and equivalent slope is reported in Figure 10. It can be noticed that for most of the data, the two calculations lead to similar results; however, for some data, the imaginary slope is far larger than the equivalent one. This is due to the calculation of the breaking depth and its distance from the dike toe, especially for very mild foreshores: the area underneath the foreshore will actually increase significantly, giving more weight to the foreshore slope than the dike slope (see Figure 2). In general, however, EPR models show similar dependence of the overtopping rates on the imaginary and equivalent slopes: namely, the larger the slope is, the smaller the overtopping. Larger equivalent or imaginary slopes correspond to cases where the foreshore slope is rather influential; hence, heavier wave breaking and bigger wave energy dissipation are expected. This leads to reduced overtopping discharge. As an example, in countries such as Belgium or The Netherlands, beach nourishment is employed in such cases as a soft countermeasure to reduce overtopping and flooding. In fact, the resulting very long and mild beaches before sea dikes induce heavy wave breaking that decrease overtopping rates.

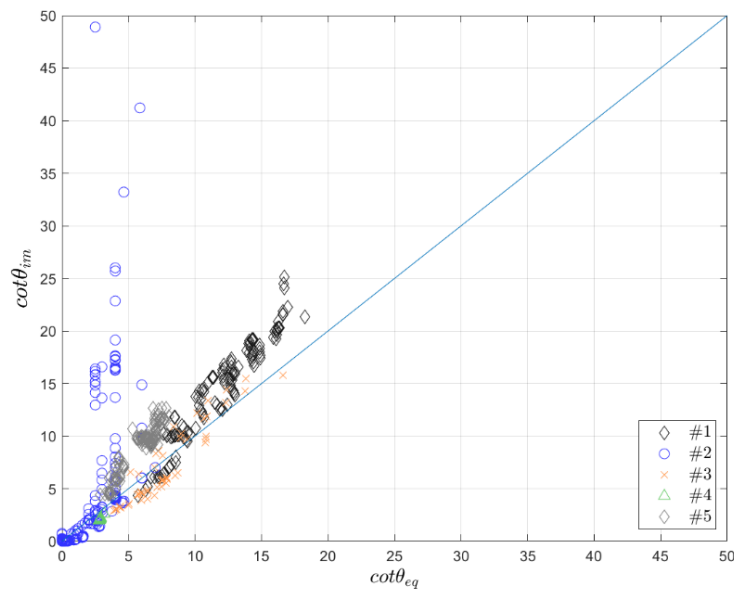


Figure 10. Imaginary slope (Mase et al., 2013) versus equivalent slope (Altomare et al., 2016).

Differences can be noticed also in the calculated run-up values, see Figure 11, where the maximum run-up is scaled by the crest freeboard. Looking at the EPR model expressions in Table 2, it can be observed that when the deficit in freeboard is calculated based on Equation (7), it appears in the exponential function with an exponent equal to 1. Instead, when Equations (16)–(18) and the equivalent slope are employed, the deficit in freeboard presents an exponent equal to 1.5 in all models but one.

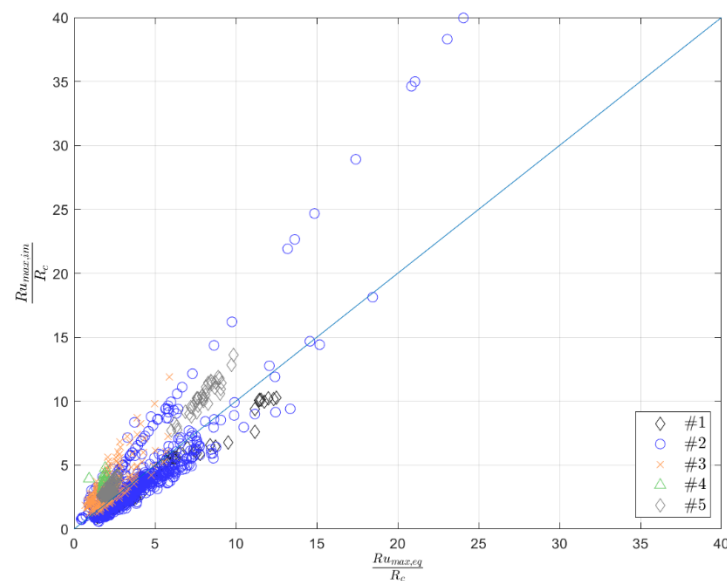


Figure 11. Comparison of dimensionless run-up values calculated with formulations proposed in EurOtop (2018) and Mase et al. (2013) and employing the equivalent ($Ru_{max,eq}$) and imaginary slope ($Ru_{max,im}$), respectively.

The uncertainties in the accuracy of the wave run-up formula and the different slope concept employed in each case are responsible for such discrepancy. The variation of Q^* on the deficit in freeboard is depicted in Figure 4 (right plot) and Figure 12, for $Ru_{max,im}$ and $Ru_{max,eq}$, respectively. Further research aiming to derive a unified formula for wave run-up is advised, covering the whole range of sloping dikes and vertical walls, for all water conditions and even an emergent toe. For it, ad hoc experimental tests and new measurements will be required. Despite the clear difference

and independently of the way it is calculated, the deficit in freeboard remains the most important explanatory variable for the process at stake. From a practical point of view, model expression III might be preferred to model expression VII for two main reasons: (1) uncertainty in the calculation of the breaking depth required to define the imaginary slope in model expression VII; and (2) model expression VII requires using both deep-water and local wave conditions. The former ones are needed to calculate the run-up based on Equation (7), and therefore, the deficit in freeboard; the latter ones define the dimensionless freeboard. Instead, the terms of the right-hand side in model expression III are expressed as a function of local wave conditions only, and due to the almost identity of volume flux expressed either with deep-water or local wave conditions (Figure 9), the latter ones can be used to scale the mean discharge (left-hand side in model III).

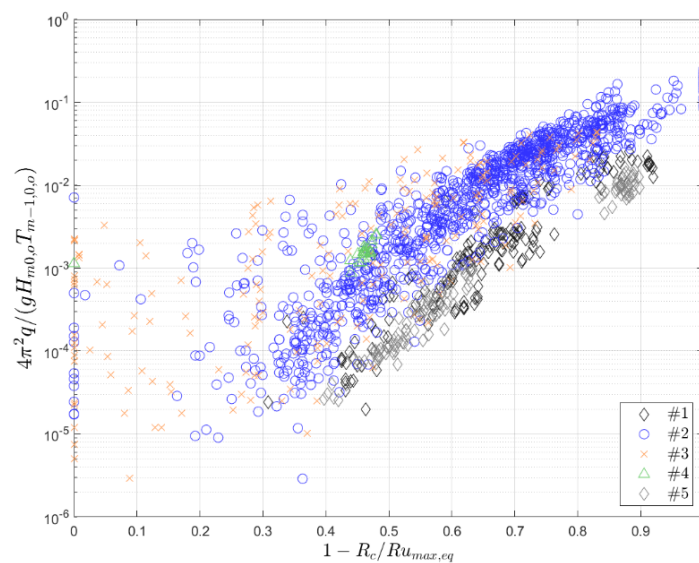


Figure 12. Dimensionless freeboard versus deficit in freeboard, where maximum run-up is calculated employing the formulas from EurOtop (2018) and the equivalent slope from Altomare et al. (2016).

8. Conclusions

The Evolutionary Polynomial Regression technique is employed in the present work to mine information from five different datasets gathered together to identify the most important explanatory variables that describe the mean overtopping discharge of smooth sea dikes and vertical walls. First, scaling laws for wave overtopping are discussed. A set of unified formulas is derived to cover a broad range of hydrodynamic conditions and geometrical characteristics and to fill the existing gaps between the most used semi-empirical formulas from the literature. Each EPR model expression is analyzed in terms of the physics of the phenomenon at stake, and expressions with eventual model overfitting are excluded. The main finding can be summarized as follows:

- Mean overtopping discharge is scaled by the volume flux, including both wave height and period ([26]): dimensionless discharge can be defined as $Q^* = 4\pi^2q/gH_{m0,\rho}T_{m-1,0,\rho}$. In this way, the wave period is always considered as an explanatory variable for wave overtopping assessment also for those cases where it is usually not taken into account; see Equation (9) for non-breaking waves.
- Employing deep-water wave characteristics rather than local wave characteristics at the dike toe to define the volume flux and, consequently, the scaling law for overtopping, leads to differences that are negligible. The reader can employ either the former ones or the latter ones, depending on the availability and accuracy of the employed information.
- EPR analysis confirms that an exponential structure, the most employed in literature (see [19]), leads to the best model representation.

- Both the dimensionless freeboard, $R_c/H_{m0,t}$, and the deficit in freeboard, $1-R_c/Ru_{max}$, are key explanatory variables for wave overtopping assessment. The main difference is that the definition of the deficit in freeboard requires the assessment of wave run-up, which is not only depending on the wave height but also on the wave period, local water depth, foreshore slope, and dike slope.
- Therefore, the study confirms previous findings from [13,22,23], and partly from [12], where other variables such as dike and foreshore slope and local water depth play an important role for the calculation of mean discharge rates. By combination these variables in one equivalent or imaginary slope concept, the model accuracy increases.
- EPR results confirm that the imaginary or equivalent slopes provided improve the estimate of mean discharge values, confirming somehow the physical meaning of such concepts. Overtopping reduces when the slope (defined as a cotangent of the angle with the horizontal) increases (heavier wave braking and dissipation over the foreshore slope).

Although the set of formulas are derived only for smooth structures, their application can be extended to rough and permeable structure (e.g., breakwaters) for a preliminary and conservative assessment of the mean discharge value.

Further research is required to define a unified formula for wave run-up and a unique concept for equivalent or imaginary slope. Nevertheless, the results of the present research provide already a clear indication of the determining variables and groups of variables for overtopping assessment.

Finally, the proposed set of formulas does not replace the most and widely used semi-empirical models for mean wave overtopping ([19]), but it aims to offer a simple method for a preliminary assessment of wave overtopping for any kind of structural layout and hydrodynamic conditions of smooth sea dikes and vertical walls, overcoming the existing model gaps.

Author Contributions: Conceptualization, C.A.; methodology, C.A. and D.B.L.; software, D.B.L.; validation, C.A.; formal analysis, C.A.; investigation, C.A.; resources, C.A., H.M., X.G.; data curation, C.A., H.M.; writing—original draft preparation, C.A.; writing—review and editing, H.M., X.G., D.B.L.; visualization, C.A.; supervision, X.G.; project administration, X.G.; funding acquisition, C.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No.: 792370.

Conflicts of Interest: The authors declare no conflict of interest.

Disclaimer: The presented results reflect only the authors' view and Research Executive Agency (REA) is not responsible for any use that may be made of the information it contains.

References

1. Sandoval, C.; Bruce, T. *Wave Overtopping Hazard to Pedestrians: Video Evidence from Real Accidents in Coasts, Marine Structures and Breakwaters 2017*; ICE Virtual Library: London, UK, 2018; pp. 501–512. [[CrossRef](#)]
2. Mares-Nasarre, P.; Argente, G.; Gómez-Martín, M.E.; Medina, J.R. Overtopping layer thickness and overtopping flow velocity on mound breakwaters. *Coast. Eng.* **2019**, *154*. [[CrossRef](#)]
3. Nørgaard, J.Q.H.; Lykke Andersen, T.; Burcharth, H.F. Distribution of individual wave overtopping volumes in shallow water wave conditions. *Coast. Eng.* **2014**, *83*, 15–23. [[CrossRef](#)]
4. Victor, L.; van der Meer, J.W.; Troch, P. Probability distribution of individual wave overtopping volumes for smooth impermeable steep slopes with low crest freeboards. *Coast. Eng.* **2012**, *64*, 87–101. [[CrossRef](#)]
5. Zanuttigh, B.; van der Meer, J.; Bruce, T.; Hughes, S. Statistical Characterisation of Extreme Overtopping Wave Volumes. In *From Sea to Shore—Meeting the Challenges of the Sea*; ICE Publishing: London, UK, 2014; pp. 442–451.
6. van Bergeijk, V.M.; Warmink, J.J.; van Gent, M.R.A.; Hulscher, S.J.M.H. An analytical model of wave overtopping flow velocities on dike crests and landward slopes. *Coast. Eng.* **2019**, *149*, 28–38. [[CrossRef](#)]
7. Schüttrumpf, H.; Oumeraci, H. Layer thicknesses and velocities of wave overtopping flow at seadikes. *Coast. Eng.* **2005**, *52*, 473–495. [[CrossRef](#)]

8. Quvang, J.; Nørgaard, H.; Lykke, T.; Burcharth, H.F.; Jan, G. Analysis of overtopping flow on sea dikes in oblique and short-crested waves. *Coast. Eng.* **2013**, *76*, 43–54. [[CrossRef](#)]
9. Romano, A.; Bellotti, G.; Briganti, R.; Franco, L. Uncertainties in the physical modelling of the wave overtopping over a rubble mound breakwater: The role of the seeding number and of the test duration. *Coast. Eng.* **2015**, *103*, 15–21. [[CrossRef](#)]
10. Williams, H.E.; Briganti, R.; Pullen, T. The role of offshore boundary conditions in the uncertainty of numerical prediction of wave overtopping using non-linear shallow water equations. *Coast. Eng.* **2014**, *89*, 30–44. [[CrossRef](#)]
11. Owen, M.W. *Design of Seawalls Allowing for Wave Overtopping*; Report No. EX 924; Hydraulics Research: Wallingford, UK, 1980.
12. Gallach Sanchez, D. *Experimental Study of Wave Overtopping Performance of Steep Low-Crested Structures*; Ghent University: Ghent, Belgium, 2018.
13. Altomare, C.; Suzuki, T.; Chen, X.; Verwaest, T.; Kortenhaus, A. Wave overtopping of sea dikes with very shallow foreshores. *Coast. Eng.* **2016**, *116*, 236–257. [[CrossRef](#)]
14. Van der Meer, J.W.; Bruce, T. New Physical Insights and Design Formulae on Wave Overtopping at Sloping and Vertical Structures. *J. Waterw. Port Coast. Ocean Eng.* **2013**, *140*. [[CrossRef](#)]
15. Van Doorslaer, K.; De Rouck, J.; Audenaert, S.; Duquet, V. Crest modifications to reduce wave overtopping of non-breaking waves over a smooth dike slope. *Coast. Eng.* **2015**, *101*, 69–88. [[CrossRef](#)]
16. Chen, W.; van Gent, M.R.A.; Warmink, J.J.; Hulscher, S.J.M.H. The influence of a berm and roughness on the wave overtopping at dikes. *Coast. Eng.* **2020**, *156*, 103613. [[CrossRef](#)]
17. Troch, P.; Geeraerts, J.; Van de Walle, B.; De Rouck, J.; Van Damme, L.; Allsop, W.; Franco, L. Full-scale wave-overtopping measurements on the Zeebrugge rubble mound breakwater. *Coast. Eng.* **2004**, *51*, 609–628. [[CrossRef](#)]
18. Van Gent, M.R.A. *Physical Model Investigations on Coastal Structures with Shallow Foreshores: 2d Model Tests with Single and Double-Peaked Wave Energy Spectra*; Deltares (WL): Delft, The Netherlands, 1999.
19. Van der Meer, J.W.; Allsop, N.W.H.; Bruce, T.; De Rouck, J.; Kortenhaus, A.; Pullen, T.; Schüttrumpf, H.; Troch, P.; Zanuttigh, B. *Manual on Wave Overtopping of Sea Defences and Related Structures. An Overtopping Manual Largely Based on European Research, but for Worldwide Application*; EurOtop, 2018; Available online: www.overtopping-manual.com (accessed on 24 July 2020).
20. Allsop, W.; Bruce, T.; Pearson, J.; Besley, P. Wave overtopping at vertical and steep seawalls. *Proc. Inst. Civ. Eng. Marit. Eng.* **2005**, *158*, 103–114. [[CrossRef](#)]
21. Victor, L.; Troch, P. Wave Overtopping at Smooth Impermeable Steep Slopes with Low Crest Freeboards. *J. Waterw. Port Coast. Ocean Eng.* **2012**, *138*, 372–385. [[CrossRef](#)]
22. Mase, H.; Tamada, T.; Yasuda, T.; Hedges, T.S.; Reis, M.T. Wave Runup and Overtopping at Seawalls Built on Land and in Very Shallow Water. *J. Waterw. Port Coast. Ocean Eng.* **2013**, *139*, 346–357. [[CrossRef](#)]
23. Goda, Y. Derivation of unified wave overtopping formulas for seawalls with smooth, impermeable surfaces based on selected CLASH datasets. *Coast. Eng.* **2009**, *56*, 385–399. [[CrossRef](#)]
24. Tamada, T.; Inoue, H.; Inoue, M. Proposal of diagrams for the estimation of wave overtopping rate on gentle slope-type seawall and consideration on time-related variation. *Proc. Civ. Eng. Ocean* **2001**, *17*, 311–316. [[CrossRef](#)]
25. Goda, Y. A performance test of nearshore wave height prediction with CLASH datasets. *Coast. Eng.* **2009**, *56*, 220–229. [[CrossRef](#)]
26. Ibrahim, M.S.I.; Baldock, T.E. Swash overtopping on plane beaches – Reconciling empirical and theoretical scaling laws using the volume flux. *Coast. Eng.* **2020**, *157*, 103668. [[CrossRef](#)]
27. Hofland, B.; Chen, X.; Altomare, C.; Oosterlo, P. Prediction formula for the spectral wave period $T_{m-1,0}$ on mildly sloping shallow foreshores. *Coast. Eng.* **2017**, *123*. [[CrossRef](#)]
28. Haykin, S. *Neural Networks: A Comprehensive Foundation*; Prentice Hall: Upper Saddle River, NJ, USA, 1999.
29. Koza, J.R. Genetic programming as a means for programming computers by natural selection. *Stat. Comput.* **1994**, *4*, 87–112. [[CrossRef](#)]
30. Davidson, J.W.; Savic, D.; Walters, G.A. Method for the identification of explicit polynomial formulae for the friction in turbulent pipe flow. *J. Hydroinform.* **1999**, *1*, 115–126. [[CrossRef](#)]
31. Davidson, J.W.; Savic, D.; Walters, G.A. *Approximators for the Colebrook-White Formula Obtained through a Hybrid Regression Method*; Taylor & Francis, Inc.: Abingdon, UK, 2000; pp. 983–989.

32. Draper, N.R.; Smith, H. *Applied Regression Analysis*; John Wiley & Sons: New York, NY, USA, 1966; 407 S., 43 Abb., 2 Tab., 180 Literaturangaben, Preis: S 90. Toutenburg, H. **1969**, *11*, 427. [[CrossRef](#)]
33. Goldberg, D.E. *Genetic Algorithms in Search, Optimization and Machine Learning*, 1st ed.; Addison-Wesley Longman Publishing Co., Inc.: Boston, MA, USA, 1989.
34. Giustolisi, O.; Savic, D.A. A symbolic data-driven technique based on evolutionary polynomial regression. *J. Hydroinform.* **2006**, *8*, 207–222. [[CrossRef](#)]
35. Giustolisi, O.; Savic, D.A. Advances in data-driven analyses and modelling using EPR-MOGA. *J. Hydroinform.* **2009**, *11*, 225–236. [[CrossRef](#)]
36. Laucelli, D.; Giustolisi, O. Scour depth modelling by a multi-objective evolutionary paradigm. *Environ. Model. Softw.* **2011**, *26*, 498–509. [[CrossRef](#)]
37. Altomare, C.; Gironella, X.; Laucelli, D. Evolutionary data-modelling of an innovative low reflective vertical quay. *J. Hydroinform.* **2013**, *15*, 763. [[CrossRef](#)]
38. Rezanian, M.; Javadi, A.A.; Giustolisi, O. Evaluation of liquefaction potential based on CPT results using evolutionary polynomial regression. *Comput. Geotech.* **2010**, *37*, 82–92. [[CrossRef](#)]
39. Giustolisi, O.; Doglioni, A.; Laucelli, D. Determination of friction factor for corrugated drains. *Proc. Inst. Civ. Eng. Water Manag.* **2008**, *161*, 31–42. [[CrossRef](#)]
40. Van der Meer, J.W.; Verhaeghe, H.; Steendam, G.J. The new wave overtopping database for coastal structures. *Coast. Eng.* **2009**, *56*, 108–120. [[CrossRef](#)]
41. Tamada, T.; Mase, H.; Yasuda, T. Random Wave Runup Formulae for Seawall with Composite Cross Section. *J. Jpn. Soc. Civ. Eng. Ser. B2 Coast. Eng.* **2009**, *65*, 936–940. [[CrossRef](#)]
42. Altomare, C.; Suzuki, T.; Verwaest, T. Influence of directional spreading on wave overtopping of sea dikes with gentle and shallow foreshores. *Coast. Eng.* **2020**, *157*, 103654. [[CrossRef](#)]
43. Altomare, C.; Gironella, X.; Suzuki, T.; Viccione, G.; Saponieri, A. Overtopping Metrics and Coastal Safety: A Case of Study from the Catalan Coast. *J. Mar. Sci. Eng.* **2020**, *8*, 556. [[CrossRef](#)]
44. Suzuki, T.; De Roo, S.; Altomare, C.; Mostaert, F. Methodology for Safety Assessment 2015: Updated Methodologies from the Report v6.0 to v10.0, Flanders Hydraulics Research, Antwerp. 2017. Available online: <http://documentatiecentrum.watlab.be/owa/imis.php?module=ref&refid=290335> (accessed on 24 July 2020).
45. Chambers, J.M.; Cleveland, W.S.; Kleiner, B.; Tukey, P.A. *Graphical Methods for Data Analysis*; Chapman and Hall/Cole Publishing Company: Wadsworth, OH, USA, 1983.
46. Bruce, T.; Pullen, T.; Allsop, W. Wave Overtopping at Vertical and Steep Structures. In *Handbook of Coastal and Ocean Engineering*; World Scientific: Singapore, 2018; pp. 411–439.
47. Schüttrumpf, H.; Van Gent, M.R. Wave overtopping at sea dikes. In Proceedings of the Coastal Structures 2003, Portland, OR, USA, 26–30 August 2003; American Society of Civil Engineers: Reston, VA, USA, 2003; pp. 431–443.
48. Rattanapitikon, W.; Shibayama, T. Breaking Wave Formulas for Breaking Depth and Orbital to Phase Velocity Ratio. *Coast. Eng. J.* **2006**, *48*, 395–416. [[CrossRef](#)]

