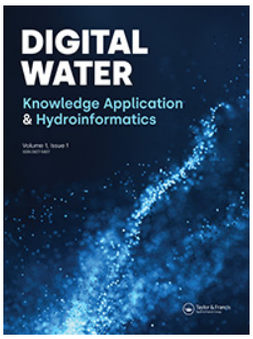


Taylor & Francis
Taylor & Francis Group

Digital Water

Knowledge Application & Hydroinformatics

ISSN: 2837-5807 (Online) Journal homepage: www.tandfonline.com/journals/tdwa20

Evaluating coastal safety using individual wave overtopping volumes: insights from evolutionary polynomial regression

Corrado Altomare, Luigi Berardi, Simone Ripani & Xavier Gironella

To cite this article: Corrado Altomare, Luigi Berardi, Simone Ripani & Xavier Gironella (2025) Evaluating coastal safety using individual wave overtopping volumes: insights from evolutionary polynomial regression, Digital Water, 3:1, 1-29, DOI: 10.1080/28375807.2025.2531945

To link to this article: https://doi.org/10.1080/28375807.2025.2531945

© 2025 The Author(s). Published with license by Taylor & Francis Group, LLC.

Published online: 22 Jul 2025.

Submit your article to this journal

Article views: 627

View related articles

View Crossmark data

Citing articles: 1 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=tdwa20

DIGITAL WATER
2025, VOL. 3, NO. 1, 1–29
https://doi.org/10.1080/28375807.2025.2531945

Routledge
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Evaluating coastal safety using individual wave overtopping volumes: insights from evolutionary polynomial regression

Corrado Altomare[a], Luigi Berardi[b], Simone Ripani[b] and Xavier Gironella[a]

[a]Maritime Engineering Laboratory, Universitat Politècnica de Catalunya BarcelonaTech, Barcelona, Spain; [b]Maritime Engineering Laboratory, Università degli Studi G. D'Annunzio, Pescara, Italy

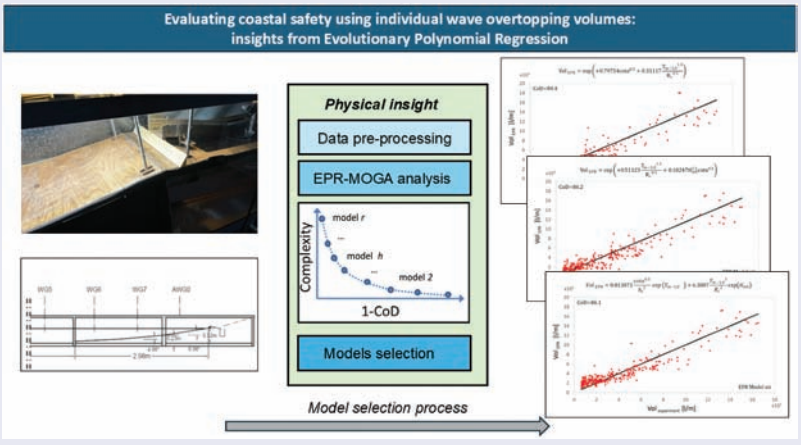
ABSTRACT

Background: Wave overtopping threatens coastal safety and is expected to worsen with sea-level rise and stronger storms. Existing Weibull-based predictors require irregular sea states and offer limited insight for flume studies. This work seeks an interpretable, data-driven alternative to estimate maximum overtopping volumes under extreme-event conditions.

Methods: A database of 316 tests from a 1:50-scale wave flume was analysed. Focused wave groups derived from NewWave theory replaced irregular spectra to target large overtopping events on shallow-water dikes with emergent toes. Hydraulic and structural inputs (wave height H, wave period T, crest freeboard Rc, local water depth h, and slope tan α) fed an Evolutionary Polynomial Regression algorithm with multi-objective genetic optimisation (EPR-MOGA) to identify compact predictive expressions.

Results: the EPR models achieve coefficients of determination (CoD) up to 89.7% with simple expressions involving three to five key variables. Results show that wave period, wave height, structural freeboard, and local water depth at the dike toe are key factors influencing overtopping volumes. The group $T_{m-1,0}/R_c$ alone accounts for 44.5% of variance in overtopping volumes, while the influence of focus phase, wave asymmetry and skewness is negligible. The model's accuracy improves slightly when including the focus location, while other factors like focus phase, wave asymmetry, and wave skewness are less significant.

Conclusion: EPR-based polynomial models provide accurate, transparent predictions of maximum overtopping volumes for experimental configurations where probabilistic Weibull methods are unsuitable. By relying on focused wave groups and a small set of readily measured variables, the approach offers a practical tool for design and risk assessment of shallow-water coastal dikes under extreme events.



ARTICLE HISTORY
Received 8 January 2025
Accepted 13 June 2025

CONTACT Corrado Altomare corrado.altomare@upc.edu Universitat Politècnica de Catalunya Barcelonatech, Calle Jordi Girona1 – 3, Edifici D1, Campus Nord, Barcelona 08012, Spain

© 2025 The Author(s). Published with license by Taylor & Francis Group, LLC.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

HIGHLIGHTS

Evaluating coastal safety using individual wave overtopping volumes: insights from Evolutionary Polynomial Regression Corrado Altomare, Luigi Berardi, Simone Ripani, Xavier Gironella

- EPR models achieve up to 89.7% CoD in predicting overtopping volumes.
- Focused wave groups replace random seas to improve repeatability and study extremes.
- Simple expressions use ≤5 variables and 2 terms with >86% accuracy.
- $T_{m-1,0}$/Rc alone explains 44.5% of overtopping volume variance.
- Variables ϕ, As, and Sk have negligible impact on model output.

1. Introduction

The accurate assessment of wave overtopping volumes is a necessary aspect in the design of coastal defences, particularly in the context of increasing flood risks driven by climate change. Coastal urban zones, especially low-lying and highly populated areas, are facing heightened vulnerability due to rising sea levels and increasing storm intensity, as evidenced by recent extreme events like Storm Gloria in January 2020 (Amores et al. 2020) and Storm Ciarán in November 2023. These episodes highlight the importance of quantifying wave overtopping not just in terms of mean discharge, but also in relation to the large overtopping volumes associated with individual sea storm. Criteria based on average flow rates are still used as a design criterion for any coastal structure. Notwithstanding it, there is ongoing debate about whether these adequately represent the real hazard posed by wave overtopping. Studies Endoh and Takahashi 1995, as well as, more recently, Sandoval and Bruce 2017, detail the limitations of using a single average value to characterise the risks. Instead, individual overtopping events may present distinct hazards, such as sweeping pedestrians off their feet or necessitating specific evacuation measures. Accurate predictions of these overtopping volumes are essential for designing effective coastal defences, as emphasised by researchers like Hughes and Thornton (2016) and Whittaker et al. (2016).

Recent studies, such as Dong et al. (2021, 2024), although specifically focussed on recurve walls and vertical structures, emphasise the role of geometrical parameters in wave overtopping. At the same time, broader reviews of coastal protection strategies underscore the potential of hybrid sea defences to integrate natural and engineered components, although there is a lack of robust data linking wave–structure interaction to overtopping behaviour (Xu et al. 2025). Machine learning methods, including Gaussian Process surrogate modelling, have also been applied successfully to overtopping prediction, offering improved accuracy and insight into variable importance across varied structure types (Kent et al. 2024). Despite these advancements, many of the existing models rely on empirical approaches tailored to irregular sea states and often exclude the dynamics of focused wave interactions that can lead to extreme individual events.

This study contributes to bridging this gap by using focused wave groups in a controlled flume environment to isolate and predict large overtopping volumes associated with extreme storm waves. By applying an Evolutionary Polynomial Regression (EPR-MOGA) framework, we aim to develop interpretable and accurate expressions for individual overtopping events, independent of Weibull assumptions and generalised empirical formulas.

In this context, the present study aims to offer simple but accurate predictive models for the characterisation of overtopping volumes associated with large wave groups within a random sea state. To achieve this, the Evolutionary Polynomial Regression (EPR) algorithm is employed (Giustolisi and Savic 2006). EPR is a data-mining tool that combines numerical and symbolic regression, enabling the formulation of a robust model for wave overtopping prediction. Data-modelling techniques are part of the digital transition process in the water sector, which is gaining importance due to the increasing availability of large amount of data available from advanced and distributed sensors at both laboratory and real scale. EPR allows exploring the physical insight of complex phenomena through the use of genetic algorithms combined with linear regressions. It offers a transparent and structured approach that retains the advantages of symbolic regression while eliminating computational inefficiencies. EPR creates simple but effective (generalisable)

models and for this reason it is applied in many fields of engineering and not only, looking at the accuracy but also the parsimony of the model. In this specific case, in the field of coastal engineering, it is applied to predict wave overtopping discharge.

This study aims to develop predictive models that are simple, interpretable, and accurate for estimating individual wave overtopping volumes generated by focused wave groups. The specific objectives are: (i) to identify the key hydraulic and structural parameters that govern overtopping under extreme conditions; (ii) to formulate transparent polynomial-based models using the Evolutionary Polynomial Regression with Multi-Objective Genetic Algorithm (EPR-MOGA); and (iii) to evaluate model performance through the coefficient of determination (CoD) (Giustolisi and Savic 2006) and criteria for model parsimony. In particular, we investigate how parameters such as crest freeboard, local wave height, wave period, and the slopes of the foreshore and dike contribute to predicting the magnitude of individual overtopping events.

The novelty of this research lies in the use of a highly controlled experimental dataset composed of focused wave group tests on smooth dikes with emergent toes in shallow water, a condition not typically covered by existing empirical or machine learning overtopping models. This approach enables the isolation of extreme overtopping events, providing detailed, high-resolution data that is both repeatable and relevant for extreme climate scenarios. Focused wave groups offer several advantages, including better repeatability and the ability to closely study large wave interactions. The NewWave theory (Whittaker et al. 2016, 2018), widely used in offshore engineering, is adopted to generate statistically probable wave groups associated with the largest waves in a sea state. This approach allows for a more detailed examination of wave–structure interactions during extreme events (Whittaker et al.. 2017), a method that has increasingly been adopted in coastal engineering studies, as demonstrated in research on coastal defences (Altomare and Gironella 2024) and wave energy converters (Ropero-Giralda et al. 2020). Ultimately, the goal of this research is to provide a reliable and efficient method for assessing wave overtopping volumes in various coastal scenarios, particularly those involving large wave groups approaching sea dikes on varying sloping foreshores, by considering focused wave properties and deep-water hydrodynamic characteristics (Yuhi et al. 2021, Lashley et al. 2021).

Furthermore, unlike conventional approaches relying on Weibull-based statistical extrapolation or data-hungry black-box models, this study produces transparent, equation-based predictors that are both physically interpretable and readily applicable in design contexts where only limited input parameters are available. As such, the findings contribute to both the methodological and practical advancement of overtopping prediction under extreme wave conditions.

The article is structured into some key sections that systematically present the research findings and methodologies. It begins with an introduction, which outlines the significance of wave overtopping and the objectives of the study. Then, the experimental campaign and collected data are described. Following this, the methodology section details the Evolutionary Polynomial Regression (EPR) approach used for modelling, including data collection and analysis techniques for the specific application in hands. The results and discussion section presents the findings, highlighting the relationships between various hydraulic and structural parameters and their impact on overtopping volumes. Finally, the article concludes with a section that summarises the main findings and suggests directions for future research.

2. Literature review of individual wave overtopping volumes

Traditionally, coastal structures have been designed using the allowable mean overtopping discharge as the primary criterion. However, recent design manuals also specify limits for the maximum individual overtopping volumes to ensure the safety of people, vehicles, and properties located behind these structures (EurOtop 2018). Individual overtopping wave events associate to one sea storm, which are random in nature, are often described by their exceedance probability distribution. This distribution is typically related to the mean overtopping discharge, the probability of overtopping, and storm duration. Empirical methods derived through dimensional analysis and scaling arguments–are widely used to estimate overtopping parameters. These methods relate overtopping responses to wave and structural parameters. One of the most comprehensive datasets for overtopping rates is the CLASH database, developed during the CLASH European project

(Pullen et al. 2009), although it primarily focuses on average overtopping rates and provides limited data on wave-by-wave analysis.

2.1. Probability distribution of individual wave overtopping volumes

It is particularly important to understand extreme individual events when assessing potential impacts, such as on pedestrian safety or the stability of the landward side of coastal structures. There are various methods for quantifying these events, including numerical modelling, machine learning techniques and empirical formulations. Unlike some other methods, empirical formulations offer the advantage of being easily applicable and of explicitly demonstrating the relationships between variables. Consequently, research often focuses on developing and refining these relationships. Traditionally, studies on individual volume distributions have primarily focused on conditions without significant wave breaking on the foreshore or limited shallow water influence (Nørgaard et al. 2014, Jorge et al. 2019). However, existing formulations derived for dikes or other structures in intermediate to deep water may not be sufficient when a shallow foreshore is present in front of a structure. The hydrodynamic processes in shallow water, such as depth-induced wave breaking and non-linear wave interactions leading to significant low-frequency wave energy, differ considerably from deep water conditions. Lately, de Ridder et al. (2025) have focused on investigating extreme individual wave overtopping events, including water layer thickness, front velocity, and volumes, at rubble mound breakwaters with a smooth crest situated in shallow water conditions. Their findings emphasise the need for new empirical expressions that are tailored to the unique characteristics of coastal structures in shallow water environments. However, they only analyse rubble mound breakwaters. To the authors' knowledge, no similar study on sea dikes in extremely shallow waters, even with an emerged toe, has yet been published.

The aforementioned gap in the literature is particularly relevant when selecting appropriate statistical models to describe overtopping events across different structure types and water depths. To describe the probability distribution of individual overtopping volumes, the two-parameter Weibull distribution is often applied. The cumulative distribution function (CDF) of the Weibull distribution is expressed as:

$$P_v = 1 - \exp\left[-\left(\frac{V}{a}\right)^b\right] \tag{1}$$

where: P_v is the probability that an individual overtopping volume V will not exceed; a is the scale parameter; b is the shape parameter, which defines the tail of the distribution and the likelihood of larger overtopping events.

2.1.1. Shape parameter formulations in overtopping studies

Several studies have provided formulations for the Weibull shape parameter b under different conditions. Early studies relied on constant shape parameters or focused on wave steepness, while later research emphasised the importance of variables such as relative crest freeboard and dike slope angle, among others. Recent studies propose more complex formulas that incorporate multiple factors and consider specific structure types and overtopping regimes (e.g. low overtopping volumes).

Franco et al. 1994 recommended a constant shape parameter $b = 0.75$ for dikes and vertical breakwaters in non-breaking wave conditions; Besley (1999) suggested a constant shape parameter b = 0.853 for sloped structures, while Bruce et al. (2009) found b = 0.74 for mound breakwaters, and determined that the type of armour and number of layers did not significantly affect b.

EurOtop (2018), recommends the following parameters for vertical breakwaters, based on Besley (1999):

2.1.1.1. Non-impulsive wave conditions.

$$b = 0.66 \quad \text{for} \quad s_{0p} = 0.02 \tag{2}$$

$$b = 0.82 \quad \text{for} \quad s_{0p} = 0.04 \tag{3}$$

where:

$$s_{0p}$$

is the deep-water incident wave steepness.

2.1.1.2. Impulsive wave conditions.

$$b = 0.85 \tag{4}$$

Victor et al. (2012) proposed the following formula of the shape parameter for steep, low-crested smooth structures:

$$b = \exp\left[-2.0\frac{R_c}{H_{m0}}\right] + (0.56 + 0.15\cot\alpha) \tag{5}$$

where

R_c is the crown wall crest freeboard, H_{m0} is the incident significant wave height, and α is the seaward dike slope angle.

Hughes et al. (2012) proposed the following formula, valid for $-2 < \frac{R_c}{H_{m0}} < 4$:

$$b = \left[\exp\left(-0.6\frac{R_c}{H_{m0}}\right)\right]^{1.8} + 0.64 \tag{6}$$

Zanuttigh et al. (2013) tatisticalCO used mean overtopping discharge to estimate the shape parameter, leading to the formulas recommended in EurOtop2018:

2.1.1.3. Smooth structures.

$$b = 0.73 + 55\left(\frac{q}{gH_{m0}T_{m-1,0}}\right)^{0.8} \tag{7}$$

2.1.1.4. Rubble mound structures.

$$b = 0.85 + 1500\left(\frac{q}{gH_{m0}T_{m-1,0}}\right)^{1.3} \tag{8}$$

where: q is the mean overtopping discharge, g is the acceleration due to gravity, and $T_{m-1,0}$ is the spectral period.

Nørgaard et al. (2014) developed formulas to estimate b for rock-armoured mound breakwaters under depth-limited breaking wave conditions:

$$b = 0.75 \text{for} \frac{H_{m0}}{H_{1/10}} \leq 0.848 \text{ and } \frac{H_{m0}}{h} \leq 0.2 \tag{9}$$

$$b = -6.1 + 8.08\frac{H_{m0}}{H_{1/10}} \text{for} \frac{H_{m0}}{H_{1/10}} > 0.848 \text{ and } \frac{H_{m0}}{h} > 0.2 \tag{10}$$

where:

$H_{\frac{1}{10}}$ is the average of the $1/10$ highest incident waves, and h is the water depth.

Gallach-Sánchez 2018 built on Victor et al. (2012), proposing the following formula:

$$b = (0.59 + 0.23\cot\alpha)\exp\left(-2.2\frac{R_c}{H_{m0}}\right) + 0.83 \tag{11}$$

A recent study from Salauddin et al. (2022) found no apparent differences in Weibull b values with changes in incident wave steepness and the impulsiveness parameter: the authors propose a new formula for

estimating b at vertical walls under both impulsive and non-impulsive wave conditions that somehow combines in two terms at the right-hand side what expressed in previously shown equations:

$$b = 0.42 + 1.53 \exp\left(-0.95 \frac{R_c}{H_{m0}}\right) + 450 \left(\frac{q}{g H_{m0} T_{m-1,0}}\right)^{1.1} \tag{12}$$

2.1.2. Scale parameter formulations in overtopping studies

The scale parameter a normalises the distribution and is generally expressed, for wave overtopping volumes, as follows:

$$a = A\overline{V} \tag{13}$$

where A is the dimensionless scale parameter and $\overline{V}$ is the mean of the measured individual wave overtopping volumes. While the Weibull shape parameter b has been extensively studied, with various formulations proposed, there is limited information on specific formulations for the scale parameter a. The focus in most studies has been on its relationship with the shape parameter and the average overtopping discharge. EurOtop (2018) provides the following equation to connect the scale parameter with the shape parameter:

$$a = \frac{1}{\Gamma(1 + \frac{1}{b})} \left(\frac{q T_m}{P_{ov}}\right) \tag{14}$$

where: Γ is the gamma function, T_m is the average period calculated from the wave record, and P_{ov} is the probability of overtopping. Equation 14 suggests that the scale parameter is proportional to the mean overtopping volume, as the right-hand side of the equation represents the theoretical mean of the Weibull distribution, which is assumed to equal the measured mean overtopping volume. Jorge *et al.* (2019), in their study of mound breakwaters with low overtopping volumes, propose a method to determine the scale parameter using a quadratic utility function. They suggest that the scale parameter is not directly derived from the mean overtopping volume but is instead obtained by fitting the Weibull distribution to measured data.

In general, it can be stated that the key variables influencing the scale parameter are:

- average overtopping discharge: a higher discharge directly leads to a larger scale parameter.
- average period: a longer mean wave period also contributes to a larger scale parameter.
- overtopping probability: a lower overtopping probability (i.e. fewer waves overtopping the structure) results in a larger scale parameter, i.e. the expected maximum volumes of events with lower P_{ov} are larger than those with higher P_{ov}.

2.1.3. Maximum individual overtopping volume

The maximum overtopping volume in one sea state (V_{max}) is calculated using the following equation:

$$V_{max} = a[\ln(N_{ow})]^{\frac{1}{b}} \tag{15}$$

where N_{ow} is the number of waves overtopping the coastal defence structure. The scale parameter a directly influences V_{max}. Consequently, changes in the variables affecting a will also impact the estimation of maximum overtopping volume. Summarising all above references, it can be concluded that i) the maximum overtopping volume is directly proportional to the mean overtopping volume; ii) the relationship between these two volumes is further influenced by parameters like the number of overtopping waves, mean overtopping discharge, probability of overtopping, relative crest freeboard, slope angle and wave steepness; iii) accurately estimating V_{max} requires a thorough understanding of these variables and their complex interactions. It would therefore be beneficial to identify a method that employs straightforward and precise predictive models for wave overtopping volumes, which do not necessitate the utilisation of Weibull distributions. This could prove to be an attractive preliminary approach to estimating the critical volume for coastal infrastructures.

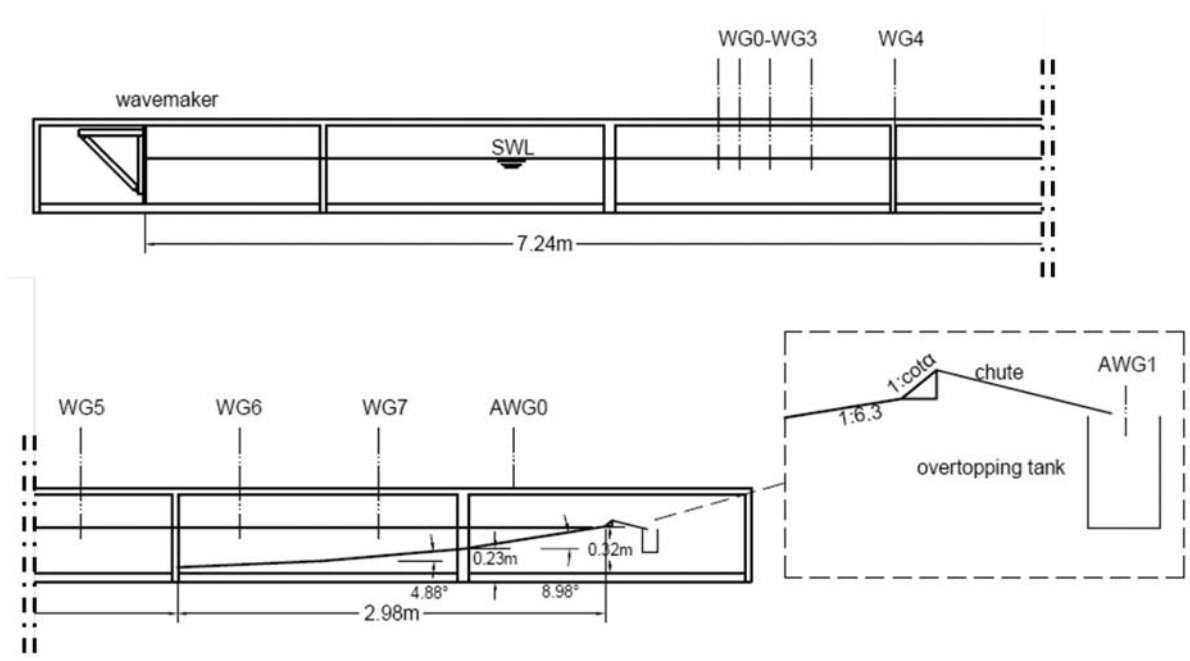


Figure 1. Ciemito 2D lateral view with detail of sea dike and overtopping box.

3. Experimental campaign

3.1. Experimental setup

Experimental tests were carried out in the small-scale wave flume, named CIEMito, at the Maritime Engineering Laboratory (LIM) of the Universitat Politècnica de Catalunya –BarcelonaTech (UPC). The flume has dimensions of 18 m in length and 0.38 m in width, and uses a piston-type wavemaker for wave generation. Active wave absorption was not used in the experimental setup. As discussed later in the manuscript, focused wave groups were employed to concentrate wave energy at a specific location and time. This approach does not require active absorption systems, as the generated time series are short in duration and wave reflections at the paddle location are negligible. Consequently, the analysis of wave–structure interactions can be performed with high precision, without the need to separate incident and reflected wave components.

The tests employed a 1:50 scale model constructed from plywood, featuring a sloping beach with a 1:15 slope beginning at 7.2 m from the wavemaker, followed by a steeper 1:6.3 slope leading to the dike or promenade toe (see Figure 1).

The structure was tested with various geometries, including a vertical wall and sloping dikes with slopes of 1:2 (vertical:horizontal) and 1:0.5, respectively. In all cases, the dike height was 0.04 m. The dike was positioned either at the end of the 1:6.3 slope or set back by 0.05 m to create a horizontal emerged berm at the dike toe. The still water level was below the structural toe, leaving the dikes exposed.

Scale effects were evaluated through Reynolds and Weber numbers, confirming that viscous and surface tension forces were negligible ($Req > 10^3$, $Weq > 10$). These findings align with the analysis by Altomare and Gironella (2024), who examined the same dataset and concluded that scale effects do not significantly influence overtopping volumes.

To measure water surface elevations, eight resistive wave gauges (WG0-WG7) and one ultrasonic proximity sensor (AWG0) were installed along the flume, at distances between 4.00 m (WG0) and 9.58 m (AWG0) from the resting wave generator. Furthermore, an overtopping tank equipped with a second ultrasonic sensor (AWG1) was employed to record the volume of overtopping. The variation in water level within the overtopping tank was then multiplied by the base area of the tank to be converted into volume. The overtopping was collected using a 0.09 m wide chute. Dividing the collected overtopping volume by the chute width, the volume per unit width (expressed in l/m) was finally assessed. For this study, we only

Table 1. Hydrodynamic and geometrical characteristics upscaled to prototype conditions.

Parameter	Value range
H_{m0} (m)	2.30–3.80
$T_{m-1,0}$ (s)	9.33–12.87
ht (m)	From −0.70 to −2.20
Rc (m)	2.55–4.20
Dike slope (cotα)	0, 0.5, 2
Skewness	0–3.70
Asymmetry	−2.15–1.05
V (l/m)	659.70–16512.56

looked at data without toe berm. This is because only a few experiments with toe berm were carried out. This meant that the total number of data points was 316. The complete dataset, including data for different foreshore slope, and all technical detail about the measurement system and setup are described in Altomare (2024).

3.2. Wave conditions and test matrix

Hughes and Thornton (2016) and Whittaker *et al.* Whittaker et al. (2018) highlighted the need to consider individual wave properties for designing coastal defences. In the present study, the NewWave theory (Tromans et al. 1991) is employed. By linking the expected shape of the maximum wave to the overall characteristics of the sea state, the theory enables a more accurate representation of extreme events. The use of focused wave groups, rather than long-duration irregular wave time series, offers several advantages: in particular, focused wave groups improve the repeatability and accuracy of experimental measurements and enhance the precision of models used to examine significant wave interactions (Hofland et al. 2014, Formentin et al. 2024). Whittaker et al. (2016) validated the effectiveness of focused wave groups in studying wave–structure interaction problems, demonstrating that a single wave group can reproduce extreme coastal responses within a specific sea state.

The time series for each focused wave group was generated according to the NewWave theory, following the method outlined by Whittaker et al. (2017), which describes the most probable shape of a large wave in a given sea state. While originally developed to describe compact wave trains on horizontal bottoms, in this study the theory was adapted to account for wave shoaling and breaking near coastal structures, following the work from Whittaker et al. (2016). A NewWave-type focused wave group consists of N infinitesimal wave components, defined as follows:

$$\eta(x,t) = A\sum_{i=1}^{N}\sigma^2 S_{\eta\eta}(\omega_i)\cos\left(k_i(x-x_f)-\omega_i(t-t_f)+\phi\right)\Delta\omega \tag{16}$$

where $S_{\eta\eta}$ is the power spectral density, t is time, ω is the angular frequency, σ is the standard deviation of the random sea state, being $\sigma^2 = \sum S_{\eta\eta}(\omega_i)\Delta\omega$, and k_i is the wavenumber of the i-th wave component with angular frequency ω_i (based on the linear dispersion relation $\omega^2 = gk\cdot\tanh(kh)$, where g is the acceleration due to gravity and h is the water depth), and x is the horizontal distance from the wave generation. All wave components come into phase at the focus location x_f and focus time t_f to form a large wave with a linear focus amplitude equal to A. A comprehensive range of focusing patterns can be attained by incorporating the phase angle ϕ of the group at focus (e.g. crest, trough, . . .). The energy concentration within the group remains unaffected by the value of ϕ and depends only on the variance σ^2. Yet, the shape of the wave, whether it is asymmetric and/or skewed, depends on the phase at focus and can influence the breaking process and, consequently, the impact exerted on the structure. All 316 data here analysed correspond to cases having overtopping volumes larger than 600 l/m (upscaled to prototype conditions): cases with lower values were excluded from the analysis to concentrate exclusively on extreme wave conditions and significant overtopping events. It is important to note that 600 l/m represents the tolerable overtopping volume limit for pedestrians, as indicated in EurOtop (2018).

Table 2. Focused wave location (expressed in prototype scale) and phase.

Parameter	Value range
$x_f(m)$ (relative to the dike toe)	32, 79, 127.5, 183
ϕ (°)	0, 90, 180, 270

In order to define the amplitude associated with the focused wave group, once the target significant wave height, $H_{m,0}$, has been defined, being equal to 4σ, the target value of A is calculated from the standard deviation, σ, if Rayleigh distribution of waves is assumed:

$$\mathrm{A} = \sqrt{2\sigma^2 \ln N} \tag{17}$$

Table 1 presents the parameters investigated in this study, with wave conditions upscaled to prototype conditions (Altomare and Gironella 2024). The significant wave height and spectral period values correspond to measurements taken at the WG0 location (i.e. close to wave generation). This approach was chosen due to the challenges in accurately determining wave conditions at the toe, particularly in cases of wave breaking or very shallow water. The table presents the values of two key parameters for describing the evolution of waves as they propagate into coastal regions: wave asymmetry and skewness. As waves approach shallower zones, their nonlinearity increases, resulting in asymmetries along both horizontal and vertical axes. In order to capture these nonlinear effects and characterise wave asymmetries, it is useful to consider higher-order moments of the free surface. Skewness (Sk) and asymmetry (As) are used to quantify these distortions, with Sk representing horizontal asymmetry and As describing vertical asymmetry. The two quantities are defined as follows:

$$Sk = \frac{\overline{\widetilde{\eta}^3}}{\left(\overline{\widetilde{\eta}^2}\right)^{3/2}}, \quad As = \frac{\overline{\widetilde{\eta}_H^3}}{\left(\overline{\widetilde{\eta}^2}\right)^{3/2}} \tag{18}$$

where $\widetilde{\eta}_H$ denotes the Hilbert transform of $\widetilde{\eta}$.

The variable h_t denotes the local water depth measured at the toe of the dike. Finally, explored value of focus location and phase of the focus wave groups are summarised in Table 2. In addition to the values presented in Table 1, the potential influence of focused wave location and phase on overtopping volumes is investigated.

4. The evolutionary polynomial regression paradigm

According to the colour-coding classification (Ljung 1999) of system identification models, the Evolutionary Polynomial Regression (EPR) is a grey-box modelling strategy for data-driven analysis, introduced in recent years (Giustolisi and Savic 2006) and successfully applied to modelling a wide range of complex engineering problems (Berardi *et al.* 2008, Javadi and Rezania 2009, Altomare et al. 2013, Vassallo et al. 2016). More specifically, the EPR-MOGA paradigm was selected to accomplish the analyses presented herein since it proved promising in examining the wave reflection of an absorbing gravity wall in Altomare et al. (2013), where similar kind of laboratory data were used. This approach allows the modelling of complex phenomena that are not described using physically based models, i.e. because they would require large amounts of field/laboratory data, often not available due to the high costs or technical difficulties. Grey-box models, as conceptual models, allow introducing elements of prior knowledge about the phenomenon (physical insight) in all phases of the modelling process: data pre-processing, search and selection of candidate covariates, settings of the expected expressions, analysis, and selection of final models. Indeed, EPR modelling strategy is conceived to identify the patterns in input–output data collected at the system under observation and returns expressions that are readily interpretable by the user. Therefore, EPR models are of direct relevance to get additional knowledge on complex phenomena, especially in technical/engineering areas where experts are used to look at explicit formulations. In brief, EPR combines the effectiveness of evolutionary search for developing the mathematical structure of symbolic expressions, through genetic algorithms, with the advantage of numerical regression for estimating model parameters, by solving a Least Squares linear

problem. Such two-steps strategy makes biunique the relationship between each mathematical structure of the developing model and its estimates. One of the general model structures that EPR can manage is reported in the following:

$$\mathbf{Y} = a_0 + \sum_{j=1}^{m} a_j \cdot (\mathbf{X}_1)^{\mathbf{ES}(j,1)} \cdot \ldots \cdot (\mathbf{X}_k)^{\mathbf{ES}(j,k)}$$

$$\cdot f\left((\mathbf{X}_1)^{\mathbf{ES}(j,k+1)} \cdot \ldots \cdot (\mathbf{X}_k)^{\mathbf{ES}(j,2k)}\right)$$

$$= a_0 + \sum_{j=1}^{m} a_j \cdot \mathbf{Z}_j \tag{19}$$

where m is the number of pseudo-polynomial terms, a_j are numerical parameters to be estimated, $\mathbf{X}_i$ are candidate explanatory variables (also defined as attributes), and $\mathbf{ES}(j, z)$ (with $z = 1, \ldots, 2k$) is the exponent of the zth input within the jth term. The function f is a function selected by the user among a set of possible alternatives (including no function selection, which means that the factor $f(\ldots)$ disappears from pseudo-polynomial terms). The exponents $\mathbf{ES}(j, z)$ are selected from a user-defined set of candidate values. It is worth noting that including 0 among the exponents allows the relevant input $\mathbf{X}_i$ to disappear in the resulting formula.

On the one hand, this model coding allows reducing the complexity of final mathematical expressions. On the other hand, it takes the general model structure and the inputs only as candidates for the final model. In fact, the EPR settings do not represent a rigid prior selection of the mathematical structures, number of variables, and parameters. These values are set by the users based on prior knowledge of the phenomenon and/or its physical interpretation.

It is worth noting that EPR mathematical structures, e.g. Equation 19, are linear with respect to their parameters, although not necessarily linear in their attributes (due to both exponents different from zero and possible selection of function f).

The EPR model coding allows keeping in the final model expressions only the most significant variables to describe the pattern in data. This, in turn, allows eliminating the less/non-influential variables or those that depend on the selected ones. In addition, it is possible to identify groups of variables that present recurrent correlations in the model expression, being the more significant to capture the phenomenon.

The search for optimal models is driven by the minimisation of the complexity of model structure, based on model coding, and the fitting to data. Indeed, consistently with Occam's razor principle, pursuing the parsimony in input variables and number of pseudo-polynomial terms in the final expressions aims at improving the interpretability, as a key requirement for model generalisation, and mitigating the risk of data overfitting. This is a common problem in black-box approaches, which do not allow for any physical/expert interpretability of the models. From such perspective, EPR represents a methodological tool for knowledge discovery, in particular to study the phenomena at the experimental scale, as in the present work. In fact, even though today there are almost no limitations in harvesting and archiving data, at the laboratory scale, the acquisition of non-significant data might represent a cost and an unjustified increase in complexity in the experimental set.

The latest version of the EPR paradigm, which is used herein, incorporates a multi-objective search (EPR-MOGA) Giustolisi and Savic (2009) in order to develop multiple models by simultaneously optimising the fitness to the training data and the parsimony of the resulting mathematical expressions, through a multi-objective genetic algorithm (David 1989), based on the Pareto dominance criterion, called OPTImized Multi-Objective Genetic Algorithm (OPTIMOGA) Laucelli and Giustolisi (2011).

Once the EPR-MOGA setting options have been defined (candidate model attributes, candidate exponents for attributes, maximum number of parameters, etc.), it returns a set of Pareto-optimal models considering up to three conflicting objective functions: (1) maximising the model accuracy for best fitting to data; (2) minimising the number of coefficients, i.e. terms of the pseudo-polynomials; and (3) minimising the number of input variables. The availability of multiple alternative model expressions allowed by the

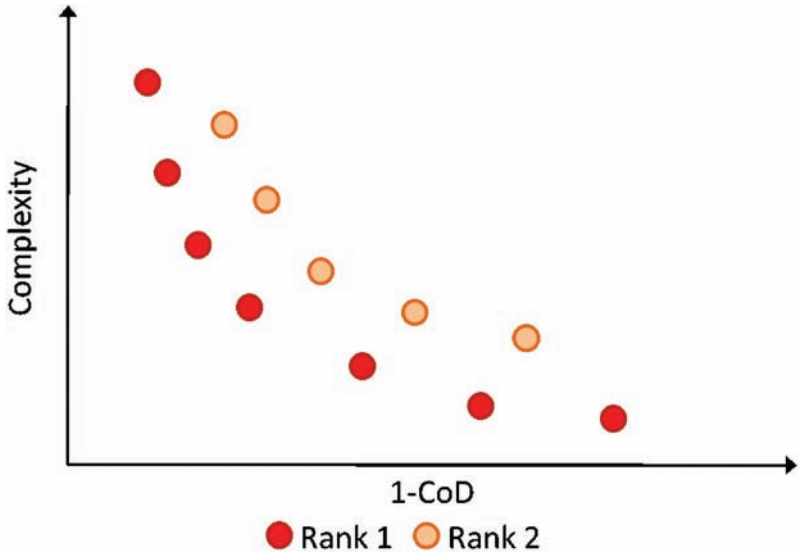


Figure 2. Representation of rank values of Pareto optimal solutions.

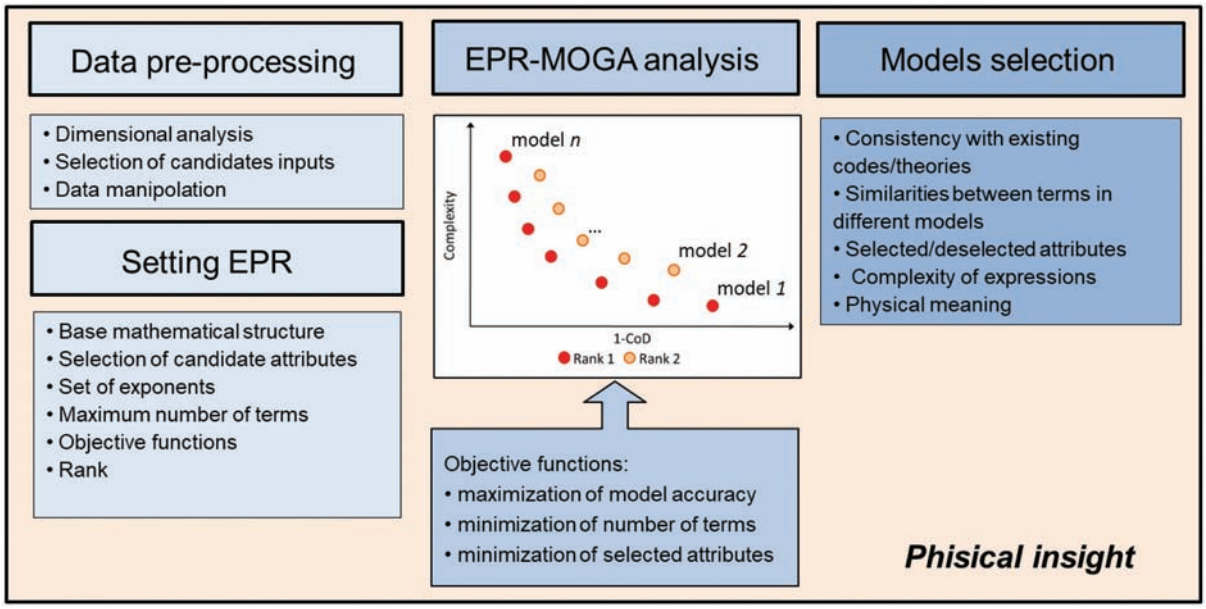


Figure 3. EPR-MOGA approach for data-driven modelling.

multi-objective search paradigm enables comparing different optimal models, identifying recursive variables or groups of variables, also in light of previous studies and knowledge of the phenomenon.

As a latest innovation in EPR-MOGA modelling, the user can get a larger range of model expressions by considering the rank-2 Pareto efficient solutions. This allows the genetic algorithm not to exclude dominated models that differ little from the solutions that are on the Pareto front. In more details, according to the methodology proposed by Fonseca and Fleming (1999), implemented in the OPTIMOGA algorithm, each solution e is assigned with a rank:

$$\text{rank}(e, \tau) = 1 + p^{(\tau)} \tag{20}$$

where $p(\tau)$ is the number of individuals dominating e, at generation τ according to the Pareto dominance criterion. Figure 2 exemplifies different solutions assigned with rank = 1 and 2. Each solution with rank > 1 is dominated by the red points in the left bottom area. This ultimately results in a more robust selection of

final models among a richer set of alternatives, preserving models that show worse statistical performances than rank 1 models, but may be useful to understand the phenomenon at hand.

Figure 3 summarises the decision support framework for model selection allowed by EPR-MOGA, emphasising its flexibility in introducing physical prior insight in all phases.

In the first phase (pre-processing data) the user can select the data that he/she considers most meaningful for the analysis, as well as manipulate them before the analysis to drive the search towards known functional terms. Furthermore, the general mathematical structure of the model, the number of variables and maximum number of pseudo-polynomial terms are based on experiences/known models and *a priori* knowledge of the phenomenon. Such setting also allows directing the search based on the purpose of the analysis. For example, minimising the number of additive terms would direct the search towards expressions consisting in the summation of single variables. This, in turn would push the search towards few groups of recurring variables. Vice versa, minimising the number of variables without constraints on number of additive terms enables a sort of sensitivity analysis to understand the influence of each single variable on the phenomenon, selected in the most parsimonious expressions. In the model selection phase, the key aspects are to obtain a generalisable and physically interpretable formulation, improving the robustness of the resulting models beyond mere statistical criteria. Therefore, recurring groups of variables and mathematical structures among the models on the Pareto front allow defining a hierarchy in the variables that explain the phenomenon while identifying those that are not meaningful or not independent form the others, i.e. never appear in the final expressions. The same principle applies to terms (i.e. parameters a_j in Equation19); for example, if the bias (a_0) not even appear in the models, it means that the variation of the output variable (Y) is explained by the input variables (Xi). Therefore, bias can be removed from next EPR-MOGA model search runs.

5. Results and discussions

5.1. Primary dependences of overtopping volumes

Prior to investigating the significant dependencies of overtopping volumes on the parameter space defined by hydraulic and structural characteristics and focused wave characteristics (see Tables 1 and 2), it is essential to examine the primary dependence of V on parameters such as crest freeboard, wave height, and wave period. This is because these parameters are expected to have an influence on the results, as widely documented in the literature. In order to achieve this, the variables of wave height and crest freeboard are

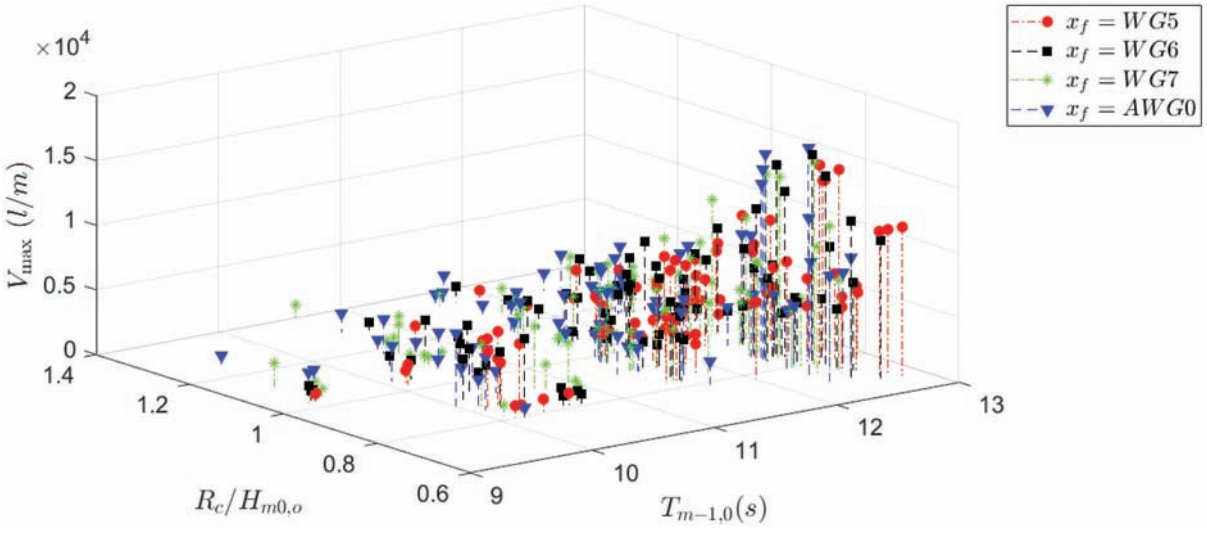


Figure 4. Overtopping volume dependence on dimensionless freeboard and spectral wave period. Results are gathered for focus location.

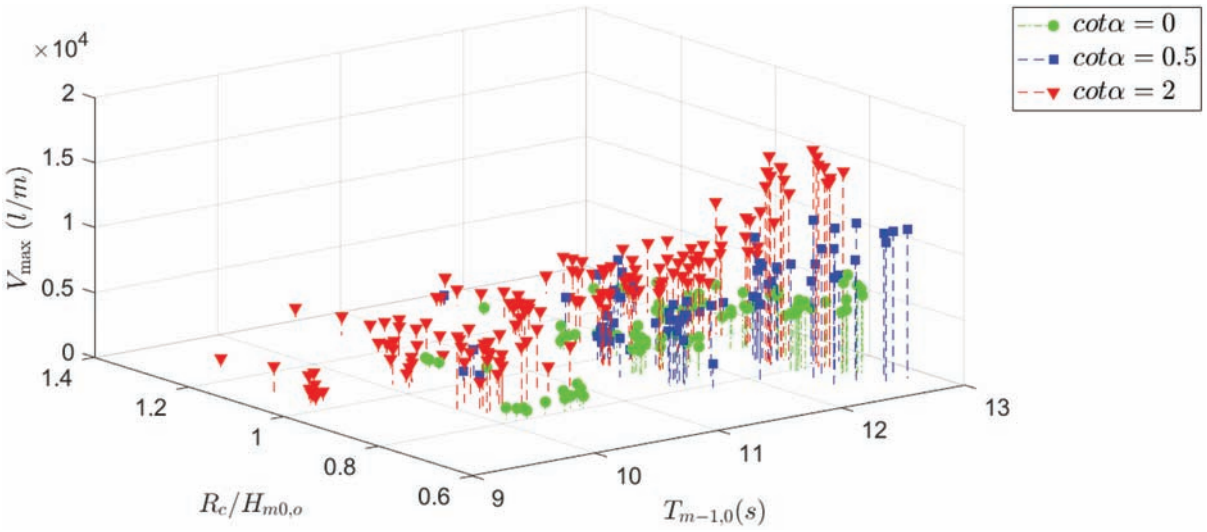


Figure 5. Overtopping volume dependence on dimensionless freeboard and spectral wave period. Results are gathered for dike slope.

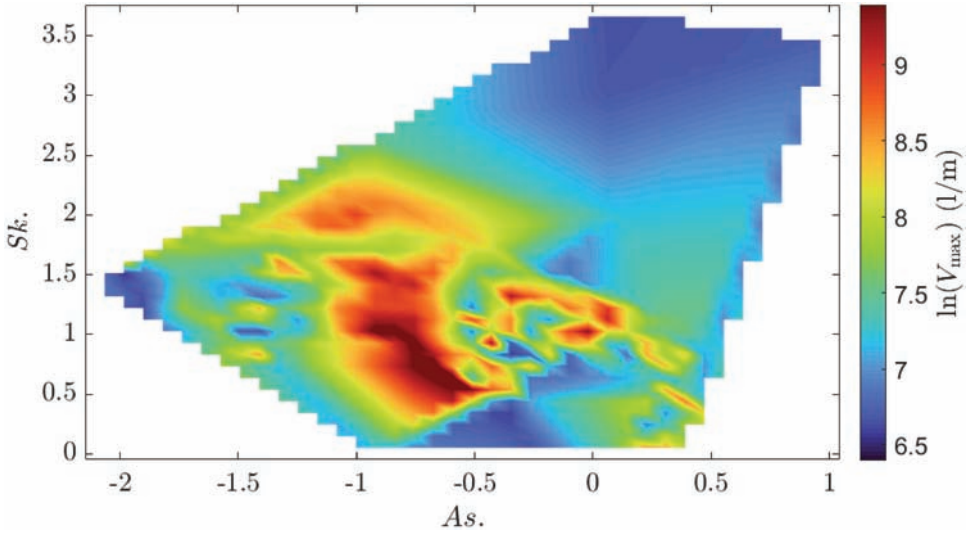


Figure 6. Overtopping volume dependence on wave skewness and asymmetry.

combined to form a dimensionless group, R_c/H_{m0}, as widely used for assessment of average wave overtopping discharge (EurOtop 2018, Lashley et al. 2021).

The variation of overtopping volumes on dimensionless crest freeboard and deep-water spectral wave period is depicted in Figures 4 and 5. Data are gathered for different focus locations, in Figure 4, while Figure 5 shows data divided per dike slope. A cursory examination of the data reveals a clear correlation between larger wave periods and larger volumes. This finding is also corroborated by (Altomare and Gironella 2024). Wave period is included in the calculation of the shape parameter (see Eqq. 8 or 12), where an increase in the wave period leads to decrease of the b, which finally will result in increase of a (see Equation 14). Besides, a is directly proportional to the wave period.

The greater the dimensionless crest freeboard, the smaller the measured overtopping volume (EurOtop 2018). The mean discharge appears in the Equation 14 and is directly related to the scale parameter of the Weibull distribution.

A more detailed examination of the data reveals that while there is no clear correlation between volumes and focused locations, at least in dimensional form. Altomare and Gironella (2024) have shown that by

considering a dimensionless focus location, proportional to the deep-water wavelength, enhances the dependence of overtopping volume on the latter. However, this correlation is not explored further here since dimensional groups will be finally considered in order to explore EPR model expressions.

Instead, there is a notable dependence on the dike slope, see Figure 5: steeper the dike, smaller the measured volume, as also confirmed by existing literature.

Finally, Figure 6 shows the variation of overtopping volumes versus wave asymmetry (*As*) and wave skewness (*Sk*). The colourmap represents the natural logarithm of the maximum overtopping volume, $\ln(V_{\max})$, highlighting regions of high and low overtopping intensity. This means that wave groups that are highly skewed and moderately asymmetric lead to the largest volumes.

While some localised peaks in overtopping volumes are observed at moderate values of asymmetry and skewness – namely for values of asymmetry between −1 and −0.5 and skewness between 0.5 and 1 –, no consistent global trend emerges from the figure. This irregular pattern suggests that, within the tested conditions, asymmetry and skewness do not exert a dominant or systematic influence on overtopping behaviour. Their role in determining overtopping volumes appears secondary compared to more influential parameters such as wave period, crest freeboard, and local water depth.

5.2. EPR setting and results

EPR has been successfully applied in coastal engineering and other engineering disciplines (Giustolisi et al. 2008, Rezania et al. 2010, Laucelli and Giustolisi 2011, Altomare et al. 2013). In this study, EPR is used to analyse the data described in Section 3, with the general mathematical structure of the models set *a priori* based on existing literature. The variables employed in the EPR models are chosen based on physical principles and derived from studies by Altomare et al. (2016), Goda (2009) and EurOtop (2018). EPR-MOGA modelling generates a set of expressions that assist in the identification of significant and recurrent groups of explanatory variables, thus facilitating the interpretation of the physical phenomenon and model comparisons in terms of accuracy and simplicity.

The analysis was conducted using two databases: one with dimensionless variables and the other with dimensional variables. In both cases, the phase angle was expressed in radians rather than degrees. For clarity, the results from the dimensionless variable database are not shown here, because EPR models using dimensionless variables showed lower accuracy than those with dimensional variables (i.e. Coefficient of Determination lower than 55%). Therefore, only the results for dimensional variables are presented and discussed, with the identified models achieving a CoD between 80% and 90%. The 316 data related to hydraulic and geometrical parameters have been used as training set for EPR-MOGA modelling. The explanatory variables included in the analysis and candidate covariates in the model expressions are spectral wave height and period at the wave generation (≈ deep waters), focus location and phase, local water depth at the dike toe, dike slope, structural freeboard, wave asymmetry and skewness. Given the prevalence of exponential and Weibull functions in the literature on wave overtopping (see section 1), an exponential function hypothesis has been retained for EPR modelling.

In order to determine the model expression, a number of hypotheses have been explored. These include the structure of the expression (for example, an exponential function, as previously explained), the maximum number of terms, the values of exponents to be used, the number of

Table 3. EPR-MOGA settings.

EPR-MOGA-XL setting	Value
Expression structure	$(1)Y = \exp(\sum(a_i X_1 X_2) + a_0), (2)Y = \sum(a_i \cdot X_1 \cdot X_2 \cdot f(X_1 \cdot X_2)) + a_0$
Inner function	No function, Exponential
Regression type	I-O Regression (no Time)
Max number of terms	3
Exponents	0, 0.5, 1, 1.3, 1.5, 2, −0.5, −1, −1.3, −1.5, −2
Number of Attributes	9
Attributes	$x_f; \phi; H_{m0}; T_{m-1,0}; h_t; \cot\alpha; A_s; S_k; R_c$
Soft ranking	2
Bias (a_0)	YES, NO
Regression Method	LSaj > 0, LS
Optimization Strategy	Min (a_j, X_i, SSE)

Table 4. EPR selected model equations and CoD.

Model no.	CoD (%)	Model expressions
i	84.4	$\exp\left(+0.79754\cot\alpha^{0.5}+0.31117\frac{T_{m-1,0}^{1.5}}{R_c^{0.5}}\right)$
ii	86.2	$\exp\left(+0.51123\frac{T_{m-1,0}^{1.3}}{R_c^{0.5}}+0.18247{H_{m0}}^{1.3}\cot\alpha^{0.5}\right)$
iii	89.7	$\exp\left(+0.50977\frac{T_{m-1,0}^{1.3}}{R_c^{0.5}}+5.9\frac{H_{m0}^{1.3}\cot\alpha^{0.5}}{T_{m-1,0}^{1.3}}-0.049148\frac{x_f\cot\alpha^{0.5}R_c^2}{H_{m0}^2T_{m-1,0}^{1.3}}\right)$
iv	44.5	$\exp\left(+0.55613\frac{T_{m-1,0}^{1.3}}{R_c^{0.5}}\right)$
v	84.1	$\exp\left(+0.79754\cot\alpha^{0.5}+0.31117\frac{T_{m-1,0}^{1.5}}{R_c^{0.5}}\right)$
vi	88.6	$\exp\left(+6.613\frac{h_t}{H_{m0}^{1.3}}+0.70754\cot\alpha^{0.5}+0.7668T_{m-1,0}\right)$
vii	86.2	$\exp\left(+0.51123\frac{T_{m-1,0}^{1.3}}{R_c^{0.5}}+0.18247{H_{m0}}^{1.3}\cot\alpha^{0.5}\right)$
viii	84.4	$\exp\left(+0.78459\cot\alpha^{0.5}+0.54905\frac{T_{m-1,0}^{1.3}}{R_c^{0.5}}-0.54416\right)$
ix	83.7	$\exp\left(+0.76988\cot\alpha^{0.5}+0.2972\frac{T_{m-1,0}^{1.5}}{R_c^{0.5}}+0.35189\right)$
x	86.6	$\exp\left(+0.51039\frac{T_{m-1,0}^{1.3}}{R_c^{0.5}}+2.0708\frac{H_{m0}^2\cot\alpha^{0.5}}{T_{m-1,0}^{1.3}}+0\right)$
xi	84.9	$+0.013095\frac{\cot\alpha^{0.5}}{h_t^2}\exp T_{m-1,0}+2228.3051h_t+493.9423H_{m0}^2$
xii	86.1	$+0.013071\frac{\cot\alpha^{0.5}}{h_t^2}\exp T_{m-1,0}+6.3087\frac{T_{m-1,0}^2}{R_c^2}\exp H_{m0}$
xiii	79.4	$+0.015\frac{\cot\alpha^{0.5}}{h_t^2}\exp T_{m-1,0}+86.8901\exp H_{m0}$
xiv	86.5	$+5.1055\frac{\cot\alpha^{0.5}}{h_t^2}\exp H_{m0}^{0.5}T_{m-1,0}^{0.5}+0.11547\frac{1}{R_c^2}\exp T_{m-1,0}$
xv	87.8	$+0.011976\frac{\cot\alpha^{0.5}}{h_t^2}\exp T_{m-1,0}+2767.188h_t+32.0108H_{m0}^{1.5}\exp T_{m-1,0}^{0.5}$

attributes (that is to say, dimensional variables), the presence or absence of a bias in the formula, the regression method (i.e. Least Squares (LS) or Non-negative Least Squares (LSaj > 0), and the optimisation strategy (i.e. minimisation of the Sum of Squared Errors (SSE) along with the minimisation of the number of variables (a_j) or the number of pseudo-polynomial terms (X_i)). Keeping lower the maximum number of terms, prevents from overfitting due to over-parameterisation of the model. Non-negative Least Squares (LSaj > 0) method for parameter estimate prevents from having negative parameters to explain inverse dependence between inputs and outputs; if LSaj > 0 is selected, inverse dependence is described by negative exponents only. Table 3 summarises the EPR-MOGA settings that have been used in different simulations and resulted in the best model performance.

A subset of 15 model expressions, selected from a total of 176, is presented in Table 4. The values of CoD are reported, resulting overall very similar to each other, except for model *iv*. Figures of each model showing the predicted volume versus the measured one are reported in Appendix A. The remaining 161 model expressions have been excluded on the grounds of complexity, specifically the presence of an excessive number of terms or groups is likely to promote overfitting to data rather than explaining the most significant patterns behind the physical phenomenon. Additionally, the groups of variables have been deemed inconsistent with the physical explanation of the phenomena, as supported by the available literature. Among the selected model expressions, six subgroups are identified, the first four without using an inner function in the general expression (1): models (i), (ii) and (iii) employing no bias and LS regression method; models (iv), (v), (vi) and (vii) without bias and with LSaj > 0; model (viii) with bias and LS regression method; models (ix) and (x) employing bias and LSaj > 0. The other two subgroups use expression structure (2) with inner exponential function: models (xi), (xii) and (xiii) having no bias and LSaj > 0; models (xiv), and (xv) with and and LSaj > 0.

It is noteworthy that models (i) and (ii) are identical to models (v) and (vii), respectively. These models are intentionally repeated despite slight differences in the EPR settings, as their repetition underscores the importance of certain explanatory groups and structural patterns.

Model (i) (= model (v)), includes only three variables and is able to describe the 84% of the phenomenon. The models that represent the best compromise between accuracy and simplicity are the models (vii) and (xii), for the expression structures (1) and (2), respectively. These models show that from four to five variables ($\cos(\alpha)$, H_{m0}, $T_{m-1,0}$, and R_c, h_t) can describe approximately 86% of the phenomenon with only two additive terms. By including the variable x_f, like in model (iii), there is a marginal improvement, raising

the explained variance to 89.7%, but an increase in complexity. Notably, the group $\frac{T_{m-1,0}}{R_c}$ alone explains 44.5% of the phenomenon (as seen in model (iv)): this group can be found in 10 of the 15 selected model expressions.

The models generally show an increasing volume with increasing wave period, as shown in Figures 4 and 5 and also evident from literature (see discussion in the previous section). There is an inverse dependence of the volumes on the crest freeboard and the local water depth: the former is expected, as smaller freeboards lead to higher overtopping rates and hence higher volumes (see Equation 14). Actually, in an overtopping Weibull distribution, both the scale parameter and the shape parameter affect the distribution of overtopping volumes, albeit with differing effects. In fact, an increase in the scale parameter directly results in an overall increase in the magnitude of the overtopping volumes, as discussed in the previous section (see also Eq. 13 and 14). Conversely, a reduction in the shape parameter results in a skewing of the distribution, leading to the emergence of a heavier tail. A reduction in the shape parameter increases the probability of extreme events but does not uniformly scale all volumes in the same manner as the scale factor. In general, it can be observed that an increase in the scale parameter has a more direct and proportional effect on the magnitude of all overtopping volumes, including the maxima. Decreasing the shape parameter increases the likelihood of extreme events by making the tail heavier, but its effect on maximum volumes is less direct compared to the scale factor. Therefore, the inverse relationship between V and crest freeboard, as well as the direct relationship between V and slope, observed in the EPR model, are consistent with the increase of the scale parameter of the Weibull distribution, which is directly proportional to mean overtopping rates. The direct relationship between V and $\cot\alpha$ is evident in all model expressions, except (iv). Higher values of $\cot\alpha$ correspond to gentler dike slopes, which are more susceptible to overtopping compared to steep vertical walls ($\cot\alpha = 0$).

The influence of the local water depth at the toe of the structure needs to be treated carefully. In fact, the inverse relationship between volumes and square of the water depth has a physical meaning only in the case of negative water depth (i.e. emergent toe), as in the cases used to train the EPR. Shallower water depths below the structural toe would lead to larger values of h_t^2. Under such conditions, wave breaking is more intense, so more wave energy is dissipated before reaching the structure and overtopping is reduced. Therefore, if model (xii) is used to calculate the overtopping volume, it should only be used for cases with emergent toe conditions.

Finally, all the analyses confirm that the variables ϕ, As and Sk are less significant than the others in explaining the observed phenomenon. Therefore, they were neglected in the following analyses.

6. Methodological workflow: data, processing, and modelling

To enhance the interpretability and reproducibility of the modelling process, a step-wise methodological flow is outlined below. This includes the dataset characteristics, pre-processing procedures, modelling techniques, and post-processing steps, with emphasis on the software used.

6.1. Dataset collection

- **Experimental Source**: 316 laboratory tests conducted at the CIEMito wave flume, UPC.
- **Wave Input**: Focused wave groups based on NewWave theory.
- **Parameters Recorded**: Spectral wave height H_{m0}, spectral period $T_{m-1,0}$, crest freeboard R_c, dike slope $\cot\alpha$, local toe depth h_t, focus location x_f, focus phase ϕ, wave asymmetry (As), and skewness (Sk).

6.2. Pre-processing

- Data filtering to retain only extreme overtopping events ($V > 600$ l/m).
- Creation of both dimensional and dimensionless datasets.

- Variable transformation using physically motivated exponents.

6.3. Modelling via EPR-MOGA

- **Modelling Tool**: Evolutionary Polynomial Regression (EPR) with Multi-Objective Genetic Algorithm (MOGA), using the EPR-MOGA-XL software.
- **Objectives**: Maximise accuracy (CoD), minimise complexity and/or number of terms and variables.
- **Expression Forms**: Symbolic pseudo-polynomial models.
- **Candidate Variables**: H_{m0}, $T_{m-1,0}$, h_t, R_c, $\cot\alpha$, x_f, ϕ, As, Sk.

6.4. Model selection and validation

- Top models achieved CoD of selected models ranging from 79.4% to 89.7%.
- In the selected models, the best trade-off encompasses two terms and four variables.
- Key group: $T_{m-1,0}/R_c$ explained 44.5% of the variance.

6.5. Post-processing and interpretation

- Most significant variables: $T_{m-1,0}$, R_c, $\cot\alpha$, h_t.
- Negligible influence: ϕ, Asymmetry, Skewness.
- Consistency with physical expectations and prior literature.

6.6. Software and tools

- **EPR-MOGA-XL**: Developed by Prof. O. Giustolisi.
- **Computational Environment**: MATLAB.
- **Visualisation**: MATLAB-based plotting and LaTeX for documentation.

7. Conclusions

Accurate estimation of wave overtopping maximum volumes is essential for the effective design of coastal defences. Existing manuals, such as EurOtop (2018), provide criteria based on average discharge rates, yet there is growing concern that these may not fully capture the risks posed by extreme overtopping events. This study sought to enhance the understanding of extreme overtopping events (i.e. individual overtopping volumes) for given sea states and structural layouts by employing the Evolutionary Polynomial Regression (Giustolisi and Savic 2009).

The proposed EPR-MOGA approach provides a flexible and insightful data-driven modelling process, enabling researchers to refine their models based on prior knowledge and the significance of variables.

This study contributes to the literature by introducing a novel application of EPR-MOGA to model overtopping volumes using a focused wave group dataset. Unlike traditional approaches that rely on irregular sea states and statistical distributions, the proposed method isolates deterministic extreme events in a controlled flume environment, allowing for high-resolution analysis.

To this end, 316 tests from an experimental campaign carried out in a wave flume facility were analysed and used to train the EPR model. Four different dike slopes with emergent toe were studied, with slopes varying from 1:2 to vertical walls. Focused wave groups were generated rather than irregular wave trains, allowing the experimental campaign to be optimised and to concentrate on the largest wave groups in a real sea state, which would result in the largest or one of the largest individual wave overtopping volumes. Focused location and focus phase have been considered among other explanatory variables, such as deep-water spectral wave height and period, local water depth, crest freeboard and wave asymmetry and skewness.

The study confirms that overtopping volumes are influenced by several physical parameters, including wave period, crest freeboard, and local water depth. Lower freeboards are associated with higher overtopping rates, and the relationship between water depth and overtopping is significant, particularly in cases with emergent toe conditions.

The inclusion of additional variables in the EPR models leads to a marginal improvement in explained variance. Although the complexity of the models increases, their predictive capabilities are enhanced. Specifically, the group $T_{m-1,0}/R_c$ significantly contributes to explaining the phenomenon, accounting for 44.5% of the variance in overtopping volumes.

The analysis confirms that gentler dike slopes (higher values of cot α) are more susceptible to overtopping compared to steeper slopes.

Certain variables, such as ϕ, A_s, and S_k, were found to be less significant in explaining overtopping volumes and can be excluded from further analyses, thereby streamlining the modelling process.

It is important to acknowledge that the EPR models developed in this study are based on a specific experimental dataset involving focused wave groups acting on smooth dikes with an exposed toe in shallow water. While this setup enabled the models to be tested repeatedly and in great detail, it also limits their general applicability. Future research should therefore consider expanding the dataset to include a wider range of structural configurations, water depths and wave conditions, with the aim of enhancing the robustness and applicability of the models. Furthermore, although this study concentrated on creating parsimonious and interpretable models that did not rely on probabilistic distributions, future studies could profit from directly comparing their predictive performance with that of empirical or Weibull-based methods (at present still lacking for the studied geometry and hydrodynamic conditions), thereby further highlighting the advantages of the proposed approach.

The findings of this study offer valuable insights for advancing coastal resilience. By developing simple and interpretable predictive models for individual wave overtopping volumes—based on focused wave group experiments—this work provides a practical tool for coastal engineers and planners to conduct preliminary assessments of extreme wave–structure interactions. This is particularly important for the design and evaluation of coastal infrastructure in low-lying, vulnerable regions where overtopping hazards are expected to intensify due to sea level rise and increased storm activity. The ability to isolate and quantify the influence of key parameters—such as wave period, crest freeboard, and local water depth—enables more targeted and adaptive design strategies. For example, the proposed models can support the refinement of design thresholds for sea walls, dikes, and hybrid defences by focusing on individual peak overtopping events, rather than relying solely on average discharge values. This is especially beneficial in contexts where empirical formulations developed for irregular sea states are not applicable. Overall, the methodology presented here contributes to more robust and evidence-based approaches in coastal protection planning.

To enhance the robustness and generalisability of the proposed models, future studies should consider training the EPR-MOGA framework on more diverse datasets, including different structure types, water depths, and wave conditions, for which this modelling approach is particularly well suited.

In conclusion, the application of EPR-MOGA modelling has enabled the development of transparent and accurate predictive expressions for overtopping volumes, without relying on Weibull-based statistical assumptions that require estimation of shape and scale parameters. These EPR-derived models are parsimonious and flexible and may be extended to other dike geometries and wave conditions, provided they remain within the range of the training dataset. As such, this work lays the groundwork for broader, physically interpretable modelling approaches in overtopping research and practice.

Acknowledgments

The authors would like to express their gratitude to Professor O. Giustolisi, who kindly supplied the EPR MOGA-XL tool, and to Andrea Marzeddu and Quim Sospedra, who provided invaluable technical support during the test campaign.

Author contributions

Corrado Altomare: Conceptualisation, Methodology, Formal analysis, Investigation, Writing original draft, Writing – review & editing, Funding Acquisition, Visualisation. Xavi Gironella: Supervision, Funding Acquisition, Review & editing. Simone Ripani: Software, Investigation, Writing, review & editing. Luigi Berardi: Resources, Software, Investigation, Writing –review & editing, Visualisation.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

Dr C. Altomare acknowledges funding from the Spanish government and the European Social Fund (ESF) under the *Ramón y Cajal 2020* programme [RYC2020-030197-I/AEI/10.13039/501100011033] and within the GLORIA project [PID2020-115030RJ-I00] titled *Adquirir conocimientos sobre el riesgo de rebase para las zonas costeras urbanizadas* [MCIN/AEI/10.13039/501100011033].

ORCID

Corrado Altomare http://orcid.org/0000-0001-8817-0431
Xavier Gironella http://orcid.org/0000-0002-8862-5704

Data availability statement

Data description is available at: https://doi.org/10.34810/data1818. The full set of acquired data will be available upon request

Declaration of generative AI and AI-assisted technologies in the writing process

In the course of preparing the present work, the author(s) employed the DeepL Write tool to verify grammar and spelling. Following the utilisation of this tool/service, the author(s) undertook a thorough review and editing of the content, assuming full responsibility for the publication's content. Additionally, ChatGPT was employed to assist with formatting in LaTeX.

References

Altomare, C., *et al.* 2016. Wave overtopping of sea dikes with very shallow foreshores. *Coastal engineering*, 116, 236–257. doi:10.1016/j.coastaleng.2016.07.002.

Altomare, C., 2024. Xavier Gironella i Cobos, and Joaquim Sospedra i Iglesias. GLORIA-CIEMito. doi:10.34810/data1818.

Altomare, C. and Gironella, X., 2024. Characterization of overtopping volumes from focused wave groups over smooth dikes with an emerged toe: insights from physical model tests. *Journal of marine science and engineering*, 12 (7). doi:10.3390/jmse12071143.

Altomare, C., Gironella, X., and Laucelli, D., 2013. Evolutionary data-modelling of an innovative low reflective vertical quay. *Journal of hydroinformatics*, 15, 763. doi:10.2166/hydro.2012.219.

Amores, A., *et al.* 7 2020. Coastal impacts of storm gloria (January 2020) over the North-Western Mediterranean. *Natural hazards and earth system sciences*, 20, 1955–1968. doi:10.5194/nhess-20-1955-2020.

Berardi, L., *et al.* 5 2008. Development of pipe deterioration models for water distribution systems using EPR. *Journal of hydroinformatics*, 10 (3), 265–265. doi:10.2166/hydro.2008.012.

Besley, P., 1999. *Overtopping of seawalls: design and assessment manual*. Great Britain: R & D technical report. Environment Agency, WRc Environment, and HR Wallingford (Firm). https://books.google.es/books?id=flbVMgEACAAJ.

Bruce, T., *et al.* 2 2009. Overtopping performance of diff t armour units for rubble mound breakwaters. *Coastal engineering*, 56, 166–179. doi:10.1016/j.coastaleng.2008.03.015.

David, E.G., 1989. *Genetic algorithms in search, optimization, and machine learning*. New York: Addison-Wesley.

de Ridder, M.P., *et al.* 2025. Individual overtopping volumes, water layer thickness and front velocities at rubble mound breakwaters with a smooth crest in shallow water. *Coastal engineering*, 198, 104701. doi:10.1016/j.coastaleng.2025.104701.

Dong, S., *et al.* 2021. Spatial distribution of wave-by-wave overtopping behind vertical seawall with recurve retrofitting. *Ocean engineering*, 238, 109674. doi:10.1016/j.oceaneng.2021.109674.

Dong, S., *et al.* 2024. Improved prediction of wave overtopping rates at vertical seawalls with recurve retrofitting. *Ocean engineering*, 302, 117647. doi:10.1016/j.oceaneng.2024.117647.

Endoh, K. and Takahashi, S., 1995. *Numerically modeling personnel danger on a promenade breakwater due to overtopping waves*. 1016–1029. Reston, Virginia, USA: ASCE. doi:10.1061/9780784400890.075.

Fonseca, C. and Fleming, P., 1999. Genetic algorithms for multiobjective optimization: formulation discussion and generalization. *The fifth international conference on genetic algorithms*, 93, 02.

Formentin, S.M., *et al.* 6 2024. Image clustering for overtopping volume measurements. *Physics of fluids*, 36 (6), 065137. doi:10.1063/5.0207486.

Franco, L., de Gerloni, M., and van der Meer, J.W., 1994. *Wave overtopping on vertical and composite breakwaters*. 1030–1045. Reston, Virginia, USA: ASCE. doi:10.1061/9780784400890.076.

Gallach-Sánchez, D., 7 2018. *Experimental study of wave overtopping performance of steep low-crested structures*. Thesis (PhD). Ghent, Belgium: Ghent University.

Giustolisi, O., Doglioni, A., and Laucelli, D., 2008. Determination of friction factor for corrugated drains. *Proceedings of the institution of civil engineers: Water management*, 161, 31–42. doi:10.1680/wama.2008.1611.31.

Giustolisi, O. and Savic, D.A., 7 2006. A symbolic data-driven technique based on evolutionary polynomial regression. *Journal of hydroinformatics*, 8, 207–222. doi:10.2166/hydro.2006.020b.

Giustolisi, O. and Savic, D.A., 7 2009. Advances in data-driven analyses and modelling using epr-moga. *Journal of hydroinformatics*, 11, 225–236. doi:10.2166/hydro.2009.017.

Goda, Y., 4 2009. Derivation of unifi wave overtopping formulas for seawalls with smooth, impermeable surfaces based on selected clash datasets. *Coastal engineering*, 56, 385–399. doi:10.1016/j.coastaleng.2008.09.007.

Hofland, B., Wenneker, I., and Van Steeg, P., 2014. Short test durations for wave overtopping experiments. *CoastLab 2014* Varna, Bulgaria. 1–10.

Hughes, S.A., *et al.* Oct 2012. Improvements in describing wave overtopping processes. *Coastal engineering proceedings*, 1 (33), 35. doi:10.9753/icce.v33.waves.35.

Hughes, S.A. and Thornton, C.I., 2016. Estimation of time-varying discharge and cumulative volume in individual overtopping waves. *Coastal engineering*, 117, 191–204. doi:10.1016/j.coastaleng.2016.08.006.

Javadi, A.A. and Rezania, M., Mar 2009. Applications of artificial intelligence and data mining techniques in soil modeling. *Geomechanics and geoengineering*, 1 (1), 53–74. doi:10.12989/gae.2009.1.1.053.

Jorge, M., *et al.* 2019. Distribution of individual wave overtopping volumes on mound breakwaters. *Coastal engineering*, 149, 15–27. doi:10.1016/j.coastaleng.2019.03.006.

Kent, P., *et al.* 2024. Resilient coastal protection infrastructures: probabilistic sensitivity analysis of wave overtopping using gaussian process surrogate models. *Sustainability*, 16 (20). doi:10.3390/su16209110.

Lashley, C.H., *et al.* 2021. Formulating wave overtopping at vertical and sloping structures with shallow foreshores using deep-water wave characteristics. *Journal of waterway, port, coastal, and ocean engineering*, 147 (6), 04021036. doi:10.1061/(ASCE)WW.1943-5460.0000675.

Laucelli, D. and Giustolisi, O., 2011. Scour depth modelling by a multiobjective evolutionary paradigm. *Environmental modelling & software*, 26, 498–509. doi:10.1016/j.envsoft.2010.10.013.

Ljung, L., 1999. *System identifi ation: theory for the user*. Prentice Hall information and system sciences series. Prentice Hall PTR. New Jersey, U.S.A. ISBN 9780136566953. https://books.google.es/books?id=nHFoQgAACAAJ.

Nørgaard, J.Q.H., Andersen, T.L., and Burcharth, H.F., 2014. Distribution of individual wave overtopping volumes in shallow water wave conditions. *Coastal engineering*, 83, 15–23. doi:10.1016/j.coastaleng.2013.09.003.

Pullen, T., *et al.* 2009. Field and laboratory measurements of mean overtopping discharges and spatial distributions at vertical seawalls. *Coastal engineering*, 56 (2), 121–140. doi:10.1016/j.coastaleng.2008.03.011.

Rezania, M., Javadi, A.A., and Giustolisi, O., 2010. Evaluation of liquefaction potential based on cpt results using evolutionary polynomial regression. *Computers and Geotechnics*, 37, 82–92. doi:10.1016/j.compgeo.2009.07.006.

Ropero-Giralda, P., *et al.* 2020. Efficiency and survivability analysis of a point-absorber wave energy converter using dualsphysics. *Renewable energy*, 162, 1763–1776. doi:10.1016/j.renene.2020.10.012.

Salauddin, M., *et al.* 9 2022. New insights in the probability distributions of wave-by-wave overtopping volumes at vertical breakwaters. *Scientific reports*, 12, 16228. doi:10.1038/s41598-022-20464-5.

Sandoval, C. and Bruce, T., 2017. *Wave overtopping hazard to pedestrians: video evidence from real accidents*. 501–512. London, United Kingdom: ICE. doi:10.1680/cmsb.63174.0501.

Tromans, P.S., Anaturk, A.R., and Hagemeijer, P., 1991. A new model for the kinematics of large ocean waves-application as a design wave. In: *Volume All Days of International Ocean and Polar Engineering Conference*. Edinburgh, United Kingdom, 8. ISOPE–I–91–154.

Van der Meer, J.W., Allsop, N.W.H., Bruce, T., De Rouck, J., Kortenhaus, A., Pullen, T., Schuttrumpf, 2018. *EurOtop-Manual on wave overtopping of sea defences and related structures. An overtopping manual largely based on European research, but for worldwide application*. Available from: www.overtopping-manual.com.

Vassallo, R., *et al.* 4 2016. Relationships between rain and displacements of an active earthfl w: a data-driven approach by eprmoga. *Natural hazards*, 81, 1–16. doi:10.1007/s11069-015-2140-9.

Victor, L., van der Meer, J.W., and Troch, P., 6 2012. Probability distribution of individual wave overtopping volumes for smooth impermeable steep slopes with low crest freeboards. *Coastal engineering*, 64, 87–101. doi:10.1016/j.coastaleng.2012.01.003.

Whittaker, C.N., *et al.* 2016. The average shape of large waves in the coastal zone. *Coastal engineering*, 114, 253–264. doi:10.1016/j.coastaleng.2016.04.009.
Whittaker, C.N., *et al.* 3 2017. Optimisation of focused wave group runup on a plane beach. *Coastal engineering*, 121, 44–55. doi:10.1016/j.coastaleng.2016.12.001.
Whittaker, C.N., *et al.* 2018. Extreme coastal responses using focused wave groups: over-topping and horizontal forces exerted on an inclined seawall. *Coastal engineering*, 140, 292–305. doi:10.1016/j.coastaleng2018.08.004.
Xu, X., *et al.* 2025. Advances in understanding the challenges and opportunities of hybrid sea defence approaches for coastal resilience. *Environmental challenges*, 19, 101130. doi:10.1016/j.envc.2025.101130.
Yuhi, M., *et al.* 1 2021. Refinement of integrated formula of wave overtopping and runup modeling. *Ocean engineering*, 220, 108350. doi:10.1016/j.oceaneng.2020.108350.
Zanuttigh, B., van der Meer, J., and der Meerder Meer, V., 2013. Statistical characterisation of extreme overtopping wave volumes. Available from: https://api.semanticscholar.org/CorpusID:218514566.

Appendix A. EPR models

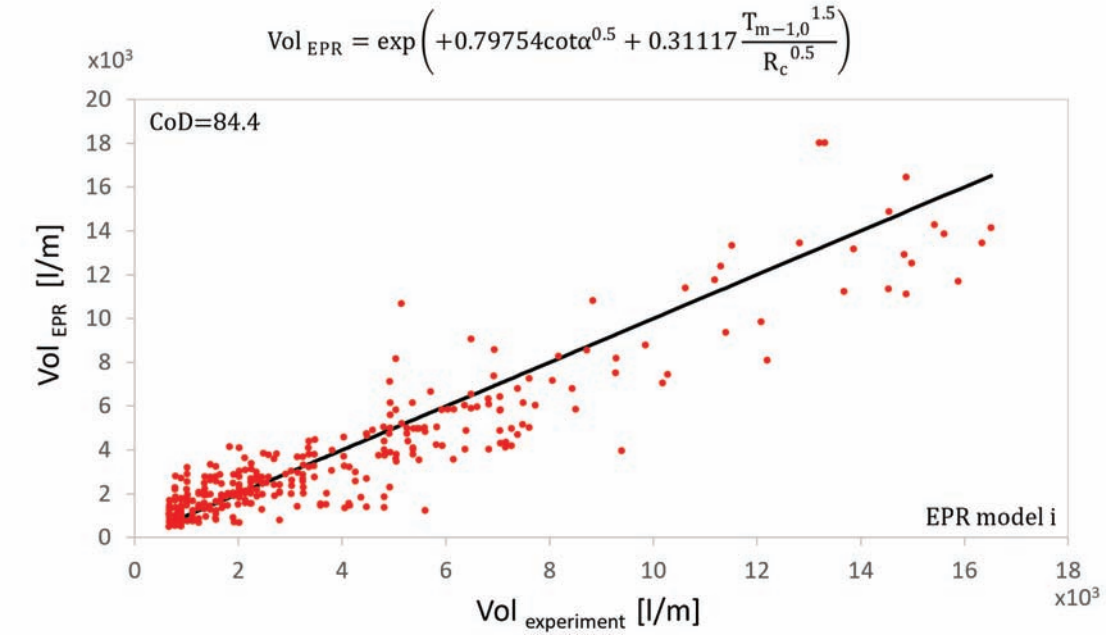


Figure A1. EPR model i.

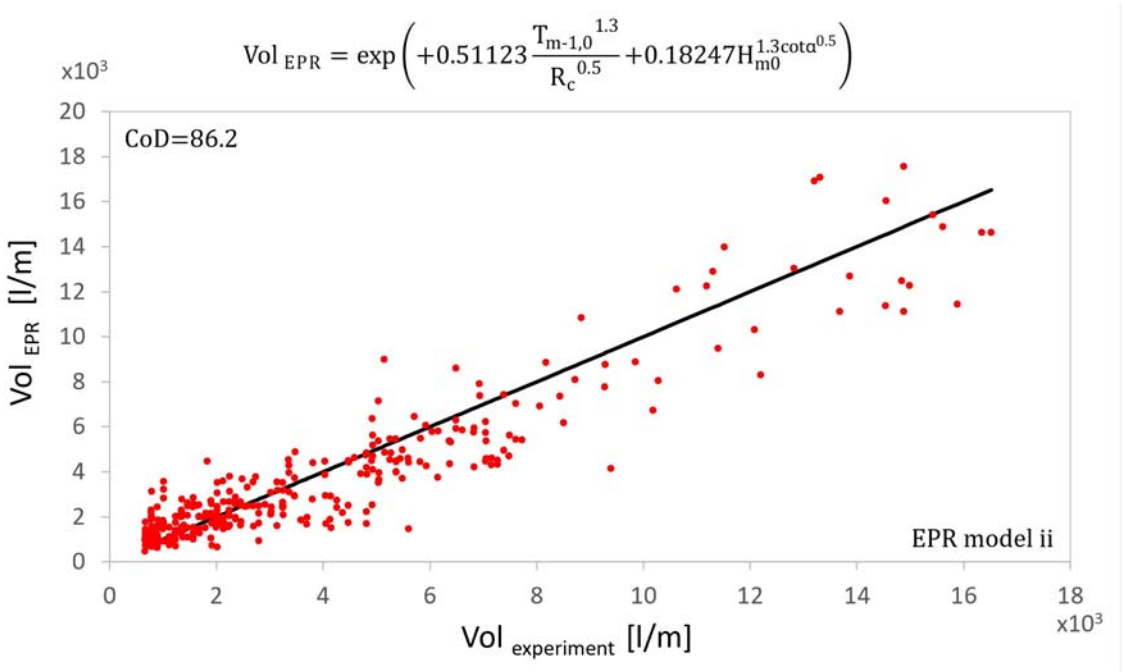


Figure A2. EPR model ii.

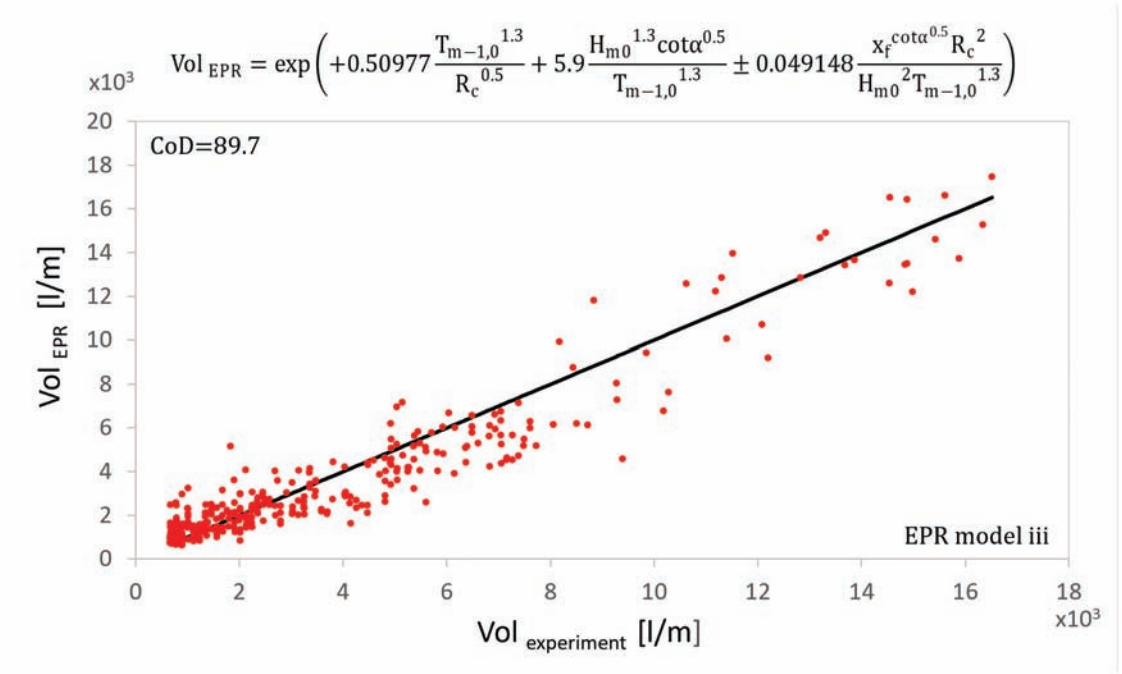


Figure A3. EPR model iii.

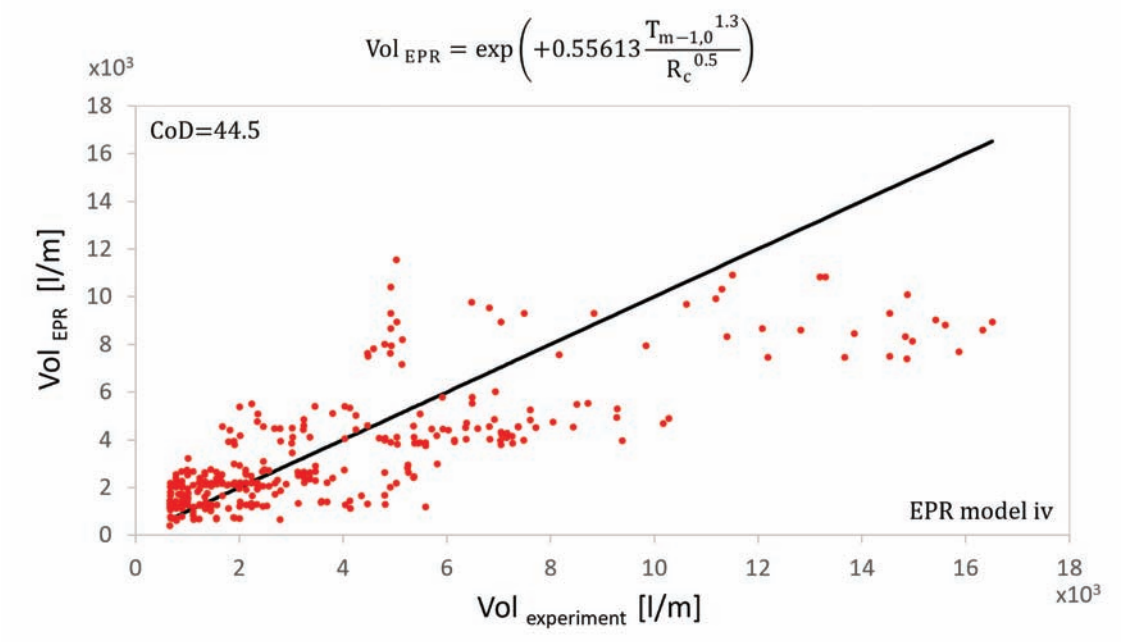


Figure A4. EPR model iv.

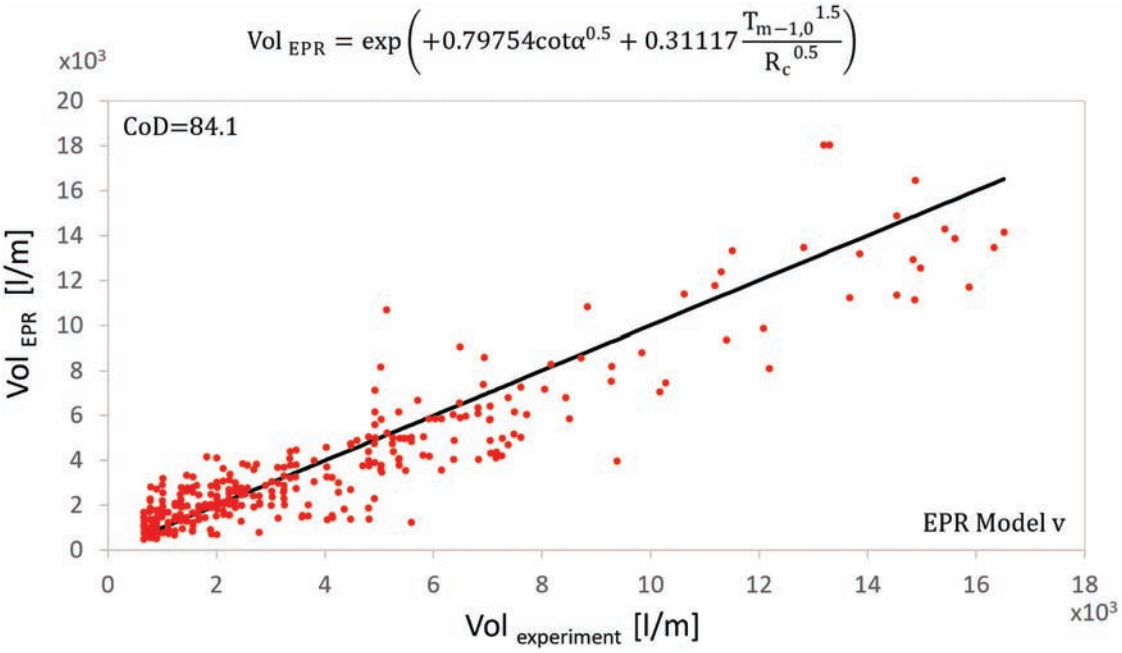


Figure A5. EPR model v.

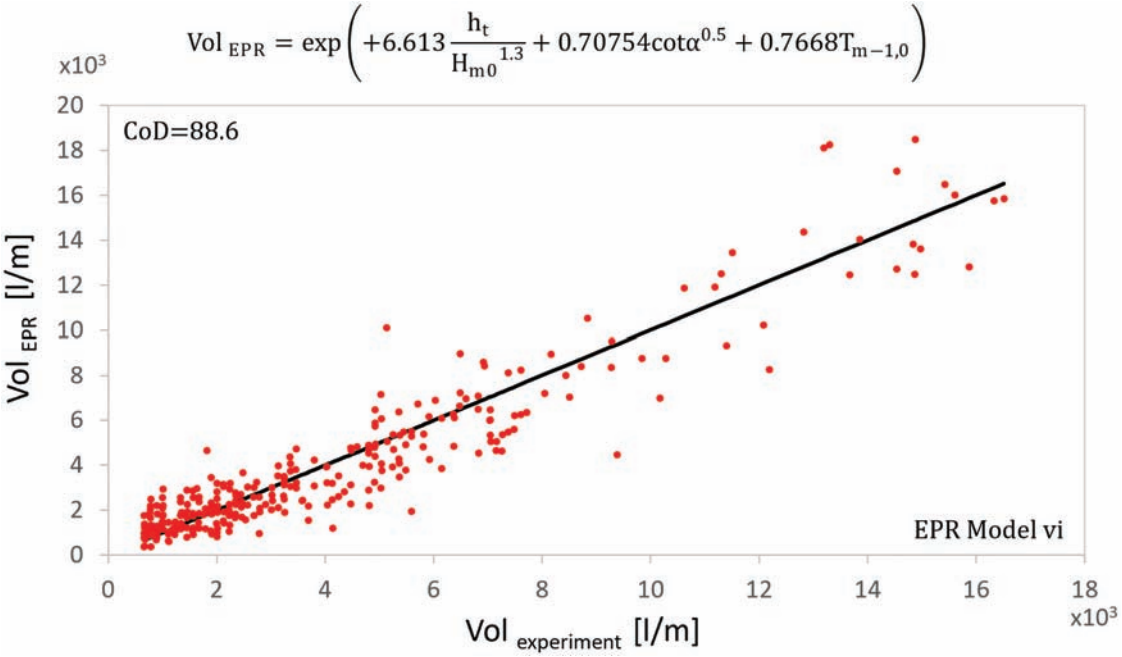


Figure A6. EPR model vi.

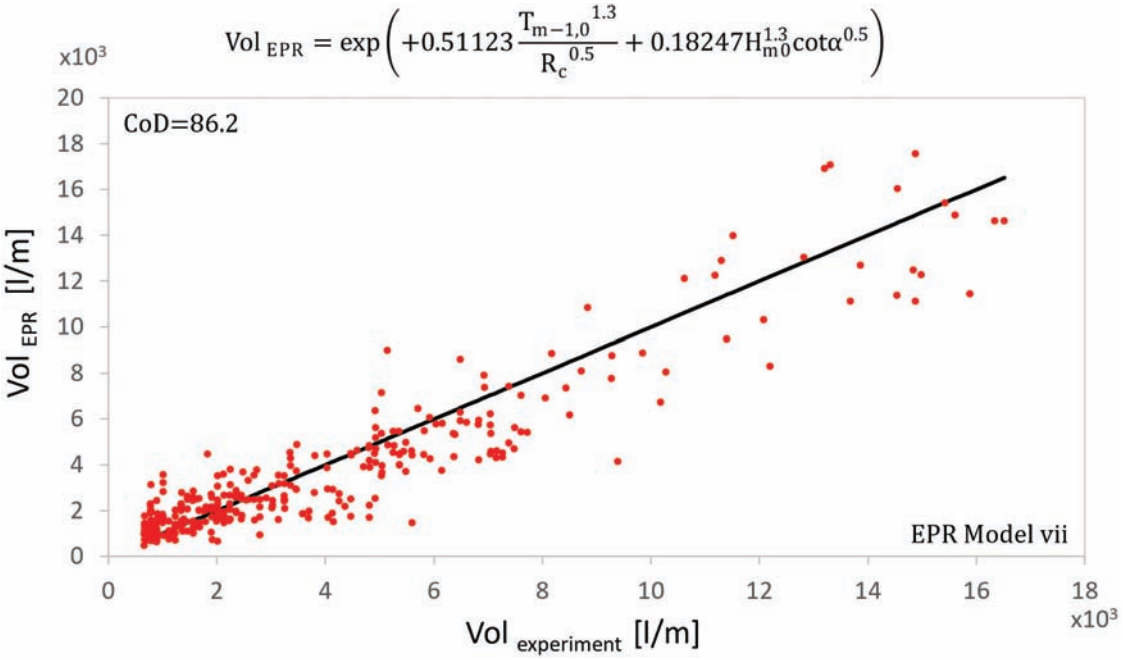


Figure A7. EPR model vii.

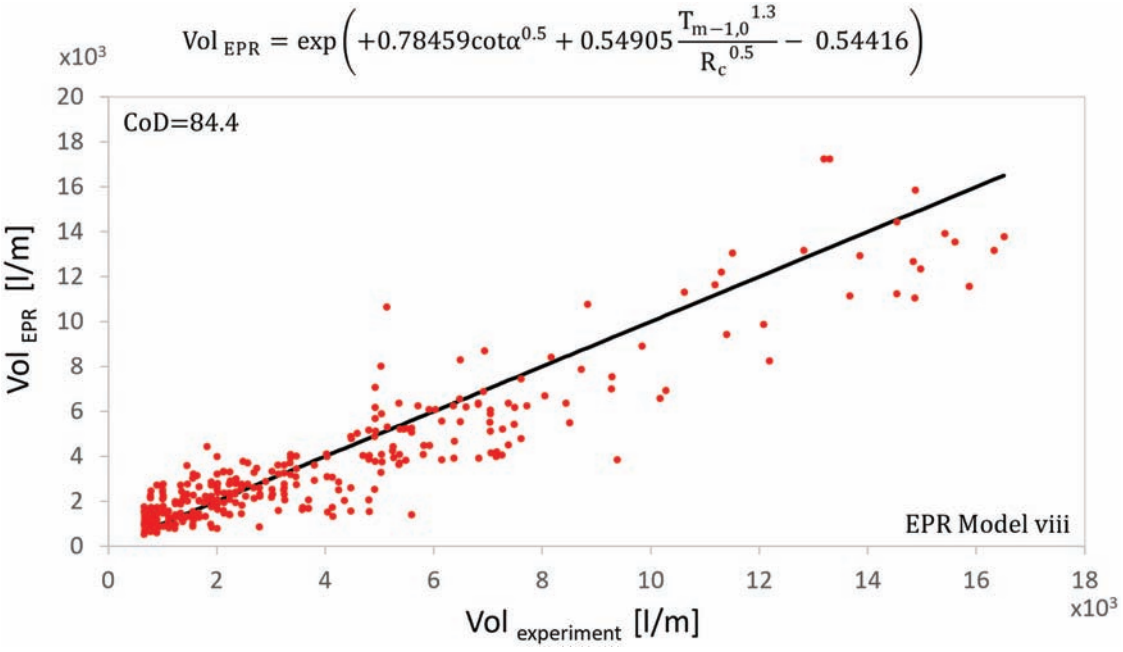


Figure A8. EPR model viii.

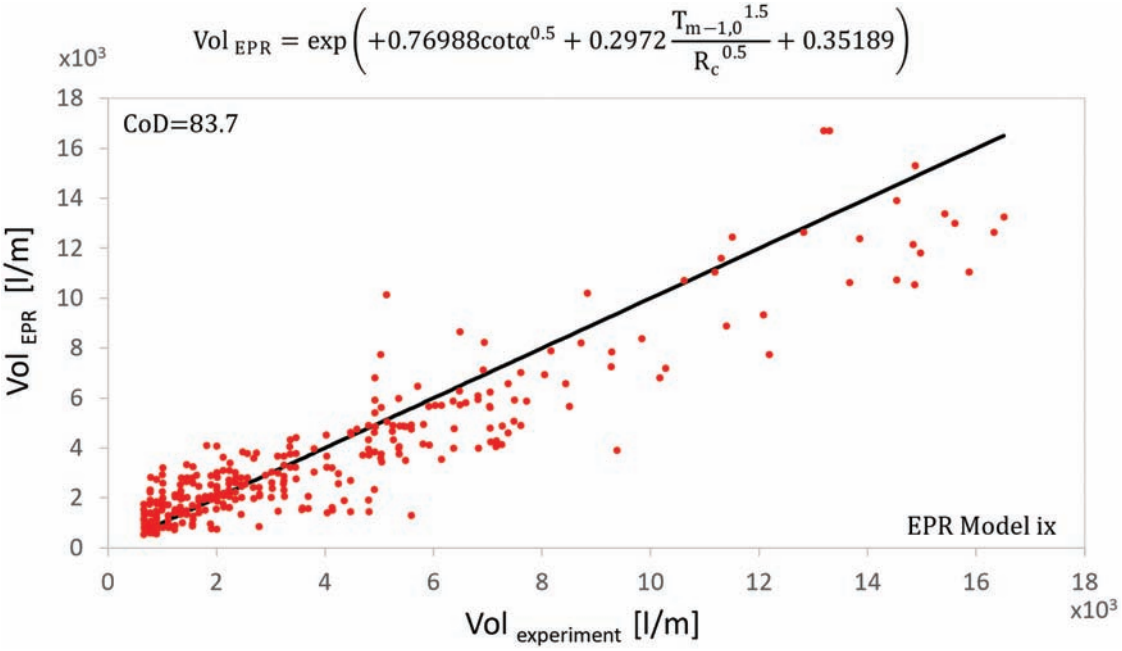


Figure A9. EPR model ix.

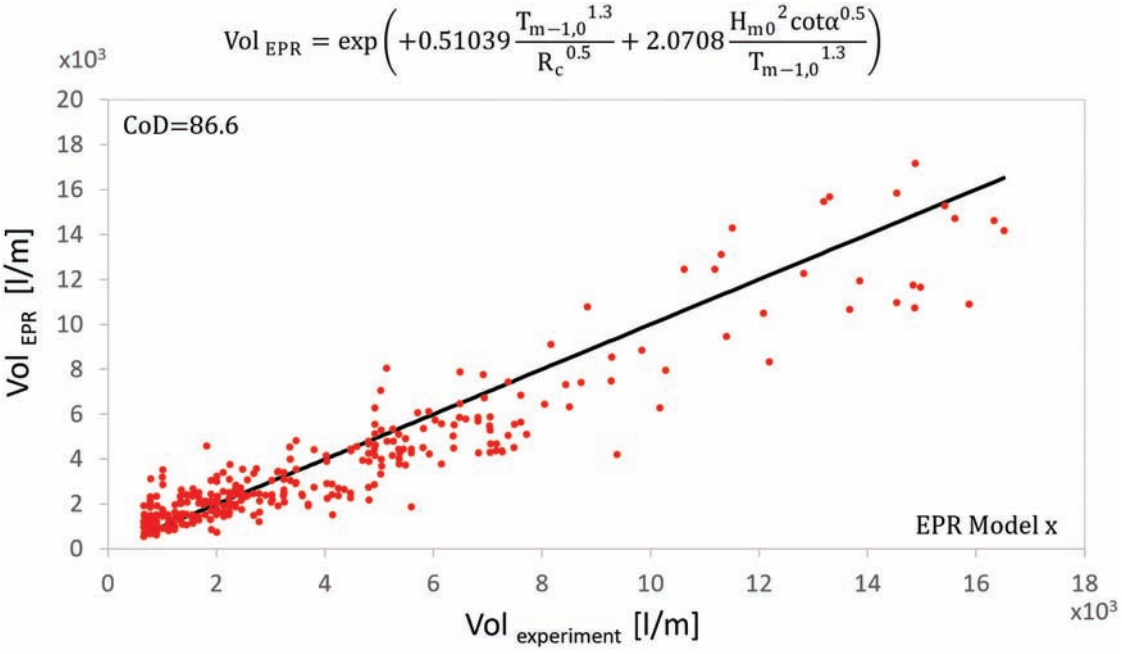


Figure A10. EPR model x.

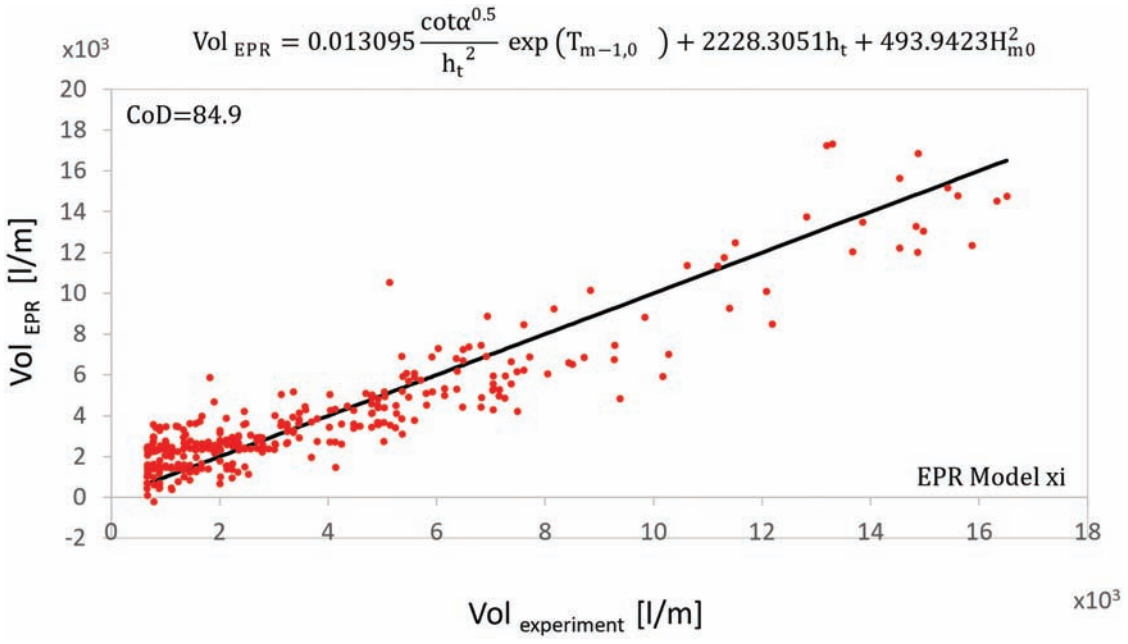


Figure A11. EPR model xi.

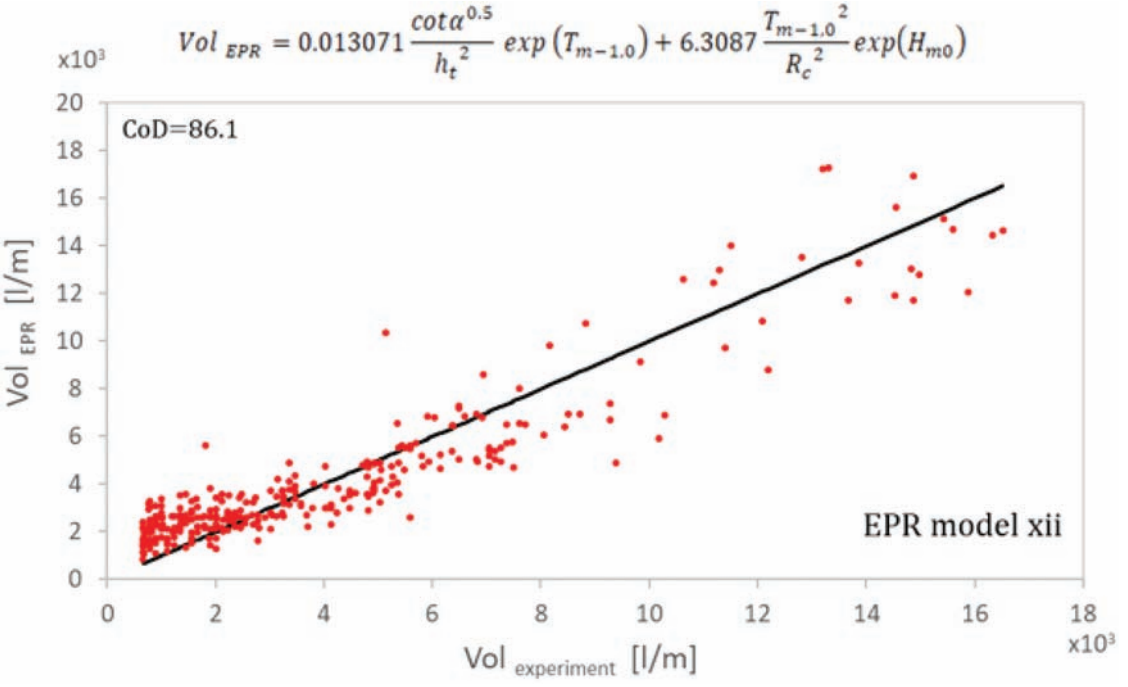


Figure A12. EPR model xii.

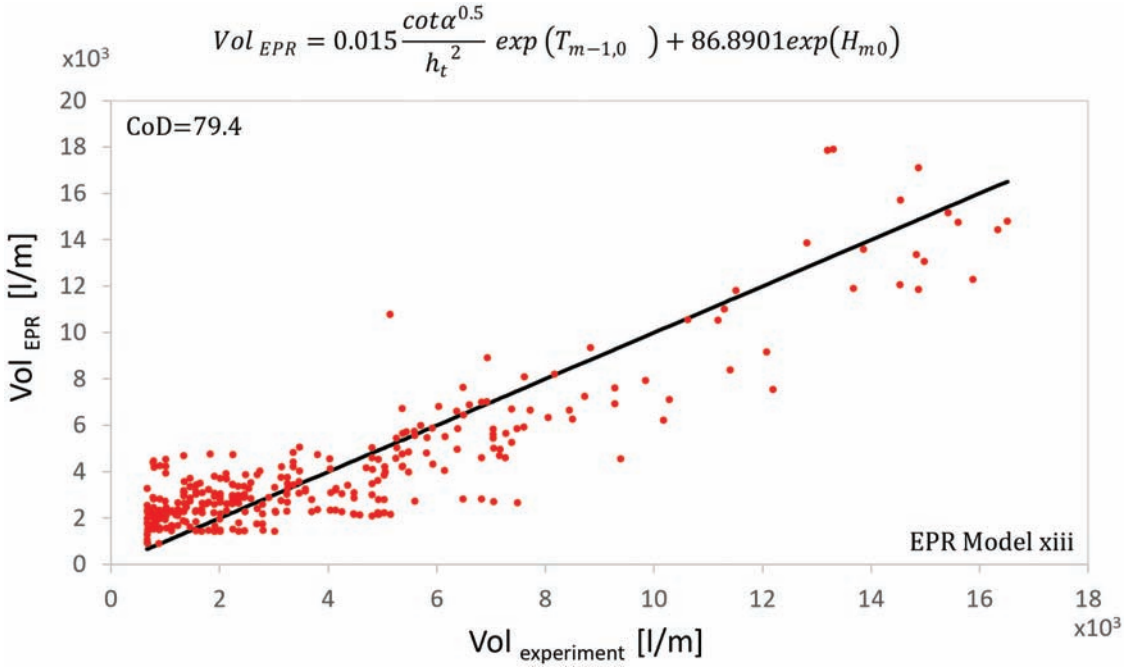


Figure A13. EPR model xiii.

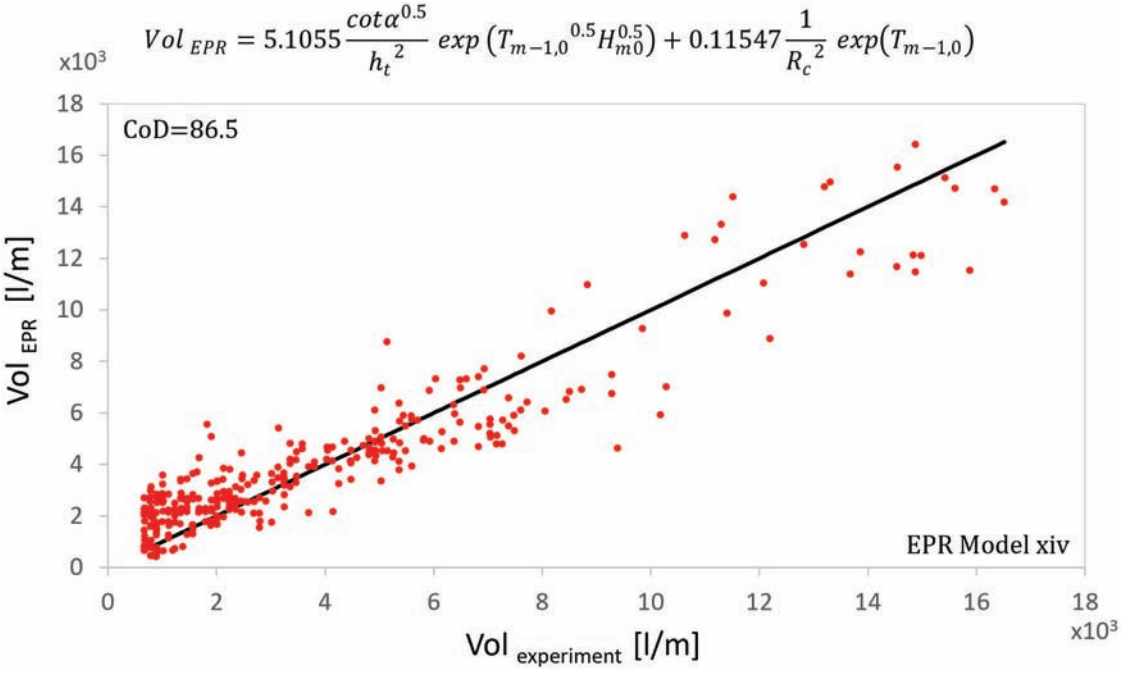


Figure A14. EPR model xiv.

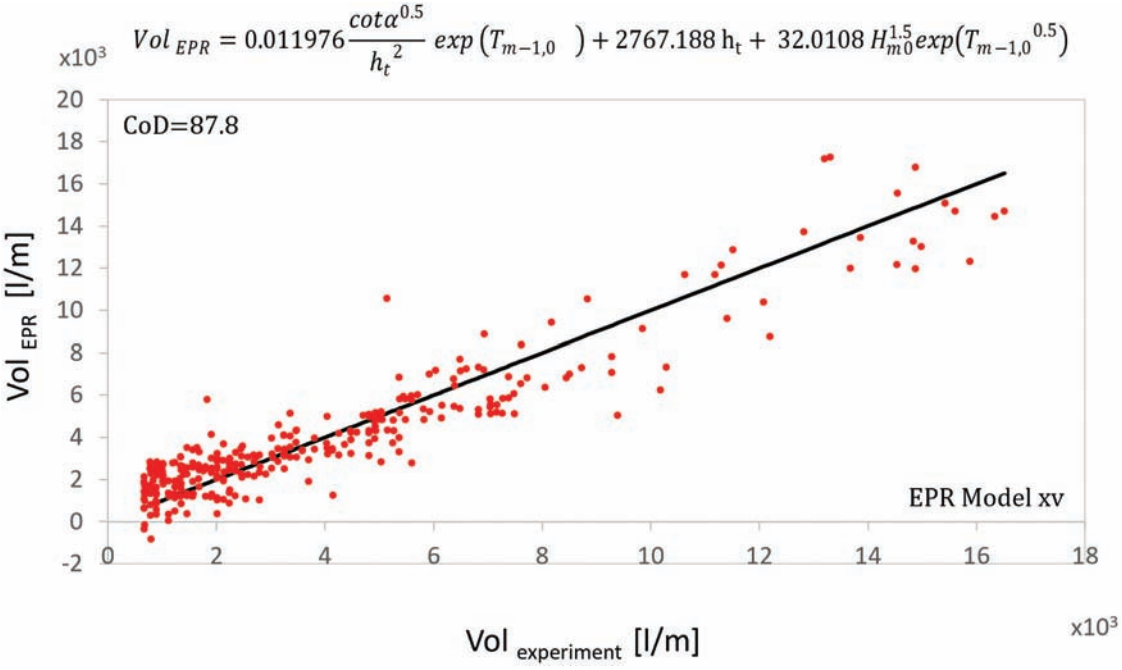


Figure A15. EPR model xv.