Prediction of Wave Runup on Beaches Using Gene-Expression Programming and Empirical Relationships

Hannah E. Power[1]*, Bahram Gharabaghi[2], Hossein Bonakdari[2], Bryson Robertson[3], Alexander L. Atkinson[4], and Tom E. Baldock[4]

[1] School of Environmental and Life Sciences, University of Newcastle, Callaghan, NSW, 2308, Australia; ORCID: 0000-0003-4332-8607
[2] Water Resources Engineering, University of Guelph, 50 Stone Rd East, Guelph, ON, N1G 2W1, Canada; ORCID: 0000-0003-0454-2811 (BG), 0000-0001-6169-3654 (HB)
[3] Mechanical Engineering, University of Victoria, PO Box 1700 STN CSC, Victoria, BC, V8W 3P6, Canada, ORCID: 0000-0002-7844-8708
[4] School of Civil Engineering, University of Queensland, St Lucia, Qld., 4067, Australia; ORCID: 0000-0002-1415-8381 (AA), 0000-0001-7040-9643 (TB)
* Corresponding author: hannah.power@newcastle.edu.au; +61-2-4921-5606

Keywords: wave runup; beaches; field data; lab data; Gene Expression Programming (GEP); sensitivity analysis

Abstract

This paper assesses the accuracy of seven empirical models and an explicit Gene-Expression Programming (GEP) model to predict wave runup against a large dataset of runup observations. Observations consist of field and laboratory measurements and include a wide array of beach types with varying sediment sizes (from fine sand to cobbles) and bed roughness (from smooth steel to asphalt). We show that the best performing models in the literature are prone to significant errors (minimum RMSE of 1.05 m and NMSE of 0.23) when used with unseen data, i.e., uncalibrated models; however, overall error values and correlations are significantly reduced when models are optimised for the dataset. The best performing empirical models use a Hunt type scaling with an additional parameter for wave induced setup. The predictive ability of the explicit GEP model, which better captures the complex nonlinear effects of the key factors on the wave runup length, resulted in a statistically significant improvement in predictive capacity in comparison to all other empirical models assessed here, even on unseen data. Wave height, wavelength, and beach slope are shown to be the three primary factors influencing wave runup, with grain size/bed roughness having a smaller, but still significant influence on the runup. The r^2 of the best optimised existing models (which takes the form of Holman (1986) and Atkinson et al. (2017) their M2 model) was 0.77, with a RMSE of 0.85 m. These were improved to an r^2 of 0.82 (6% increase) and RMSE of 0.75 m (12% decrease) in the GEP-based model. The sensitivity of the proposed GEP-based model to each input variable is assessed via a partial derivative sensitivity analysis. The results demonstrate a higher sensitivity in the model to small values of each input and that wave steepness and beach slope are the primary factors influencing wave runup.

1. Introduction

Wave runup represents the landward limit of wave action on a beach. It consists of a combination of two processes: swash and wave generated setup (Holman 1986). The wave runup region is of crucial importance for coastal engineering and management applications: from wave overtopping of coastal barriers and engineering structures to predicting coastal erosion and recovery during and subsequent to storm events to coastal engineering structure design. It is also the region of sediment exchange between the subaqueous and subaerial beach.

The most common way to predict runup is to use empirical formulae to predict the vertical level of wave runup relative to the still water level (or mean offshore ocean water level) and this approach has been used since the 1950's (Wassing 1957, Hunt 1958). These formulae often aim to predict statistical parameters such as R_{max} and $R_{2\%}$. R_{max} represents the maximum runup elevation observed within a given time period (and is therefore a function of the sampling period), while $R_{2\%}$ represents the level exceeded by 2% of runup events within a time period and, given constant wave conditions, should be independent of sampling period.

The majority of empirical formulae are designed to be able to predict wave runup levels using easily obtainable parameters such as offshore wave conditions and beach conditions. Most models typically include some combination of wave height and wavelength (*H* and *L*) and beach (or swash zone) slope (tan*β*). Models have been developed using either laboratory and/or field data and by fitting models to the runup observed. The limitations of both of these techniques were recently discussed by Atkinson et al. (2017).

One common model scaling that is used in several empirical models is the scaling developed by Hunt (1958) of $\alpha \tan\beta \sqrt{HL}$ where α is a scaling parameter adjusted to improve the data-model fit (e.g., Hunt 1958, Mase 1989, Van Der Meer and Stam 1992). In some cases, an additional term is added to represent the wave setup: $\alpha \tan\beta \sqrt{HL} + \gamma H$ where α and γ are independent scaling parameters (e.g., Holman 1986, Hedges and Mase 2004, Atkinson et al. 2017). Few empirical models include bed roughness or grainsize in their formulations; one recent exception is Poate et al. (2016). Further detail on the range and variation of empirical runup models can be found in recent works by Blenkinsopp et al. (2016) and Atkinson et al. (2017) who provided detailed overviews of the wide range of empirical models that are frequently used to predict wave runup.

Atkinson et al. (2017) recently assessed the accuracy of 11 empirical runup models using a field data set collected on 11 intermediate beaches on the east Australian coast. They observed large variability in the natural runup measurements and high variability in the accuracy of different runup model predictions. Three $R_{2\%}$ models were identified as the most accurate: Holman (1986), Vousdoukas (2012), and Atkinson et al. (2017) M2 which each gave errors of ~25% of $R_{2\%}$; however, there were significant differences in the RMSE for different models on the same beach and significant differences in RMSE for the same model on different beaches. Models derived from field data performed significantly better when predicting runup in the field than models derived from lab data. Most significantly, no single model gave the best runup estimates for all beaches in their dataset, suggesting that un-calibrated model predictions on an arbitrarily selected beach can be prone to significant error. Site-specific nearshore wave transformation effects and local variations in surf zone profiles may contribute to these errors, however, these are typically not included to maintain ease of use of empirical equations.

In an attempt to further improve wave runup predictions on beaches, this paper makes use of artificial intelligence (AI) based techniques which have grown substantially in various engineering sciences, particularly water and ocean engineering, due to their ability to solve complex nonlinear problems. The most popular AI-based techniques in this field are artificial neural networks (ANNs), adaptive neuro-fuzzy inference systems, and support vector machines. Although these three methods have been shown to have acceptable performance for solving water engineering problems (e.g., Azimi et al. 2016, Sharafi et al. 2016, Ebtehaj et al. 2018, Moradi et al., 2018, Gholami et al., 2018), their main drawback is their inability to provide an explicit expression to employ in future work, limiting transferability and application. To overcome this limitation, the AI technique Gene-Expression Programming (GEP) was developed and is increasingly being used as an efficient method for modelling nonlinear and complex processes (Ferreira 2001). One of the key benefits of the GEP technique is that it produces an explicit predictive expression as opposed to AI techniques like ANNs which do not produce an explicit expression and are thus limited in their transferability. Examples of applications that have used GEP include global climate analysis (Barbulescu and Băutu 2009), flow discharge (Azimi et al. 2017), shear stress distribution (Khozani et al. 2017), scour depths (Azamathulla 2012), sediment transport (Ebtehaj and Bonakdari 2017), and wave breaking (Robertson and Gharabaghi 2017; Robertson et al. 2017).

1.1. Gene expression programming

The GEP technique combines two popular genetic-based techniques: the genetic algorithm technique of describing complex relations by simpler, fixed length, linear structures called chromosomes (referred to as the genotype, e.g., Table 1) and the genetic programming technique of using expression tree (ET) structures with different sizes and shapes (referred to as the phenotype, e.g., Table 1). This harnesses the advantages of each of the two methods and overcomes their individual constraints. In the GEP method, the solution to the problem being investigated is described via chromosomes that consist of one or more genes. Each gene consists of a head and a tail. The head contains both symbols that represent functions (addition, subtraction, etc.) and symbols that represent the variables in the problem (in the case of wave runup, offshore wave height, wavelength, etc.), while the tail consists of only variables (see Table 1). Each gene codes for an ET of varying length that will always describe a mathematically valid algebraic expression. The length of the ET will depend on the order and positioning of the elements in the gene: the shortest ET will derive from a gene that has a variable as the first element of the gene, and the longest ET will derive from a gene that has only functions in the head. As an example, the single gene described in Table 1 creates a valid and complete ET using only the first 15 elements, i.e., the last six elements in the gene are not expressed. Two or more genes can be combined to make a chromosome, where each gene in the chromosome codes for a sub-ET with the sub-ETs being combined through a chosen linking function (e.g., addition, multiplication, etc.; see example in Table 1).

The first step in GEP is to randomly initialise a population of individuals, whereby each individual consists of one chromosome containing a specified number of genes that are combined using a sub-ET linking function. Each gene has a fixed head size and is made up of the user-specified functions and variables. Each individual in the initial population is then represented as an ET and the efficiency or compatibility degree of each individual is evaluated using a fit function (e.g., root-mean-square error, correlation, etc.). The individuals with greater fitness are selected for replication while some individuals are replicated with random modifications using different genetic operators. The linear structure of the chromosomes in the GEP method makes it possible to use a range of genetic operators, such as random mutation and recombination, to construct valid structures and to evolve existing structures in the same way in which genetic material in natural organisms is replicated and mutated. Examples of ways in which genes and chromosomes can be modified through genetic operators and thus evolve from one generation to the next are shown in Table 2. This process of replication and modification results in new offspring being generated creating a subsequent generation of the population. This process can be repeated until a certain criterion is met (e.g., number of generations, minimum fit function value, etc.). A flowchart schematising the GEP model technique is shown in Figure 1 and a full explanation of the GEP algorithm can be found in Ferreira (2001).

The main objective of the current study is the development of an explicit, GEP-derived expression to predict wave runup and to gain insights into the factors controlling wave runup using this novel method. Moreover, this study assesses the performance of seven empirical runup models against a large dataset of runup observations. To the authors' knowledge, this is the largest runup dataset compiled in the literature. The empirical models are also optimised to obtain the lowest error by varying their coefficients. A GEP-based explicit expression is developed and tested against the dataset, with corresponding analysis on the relative importance of the variables incorporated in the model. The paper is organised as follows. The runup dataset is described in Section 2. Empirical runup models are detailed in Section 3 and the GEP model is described in Section 4. Results are presented in Section 5 and are discussed in Section 6. Final conclusions follow in Section 7.

2. Runup dataset

In order to accurately assess the performance of existing empirical runup models, a large dataset of 1390 observations of $R_{2\%}$ was collated. The dataset primarily consisted of field data, with a small proportion of laboratory data (148 observations).

Field datasets, consisting of $R_{2\%}$, offshore wave conditions, and beach face slopes, were obtained from datasets freely available online and data provided by colleagues (Table 3). Details of the field experiments can be found in Stockdon et al. (2006), Poate et al. (2016), Nicolae Lerma et al. (2017), and Atkinson et al. (2017). The compiled field dataset represents a wide range of beach conditions from low to high energy (0.35m≤H_s≤7.17m, where H_s is significant wave height) with a wide range of peak wave periods (3.7s≤T_p≤23.7s) and beach slopes (0.0090≤tanβ≤0.29), and from fine sand (D_{50}=0.2mm) to pebbles (D_{50}=5cm). The approximation H_s=1.41H_{rms} was used to convert between offshore significant wave height and offshore root mean square wave height (H_{rms}) where required. Details of each field dataset, including wave parameters, beach slope, and grain size, are shown in Table 3.

The laboratory datasets consisted of both large- and small-scale irregular wave data on plane slopes. The dataset of Mase (1989) consists of small-scale data with a range of wave conditions and beach slopes on a smooth steel bed. Published and unpublished data from laboratory experiments (SASME experiments by Baldock and Huntley, 2002) were incorporated and consisted of a range of wave conditions on a fixed polyethylene slope. Data from Howe (2016) was from the Large Wave Flume (Großer Wellenkanal, GWK) at Leibniz Universität Hannover with a 1:6 slope and varying wave conditions with two different bed types: a solid asphalt bed and a polyethylene bed. Details of each laboratory dataset, including wave parameters, beach slopes, and hydraulic roughness lengths, are shown in Table 4.

The hydraulic roughness length, r, was used to compare bed roughness between field and laboratory data. For the field data, grain size was converted to a roughness length using r=2.5D_{50} (Nielsen, 1992, p. 105). For the laboratory data, roughness lengths for smooth steel and polyethylene were obtained from standard look-up tables, and Howe (2016) provided an equivalent D_{50} for the solid asphalt bed (Table 4).

The full compiled dataset is available as supplementary material to this manuscript.

3. Empirical runup models

The performance of several empirical models was assessed by comparing predicted runup values to the observed values from the combined field and laboratory dataset. A total of seven empirical models were chosen for assessment: Holman (1986), Nielsen and Hanslow (1991), Stockdon et al. (2006), Vousdoukas et al. (2014), Poate et al. (2016) their Eq. (12), Poate et al. (2016) their Eq. (9), and Atkinson et al. (2017) M2 (their Eq. (15)). The empirical models, their applicability, and the data type from which they were derived (i.e., laboratory or field data) are detailed in Table 5. This subset of models was chosen for one or both of the following reasons: (1) the model was identified as an accurate model in a recent assessment of runup models (e.g., Holman, Vousdoukas, and Atkinson M2 models were designated as the most accurate field derived models in Atkinson et al. (2017); Stockton was recommended by Blenkinsopp (2016), and Poate et al. (2016) found variations on their Eq. (12) to be the best predictor that did not incorporate grainsize); and/or (2) the model takes a different approach to parameterising runup or incorporates a variable that is not used in the models identified by reason (1) (e.g., Nielsen and Hanslow vary their formulation for runup depending on the beach slope while Poate Eq. (9) incorporates grainsize). As the majority of the dataset compiled in this study consists of field data, empirical models derived from laboratory data were not assessed here as Atkinson et al. (2017) showed that models derived from field data were, on average, more accurate for predicting runup in the field.

Model performance was quantified through the use of the following statistical parameters: root mean square error (RMSE), normalised mean square error (NMSE), the coefficient of determination (r^2), mean prediction error or bias (B), and scatter index (SI). These are calculated using:

$$RMSE = \sqrt{\frac{\sum_i \left(R_{pred,i} - R_{obs,i}\right)^2}{N}}, \tag{1}$$

$$NMSE = \frac{1}{N} \sum_i \frac{\left(R_{pred,i} - R_{obs,i}\right)^2}{\overline{R_{pred}} \cdot \overline{R_{obs}}}, \tag{2}$$

$$r^2 = \left(\frac{N\sum_i(R_{pred,i}R_{obs,i}) - \sum_i R_{pred,i}\sum_i R_{obs,i}}{\sqrt{\sum_i R_{obs,i}}}\right)^2, \quad (3)$$

$$B = \frac{\sum_i(R_{pred,i} - R_{obs,i})}{N}, \text{ and} \quad (4)$$

$$SI = \frac{RMSE}{\sum_i R_{obs,i}/_N}, \quad (5)$$

where R_{obs} is the observed 2% runup exceedance level, R_{pred} is the predicted 2% runup exceedance level, and N is the number of observations. Low values of *RMSE, NMSE*, and *SI* (i.e., values approaching 0) represent better model performance, as do values of r^2 approaching 1. Negative values of B indicate the model underestimated the observed values while positive values indicate overestimations with $B = 0$ indicating no net over- or under-estimation. To further assess the uncertainty in model predictions, the standard deviation of prediction errors (S_e) is calculated using:

$$S_e = \sqrt{\frac{\sum_i\left((R_{pred,i} - R_{obs,i}) - B\right)^2}{N-2}} \quad (6)$$

where B represents the mean prediction error as described above in Eq. (4). Using a Wilson score method without continuity correction, an uncertainty band can be defined around the predicted values of the mean prediction error (B) through the use of $B \pm 1.96 S_e$ to obtain an approximately 95% error uncertainty band (Atieh et al., 2017; Ebtehaj et al. 2018).

4. GEP modelling

For the purposes of GEP modelling, runup was converted to dimensionless runup ($R_{2\%}/H_s$, relative to the still water level) and was considered to be a function of three dimensionless variables:

$$R_{2\%}/H_s = f\left(H_s/L_p, \tan\beta, r/H_s\right) \quad (7)$$

Several runs were performed to ensure adequate robustness and generalisation of the model derived. The fitting parameters of the GEP method were selected based on the previous studies of the authors (Ebtehaj et al. 2015a; 2017) and a number of preliminary runs. The population size (number of chromosomes, see Section 1.1) determines the execution time so that a model with higher population leads to longer execution time. Due to the problem complexity and the number of possible solutions, the population sizes considered were 30, 40, 50, 100, 150 and 200, with a final size of 50 chosen to reduce model size but ensure model accuracy. To evolve the chromosome architecture, it is necessary to characterize the number of sub-ETs and the degree of gene complexity, which, in the evolved model, are defined by the gene numbers and the head length respectively. The optimal values for the number of genes and the head length were obtained via trial and error and, in this study, were taken as 5 and 8 respectively. The optimal values of other parameters used in the GEP model are presented in Table 6. Due to the successful performance of the root relative squared error (RRSE) function in recent studies in hydraulic and hydrology fields (Khozani et al., 2017; Gholami et al., 2018), the fitness of the GEP model was determined by the fitness function of the program, f_i:

$$f_i = \frac{1000}{1+\mathrm{RRSE}_j} \quad (8)$$

where model fitness ranges from 0 to 1000, such that f_i=1000 is a perfect fit.

5. Results

5.1. Empirical model assessment

Figure 2 and Table 7 show the comparison between the observed and predicted runup using the seven empirical relationships assessed in this paper, and statistical measures of the performance of the empirical relationships against the full dataset. The relationships of Holman (1986) and Atkinson et al. (2017) had the joint lowest NMSE (0.23) and the lowest RMSE (1.05 and 1.06 respectively) with both also having small scatter indices (0.45 and 0.46 respectively) and 95% confidence intervals of the error uncertainty bands (-2.26 - 1.70 and -2.23 - 1.88 respectively). Poate et al. (2016) Eq. (12) had the lowest absolute bias (0.09) and the highest r^2 value (0.72), with values for RMSE, NMSE, and the scatter index only marginally greater than those for the Holman (1986) and the Atkinson et al. (2017) M2 models. Poate et al. (2016) Eq. (9) had the lowest r^2 value (0.56) and the Nielsen and Hanslow (1991) model had the highest RMSE (2.14), NMSE (0.55), bias (1.27), and scatter index (0.92). All models, excluding Nielsen and Hanslow (1991) and Poate et al. (2016) Eq. (9), underpredicted runup (see *B* in Table 7). The models that are based on a Hunt type scaling $(\alpha \tan \beta \sqrt{HL}$; i.e., Holman (1986), Nielsen and Hanslow (1991), Stockdon (2006), Vousdoukas et al. (2014), and Atkinson et al. (2017) M2) display two clear diverging trends in the model-data comparisons (see Figure 2a-d and Figure 2g), which are particularly evident at high runup values. This is not apparent in the Poate et al. (2016) models which take a different approach to scaling runup ($\alpha \tan \beta^{0.5}\, HT$, Figure 2e-f) with grain size also incorporated in Poate et al. (2016) Eq. (12). The two data points with very large predicted $R_{2\%}$ values were collected at collected at Chesil Beach (February 2014) during a storm event and have offshore wave parameters of H_s=7.17 m and T_p=23.7 s. Quality controlled data from the nearest offshore wave buoy show that T_p exceeded 24 s during this storm (i.e., these data points are not anomalous) and, consequently, the large values of T_p result in very large runup predictions using the empirical formula found in the literature due to their dependence on wave period.

5.2. Empirical model optimisation

In addition to assessing empirical model performance, model coefficients were optimised to obtain an improved fit with the newly compiled dataset presented in this study. Coefficients were optimised through an unconstrained optimisation with the objective of minimising the sum of the differences between predicted and observed 2% runup exceedance levels. The optimised empirical equations are shown in Table 8.

In all cases, the empirical runup relationships with optimised coefficients showed improved or equal performance when compared with the unmodified empirical relationship for all measures of performance except bias (see Figure 3 cf. Figure 2, and Table 7 cf. Table 8). Only one optimised model, Poate et al. (2016) Eq. (12), showed an increased absolute bias of -0.14 when compared to the unmodified equation (B=-0.09). For all other models, absolute bias was reduced with optimised coefficients. The best performing optimised model, on all measures of performance, was the Holman (1986) formulation that has the standard Hunt type scaling plus setup (also used in the Atkinson et al. M2 model). Interestingly, the two scaling coefficients for the optimised model have very different weightings relative to the original model. In the optimised model, the setup term (the γH term in $\alpha\, tan\, \beta \sqrt{HL} + \gamma H$, where α and γ are independent scaling coefficients) is more heavily weighted than the runup term (the $\alpha\, tan\, \beta \sqrt{HL}$ term; where α=0.83 and γ=0.2 in the original equation and α=0.5 and γ=0.7 in the optimised equation). While the γH term is included in empirical equations of this form to represent setup, the increase in relative weighting of this term may not be solely due to setup but does represent an increase in the relative importance of wave height for predicting runup relative to the other parameters in these forms of empirical equations (i.e., tanβ and L). It also, therefore, suggests that beach slope and wave period are potentially less important than the Hunt approach would suggest. This is consistent with the Nielsen and Hanslow model for beaches with $\tan\beta < 0.1$ that does not include beach slope in the runup equation and may reflect the narrower relative range of periods observed in the field than in the Hunt laboratory data.

On average, optimised equations showed increased r^2 values by 0.06 and decreased RMSE and NMSE values by 0.40 m and 0.16 respectively, relative to the values obtained using the original model formulations. It is worthwhile noting that optimising the Holman (1986) formulation reduces the NMSE down from the typical error of 25% (identified by Atkinson et al. (2017) for the best models when applied to unseen data) to about 15%. This is the order of the error for individual models when optimised to the sub-datasets used to derive those same models (where reported). Thus, the present collated dataset has significant value to the research community in terms of model optimisation.

Interestingly, the propensity for the models that are based on a Hunt type scaling to display two clear trends in the original model-data comparisons (see Figure 2) was reduced such that the model-data comparisons for the optimised Holman (1986) type formulation collapse to show a single pattern (Figure 4). The optimised Nielsen and Hanslow (1991), Stockdon et al. (2006), and Vousdoukas et al. (2014) models still display two trends in the model-data comparisons (Figure 3).

5.3. GEP model

The GEP model was trained using a subset (79%) of the dataset to avoid overfitting. The training data was chosen at random, and the remainder of the dataset was used to validate model performance as testing data (e.g., Thompson et al., 2016; Trenouth et al., 2016; Atieh et al., 2017). All variables were non-dimensionalised such that the input variables were: H_s/L_p, tanβ, and r/H_s; and the output variable was $R_{2\%}/H_s$. Model performance of the GEP model was quantified through the use of the same statistical parameters used to assess the empirical models (RMSE, NMSE, r^2, B, SI, and S_e) and performance was evaluated for each of the training, testing, and complete datasets. The final GEP model is shown in Eq. (9) (where x_1=H_s/L_p, x_2=tanβ, and x_3=r/H_s) and Figure 5a.

$R_{2\%}/H_s$ = (x2+(((x3.*3)/exp(-5)).*((3.*x3).*x3))) + ((((x1+x3)-2)-(x3-x2))+((x2-x1)-x3)) +
(((x3.^x1)-(x3.^(1/3)))-(exp(x2).^(x1.*3))) + sqrt((((x3+x1)-x2)-(x2+log10(x3)))) +
((((x2.^2)./(x1.^(1/3))).^(x1.^(1/3)))-sqrt(x3)) + ((x2+((x3./x1).^(1/3)))+(log(2)-
(1./(1+exp(-(x2+x3)))))) + ((sqrt(x3)-(((3.^2)+3).*(x2.^2))).^2) +
((((x3.*-5).^2).^2)+(((x3+x3).*x1)./(x2.^2))) +
log((sqrt(((x2.^2)+(x3.^(1/3))))+((x2+3).^(1/3)))) +
((((x1./x3).*(-5.^2)).*(x3.^2))-log10((1./(1+exp(-(x2+x3)))))) + (x1.^x3) +
exp(-((((x3./x1).^exp(4))+(exp(x3).^3)).^2)) + exp((log((x2-x3))-log(exp(-((-1+x1).^2))))) +
((sqrt(4).*(((x3./x2)-x2)-(0-x1))).^2) + (2.*((((-5.*x3)+x1).*(2-x3))-2)) +
((sqrt(4).*(((x3./x2)-x2)-(0-x1))).^2) + ((((-5+x1)-x2).*(x2-x3)).*((x1-x2)-(-4.^-5))) +
(exp(-((x2+(-5-x1)).^2))+((x2+5).*(x3.^2))) +
sqrt(1./(1+exp(-((exp(x1)-exp(-((x3+x3).^2)))+((x1.^x3)-(x3.*4)))))) +
((exp(-((((exp(-(((sqrt(x3).*4) + (1./(1+exp(-(x2+2))))).^2))).^2)+x1).^2))).^3); (9)

The GEP model consistently outperformed the empirical runup relationships as shown by all statistical parameters, with RMSE and NMSE for the full dataset of 0.75 m and 0.10 respectively and an r^2 value of 0.82 (Table 8). These values represent a decrease in RMSE and NMSE of 12% and 29% respectively relative to the best performing optimised empirical model and of 29% and 57% respectively when compared to the best performing original empirical models. Additionally, the GEP achieves a 12% decrease in 95% prediction error uncertainty band when compared to the best performing optimised empirical model and a 25% decrease when compared to the best performing original empirical model (see Table 7 and Table 8). Only 15.5% of the testing dataset had predictions outside a 50% tolerance window. Further, the GEP model also outperformed both the original and optimised equations of Poate on the gravel beach subset of the dataset in all measures except bias (see Table 7 and Table 8). It should be noted that the GEP model was developed for the data ranges specified in Table 3 and Table 4 and is, therefore, untested outside of these data ranges, however, this is typically the case for the empirical relationships presented in the literature.

Figure 6 shows non-dimensional runup plotted against the three input variables used in the GEP model, with data delineated by the beach and study from which the data were obtained. In general, $R_{2\%}/H_s$ increases with decreasing H_s/L_p, with more rapid changes and greater variability in $R_{2\%}/H_s$ for smaller values of H_s/L_p. The reverse is true for beach slope with increasing $R_{2\%}/H_s$ with increasing tanβ. Trends between $R_{2\%}/H_s$ and r/H_s are less clear and there is a high degree of variability within the dataset. It is also clear that, despite conforming to the overall trends, individual datasets display highly variable behaviour. Figure 7 shows a coplot of the full dataset and the GEP model (Eq. 8). This further confirms the observations from Figure 6 and shows that the trends observed in the data are well described by the GEP model (solid black lines) with the GEP showing increasing $R_{2\%}/H_s$ with decreasing H_s/L_p and increasing $R_{2\%}/H_s$ with increasing tanβ.

To further investigate these trends, a multiway ANOVA with random effects was used to assess the relative importance of the three independent input variables (H_s/L_p, tanβ, and r/H_s) on the model outputs ($R_{2\%}/H_s$). Data were grouped into 20 equal divisions (i.e., 5 percentile ranges) for each variable range and the full factor space for the three variables each with 20 equal divisions was generated (i.e., 20^3 combinations). Combinations of variables with no corresponding data points were then removed from analysis, leaving n=581 combinations of variables, and the mean model output was calculated for each combination. The variance components estimate for each independent variable were obtained from the ANOVA and converted to a percentage of total variance. This procedure was repeated for 10 (n=318), 12 (n=402), 14 (n=452), 16 (n=513), and 18 (n=548) equal divisions to ensure the number of divisions was not affecting results. The mean percentage of total variance attributable to each of the three independent variables was consistent across the ANOVA models with different data groupings, with 23-25% (average of 24%) of variability attributable to H_s/L_p, 51-57% (53%) attributable to tanβ, and 15-19% (18%) attributable to r/H_s. The remaining 3-7% (6%) was not attributable to any of the three independent variables.

Non-dimensional runup is shown against wave steepness and beach slope in Figure 8 as these are the two parameters identified as having the highest impact on the output parameters. A thin-plate spline interpolated surface is also shown to assist with the visualisation of the trends. It is clear that the trends described by the GEP (i.e., increasing non-dimensional runup with decreasing wave steepness and increasing non-dimensional runup with increasing beach slope) are seen in the data. Significant scatter is observed within the data, some of which may be attributable to grainsize/roughness effects which are not included in the figure, however, the results of the ANOVA analysis also suggest that there are additional factors that cause variability in runup levels beyond the variables investigated here (i.e., 3-7% of the variability could not be accounted for). Additional parameters, not accounted for in this study given the limits of the dataset, such as beach type, surf zone width, slope, and type, tidal phase, and nearshore bathymetric profiles that would alter wave transformation between the offshore and the surf zone, may also account for some of the scatter in this figure. In particular, Atkinson et al. (2017) found that model performance varied considerably at different tidal stages.

To statistically identify the best performing model of all the models tested here, the Akaike information criterion (AIC) was used as it incorporates the number of model parameters by effectively penalising models with greater numbers of parameters (Burnham and Anderson, 2004). This therefore ensures a “fairer” comparison between the relatively simple empirical models that exist in the literature and the more complex GEP model developed here. The approach taken here is to use the least squares (LS) based version of the AIC formula which is expressed as:

$$\mathrm{AIC} = N\log(MSE) + 2K \tag{10}$$

where

$$MSE = \frac{\sum_i \left(R_{pred,i} - R_{obs,i}\right)^2}{N} \tag{11}$$

and K is the total number of parameters in the model including any intercepts. For the GEP model, a conservative approach was taken to determining the number of parameters by counting every constant and variable in Eq. (9). While the GEP model is not derived using a LS technique, this approach is considered approximately correct for our data as the criterion for fitting, RRSE (Eq. 8), is similar to the MSE used in LS modelling and the residuals are approximately normally distributed. The AIC value is computed for each model in a given set of models investigated and the models can be ranked from best (lowest AIC value) to worst. To enable effective comparison between AIC values, they are rescaled to:

$$\Delta_i = \mathrm{AIC}_i - \mathrm{AIC}_{\min} \tag{12}$$

where $\mathrm{AIC}_{\min}$ is the minimum AIC value of all the AIC values for each individual model (AIC_i). Thus the best model has $\Delta_i = 0$ and all other models have $\Delta_i > 0$. Models with values of $\Delta_i \leq 2$ have substantial statistical support, i.e. they would be considered equivalent to the best model. Models with $4 \le \Delta_i \le 7$ have considerably less support, and models with $\Delta_i > 10$ have essentially no support, i.e. they would be considered definitely inferior to the best model (for further detail see Burnham and Anderson, 2004). It is clear from the results shown in Table 9 that the GEP model is by far the statistically best model with all other models having Δ_i values far in excess of $\Delta_i > 10$. Even the best performing empirical model (the optimised Holman (1986) formulation) had a Δ_i far in excess of $\Delta_i = 10$ with $\Delta_i = 76.58$. Further, the Akaike weights (w_i) are calculated using:

$$w_i = \frac{\exp(-\Delta_i/2)}{\sum_{r=1}^{R} \exp(-\Delta_r/2)} \tag{12}$$

for each individual model in the model set (r = 1,…,R) where Δ_i is given in Eq. (12) above. The resultant values (w_i) can be interpreted as the probabilities that a given model, i, in the model set is the best model for the data. The w_i values clearly show that the GEP model with probability $w_i = 1$ is the best model. The sum of all other w_i values (i.e., the sum of the probabilities of any one of all the other models being correct) is 2.35×10^{-17} (Table 9).

6. Discussion

The dataset compiled here represents a wide range of wave and beach conditions that represent field conditions from below average (e.g., Atkinson et al. 2017) to energetic storm conditions (e.g., Poate et al. 2016) as well as both small and large scale laboratory conditions (Mase 1989, Baldock and Huntley 2002, Howe 2016). While the majority of the dataset compiled here is from field data sets or large scale lab data (89% of the dataset), there is potential for non-scalable physical characteristics to influence the measured runup in the two small scale laboratory datasets (Mase 1989, Baldock and Huntley 2002) and this is not accounted for in this study.

The results of the model optimisation process confirm that, while the empirical models presented in the literature work well for the datasets against which they were calibrated, they are not universally applicable, with every empirical model tested improving in performance after model optimisation (see Table 8 c.f. Table 7). This suggests that the models are not fully capturing all the relevant factors controlling wave runup on beaches or that simplifying complex natural processes (such as varying wave spectra and non-planar beach slopes) to simple parameterisations (H_s, L_p, $\tan\beta$, and r) is insufficient. Additionally, model predictions changed drastically between datasets, with some datasets obtained from the same beach displaying vastly different predictions using the same model. For example, the Holman model (both original and optimised) over-predicts data from Duck82 but under-predicts data from Duck94 (Figure 4). Despite this, the optimised form of the Holman equation, which is a Hunt style scaling parameter with an additional term for setup, was found to be the most accurate of the optimised empirical formulations.

The GEP model provided a significant improvement in predictive ability when compared to the existing empirical relationships and is shown to be by far the statistically best model. Statistical measures for the GEP model compared to the whole dataset, such as correlation coefficient (r^2=0.82) and normalised mean square error (NMSE=0.10), were consistently better than any observed for the empirical relationships (r^2=0.77 and NMSE=0.14 for the best performing optimised model; Table 8) with improvements of 6% and 29% respectively.

6.1. GEP sensitivity analysis

Any explicit expression provided to compute a parameter depends on a number of independent parameters. Therefore, the importance of each of these independent parameters on the proposed explicit expression must be investigated. In this study, the partial derivative sensitivity analysis (PDSA) method (Ebtehaj et al. 2015b; Azimi et al. 2017) is employed to study the trends in the variation of the ouput variable ($R_{2\%}/H_s$) due to variations in of each of the input variables (x_i: H_s/L_p, tanβ, and r/H_s). Here, sensitivity is assessed by calculating the partial derivative (PD) of $R_{2\%}/H_s$ with respect to each input parameter (i.e., $\partial(R_{2\%}/H_s)/\partial(H_s/L_p)$, $\partial(R_{2\%}/H_s)/\partial(\tan\beta)$, and $\partial(R_{2\%}/H_s)/\partial(r/H_s)$) and calculating the corresponding value for each data point (Figure 5). The absolute magnitude of the PD value indicates the degree of influence of a given input parameter (x_i) on $R_{2\%}/H_s$ (i.e., a larger absolute PD value indicate a greater degree of influence of x_i on $R_{2\%}/H_s$) and the sign of the PD value, positive or negative, represents the sign of the trend (i.e., a positive (negative) PD value results in an increase (decrease) of $R_{2\%}/H_s$ with increasing x_i).

The PDSA results of the GEP expression proposed here are shown in Figure 5b-e. The sensitivity trends of all three independent variables on $R_{2\%}/H_s$ are not constant, such that different ranges of each of the independent variables result in varying sensitivities of the output variable, i.e., the trends are highly non-linear. The sensitivity of the proposed model to H_s/L_p is greatest for low values of H_s/L_p with large negative sensitivities indicating large decreases in $R_{2\%}/H_s$ for small increases in H_s/L_p (Figure 5b). As H_s/L_p increases, the magnitude of the sensitivity decreases such that increases in H_s/L_p result in smaller decreases of $R_{2\%}/H_s$. The majority of sensitivity values for tanβ are positive, with a slight trend of increasing positive sensitivity with increasing tanβ thus implying more rapid increases in $R_{2\%}/H_s$ with increases in tanβ for larger values of tanβ (Figure 5c). The sensitivity of the model to low values of r/H_s is highly variable with no specific trends in the sensitivity of r/H_s in this range of r/H_s (Figure 5d). For larger values of r/H_s (r/H_s>0.03), the sensitivity indicates a decrease in $R_{2\%}/H_s$ for increases in r/H_s (Figure 5e) with the magnitude of sensitivity decreasing such that increases in r/H_s at larger values of r/H_s result in smaller decreases of $R_{2\%}/H_s$. The high variability in the influence of r/H_s may be due in part to variability in the way that r is defined for the different field sites (i.e., D_{50}). Given the compiled nature of the dataset used here, sediment sampling and analysis techniques are unlikely to be consistent across all datasets and, additionally, the representation of a sediment grainsize distribution with one value may not be fully appropriate for all sites.

The sensitivity analyses further confirm that H_s/L_p and tanβ are the primary factors influencing wave runup, with r/H_s having a smaller, but still significant influence on the runup. This is in agreement with previous research that identified wave height, length, and beach slope as the primary factors affecting wave runup (e.g., Holman, 1986, Stockton et al., 2006, Blenkinsopp et al., 2016, and Atkinson et al., 2017). The relative impact analysis also confirms that roughness or grainsize influences wave runup in agreement with the results of Poate et al. (2016). While the relative impact analysis allows for an assessment of the relative importance of the input variables, it does not provide insights into the functional form of the degree of influence. Figure 6, Figure 7, and Figure 8 show that a linear influence may be an incorrect assumption and this will be the focus of future work. The relative influence of each parameter changes dramatically across the parameter space as illustrated in Figure 6, Figure 7, and Figure 8. These figures show that the influence of wave steepness on runup is more significant at higher beach slopes and the influence of beach slope is most significant for lower values of wave steepness. Both trends are consistent with swash processes dominating over surf zone processes, i.e., reflective beaches.

7. Conclusions

This study has compiled a large dataset of wave runup observations collected under a wide range of conditions (laboratory and field, $0.019 \leq H_s \leq 7.17$ m, $0.81 \leq T_p \leq 23.7$ s, $0.009 \leq \tan\beta \leq 0.29$, $0.000003 \leq r \leq 0.125$ m; N = 1390) and used this novel dataset to assess the accuracy of commonly used empirical runup models and to develop a data-driven explicit GEP model to predict wave runup. We show that the best performing empirical models in the literature are prone to significant errors (minimum NMSE 0.23) when used with unseen data, i.e., uncalibrated models. Overall error values are significantly reduced (NMSE decreases between up to 58%) and correlations increased (by up to 23%) when individual models are optimised for the whole dataset. The best performing empirical model uses a Hunt type scaling with an additional parameter for wave induced setup as proposed by Holman (1986). These model types were also among the best performers in their non-optimised (original) form. The predictive ability of the explicit GEP model developed here was shown to be statistically significantly better than all other empirical models, confirming the impressive predictive ability of GEP models, albeit with a more complex expression. This highlights that the runup process is more complex than what is suggested by the simple empirical models that are widely used in the literature. Nevertheless, calculation of the runup from the explicit GEP model is still a trivial task with regard to parametric modelling of wave runup. Wave height, wavelength, and beach slope are confirmed to be the three primary factors influencing wave runup, with grain size having a smaller, but still significant influence on the runup. The high-degree of non-linearity between the key input variables and runup over the wide range of the data set is described and the new model developed here is shown to better account for this non-linearity. Sensitivity analysis demonstrates the importance of wave steepness and beach slope as key parameters for predicting runup and that normalised runup increases with increasing Iribarren number, i.e., as surf zone energy dissipation reduces.

Acknowledgements

The authors are grateful to those colleagues who have made their runup data freely available and to several colleagues who provided additional runup data or associated data directly: Hilary Stockdon and Rob Holman assisted with grainsize data for beaches in the U.S., Tim Poate provided a large dataset of runup from the U.K., Alexandre Nicolae Lerma provided data from Truc Vert, France, and Dan Howe provided data from the GWK flume. The authors gratefully acknowledge Michael Kinsela from NSW Office of Environment and Heritage and Manly Hydraulics Laboratory for providing wave and tide data for the NSW field experiments and the NSW State Government for providing details on permanent survey marks. Kim Colyvas from the University of Newcastle Statistical Support Service provided very helpful advice on the statistical analyses. This work was partially supported by ARC Grants LP100100375, DP13101122, and DP14101302. AA is supported by an Australian Postgraduate Award.

References

Atieh, M., Taylor, G., Sattar, A. M., & Gharabaghi, B., 2017, Prediction of Flow Duration Curves for Ungauged Basins. Journal of Hydrology. 545, p. 383–394.

Atkinson, A. L., Power, H. E., Moura, T., Hammond, T., Callaghan, D. P., and Baldock, T. E., 2017, Assessment of runup predictions by empirical models on non-truncated beaches on the south-east Australian coast: Coastal Engineering, v. 119, p. 15-31.

Azamathulla, H. M. (2012). Gene-expression programming to predict scour at a bridge abutment. *Journal of Hydroinformatics*, *14*(2), 324-331.

Azimi, H., Bonakdari, H., Ebtehaj, I. (2017). A Highly Efficient Gene Expression Programming Model for Predicting Discharge Coefficient in a Side Weir along a Trapezoidal Canal. Irrigation and Drainage. 66(4):655-666.

Azimi, H., Bonakdari, H., Ebtehaj, I., & Michelson, D. G. (2016). A combined adaptive neuro-fuzzy inference system–firefly algorithm model for predicting the roller length of a hydraulic jump on a rough channel bed. *Neural Computing and Applications*, 1-10.

Baldock, T. E., and Huntley, D. A., 2002, Long-wave forcing by the breaking of random gravity waves on a beach: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 458(2025), 2177-2201.

Blenkinsopp, C. E., Matias, A., Howe, D., Castelle, B., Marieu, V., and Turner, I. L., 2016, Wave runup and overwash on a prototype-scale sand barrier: Coastal Engineering, v. 113, p. 88-103.

Barbulescu, A., & Băutu, E. (2009). ARIMA and GEP models for climate variation. *International Journal of Mathematics & Computation™*, *3*(J09), 1-7.

Burnham, K. P. and Anderson, D. R., 2004, Multimodel Inference: Understanding AIC and BIC in Model Selection, Sociological Methods & Research, v. 33(2), 261-304.

Dreyfus, G., 2005, *Neural networks: methodology and applications*. Springer.

Ebtehaj, I., Bonakdari, H., Moradi, F., Gharabaghi, B., & Khozani, Z. S. (2018). An integrated framework of Extreme Learning Machines for predicting scour at pile groups in clear water condition. Coastal Engineering, 135, 1-15.

Ebtehaj, I., & Bonakdari, H. (2016). Bed load sediment transport estimation in a clean pipe using multilayer perceptron with different training algorithms. *KSCE Journal of Civil Engineering*, *20*(2), 581-589.

Ebtehaj, I., Bonakdari, H., Zaji, A. H., Azimi, H., & Sharifi, A. (2015). Gene expression programming to predict the discharge coefficient in rectangular side weirs. Applied Soft Computing, 35, 618-628.

Ebtehaj, I., Bonakdari, H., Zaji, A. H., Azimi, H., & Khoshbin, F. (2015b). GMDH-type neural network approach for modeling the discharge coefficient of rectangular sharp-crested side weirs. Engineering Science and Technology, an International Journal, 18(4), 746-757.

Ebtehaj, I., Bonakdari, H. (2017). No-deposition Sediment Transport in Sewers Using of Gene Expression Programming. Soft Computing in Civil Engineering. 1(1):26-42.

Ferreira, C. (2001). Gene Expression Programming: A New Adaptive Algorithm for Solving Problems. Complex Systems, 13:2, 87-129

Gazendam, E., Gharabaghi, B., Ackerman, J., & Whiteley, H., 2016, Integrative Neural Networks Models for Stream Assessment in Restoration Projects. Journal of Hydrology, v. 536, p. 339-350.

Gholami, A., Bonakdari, H., Zeynoddin, M., Ebtehaj, I., Gharabaghi, B., & Khodashenas, S. R. (2018). Reliable method of determining stable threshold channel shape using experimental and gene expression programming techniques. Neural Computing and Applications, 1-19.

Gurney, K., 1997, *An introduction to neural networks*. CRC press.

Harvey, R., McBean, E., & Gharabaghi, B., 2014, Predicting the Timing of Water Main Failure Using Artificial Neural Networks. Journal of Water Resources Planning and Management, v. 140, no. 4, p. 425-434.

Hedges, T. S., and Mase, H., 2004, Modified Hunt's Equation Incorporating Wave Setup: Journal of Waterway, Port, Coastal, and Ocean Engineering, v. 130, no. 3, p. 109-113.

Holman, R. A., 1986, Extreme value statistics for wave run-up on a natural beach: Coastal Engineering, v. 9, no. 6, p. 527-544.

Howe, D., 2016, *Bed shear stress under wave runup on steep slopes*, PhD Thesis, Water Research Laboratory, Faculty of Engineering, UNSW.

Hunt, I. A. 1958, Design of Seawalls and Breakwaters, US Lake Survey.

Khozani, Z.S., Bonakdari, H., Ebtehaj, I. (2017). An analysis of shear stress distribution in circular channels with sediment deposition based on Gene Expression Programming. International Journal of Sediment Research.32(4):575-584.

Mase, H., 1989, Random wave runup height on gentle slope, *Journal of Waterway, Port, Coastal and Ocean Engineering,* 115(5), 649-661.

Moradi, F., Bonakdari, H., Kisi, O., Ebtehaj, I., Shiri, J., & Gharabaghi, B. (2018). Abutment scour depth modeling using neuro-fuzzy-embedded techniques. Marine Georesources & Geotechnology, 1-11.

Nicolae Lerma, A., Pedreros, R., Robinet, A., Senechal, N., 2017, Simulating wave setup and runup during storm conditions on a complex barred beach, *Coastal Engineering,* 123, 29-41.

Nielsen, P., 1992, *Coastal Bottom Boundary Layers and Sediment Transport*, World Scientific, River Edge, N. J., 324pp.

Nielsen, P., and Hanslow, D. J., 1991, Wave Runup Distributions on Natural Beaches: Journal of Coastal Research, v. 7, no. 4, p. 1139-1152.

Poate, T. G., McCall, R. T., and Masselink, G., 2016, A new parameterisation for runup on gravel beaches: Coastal Engineering, v. 117, p. 176-190.

Robertson, B., Gharabaghi, B., & Power, H. E. (2017). Predicting breaking wave conditions using gene expression programming. *Coastal Engineering Journal*, *59*(03), 1750017.

Robertson, B., Gharabaghi, B., and Hall, K., 2015, Prediction of Incipient Breaking Wave-Heights Using Artificial Neural Networks and Empirical Relationships: Coastal Engineering Journal, v. 57, no. 4, p. 27.

Sabouri, F., B. Gharabaghi, A. A. Mahboubi, and E. A. McBean, 2013, Impervious surfaces and sewer pipe effects on stormwater runoff temperature, Journal of Hydrology, vol. 502, pp. 10–17.

Sharafi, H., Ebtehaj, I., Bonakdari, H., & Zaji, A. H. (2016). Design of a support vector machine with different kernel functions to predict scour depth around bridge piers. *Natural Hazards*, *84*(3), 2145-2162.

Singh, H., Thompson, A., and Gharabaghi, B., 2017, Event Runoff and Sediment-Yield Neural Network Models for Assessment and Design of Management Practices for Small Agricultural Watersheds. Journal of Hydrologic Engineering, v.22, no. 2, p. 1-12.

Stockdon, H. F., Holman, R. A., Howd, P. A., and Sallenger, A. H., 2006, Empirical parameterization of setup, swash, and runup: Coastal Engineering, v. 53, p. 573-588.

Trenouth, W. R., & Gharabaghi, B., 2016, Highway runoff quality models for the protection of environmentally sensitive areas. Journal of Hydrology, v. 542, p. 143–155.

Thompson, J., Sattar, A., Gharabaghi, B., & Warner, R., 2016, Event-Based Total Suspended Sediment Particle Size Distribution Model. Journal of Hydrology, v. 536, p. 236-246.

Vousdoukas, M. I., 2014, Observations of wave run-up and groundwater seepage line motions on a reflective-to-intermediate, meso-tidal beach: Marine Geology, v. 350, p. 52-70.

Wassing, F. 1957, Model investigation on wave run-up carried out in the Netherlands during the past twenty years, in: Proccedings of the 6th International Coastal Engineering Conference, American Society of Civil Engineers, pp. 700-714.

Figures

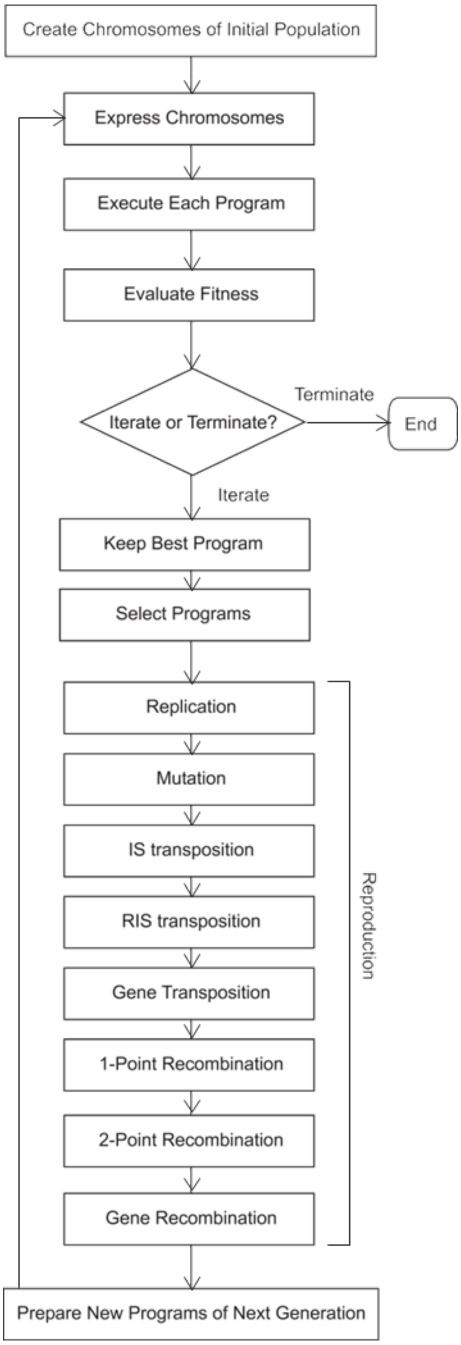


Figure 1. Flowchart detailing the GEP technique (from Ferreira, 2001). The reproduction processes shown here are detailed in Table 2.

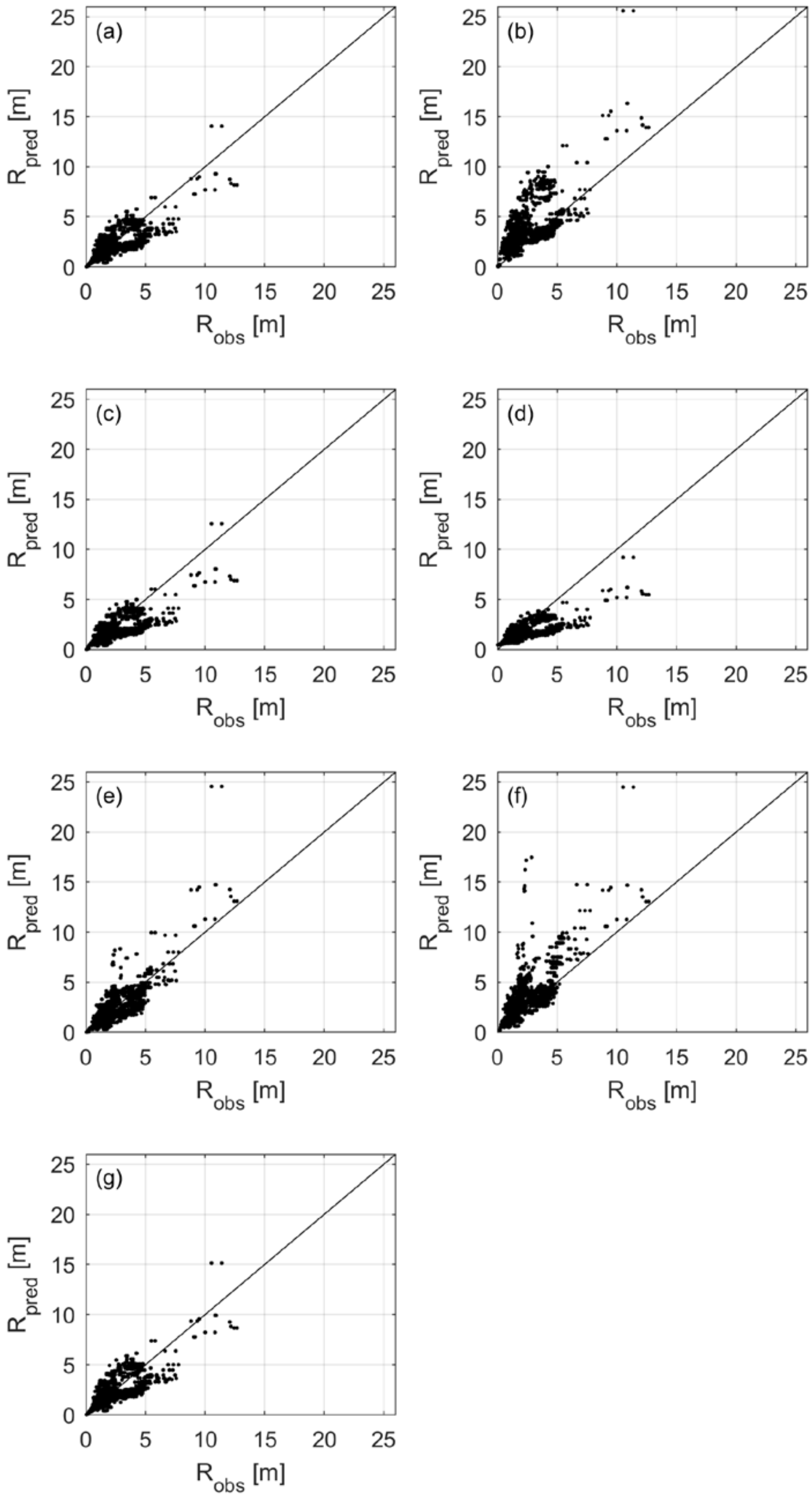


Figure 2. Comparisons of observed (R_{obs}) and predicted (R_{pred}) 2% runup exceedance levels for the seven empirical models tested: (a) Holman (1986), (b) Nielsen and Hanlsow (1991), (c) Stockdon et al. (2006), (d) Vousdoukas et al. (2014), (e) Poate et al. (2016) their Eq. 12, (f) Poate et al. (2016) their Eq. 9, and (g) Atkinson (2017) M2.

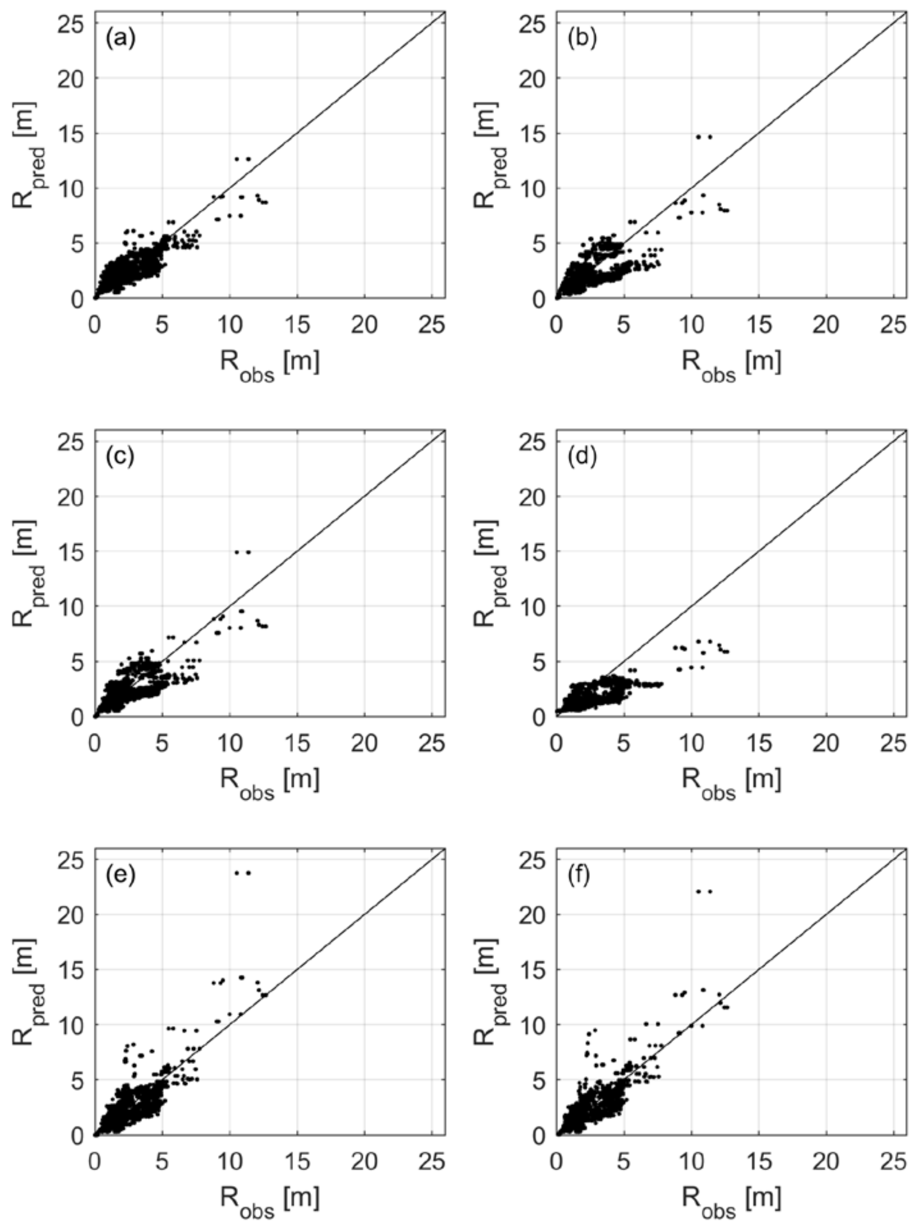


Figure 3. Comparisons of observed (R_{obs}) and predicted (R_{pred}) 2% runup exceedance levels for the six optimised empirical models: (a) Holman (1986)/Atkinson (2017) M2, (b) Nielsen and Hanlsow (1991), (c) Stockdon et al. (2006), (d) Vousdoukas et al. (2014), (e) Poate et al. (2016) their Eq. 12, and (f) Poate et al. (2016) their Eq. 9.

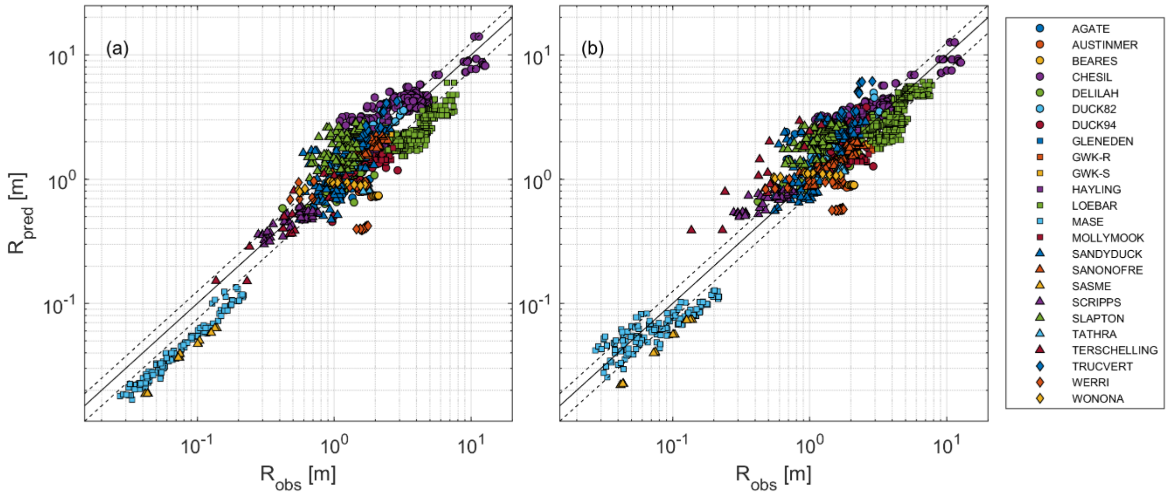


Figure 4. Comparisons of observed (R_{obs}) and predicted (R_{pred}) 2% runup exceedance levels for the (a) original and (b) optimised Holman empirical model (see Figure 2a and Figure 3a respectively) shown on a log scale with points coded by dataset. The 1:1 line is shown in black and the 25% error lines are shown by the dashed lines.

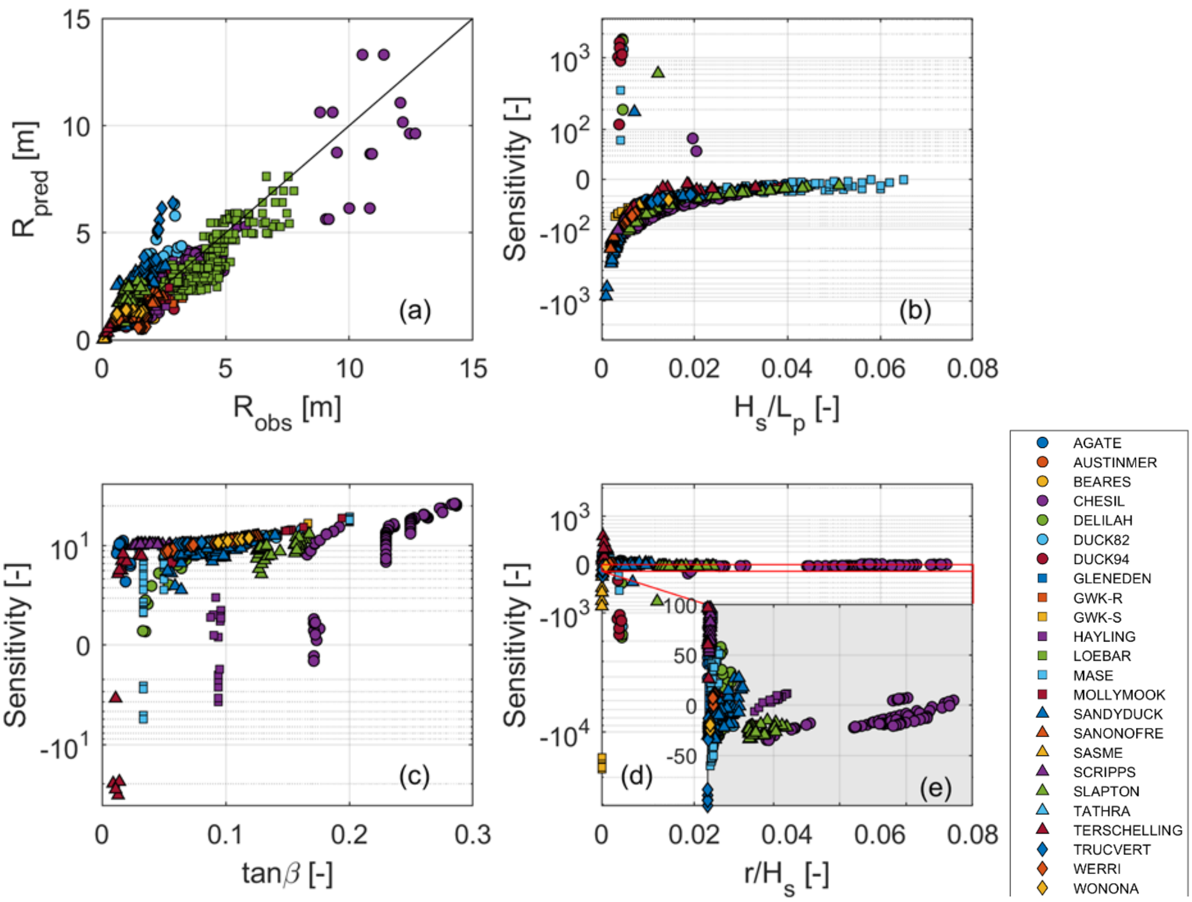


Figure 5. GEP Model predictions and sensitivity analysis: (a) comparison of observed and predicted runup as calculated using the GEP model; and partial derivative sensitivity for each of the three input variables: (b) H_s/L_p, (c) tanβ, and (d) r/H_s. The inset panel (e) in panel (d) shows the area delineated by the red box in panel (d). The 1:1 line is shown in black in (a).

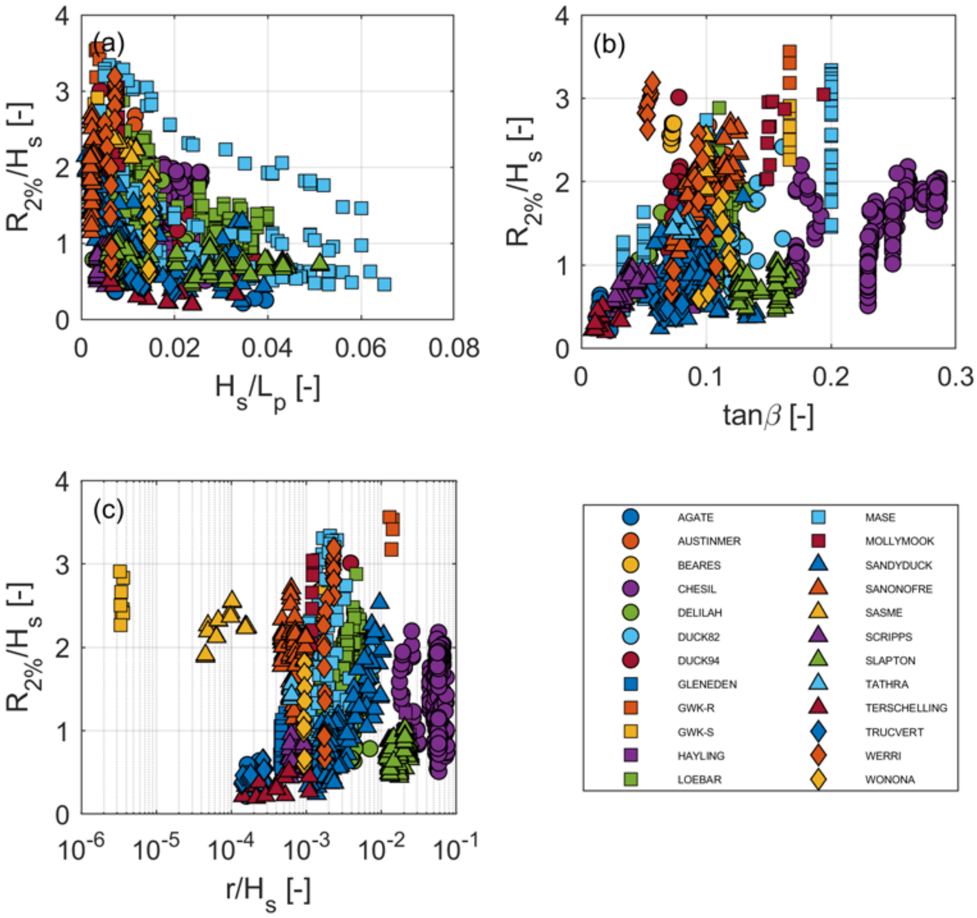


Figure 6. Non-dimensional runup plotted against (a) wave steepness, (b) beach slope, and (c) r/H_s for all datasets.

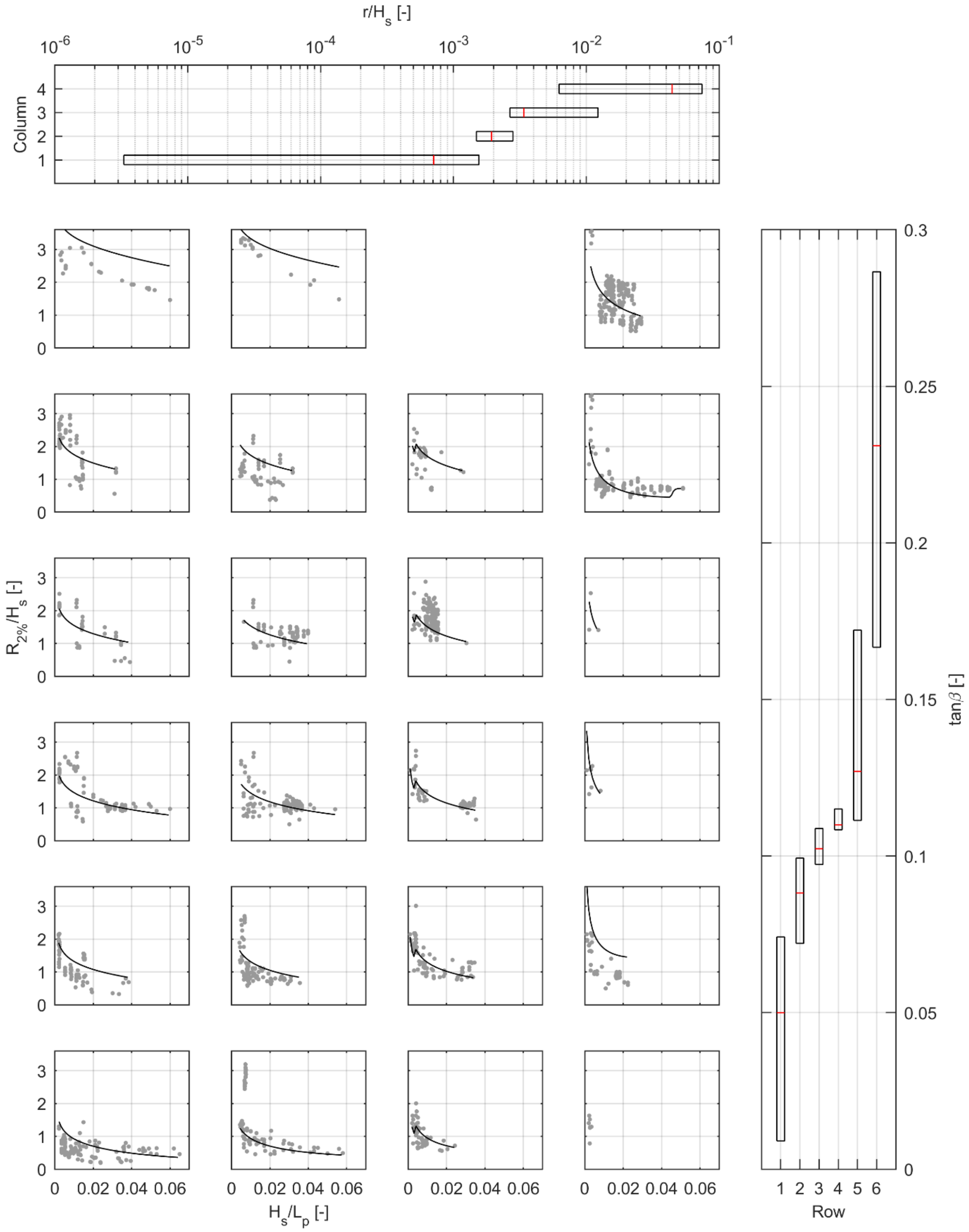


Figure 7. Coplot of the full dataset (grey points) for the three non-dimensional variables: H_s/L_p, tanβ, and r/H_s. Each column of panels represents one of four subdivisions of r/H_s shown in the top plot and each row of panels represents one of six subdivisions of tanβ shown in the right hand side plot. The subdivisions of tanβ and r/H_s each contain an equal number of data points with 10% overlap between each subdivision. The median value for each subdivision is shown by the red lines in the top and right hand side plots. The black solid lines in each panel represent the GEP model output ($R_{2\%}/H_s$) for the data range of H_s/L_p shown in each panel calculated using the median values tanβ and r/H_s for that subdivision, i.e., the median of all data presented in a given row or column

respectively not the median value of the data presented in each panel. The empty panel indicates there were no data for that combination of values of tanβ and r/H_s.

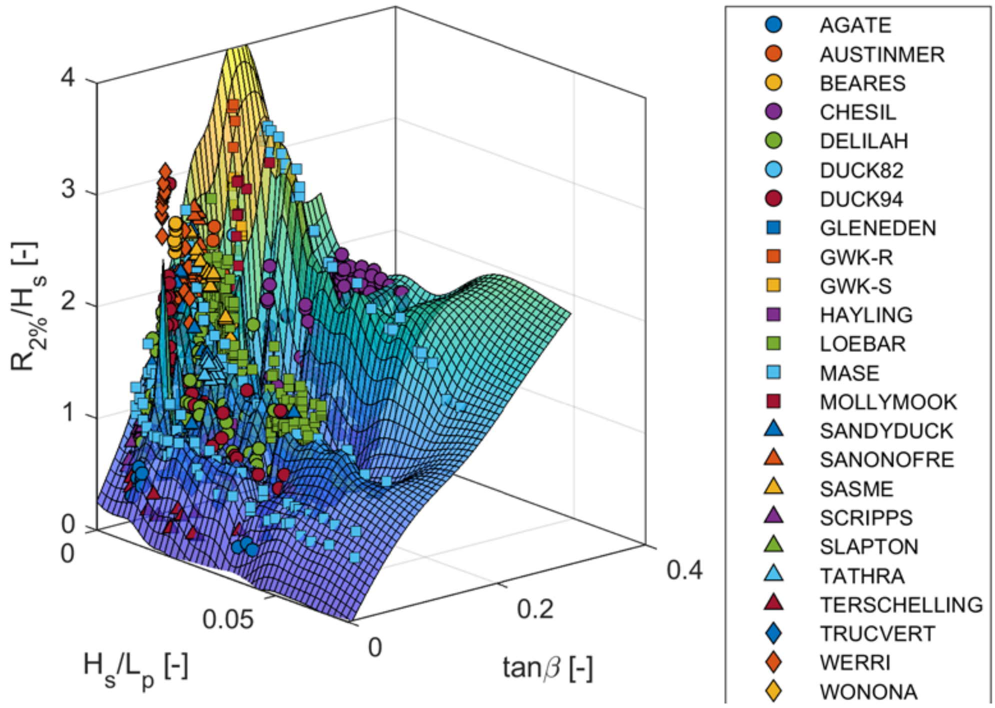


Figure 8. Three dimensional scatter plot of non-dimensional runup against wave steepness and beach slope for all datasets. A semi-transparent, thin-plate spline interpolated surface is also shown to assist with the visualisation.

Tables

Table 1. Examples of how genes and chromosomes can be converted to algebraic expressions. The genotype is translated to the phenotype by reading from left to right and creating the expression trees (ETs) shown. The single gene creates a single ET while the chromosome with two genes creates two sub-ETs that are combined through addition (in this example) to form a final ET. The ETs can then be converted to algebraic expressions for evaluation. Genes can be inferred from expression trees by reading the expression tree from top-to-bottom and left-to-right. Note that “Q” represents the square root function and the tail of each gene is shown in bold type. Modified from Ferreira (2001).

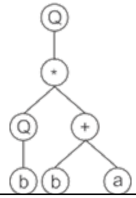
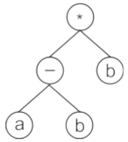
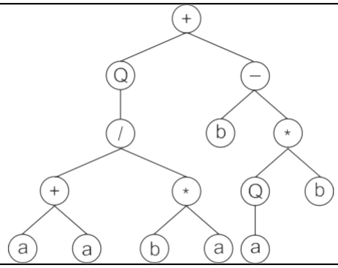
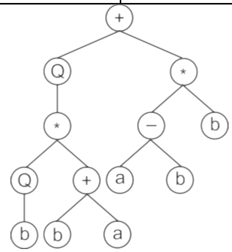
	Single gene	**Chromosome with two genes**
Genotype	`012345678901234567890` `+Q-/b*+*Qb`**`aabaabbaaab`**	`012345678012345678` `Q*Q+`**`bbaaa`**`*-ba`**`baabb`**
Head length	10	4
Tail length	11	5
Gene length	21	9
Sub-ET linking function	N/A	Addition
Sub-expression trees	N/A	Sub-ET$_1$  Sub-ET$_2$ 
Phenotype (expression tree)		
Algebraic expression	$\sqrt{2a/ba}+b-\sqrt{a}b$	$\sqrt{\sqrt{b}(b+a)}+(a-b)b$

Table 2. Examples of how genomes and chromosomes are reproduced through various methods of genome modification. One chromosome consisting of one gene is shown for replication, one chromosome with two genes is shown for mutation, insertion sequence transposition, root insertion sequence transposition, and gene transposition, and two chromosomes of two genes each are shown for one-point recombination, two-point recombination, and gene recombination. Tails of genes are shown in bold and modifications are highlighted in red. Modified from Ferreira (2001).

	Original genome	**Reproduced genome**
Replication	`012345678901234567890` `+Q-/b*+*Qbaabaabbaaab`	`012345678901234567890` `+Q-/b*+*Qbaabaabbaaab`
Mutation	`012345678012345678012345678` `-+-+abaaa/bb/ababb*Q*+aaaba`	`012345678012345678012345678` `Q+-+abaaa/bbQababb*b*+aaaba`
Insertion sequence transposition	`012345678901234567890012345678901234567890` `*-+*a-+a*bbabbaabababQ**+abQbb*aabbaaaabba`	`012345678901234567890012345678901234567890` `*-+*a-bba+babbaabababQ**+abQbb*aabbaaaabba`
Root insertion sequence transposition	`012345678901234567890012345678901234567890` `-ba*+-+-Q/abababbbaaaQ*b/+bbabbaaaaaaaabbb`	`012345678901234567890012345678901234567890` `-ba*+-+-Q/abababbbaaa+bbQ*b/+bbaaaaaaaabbb`
Gene transposition	`012345678012345678012345678` `*a-*abbab-QQ/aaabbQ+abababb`	`012345678012345678012345678` `-QQ/aaabb*a-*abbabQ+abababb`

One-point recombination	`012345678012345678`
	`-b+Qbbabb/aQbbbaab`
	`/-a/ababb-ba-abaaa`
		`012345678012345678`
		`/-aQbbabb/aQbbbaab`
		`-b+/ababb-ba-abaaa`

Two-point recombination	`0123456789001234567890`
	`+*a*bbcccac*baQ*acabab`
	`*cbb+cccbcc++**bacbaab`
		`0123456789001234567890`
		`*cbb+ccccac*ba*bacbaab`
		`+*a*bbccbcc++*Q*acabab`

Gene recombination	`012345678012345678012345678`
	`/aa-abaaa/a*bbaaab/Q*+aaaab`
	`/-*/abbabQ+aQbabaa-Q/Qbaaba`
		`012345678012345678012345678`
		`/aa-abaaaQ+aQbabaa/Q*+aaaab`
		`/-*/abbab/a*bbaaab-Q/Qbaaba`

Table 3. Field datasets used in this study showing field site and experiment, the mean (and range) of significant wave height (H_s), peak wave period (T_p), and swash zone slope (tanβ), beach face grain size (GS), and the number of data points from each location. The bottom row shows the mean (and range) of each parameter for the full field dataset.

Dataset	Beach (experiment)	H_s [m]	T_p [s]	*tanβ* [-]	*GS* [m]	Data points
Nicolae Lerma et al. (2017)	Truc Vert, France	4.30 (3.09-6.37)	14.6 (13.3-16.3)	0.069 (0.060-0.081)	0.00035	17
Atkinson et al. (2017)	Austinmer, Australia	0.73 (0.72-0.74)	6.4 (6.4-6.4)	0.11 (0.10-0.12)	0.000445	5
Atkinson et al. (2017)	Beares, Australia	0.79 (0.78-0.79)	8.6 (8.6-8.6)	0.073 (0.071-0.074)	0.000511	9
Atkinson et al. (2017)	Mollymook, Australia	0.89 (0.89-0.91)	8.5 (8.5-8.5)	0.16 (0.15-0.19)	0.000426	10
Atkinson et al. (2017)	Tathra, Australia	1.16 (1.14-1.19)	7.0 (7.0-7.0)	0.082 (0.07.-0.089)	0.00029	9
Atkinson et al. (2017)	Werri, Australia	0.65 (0.55-0.75)	8.0 (7.0-9.3)	0.074 (0.051-0.11)	0.000511	26
Atkinson et al. (2017)	Wonoona, Australia	0.93 (0.92-0.93)	6.4 (6.4-6.4)	0.11 (0.093-0.12)	0.000346	12
Poate et al. (2016)	Chesil, U.K.	2.54 (1.68-7.17)	10.1 (6.3-23.7)	0.24 (0.17-0.29)	0.05	214
Poate et al. (2016)	Loe Bar, U.K.	3.09 (1.43-5.69)	9.8 (7.1-20.0)	0.11 (0.098-0.12)	0.003	324
Poate et al. (2016)	Slapton Sands, U.K.	1.65 (1.04-2.09)	8.3 (4.8-11.7)	0.14 (0.13-0.17)	0.01	98
Poate et al. (2016)	Hayling Island, U.K.	2.53 (2.08-3.50)	11.8 (9.1-22.2)	0.094 (0.088-0.096)	0.02	26
Stockdon et al. (2006)	Agate, U.S.A.	2.48 (1.85-3.14)	11.9 (7.1-14.3)	0.016 (0.012-0.023)	0.0002	14
Stockdon et al. (2006)	Duck, U.S.A. (Delilah)	1.40 (0.52-2.51)	9.3 (4.7-14.8)	0.091 (0.033-0.13)	0.0015	138
Stockdon et al. (2006)	Duck, U.S.A. (Duck82)	1.71 (0.48-4.08)	11.9 (6.3-16.5)	0.12 (0.090-0.16)	0.0015	36
Stockdon et al. (2006)	Duck, U.S.A. (Duck94)	1.89 (0.73-4.06)	10.5 (3.8-14.8)	0.079 (0.056-0.095)	0.0015	52
Stockdon et al. (2006)	Gleneden, U.S.A.	2.06 (1.83-2.25)	12.4 (10.4-16.0)	0.080 (0.030-0.11)	0.0004	42
Stockdon et al. (2006)	Duck, U.S.A. (SandyDuck)	1.37 (0.35-3.57)	9.5 (3.7-15.4)	0.094 (0.053-0.14)	0.0015	95
Stockdon et al. (2006)	San Onofre, U.S.A.	0.81 (0.51-1.07)	14.9 (13.0-17.0)	0.10 (0.074-0.13)	0.0002	59
Stockdon et al. (2006)	Terschelling, The Netherlands	1.83 (0.51-3.93)	8.25 (4.8-10.6)	0.017 (0.009-0.032)	0.000225	14
Stockdon et al. (2006)	Scripps, U.S.A. (USWASH)	0.69 (0.54-0.84)	10.0 (10.0-10.0)	0.98 (0.025-0.055)	0.0002	41
		2.10 (0.35-7.17)	10.0 (3.7-23.7)	0.12 (0.0090-0.29)	0.011 (0.0002-0.05)	1242

Table 4. Laboratory datasets used in this study showing slope type, the mean (and range) of significant wave height (H_s) and peak wave period (T_p), the value or range of the swash zone slope (tanβ), hydraulic roughness length (r), and the number of data points from each dataset. The bottom row shows the mean (and range) of each parameter for the full laboratory dataset.

Dataset	Slope type	H_s [m]	T_p [s]	tanβ [-]	r [m]	Data points
Baldock and Huntley (2002)	Plane	0.039 (0.019-0.066)	1.61 (1.03-1.98)	0.1	0.000003	16
Howe (2016)	Plane	0.86 (0.82-0.91)	11.8 (9.8-13.7)	0.167	0.000003 & 0.01175	12
Mase (1989)	Plane	0.062 (0.026-0.11)	1.52 (0.81-2.50)	0.033 – 0.2	0.0001	120
		0.12 (0.019-0.091)	2.4 (0.81-13.7)	0.10 (0.33-0.20)	0.00040 (0.000003-0.01175)	148

Table 5. Summary of published empirical models for predicting 2% runup exceedance heights assessed in this study.

Authors	Empirical model	Applicability	Study conditions
Holman (1986)	$0.83 \tan\beta \sqrt{H_s L_p} + 0.2 H_s$		Field
Nielsen and Hanslow (1991)	$1.98 L_R + Z_{100\%}$ ** $L_R = 0.85 \tan\beta \sqrt{H_s L_s}$ $L_R = 0.085 \sqrt{H_s L_s}$	$\tan\beta \geq 0.1$ $\tan\beta < 0.1$	Field
Stockdon et al. (2006)	$0.043\sqrt{H_s L_p}$ $1.1\left(\frac{\sqrt{H_s L_p (0.563\beta^2 + 0.004)}}{2} + 0.35\beta\sqrt{H_s L_p}\right)$	$\xi_p < 0.3$ $\xi_p \geq 0.3$	Field
Vousdoukas et al. (2014)	$0.53\beta\sqrt{H_s L_p} + 0.58 \tan\beta\, H_s + 0.45$		Field
Poate et al. (2016) Eq. (12)	$0.33 \tan\beta^{0.5}\, T_p H_s$		Field + modelled
Poate et al. (2016) Eq. (9)	$0.21 D_{50}^{-0.15} \tan\beta^{0.5}\, T_{m-1m0} H_s$ **		Field + modelled
Atkinson et al. (2017) M2	$0.92 \tan\beta \sqrt{H_s L_p} + 0.16 H_s$		Model of models (derived from models fitted to field + large scale lab)

Note that: * indicates that the $Z_{100\%}$ is approximated as the tide varying water level (i.e., SWL); and ** indicates that the standard value of T_p was used in place of $T_{m\text{-}1m0}$ due to data availability.

Table 6. The optimal parameters used for the GEP model.

Parameter	**Optimal value**
Number of generations	190000
Population size	50
Function set	+, -, ×, /, exp, $\sqrt[2]{\ }$, $\sqrt[3]{\ }$, x^2, log, ln
Number of genes	5
Head size	8
Linking function	Addition
Mutation rate	0.004
Inversion rate	0.15
One point recombination rate	0.35
Two point recombination rate	0.35
Gene recombination rate	0.1
Gene transposition rate	0.1

Table 7. Performance of unmodified empirical relationships against full dataset as described by root mean square error (RMSE), normalised mean square error (NMSE), correlation (r^2), bias (B), scatter index (SI), standard deviation of prediction errors (S_e), and approximate 95% confidence interval of error uncertainty band ($B \pm 1.96 S_e$). Italicised numbers represent the empirical model that performed best or equal best according to each metric. Numbers in brackets represent the values for the gravel subset of the field dataset (i.e., $GS \geq 0.002$ m).

Authors	*RMSE* [m]	*NMSE* [-]	r^2 [-]	*B* [m]	*SI* [m]	S_e [m]	$B \pm 1.96 S_e$ [m]
Holman (1986)	*1.05*	*0.23*	0.68	-0.28	*0.45*	*1.01*	*-2.26 - 1.70*
Nielsen and Hanslow (1991)	2.14	0.55	0.57	1.27	0.92	1.72	-2.11 - 4.65
Stockdon et al. (2006)	1.21	0.36	0.63	-0.55	0.52	1.08	-2.67 - 1.58
Vousdoukas et al. (2014)	1.36	0.49	0.63	-0.68	0.59	1.18	-2.99 - 1.63
Poate et al. (2016) Eq. (12)	1.12 (1.38)	0.24 (0.16)	*0.72* (0.75)	*-0.09* (0.11)	0.48 (0.39)	1.12	-2.28 - 2.11
Poate et al. (2016) Eq. (9) *	2.02 (1.70)	0.45 (0.19)	0.56 (0.71)	1.01 (0.79)	0.79 (0.48)	1.75	-2.43 - 4.44
Atkinson et al. (2017) M2	1.06	*0.23*	0.66	-0.18	0.46	1.05	-2.23 - 1.88

Note that: * indicates that the relationship was only tested against the field data subset of the full dataset due to the inclusion of grain size in the relationship.

Table 8. Empirical runup relationships with optimised coefficients and performance of optimised empirical relationships against full dataset as described by root mean square error (RMSE), normalised mean square error (NMSE), correlation (r^2), bias (B), scatter index (SI), standard deviation of prediction errors (S_e), and approximate 95% confidence interval of error uncertainty band ($B\pm1.96S_e$). Performance of the GEP model is also shown. Italicised numbers represent the model that performed best or equal best according to each metric out of the optimised empirical models. Numbers in brackets represent the values for the gravel subset of the field dataset (i.e., *GS*≥0.002 m).

Authors	Empirical model	Applicability	RMSE [m]	NMSE [-]	r^2 [-]	B [-]	SI [-]	S_e [m]	$B\pm1.96S_e$ [m]
Holman (1986) & Atkinson et al. (2017) M2	$0.50\tan\beta\sqrt{H_sL_p}+0.70H_s$		*0.85*	*0.14*	*0.77*	*0.01*	*0.37*	*0.85*	*-1.67 - 1.68*
Nielsen and Hanslow (1991)	$1.85L_R+Z_{100\%}$** $L_R=0.52\tan\beta\sqrt{H_sL_s}$ $1.98L_R+Z_{100\%}$** $L_R=0.042\sqrt{H_sL_s}$	$\tan\beta\geq0.1$ $\tan\beta<0.1$	1.19	0.29	0.59	-0.22	0.51	1.16	-2.60 - 1.93
Stockdon et al. (2006)	$0.048\sqrt{H_sL_p}$ $1.2\left(\frac{\sqrt{H_sL_p(0.657\beta^2+0.008)}}{1.974}+0.352\beta\sqrt{H_sL_p}\right)$	$\xi_p<0.3$ $\xi_p\geq0.3$	1.07	0.23	0.65	-0.15	0.46	1.06	-2.23 - 1.93
Vousdoukas et al. (2014)	$0.005\beta\sqrt{H_sL_p}+4.563\tan\beta\,H_s+0.458$		1.02	0.21	0.68	-0.18	0.44	1.12	-2.93 - 1.44
Poate et al. Eq. (12)	$0.309\tan\beta^{0.48}\,T_pH_s$		1.10	0.24	0.72	-0.14	0.47	1.09	-2.28 - 1.99
Poate et al. Eq. (9)*	$0.359D_{50}^{-0.1}\tan\beta^{0.797}\,T_pH_s$ *		1.11	0.19	0.69	-0.08	0.43	1.11	-2.26 - 2.09
GEP model (Eq. 9)	Training Dataset		0.71	0.09	0.84	0.15	0.31	0.70	-1.21 – 1.52
GEP model (Eq. 9)	Testing Dataset		0.89	0.17	0.86	-0.45	0.37	0.77	-1.96 – 1.06
GEP model (Eq. 9)	Whole Dataset		0.75 (0.81)	0.10 (0.06)	0.82 (0.81)	0.03 (-0.22)	0.33 (0.23)	0.75 (0.78)	-1.45 – 1.51 (-1.75 – 1.30)

Note that * indicates that the relationship was only tested against the field data subset of the full dataset due to the inclusion of grain size in the relationship; and ** indicates that the $Z_{100\%}$ is approximated as the tide varying water level (i.e., SWL). Also note that the M2 model of Atkinson et al. has the same functional form as that derived by Holman (1986).

Table 9. Akaike information criterion (AIC) analysis showing the mean squared error value (MSE), the number of model parameters (*K*), the AIC value, the rescaled AIC values (Δ_i**)**, and the Akaike Weights (w_i). See text in Section 5.3 for explanation of parameters.

	Authors	**MSE [m^2]**	**K**	**AIC**	**Δ_i**	**w_i**
Original equations	Holman (1986)	1.10	3	133.62	642.28	3.39E-140
	Nielsen and Hanslow (1991)	4.58	4	2123.40	2632.06	0.00E+00
	Stockdon et al. (2006)	1.47	7	549.41	1058.07	1.75E-230
	Vousdoukas et al. (2014)	1.84	4	859.09	1367.75	9.91E-298
	Poate et al. (2016) Eq. (12)	1.26	2	326.57	835.23	4.28E-182
	Poate et al. (2016) Eq. (9) *	4.07	2	1954.36	2463.02	0.00E+00
	Atkinson et al. (2017) M2	1.13	3	174.55	683.21	4.39E-149
Optimised equations	Holman (1986) & Atkinson et al. (2017) M2	0.73	3	-432.09	76.58	2.35E-17
	Nielsen and Hanslow (1991)	1.44	4	517.54	1026.20	1.46E-223
	Stockdon et al. (2006)	1.15	7	202.95	711.61	3.00E-155
	Vousdoukas et al. (2014)	1.80	4	824.24	1332.91	3.66E-290
	Poate et al. (2016) Eq. (12)	1.21	2	264.39	773.05	1.36E-168
	Poate et al. (2016) Eq. (9) *	1.23	2	295.71	804.37	2.16E-175
	GEP model (Eq. 9)	0.57	137	-508.66	0.00	1.00E+00

Note that * indicates that the relationship was only tested against the field data subset of the full dataset due to the inclusion of grain size in the relationship.