Assessment of runup predictions by empirical models on non-truncated beaches on the south-east Australian coast

Alexander Atkinson[1*], Hannah E. Power[2], Theo Moura[1], Tim Hammond[1,3], David P. Callaghan[1] and Tom E. Baldock[1]

[1]School of Civil Engineering, University of Queensland, St Lucia, QLD, 4072, Australia
[2]School of Environmental and Life Sciences, University of Newcastle, Callaghan, NSW 2308, Australia.
[3] Bureau of Meteorology, Melbourne, Australia.

Abstract

This paper assesses the accuracy of 11 existing runup models against field data collected under moderate wave conditions from 11 non-truncated beaches in New South Wales and Queensland, Australia. Beach types spanned the full range of intermediate beach types from low tide terrace to longshore bar and trough. Model predictions for both the 2% runup exceedance ($R_{2\%}$) and maximum runup (R_{max}) were highly variable between models, with predictions shown to vary by a factor of 1.5 for the same incident wave conditions. No single model provided the best predictions on all beaches in the dataset. Overall model root mean square errors are of the order of 25% of the $R_{2\%}$ value. Models for $R_{2\%}$ derived from field data were shown to be more accurate for predicting runup in the field than those developed from laboratory data, which overestimate the field data significantly The most accurate existing models for predicting $R_{2\%}$ were those developed by Holman (1986) and Vousdoukas *et al.* (2012), with mean RMSE errors of 0.30m or 25%. A new "model of models" for $R_{2\%}$ was developed from a best fit to the predictions from six existing field and one large scale laboratory $R_{2\%}$ data-derived models. It uses the Hunt (1958) scaling parameter $tan\beta\sqrt{H_o L_o}$ and incorporates a setup parameterisation. This model is shown to be as accurate as the Holman and Vousdoukas *et al.* models across all tidal stages. It also yielded the smallest maximum error across the dataset. The most accurate predictions for R_{max} was given by Hunt (1958) but this still tended to under predict the observed maximum runup obtained for 15-minute records. Mase's (1989) model has larger errors but yields more conservative estimates. Greater observed values of R_{max} are expected with increased record length, leading to greater differences with predicted values. Given the large variation in predictions across all models,

however, it is clear that predictions by uncalibrated runup models on a given beach may be prone to significant error and this should be considered when using such models for coastal management purposes. It should be noted that in extreme events, which are lacking in the dataset, runup may truncated by beach scarps, cliffs, and dunes, or by overtopping, and, as a result, the probability density functions will have different tail shapes. The uncertainty already present in current models is likely to increase in such conditions.

Keywords: Runup; swash; model accuracy, remote sensing; beaches.

.

1. Introduction

Runup is the final stage of a wave's landward propagation, and thus the determinant of the most landward position a wave can reach before receding seaward. Runup above the local ocean level outside the surf zone results from a combination of two processes: wave induced set up and swash (i.e., Holman, 1986). Past research has focussed on modelling maximum wave runup values, most commonly R_{max} (Hunt (1958), Mase (1989) and Douglass (1992)) and $R_{2\%}$ (Wassing (1957), Mase (1989), Hedges and Mase (2004), Holman (1986), Nielsen and Hanslow (1991), van der Meer and Stam (1992), Ruggiero *et al*. (2001), Soldini *et al*. (2013), Stockdon *et al*. (2006), and Vousdoukas *et al*. (2012)). R_{max} is the greatest elevation obtained by a single runup event within a given time period and is therefore a function of record length. $R_{2\%}$ is a statistical measure of the elevation exceeded by only 2% of all runup or swash events within a time period.

The importance of being able to predict maximum runup values for different wave and beach conditions is obvious with regard to hazard risk assessment. Typical applications include assessing overtopping swash flows (Peregrine and Williams, 2001; Baldock *et al*., 2012), forecasting beach erosion with respect to climate change (Bruun, 1954; Kriebel and Dean, 1993), or for design purposes, such as for beach nourishment or the positioning of temporary structures near the shoreline. A common goal has been to develop empirical models for predicting runup elevations that makes use of readily available or easily obtainable parameters. Models typically include a combination of (though not necessarily all of) wave height and length (H and L, respectively) and the beach (or swash zone) slope (β). Other important factors

to take into account may be related to time-varying ocean water levels (i.e. tidal elevation), which can change the surf zone characteristics considerably.

Empirical runup models have been developed from both laboratory (e.g. Wassing, 1957, Hunt, 1958, Mase, 1989) and field data (e.g. Holman, 1986, Nielsen and Hanslow, 1991, Stockdon *et al*., 2006). Laboratory conditions are useful for separating the influence of different variables and excluding 3D effects. However, scale effects can be present in smaller scale models and often result in distorted dimensions of some variables, sediment size being a common example (Hughes, 1993), such that when scaled up to a prototype, a grain of sand may be more representative of gravel, which could result in different runup distributions. Field data on the other hand makes detection of the most influential variables more difficult and obtaining a wide variety of representative conditions is not always possible, such that the range of beaches and/or conditions used to create a model can be limited (e.g. Holman, 1986, Douglass, 1992, Ruggiero *et al*., 2001, Vousdoukas *et al*., 2012). Thus, individual models may not be applicable to beaches or wave conditions far beyond the parameter space initially used to develop the model. The method by which the runup is measured may also influence the recorded values. Recently, there has been a trend to measure runup through video analysis (e.g., Holman, 1986; Ruggiero *et al*., 2001; Stockdon *et al*. 2006 and Vousdoukas 2012), while previous work used resistance wires (e.g., Mase, 1989, Hedges and Mase, 2004; Van der Meer and Stam, 1992) or simply counted the number of waves passing known locations up the beach (Nielsen and Hanslow, 1991). Holman and Guza (1984) provide a discussion on the pros and cons between the resistance wire and image analysis techniques.

The use of large data sets and subsequent fitting to an empirical model by coefficients can result in reduced accuracy when considering specific parameter spaces (Stockdon *et al*. 2006; 2014). Many other potential influences on runup have also been excluded because they are unknown or cannot be easily parameterised. For example, the nearshore bathymetric profiles of field sites used to develop runup models differ significantly, but this is not typically included as a model parameter. Variation in the nearshore bathymetric profile could lead to variable wave energy attenuation due to different shoreface slopes (Wright, 1976) and result in varying correlations between runup elevations and offshore wave conditions. Therefore, a model developed using data collected on the north-east coast of the United States (e.g. Duck, NC, Holman, 1986) may provide less accurate runup predictions on Australian beaches compared

with a model developed locally (e.g. Nielsen and Hanslow, 1991), or a model developed from planar laboratory beaches.

As a result, runup models that have been developed using data from a specific site, or from a limited range of field data sites, may not be the best model for other locations. Here, we investigate the performance of a range of available runup models applied to data from the southeast Australian coast. We determine model limitations and error margins that should be considered when using empirical models to forecast runup when no data for calibration is available.

Despite the absence of tidal water levels as a parameter in the runup models assessed below, the potential effect of tidal elevation on runup exceedance values is of interest because the surf zone conditions often differ between high and low tide (Short, 1993 & 2000). Vousdoukas *et al*. (2012) observed that runup models tended to over predict runup at low tidal stages and under predict during higher tidal stages, suggesting variability in wave energy dissipation at different stages of the tidal cycle. The south-east Australian coast is microtidal with a very steep lower shoreface, suggesting that the tidal influence may be less than in other regions (Short, 1993). Despite this, Power *et al*. (2013a) observed differences in model performance depending on whether the tide was rising or falling; however, their observations were limited to only a few beaches, suggesting this is a factor worth investigating further with additional data from a wider range of beaches.

A caveat that must be noted for models derived from field data is that the most extreme runup events often occur in scenarios where the runup is truncated by a scarp or cliff, or overtopping occurs. None of the current empirical models are valid for these morphological conditions. After the 2011 Tohoku Tsunami, runup on coasts lined by 16m cliffs had the highest watermarks of 21m above mean sea level (Sato et al., 2013). Callaghan et al. (2007) reported extensive destruction of buildings due to wave impacts on cliffs overtopping 20m high cliff faces. Runup data used for model calibration are, by necessity, obtained from conditions where the runup is not truncated and therefore may not include extreme events. This is also the case in the present study, where the existing and new models have been derived for runup on non-truncated largely planar beaches, with no impact on dunes or cliffs. Therefore, extreme conditions are typically, but not always (Fiedler *et al*., 2015), outside the parameter space used to develop the empirical models, potentially limiting their use to less severe wave conditions.

However, the widely observed and consistent scaling on H and L allows a degree of extrapolation and, therefore, application of such models to more extreme conditions.

The present paper addresses these issues and examines the accuracy of a suite of runup models, assessed for moderate wave conditions on 11 largely planar beaches along the south-east Australian coast. Beach states ranged from longshore bar and trough to low tide terrace (Wright and Short, 1984). The geographically diverse dataset and range of wave and beach conditions allows for a comprehensive assessment of the accuracy and typical error margins of common empirical runup models applied to beaches falling within this range of beach states. A total of 11 runup models are assessed, *viz.*: Wassing (1957), Hunt (1958), Mase (1989), Holman (1986), Nielsen and Hanslow (1991), Douglass (1992), van der Meer and Stam (1992), Ruggiero *et al*. (2001), Hedges and Mase (2004), Stockdon *et al*. (2006), and Vousdoukas *et al*. (2012). These models have been developed using both laboratory and field data. Consistent differences in performance between the models developed from small scale laboratory data and those developed from field data are identified. Excluding small scale laboratory derived models, two different "*model of models*" are also developed by taking the best fit to predictions made by other models for a range wave and beach conditions, represented by the Iribarren number (Iribarren and Nogales, 1949). The model of models is then assessed in conjunction with the assessment of the existing empirical models. The paper is organised as follows. A brief outline of each of the models assessed follows in Section 2. Details of the field data sites, data collection, and analysis techniques are provided in Section 3. A comprehensive analysis of the results is presented in Section 4, followed by a discussion in Section 5. Concluding remarks are made in in Section 6.

2. Selected runup models

Numerous empirical runup models are available which have been derived from laboratory or field data. A total of 11 models are described in this section, distinguished by the data type from which they were derived, i.e. laboratory data or field data. For later reference in figure captions, each model is given an abbreviation (Table 1).

2.1. Runup models derived from laboratory data

The laboratory data used for model development consisted of regular (monochromatic) wave and random wave experiments. Wassing (1957) used small scale flume data with

combinations of waves with and without wind to assess wave runup distributions on an impermeable plane slope to obtain:

$$R_2 = 8tan\beta H_{st} \quad (1)$$

where β is the beach slope and H_{st} is the significant wave height at the toe of the slope, however, this is taken to be the deep water significant wave height, H_s, for the purpose of assessing runup predictions using easily obtainable offshore values. It should be noted that the above formula (1) was derived for waves with steepness in the range $0.05 < H/L < 0.07$.

Hunt (1958) analysed regular wave data from small and large scale laboratory experiments on structural slopes with monochromatic waves and found the relationship:

$$R_{max} = \xi H_0 = tan\beta\sqrt{H_0 L_0} \quad (2)$$

where ξ is the surf similarity parameter, or Iribarren number ($\xi = \frac{tan\beta}{\sqrt{H_0/L_0}}$; Iribarren and Nogales, 1949; Battjes, 1974) and H_0 and L_0, the deep water wave height and wavelength respectively. The deep water wavelength is calculated assuming linear wave theory, $L = \frac{gT^2}{2\pi}$ (where g is the gravitational acceleration, and T is the wave period). Although Hunt only assessed results from regular wave experiments Hunt suggested the steady wave height be considered as the significant wave height for design purposes but suggests that larger scale and/or wind generated (random) waves will affect runup distributions, requiring different coefficients, although the concepts and relationships are described as fundamental. This formulation, or scaling, forms the basis for nearly all subsequent models. The use of the Iribarren number and the proportionality of runup with the square root of wave height and wave length was proposed originally by Hunt because of the relationship with energy dissipation and wave reflection identified by Iribarren and Nogales (1949).

Mase (1989) examined runup distributions in a small-medium sized wave flume (27m long x 0.5m wide x 0.75m deep) with random (Pierson-Moskowitz) wave conditions on a variety of planar impermeable slopes ($0.03 < \tan\beta < 0.2$), providing:

$$R = aH_s{\xi_s}^b \quad (3)$$

where ξ_s is calculated with respect to the deep water significant wave height (H_s) and significant period (T_s), a and b are changeable coefficients obtained through least-squares fitting to their data depending on the run-up prediction desired: for $R_{2\%}$, a = 1.86, b = 0.71; for R_{max}, a = 2.32, b = 0.77. Analysis was conducted on experiments that each had between 650 and 900 individual runup events recorded. The effect of the added exponent, b, causes the runup predictions plotted against the Iribarren number to deviate from the straight line produced by the Hunt formula and increases the predictions of $R_{2\%}$ and R_{max} for all values of ξ_s < 3.

Van der Meer and Stam (1992) analysed large scale wave flume data for wave run-up on smooth and rocky slopes to obtain:

$$R_2 = Cp tan\beta \sqrt{H_{st} L_p} \quad (5)$$

where H_{st} is the significant wave height at the toe of the slope (again, assumed as the deep water significant wave height, H_s, for the purposes of runup predictions using offshore wave heights), L_p is the peak period deep water wave length, and Cp is a constant varying between 1.3<Cp<1.7, depending on the type of swell arriving (lower values for developed ground-swell and higher values for sea states). They found Cp = 1.5 provided the most reliable predictions for a large set of wave run-up data on impermeable slopes with wave conditions in the range 0.5<ξ_p<2. The majority of wave conditions for the present research fell within this range (see Table 3), so Cp = 1.5 has been adopted in equation (5) in the present analysis.

Hedges and Mase (2004) used the data of Mase (1989) and adapted the model of Hunt (1958) to incorporate wave set up and remove the exponent fitted around the original model of Mase (above):

$$R_2 = 1.49 tan\beta \sqrt{H_s L_p} + 0.34 H_s \quad (4)$$

where H_s is the deep water significant wave height, L_p is the wavelength corresponding to the deep water peak wave period, T_p, and the second term on the right corresponds to the wave setup.

2.2. Runup models derived from field data

Holman (1986) examined field data in 35-minute-long time-series from Duck, North Carolina. Holman aimed to improve on the original formulations by Hunt and Wassing by incorporating wave setup prior to the work of Hedges and Mase (2004). For deep water waves,

$R_{2\%}$ is calculated using the Iribarren number with peak wave period and significant wave height:

$$R_2 = 0.83tan\beta\sqrt{H_s L_p} + 0.2H_s \quad (6)$$

where the second term on the right represents the setup. In Holman's original paper the coefficients were reported as 0.83 (± 0.06) and 0.2 (± 0.1), for offshore wave conditions.

Nielsen and Hanslow (1991) analysed runup distributions from six beaches in New South Wales, Australia, to extract a runup length scale, L_R. Assuming a Rayleigh distribution of runup excursions, L_R is the gradient of the log-ranked distribution (see Figure 4 for an example). Using a linear best fit to the log-ranked runup values, a value of percentage exceedance runup (i.e. $R_{2\%}$) is calculated by:

$$R_2 = 1.98L_R + Z_{100} \quad (7)$$

where 1.98 is the coefficient to obtain the 2% runup exceedance and $Z_{100\%}$ is the highest vertical level passed by all swash events in a time period (Figure 4). As $Z_{100\%}$ is not known *a priori*, a forecastable parameter is required instead. This may reduce the model accuracy. However, as the predictive accuracy of models for runup is being assessed, the tide varying water level reduced to the common datum (metres above the Australian Height Datum which is approximately mean sea level) will be taken as $Z_{100\%}$. This is an approximation, however, since from laboratory data it is known that the minimum run-down location is frequently below the still water level (Baldock and Huntley, 2002) and there is some evidence from large scale wave flumes that $Z_{100\%}$ occurs below the still water level (Blenkinsopp *et al*., 2015). As noted by Nielsen (2009), $Z_{100\%}$ is also dependent on beach slope as a result of the merging of bores in the inner surf zone, and is below the still water level on mildly sloping beaches.

Based on the data set of Nielsen and Hanslow (1991), Nielsen (2009) provides two updated formulae to calculate the runup length scale, dependent on the swash zone slope:

$$\begin{aligned} L_R &= 0.6tan\beta\sqrt{H_{rms}L_s} \quad if tan\beta \geq 0.1 \\ L_R &= 0.06\sqrt{H_{rms}L_s} \quad if tan\beta < 0.1 \end{aligned} \quad (8)$$

where L_s is offshore deep water wavelength calculated using the significant wave period, T_s, and H_{rms} is the root-mean-square offshore (deep water) wave height.

Douglass (1992) suggested that since slopes on natural beaches are difficult to predict *a priori*, a formula that does not incorporate the beach slope would be useful. Douglass analysed Holman's (1986) field data set which was analysed in 17-minute records, and proposed the following:

$$R_{max} = 0.12\sqrt{H_{m0}L_p} \tag{9}$$

where H_{m0} is the spectral significant wave height. In the present research H_s was used in place of H_{m0} (Table 3), where $H_{m0} \approx H_s$. Douglass found that runup behaved independently of beach slope, however, the model was only developed on data taken from one beach with wave heights varying from 0.8m < H_{m0} < 4m, and period from 6s < T_p < 16s, measured in 8m water depth with swash zone slopes 0.07 < tanβ < 0.16.

Ruggiero *et al*. (2001) analysed runup values from 74 two-hourly timestacks (tide de-trended) in order to investigate the different $R_{2\%}$ values observed between the dissipative beaches of Oregon and the more variable conditions at Duck. They found it was necessary to incorporate the beach slope to account for different beach states and provided a single model using deep water wave values for predicting $R_{2\%}$:

$$R_2 = 0.27\sqrt{\beta H_s L_p} \tag{10}$$

where the foreshore beach slope, β, is defined as the mean slope over the beach face encompassed by $\pm 2\sigma$ around the time averaged shoreline, where σ is the standard deviation of the continuous shoreline elevation.

Stockdon *et al*. (2006) collated data from numerous experiments, comprising of ten datasets from six beaches and a total of 491 individual 17 minute timestacks. However, the majority (approx. 91%) were from the same beach at Duck, NC. Mean significant wave height, peak wave period, and swash zone slopes ranged from 0.7m < $H_{s,local}$ < 2.5m, 8.0s < T_p < 14.9s, and 0.01 < β < 0.11, respectively. It should be noted here that the significant wave height

($H_{s,local}$) described were those provided by local buoys in variable water depths (7 to 20m). These local values were then reverse-shoaled to provide an offshore (deep water) value used in the formulations below. They proposed two models, depending on the offshore Iribarren number, ξ_p:

$$R_2 = 1.1\left(\frac{\sqrt{H_s L_p(0.563\beta^2+0.004)}}{2} + 0.35\beta\sqrt{H_s L_p}\right) \qquad if\,\xi_p \geq 0.3 \tag{11}$$

where the term $0.35\beta\sqrt{H_s L_p}$ represents the setup,

or

$$R_2 = 0.043\sqrt{H_s L_p} \qquad if\,\xi_p < 0.3 \tag{12}$$

where β is the swash zone beach slope defined in the same manner as Ruggiero *et al*. (2001). However, for simplicity in the present study, the definition of β is taken in the same way as in the remainder of the studies from which the models were derived, such that β is the daily averaged slope between the lowest and highest shoreline location measured. This is not expected to yield significantly different values of β from those that would result from the Ruggiero *et al*. definition.

Finally, Vousdoukas *et al*. (2012) assessed a dataset comprising of 456 10-minute timestacks to investigate their applicability on the European Atlantic coast (south coast Algarve, Portugal) using deep water significant wave height and peak period. Similarly, to the methodology of the present paper, they manually selected the runup maxima, rather than relying on automatic value extraction (*cf*. Stockdon *et al*. 2006) and provided a model that is based on runup measurements from a typically reflective beach:

$$R_2 = 0.53\beta\sqrt{\left(H_s L_p\right)} + 0.58 tan\beta H_s + 0.45 \tag{13}$$

It should be noted that this model yields a finite value for the runup as the offshore wave height tends to zero.

2.3 Model comparison and model of models

Many of the models contain the same parameters (swash zone slope, wave height and wavelength). Frequently, the difference between models is simply a different coefficient in front of a Hunt-type formula and a second term to incorporate wave setup. To our knowledge, there is no theoretical basis for the proportionality to the square root of the wave height as observed for breaking wave runup. However, in general, it might be expected that short wave swash is related to the bore height at the shore, which is related in a complex manner to the offshore wave height and wave period and dependent on whether the surf zone is saturated or unsaturated (Power *et al.*, 2013b). We would though expect a relationship weaker than linear because of surf zone dissipation, as proposed by Hunt (1959). It is noted that solutions for non-breaking waves do exist, which indicate a linear relationship with wave height (see e.g. Antuono and Brocchini, 2008). Considering all of these models have been fitted to be the most accurate for the data sets upon which they were built, it is of interest to investigate what a combined '*model of models*' may provide. The '*model of models*' is developed by analysing the predictions from all the field derived models discussed above plus Van der Meer and Stam's (1992) model (developed from data from a large scale wave flume), for a range of the parameter values used in each model, with the predictions then plotted versus the Hunt scaling of $tan\beta\sqrt{H_s L_p}$. A least squares analysis provided new coefficients for the line of best fit to all the predictions from all models. The '*model of models*' is derived from the predictions of large scale laboratory and field-derived run-up models only, since, as is clearly shown in section 4, the models derived from smaller scale laboratory experiments consistently predict higher runup than the field-derived models, and over-estimate the field data reported below.

The use of different wave height and period definitions between models is accounted for through the usual approximations: $H_s \approx H_{rms}\sqrt{2}$ and $H_s \approx H_{m0}$. The wave period used is taken as the peak period, as commonly applied in most models, with the exception of Mase (1989), Nielsen and Hanslow (1991), and Nielsen (2009). However, in the absence of a method of converting T_p to T_s, the period used to define the wavelength will be taken as T_p in those models also.

The scatter in the predictions from the individual models is considerable (Figure 1), highlighting the variability of wave runup predictions and the potential dependence of each model on local conditions. Taking all the model predictions, two least-squares lines of best fit are calculated (one forced through the origin and the other with a calculated intercept) to

provide two new '*model of models*' that are derived without any tuning to the current data (Figure 1).

$$\text{MM1} \qquad R_2 = 0.99 tan\beta \sqrt{H_s L_p} \tag{14}$$

$$\text{MM2} \qquad R_2 = 0.92 tan\beta \sqrt{H_s L_p} + 0.16 H_s \tag{15}$$

where (14) corresponds to the line of best fit forced through the origin and it is noted that it is practically identical to Hunt's formula (2). Equation (15) is not forced through the origin, and interestingly is very similar to Holman (1986) The second term of equation (15) represents the setup (*c.f.* Holman, 1986), with the coefficient derived from the intercept of $R_{2\%}/H_s$ at tanβ $(H_s L_p)^{0.5}/H_s = 0$. Eq. (14) provides only a marginally worse fit to the predicted dataset than (15), with R^2 values of 0.71, 0.72, respectively. All the current $R_{2\%}$ data is provided in Figure 1, which is zoomed in and highlights the variability inherent in the existing models and data when plotted against the Iribarren number. This provides an initial indication of model performance to be analysed in more detail below.

3. Field data

3.1. Study Sites

Video observations and survey data have been used in conjunction to measure wave runup excursions on 11 beaches along the coasts of New South Wales and south east Queensland, Australia. Table 2 details the locations, abbreviated site names, the number of data runs (timestacks) and the total number of swash events observed at each location. Figure 2 illustrates the range of cross shore profiles. Table 3 provides the average beach slope and offshore wave conditions measured from the nearest offshore wave rider buoy during the period of data collection.

Beach profiles were measured, using either a total station or real-time kinematic (RTK) GPS, from the upper beach to the maximum water depth in the inner surf zone achievable without risking damage to the RTK survey equipment. All of the empirical models analysed in this study use the swash zone slope as their measure of beach slope. In contrast, Nielsen and Hanslow (1991) suggested that the surf zone slope would also be useful as this represents a larger region of the beach that influences wave transformation; however, they acknowledge

that obtaining accurate profile data to and beyond the main breaker bar is often too difficult with manual survey techniques. Stockdon *et al*. (2006) found that parameterisations including surf zone slope did not improve runup predictions derived using the beach face (swash zone) slope only. Recently, Blenkinsopp *et al*. (2015) performed an analysis of the differences when using the swash zone slope and surf zone slope in runup predictions for experiments in the BARDEX II large scale wave flume test series, and found the models based on swash zone slope to perform significantly better than those using the surf zone slope.

Offshore wave and tide conditions were obtained for each site during the relevant period of data collection. For wave conditions, deep water wave rider buoy data (typically in water depths of around 80m) were obtained from the nearest appropriately situated buoy. Tide conditions were obtained from the nearest tide gauge; it is noted that the atmospheric ocean tide along this coast is very homogeneous and has very little lag between different locations.

Most locations were characterised as being in the intermediate range of beach states (low tide terrace to longshore bar and trough) as described by the Wright and Short Australian beach model (Wright and Short, 1984) which is typical of this part of the Australian coast. As is also typical for most beaches along the south east Australian coast, the study sites are microtidal, swell dominated environments with the annual predominant swell direction from approximately south east. Sediment size is also provided in Table 2 in the last column, along with the beach type. Sediment size analysis was performed using laser diffraction analysis. This is known to produce larger measurements than traditional sieve analysis (Rodríguez and Uriarte, 2009). All sediment sizes were in the range 246μm < d_{50} < 511μm (Gravois and Baldock, 2013). Gravois and Baldock (2013) performed a comparative sieve analysis of six representative sediment samples from the NSW data and found the grain sizes were found to be of the order 11-18% larger when measured by laser diffraction.

3.2. Data collection and pre-processing

All survey data were reduced to Australian Height Datum (AHD) using permanent survey marks near each beach. The swash zone slope, β, was taken as the average slope during the duration of the dataset obtained at each beach, calculated between the lowest and highest shoreline point in the swash zone for each 15-minute record.

It is unclear which definition of wave height and wave period is most appropriate for use in Hunt's model, which was developed using only regular waves. Therefore, the model was assessed using the three most common descriptors (T_p, T_z, and T_s, being the peak, zero-crossing and significant periods respectively). T_p was consistently found to provide the most accurate predictions, so was deemed the most appropriate and the results reported here use this definition. For the wave height, H_s, was used as this roughly corresponds to visual estimates of wave heights (Masselink *et al*., 2011), and this is the parameter adopted by most other field-based models.

Optical remote sensing methods were used to measure the runup following the methods of Aagaard and Holm (1989) and Power *et al*. (2011). A high definition (1920 x 1080 pixel) video camera was positioned on an elevated vantage point (typically a headland if available, or fore dune) to capture the inner surf and swash zone. In order to rectify the video frames, a square rectification box of markers of known location was placed in the camera field of view. These markers were typically located at the upper and lower limit of the swash zone and were surveyed along with a beach transect through the middle of the rectification box to provide the profile data. Video data were analysed at 5Hz and each frame was rectified using standard Matlab geo-rectification techniques to produce a plan view image. In order to analyse the swash zone at high resolution, timestacks were constructed using a line of pixels that corresponded to the surveyed transect for each rectified frame (Figure 3). Further details on producing a timestack in this way are provided by Power *et al*. (2011). For studies focusing on runup occurring in the field, the time interval for which the runup statistics are calculated should be of order 15 minutes, to assume stationarity with respect to the tide (Hughes and Moseley, 2007), which was the interval adopted here.

3.3. Data processing and analysis

A total of 297 15 minute timestacks and over 22,000 individual swash events (Table 2) were analysed. Analysis of the timestacks was performed using Matlab, where the maximum excursions for individual swash were manually selected to minimise errors introduced by an automated shoreline detection algorithm. Following Power *et al*. 2011, this gives an absolute accuracy of order 0.38m cross-shore, or 0.04m vertically on a beach with a 1:10 gradient. Only swash events where a maximum occurred were selected; swash events overtaken by following bores before reaching a runup maximum were excluded (Figure 3). Cross-shore coordinates of

all maxima were extracted from the timestacks and the values converted to vertical runup elevations using the beach survey data.

The runup maxima from each 15 minute timestack were analysed to provide values of R_{max} (the single highest excursion in each 15-minute segment) and $R_{2\%}$. $R_{2\%}$ was calculated following the method of Nielsen and Hanslow (1991), assuming a Rayleigh distribution and log-ranking the runup maxima to obtain the run-up length scale, L_R, from the line of best fit for each timestack. Due to the assumption of a Rayleigh distribution, data runs that did not conform to the Rayleigh distribution were excluded. Exclusions were made based on the coefficient of determination for the line of best fit (rejection if: $R^2 < 0.9$) (e.g., Figure 4). Hughes *et al*. (2010) discussed the applicability of both the Rayleigh distribution and Normal distribution for describing swash maxima. Hughes *et al*. (2010) found runup distributions to be consistently represented more closely by a Normal distribution than a Rayleigh distribution. Stockdon *et al*. (2006) also compared their runup distributions to a Normal distribution, with reasonable agreement. An analysis using the two different distributions for obtaining the statistical $R_{2\%}$ value has been conducted here, and for this data set there does not appear to be a significant difference between the two (Figure 4, Table 4). When a dataset was rejected for lack of coherence to a Rayleigh distribution, it would be similarly rejected for lack of coherence to a normal distribution. For example, in the case shown in Figure 4, the flat section in the distribution was due to 21 out of 48 measured maxima occurring on a beach berm. However, in either case, from the 297 total data runs assessed here, only 4 were rejected, less than 2%.

As noted, the observed R_{max} depends on record length, or more precisely the number of waves in the record, N. Therefore, the record length used for the development of the R_{max} empirical models should be considered during application. Douglass (1992) used Holman's (1986) dataset, with record lengths of 17-minutes, and therefore N is likely to be similar to those for the 15-minute record length adopted here. Hunt's (1958) analysis was based on regular waves, and therefore independent of record length. The analysis performed by Mase (1988) was completed using records consisting of $650 < N < 900$ individual swash events (Mase and Iwagaki, 1983). For the present dataset, the number of individual swash events was in the range $45 < N < 128$ (Table 2). Given the order of magnitude reduction in record length compared with Mase's dataset, a ratio is proposed that may provide a nominal adjustment. Assuming a Rayleigh distribution, R_{max} scales with the square root of the natural logarithm, (ln

$N)^{0.5}$. Given this relationship, the ratio of the expected R_{max} for the two different record lengths of Mase (1988) and the present data may be approximated by:

$$\frac{\sqrt{\ln N_{Mase}}}{\sqrt{\ln N_{Data}}} = \frac{\sqrt{\ln 650}}{\sqrt{\ln 45}} \text{ to } \frac{\sqrt{\ln 900}}{\sqrt{\ln 128}} = 1.3 \text{ to } 1.2$$

In other words, the R_{max} measured for the present dataset, given the shorter record lengths, may yield average values that are of the order 1.3 times lower than a record length consisting of $650 < N < 900$ individual runup events. This correction was applied to the Mase (1988) model for R_{max} predictions.

For each model, the root-mean-square error (RMSE) between measured and predicted runup ($R_{2\%}$ and R_{max}) at each beach was calculated, as well as the mean value of the RMSE obtained from all the beaches. The distribution, mean, skewness and kurtosis of the model errors were also evaluated for each model for the entire dataset (i.e., combining the data from all beaches) by analysing the data to provide distributions that were split into ten equally spaced error bins that encompassed the range of each model. Skewness (Skew) provides a measure of the symmetry of the error distribution, where negative or positive values indicate skew or decreased weighting towards values lower or higher than the mean, respectively, and a zero value corresponds to a perfectly symmetric distribution. Kurtosis (Kurt) provides an indication of the behaviour of the distribution's peak and tails. Distributions with Kurt<3 are more broad, tending towards a uniform distribution (i.e. MM1, Figure 6) whereas Kurt>3 are peakier distributions with an increased chance of outcomes occurring over a narrower band (i.e. Ma, Figure 6). A normal distribution has Skew=0 and Kurt=3 (i.e. St, Figure 6), so for model accuracy, the most desirable combination of these three indicators would be a zero mean and skew and a high kurtosis.

4. Results

4.1.Combined data set from all beaches

Firstly, model performance is considered for the total data set, i.e. combining the data from all beaches. The analysis is performed initially irrespective of tidal stage (section 4.1.1) and subsequently for low, mid and high tide conditions to assess any tidal dependency (section 4.1.2). Tidal stage for each experiment was determined by taking the maximum tidal range for

each day and splitting it into thirds (i.e., low tide conditions being those data runs that occurred within the lowest third of the tidal range). Model abbreviations are provided in Table 1.

4.2 Model performance irrespective of tidal stage

Figure 5 provides box-whisker plots of the error in predicted run-up (R_{pred}) compared to the observed (R_{obs}) run-up ($error = R_{pred} - R_{obs}$) for $R_{2\%}$ and R_{max} for the total data set. The median error (indicated by the red line in each box) varies between models, with models both under predicting (St, VS, Ru, Vo, MM1) and over predicting (Ma, Wa and HM) the observations. The models NH, Ho and MM2 have median errors close to zero. The locations of the upper and lower quartiles (top and bottom of the boxes) and the spread of the rest of the data (whiskers) indicate the full distribution of the errors.

The error distributions for each model for the entire data set are provided as histograms in Figures 6 and 7, along with the mean error (Mean), skewness (Skew) and kurtosis (Kurt). The smallest mean errors are given by Nielsen and Hanslow (1991) and MM2. Holman (1986) and MM2 have a very similar mean error and lower values of skewness than Nielsen and Hanslow (1991). Nielsen and Hanslow (1991), Stockdon *et al* (2006) and Vousdoukas *et al*. (2012) have a greater number of errors resulting from the under prediction of the measured data. The models of Mase (1989) and Hedges and Mase (2004) produced the least consistent results with the largest mean errors and positive skews. The combination of the high positive mean and high kurtosis of Mase indicates a strong likelihood to over predict (Mean=0.63m & 0.83m, Skew=0.60 & 0.56, and Kurt=3.33 & 2.8, respectively). Vousdoukas *et al*. (2012) was the only $R_{2\%}$ model to yield a negative mean and negative skew, suggesting that it will have a tendency to underestimate the wave runup. The other $R_{2\%}$ models with negative means all had positive skews (NH, St, VS, Ho, Ru, MM1 and MM2). The two '*model of models*', MM1 and MM2 (14 and 15) performed comparably to the existing better performing models. In fact, MM2 (15), which includes a setup term, provided the most consistent runup predictions of the two. It had the joint smallest mean error (-0.05m), slight positive skewness (0.17) and small kurtosis (2.77), indicating a reduced chance of outliers, which suggests that this model may be one of the most reliable. The higher skewness (0.26) and relatively high kurtosis (3.30) given by Nielsen and Hanslow (1991) suggests a positive bias and an increased chance of outliers compared to MM2.

The models for R_{max} performed more variably, but it is noted that the record length needed to achieve stationary conditions with respect to the tide limits the number of events in each record. Thus R_{max} is more variable between data records than $R_{2\%}$. Hunt (1958) yielded a mean error of -0.25m, with a slight positive skew (0.06) and a low kurtosis (2.65). The models of Mase (1989) and Douglass (1992) had comparable mean errors of 0.42m and 0.46m, respectively. Mase's model had the strongest positive skewness (0.86) and a kurtosis of 3.38. Douglass (1992) had a negative skew (-0.15) and the highest kurtosis (4.49).

Root-mean-square errors at individual beaches

The root-mean-square error (RMSE) was calculated for each model at each individual beach. These are illustrated in a stacked histogram that shows both the RMSE error at each beach (colour bar height = RMSE) (and the sum of these errors) and the overall mean-RMSE for each model (Figure 8). The most accurate $R_{2\%}$ models (determined by lowest mean RMSE (right axis) and lowest summed RMSE (left axis)) were jointly Holman (1986), Vousdoukas *et al*. (2012) and the new model (MM2), followed by Stockdon *et al*. (2006). Including the four aforementioned models, three other models had a mean-RMSE below 0.40m (Nielsen and Hanslow, 1991, Ruggiero *et al*., 2001 and MM1). The least accurate model was Hedges and Mase (2004), with a mean RMSE of 0.83m. The differences in the RMSE for different models on the same beach, and the differences in RMSE for the same model on different beaches, are quite striking. This suggests that un-calibrated model predictions on an arbitrarily selected beach can be prone to significant error.

The difference between the $R_{2\%}$ models developed using field or laboratory data (white and black mean-RMSE dots respectively on Figure 8) show that the laboratory models exhibit higher mean-RMSE values, suggesting that the laboratory derived models are less predictively accurate than those developed in the field when used for field runup predictions. With the exception of Van der Meer and Stam (1992), Figures 6 and 7 indicate that the models developed in laboratories tend to over predict the runup, which is perhaps not unexpected given the 2D and 3D conditions in the laboratory and field, respectively. Blenkinsopp *et al*. (2015) found field-derived models to under predict large scale laboratory measurements of runup (exceptions were Nielsen and Hanslow (1991) and Stockdon *et al*. (2006)) and most laboratory derived models to over predict (with the exception of Van der Meer and Stam (1992)). This is consistent with the present observations.

The R_{max} model with the lowest RMSE was Hunt (1958), followed by Douglass (1992) and then Mase (1989). The difference between the R_{max} models developed using field data and laboratory data (white and black mean-RMSE dots respectively, Figure 9) was variable. Hunt (1958) generally performed better than Douglass (1992), but tended to underestimate R_{max} (Figures 6 & 7). Mase (1989) tended to have the greatest errors, with a greater tendency to overestimate R_{max} and with an error distribution that sits higher than the other two models (Figures 6 & 7).

4.3 Influence of tide on model performance

Model performance at low, mid and high tides was observed to be varied and is examined in greater detail below. Table 5 provides the mean, skewness and kurtosis for the error distributions of each model under each tide condition and Figure 9 provides the corresponding box-whisker plots for the error distributions. The tidal stage appears to have more influence over the skewness and kurtosis (Table 5) than on the mean error for most of the $R_{2\%}$ models. The mean errors appear to be relatively insensitive to tidal stage, although the range of the quartile errors is clearly reduced on the low tide. With the exception of Hedges and Mase (2004), Wassing (1957) and MM1, all skewness values increase for the $R_{2\%}$ models for the higher tide levels, which is apparent in the box and whisker plots, where the median lines are closer to the lower quartile for the lower tide levels (Figure 9). The kurtosis at different tidal stage varies more markedly. Nielsen and Hanslow (1991), Stockdon *et al*. (2006), and Ruggiero (2001) were the most sensitive to tidal stage, with a greater chance of outlier-errors on high and low tides for Nielsen and Hanslow (1991) and on low tides for Stockdon *et al*. (2006) and Ruggiero (2001). Out of the eleven $R_{2\%}$ models, six exhibited an increased kurtosis going from high to low tides (NH, St, Ru, HM, Vo and MM1) and eight exhibited an increased kurtosis from mid to low tides (NH, St, Ho, Ru, HM, Vo, MM1 and MM2), suggesting an increased chance of larger outlier errors for these models on lower tides.

Root-mean-square errors at individual beaches

Stacked histograms of the RMSE error at each beach and the overall mean-RMSE for each model for each tidal stage are shown in Figures 10 and 11. For high tides, the most accurate $R_{2\%}$ model considering the mean-RMSE was MM2. For mid tides, the most accurate model was Vousdoukas *et al*. (2012), which was followed jointly by Stockdon *et al*. (2006), Holman (1986) and MM2. For low tides, Stockdon *et al*. (2006) was the most accurate,

followed jointly by Ruggiero (2001) and Vousdoukas *et al*. (2012). The performance of most models was variable between tidal stages with respect to their relative accuracies (determined by lowest mean-RMSE) relative to one another, however, the group of models previously noted were consistently found to perform well. In contrast, the laboratory developed models of Wassing (1957), Van der Meer and Stam (1992), Mase (1989), and Hedges and Mase (2002) regularly gave the poorest predictions, with the Mase (1989) and Hedges and Mase (2002) models consistently being the two least accurate models in every scenario. The greatest differences that occurred were for Nielsen and Hanslow (1991), joint-4th and 5th most accurate for low and mid tides respectively but only the 9th most accurate during high tides. The difference in the absolute error between the laboratory and field derived models (Figure 10) was similar to that for the total dataset (Figure 8), with the laboratory derived models typically producing greater errors than the field derived models. The only exception was on high tides, where Nielsen and Hanslow (1991) was less accurate than Van der Meer and Stam (1992) and Wassing (1957).

The difference between the performance of the R_{max} models at different tidal stages was very small (Figure 11). Hunt (1958) was always the most accurate followed by Douglass (1991) for mid and low tides and Mase (1989) for high tides.

Finally, the maximum observed RMSE for each model over the entire dataset is plotted in Figure 12. This is the error indicated by the magnitude of the largest coloured block for each model in Figures 8, 10 and 11, i.e., it is the RMSE at the beach where that model performs least well. This provides an additional assessment of the general performance of the models. The new model, MM2 had the smallest maximum RMSE considering all tidal stages, followed by Stockdon *et al*. (2006), Ruggiero *et al*. (2001) and Holman (1986). Taking these four models as having similar performance, the maximum RMSE is of order 0.6-0.7m which provides a measure of the typical accuracy of un-calibrated runup models on typical SE Australian beaches during these wave conditions. There is more variability if considering the three different tidal levels and this is discussed further in section 5. For the maximum runup, Hunt (1958) consistently yielded the smallest maximum RMSE for all cases, irrespective of tidal level. The results of the Mase (1989) and Douglass (1991) models reflect those of the other tidal RMSE results described above.

5. Discussion

5.1. Wave and beach conditions

The wave conditions assessed here encompassed a range of conditions with averages ($H_s \approx$ 1.48m and $T_p \approx$ 8.91s) that were slightly below mean conditions typical of the region of $H_s \approx$ 1.55m and $T_p \approx$ 9.5s (Lord and Kulmar, 2001). Extreme wave conditions were not present so the testing of these models has been limited to near-mean conditions. Nevertheless, an analysis of model performance in these conditions is of benefit for forecasting runup on beaches under non-extreme conditions as well as allowing for a further assessment of various empirical runup models. Recent runup measurements by Fiedler *et al*. (2015) indicate that the Hunt (1958) scaling holds over a large range of wave conditions, up to extreme offshore wave heights. Further, several of the field based models were developed from datasets with a wide range of wave conditions that included some storm data (Stockdon *et al*. (2006), Holman (1986), Ruggiero *et al*. (2001) and Vousdoukas *et al*. (2012)). Therefore, higher energy conditions are also built into the two "*model-of models*".

In absolute terms, maximum RMSE are order 0.6-0.7m on any beach for the best performing models, but this is for average wave conditions. To provide a measure of relative accuracy, the RMSE values for each beach (Figure 8) have been normalised by the corresponding observed mean runup value ($R_{2\%}$ or R_{max}) for that beach dataset, which indicates the accuracy of each model as a percentage (Figure 13). When analysed in this regard, Vousdoukas et al. (2012) performed best with RMSE/$R_{2\%}$ = 0.23, followed by Ho, MM2 and St, each with RMSE/$R_{2\%}$ = 0.26. The mean values of $R_{2\%}$ or R_{max} range from 0.7m - 2.3m for the different datasets. The four best performing $R_{2\%}$ models have relative errors of order 25%. For R_{max}, Hunt (1958) has a relative error of order 30%, while the models of Mase (1989) and Douglass (1992) performed comparably, with relative errors of 46% and 50%, respectively.

The most recent runup model assessed using the present data was Vousdoukas *et al*. (2012), which was found to be the joint most accurate overall (but biased toward low predictions) with Holman (1986) and the new "*model of models*", MM2, with respect to the stacked and mean root-mean-square error (Figure 8), as well as for mid tides. In addition to being joint most accurate for all tides, Holman (1986) was also the second most accurate for mid and high tides. The new '*model of models*', MM2 was also the most accurate model on high tides. Stockdon *et al*. (2006) was the most accurate for the low tide dataset (Figure 11), and the joint-third most

accurate, with MM2, for mid tides. Recently, Stockdon *et al*. (2014) assessed the accuracy of Stockdon *et al*. (2006) against data and numerical model simulations (Xbeach), with the former being most accurate. Root mean square errors for the parameterized runup predictions were in the range 0.26-0.36m, which is in close agreement with the present data (RMSE of 0.32m). Soldini *et al*. (2013) also performed numerical analysis of runup over gently sloping beaches ($\beta < 0.04$) and found the majority of their results to match reasonably with Stockdon *et al*.'s model for $\xi_p < 0.3$. When considering the maximum observed error (Figure 12), MM2 provides some of the lowest maximum errors, suggesting consistently good performance, which is supported by the error histograms (Figure 6) and box-whisker plots (Figures 5 and 9). The average of the maximum error of each model at each beach is plotted in Figure 14. This further shows that MM2 has the smallest average-maximum error, followed by Holman, Nielsen and Hanslow, and then Vousdoukas. MM2 provided a positive mean of the maximum errors overall, whereas the other three models provided negative means, suggesting a tendency toward under-predictions, which is again reflected in the box and whisker plots in Figure 5.

Vousdoukas *et al*. (2012) found Stockdon *et al*.'s model to under predict runup compared to their parameterisation, however, both models underestimate the present data slightly. The mean error for Vousdoukas *et al*. (2012) and Stockdon *et al*. (2006) was the same at -0.19m, but Vousdoukas *et al*. (2012) shows a negative skewness (-0.44), whereas Stockdon *et al*. (2006) has a slight positive skew (0.10), (Figure 5), which may also be reflected in Figure 14, with Stockdon *et al*.'s model having a greater mean of the maximum runup errors. Considering the error distributions for the entire dataset (Figures 5 and 6), it is difficult to pick any significant differences between the two. The errors from Vousdoukas *et al*. (2012) had a median (red line) value slightly closer to zero and the two models have oppositely uneven whisker lengths (longer downward whiskers for Vousdoukas *et al*. (2012), whereas Stockdon *et al*. (2006) exhibits longer upward whiskers). Stockdon *et al*. (2006) however yields a lower maximum RMSE (Figure 12).

The underestimation by Stockdon *et al* (2006) may be, in part, due to the dataset from which it was developed, where the majority (91%) of the data came from a single beach (Duck, North Carolina, USA). The offshore profile (lower shoreface) of Duck has a shallower gradient compared to those in the region of the present study. Shallower shoreface gradients correspond to a longer shoaling region, resulting in increased wave energy dissipation with increasing wavelength. The mean offshore period the Stockdon *et al*.'s (2006) dataset varied between 8s

< T_p < 14.9s, with the majority over 10s. In comparison, the present paper analysed runup produced by peak periods mostly lower than 10s (exceptions are NS2 and NS3, although Stockdon *et al*. (2006) did not perform significantly better for those two beaches). The correspondingly greater wavelengths associated with Stockdon *et al*.'s data, along with the shallower bathymetry may have resulted in increased wave energy attenuation from greater shoaling distances. Although the tidal range along the NSW coast is small, Stockdon *et al*.'s (2006) model was observed to perform the best of all models on low tide, conversely, their model was also at its least accurate (mean RMSE = 0.39) for the high tide dataset. This is consistent with the presence of a shallower foreshore gradient at low tide for the present dataset. The shoreface along the south coast of Algarve where the measurements of Vousdoukas *et al*. (2012) were conducted drops off to 50m depth over approximately 8km, which is similar to the steep drop off observed along the NSW coast (Wright, 1976). Stockdon *et al*. (2006) noted that due to their model coefficients being tuned to provide a best fit to all the data they analysed, systematic errors were introduced considering parameter-specific datasets due to the broad nature of the model. Stockdon *et al*. (2014) also noted that Stockdon *et al*.'s (2006) model omits many influential surf zone parameters, such as long and cross shore morphodynamic variability. These shortcomings are present in all of the models assessed here and are necessary in order to produce a simple and easy to use predictive model. The spread of the measured $R_{2\%}$ data when plotted against the Iribarren number in Figure 1 may be a further indication of missing parameters from the formulation. It is important to bear in mind these limitations when using such tools.

5.2 "Model of models" performance

The two newly proposed *model of models*, MM1 and MM2 (Equations 14 and 15), which represent the best fit to predictions from all the other field-derived and Van der Meer and Stam (1992) models, both performed well, with MM2 being the most accurate of the two models overall. MM2 yielded the joint smallest mean error of -0.05m, comparable with Nielsen and Hanslow (1991), and somewhat smaller than Stockdon *et al*. (2006) and Vousdoukas *et al*. (2012). MM2, Holman (1986) and Vousdoukas (2012) all had the same mean root-mean-square error, which was marginally smaller than that of Stockdon *et al*. (2006). MM2 was found to have the smallest maximum error across all beaches. Two other *Model of models* were also formulated by including all the laboratory-derived $R_{2\%}$ models, but the accuracy was lower than when just the field and large scale laboratory data derived models were included,

consistent with the distinct differences in mean error for the two different groups of models (Figures 6, 8 and 10).

5.2. *Runup length scale coefficient (Nielsen and Hanslow, 1991)*

Originally, Nielsen and Hanslow (1991) reported a runup length scale coefficient of L_r = 0.05 for beaches where tanβ<0.1. Later, Nielsen (2009) revised the coefficient to 0.06 (Equation 8) to better suit additional data, corresponding to the value used in the calculations presented above. The data from the present research is now also added to the existing plot of Nielsen (2009, his Figure 2.4.6, p.129), Figure 15. The average of all the data for beaches where tanβ<0.1 provided a coefficient of 0.064, shown as the fine dotted line (Figure 15), slightly greater than Nielsen's value. However, considering the scatter, the rounded value of 0.06 remains a reasonable approximation and this data likewise supports the value of 0.06 rather than 0.05.

6. Conclusions

Runup measurements ($R_{2\%}$ and R_{max}) from 11 different beaches along the south-east Australian coast have been compared to predictions from a range of commonly used runup models. The variations in the predictions from these models for the same incident wave conditions were also assessed, and were shown to be considerable (order of a factor 1.5). Two new *model of models* have been derived by plotting predictions from all the field-derived runup models versus the runup scaling ($tan\beta\sqrt{H_s L_p}$) introduced by Hunt (1958) and fitting new coefficients. Following Holman (1986), one of these models includes a setup parameterisation.

The most accurate models for predicting $R_{2\%}$, irrespective of tidal stage, were jointly Holman (1986), Vousdoukas *et al*. (2012) and the new *model of models* that includes setup, MM2. MM2 was also found to be the most accurate model for the high tide dataset whereas the most accurate models for the mid and low tide datasets were Vousdoukas *et al*. and Stockdon *et al*., respectively. Vousdoukas' model was also found to be the most accurate model independent of $R_{2\%}$ (Figure 13). However, when considering the maximum errors across the entire dataset (Figure 12) as well as the average of the maximum errors from all beaches (Figure 14), MM2 returned the lowest value which was also positive, whereas the models of Holman and Vousdoukas returned small, but negative values, indicating a greater tendency to under

predict compared with MM2. The most accurate model for R_{max} was found to be that of Hunt (1958) which exhibited the lowest skewness and kurtosis, but still tended to under predict the observed maximum runup. Mase's (1989) model provides more conservative estimates. It is important to note that the observed R_{max} depends on the record length and a correction may be needed for data and models derived from a different record length.

The most appropriate models for predicting $R_{2\%}$ values in the field are those developed from field data. Most models derived from laboratory data tend to significantly overestimate the field data obtained in this study and all yield root-mean-square errors (RMSE) greater than all the field-derived models. The variability of the natural runup values is quite striking and is reflected in the variability in the accuracy of the different runup model predictions, and differences in the RMSE for different models on the same beach, and the differences in RMSE for the same model on different beaches. No single model provided the best runup estimates for all beaches within the present data set. This suggests that un-calibrated model predictions on an arbitrarily selected beach can be prone to significant error. The top three models irrespective of tidal stage (Holman, 1986; Vousdoukas *et al*., 2012, MM2) gave mean RMSE of order 0.30m for root-mean-square $R_{2\%}$ values of 1.56m, or an overall error of about 25% of $R_{2\%}$.

7. Acknowledgments

The authors gratefully acknowledge Manly Hydraulics Laboratory (NSW) and DSITI (QLD) for providing wave and tide data for these field experiments and NSW and QLD State Governments for providing details on permanent survey marks. This work was supported by ARC grants LP100100375, DP13101122 and DP14101302.

8. References

Aagaard, T. & Holm, J. 1989. Digitization of Wave Run-up Using Video Records. Journal of Coastal Research, 5**,** 547-551.

Antuono, M. and Brocchini, M., 2008. Maximum run-up, breaking conditions and dynamical forces in the swash zone: a boundary value approach. Coastal Engineering, 55(9), pp.732-740.

Baldock, T.E., Peiris, D. & Hogg, A.J., 2012. Overtopping of solitary waves and solitary bores on a plane beach. In Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 468 (2147), 3494-3516

Battjes, J. A. 1974. Surf Similarity. Proceedings of 14th Coastal Engineering Conference. Copenhagen, Denmark, American Society of Civil Engineers, New York.

Blenkinsopp, C., Matias, A., Howe, D., Castelle, B., Marieu, V. & Turner, I. 2015. Wave runup and overwash on a prototype-scale sand barrier. Coastal Engineering.

Bruun, P. 1954. Coast erosion and the development of beach profiles, Beach Erosion Board Corps of Engineers.

Callaghan, D.P., Callaghan, J., Nielsen, P. and Baldock, T.E., 2006. Generation of extreme wave conditions from an accelerating tropical cyclone. In Proceedings of the Coastal Engineering Conference (Vol. 1, pp. 752-760). World Scientific.

Douglass, S. 1992. Estimating Extreme Values of Run-Up on Beaches. Journal of Waterway, Port, Coastal and Ocean Engineering, 118**,** 220–224.

Fiedler, J. W., Brodie, K. L., McNinch, J. E., and Guza, R. T. 2015, Observations of runup and energy flux on a low-slope beach with high-energy, long-period ocean swell, Geophys. Res. Lett., 42, 9933–9941.

Gravois and Baldock, 2013. Sediment Sizing of NSW Beach Sand Particle size distribution analysis of dune, berm and swash sand samples from NSW beaches., Prepared for OEH, NSW State Gov.

Hedges, T. & Mase, H. 2004. Modified Hunt's equation incorporating wave setup. Journal of Waterway, Port, Coastal, and Ocean Engineering, 130**,** 109-113.

Holman, R. A. 1986. Extreme value statistics for wave run-up on a natural beach. Coastal Engineering, 9**,** 527-544.

Holman, R.A. and Guza, R.T., 1984. Measuring run-up on a natural beach. Coastal Engineering, 8(2), pp.129-140.

Hughes, S. A. 1993. Physical models and laboratory techniques in coastal engineering, World Scientific.

Hughes, M. G. & Moseley, A. S., 2007. Hydrokinematic regions within the swash zone. Continental Shelf Research, 27 (15), 2000-2013

Hughes, M. G., Moseley, A. S. & Baldock, T. E. 2010. Probability distributions for wave runup on beaches. Coastal Engineering, 57**,** 575-584.

Hunt, I. A. 1958. Design of seawalls and breakwaters, US Lake Survey.

Iribarren, C.R., Nogales, C., 1949. Protection des ports. Sect. 2, comm. 4, 17th Int. Nav. Cong., Lisbon, pp.31–80.

Komar, P. D. 1976. Beach Processes and Sedimentation, United States of America, Prentice-Hall.

Kriebel, D. L. & Dean, R. G. 1993. Convolution method for time-dependent beach-profile response. Journal of Waterway, Port, Coastal and Ocean Engineering, 119, 204-226.

Lord, D. & Kulmar, M. The 1974 Storms Revisited: 25 Years Experience in Ocean Wave Measurement Along the South East Australian Coast. Coastal engineering conference 2001. ASCE American Society of Civil Engineers, 559-572.

Mase, H. 1989. Random Wave Runup Height on Gentle Slope. Journal of Waterway, Port, Coastal, and Ocean Engineering, 115**,** 649-661.

Mase, H. and Iwagaki, Y., 1984. Run-up of random waves on gentle slopes. Coastal Engineering Proceedings, 1(19).

Masselink, G., Hughes, M. & Knight, J. 2011. Introduction to coastal processes and geomorphology, Hodder Education.

Nielsen, P. 2009. Coastal and estuarine processes, Singapore, World Scientific.

Nielsen, P. & Hanslow, D. J. 1991. Wave Runup Distributions on Natural Beaches. Journal of Coastal Research, 7**,** 1139-1152.

Peregrine, D.H. & Williams, S.M., 2001. Swash overtopping a truncated plane beach. Journal of Fluid Mechanics, 440, 391-399.

Power, H. E., Holman, R. A. & Baldock, T. E. 2011. Swash zone boundary conditions derived from optical remote sensing of swash zone flow patterns. Journal of Geophysical Research, 116.

Power *et al*., 2013a: Power, H. E., Atkinson, A. L., Hammond, T. & Baldock, T. E. 2013. Accuracy of wave runup formula on contrasting southeast Australian beaches. Coasts and Ports 2013. Sydney, Australia.

Power *et al*., 2013b: Power, H.E., Baldock, T.E., Callaghan, D.P. and Nielsen, P., 2013. Surf zone states and energy dissipation regimes — a similarity model. Coastal Engineering Journal, 55(01), 1350003.

Ruggiero, P., Komar, P. D., McDougal, W. G., Marra, J. J. & Beach, R. A. 2001. Wave runup, extreme water levels and the erosion of properties backing beaches. Journal of Coastal Research**,** 407-419.

Rodríguez, J.G., and Uriarte, A. (2009). Laser Diffraction and Dry-Sieving Grain Size Analyses Undertaken on Fine- and Medium-Grained Sandy Marine Sediments: A Note., Journal of Coastal Research, Vol. 25, No. 1, pp. 257-264

Sato, S., Yeh, H., Isobe, M., Mizuhashi, K., Aizawa, H. and Ashino, H., 2013. Coastal and nearshore behaviors of the 2011 Tohoku Tsunami along the central Fukushima Coast. In Proc. Coastal Dynamics.

Short, A.D., 1993. Beaches of the New South Wales coast, Sydney University Press, 358 pp.

Short, A.D., 2000. Beaches of the Queensland Coast: Cooktown to Coolangatta, Sydney University Press, 369 pp.

Soldini, L., Antuono, M., Brocchini, M. 2013. Numerical modelling of the influence of the beach profile on wave run-up. Journal of Waterway, Port, Coastal and Ocean Engineering, 139**,** 61-71.

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

Stockdon, H.F., Thompson, D.M., Plant, N.G. and Long, J.W., 2014. Evaluation of wave runup predictions from numerical and parametric models. Coastal Engineering, 92, pp.1-11.

Van der Meer, J. W. & Stam, C.-J. M. 1992. Wave runup on smooth and rock slopes of coastal structures. Waterway, Port, Coastal, and Ocean Engineering, 118**,** 534–550.

Vousdoukas, M. I., Wziatek, D. & Almeida, L. P. 2012. Coastal vulnerability assessment based on video wave run-up observations at a mesotidal, steep-sloped beach. Ocean Dynamics, 62**,** 123-137.

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

Wright, L. 1976. Nearshore wave-power dissipation and the coastal energy regime of the Sydney-Jervis Bay region, New South Wales: a comparison. Marine and Freshwater Research, 27**,** 633-640.

Wright, L. D. & Short, A. D. 1984. Morphodynamic variability of surf zones and beaches: A synthesis. Marine Geology, 56**,** 93-118.

TABLES

Table 1: Abbreviations for models used in this study.

Wassing (1957)	Wa
Hunt (1958)	Hu
Mase (1989)	Ma
Hedges and Mase (2004)	HM
Van der Meer and Stam (1992)	VS
Holman (1986)	Ho
Nielsen and Hanslow (1991) & Nielsen (2009)	NH
Douglass (1992)	Do
Ruggiero *et al*. (2001)	Ru
Stockdon *et al*. (2006)	St
Vousdoukas *et al*. (2012)	Vo
Model of Models 1 (Present paper)	MM1
Model of Models 2 (Present paper)	MM2

Table 2: Location, date (range), number of records, number of swash maxima, average number of swash events per record, beach type and sediment size (where available) is provided in square brackets under the beach type.

Location	Abbr.	Date	Records (total, rising tide, falling tide)	Number of swash events	Average number of swash maxima per record	Beach type as per Short (1993 & 2000) and sediment size [μm]
Norries Head Beach, NSW, 28°20'8.44"S, 153°34'32.73"E	NHB	9/11/2010	21, 7, 14	1678	80	LTT + sand waves [270 (Cabarita)]
South Boganger Beach, NSW, 28°21'27.24"S, 153°34'32.68"E	SBB	10/11/2010-11/11/2010	40, 14, 26	3059	76	Inner TBR; outer RBB [270 (Cabarita)]
Stockton, NSW, 32°52'35.30"S, 151°48'32.24"E	ST	13/06/2011	20, 7, 13	1373	69	Inner TBR; outer RBB [266]
Yamba, NSW, 29°26'36.37"S, 153°21'59.00"E	YA	10/06/2011	18, 18, 0	813	45	TBR [246]
North Stradbroke, QLD, 27°26'20.30"S, 153°32'27.40"E	NS	23/2/2011 - 6/3/2011	110, 52, 58	6583	60	Inner TBR; outer RBB/LBT [280]
Woonona, NSW, 34°20'54.38"S, 150°55'16.72"E	WO23	23/03/2013	12, 4, 8	1488	124	North: TBR/RBB [346]
Austinmer, NSW, 34°18'23.07"S, 150°56'6.99"E	AU24	24/03/2013	5, 0, 5	607	121	TBR [445]
Werri, NSW, 34°44'29.06"S, 150°49'57.88"E	WE25	25/03/2013	14, 4, 10	1463	105	South: TBR [511]
Mollymook, NSW, 35°19'43.22"S, 150°28'36.82"E	MO26	26/03/2013	10, 0, 10	1275	128	Centre: TBR [426]
Tathra, NSW, 36°43'36.46"S, 149°58'58.17"E	TA28	28/03/2013	9, 0, 9	684	76	South TBR [290]
Beares, NSW,	BE29	29/03/2013 – 30/03/2013	17, 8, 9	1475	87	TBR [511]

36°26'5.65"S, 150° 4'35.79"E						
Werri, NSW, 34°43'46.79"S, 150°50'17.28"E	WE02	2/04/2013	12, 0, 12	731	61	North: RBB [511]

Table 3: Location, average beach slope (tanβ), offshore significant wave height (H_s), peak offshore wavelength (L_p), significant offshore wavelength (L_s), and Iribarren numbers with respect to the peak (ξ_p) and significant (ξ_s) wave periods.

	tanβ [-]	H_s [m]	L_p [m]	L_s [m]	ξ_p [-]	ξ_s [-]
NHB	0.06	1.5	81	69	0.44	0.41
SBB1	0.05	1.54	100	77	0.40	0.35
SBB2	0.06	1.6	127	100	0.53	0.47
ST	0.09	3.27	125	102	0.56	0.50
YA	0.05	1.91	191	157	0.50	0.45
NS1	0.06	3.08	151	79	0.42	0.30
NS2	0.06	1.16	209	95	0.81	0.54
NS3	0.02	0.74	188	61	0.32	0.18
WO23	0.11	1.55	86	79	0.82	0.79
AU24	0.11	1.01	83	79	1.00	0.97
WE25	0.09	1.13	136	63	0.99	0.67
MO26	0.16	1.13	120	70	1.65	1.26
TA28	0.08	1.12	94	60	0.73	0.59
BE29	0.08	1.04	127	100	0.88	0.78
BE30	0.07	1.11	136	107	0.77	0.69
WE02	0.05	1.25	100	82	0.45	0.40

Table 4: Coefficient of determination (R^2) for selected observed runup observations, showing a similar goodness of fit to both a Rayleigh and Normal distribution.

		R^2		
Beach	Time	Log-Ranked	Normal	Rayleigh
WO23	8:20	0.99	0.87	0.87
AU24	8:54	0.99	0.84	0.84
WE25	9:05	0.99	0.89	0.88
NS3	10:38	0.89	0.71	0.76
SBB2	13:59	0.89	0.7	0.81

Table 5: Mean, skewness and kurtosis values for model errors on high, mid and low tide. Refer to Table 1 for model abbreviations.

	High tide			Mid tide			Low tide		
$R_{2\%}$	Mean	Skew	Kurt	Mean	Skew	Kurt	Mean	Skew	Kurt
NH	-0.22	1.48	4.22	-0.17	0.54	3.03	0.10	-1.10	4.96
Ma	0.67	1.28	3.84	0.47	0.75	3.82	0.72	0.43	2.69
St	-0.25	1.06	3.56	-0.28	0.45	2.44	-0.11	-1.01	4.94
Wa	0.44	0.45	2.79	0.20	0.72	2.56	0.28	0.41	2.38
VS	-0.32	0.90	3.30	-0.41	0.44	3.75	-0.18	0.08	3.12
Ho	-0.08	1.22	4.00	-0.17	0.51	2.89	0.00	-0.40	3.16
Ru	-0.33	1.14	3.45	-0.32	0.48	2.39	-0.14	-0.89	4.85
HM	1.02	-0.31	1.89	0.71	0.67	2.79	0.83	0.77	2.96
Vo	-0.29	0.29	2.51	-0.26	0.00	2.47	-0.09	-1.17	4.93
MM1	-0.22	0.39	2.50	-0.31	0.62	2.70	-0.19	-0.54	3.14
MM2	-0.04	1.05	3.69	-0.14	0.57	2.85	0.01	-0.26	2.93
R_{max}									
Hu	-0.24	0.18	2.71	-0.34	0.60	2.72	-0.19	-0.35	3.16
Ma	0.35	0.90	4.13	0.24	0.96	3.73	0.57	0.67	2.71
Do	0.29	0.37	3.56	0.35	0.72	2.99	0.61	-0.23	2.83

FIGURES

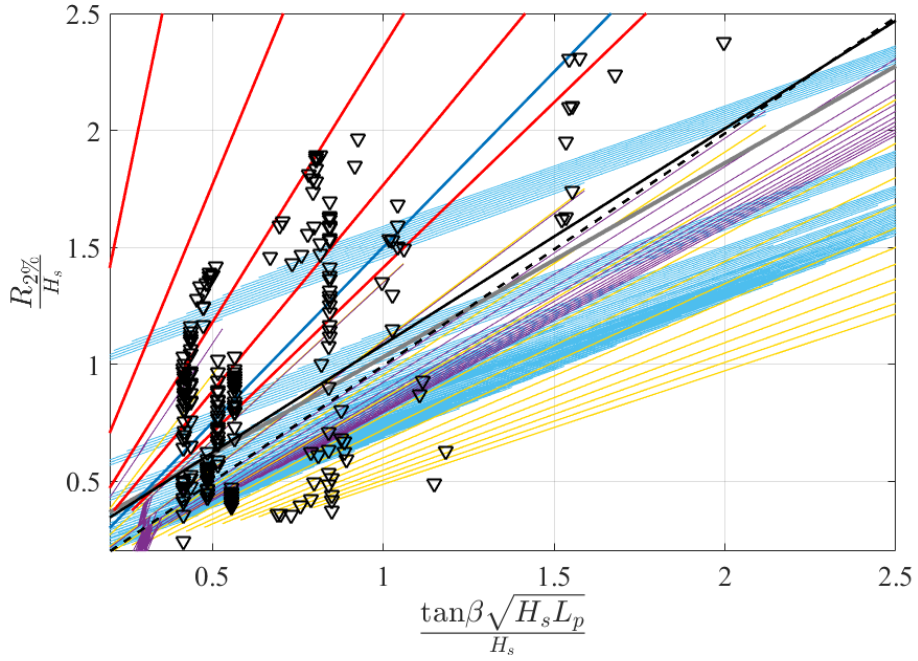


Figure 1: Field and large scale laboratory derived runup model predictions replotted versus Hunt (1958) scaling or Iribarren number calculated using deep water significant wave height and peak period. Parameter range: $H_s/L_p \leq 0.14$, approx. and 0m < H_s < 5m, 0s < T_p < 15s, 0 < β < 0.2. Models shown as: dark blue (VS), grey (Ho), red (NH), yellow (Ru), purple plus signs (St), and light blue (Vo). The dashed and solid black lines of best fit correspond to MM1 and MM2, equations (14) and (15) respectively. The present $R_{2\%}$ data is shown as inverted black triangles.

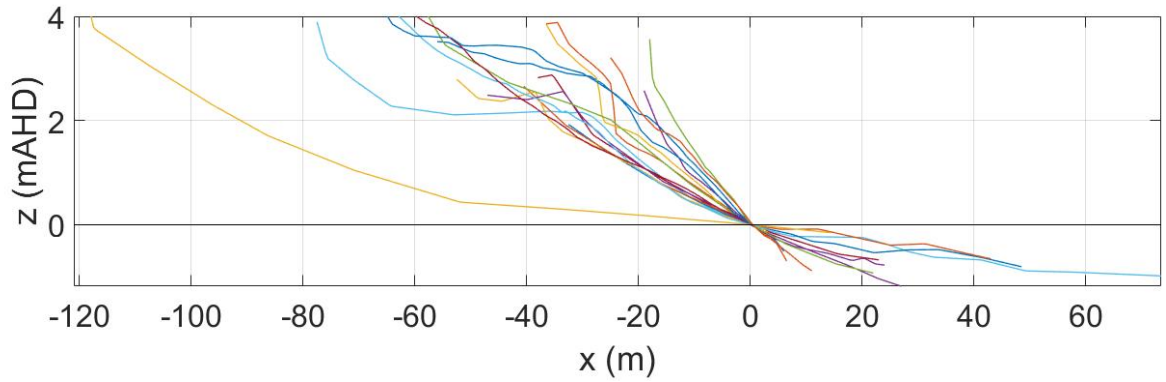


Figure 2: Representative cross-shore beach profiles for each beach on each day of data collection. The shallowest swash zone slope was observed at North Stradbroke (NS3) due to data collection over a low tide terrace, the steepest swash zone slope was at Mollymook (MO26).

Figure 3: A portion of a timestack image from the North Stradbroke dataset, illustrating the maxima (red circles) and a case where one swash (A) was overtaken by a following swash (x ≈ 18m, $t \approx 270$s). The overtaken swash events are excluded from the runup analysis.

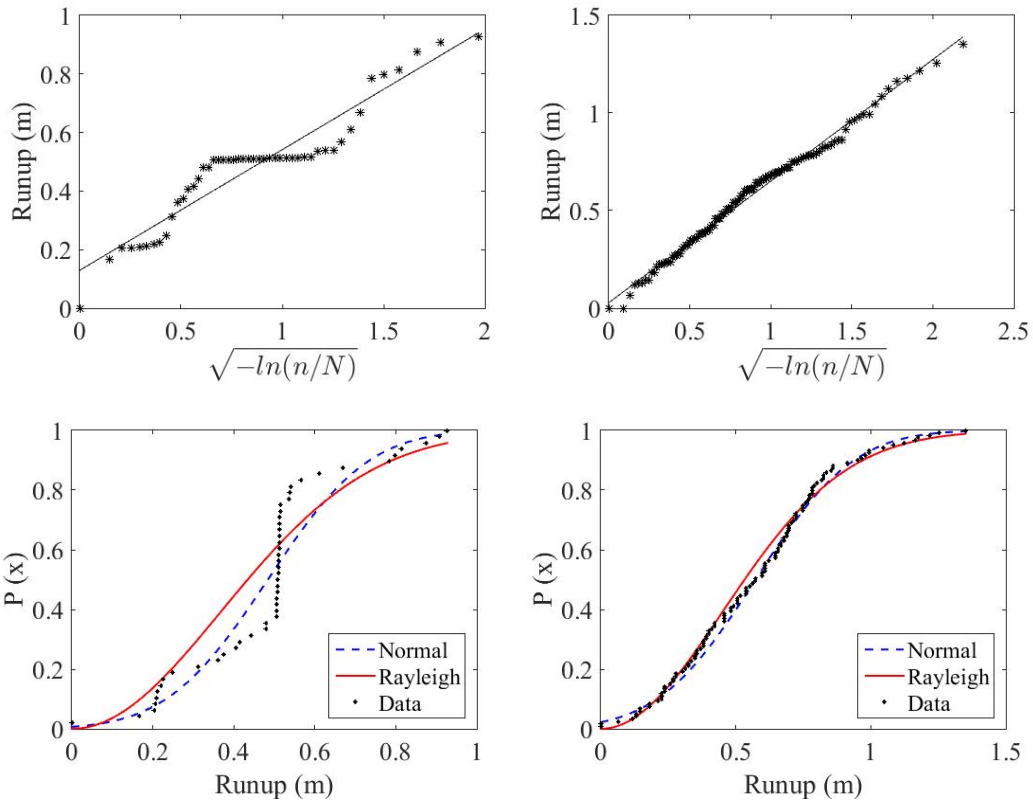


Figure 4: Example log-ranked plots (top) and runup distributions from two 15-minute timestacks from the North Stradbroke (NS3, Table 4) dataset where $R^2 \approx 0.89$ (left, rejected) and from Woonona (WO23, Table 4) where $R^2 \approx 0.99$ (right, accepted). From 297 total data runs, 4 were rejected, less than 2%. Bottom plots provide a comparison of the data with both normal and Rayleigh probability density functions.

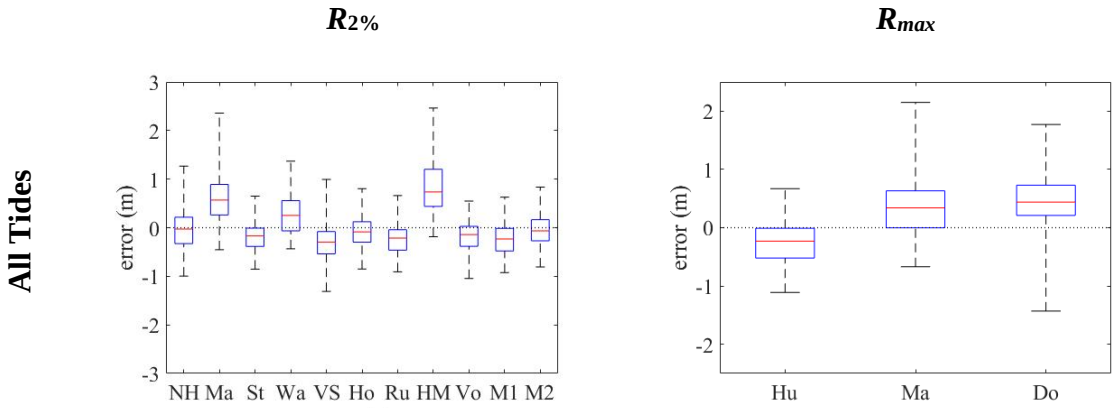
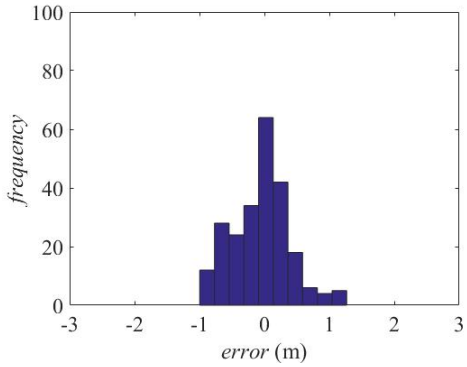


Figure 5: Box and whisker plots of model error in vertical runup for all beaches and data. The red lines indicate the median value and the upper and lower box boundaries indicate the upper and lower quartiles. The whiskers indicate the full range of the errors. Model abbreviations are given in Table 1.

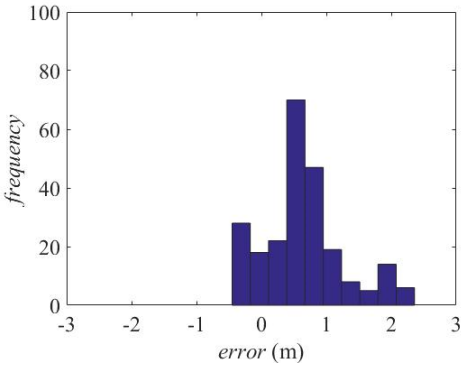
Nielsen and Hanslow

Mean=-0.05m; Skew=0.26; Kurt=3.30



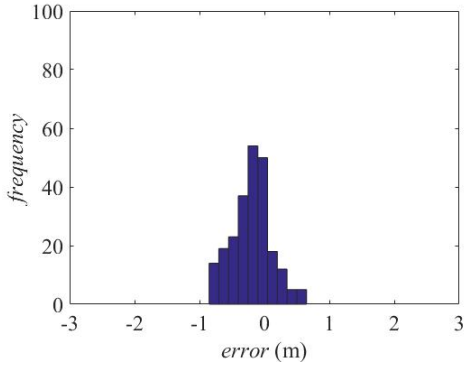
Mase ($R_{2\%}$) *

Mean=0.63m; Skew=0.60; Kurt=3.33



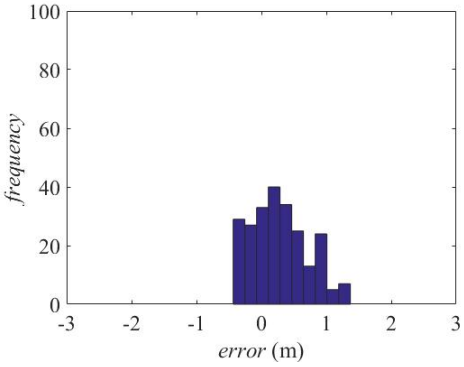
Stockdon *et al*.

Mean= -0.19m; Skew=0.09; Kurt=3.00



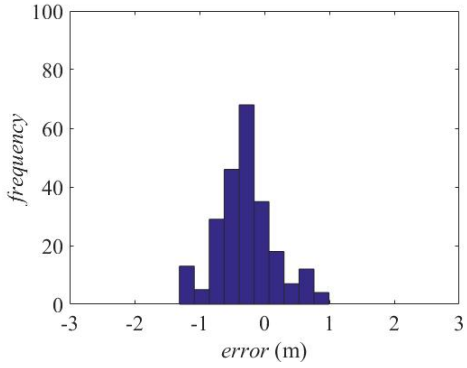
Wassing *

Mean= 0.28m; Skew=0.41; Kurt=2.44



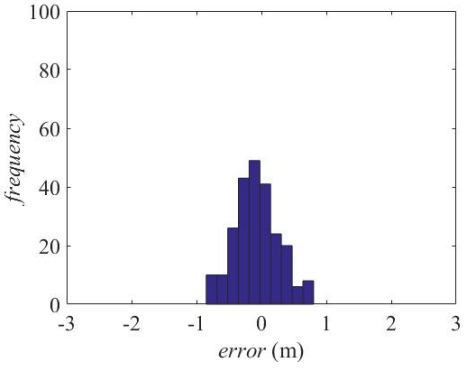
Van der Meer and Stam (1992) *

Mean=-0.28m; Skew=0.28; Kurt=3.51



Holman (1986)

Mean=-0.07m; Skew=0.13; Kurt=2.84

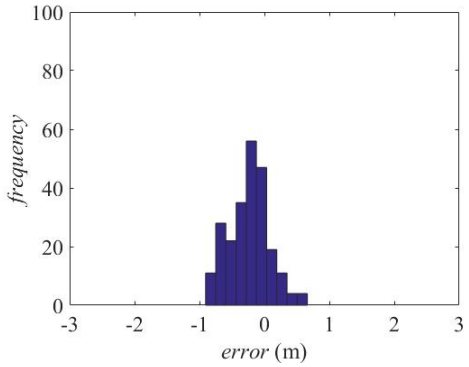


Ruggiero *et al*.

Mean= -0.23m; Skew=0.14; Kurt=2.92

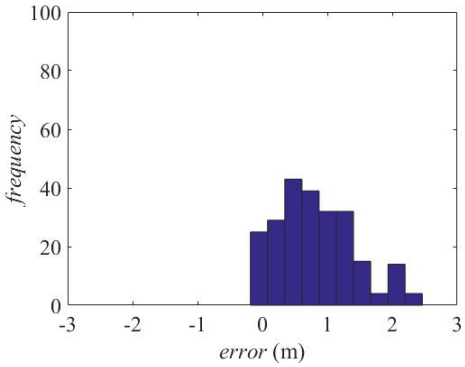
Hedges and Mase (2004) *

Mean=0.83m; Skew=0.56; Kurt=2.80



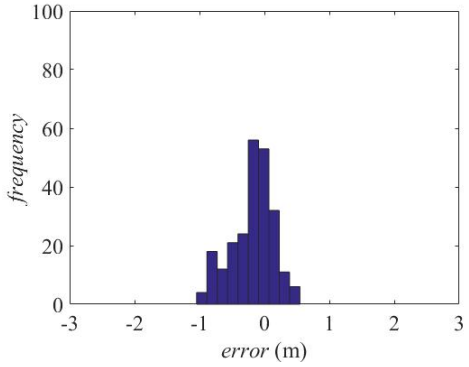
Vousdoukas *et al*. (2012)

Mean= -0.19m; Skew= -0.43; Kurt=2.90



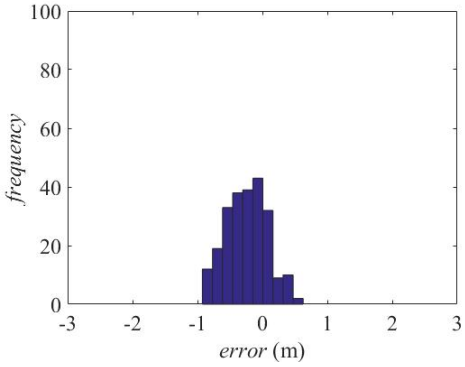
Model MM1

Mean=-0.24m; Skew=0.10; Kurt=2.65



Model MM2

Mean= -0.05m; Skew=0.17; Kurt=2.77



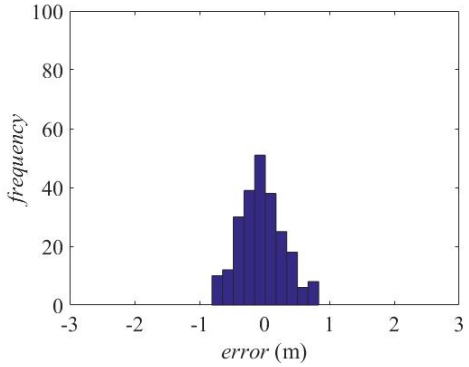
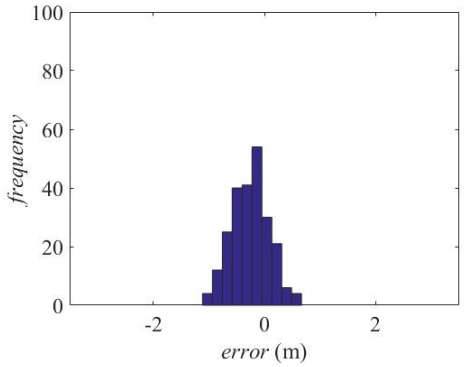


Figure 6: Histograms of model error for the $R_{2\%}$ models. The mean, skewness and kurtosis are also provided, see text for further descriptions. * denotes model developed from laboratory experiments.

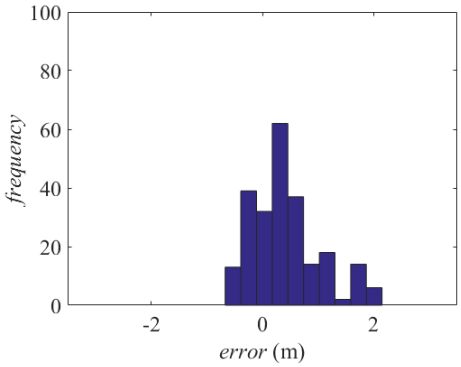
Hunt *

Mean= -0.25m; Skew= 0.06; Kurt=2.65



Mase (R_{max}) *

Mean=0.42m; Skew=0.89; Kurt=3.38



Douglass

Mean=0.46m; Skew= -0.15; Kurt=4.49

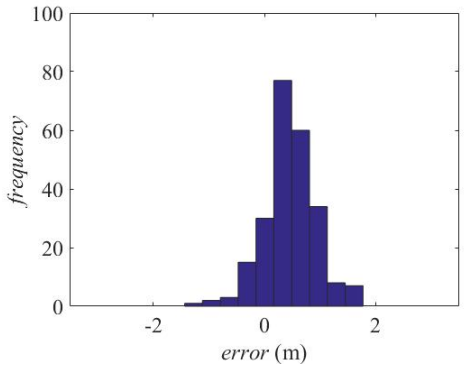


Figure 7: Error distributions for the R_{max} models. The mean, skewness and kurtosis are also provided, see text for further descriptions. * denotes model developed from laboratory experiments.

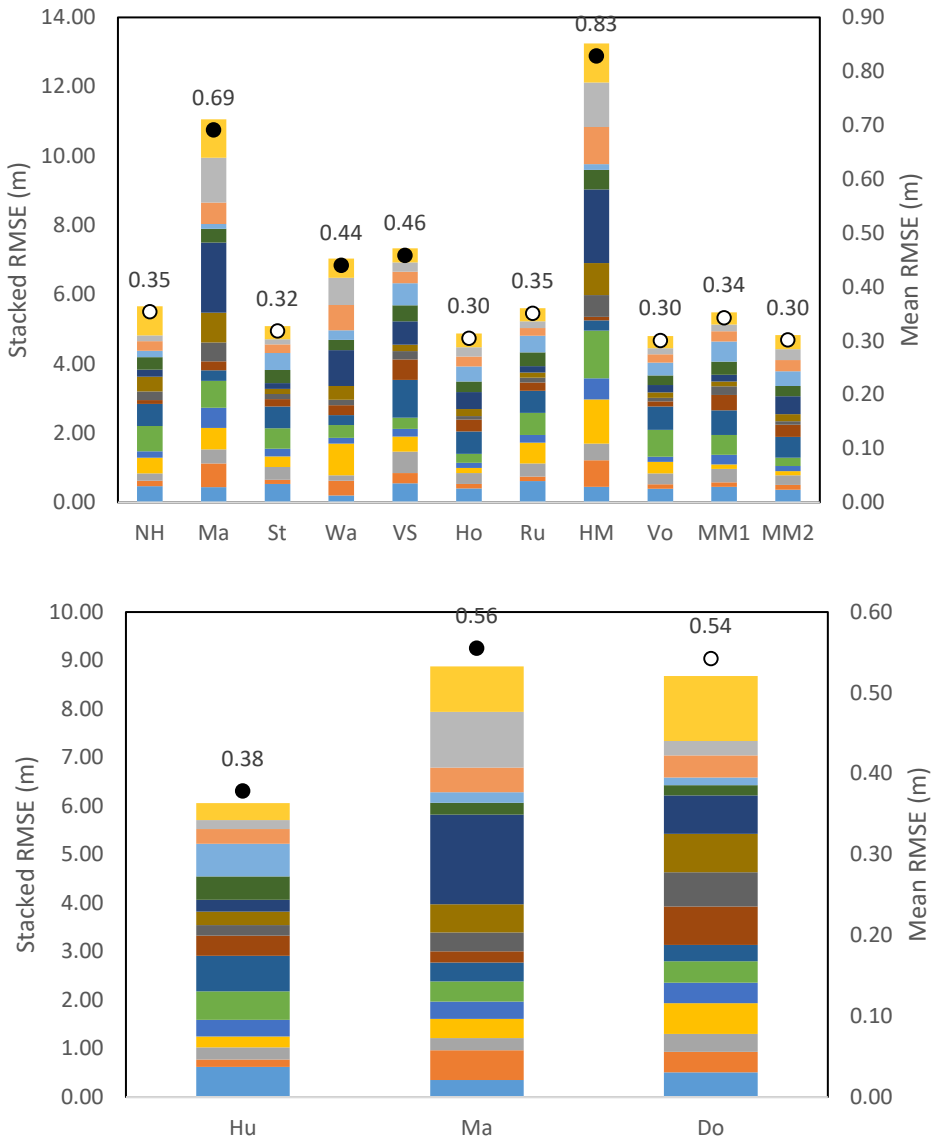


Figure 8: Stacked (left axis) and mean (right axis) RMSE for each model ($R_{2\%}$ top plot, R_{max} bottom plot) for the entire dataset. The colour changes in the stacked columns represent each beach, ordered from bottom to top as: AU24, BE29, BE30, MO26, NHB, NS1, NS2, NS3, SBB1, SBB2, ST, TA28, WE02, WE25, WO23 and YA, refer to Table 2 for abbreviations. Dots indicate mean RMSE and if model was developed from laboratory (black) or field (white) data. Refer to Table 1 for model abbreviations.

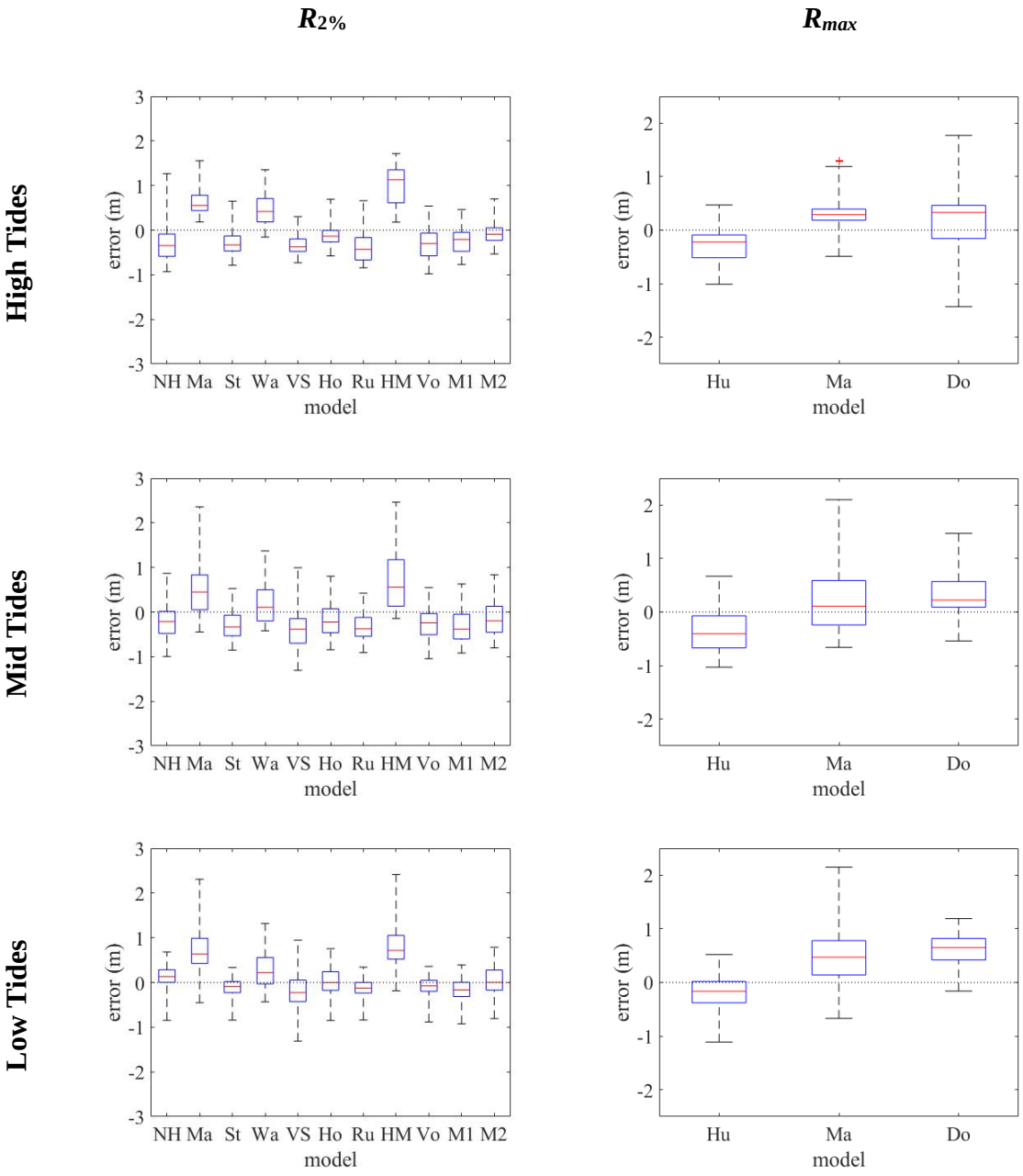
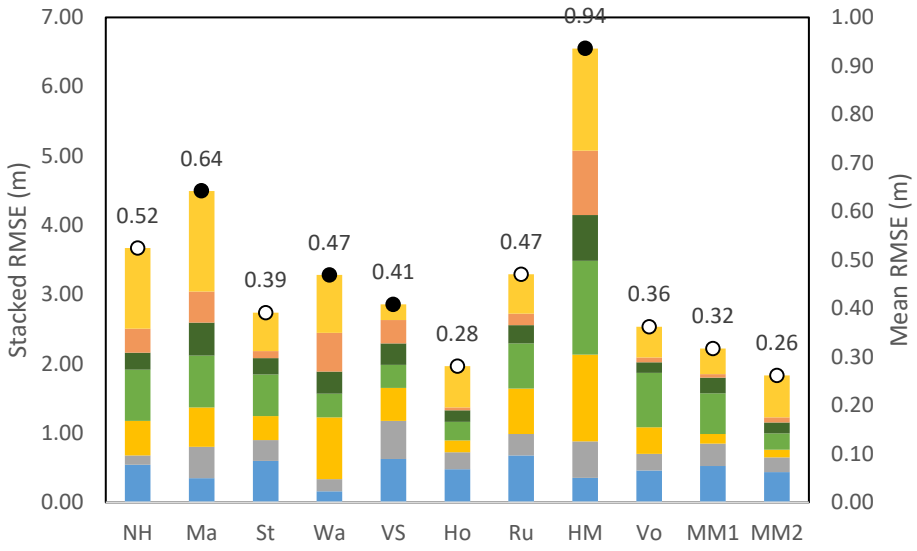
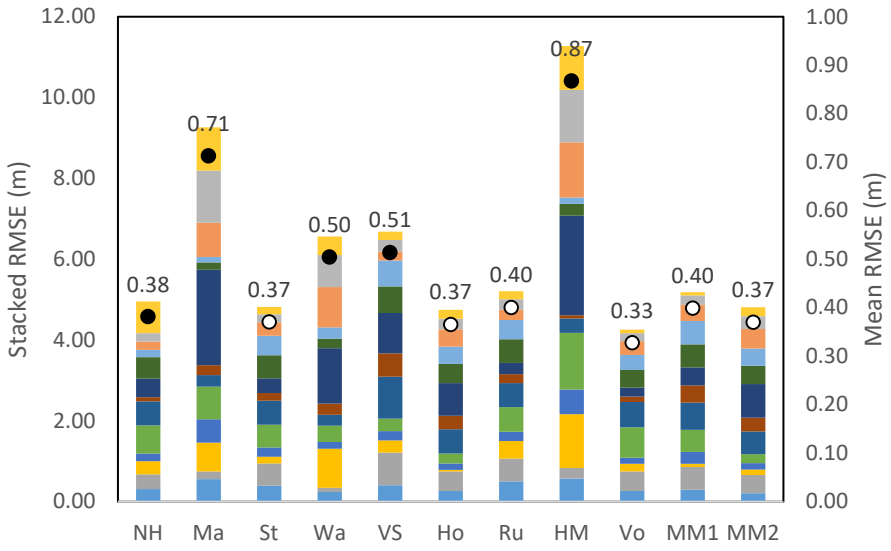


Figure 9: Box and whisker plots of model error (vertical) for all beaches (high, mid and low tides). The red lines indicate the median value and the upper and lower box boundaries indicate the upper and lower quartiles. The whiskers indicate the full range of the model errors. Refer to Table 1 for model abbreviations.


Stacked RMSE (m)
Mean RMSE (m)
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0.52
0.64
0.39
0.47
0.41
0.28
0.47
0.94
0.36
0.32
0.26
NH
Ma
St
Wa
VS
Ho
Ru
HM
Vo
MM1
MM2


Stacked RMSE (m)
Mean RMSE (m)
12.00
10.00
8.00
6.00
4.00
2.00
0.00
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0.38
0.71
0.37
0.50
0.51
0.37
0.40
0.87
0.33
0.40
0.37
NH
Ma
St
Wa
VS
Ho
Ru
HM
Vo
MM1
MM2

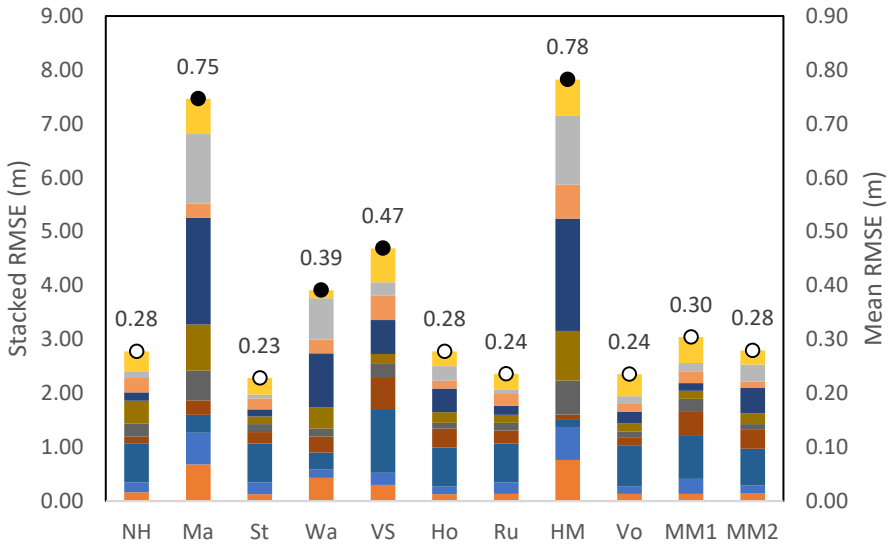
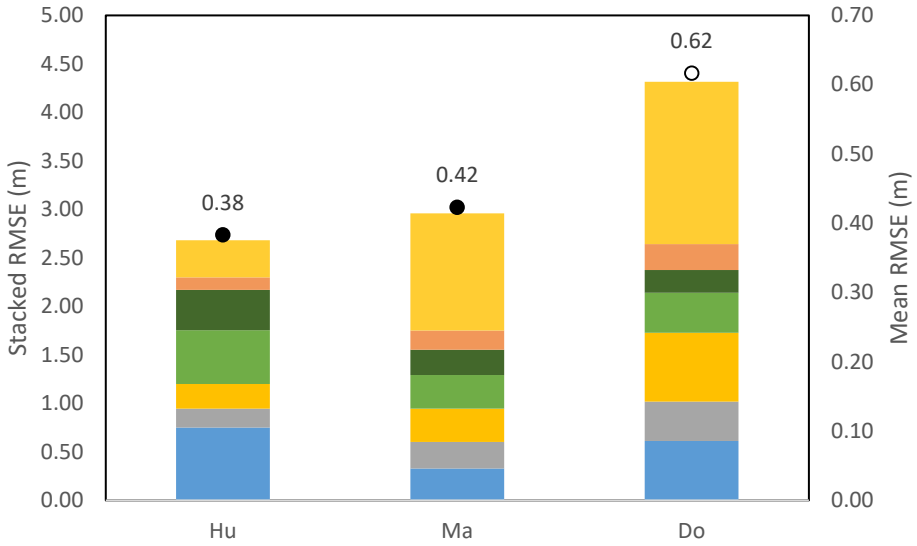
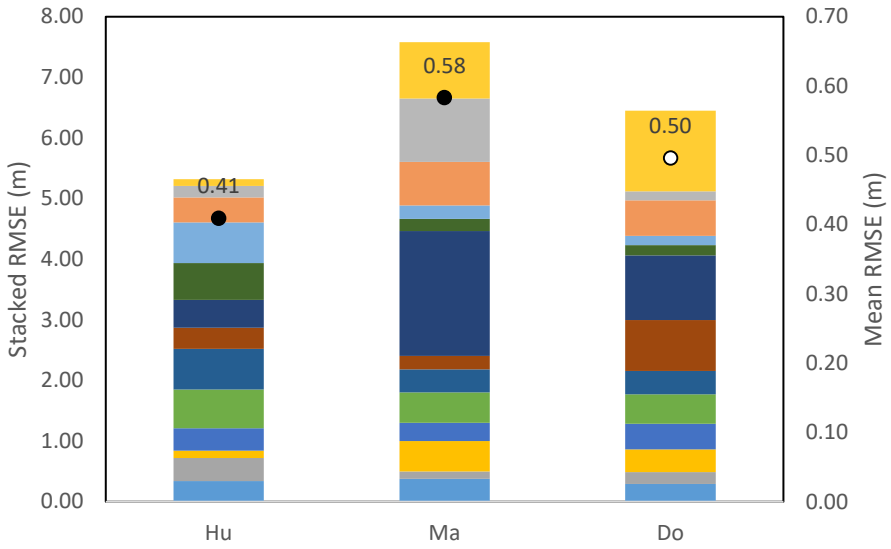

Stacked RMSE (m)
Mean RMSE (m)
9.00
8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0.28
0.75
0.23
0.39
0.47
0.28
0.24
0.78
0.24
0.30
0.28
NH
Ma
St
Wa
VS
Ho
Ru
HM
Vo
MM1
MM2

Figure 10: Stacked (left axis) and mean (right axis) RMSE for each $R_{2\%}$ model (high tide top plot, mid tide middle plot and low tide bottom plot). The colour changes in the stacked columns represent each beach, ordered from bottom to top as: AU24, BE30, MO26, NS1, TA28, WE25 and YA (top plot); AU24, BE30, MO26 NHB, NS1, NS2, NS3, ST, TA28, WE02, WE25, WO23 and YA (middle plot); BE29, NHB, NS2, NS3, SBB1, SBB2, ST, WE25, WO23 and YA (bottom plot); refer to Table 2 for abbreviations. Dots indicate mean RMSE, and if model was developed from laboratory (black) or field (white) data. Refer to Table 1 for model abbreviations.


5.00
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
Stacked RMSE (m)
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Mean RMSE (m)
0.38
0.42
0.62
Hu
Ma
Do


8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
Stacked RMSE (m)
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Mean RMSE (m)
0.41
0.58
0.50
Hu
Ma
Do

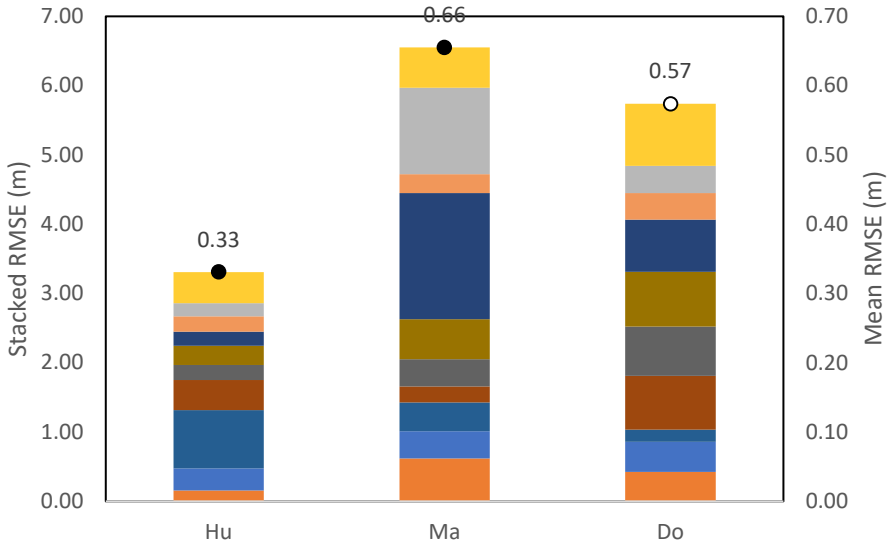

7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
Stacked RMSE (m)
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Mean RMSE (m)
0.33
0.66
0.57
Hu
Ma
Do

Figure 11: Stacked (left axis) and mean (right axis) RMSE for each R_{max} model (high tide top plot, mid tide middle plot and low tide bottom plot). The colour changes in the stacked columns represent each beach, ordered from bottom to top as: AU24, BE30, MO26, NS1, TA28, WE25 and YA (top plot); AU24, BE30, MO26 NHB, NS1, NS2, NS3, ST, TA28, WE02, WE25, WO23 and YA (middle plot); BE29, NHB, NS2, NS3, SBB1, SBB2, ST, WE25, WO23 and YA (bottom plot); refer to Table 2 for abbreviations. Dots indicate mean RMSE, and if model was developed from laboratory (black) or field (white) data. Refer to Table 1 for model abbreviations.

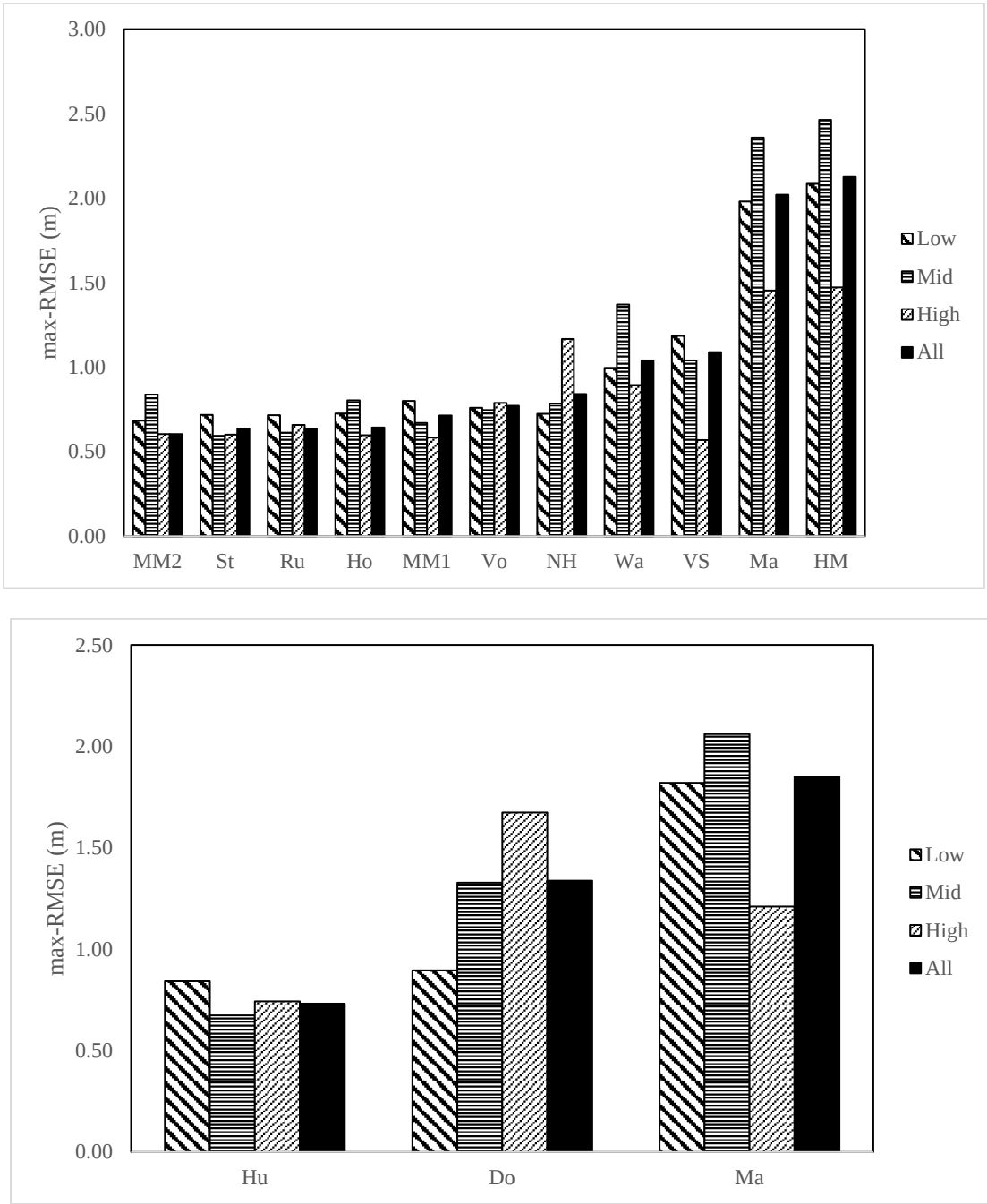


Figure 12: Maximum RMSE of each model from any beach, sorted from lowest to highest values on all tides, for $R_{2\%}$ (top) and R_{max} (bottom) for: low tides (coarse downward-right stripes), mid tides (horizontal stripes), high tides (fine upward-right stripes) and all tides (solid black).

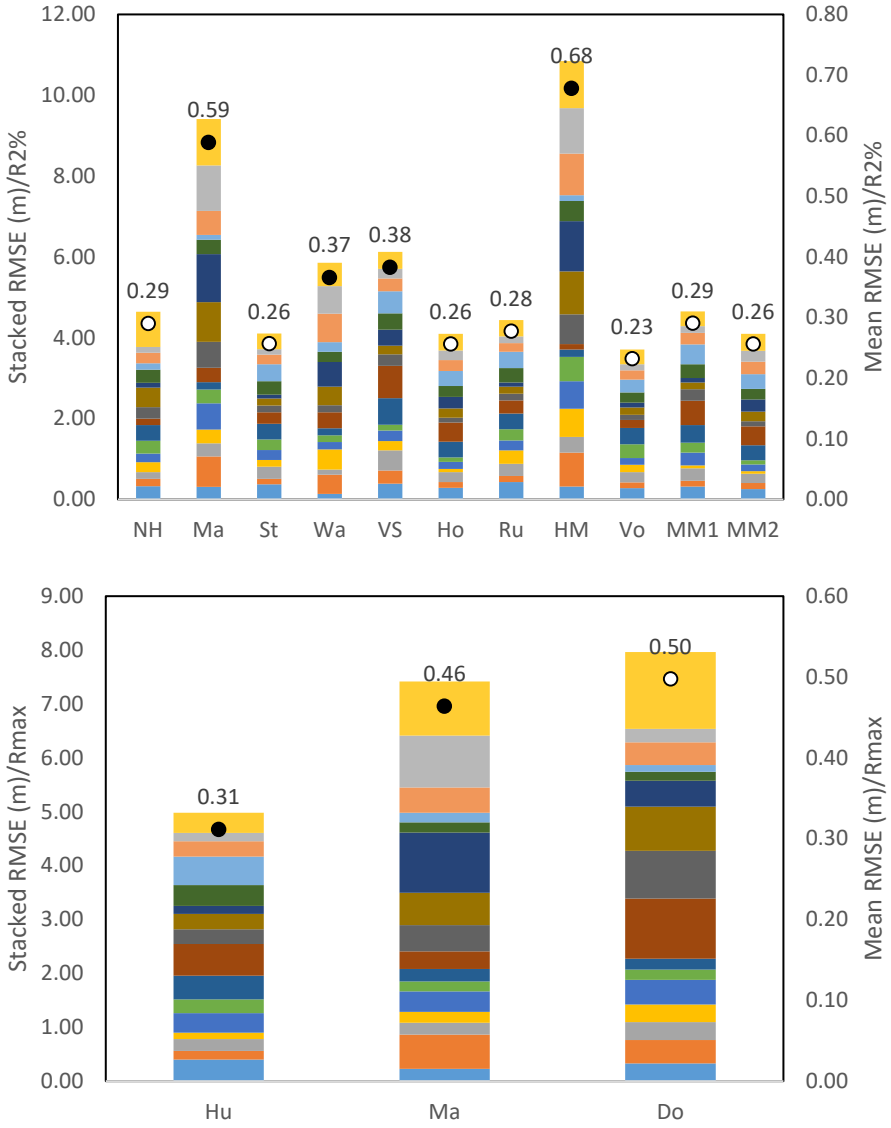


Figure 13: Stacked (left axis) and mean (right axis) normalised RMSE for each model ($R_{2\%}$ top plot, R_{max} bottom plot) for the entire dataset. The colour changes in the stacked columns represent each beach, ordered from bottom to top as: AU24, BE29, BE30, MO26, NHB, NS1, NS2, NS3, SBB1, SBB2, ST, TA28, WE02, WE25, WO23 and YA, refer to Table 2 for abbreviations.

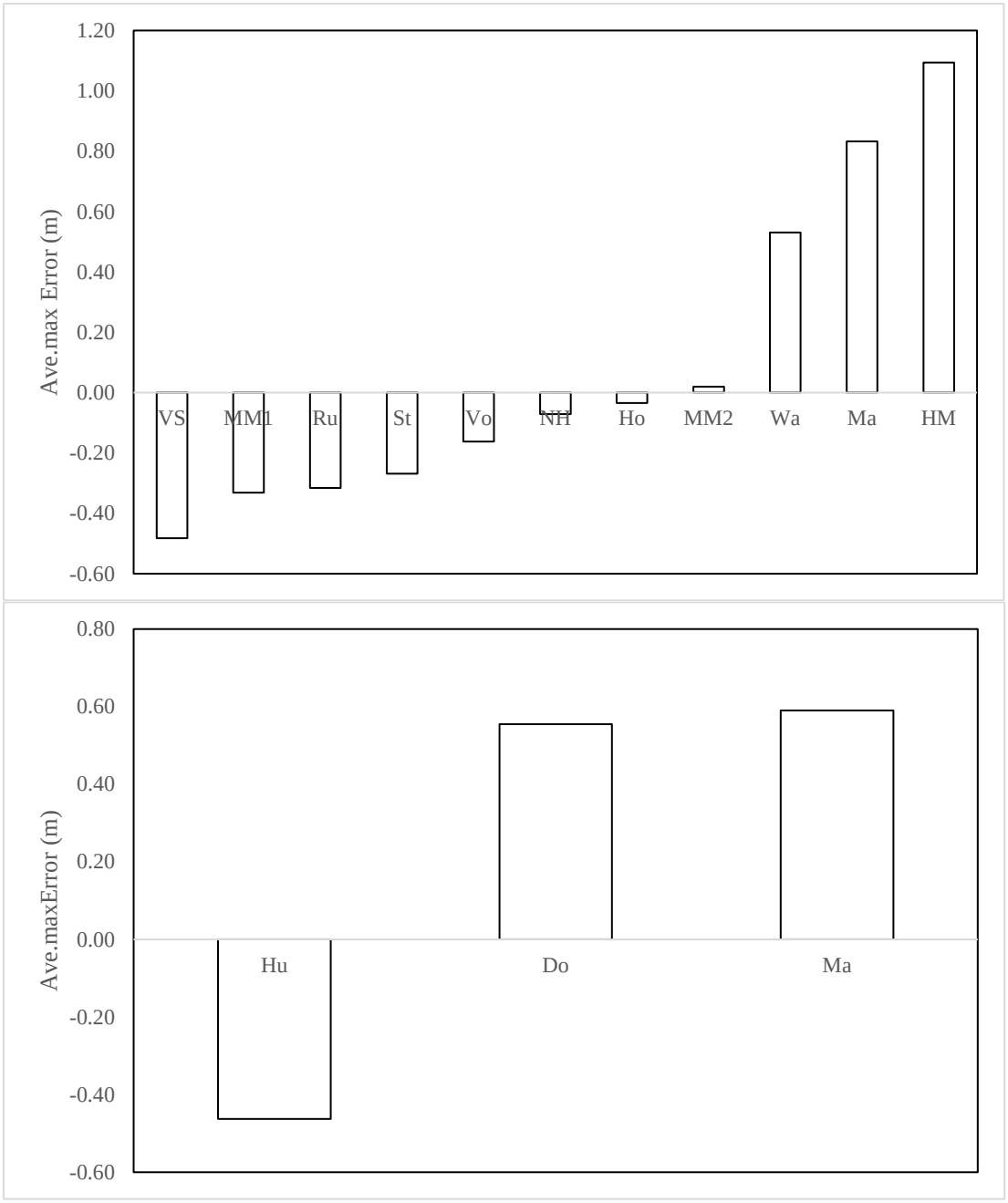


Figure 14: Average maximum error for each model averaged over all beaches, $R_{2\%}$(top) and R_{max} (bottom).

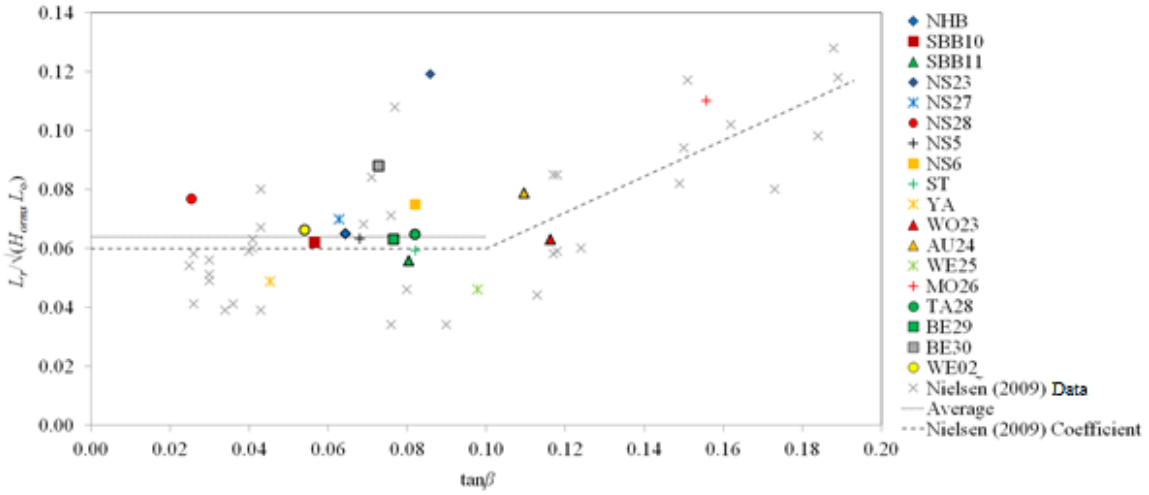


Figure 15: Runup length scale from present data plotted with the original data from Nielsen and Hanslow (1991) from six beaches in NSW and 37 individual records, shown as grey crosses, and Nielsen's (2009) lines of best fit (dotted). The new data is provided as coloured shapes and represents the mean value from the records obtained each day at each beach. The best fit line for the present data and $\beta<0.1$ is indicated (solid).