A New Probabilistic Wave Breaking Model for Dominant Wind-sea Waves Based on the Gaussian

Caio Eadi Stringari[1,1], Marc Prevosto[2,2], Jean François Filipot[1,1], Fabien Leckler[1,1], and Pedro Veras Guimaraes[1,1]

[1]France Energies Marines
[2]IFREMER

November 30, 2022

Abstract

This paper presents a novel method for obtaining the probability wave of breaking Pb of deep water, dominant wind-sea waves (that is, waves made of the energy within +-30% of the peak wave frequency) derived from Gaussian wave field theory. For a given input wave spectrum we demonstrate how it is possible to derive a joint probability density function between wave phase speed (c) and horizontal orbital velocity at wave crest (u) from which a model for Pb can be obtained. A non-linear kinematic wave breaking criterion consistent with the Gaussian framework is further proposed. Our model would allow, therefore, for application of the classical wave breaking criterion (that is, wave breaking occurs if u/c > 1) in spectral wave models which, to the authors' knowledge, has not been done to date. Our results show that the proposed theoretical model has errors in the same order of magnitude as six other historical models when assessed using three field datasets. With optimization of the proposed model's single free parameter, it can become the best performing model for specific datasets. Although our results are promising, additional, more complete wave breaking datasets collected in the field are needed to comprehensively assess the present model, especially in regards to the dependence on phenomena such as direct wind forcing, long wave modulation and wave directionality.

A New Probabilistic Wave Breaking Model for Dominant Wind-sea Waves Based on the Gaussian Field Theory

C. E. Stringari[1], M. Prevosto[2], J.-F. Filipot[1], F. Leckler[1], P. V. Guimarães[1,3]

[1]France Energies Marines, Plouzané, France
[2]Institut Français de Recherche pour l'Exploitation de la Mer, Plouzané, France
[3]PPGOceano, Federal University of Santa Catarina, Florianópolis, 88040-900, Brazil

Key Points:

- A new probabilistic wave breaking model based on Gaussian field theory is presented for dominant, wind-sea waves.
- Wave breaking probabilities are modeled from the joint probability density between wave phase speed and particle orbital velocity.
- The proposed model performs well when compared to six other historical models using three field datasets.

Corresponding author: C.E. Stringari, `Caio.Stringari@france-energies-marines.org`

Corresponding author: J.F. Filipot, `Jean.Francois.Filipot@france-energies-marines.org`

Abstract

This paper presents a novel method for obtaining the probability wave of breaking (P_b) of deep water, dominant wind-sea waves (that is, waves made of the energy within ±30% of the peak wave frequency) derived from Gaussian wave field theory. For a given input wave spectrum we demonstrate how it is possible to derive a joint probability density function between wave phase speed (c) and horizontal orbital velocity at wave crest (u) from which a model for P_b can be obtained. A non-linear kinematic wave breaking criterion consistent with the Gaussian framework is further proposed. Our model would allow, therefore, for application of the classical wave breaking criterion (that is, wave breaking occurs if $u/c > 1$) in spectral wave models which, to the authors' knowledge, has not been done to date. Our results show that the proposed theoretical model has errors in the same order of magnitude as six other historical models when assessed using three field datasets. With optimization of the proposed model's single free parameter, it can become the best performing model for specific datasets. Although our results are promising, additional, more complete wave breaking datasets collected in the field are needed to comprehensively assess the present model, especially in regards to the dependence on phenomena such as direct wind forcing, long wave modulation and wave directionality.

Plain Language Summary

Waves will break if the speed of the water particles on the wave crest is greater than the speed of the wave itself, causing the wave crest to overtake the front part of the wave, leading to wave breaking. Precisely simulating real ocean waves requires, therefore, a particle-by-particle description of the water motion, which is too expensive for the current computers to handle in real-world applications. Instead, wave models describe waves by means of their statistical properties, that is, averaged over a large number of waves. In this paper, we present a mathematical formulation that allows to calculate the combined probability between the speed of particles on the wave crest and the wave speed based only on statistical properties. From these combined probabilities, we model the probability of wave breaking. Our results indicate that our model performed relatively well when compared to six other models using three historical datasets. Because of a lack of observed data to assess our model, we recommend that future research should focus on collecting more wave breaking data measured in the field. Future advances on this line of

research could lead, for example, to improvements on operational weather forecast models.

1 Introduction

A robust description of wave breaking is a crucial aspect of wave modelling. It is via wave breaking that most of the wave energy is dissipated and a precise formulation of this phenomenon is required to obtain reliable models. Despite of its importance, energy dissipation due to wave breaking is still modelled as a semi-empirical process due to the difficulty to represent physically-derived wave breaking criteria on phase-averaged wave models (Battjes & Janssen, 1978; Thornton & Guza, 1983; Banner et al., 2000; Filipot et al., 2010; Filipot & Ardhuin, 2012; Banner et al., 2002; Ardhuin et al., 2010; Banner et al., 2014; Zieger et al., 2015; Ardag & Resio, 2020). The available probabilistic (that is, parametric, or empirical) formulations included in these models have been derived from limited datasets and without rigorous theoretical frameworks and, therefore, they currently lack a solid physical background. While the current operational (spectral) models are capable of reproducing field observations of integrated spectral parameters (for example, significant wave height, peak wave period and peak wave direction) with good accuracy, it remains unclear if their wave breaking parameterizations are entirely reliable. This knowledge gap partly occurs because limited research has focused on wave breaking statistics derived from field data, especially when it comes to wave breaking observations distributed as a function of wave scales (for example, wave frequency or wave phase speed). The research developed here has, therefore, important implications for air-sea flux parameterizations (Kudryavtsev et al., 2014), safety at sea (Kjeldsen et al., 1980) and design of offshore structures (Filipot et al., 2019), all of which directly rely on the properties of breaking waves.

Historically, parametric wave breaking models have been constructed from two different approaches: the first approach considers wave statistics (wave steepness, most frequently) derived from a wave-by-wave analysis of the surface elevation timeseries collected at a single point location where wave breaking occurrences are synchronously identified (using video data, most frequently). The wave breaking probability (that is, the ratio between the total number of breaking waves over the total number of waves during a given period of time) can then be expressed as a bulk quantity (Thornton & Guza, 1983; Chawla & Kirby, 2002; Alsina & Baldock, 2007; Janssen & Battjes, 2007) or can be distributed over wave frequency (f), wavenumber (k), or wave speed (c) ranges, referred as to "wave scales" by the wave modelling community (Eldeberky & Battjes, 1996; Banner et al., 2002; Filipot et al., 2010).

The second approach follows from Phillips (1985) who defined the distribution $\Lambda(c)dc$ as the "average total length per unit surface area of breaking fronts that have velocities in the range c to $c+dc$". This approach therefore relates to the analysis of sea surface images in which individual wave breaking patches are tracked in space and time. The main motivation for introducing this new concept was clearly stated in Phillips (1985): "There is clearly some association of the breaking events with waves of different scales, but it is difficult to make the association in an unambiguous way if we consider only the surface configuration at one given instant. A breaking crest may indeed be a local maximum in the instantaneous surface configuration but there is no guarantee that a local wavelength of the breaking wave can be defined clearly. It seems more satisfactory to use the velocity c of the breaking front as a measure of the scale of the breaking". This quotation clearly identify the limitations of directly relying on the analysis of single point elevation timeseries. Different parameterizations have been proposed to quantity $\Lambda(c)dc$ from theoretical (Phillips, 1985) or empirical considerations (Melville & Matusov, 2002; Sutherland & Melville, 2013; Romero, 2019). However, Phillips' (1985) framework remains controversial, particularly regarding its practical application, given that different interpretations of his concepts can generate differences of several orders of magnitude in the calculations of $\Lambda(c)dc$ and its moments (Banner et al., 2014). For a detailed review of commonly used parametric wave breaking models please refer to Appendix A.

Interestingly, while the ratio between the horizontal orbital velocity at the crest (u) to wave phase speed (c) appears the most reliable parameter to determine wave breaking occurrence (Saket et al., 2017; Barthelemy et al., 2018; Derakhti et al., 2020; Varing et al., 2020), it was not used by any of the approaches mentioned above. This paper provides a new promising wave breaking model by revisiting Rice (1944) and Longuet-Higgins (1957) statistical descriptions of Gaussian processes (that is, for linear waves) to obtain the theoretical joint probability density between c and u ($p(c, u)$). We then model P_b assuming a kinematic wave breaking criterion consistent with non-linear waves, that is, a wave breaks if the fluid velocity at the wave crest is greater than the wave phase speed ($u > c$). This study focuses on analysing dominant waves, defined as waves that have frequencies within ±30% of the spectral peak frequency of the wind-sea (Banner et al., 2000). Future research will be dedicated to extend our efforts to broader wave scales. This paper is organized as follows: Section 2 describes the proposed model, Section 3 presents three historical datasets used to evaluate the model, Section 4 presents the results, Section 5 discusses and Section 6 concludes.

2 Definition of a Probabilistic Wave Breaking Model Based on Gaussian Field Theory

The kinematic wave breaking criterion $u/c = 1$ has been historically used as the onset of wave breaking for non-linear, real waves (see Perlin et al. (2013) for a review). Recently, Barthelemy et al. (2018) found and Derakhti et al. (2020) confirmed via numerical simulations that waves will inevitably start to break shortly after u/c exceeds 0.85 in deep and shallow water. Further numerical simulations showed that wave breaking occurs when the maximum orbital velocity (u_{max}) equals c somewhere along the wave profile and not necessarily at the wave crest (Varing et al., 2020). Although the relationship u/c provides a solid physical background to establish the onset of wave breaking, this approach has never been applied to spectral wave models because it requires phase-resolving the wave field. In the sections below, we circumvent this difficulty by defining a wave breaking probability model using the joint probability density between c and u corresponding to a given wave energy spectrum ($E(f)$). The efforts in this paper are consistent with part of the recent work from Ardag and Resio (2020) in the sense that both works aim to solidify the use of the kinematic wave breaking criterion as the standard approach for modelling wave breaking.

2.1 Theoretical Derivation of the Joint Probability Density Distribution of Orbital Velocity at the Wave Crest and Phase Speed

Longuet-Higgins (1957) published a very complete work on the statistics of Gaussian wave fields. In particular, Longuet-Higgins (1957) studied the probability density of the speed of zero-crossings along a given line that is of interest for us in this work. In his paper, the speed of zero-crossings were applied in particular to the zero-crossings of the space derivative of a Gaussian process, that is, the velocities of the local maxima in space (Longuet-Higgins (1957), pp. 356-357). The present work describes how the same methodology can be extended to derive the joint density of the speed of space local maxima (or local crests) and simultaneous wave horizontal orbital velocity for a one-dimensional

Gaussian sea state. For simplicity, this paper follows the same notations as those of Longuet-Higgins (1957) and the reader is directed to Section 2.5 in Longuet-Higgins (1957) for further details.

As explained in Longuet-Higgins (1957), if $\xi_1(x,t)$ is a stationary-homogeneous process and we are interested in the points (for example, in space) were this process crosses a level x_1, the joint distribution of the space derivative of ξ_1 noted ξ_2, with other related processes ξ_3, ξ_4, ... at $\xi_1 = x_1$ is given by:

$$p(\xi_2,\xi_3,\xi_4,...)_{x_1} = \frac{|\xi_2|\, p(\xi_1,\xi_2,\xi_3,\xi_4,...)|_{\xi_1=x_1}}{N_0(x_1)} \tag{1}$$

where $N_0(x_1)$ is the number of crossings of the level x_1 by ξ_1 (see Equation 2.2.5 in Longuet-Higgins (1957)). In this paper we are interested in joint distributions at the local maxima in space of the wave elevation process ξ_0. Therefore, ξ_1 is the space derivative of the wave process and local maxima correspond to down-crossings of the zero level by $\xi_1 = \partial\xi_0/\partial x$.

$$\xi_1 = \frac{\partial\xi_0}{\partial x}\ ,\ \xi_2 = \frac{\partial^2\xi_0}{\partial x^2} = \frac{\partial\xi_1}{\partial x}. \tag{2}$$

In the case of Gaussian processes, $N_0^-(x_1)$ is:

$$N_0^-(x_1) = \frac{1}{2\pi}\sqrt{\frac{m_4}{m_2}}\exp\left(-\frac{x_1^2}{2m_2}\right)\ ,\ N_0^- = N_0^-(0) = \frac{1}{2\pi}\sqrt{\frac{m_4}{m_2}} \tag{3}$$

where $m_0,\ m_1,\ \ldots,\ m_i$ are the i-th wavenumber spectral moments and the minus sign indicates that we consider only down-crossings.

2.1.1 Speed of Local Maxima (Phase Speed)

Following Longuet-Higgins (1957), if we are interested in the speed c of the local maxima in space, that is, the speed of the down-crossings of ξ_1, we have:

$$c = -\frac{\partial\xi_1/\partial t}{\partial\xi_1/\partial x} = -\frac{\xi_3}{\xi_2} \text{ with } \xi_2 = \frac{\partial^2\xi_0}{\partial x^2} \text{and } \xi_3 = \partial\xi_1/\partial t. \tag{4}$$

Using Equation 1,

$$p(\xi_2, \xi_3)_0 = \frac{|\xi_2|\, p(\xi_1, \xi_2, \xi_3)|_{\xi_1=0}}{N_0^-} \tag{5}$$

with $p(\xi_1, \xi_2, \xi_3)$ the point joint distribution of the three Gaussian processes $\frac{\partial \xi_0}{\partial x}, \frac{\partial^2 \xi_0}{\partial x^2}, \frac{\partial^2 \xi_0}{\partial x \partial t}$ is:

$$p(\xi_1, \xi_2, \xi_3) = p(\xi_1)\, p(\xi_2, \xi_3) = \frac{\mathrm{e}^{-\frac{\xi_1^2}{2m_2}}}{2\pi\sqrt{m_2}} \frac{\mathrm{e}^{-\frac{1}{2}\left([\xi_2 \xi_3] Q_c^{-1} \begin{bmatrix} \xi_2 \\ \xi_3 \end{bmatrix}\right)}}{\sqrt{(2\pi)^3 \det(Q_c)}} \tag{6}$$

and covariance matrix:

$$Q = \begin{bmatrix} m_2 & 0 & 0 \\ 0 & m_4 & m_3' \\ 0 & m_3' & m_2'' \end{bmatrix} = \begin{bmatrix} m_2 & 0 \\ 0 & Q_c \end{bmatrix}. \tag{7}$$

Note that following Longuet-Higgins (1957) notations, m_i'' indicates the mixed wavenumber-frequency i-th spectral moment, where the number of quotes indicates the order of the frequency spectral moment, for example,

$$m_3' = \int_0^\infty 2\pi f^1 k^3 E(k) dk, \tag{8}$$

where $E(k)$ is a given wavenumber spectra.

Classically, to introduce c in the joint density and obtain $p(c, \xi_3)_0$, we apply a change of variables

$$\xi_2 = -\frac{\xi_3}{c}\ ,\ \xi_3 = \xi_3 \tag{9}$$

and after the integration of $p(c, \xi_3)_0$ over all the domain of definition of ξ_3, we obtain the distribution of c (Longuet-Higgins (1957), Eq. 2.5.19):

$$p(c)_0 = \frac{1}{2} \frac{m_4 m_2'' - m_3'^2}{\sqrt{m_4}\left(c^2 m_4 + 2c m_3' + m_2''\right)^{3/2}} \tag{10}$$

Note that the sign on c (or on $m_3^{'}$) depends on the convention on the wave propagation direction. We have kept the convention used by Longuet-Higgins (1957) here.

2.1.2 Introducing the Orbital Velocity

As indicated in Equation 1, we can introduce in the formula a variable which represents the horizontal orbital velocity. For Gaussian waves the horizontal orbital velocity u is defined as

$$u = \mathcal{H}_t \left(\frac{\partial \xi_0}{\partial t} \right) \tag{11}$$

with $\mathcal{H}_t$ the Hilbert transform in time domain. Which means that

$$\xi_0 = \sum_i a_i \cos\left(k_i x - \omega_i t\right) \tag{12}$$

is transformed in

$$u = \sum_i a_i \omega_i \cos\left(k_i x - \omega_i t\right), \tag{13}$$

with a_i the wave amplitude, k_i the wavenumber and ω_i the angular wave frequency of the wave component i. As the Hilbert transform is a linear operator, u is also Gaussian. As previously, at the local maxima we have:

$$p\left(\xi_2, \xi_3, u\right)_0 = \frac{|\xi_2|\, p\left(\xi_1, \xi_2, \xi_3, u\right)|_{\xi_1=0}}{N_0^-} \tag{14}$$

with a new covariance matrix for ξ_1, ξ_2, ξ_3 and u:

$$Q = \begin{bmatrix} m_2 & 0 & 0 & 0 \\ 0 & m_4 & m_3^{'} & m_2^{'} \\ 0 & m_3^{'} & m_2^{''} & m_1^{''} \\ & m_2^{'} & m_1^{''} & m_0^{''} \end{bmatrix} = \begin{bmatrix} m_2 & 0 \\ 0 & Q_c \end{bmatrix}. \tag{15}$$

As previously, we can apply a similar change of variables

$$\xi_2 = -\frac{\xi_3}{c}\ ,\ \xi_3 = \xi_3\ ,\ u = u, \tag{16}$$

or the easiest to deal with,

$$\xi_3 = -c\xi_2\ ,\ \xi_2 = \xi_2\ ,\ u = u \tag{17}$$

and integrate $p\left(c, \xi_2, u\right)_0$ over all the domain of definition of ξ_2. The result is a more complicated but again semi-analytical. The body of the integral has the form

$$\mathrm{e}^{-\frac{1}{2}\left[\xi(c)\xi_2^2+\beta(c,u)\xi_2+\alpha(u)\right]}\xi_2^2 \tag{18}$$

and its integration in ξ_2 on the down-crossings space $]-\infty, 0]$ gives

$$I\left(c, u\right) = \frac{\left(\left(2\phi^2+1\right)\sqrt{\pi}\left(\mathrm{erf}\left(\phi\right)+1\right)\mathrm{e}^{\phi^2}+2\phi\right)}{\sqrt{2}\xi^{3/2}}\mathrm{e}^{-\alpha/2} \tag{19}$$

with

$$\phi = \phi\left(c, u\right) = \frac{1}{2\sqrt{2}}\frac{\beta(c, u)}{\sqrt{\xi\left(c\right)}}, \quad \alpha = \alpha(u), \tag{20}$$

$$\Delta = \det(Q_c) = m_3^{'}\left(m_2^{'}m_1^{''} - m_3^{'}m_0^{''}\right) + m_4\left(m_0^{''}m_2^{''} - m_1^{''2}\right) + m_2^{'}\left(m_3^{'}m_1^{''} - m_2^{'}m_2^{''}\right), \tag{21}$$

$$\alpha(u) = \frac{m_4 m_2^{''} - m_3^{'2}}{\Delta}u^2, \tag{22}$$

$$\beta(c, u) = 2\frac{m_3^{'}m_1^{''} - m_2^{'}m_2^{''}}{\Delta}u + 2\frac{m_4 m_1^{''} - m_2^{'}m_3^{'}}{\Delta}uc \tag{23}$$

and

$$\xi\left(c\right) = \frac{m_0^{''}m_2^{''} - m_1^{''2}}{\Delta} + 2\frac{m_3^{'}m_0^{''} - m_2^{'}m_1^{''}}{\Delta}c + \frac{m_4 m_0^{''} - m_2^{'2}}{\Delta}c^2. \tag{24}$$

The joint probability density of (c, u) is then:

$$p\left(c,u\right)=\frac{1}{N_0^-}\frac{1}{(2\pi)^2\sqrt{m_2\Delta}}I\left(c,u\right)=\frac{I\left(c,u\right)}{2\pi\sqrt{m_4\Delta}}. \tag{25}$$

Note again that the sign on c and u (or on $m_2^{'}$ and $m_3^{'}$) depends on the convention on the wave propagation direction and Longuet-Higgins (1957)'s convention is still used here. The coefficients (α,β,ξ) can be calculated directly numerically and Δ is the determinant of Q_c, the sub-matrix of Q, and after the inverse of Q_c is calculated:

$$Q_c^{-1}=\begin{bmatrix} R & \boldsymbol{s} \\ \boldsymbol{s}^t & r \end{bmatrix} \tag{26}$$

we find

$$\alpha(u)=ru^2, \tag{27}$$

$$\beta(c,u)=2\begin{bmatrix} 1 & c \end{bmatrix}\boldsymbol{s}u, \tag{28}$$

$$\xi\left(c\right)=\begin{bmatrix} 1 & c \end{bmatrix}R\begin{bmatrix} 1 \\ c \end{bmatrix}. \tag{29}$$

An example of the joint density of the couple (phase speed, horizontal particle velocity) at local maxima in space is shown in Figures 1-a and b for a JONSWAP spectrum.

2.2 Modelling P_b from $p(c,u)$

By using Equation 25 applied to the dominant spectral wave band (that is, that contained in the interval $[0.7f_p,\ 1.3f_p]$, where f_p is the peak wave frequency), the probability of dominant wave breaking can be computed by integrating Equation 25 over all phase speeds and for orbital velocities over a threshold Ac, with A a constant that will be in the next section:

$$P_b=\int_{u>Ac}\int_0^\infty p(c,u)dcdu. \tag{30}$$

P_b will be modelled following Equation 30 hereafter. Note that from the definitions in Equation 3, the proposed P_b is defined as number of breaking local maxima over the total number of local maxima. From the analysis of $p(c, u)$ we observed that spurious, non-moving local maxima may exist around $c = 0$ and $u = 0$; therefore, to avoid artificially increasing P_b, we adopted a practical integration range of $c, u \in [0.05, +\infty]$ here. Note that this range may, however, only be valid for very narrow spectra. Further, we draw attention that, following from Equation 1, our P_b model is defined in space domain, whereas all the previous P_b models and data are (at least partially) defined in time domain (see Appendix A for details). For the very narrow spectral band used here, the differences between temporal and spatial definitions of Pb are negligible. This is discussed further in Section 5.

Finally, the proposed model can be extended to accommodate two-dimensional spectra without changes on how $p(c, u)$ is calculated. This is done by applying an appropriated spreading function to any given one-dimensional spectra (or directly inputting a directional spectra) and by recalculating the moments in Equations 8 to take directionality into account or, more explicitly,

$$m_i = \int_0^{2\pi} \int_0^{\infty} \left(f \cos\theta \cos\alpha + f \sin\theta \sin\alpha\right)^i E(f, \theta) df d\theta. \tag{31}$$

An example considering the simplified cosine spreading law ($D(\theta) = cos(\theta - \bar{\theta})^{2s}$) with $s = 20$, $\bar{\theta} = 0$ and α=0 applied to same JONSWAP spectrum shown in Figure 1-a is shown in Figure 1-c. Note that the differences in $p(c, u)$ between the one-dimensional (Figure 1-b) and the two-dimensional (Figure 1-d) spectra are negligible for the present assumptions. This relatively simple extension allows for the consideration of two-dimensional wave spectral but we caution the reader that it may not be fully complete. A follow-up publication will be dedicated to include and assess the effects of wave directionality in our method more rigorously.

2.3 Definition of a Gaussian-equivalent Non-linear Wave Breaking Criterion

The previously introduced joint probability density distribution $p(c, u)$ is based on Gaussian theory and therefore assumes that waves are linear. Breaking waves are, however, highly non-linear. For real non-linear waves, as detailed in the introduction, it is widely accepted that wave breaking starts when the water particle horizontal velocity at its crest (u_{nl}) reaches the wave phase speed (c_{nl}). A non-linear wave breaking crite-

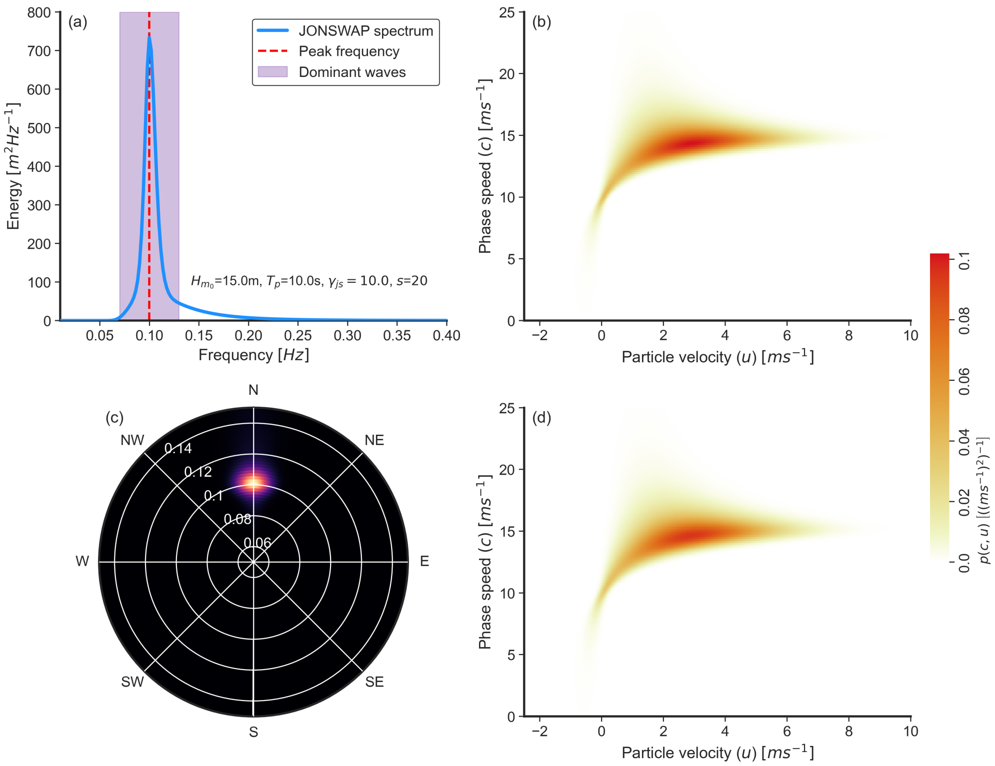

Figure 1. Example of the application of the method. a) JONSWAP spectrum for H_{m_0}=15m, T_p=10s and shape parameter γ_{js}=10. b) Obtained joint probability density between the wave phase speed (c) and the horizontal particle velocity at wave crest (u) calculated using Equation 25. Note that the joint probability density was computed using only the spectral energy between $0.7f_p$ and $1.3f_p$, that is, corresponding to the dominant wave band only. c) Directional spectrum for the same parameters as in a) and directional spreading $D(\theta) = cos(\theta - \bar{\theta})^{2s}$ with $s = 20$ and $\bar{\theta}$ = 0. d) Obtained $p(c, u)$ considering only the spectral energy in the direction $\alpha = 0$.

rion can be thus be defined as $A_{nl} = u_{nl}/c_{nl} = 1$. Therefore, we assume that it is possible to obtain an equivalent kinematic criterion, $A_{lin} = constant$ that relates Gaussian waves to non-linear waves.

Based on numerical experiments, Cokelet (1977) provided the potential and kinetic energy of a fully non-linear regular wave in deep-water at the onset of wave breaking (see the last row of his Table A.0). Based on his results, we define the kinematic criterion as the linear wave that has total energy equals to the nearly breaking non-linear regular wave computed by Cokelet (1977). Following Cokelet (1977), where k, g and ρ are expressed as non-dimensional variables, a deep-water wave at the breaking onset (see last row of his table A.0) has kinetic energy $T = 3.827 \times 10^{-2}$ and potential energy $V = 3.457 \times 10^{-2}$. The energy-equivalent linear wave (denote with subscript eq) has, therefore, amplitude:

$$a_{eq} = \sqrt{2 \times E} = \sqrt{2 \times (V + T)} = 0.3817. \tag{32}$$

For this particular case, the linear dispersion relation reads:

$$\omega^2 = gk = 1, \tag{33}$$

the fluid velocity at crest of the energy-equivalent linear wave is:

$$u_{eq} = \omega a_{eq} = 0.3817, \tag{34}$$

and the phase speed of the linear wave is:

$$c_{eq} = \sqrt{\frac{g}{k}} = 1. \tag{35}$$

Given these constants, we obtain:

$$A_{lin} = \frac{u_{eq}}{c_{eq}} = \frac{0.3817}{1} = 0.3817. \tag{36}$$

Following this approach, we define the correction coefficient $A = A_{lin} = 0.382$ that will be used as reference value hereafter for our tests. This result is consistent with recent findings from Ardag and Resio (2020) who reported from the re-analysis of Duncan's (1981) experimental results, a wave breaking threshold between 0.75 and 1.02 (see their Figure 1). Note, however, that these authors defined their wave breaking threshold as u/c_g, where c_g is the group velocity and u was obtained from linear wave theory. Replacing wave group velocity (c_g) by the wave phase speed (c) yields a range of possible values between 0.35 and 0.50, which is consistent with A_{lin}.

Figure 2 illustrates the sensitivity in wave breaking probability with changes in the wave breaking threshold A. For the given $p(c, u)$ in Figure 2-a, letting A to vary from 0 to 1 resulted in a exponential increase in P_b at $A \leq 0.2$ (Figure 2-b), which may be unrealistic. When setting A=A_{lin}=0.382 and letting the significant wave height (H_{m_0}) and wave peak period (T_p) vary in the definition of the JONSWAP spectrum, the results indicate that steeper waves are more probable to break, which is expected (Figure 2-c). Finally, note that the wave breaking threshold A might be sensitive to other wave and atmospheric parameters such as wave directionality or direct wind forcing (or, equivalently, wave age). In the next sections, the accuracy of our model is assessed using field observations and our results are compared with other parametric wave breaking formulations.

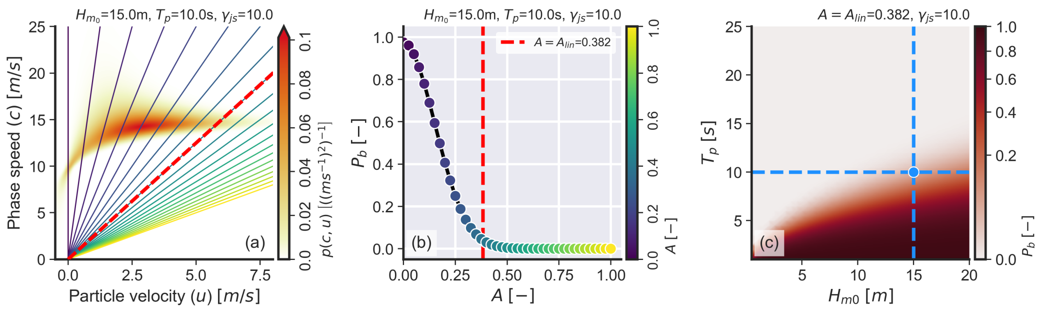

Figure 2. a) Example of joint probability density between u and c obtained from Equation 30. The colored lines indicate different values of A and the red dashed line shows A=A_{lin}=0.382. b) Possible values of P_b for varying A calculated using the joint PDF from a). The vertical dashed line shows A=A_{lin}=0.382. c) Obtained P_b for varying H_{m_0} and T_p and fixed A (0.382) and γ_{js} (10). The dashed blue lines and marker indicate the H_{m_0} and T_p values used in a) and b). Note that as in Figure 1, these results only consider dominant waves, that is, they were calculated from the spectrum between $0.7f_p$ and $1.3f_p$.

3 Field Data

Three historical datasets were used to evaluate the present model. Further, six historical models (detailed in Appendix A) were chosen to contextualize our model in relation to the state-of-the-art. These historical models range from baseline models in which the only inputs are known environmental parameters (wind speed in Melville and Matusov (2002) or wave steepness in Banner et al. (2000), for example) to fairly complex models that account for combinations of several phenomena (Romero (2019), for example).

3.1 Thomson (2012) and Schwendeman et al. (2014) dataset (TSG14)

The first data are from Thomson (2012) and Schwendeman et al. (2014), hereafter TSG14, and were collected in the Strait of Juan de Fuca, Washington. These data were collected by a gray scale video camera with a resolution of 640×480 pixels installed above the wheelhouse of Research Vessel R/V *Robertson* which recorded at an acquisition rate of 30 Hz (Schwendeman et al., 2014). These data were then projected into a metric coordinate grid with resolution of 0.25m (cross wave) and 0.075m (along wave) using the method proposed by Holland et al. (1997) and were then used to obtain $\Lambda(c)$ using the spectral approach of Thomson and Jessup (2009). The data were collected in a (usually) fetch-limited region and for a young sea state; note, however, that the particular sea-states analyzed here may not be fetch-limited. Figure 3-a shows the measured wave spectra, Figure 3-b shows $\Lambda(c)$ distributions, and Table 1 shows a summary of these data. For these data, P_b was calculated using the measured $\Lambda(c)$ distributions combined with the method described below in Equation 37. Additional information regarding the data collection is available from Thomson (2012) and Schwendeman et al. (2014).

3.2 Sutherland and Melville (2013) dataset (SM13)

The second dataset is from Sutherland and Melville (2013), hereafter SM13, and was collected using the Research Platform R/P *FLIP* during a two-day field campaign in the Southern California Bight under the scope of the SoCal 2010 experiment (Sutherland & Melville, 2013). Here, we focus only on the visible imagery collected by these authors to keep consistency with the previously presented data. Stereo video data were collected by a pair of video cameras mounted on the R/P FLIP for 10 minutes at the start of each hour and $\Lambda(c)$ was obtained using a variation of the method of Kleiss and Melville (2011), that is, tracking the temporal evolution of breakers obtained via pixel intensity threshold. Figure 3-c shows the measured wave spectra, Figure 3-d shows $\Lambda(c)$ distributions, and Table 1 shows a summary of these data. Note that because wave breaking was not observed for frequencies below $0.2Hz$ and from numerical simulations (not shown) these waves corresponded to a cross-swell not forced by the wind, our analyses only consider waves in the frequency range $0.2 < f < 0.8Hz$. Additional information regarding the data collection is available from Sutherland and Melville (2013). For these and TSG14 data, P_b was calculated using the measured $\Lambda(c)$ distributions combined with the formulas from Banner and Morison (2010):

$$P_b = \frac{\int_{c_0}^{c_1} c\Lambda(c)dc}{\int_{c_0}^{c_1} c\Pi(c)dc} \tag{37}$$

where $c_0 = \frac{g}{2\pi}\frac{1}{1.3f_p}$, $c_1 = \frac{g}{2\pi}\frac{1}{0.7f_p}$, $\Pi(c) = \chi g/(2\pi c^3)$ and $\chi = 0.6$. The implication of this choice is discussed in further detail in Section 5.

3.3 Banner, Babanin and Young (2000) dataset (B00)

The third dataset is from Banner et al. (2000), hereafter B00, and was collected in the Black Sea (BS), Lake Washington (LW) and the Southern Ocean (SO). These authors directly provide values for significant wave height H_{m0}, peak period (T_p) and the wave breaking probability in their Tables 1 (Black Sea, denoted as BS here) and 2 (Southern Ocean, denoted as SO here). The majority of the data were collected in the Black Sea (13 data runs) and two data runs are from the Southern Ocean. Given that the original spectral data were not published alongside their paper, we approximate the observed spectra using the provided pairs H_{m0}, T_p assuming a JONSWAP shape with $\gamma_{js} = 3.3$ (as previously done in Filipot et al. (2010), for example). Given that in this paper we are only interested in a very narrow spectral band, the differences between observed and simulated spectra should be minimal. For more details regarding this data refer to Banner et al. (2000).

Table 1. Data summary for the two experiments described in Sections 3.1 and 3.2. Note that the parameters obtained from wave spectra were computed specifically for the bands shown in Figure 3 for TSG14 and SM13 cases. The wave height (H_p) and wave steepness (ϵ) parameters for dominant waves were calculated as per Banner et al. (2002) (see Section A1 for details). The wave age parameter was calculated as c_p/u_* .

Dataset	Date	Length	H_{m_0}	T_p	H_p	ϵ	U_{10}	u_*	c_p	Wave age	P_b
	[−]	$[min]$	$[m]$	$[s]$	$[m]$	[−]	$[ms^{-1}]$	$[ms^{-1}]$	$[ms^{-1}]$	[−]	[−]
TSG14	14/02/2011 20:33	6.5	0.75	2.88	0.66	0.160	11.50	0.373	4.50	12.07	3.54E-03
TSG14	14/02/2011 20:58	5.1	0.75	2.96	0.66	0.152	12.55	0.417	4.62	11.08	9.57E-03
TSG14	14/02/2011 21:30	6.5	0.91	2.99	0.82	0.184	15.07	0.561	4.67	8.33	6.29E-02
TSG14	14/02/2011 21:44	8.5	1.09	3.17	1.00	0.200	15.73	0.599	4.94	8.25	1.01E-01
TSG14	14/02/2011 22:29	6	1.21	3.44	1.09	0.186	17.24	0.636	5.36	8.44	1.51E-01
TSG14	14/02/2011 22:37	4.8	1.37	3.53	1.24	0.199	18.01	0.660	5.52	8.36	7.61E-02
TSG14	15/02/2011 19:04	10	0.87	3.29	0.79	0.146	14.45	0.360	5.13	14.28	3.75E-03
TSG14	15/02/2011 19:19	6	0.90	3.31	0.81	0.149	13.11	0.477	5.17	10.85	4.05E-02
SM13	06/12/2010 21:59	10	0.61	3.51	0.52	0.085	6.46	0.205	5.48	26.68	7.96E-03
SM13	06/12/2010 23:00	10	0.61	3.33	0.54	0.097	7.55	0.342	5.20	15.22	1.95E-03
SM13	07/12/2010 00:00	10	0.73	3.45	0.66	0.112	8.62	0.319	5.38	16.85	3.24E-03
SM13	08/12/2010 00:00	10	0.34	2.04	0.23	0.110	5.24	0.160	3.19	19.96	1.65E-02
B00 (SO)	10/6/1992	5	9.20	13.46	8.02	0.089	19.80	0.835	21.01	25.17	2.70E-02
B00 (SO)	11/6/1992	9	4.20	12.04	3.66	0.051	16.00	0.626	18.78	30.02	0.00E+00
B00 (BS)	1993	34-68	0.39	2.78	0.34	0.089	11.70	0.414	4.34	10.49	3.80E-02
B00 (BS)	1993	34-68	0.49	2.94	0.43	0.100	12.70	0.461	4.59	9.96	6.50E-02
B00(BS)	1993	34-68	0.53	3.33	0.47	0.084	14.00	0.524	5.20	9.93	6.00E-02
B00 (BS)	1993	34-68	0.54	3.23	0.47	0.092	14.40	0.544	5.04	9.26	5.20E-02
B00 (BS)	1993	34-68	0.38	2.27	0.34	0.131	15.00	0.574	3.55	6.18	6.30E-02
B00 (BS)	1993	34-68	0.45	2.56	0.40	0.121	14.60	0.554	4.00	7.23	6.70E-02
B00 (BS)	1993	34-68	0.45	2.44	0.40	0.134	13.70	0.509	3.81	7.49	8.40E-02
B00 (BS)	1993	34-68	1.19	5.88	1.04	0.061	8.70	0.295	9.18	31.10	0.00E+00
B00 (BS)	1993	34-68	1.32	6.24	1.15	0.060	11.20	0.391	9.74	24.91	0.00E+00
B00 (BS)	1993	34-68	0.83	6.24	0.73	0.038	9.50	0.322	9.74	30.22	0.00E+00
B00 (BS)	1993	34-68	0.89	5.88	0.78	0.045	10.70	0.368	9.18	24.91	0.00E+00
B00 (BS)	1993	34-68	0.99	3.71	0.87	0.127	10.00	0.339	5.79	17.06	3.40E-02
B00 (BS)	1993	34-68	0.88	4.00	0.77	0.097	8.70	0.295	6.24	21.14	5.80E-02

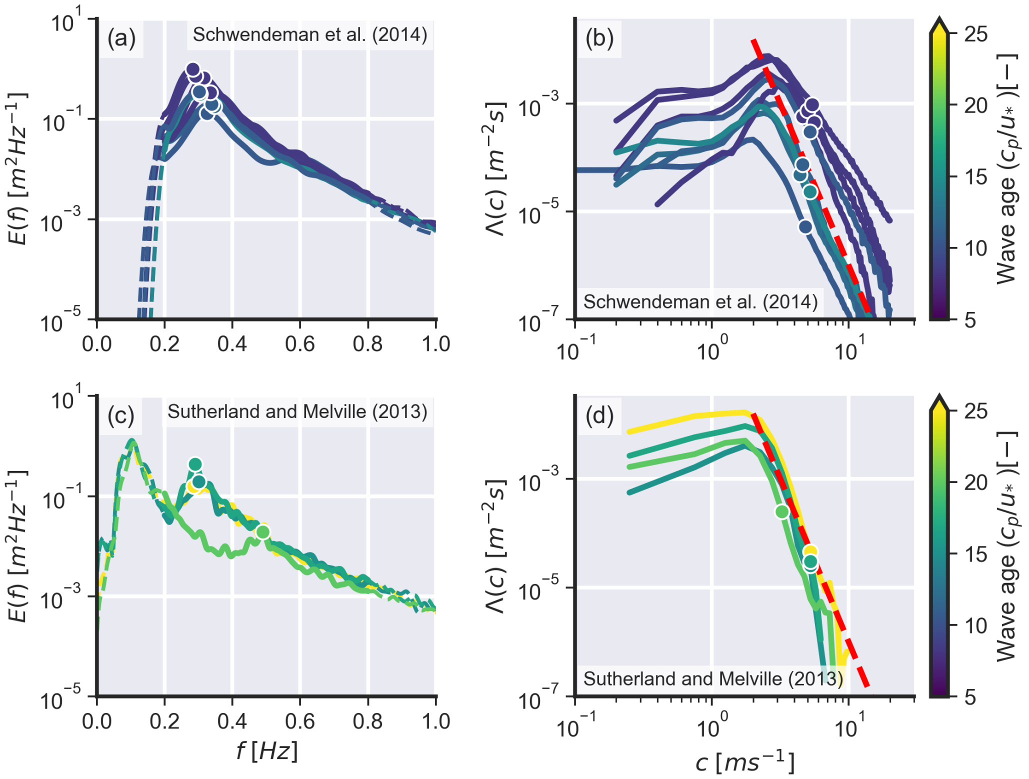

Figure 3. Field data. a) Spectral data from TSG14. b) $\Lambda(c)$ data from STG14. c) Spectral data from SM13. c) $\Lambda(c)$ data from SM13. The coloured circular markers show in a) and c) show the peak frequency (f_p) and the coloured circular markers show in b) and d) show the peak wave speed (c_p). The red dashed line in b) and d) shows the theoretical c^{-6} decay predicted by Phillips (1985). In all plots, the color scale shows the wave age (c_p/u_*).

4 Results

4.1 Comparison with Field Data

Figure 4 shows the comparison between estimated (or observed) (x-axis) and modelled (y-axis) values of P_b for each model. In general, no model was able to closely reproduce the trends seen in the combined observed data, regardless of the underlying mathematical or physical formalism. Furthermore, orders of magnitude of difference between the models and, more worryingly, between the models and the measured data were observed. In general, models based on a wave steepness-derived wave breaking criterion (Banner et al. (2000), Banner et al. (2002), for example) overestimated data derived from $\Lambda(c)$ while models based on $\Lambda(c)$ (Melville and Matusov (2002) and Sutherland and Melville (2013), for example) underestimated P_b data that was not derived from Λ (that is, B00 data). The model from Filipot et al. (2010) was found to be the most consistent model. From Figure 4-g, the formulation presented in Section 2 with $A = A_{lin} = 0.382$ underestimated the observed P_b for B00 and SM13 data (note that P_b was too low to be displayed on the plot) but performed relatively well for the majority of TSG14 data. Using the mean absolute error (MAE) as a convenient metric to assess the models, it was found that the present model has errors in the same order of magnitude as the previous models. Given the spread in the results seen in Figure 4, no model could be considered a clear winner. For the discussion of these results, see Section 5.

4.2 Model Optimization

From the analysis of Figure 2, minor changes in A can lead to major variations in P_b. Further, from the analysis of Figure 4, the proposed model underestimated P_b for $A = A_{lin} = 0.382$ particularly for S13 and B00 data. Given that it is a common practice to optimize wave breaking models for particular datasets, we present two methods to do so using TSG14 data as an example. The same could be done for B00 and SM13 data but, for brevity, this is not done here. Given that the present model is not computationally expensive, the first approach consisted of varying A from 0.1 to 0.5 in 0.001 intervals and finding the value of A that resulted in the lowest squared error ($\sqrt{\left(p^d_{b_i} - p^m_{b_i}\right)^2}$, where the superscripts d and m indicate observed and modelled data, respectively) for each data run. Figure 5-a shows the results of this procedure. The value $A = A_{opt} = 0.24$ was, on average, the optimal values of for this particular dataset. The second approach

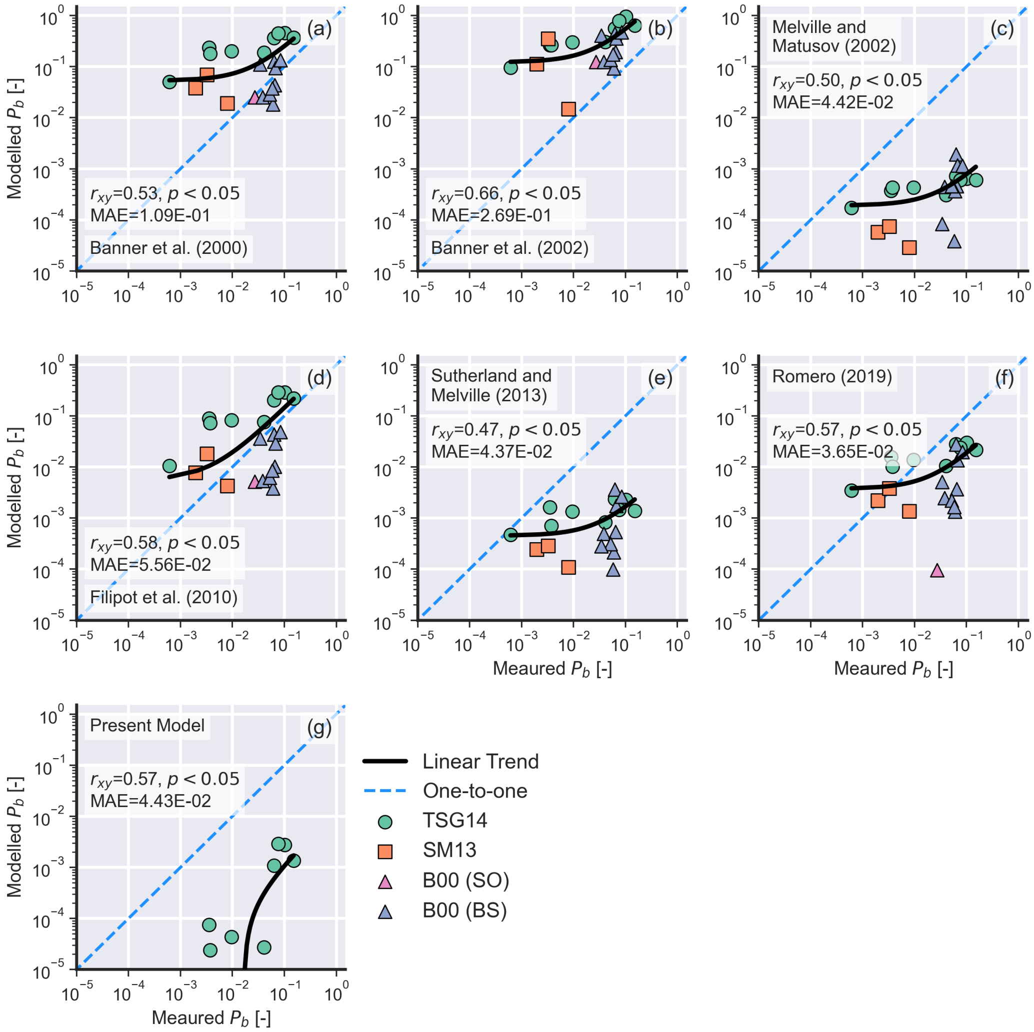

Figure 4. Compassion between measured and computed P_b for different models and data. a) Banner et al. (2000), b) Banner et al. (2002), c) Melville and Matusov (2002), d) Filipot et al. (2010), e) Sutherland and Melville (2013), f) Romero (2019), and g) present model with A=A_{lin}=0.382. The thick black line shows the linear regression between measured and modelled P_b and the blue dashed line indicates the one-to-one correspondence in all panels. Data points with modelled P_b < 10^{-5} or observed P_b = 0 are not shown in this plot. In all plots, r_{xy} is Pearson's correlation coefficient and MAE indicates the mean absolute error. Note the logarithmic scale.

consisted in parameterizing the optimal value of A for each data run as a function of a known environmental variable, in this example, the waveage c_p/u_* (Figure 5-b). The results of these two approaches are show in Figures 5-c and d, respectively. Both approaches considerably improved the model results from the baseline model presented in Figure 4, with the parametric model (Figure 5-d) performing slightly better when considering Pearson's correlation coefficient (r_{xy}) as a comparison metric.

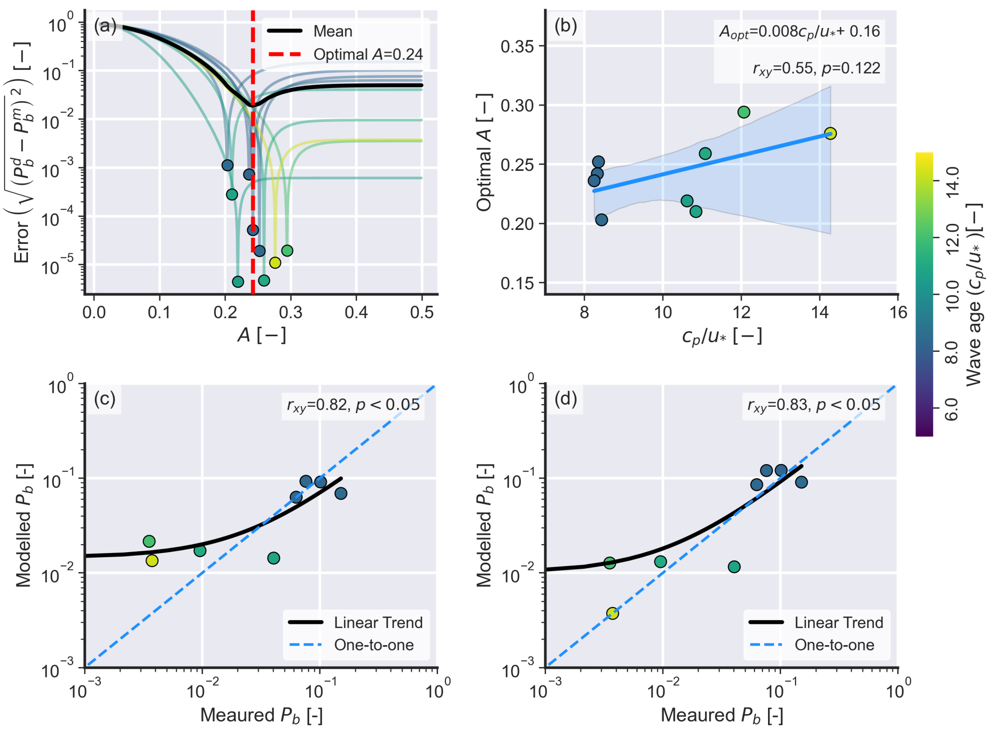

Figure 5. Results of the optimization procedures. a) Optimization curves for each data record (coloured lines) and the global averaged (black line). The vertical dashed line show s$A = A_{opt}$ $= 0.24$. b) Parametrization of A as a function of c_p/u_*. The blue swath indicates the 95% confidence interval. For this particular case, $A = 0.008c_p/u_*$ + 0.16. Note the logarithmic scale in a), c) and d). In all plots, the color scale shows the wave age (c_p/u_*). In b) to d) r_{xy} is Pearson's correlation coefficient.

5 Discussion

We have introduced a new model for obtaining the probability of wave breaking (P_b) for dominant waves based on the theoretical joint probability density distribution between wave phase speed (c) and horizontal orbital velocity at the wave crest (u) for unidirectional Gaussian wave fields. The present model has only one parameter for defining the wave breaking threshold (A), which makes it relatively easy to optimize for a given dataset (as shown in Section 4.2). While the proposed model performed relatively well for one of the investigated datasets (TSG14), it greatly underestimated P_b for the two other datasets (SM13 and B00). For the data investigated here, such underestimation did not result in a high mean absolute error (MAE) and, in fact, our model had one of the lowest MAE. Recent results of Barthelemy et al. (2018), Derakhti et al. (2020) and Varing et al. (2020) showed that waves with horizontal fluid velocity that exceeds 0.85 times the phase velocity will inevitably break. These results suggest that the breaking threshold derived from Cokelet (1977) in Section 2.3 could be reduced by ≈15%. If we apply their findings to our case, we obtain $A = 0.382 \times 0.85 = 0.324$ which would help to reduce the underestimation of P_b, but not significantly. It is more probable that other environmental phenomena such as direct wind forcing, directional spreading and long wave modulation, which are not accounted in our model, are the reason for such differences.

One of the most challenging aspects when assessing our model is, nevertheless, regarding the field data. The attribution of wave breaking occurrences to wave scales using timeseries analysis, as done in Banner et al. (2000) or Filipot et al. (2010), is difficult because several wave scales can be present at the same time and space. This lead us to use $\Lambda(c)$ observations as well as data from Banner et al. (2000) to investigate our model. Different interpretations of how $\Lambda(c)dc$ is computed from field data can, however, generate orders of magnitude of difference in its moments (Gemmrich et al., 2013; Banner et al., 2014) and, consequently, in P_b. Next, it is difficult to relate the speed of the wave breaking front to the phase speed of the carrying wave because small, slower breaking waves could merely be traveling on top of longer, much faster waves. In particular, we believe that these wave breaking events can significantly contribute to the observed $\Lambda(c)dc$ distribution as they would have c close to the peak wave phase speed. This wave breaking "sub-population" has not receive much research interest because of its apparent small contribution to energy dissipation but, for our particular case, they directly impact model validation.

Further, relating $\Lambda(c)$ to P_b is also challenging. Here, we adopted the convenient formula from Banner and Morison (2010). While this formula has some support from the literature (Ardhuin et al., 2010), the actual functional form of $\Pi(c)$ and the value for the constant χ (see Equation 37) are unknown and changes in these will lead to changes in P_b. The Gaussian framework developed in Section 2.1 provides an alternative method to obtain $\Pi(c)$ (from Equation 3, for example) but this is beyond the scope of this introductory paper and will be the focus of a future publication.

Finally, we would like to re-emphasize that our model is derived in the space domain whereas P_b data is (at least partially) obtained in the time domain. For the narrow spectral band investigated here, Monte-Carlo simulations of linear waves indicate that the difference between P_b modelled in space is less than five percent from P_b modelled in time (not shown). Given all these complications and the fact that some historical models are being compared to data that was used to create them (Banner et al. (2000) and Sutherland and Melville (2013), for example), we are unable to provide an accurate ranking of the existing models. Future research should focus, therefore, on obtaining P_b data that is unambiguous and widely available. In this regard, and despite its own limitations, wave tank experiments could bring further insight on the statistics of dominant (or not) breaking waves. Such a dataset would ultimately allow researchers to focus on models derived from physical and mathematical concepts (such as ours) rather than on empirical concepts.

6 Conclusion

We have presented a new statistical wave breaking model derived from Gaussian field theory that we have applied to obtain the probability of wave breaking for dominant, wind-sea waves. Although more mathematically complex than previous formulations, the present model relies on the ratio between the crest orbital velocity and the phase speed and uses only on a single free parameter, the wave breaking threshold A. Using theoretical results obtained by Cokelet (1977) for regular nearly breaking waves, we derived a wave breaking threshold to adapt our linear model to non-linear waves. The present model has errors in the same order of magnitude as six other historical models when assessed using three field datasets. For a particular dataset (TSG14), our model performed well, especially if the free-parameter A is fine tuned. Additional observations are however required, to further understanding and quantifying the dependence of A on environmental parameters that are not accounted for in our model (for example, wind forcing, wave directionality or modulation by long waves). Future research should be dedicated to collect more wave breaking observations in different and repeatable environmental conditions to provide reliable constraints for the optimization of the present and other wave breaking models. Still and although the research presented here is in early stages, the present model should be extendable to waves of any scale and, therefore, has the potential to be implemented in current state-of-the-art spectral wave models as a new wave breaking dissipation source term with relatively little effort.

Appendix A Historic Parametric Wave Breaking Models

A1 Banner et al. (2000)

Banner et al.'s (2000) is a popular model for calculating wave breaking probabilities for deep water, dominant waves. This model follows from observations and results from Donelan et al. (1972), Holthuijsen and Herbers (1986) and Banner and Tian (1998) who demonstrated the importance of the wave group modulation on the wave breaking onset. These authors conveniently obtained a parameterization for the probability of wave breaking (P_b) based solely on the spectral steepness of the dominant wave scale (ϵ_p), assuming that their formulas would capture the influence of the wave group modulation on the wave breaking onset. Their formulation was derived using a dataset of measurements collected in various environments ranging from lakes to open ocean conditions (Banner et al., 2000). From these observations, these authors were then able to obtain a wave breaking threshold behaviour for the dominant waves as a function of the dominant spectral wave steepness given by:

$$\epsilon_p = \frac{H_p k_p}{2} \tag{A1}$$

in which k_p is the wavenumber at peak frequency (f_p) and H_p is the significant wave height of the dominant waves calculated as:

$$H_p = 4\sqrt{\left(\int_{0.7f_p}^{1.3f_p} E(f)df\right)} \tag{A2}$$

where $E(f)$ is the spectra of wave heights as a function of frequency. For their data, P_b was then parameterized as a single equation with three free parameters (p_1, p_2, p_3):

$$P_b = p_1 + (\epsilon_p - p_2)^{p_3}, \tag{A3}$$

For the available field data, Banner et al. (2000) found optimal values of $p_1 = 22$, $p_2 = 0.055$, and $p_3 = 2.01$. Note that hereafter free parameters for the different models will be denoted as p_n where n is a sequential number.

A2 Banner et al. (2002)

This work extended Banner et al. (2000) model to shorter wave scales (up to 2.48 times the peak wave frequency). From field data Banner et al. (2002) reported that the waves were breaking if the saturation spectrum $\sigma(f) = 2\pi^4 f^5 E(f)/2g^2 = \sigma(k) = k^4 E(k)$ exceeded a threshold that was frequency dependent. These author's related this dependence to the directional spreading $\overline{\theta(k)}$ which later led Banner and Morison (2010) to explicitly define the following empirical formulation:

$$P_b(k_c) = \mathcal{H}_h(\tilde{\sigma}(k_c) - p_1) \times p_2 \times (\tilde{\sigma}(k_c) - \tilde{\sigma}_t), \tag{A4}$$

in which $\mathcal{H}_h$ is the Heaviside step function, k_c is the central wavenumber for a given wavenumber range, $\tilde{\sigma}(k_c) = \sigma(k_c)/\overline{\theta(k_c)}$ is the saturation spectrum normalized by the averaged directional spreading, $p_1 = 0.0045$ and $p_2 = 33$ are constants obtained from their observations. Following Banner et al. (2002), the directional spreading angle is calculated according to Hwang et al. (2000) (their equation 19a):

$$\theta\left(\frac{k}{k_p}\right) = \begin{cases} 0.35 + 1.05\left(1 - \frac{k}{k_p}\right) & \text{if } \frac{k}{k_p} < 1.05 \\ 0.30 + 0.087\left(\frac{k}{k_p} - 1\right) & \text{if } 1.05 \le \frac{k}{k_p} < 5 \end{cases} \tag{A5}$$

where θ is the directional spreading angle as a function of the wavenumber.

A3 Filipot et al. (2010)

This method follows from the original works of Le Méhauté (1962), Battjes and Janssen (1978) and Thornton and Guza (1983) and assumes that the probability distribution function (PDF) of breaking wave heights in the dominant wave scale is parameterized by its central frequency f_c or, equivalently, by its representative phase speed $c(f_c)$ and the product between a Rayleigh PDF for the wave heights

$$P(H, f_c) = \frac{2H}{H_{rms}^2(f_c)} \exp\left[-\left(\frac{H}{H_{rms}(f_c)}\right)^2\right] \tag{A6}$$

in which

$$H_r(f_c) = \frac{4}{\sqrt{2}}\sqrt{\int_0^\infty U_{fc}(f)E(f)df} \tag{A7}$$

and

$$U_{f_c} = 0.5 - 0.5\cos\left(\frac{\pi}{\delta}\left[\frac{f}{f_c} - 1 - \delta\right]\right) \tag{A8}$$

where δ is the bandwidth of a Hann window (in this study, $\delta = 0.6$), and a weighting function

$$W(H, f_c) = p_1\left[\frac{\beta_r}{\beta}\right]^2\left\{1 - \exp\left[-\left(\frac{\beta}{\tilde{\beta}}\right)^{p_2}\right]\right\} \tag{A9}$$

in which $\beta = kH/\tanh(kh)$, and p_1 and p_2 are free parameters. In order to extend the formulation outside the shallow water domain, these authors replaced Thornton and Guza's (1983) breaking criterion based on the wave height (H) to water depth (h) ratio ($\gamma = H/h = 0.42$) with an adaptation of Miche's (1944) wave breaking parameter:

$$\beta_r = \frac{\overline{k_r}(f_c)H_r(f_c)}{\tanh\left(\overline{k_r}(f_c)h\right)} \tag{A10}$$

in which

$$\overline{k_r}(f_c) = \frac{\int_0^\infty U_{f_c}(f)k(f)E(f)df}{\int_0^\infty U_{f_c}(f)E(f)df} \tag{A11}$$

and

$$\tilde{\beta} = b(b_3\tanh(kh)^3 - b_2\tanh(kh)^2 + b_1\tanh(kh) - b_0) \tag{A12}$$

in which $b = 0.48$, $b_3 = 1.0314$, $b_2 = 1.9958$, $b_1 = 1.5522$, and $b_0 = 0.1885$. In their model, the variable $\tilde{\beta}$ was obtained via numerical calculations of regular nearly breaking waves using the stream wave theory of Dean (1965). Finally, the wave breaking probability is obtained as:

$$P_b(f_c) = \int_0^\infty P(H, f_c)W(H, f_c)dH \leq 1. \tag{A13}$$

To keep consistency with Section A2, P_b will be only considered at the spectral peak; other definitions are, however, also possible.

A4 Models based on Phillips' (1985) $\Lambda(c)$

The major issue with the previous models is the difficulty to obtain reliable observations of the wave breaking probabilities as a spectral distribution solely from point measurements. Due to the presence of different wave scales at the time and location, it is indeed difficult to assign the breaking occurrence to a given wave frequency of wave number. To avoid this problem, Phillips (1985) proposed to use the speed of the breaking front as a proxy for the phase speed of the carrying wave. Phillips (1985) defined the parameter $\Lambda(c)dc$ as the "average total length per unit surface area of breaking fronts that have velocities in the range c to $c + dc$" and then defined the following quantities:

$$L = \int \Lambda(c)dc \tag{A14}$$

and

$$R = \int c\Lambda(c)dc \tag{A15}$$

which represent the "total length of breaking fronts per unit area" (Equation A14) and "the total number of breaking waves of all scales passing a given point per unit time" (Equation A15). Assuming that Phillips (1985) assumptions hold, it is possible to obtain parametric models for Λ from known variables (e.g., wind speed) and, consequently, for P_b (see Equation 37).

A4.1 Melville and Matusov (2002)

Melville and Matusov's (2002) model for $\Lambda(c)$ relies only on the wind speed measured at 10m (U_{10}) to obtain $\Lambda(c)$. Following Melville and Matusov (2002) and using the explicit formula given by Reul and Chapron (2003), this parameterization is written as:

$$\Lambda(c) = p_1 \left[\frac{U_{10}}{10}\right]^3 10^{-4} \exp\left[-(p_2 c)\right] \tag{A16}$$

in which p_1 and p_2 are constants. For their data, Melville and Matusov (2002) found $p_1 = 3.3$ and $p_2 = 0.64$. As discussed by Reul and Chapron (2003), this formulation approaches Phillips's (1985) theoretical c^{-6} but may overly estimates the amount of small breakers.

A42 Sutherland and Melville (2013)

Sutherland and Melville (2013) used dimensional analysis to scale $\Lambda(c)$ and obtain a parameterization that is a function of the wind drag (u_*), peak wave phase speed (c_p), significant wave height (H_s) and three constants. From Sutherland and Melville's (2013) Equation 9 and their Figure 4, $\Lambda(c)$ is calculated as:

$$\Lambda(c) = p_1 \frac{g}{c_p^3}\left(\frac{u_*}{c_p}\right)^{p_2}\left(\frac{c}{\sqrt{gH_s}}\left(\frac{gH_s}{c_p^2}\right)^{p_3}\right)^{-6} \tag{A17}$$

where $p_1 = 0.05$, $p_2 = 0.5$, and $p_3 = 0.1$ are constants obtained from the available data. Their formulation reproduces Phillips's (1985) c^{-6} frequency dependency but does not have the typical roll-off at low c as these authors chose to use infrared (other than visible) imagery to obtain and model their $\Lambda(c)$. This choice included the contribution of micro-scale breakers that do generate visible bubbles in their model, hence the difference.

A43 Romero (2019)

Recently, Romero (2019) developed and implemented a new wave breaking parameterization in WaveWatchIII which relies exclusively on $\Lambda(c)$. Differently from previous parameterizations, Romero's (2019) takes into account both the modulations due to winds and long waves on $\Lambda(c)$. His model is fairly general but depends on six free parameters that needed to be laboriously obtained by comparing WaveWatchIII's significant wave height outputs with available measured significant wave heights from buoy data. In Romero's (2019) model, Λ was modelled assuming that it is proportional to the crest lengths exceeding a slope threshold:

$$\Lambda(f,\theta) = \left(\frac{2(2\pi)^2 p_1}{g}\right) f \exp\left[-\left(\frac{p_2}{B(f,\theta)}\right)\right] M_{LW} M_W \tag{A18}$$

where p_1 = 3.5x10^{-5} and p_2 = 5x10^{-3} are constants to be obtained from the data, M_{LW} is the modulation due to long waves, M_W is the modulation due to winds and $B(f,\theta)$ is the directional wave breaking saturation spectra:

$$B(f) = \int_0^{2\pi} B(f,\theta)d\theta = E(f)\left(\frac{2\pi f^5}{2g}\right). \tag{A19}$$

The modulation due to long waves is calculated according to Guimarães (2018):

$$M_{LW} = \left[1 + p_3\sqrt{\mathrm{cmss}\,(E(f))}\cos^2(\theta - \hat{\theta})\right]^{p_4} \tag{A20}$$

where p_3 = 400 and p_4 = 3/2 are also best-fit constants found by Romero (2019). The cumulative mean square slope (cmss) is defined as:

$$\mathrm{cmss} = \int_0^{\infty} E(f)\left(\frac{(2\pi)^4 f^4}{g^2}\right) df. \tag{A21}$$

and

$$\hat{\theta} = \tan\left(\frac{\int E(f,\theta)\sin(\theta)df d\theta}{\int E(f,\theta)\cos(\theta)df d\theta}\right) \tag{A22}$$

The modulation due to the wind is computed as:

$$M_W = \frac{\left(1 + p_5 \max\left(1, \frac{f}{f_0}\right)\right)}{(1 + p_5)} \tag{A23}$$

with

$$f_0 = p_6 \frac{1}{u_*}\frac{g}{2\pi} \tag{A24}$$

where p_5 = 0.9 is a constant related to the DIA algorithm and p_6 = 3/28 is yet another constant. Finally, the conversion from $\Lambda(f)$ to $\Lambda(c)$ is done using the relation $\Lambda(c)dc$ = $\Lambda(f)df$ and the linear dispersion relation (see Romero's (2019) Eqs. 17-23 for details).

Acknowledgments

This work benefited from France Energies Marines and State financing managed by the National Research Agency under the Investments for the Future program bearing the reference numbers ANR-10-IED-0006-14 and ANR-10-IEED-0006-26 for the projects DiME and CARAVELE. The authors' thank Peter Sutherland and Jim Thompson for kindly sharing their data.

Data Availability

All data used in this publication has been previously published by Banner et al. (2000), Sutherland and Melville (2013), Schwendeman et al. (2014).

References

Alsina, J. M., & Baldock, T. E. (2007). Improved representation of breaking wave energy dissipation in parametric wave transformation models. *Coastal Engineering*, *54*, 765–769. doi: 10.1016/j.coastaleng.2007.05.005

Ardag, D., & Resio, D. T. (2020). A new approach for modeling dissipation due to breaking in wind wave spectra. *Journal of Physical Oceanography*, *50*(2), 439–454.

Ardhuin, F., Rogers, E., Babanin, A. V., Filipot, J. F., Magne, R., Roland, A., ... Collard, F. (2010). Semiempirical dissipation source functions for ocean waves. Part I: Definition, calibration, and validation. *Journal of Physical Oceanography*, *40*(9), 1917–1941. doi: 10.1175/2010JPO4324.1

Banner, M. L., Babanin, A. V., & Young, I. R. (2000). Breaking Probability for Dominant Waves on the Sea Surface. *Journal of Physical Oceanography*, *30*(12), 3145–3160. doi: 10.1175/1520-0485(2000)030⟨3145:BPFDWO⟩2.0.CO;2

Banner, M. L., Gemmrich, J. R., & Farmer, D. M. (2002). Multiscale measurements of ocean wave breaking probability. *Journal of Physical Oceanography*, *32*(12), 3364–3375. doi: 10.1175/1520-0485(2002)032⟨3364:MMOOWB⟩2.0.CO;2

Banner, M. L., & Morison, R. P. (2010). Refined source terms in wind wave models with explicit wave breaking prediction. Part I: Model framework and validation against field data. *Ocean Modelling*, *33*(1-2), 177–189. Retrieved from `http://dx.doi.org/10.1016/j.ocemod.2010.01.002` doi:

10.1016/j.ocemod.2010.01.002

Banner, M. L., & Tian, X. (1998). On the determination of the onset of breaking for modulating surface gravity water waves. *Journal of Fluid Mechanics*, *367*, 107–137.

Banner, M. L., Zappa, C. J., & Gemmrich, J. R. (2014). A note on the Phillips spectral framework for ocean whitecaps. *Journal of Physical Oceanography*, *44*(7), 1727–1734. doi: 10.1175/JPO-D-13-0126.1

Barthelemy, X., Banner, M. L., Peirson, W. L., Fedele, F., Allis, M., & Dias, F. (2018). On a unified breaking onset threshold for gravity waves in deep and intermediate depth water. *Journal of Fluid Mechanics*, *841*, 463–488. doi: 10.1017/jfm.2018.93

Battjes, J. A., & Janssen, J. (1978). Energy loss and set-up due to breaking of random waves. *Coastal Engineering*, *32*(1), 569–587.

Chawla, A., & Kirby, J. T. (2002). Monochromatic and random wave breaking at blocking points. *Journal of Geophysical Research: Oceans*, *107*(C7), 4–1.

Cokelet, E. D. (1977). Steep Gravity Waves in Water of Arbitrary Uniform Depth. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *286*(1335), 183–230. Retrieved from `http://rsta.royalsocietypublishing.org/cgi/doi/10.1098/rsta.1977.0113` doi: 10.1098/rsta.1977.0113

Dean, R. G. (1965). Stream Function Representation of Nonlinean Ocean Waves. *Journal of Geophysical Research*, *70*(18), 4561–4572.

Derakhti, M., Kirby, J. T., Banner, M. L., Grilli, S. T., & Thomson, J. (2020). A unified breaking onset criterion for surface gravity water waves in arbitrary depth. *Journal of Geophysical Research: Oceans*(2013), 1–28. doi: 10.1029/2019jc015886

Donelan, M., Longuet-Higgins, M., & Turner, J. (1972). Periodicity in whitecaps. *Nature*, *239*(5373), 449–451.

Duncan, J. H. (1981). An Experimental Investigation of Breaking Waves Produced by a Towed Hydrofoil. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *377*(1770), 331–348. Retrieved from `http://rspa.royalsocietypublishing.org/cgi/doi/10.1098/rspa.1981.0127` doi: 10.1098/rspa.1981.0127

Eldeberky, Y., & Battjes, J. A. (1996). Spectral modeling of wave breaking: Application to Bousinesq equations. *Journal of Geophysical Research*, *101*, 1253–1264.

Filipot, J. F., & Ardhuin, F. (2012). A unified spectral parameterization for wave breaking: From the deep ocean to the surf zone. *Journal of Geophysical Research: Oceans*, *117*(4), 1–19. doi: 10.1029/2011JC007784

Filipot, J. F., Ardhuin, F., & Babanin, A. V. (2010). A unified deep-to-shallow water wave-breaking probability parameterization. *Journal of Geophysical Research: Oceans*, *115*(4), 1–15. doi: 10.1029/2009JC005448

Filipot, J.-F., Guimaraes, P., Leckler, F., Hortsmann, J., Carrasco, R., Leroy, E., ... Le Dantec, N. (2019). La Jument lighthouse: a real-scale laboratory for the study of giant waves and their loading on marine structures. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *377*(2155), 20190008. doi: 10.1098/rsta.2019.0008

Gemmrich, J., Zappa, C. J., Banner, M. L., & Morison, R. P. (2013). Wave breaking in developing and mature seas. *Journal of Geophysical Research: Oceans*, *118*(9), 4542–4552. doi: 10.1002/jgrc.20334

Guimarães, P. V. (2018). *Sea surface and energy dissipation* (Unpublished doctoral dissertation). Universitè de Bretagne Loire.

Holland, K. T., Holman, R. A., Lippmann, T. C., Stanley, J., Member, A., & Plant, N. (1997). Practical Use of Video Imagery in Nearshore Oceanographic Field Studies. *IEEE Journal of Oceanic Engineering*, *22*(1), 81–92.

Holthuijsen, L., & Herbers, T. (1986). Statistics of breaking waves observed as whitecaps in the open sea. *Journal of Physical Oceanography*, *16*(2), 290–297.

Hwang, P. A., Wang, D. W., Walsh, E. J., Krabill, W. B., & Swift, R. N. (2000). Airborne Measurements of the Wavenumber Spectra of Ocean Surface Waves. Part II: Directional Distribution. *Journal of Physical Oceanography*, *30*(11), 2768–2787. doi: 10.1175/1520-0485(2001)031⟨2768:amotws⟩2.0.co;2

Janssen, T. T., & Battjes, J. A. (2007). A note on wave energy dissipation over steep beaches. *Coastal Engineering*, *54*(9), 711–716. doi: 10.1016/j.coastaleng.2007.05.006

Kjeldsen, S. P., Vinje, T. P., Myrhaug, D. P., & Brdvig, P. P. (1980). Kinematics of deep water breaking waves. In *Offshore technology conference.*

Kleiss, J. M., & Melville, W. K. (2011). The analysis of sea surface imagery for whitecap kinematics. *Journal of Atmospheric and Oceanic Technology*, *28*(2), 219–243. doi: 10.1175/2010JTECHO744.1

Kudryavtsev, V., Chapron, B., & Makin, V. (2014). Impact of wind waves on the air-sea fluxes: A coupled model. *Journal of Geophysical Research: Oceans*, *119*(2), 1217–1236.

Le Méhauté, B. (1962). On non-saturated breakers and the wave run-up. *Proceedings of the 8th International Conference on Coastal Engineering*(Figure 1), 77–92. Retrieved from `http://journals.tdl.org/icce/index.php/icce/article/viewArticle/2255`

Longuet-Higgins, M. S. (1957). The Statistical Analysis of a Random, Moving Surface. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *249*(966), 321–387. doi: 10.1098/rsta.1957.0002

Melville, W. K., & Matusov, P. (2002). Distribution of breaking waves at the ocean surface. *Nature*, *417*(6884), 58–63. doi: 10.1038/417058a

Miche, A. (1944). Mouvements ondulatoires de la mer en profondeur croissante ou décroissante. Première partie. Mouvements ondulatoires périodiques et cylindriques en profondeur constante. *Annales des Ponts et Chaussées*, *Tome 114*, 42–78.

Perlin, M., Choi, W., & Tian, Z. (2013). Breaking Waves in Deep and Intermediate Waters. *Annual Review of Fluid Mechanics*, *45*(1), 115–145. doi: 10.1146/annurev-fluid-011212-140721

Phillips, O. M. (1985). Spectral and statistical properties of the equilibrium range in wind-generated gravity waves. *Journal of Fluid Mechanics*, *156*, 505–531. doi: 10.1017/S0022112085002221

Reul, N., & Chapron, B. (2003). A model of sea-foam thickness distribution for passive microwave remote sensing applications. *Journal of Geophysical Research C: Oceans*, *108*(10), 19–1. doi: 10.1029/2003jc001887

Rice, S. O. (1944). Mathematical Analysis of Random Noise. *The Bell System Technical Journal*, *23*(3), 282 – 332. doi: https://doi.org/10.1002/j.1538-7305.1944.tb00874.x

Romero, L. (2019). Distribution of Surface Wave Breaking Fronts. *Geophysical Research Letters*, *46*(17-18), 10463–10474. doi: 10.1029/2019GL083408

Saket, A., Peirson, W. L., Banner, M. L., Barthelemy, X., & Allis, M. J. (2017). On the threshold for wave breaking of two-dimensional deep water wave groups in the absence and presence of wind. *Journal of Fluid Mechanics*, *811*, 642.

Schwendeman, M., Thomson, J., & Gemmrich, J. R. (2014). Wave breaking dissipation in a Young Wind Sea. *Journal of Physical Oceanography*, *44*(1), 104–127. doi: 10.1175/JPO-D-12-0237.1

Sutherland, P., & Melville, W. K. (2013). Field measurements and scaling of ocean surface wave-breaking statistics. *Geophysical Research Letters*, *40*(12), 3074–3079. doi: 10.1002/grl.50584

Thomson, J. (2012). Wave breaking dissipation observed with "swift" drifters. *Journal of Atmospheric and Oceanic Technology*, *29*(12), 1866–1882. doi: 10.1175/JTECH-D-12-00018.1

Thomson, J., & Jessup, A. T. (2009). A fourier-based method for the distribution of breaking crests from video observations. *Journal of Atmospheric and Oceanic Technology*, *26*(8), 1663–1671. doi: 10.1175/2009JTECHO622.1

Thornton, E. B., & Guza, R. T. (1983). Transformation of Wave Height Distribution. *Journal of Geophysical Research*, *88*(C10), 5925–5938.

Varing, A., Filipot, J.-f., Grilli, S., Duarte, R., Roeber, V., & Yates, M. (2020). A new kinematic breaking onset criterion for spilling and plunging breaking waves in shallow water. *Coastal Engineering*, 1–24.

Zieger, S., Babanin, A. V., Erick Rogers, W., & Young, I. R. (2015). Observation-based source terms in the third-generation wave model WAVEWATCH. *Ocean Modelling*, *96*(Vic), 2–25. Retrieved from `http://dx.doi.org/10.1016/j.ocemod.2015.07.014` doi: 10.1016/j.ocemod.2015.07.014