



OPEN

Generation mechanism and prediction of an observed extreme rogue wave

Johannes Gemmrich^{1✉} & Leah Cicon²

Rogue waves are individual ocean surface waves with crest height η or trough-to-crest height H that are large compared to the significant wave height H_s of the underlying sea state: $H/H_s > 2.2$ or $\eta/H_s > 1.25$. The physics of rogue wave generation and the potential of predicting the rogue wave risk are open questions. Only a few rogue waves in high sea states have been observed directly, but they can pose a danger to marine operations, onshore and offshore structures, and beachgoers. Here we report on a 17.6m high rogue wave in coastal waters with $\eta/H_s = 1.98$ and $H/H_s = 2.9$ which are likely the largest normalized heights ever recorded. Simulations of random superposition of Stokes waves in intermediate water depth show good agreement with the observation. Non-linear wave modulational instability, a well known cause for rogue waves in laboratory settings, did not contribute significantly to the rogue wave generation. A parameter obtained from a routine spectral wave forecast provides a practical risk prediction for rogue waves. These results confirm that probabilistic prediction of oceanic rogue waves based on random superposition of steep waves are possible and should replace predictions based on modulational instability.

The observation of the “Draupner” wave ($\eta/H_s = 1.55$, $H/H_s = 2.15$, $H_s = 11.9$ m) on 1 January 1995 proved that the existence of so-called rogue or freak waves are not a seafarer’s folklore¹. Technically, rogue waves are waves in the tail of the wave height or crest height probability distribution, $H/H_s > 2.2$ or $\eta/H_s > 1.25$ where H_s is the average of the highest one-third waves². Thus, rogue waves occur in any background sea state. Not surprisingly, rogue waves in high sea states receive the most attention. In particular, the “Draupner” wave, the “Andrea” wave ($\eta/H_s = 1.62$, $H/H_s = 2.30$, $H_s = 9.2$ m), and the “Killard” wave ($\eta/H_s = 1.57$, $H/H_s = 2.29$, $H_s = 11.4$ m) are well studied^{3–7}. Of main interest is the probability of encountering such a rogue wave, and which physical processes, if any, can alter the underlying probability density function for wave or crest heights^{2,8–10}. In linear theory, the sea is a random superposition of linear waves of a narrow frequency spectrum and the individual wave heights H follow the Rayleigh distribution¹¹. The probability of a wave or crest exceeding the significant wave height by a multiple z is

$$P(H/H_s > z) = e^{-2z^2}, \text{ and } P(\eta/H_s > z) = e^{-8z^2} \quad (1)$$

for wave heights and crest heights, respectively. Thus, a rogue wave with $H/H_s > 2.2$ is expected with a probability 1/16,000, or since the dominant wave period in the ocean is $O(10$ s) a rogue wave is expected about every two days. Allowing for finite amplitudes, wave crests become taller and troughs shallower, due to the presence of bound higher harmonics¹². At second order this Stokes correction affects crest height distributions, and exceedance probabilities corrected for second order have become the standard model for estimating extreme crest height probabilities^{13–15}. For typical wave steepness the probability of rogue crests increases by a factor 5 compared to small-amplitude linear theory. For narrow spectral bandwidth, wave height distributions are not affected by second order corrections. Observations from the GORM platform in the North Sea showed that the distribution of moderate wave heights $H/H_s \lesssim 2$ is consistent with second order models accounting for finite spectral bandwidth, but extreme waves $H/H_s \gg 2$ seem to follow a different distribution⁹. Most observed wave height distributions are based on relatively short records and the extreme values are limited to moderate normalized wave heights where the distributions are consistent with second order models¹⁶. Using synthesized surface elevation it was shown that the increased probabilities of extreme waves, as observed in the GORM data set, cannot be explained by second order alone, but is consistent with fourth order Stokes theory of bound waves⁸.

¹Physics and Astronomy, University of Victoria, Victoria, BC, Canada. ²School of Earth and Ocean Sciences, University of Victoria, Victoria, BC, Canada. ✉email: gemmrich@uvic.ca

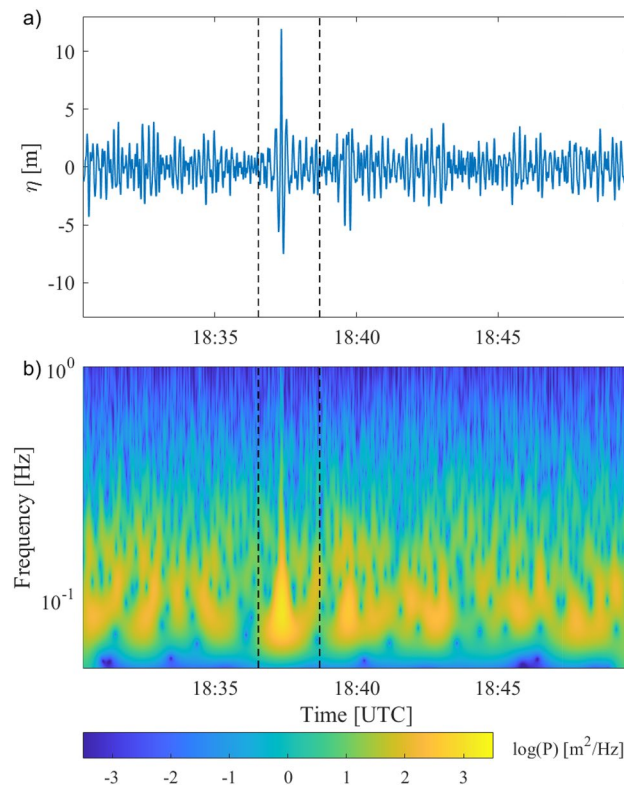


Figure 1. Rogue wave recorded on Nov 17, 2020. Vertical dashed lines indicate the wave group containing the rogue wave. **(a)** Surface elevation η . **(b)** Spectrogram of surface elevation using the Morlet wavelet.

Rogue waves can be the consequence of third order nonlinear four-wave interactions between free waves forming a modulational instability¹⁷ where individual waves can grow as sidebands, drawing energy from neighbouring waves. For long-crested waves this mechanism is well documented in wave flumes^{18–20} and it is the basis for the Benjamin–Feir Index (BFI) which is routinely being used as rogue wave risk predictor²¹. Under modulational instability the tails of the probability density function of wave crests and wave heights will be higher than predicted by second order theory. However, as waves become more directionally spread modulational instability becomes less effective, as has been shown in wave tank experiments²² and numerical simulations^{23,24}, and for observed real ocean rogue waves BFI has little predictive power²⁵.

Here we report observations of an extreme rogue wave ($\eta/H_s = 1.98$, $H/H_s = 2.91$, $H_s = 6.05$ m) off Vancouver Island, BC, Canada. According to second order theory¹³ such an event ($\eta/H_s = 1.98$, wave steepness parameter $R = 0.23$) would occur once in 1300 years. The corresponding Benjamin–Feir Index $BFI = 0.13$ indicates that modulational instability did not play a role in generating such an extreme wave. However, Monte-Carlo simulation of random superposition of fourth order Stokes waves⁸ drawn from the observed wave energy spectra yield wave and crest height distributions in good agreement with the 10 month buoy record, including the extreme rogue wave. Furthermore, we show that the crest-trough correlation $r^{25–27}$ calculated from a WAVE-WATCH III²⁸ (WW3) wave model has strong predictive power for rogue wave risk prediction from a standard wave forecast model.

Rogue wave observation

A 0.9 m CoastScout wave buoy (MarineLabs) was deployed 13/08/2020 at Amphitrite Bank (48.9° N, 125.6° W) about 7 km offshore in 45 m water depth. The buoy records surface elevation at 5 Hz from a 3-D inertial measurement unit (IMU), and from differential GPS. Data are recorded in 20 min bursts every 30 min and are available in near-real time. Here we analyze its data record until 21/05/2021, covering a wide range of sea states $0.5 \text{ m} < H_s \leq 10.1$ m. The record includes an extreme rogue wave on 17/11/2020 (Fig. 1a) with $\eta = 11.96$ m crest height (recorded by IMU and GPS sensor) when the significant wave height was $H_s = 6.05$ m. The total wave height H was 17.6 m (19.5 m) based on the preceding (following) trough. The significant wave height had increased rapidly from a minimum $H_s = 1.93$ m 18 h prior to the event. The rogue wave is the fourth crest in a group of about 10 waves. The crests of the preceding and following waves were much smaller than the rogue wave crest, with $\eta = 3.27$ m and $\eta = 4.13$ m, respectively. This is consistent with the fact that rogue waves generally occur near the centre of a group and are unexpected, i.e. there is not a gradual build-up of individual wave heights^{29,30}.

The spectrogram (Fig. 1b) reveals bursts of energy associated with wave groups and a dominant frequency $f_p = 0.078$ Hz (corresponding to a dominant period $T_p = 12.8$ s), and a second peak at $f_2 = 0.156$ Hz ($T_2 = 6.4$ s). The group containing the rogue wave has the highest energy and during the rogue wave the energy

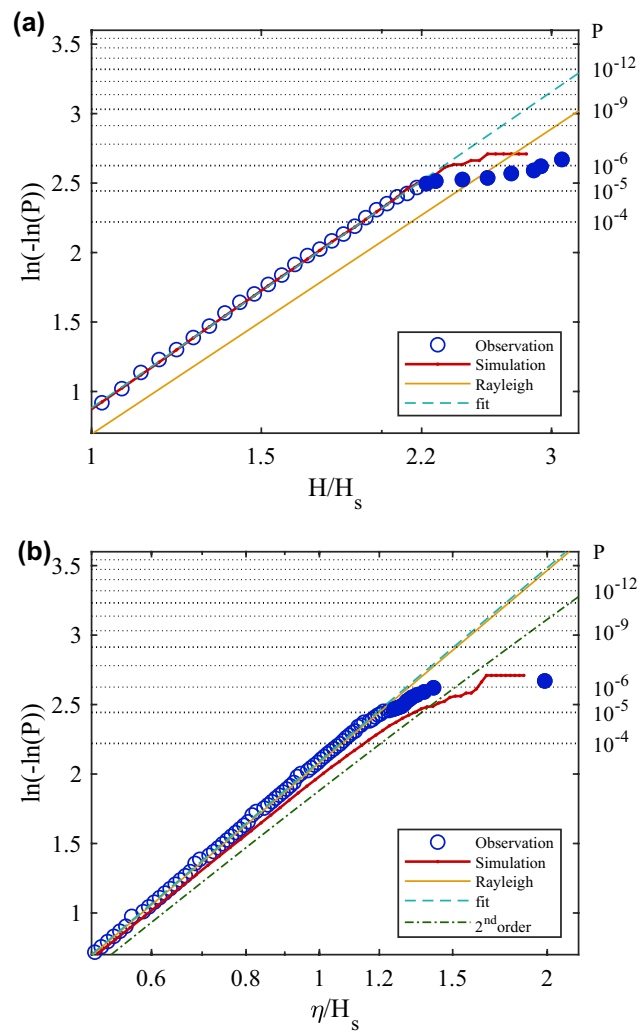


Figure 2. Exceedance probability P of (a) non-dimensional wave height H/H_s , and (b) non-dimensional crest height η/H_s . Second order curve is based on steepness $R = 0.23^{13}$. Filled symbols show observed rogue waves.

is spread continuously over a wide frequency range from $f = 0.06$ Hz to $f = 0.9$ Hz. This implies that the rogue wave is the combination of energetic waves from a broad range of scales. As the group evolves the high frequency energy rapidly decreases and the low frequency peak exhibits a continuous upshift and individual wave heights decrease. This spectral shape is unique to the wave group containing the rogue wave.

Rogue wave occurrence rates. It is convenient to plot wave and crest height exceedance probabilities P as $\ln(-\ln(P))$ versus the normalized heights η/H_s , H/H_s on a logarithmic scale (Fig. 2). On such a plot the Rayleigh distribution (Eq. 1) and distributions allowing for second order corrections fall on straight lines and any deviations are readily visible⁸. Our observations follow the standard model^{13–15} up to $H/H_s < 2.1$ and $\eta/H_s < 1.1$ but beyond that they show a dramatic increase of probabilities of high waves. As commonly observed, probabilities of wave heights H/H_s exceeding a given value are lower than given by the Rayleigh distribution, due to the finite spectral width of ocean wave data². Extrapolation of the observed distributions would give a probability of the November 17 rogue wave as 1 in 5×10^9 and the rogue crest as 1 in 10^{14} , several orders of magnitude less frequent than observed. In particular, the extreme crest height $\eta/H_s = 1.98$ would be a nearly impossible event (1 in 31 million years) if based on the extrapolation of moderate waves.

To test whether the buoy observations are consistent with the linear mechanism of wave superposition allowing for the Stokes correction, we perform a Monte-Carlo simulation^{8,30,31} based on the observed wave spectra, and add bound harmonics up to fourth order⁸ (see “Methods”). This way we obtain a synthetic surface elevation record with the same characteristics as the one recorded in the ocean, but which does not include nonlinear wave interactions and therefore no modulational instability. The wave and crest heights from the simulations are in good agreement with the observed wave height distributions. In particular the rapid increase of rogue wave probabilities over the Rayleigh distribution or second order corrections is well captured (Fig. 2). This supports random superposition of Stokes waves as the leading cause for rogue waves.

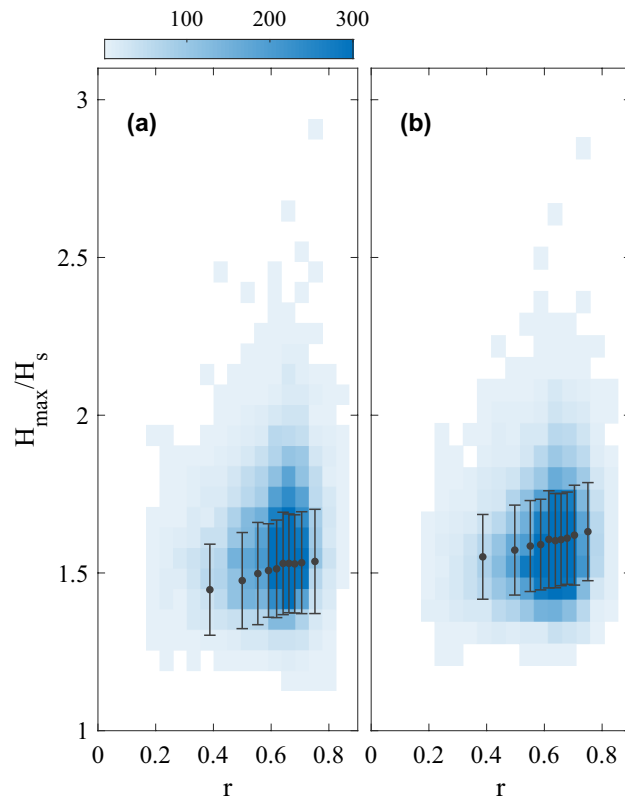


Figure 3. Maximum non-dimensional wave height H/H_s within 30 min, as function of crest-trough correlation r . Individual data are shown as blue scatter plot, with data density given by the colour bar, and constant-size bin averages and $\pm 1\sigma$ are shown in black. H_{max} , H_s , and r are taken from (a) observations, (b) Monte-Carlo simulation.

Rogue wave risk prediction

The crest-trough correlation $r^{14,32}$ (see “Methods”) is strongly correlated with the maximum normalized wave height in a record (Fig. 3). Observed and synthetic data show similar maximum wave heights for a given correlation value. The largest normalized wave heights are clearly associated with high correlation values (Fig. 3). A recent study found that in wave buoy observations the probability $P(H/H_s > 2)$ increased by about one order of magnitude for the highest r value compared to records with the lowest r^{25} . We performed a similar analysis to evaluate how the rogue wave probability p varies with sea state parameters r and BFI (see “Methods”). This analysis utilizes surface elevation measurements from the MarineLabs buoy and an additional nearby buoy, 2 km offshore in 20 m water depth referred to as the Nearshore buoy, resulting in a combined 28 months of data. Here, p is either defined as the probability of a wave exceeding $2H_s$ or $2.2H_s$.

Based on this observational data set r shows substantial correlation with p while BFI is practically ineffective (Fig. 4). The predictive power P_x of r for $p = P(H/H_s > 2)$ is 0.85 with a lower bound of 0.56, and p ranging from 9.1×10^{-6} to 6.3×10^{-5} . That is, the occurrence rate of a rogue wave exceeding $2H_s$ varies from approximately 1 in 110,000 for the lowest r value to 1 in 16,000 during records with the highest r value. For $p = P(H/H_s > 2.2)$, r has a higher predictive power of 0.99, however with a lower bound of 0.15, and p ranging from 7.2×10^{-7} to 6.9×10^{-6} . The occurrence rate of a rogue wave exceeding $2.2H_s$ varies from 1 in 1,400,000 to 1 in 140,000. One of the issues that arises when analyzing extreme events is the limited amount of data, which has the unfortunate consequence of increasing the confidence bounds of the rogue wave probability exceeding $2.2H_s$ compared to $2H_s$. BFI has a predictive power of only 0.32 with lower bound of 0.077 for $p = P(H/H_s > 2)$ and 0.81 for $p = P(H/H_s > 2.2)$ with a lower bound of -0.07. The confidence bounds on P_x indicate that BFI is not a robust prediction parameter and therefore would perform poorly as a rogue wave risk indicator.

The benefit of a spectrally derived parameter such as r relating to increased rogue wave risk, is that it can be obtained from standard spectral wave forecast models (Fig. 5). However, despite the regional WW3 model predicting H_s with high reliability with a correlation coefficient of 0.96, the models prediction of r at the near shore location of the MarineLabs and Nearshore buoy is less reliable with a correlation coefficient of 0.61 and a scatter index of 15%. Even so, the crest-trough correlation from the model r_{model} demonstrates a strong positive correlation with rogue wave probability. Additionally, r_{model} has similar high predictive power for rogue wave probability ($p = P(H/H_s > 2.2)$) as r obtained from the observations r_{buoy} .

Thus far the demonstration of the correlation between rogue wave probability and r has been restricted to coastal areas²⁵ and it has been noted that the relationships between p and sea state parameters will likely depend on location³³. Long term surface elevation records, which are the basis of the analysis presented here, and in²⁵ are

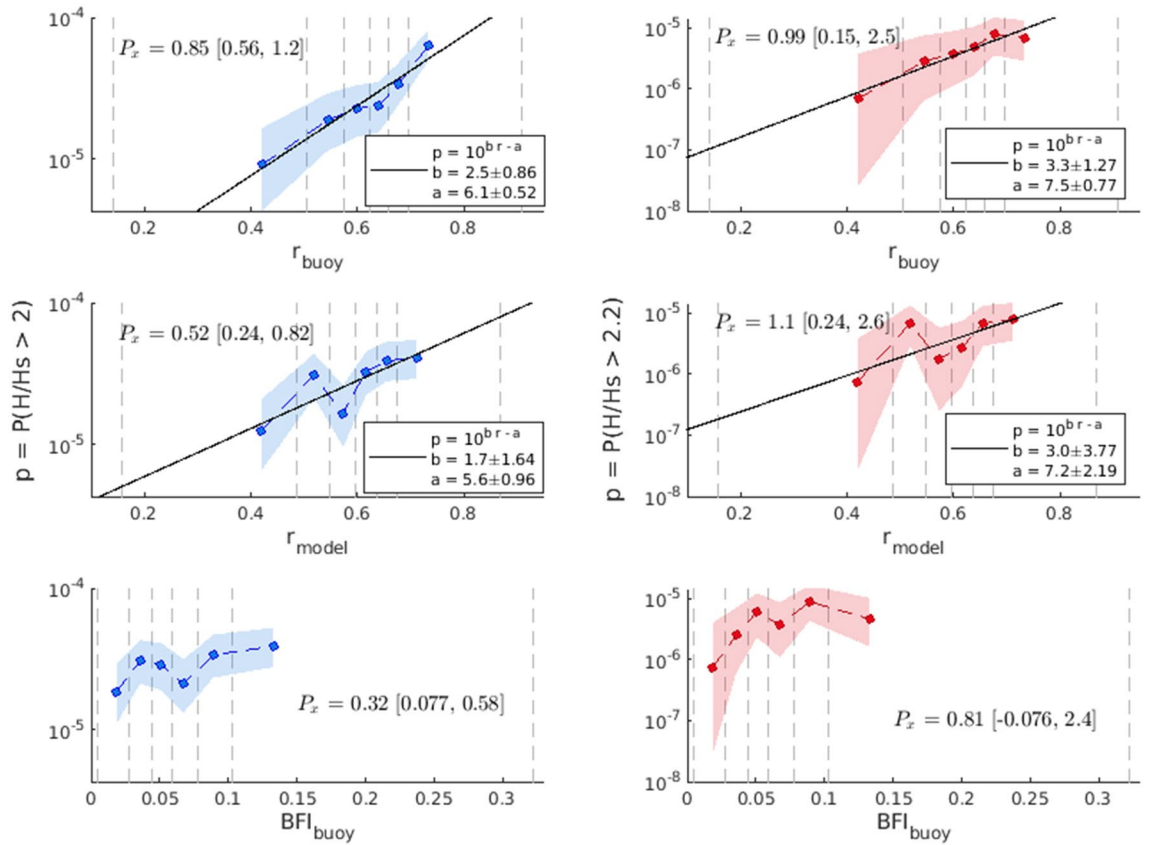


Figure 4. Rogue wave probability as functions of spectral wave parameters. Left column: Probabilities of normalized wave height >2.0 , $p = P(H/H_s > 2)$. Right column: Probabilities of normalized wave height >2.2 , $p = P(H/H_s > 2.2)$. Top row: variability with observed crest-trough correlation r_{buoy} . Middle row: variability with crest-trough correlation from the model r_{model} . Bottom row: variability with observed Benjamin Feir Index BFI_{buoy} . Shading indicates the 95% credible interval and vertical dashed lines indicate the edges of the parameter bins. Values for predictive power P_x and best fit parameters are stated.

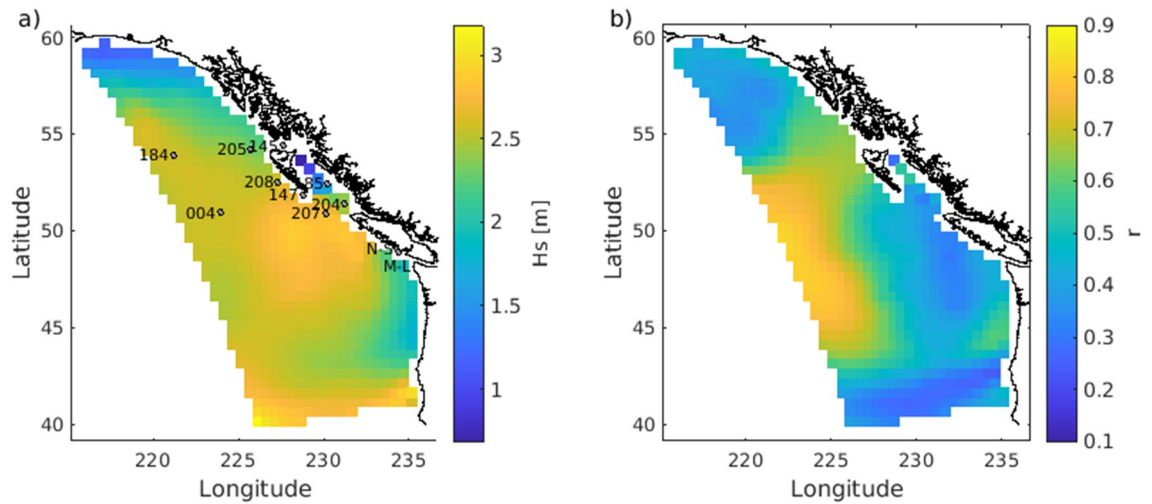


Figure 5. Example forecast output for March 22, 2020 00h UTC of (a) significant wave height H_s and (b) crest-trough correlation r . These model output fields form the basis of a rogue wave risk prediction. Grey markers in (a) are locations of buoys utilized in this study. M-L and N-S marking the location of the coastal MarineLabs and Nearshore buoys, respectively and the numbered markers are the buoys from Fisheries and Oceans Canada where the complete names are the numbers preceded by a ‘C46’ prefix.

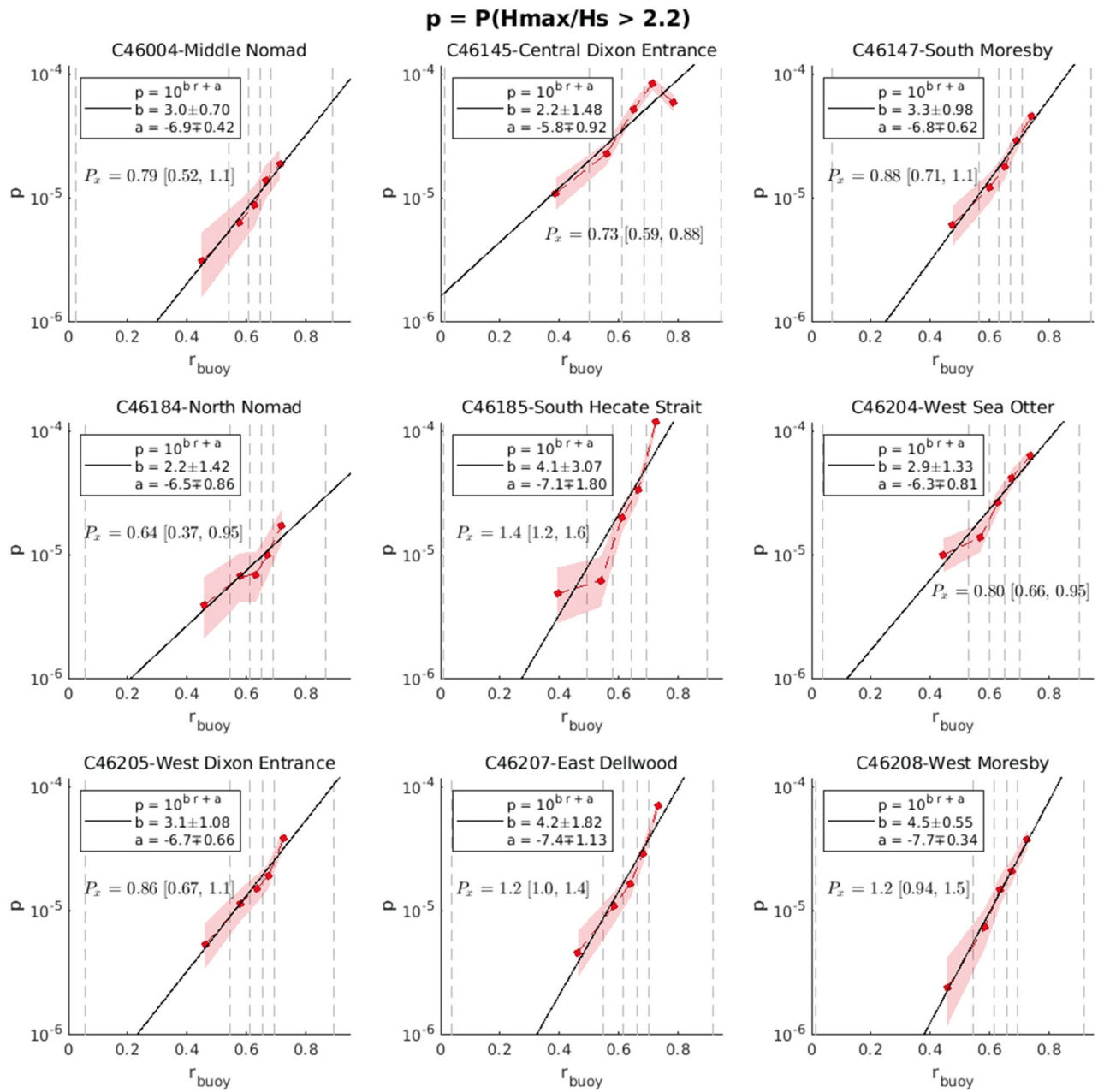


Figure 6. Variation of rogue wave probability $p = P(H_{max}/H_s > 2.2)$ with observed crest-trough correlation r_{buoy} from Fisheries and Oceans Canada buoys. Data sets include time series from 2010 to 2021 where quality control was performed and the record appears correct. Shading indicates the 95% credible interval and vertical dashed lines indicates the edges of the parameter bins.

few. However, time series of hourly wave height maxima H_{max} and significant wave height H_s are more readily available. Using these bulk statistics a similar analysis can be performed using longterm time series from buoys available through the Marine Environmental Data Section (MEDS) of Fisheries and Ocean Canada (DFO)³⁴. Records are available for a wide range of offshore buoy locations (Fig. 5) and thus the rogue wave prediction can be calibrated for the full domain of the operational wave forecast (Fig. 6). The caveat with this analysis is it does not register multiple rogue wave events in a given record segment and the number of waves in a record is estimated from the mean period. The models forecast of r also improves at offshore locations. For example, at an open ocean buoy (C46004) the correlation coefficient increased to 0.78 and the scatter index decreased to 10%.

To formalize the risk prediction forecast we took the average of the semi logarithmic fits from the Marinelabs and Nearshore data set and 9 DFO buoys to get the following equations relating rogue wave probability and r .

$$P(H/H_s > 2) = 10^{(2.54 \pm 0.48)r - (5.70 \pm 0.29)}, \text{ and } P(H/H_s > 2.2) = 10^{(3.31 \pm 0.86)r - (6.90 \pm 0.53)} \quad (2)$$

We didn't find substantial differences in the slope of the fits when grouped by depth, region or distance from the coast. The spatial and temporal fields of r can be output from the wave model and using the above equations the probability of extreme wave events can be calculated. This analysis was only performed for extreme wave heights as wave crests by definition are not affected by crest-trough correlation.

Discussion

The spectrogram of the surface elevation (Fig. 1b) highlights the rogue wave as an event where a wide range of energetic spectral wave components align. These components are steep compared to the background wave field with enhanced higher order Stokes corrections, and their superposition generates extreme crest heights. The Stokes enhancement of less energetic, and therefore less steep waves is much weaker (see Eq. 3) and their superposition results in moderate crest heights, only. This explains the dramatic increase of rogue wave probabilities for $H/H_s \gtrsim 2.1$ which is not captured by second-order Stokes corrections.

To develop a rogue wave risk prediction system we evaluated how the rogue wave probability p changes with certain sea state parameters. Based on the strong correlation with rogue wave probability at offshore and near-shore locations, we present crest-trough correlation r to be used for rogue wave risk prediction in wave forecasts. r is a logical parameter to be linked with the generation of rogue waves as it is an estimate of the auto-correlation between the crest heights and the trough depths. Therefore, in narrow-banded seas the crests and successive troughs are approximately the same size resulting in high r values. Narrow-banded seas also maintain wave groups and rogue waves are more likely to form in 'groupy' sea states. Additionally, r can be easily computed by a routine wave forecast as demonstrated in Fig. 5 and alongside a forecast for H_s provides a complete risk assessment where overlapping areas of high H_s and r pose the greatest risk.

Methods

Linear rogue wave simulation. Long time series of high-resolution wave observations under natural conditions are still rare but are required for assessing the occurrence rate of rogue waves. Many height distributions are evaluated with observed probabilities $\leq 10^{-5}$ ^{33,35} and only a few studies reach $P \leq 10^{-7}$ ^{25,27,29,36}. Alternatively, synthetic time series can provide good statistics on low-probability events and have the advantage to test non-linear (up to a desired order) or linear dynamics. Here we use the Matlab toolbox WAFO³⁷ to synthesize surface elevation from linear superposition of spectral components with sine and cosine terms that are independent and Gaussian³⁰. For each of the 11,600 data segments from the buoy a thirty-minute synthetic surface elevation h_{lin} with 5Hz resolution and the same spectral characteristics as the observations is generated. All individual waves, defined as data between consecutive zero-down crossings, are then modified with fourth order Stokes correction¹²

$$h_4 = a \cos(kx) + \left(\frac{1}{2} \varepsilon + \frac{17}{24} \varepsilon^3 \right) a \cos(2kx) + \frac{3}{8} \varepsilon^2 a \cos(3kx) + \frac{1}{8} \varepsilon^3 a \cos(4kx) \quad (3)$$

with wave amplitude a , wave steepness $\varepsilon = ak$, and the wavenumber k is obtained iteratively from the dispersion relation $\omega^2 = gk \tanh(kD)$, with $\omega = 2\pi/T$, zero-down crossing wave period T , gravitational acceleration g and water depth D . The first term on the r.h.s. of Eq. (3) represents the linear component h_{lin} , the remaining terms the Stokes correction. For each data segment the significant wave height $H_s = 4\sigma$, where σ is the standard deviation of the surface elevation, the crest-trough correlation r and BFI (see below) are calculated. For the entire observed and synthetic time series all individual crest heights and trough-to-crest wave heights are extracted and normalized by H_s of the segment.

Wave model. A WAVEWATCHIII[®] (WW3) regional wave model was calibrated for the Northeast Pacific where the domain extends from northwest and southwest corners located at approximately (60° N, 145° W) and (40° N, 134° W) to the coast (Fig. 5). The model uses a triangular unstructured grid with variable resolution of 1000 m nearshore to 5 km offshore. An implicit solving scheme was used in this study with global time step of 600 s. The model resolves 36 frequency bins from 0.035 to 0.98 Hz and 36 direction bins spaced every 10. Boundary conditions are imposed at the western and southern open boundary nodes extracted from the 0.25 Global Deterministic Wave Prediction System (GDWPS) every 5 km (https://weather.gc.ca/model_forecast/wave_e.html). The model is forced with 2.5 km resolution winds from High Resolution Deterministic Prediction System (HRDPS). The ST4 package was used to parameterize wind input and dissipation with growth parameter³⁸. In addition, nonlinear four wave interaction using discrete interaction approximation, DIA (NL1)³⁹, JONSWAP bottom friction (BT1), depth-limited breaking⁴⁰ (DB1), triad interactions (TR1)⁴¹, and a linear input term (LN1) have been used for the computations. To produce the low resolution r field in Fig. 5 the WW3 model was run and point spectra were output every . r was then calculated from Eq. (7).

Rogue wave predictors. To assess the nonlinear mechanism of rogue wave generation by the modulation instability we use the Benjamin Feir Index (BFI)¹⁷. BFI is calculated as the ratio of wave steepness ε to spectral bandwidth ν ⁴².

$$\text{Benjamin Feir Index, BFI} = \sqrt{2} \left(\frac{\varepsilon}{\nu} \right) \quad (4)$$

$$\text{Wave Steepness, } \varepsilon = \sqrt{k_p^2 m_0} \quad (5)$$

$$\text{Spectral Bandwidth, } \nu = \sqrt{\frac{m_0 m_2}{m_1^2}} - 1 \quad (6)$$

where k_p is the wavenumber associated with the peak period T_p calculated from the weighted spectrum⁴³, and $m_n = \int_0^\infty f^n S(f) df$ is the n^{th} spectral moment.

The crest-trough correlation r has been suggested as a parameter to relate to wave height distributions¹⁴. Crest-trough correlation r is the auto-correlation of the sea surface elevation at half the wave period. It can be estimated from the spectral density, $S(f)$ using the Wiener-Khinchin theorem. Following¹⁵ we compute r as,

$$r = \frac{1}{m_0} \sqrt{\rho^2 + \lambda^2}, \text{ where } \rho = \int_0^\infty S(f) \cos(2\pi f \tau) df \text{ and } \lambda = \int_0^\infty S(f) \sin(2\pi f \tau) df, \quad (7)$$

where $\tau = \frac{\bar{T}}{2}$ is the lag time at half the spectral mean period $\bar{T} = \frac{m_0}{m_1}$.

Estimation of rogue wave probabilities. We evaluate the univariate rogue wave probability p with uncertainties following the same method as Häfner et al. (2021) outlined briefly below²⁵. We define p as the probability that the next wave height will exceed the threshold $2H_s$ or $2.2H_s$. To evaluate how p varies with a wave parameter x we split x and H/H_s into N bins with approximately equal number of observations in each bin. We assume that the number of rogue waves n^+ and the number of non-rogue waves n^- in each bin are identically and independently distributed (iid) according to a binomial distribution with probability p . For the prior distribution of p we assume a Beta distribution with parameters $\alpha_0 = 1$ and $\beta_0 = 3000$ for $P(H/H_s > 2)$ and $\beta_0 = 16,000$ for $P(H/H_s > 2.2)$ from the Rayleigh distribution. Applying Bayes theorem we find the posterior probability for p is another Beta distribution, since the Beta prior for p is conjugate to the binomial distribution of n^+ .

$$P(p|n^+, n^-) = \text{Beta}(n^+ + \alpha_0, n^- + \beta_0) \quad (8)$$

Our estimate of p with uncertainties is calculated from the median and the 95% credible interval of Eq. (8) (based on 2.5th and 97.5th percentiles). The benefit of this analysis is to be able to generate uncertainties for p to determine whether the variation with x is meaningful. Here are a few things to note. Firstly, the purpose of the Beta prior is only to constrain p to a reasonable order of magnitude so the exact choice of α_0 and β_0 does not influence final results. Secondly, this analysis hinges on the assumptions that p is iid distributed in each bin, which isn't the case if p depends on more than one parameter. Therefore, these uncertainties indicate the level of confidence in our best estimate of p if we only consider one parameter at a time.

For buoys where surface elevation is unavailable, the normalized wave is given by H_{max}/H_s where H_{max} is the maximum wave in a record and the total number of waves N_0 is estimated using the mean period \bar{T} and the record length RL where, $N_0 = RL/\bar{T}$. To evaluate if this simplifying approach is acceptable we perform these simplifications on our Nearshore and MarineLabs data set to compare with results using the full sea surface elevation. Using H_{max}/H_s excludes only three rogue events in the data set. Those events correspond to segments where there were two waves with height exceeding $2H_s$. Overall, there is no appreciable difference using H_{max} and T_{01} to get N_0 , however there is the limitation that the limit of rogue waves flagged in a record is one.

To evaluate the degree of variability of p with parameter x we use the predictive power defined as, where i_{max} is the bin index where x is highest and i is bin index where x is lowest. This measures how much p varies with x if we only consider this one parameter. The uncertainty in p is quantified through Monte Carlo sampling, based on the known distributions of $p_{i_{max}}$ and p_i given in Eq. (8) and the 95% confidence interval calculated from the 2.5th and 97.5th percentiles of the distribution of p .

MEDS buoys. The offshore buoys data sets are provided by the Marine Environmental Data Section (MEDS) buoys of Fisheries and Oceans Canada (DFO) (<https://meds-sdmm.dfo-mpo.gc.ca>)³⁴. We utilized the long term time series of H_{max} , H_s and frequency spectra. The open ocean buoys (C46004 and C46184) are 6m NOMAD buoys, the rest are AXYS 3m discus buoys. They record vertical acceleration at 1 Hz sampling rate for 34 minutes every hour and output bulk wave statistics. This work only used data which had received a quality control check and the record appears correct from 01/01/2010 to 15/08/2021. Any other suspicious data sections were removed. Due to outages in service and erroneous records the time series from the MEDS buoys are variable in length and do not necessarily span the full 11 year. The buoys location are found in Fig. 5a.

Data availability

The wave model data and the MarineLabs buoy data are available from the corresponding author on reasonable request. Data access is restricted to research and educational applications. Data from the Fisheries and Oceans Canada buoys can be accessed at <https://meds-sdmm.dfo-mpo.gc.ca>.

Code availability

The generation of the synthetic surface elevation data was performed with the free Matlab toolbox “WAFO”³⁷. WAVEWATCH III[®] is open source code²⁸.

Received: 22 October 2021; Accepted: 17 January 2022

Published online: 02 February 2022

References

1. Haver, S. A possible freak wave event measured at the Draupner jacket January 1 1995. In *Proceedings of Rogue Waves* (2004).

2. Garrett, C. & Gemmrich, J. Rogue waves. *Phys. Today* **62**, 62–63 (2009).
3. Adcock, T. A. A., Taylor, P. H., Yan, S., Ma, Q. W. & Janssen, P. A. E. M. Did the Draupner wave occur in a crossing sea?. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **467**, 3004–3021. <https://doi.org/10.1098/rspa.2011.0049> (2011).
4. Cavaleri, L. *et al.* The Draupner wave: A fresh look and the emerging view. *Journal of Geophysical Research: Oceans* **121**, 6061–6075. <https://doi.org/10.1002/2016jc011649> (2016).
5. Donelan, M. A. & Magnusson, A.-K. The making of the Andrea wave and other rogues. *Scientific Reports* **7**, <https://doi.org/10.1038/srep44124> (2017).
6. Fedele, F., Brennan, J., de León, S. P., Dudley, J. & Dias, F. Real world ocean rogue waves explained without the modulational instability. *Scientific Reports* **6**, <https://doi.org/10.1038/srep27715> (2016).
7. Kokina, T. & Dias, F. Influence of computed wave spectra on statistical wave properties. *J. Mar. Sci. Eng.* **8**, 1023. <https://doi.org/10.3390/jmse8121023> (2020).
8. Gemmrich, J. & Garrett, C. Dynamical and statistical explanations of observed occurrence rates of rogue waves. *Nat. Hazard.* **11**, 1437–1446. <https://doi.org/10.5194/nhess-11-1437-2011> (2011).
9. Dysthe, K., Krogstad, H. E. & Müller, P. Oceanic rogue waves. *Annu. Rev. Fluid Mech.* **40**, 287–310 (2008).
10. Dudley, J. M., Genty, G., Mussot, A., Chabchoub, A. & Dias, F. Rogue waves and analogies in optics and oceanography. *Nat. Rev. Phys.* **1**, 675–689. <https://doi.org/10.1038/s42254-019-0100-0> (2019).
11. Longuet-Higgins, M. S. On the statistical distribution of the heights of sea waves. *J. Mar. Res.* **11**, 245–266 (1952).
12. Kinsman, B. *Wind waves: their generation and propagation on the ocean surface (Englewood Cliffs)* (Prentice-Hall, N.J., 1965).
13. Forristall, G. Z. Wave crest distributions: Observations and second-order theory. *J. Phys. Oceanogr.* **30**, 1931–1943 (2000).
14. Tayfun, M. A. Distribution of large wave heights. *J. Waterw. Port Coast. Ocean Eng.* **116**, 686–707. [https://doi.org/10.1061/\(asce\)0733-950x\(1990\)116:6\(686\)](https://doi.org/10.1061/(asce)0733-950x(1990)116:6(686)) (1990).
15. Tayfun, M. A. & Fedele, F. Wave-height distributions and nonlinear effects. *Ocean Eng.* **34**, 1631–1649. <https://doi.org/10.1016/j.oceaneng.2006.11.006> (2007).
16. Fedele, F., Herterich, J., Tayfun, A. & Dias, F. Large nearshore storm waves off the Irish coast. *Sci. Rep.* **9**, <https://doi.org/10.1038/s41598-019-51706-8> (2019).
17. Janssen, P. A. E. M. Nonlinear four-wave interactions and freak waves. *J. Phys. Oceanogr.* **33**, 863–884 (2003).
18. Onorato, M. *et al.* Extreme waves, modulational instability and second order theory: Wave flume experiments on irregular waves. *Eur. J. Mech. B. Fluids* **25**, 586–601. <https://doi.org/10.1016/j.euromechflu.2006.01.002> (2006).
19. Cherneva, Z., Tayfun, M. A. & Soares, C. G. Statistics of nonlinear waves generated in an offshore wave basin. *J. Geophys. Res.* **114**. <https://doi.org/10.1029/2009jc005332> (2009).
20. Chabchoub, A., Hoffmann, N. P. & Akhmediev, N. Rogue wave observation in a water wave tank. *Phys. Rev. Lett.* **106**, 204502. <https://doi.org/10.1103/physrevlett.106.204502> (2011).
21. Janssen, P. & Bidlot, J.-R. On the extension of the freak wave warning system and its verification. *ECMWF Technical Memorandum* 588. <https://doi.org/10.21957/UF1SYBOG> (2009).
22. Onorato, M. *et al.* Statistical properties of mechanically generated surface gravity waves: a laboratory experiment in a three-dimensional wave basin. *J. Fluid Mech.* **627**, 235–257. <https://doi.org/10.1017/s002211200900603x> (2009).
23. Socquet-Juglard, H., Dysthe, K., Trulsen, K., Krogstad, H. E. & Liu, J. Probability distributions of surface gravity waves during spectral changes. *J. Fluid Mech.* **542**. <https://doi.org/10.1017/s0022112005006312> (2005).
24. Toffoli, A. *et al.* Evolution of weakly nonlinear random directional waves: Laboratory experiments and numerical simulations. *J. Fluid Mech.* **664**, 313–336. <https://doi.org/10.1017/s002211201000385x> (2010).
25. Häfner, D., Gemmrich, J. & Jochum, M. Real-world rogue wave probabilities. *Sci. Rep.* **11**. <https://doi.org/10.1038/s41598-021-89359-1> (2021).
26. Naess, A. On the distribution of crest to trough wave heights. *Ocean Eng.* **12**, 221–234. [https://doi.org/10.1016/0029-8018\(85\)90014-9](https://doi.org/10.1016/0029-8018(85)90014-9) (1985).
27. Casas-Prat, M. & Holthuijsen, L. H. Short-term statistics of waves observed in deep water. *J. Geophys. Res.* **115**. <https://doi.org/10.1029/2009jc005742> (2010).
28. The WAVEWATCH III Development Group. User manual and system documentation of WAVEWATCH III⁺ version 6.07. Tech. Note 333, NOAA/NWS/NCEP/MMAB (2019). 465 pp. + Appendices.
29. Gemmrich, J. & Thomson, J. Observations of the shape and group dynamics of rogue waves. *Geophys. Res. Lett.* <https://doi.org/10.1002/2016gl072398> (2017).
30. Gemmrich, J. & Garrett, C. Unexpected waves. *J. Phys. Oceanogr.* **38**, 2330–2336 (2008).
31. Gemmrich, J. & Garrett, C. Unexpected waves: Intermediate depth simulations and comparison with observations. *Ocean Eng.* **37**, 262–267 (2010).
32. Häfner, D., Gemmrich, J. & Jochum, M. FOWD: A free ocean wave dataset for data mining and machine learning. *J. Atmos. Oceanic Tech.* <https://doi.org/10.1175/jtech-d-20-0185.1> (2021).
33. Cattrell, A. D., Srokosz, M., Moat, B. I. & Marsh, R. Can rogue waves be predicted using characteristic wave parameters?. *J. Geophys. Res. Oceans* **123**, 5624–5636. <https://doi.org/10.1029/2018jc013958> (2018).
34. DFO. Marine environmental data section archive. <https://meds-sdmf.dfo-mpo.gc.ca>, Ecosystem and Oceans Science, Department of Fisheries and Oceans Canada (2021). Data obtained on 2021/08/15.
35. Knobler, S., Bar, D., Cohen, R. & Liberzon, D. Wave height distributions and rogue waves in the eastern. *J. Mar. Sci. Eng.* **9**, 660. <https://doi.org/10.3390/jmse9060660> (2021).
36. Orzech, M. D. & Wang, D. Measured rogue waves and their environment. *J. Mar. Sci. Eng.* **8**, 890. <https://doi.org/10.3390/jmse8110890> (2020).
37. Brodtkorb, P. *et al.* WAFO—a Matlab toolbox for the analysis of random waves and loads. In *Proc. 10th Int. Offshore and Polar Eng. Conf., ISOPE, Seattle, USA*, vol. 3, 343–350 (2000).
38. Ardhuin, F. *et al.* Semiempirical dissipation source functions for ocean waves. Part I: definition, calibration, and validation. *J. Phys. Oceanogr.* **40**, 1917–1941. <https://doi.org/10.1175/2010jpo4324.1> (2010).
39. Hasselmann, S., Hasselmann, K., Allender, J. H. & Barnett, T. P. Computations and parameterizations of the nonlinear energy transfer in a gravity-wave spectrum. Part II: Parameterizations of the nonlinear energy transfer for application in wave models. *J. Phys. Oceanogr.* **15**, 1378–1391 (1985).
40. Battjes, J., J.A. Janssen. Energy loss and set-up due to breaking random waves. *Proceedings of 16th Conference on Coastal Engineering* 569–587 (1978).
41. Eldeberky, Y., Polnikov, V. & Battjes, J. *A statistical approach for modeling triad interactions in dispersive waves*, chap. 85, 1088–1101 (Coastal Engineering, 1997).
42. Thomson, J., Gemmrich, J., Rogers, W. E., Collins, C. O. & Ardhuin, F. Wave groups observed in pancake sea ice. *J. Geophys. Res. Oceans* **124**, 7400–7411. <https://doi.org/10.1029/2019jc015354> (2019).
43. Young, I. The determination of confidence limits associated with estimates of the spectral peak frequency. *Ocean Eng.* **22**, 669–686. [https://doi.org/10.1016/0029-8018\(95\)00002-3](https://doi.org/10.1016/0029-8018(95)00002-3) (1995).

Acknowledgements

We thank MarineLabs (Victoria, BC) and the Pacific Regional Institute for Marine Energy Development (University of Victoria) for access to the buoy data. Data collection was funded by an Innovate BC IGNITE grant. Data analysis was funded by Search And Rescue New Initiative Funds (Public Safety Canada) grant SN201907.

Author contributions

J.G. conceived the project, L.C. conducted the wave model, both authors analysed the data and co-wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022