Data Mining Pt. Reyes Buoy for Rare Wave Groups

Harleigh C. Seyffert[1]
Naval Architecture and Marine Engineering,
University of Michigan,
Ann Arbor, MI 48109
e-mail: harleigh@umich.edu

Armin W. Troesch
Naval Architecture and Marine Engineering,
University of Michigan,
Ann Arbor, MI 48109
e-mail: troesch@umich.edu

This paper addresses the existence of rare wave groups by examining time series data from the Pt. Reyes buoy. The buoy is operated by the Coastal Data Information Program (CDIP), University of California San Diego. The definition of rare wave groups, as defined by Kim and Troesch, used in this paper differs from the more commonly used wave group definition based on threshold crossings. With the time series data from the Pt. Reyes buoy, these rare wave groups are shown to be a naturally occurring phenomenon. The essential features of the data are examined, as well as the analysis methods and findings. By sifting through 17 years of wave elevation data from the Pt. Reyes buoy, this preliminary work addresses not only the question to what extent rare wave groups exist in nature but also what their probability of occurrence is.
[DOI: 10.1115/1.4031973]

Introduction

Due to their significant, possibly critical, effect on the performance of marine systems, wave groups have long been a popular topic in oceanographic and ocean engineering research. Traditional descriptions of wave groups are typically formulated as an envelope peaks over a threshold problem or a counting of discrete peaks over a threshold problem (e.g., Markov chain model). Early investigations into the statistics of wave groups were made by several researchers, e.g., Longuet-Higgins [2], Goda [3], and Kimura [4]. Liu et al. [5], among others, noted that recorded statistics of real wave groups seem to follow linear Gaussian models, at least in a qualitative sense. The instances when linear theory appears to be inappropriate are typically related to sparse data or waves in the surf zone, e.g., Ref. [5].

The various approaches are reviewed in many publications where comparisons between predictions of the theories and field data suggest some limitations or deficiencies (e.g., Masson and Chandler [6], Ochi [7], and Tucker and Pitt [8]). Challenging issues include, but are not limited to, the level of (in)dependence of successive wave peaks, low threshold levels for long wave group runs (typically less than two standard deviations), and the desire for ensemble averages when only limited temporal runs are available.

Historically, wave groups were defined as a given number of successive wave peaks that exceed some threshold value. However, identification of long group runs from field data has been challenging due to either rarity of long runs exceeding a prescribed high level, or lack of sufficient datasets containing large wave groups. In addition, when looking at ship dynamical events such as parametric roll, this traditional definition does not account for the large role that the forcing period has on the event. And equally problematic, an ensemble of wave groups restricted to successive threshold crossings will not include the more probable wave group sequence, targeted for a critical threshold, which still permits the inclusion of one or two minor down-crossings within the grouping. This more likely wave group is, through resonant excitation, also capable of generating extreme responses.

A new wave group definition, based on a derived stochastic process of a Gaussian random variable [1], relaxes the successive amplitude-threshold requirement and includes a mean-cycle frequency prerequisite. Specifically, these wave groups are identified by extreme values of a derived process based on a moving temporal summation of individual wave elevations. Rare wave groups of a given number, or run index, k, are recognized when the derived process, the summation of the k samples separated in time by a prescribed period, τ, reaches a maximum value within a given data block. Looking at a large dataset of wave elevations assumed to represent a stationary and ergodic process, one can take the truncated segments where the maximum of the derived process occurs and establish wave group statistics. For example, the expected profile of the water surface, conditioned on the derived process being a maximum, will be found by taking the ensemble average. For a Gaussian process, it has been established that this ensemble average will take the form of a wave group [1].

Data Collection. The data used in this research were collected from the Pt. Reyes buoy operated by the CDIP, the University of California, San Diego [9]. The Pt. Reyes buoy operates off the coast of San Francisco, California in approximately 550 m water depth. The buoy is equipped with a GPS tracker to record its moored location and is designed to ride the waves so buoy motion yields the wave height. Specifically, the buoy's accelerometer captures the accelerations of the body. The data are then filtered and two integrations yield the buoy displacement (surface elevation). Shown in Fig. 1 is a schematic of the buoy along with its reference axes. The buoy specifications are given in Table 1.

The buoy's ability to capture high wave frequencies is limited by its diameter. Conversely, its ability to measure low frequencies is limited by its mooring. For these limiting cases, the energy is spread over multiple frequencies. Tracking the pitch and roll indicates wave direction. The mooring for the buoy is a combination of an anchor weight at the seabed, then polypropylene rope, and then a stabilizing chain which attaches to the buoy [9].

Data Analysis. The data analysis used in this paper includes elements of the following statistical quantities:

(1) Ensemble and temporal averages, i.e., mean square and variance, based on averages of spectra, the derived process for given group index k, and wave elevations associated with a derived process maxima.
(2) Order statistics to determine extreme value statistics based on ensemble maxima.

[1]Corresponding author.
Contributed by the Ocean, Offshore, and Arctic Engineering Division of ASME for publication in the Journal of Offshore Mechanics and Arctic Engineering. Manuscript received June 18, 2015; final manuscript received October 29, 2015; published online December 15, 2015. Assoc. Editor: Yi-Hsiang Yu.

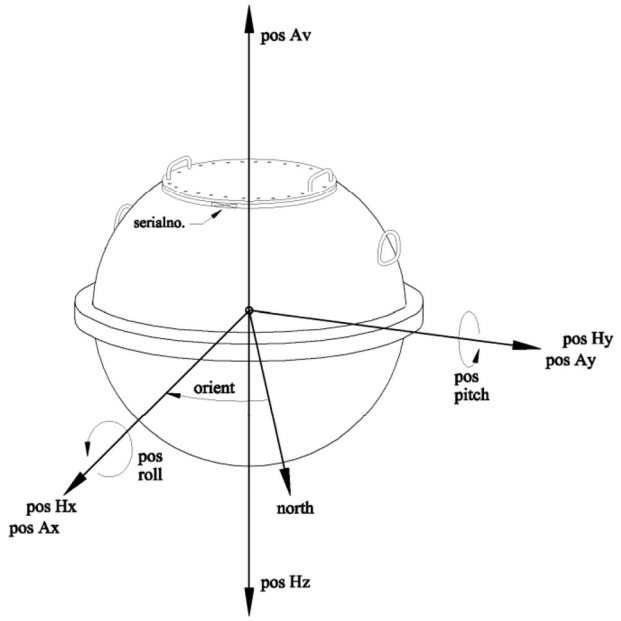

Fig. 1 Orientation of reference axes for datawell directional buoy [9]

Table 1 Buoy specifications

Parameter	Value
Heave	
Range	$-20 \rightarrow 20$ m
Resolution	1 cm
Period range	$1.6 \rightarrow 30$ s
Filtering	
Sampling frequency	1.28 Hz
Digital filtering type	Phase linear, combined band-pass double-integrating FIR filter

(3) Modeling of extreme value distributions of the derived process under the assumption of Gaussian processes.

The major method employed to identify wave groups within the available time series was the derived process [1], which has similarities to a moving average. The time series from the buoy were sorted into 30-min segments (the majority of the data was reported in 30-min increments [9]) and then linked with additional buoy data to attribute a significant wave height and peak modal period to the temporal elevation data. This allowed a filtering of the data to look at rare wave groups within certain ranges of peak modal periods and significant wave heights, as will be discussed later. A rare wave group of k waves within a time series segment can be identified by the maximum of the derived process with k peaks, separated by τ, the peak modal period. Specifically, the derived process, $z_k(t)$, for a wave group of k waves is given in Eq. (1). This would be similar to a moving average if $z_k(t)$ were normalized by the wave index k and the time separation constant τ, replaced by the time series sampling rate, 1/1.28 Hz (see Table 1)

$$z_k(t) = \sum_{j=1}^{k} \eta(t + (j-1)\tau) \tag{1}$$

For the purposes of this paper, the maximum values of the derived process for a given index $k = 1$–15 are found. Then, the locations of the various ensemble maxima of the derived process are lined up so that the peaks of the derived process occur at the same point across all ensemble samples. This is equivalent to shifting, in time, the maxima to $t = t_o$, where t_o may take an arbitrary value without loss of generality. A subsequent ensemble average shows the general trend. A large number of time series are required for adequate ensemble averaging to realize converged statistics. However, the 17 years of time series data from the Pt. Reyes buoy should be sufficient for most statistical metrics.

Results

General Environmental Conditions. Given the time series data from the Pt. Reyes buoy, it is possible to find the most rare wave groups for groups of 1–15 waves. An initial challenge was the categorization of the different ranges of significant wave heights and modal periods that make up the domain of the 17 year buoy service. In Fig. 2, the distribution of time series attributed to a particular significant wave height and peak modal period is shown, with the color bar representing the number of 30-min time segments that fit in the individual bins. Based on the CDIP buoy specification sheets, the buoys are only able to resolve certain frequencies, and therefore can only attribute a discrete number of peak modal period values to the time series [9]. Some peak modal period values have no time series in that bin only because of resolution coarseness in the data process.

Based on the distribution of the time series (i.e., Fig. 2), the following strategy was used to select bin categories: keep the significant wave height range the same and use time series with different peak modal periods. It was also important to choose bins that had enough available time series for reasonably converged statistics.

Given these criteria, two bins were chosen and are shown in Table 2 and Fig. 2.

Comparison of time series from the two bins allows for an initial evaluation of the effect on rare wave groups of wave steepness and number of wave cycles in a 30-min time series. The two spectra are shown in Fig. 3, with a representative spectra corresponding to an individual time series and the ensemble spectrum for each bin. The "ensemble spectrum" is considerably smother since ensemble averaging of "n" raw spectra has the effect of reducing the variance of any normalized spectral ordinate by $O(1/\sqrt{n})$ [10]. The significant wave height for each bin was calculated from four times the square root of the area under the ensemble spectrum curve with values for the two bins given in Table 2. The peak modal period corresponds to the location of the maximum value in the ensemble spectrum and is also given in Table 2.

For the two bins selected, the range of significant wave heights remained constant but featured time series with longer

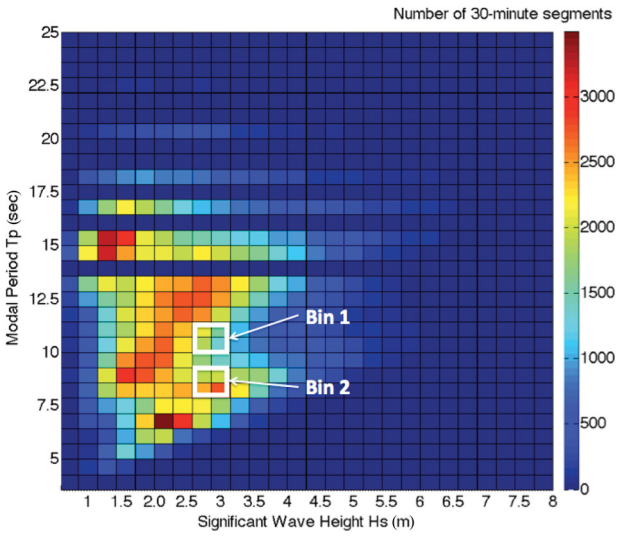

Fig. 2 Distribution of available time series for given significant wave height/peak modal period January 1997–December 2013

Table 2 Bin ranges

Parameter	Value
Bin 1	
Hs range	2.6 → 3.0 m
Tp range	10.0 → 11.5 s
Number of 30-min time series	3350
Total time	1675 hrs
4σ (ensemble average)	2.75 m
Ensemble peak period	10.65 s
Bin 2	
Hs range	2.6 → 3.0 m
Tp range	7.9 → 8.6 s
Number of 30-min time series	2430
Total time	1215 hrs
4σ (ensemble average)	2.81 m
Ensemble peak period	8.22 s

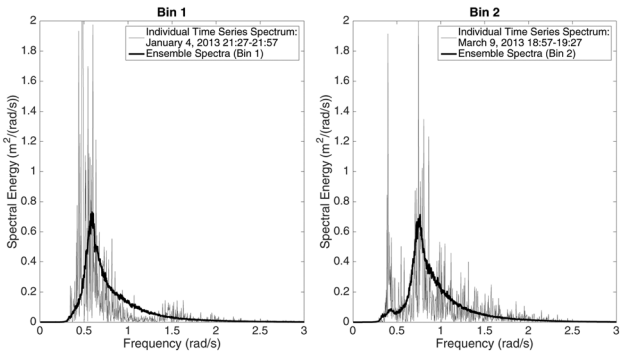

Fig. 3 Truncated double-sided spectrum for bins 1 and 2

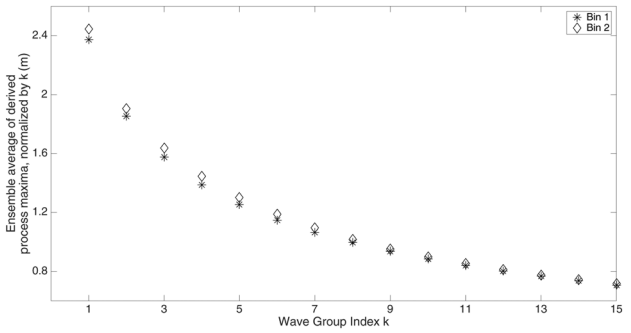

Fig. 4 Average of derived process maxima for 30-min time series in bins 1 and 2, normalized by *k*

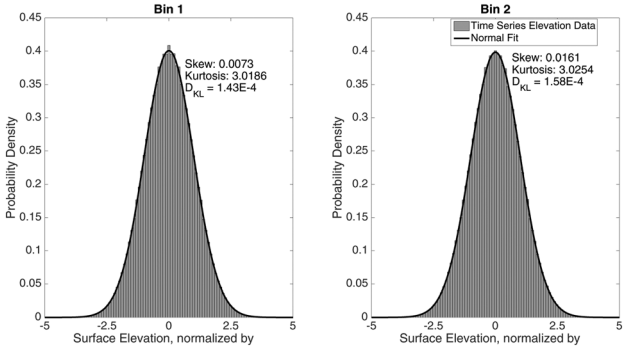

Fig. 5 Probability density function (PDF) of wave elevation time series data, overlaid with Gaussian distribution, for all time series in bins 1 and 2 ($N = 7.7 \times 10^6$ and $N = 5.6 \times 10^6$, respectively)

(10.0–11.5 s) and shorter (7.9–8.6 s) peak modal periods. This range of periods yields a noticeable difference in the number of oscillations in the two sets of time series as a whole. However, it is remarkable that wave groups with similar characteristics were identified for both bins for this range of significant wave heights.

The time separation between the points in the derived process, τ, can easily be changed to match a particular dynamic problem. For purposes of this paper, τ was chosen as the peak modal period of the ensemble spectrum for each bin. This could reflect half the roll natural period of a vessel if the objective were to find a rare wave group that could cause parametric excitation, which in turn could lead to an extreme dynamic event for that vessel (e.g., Ref. [1]).

With the two bins selected, the time series were sorted and the derived process was computed to identify maximum wave groups for groups of waves 1–15. Examples of time series shown in this paper are for wave groups of 1, 3, 6, and 9 waves.

Derived Process Maxima and Wave Group Statistics. The derived process for index k of 1–15, corresponding to wave groups of 1–15 waves, was calculated for each 30-min time series that fit into each bin. From these 30-min derived process time series, the maximum value and time of maximum value for a given wave group index were determined.

In Fig. 4, the maxima of the derived process for all time series in bins 1 and 2 are averaged and normalized by the wave group index. This is a valuable measure of the wave groups since it quantifies the inverse relationship between average group amplitude and length of group run. Indeed, the largest normalized value for the derived process occurs for a single wave, and decreases as the wave index increases. This indicates that if we start to look for wave groups with long runs, we cannot expect those groups to have large mean wave group heights.

The empirical histograms of all the wave time series elevations for each bin were also calculated and are given in Fig. 5, with the time series all normalized by their respective standard deviation (i.e., σ). A Gaussian PDF is overlaid and the skewness and kurtosis for the respective time series are given. To compare the empirical histograms of the time series data with the Gaussian distribution, the Kullback–Leibler divergence, D_{KL} [11], as given in Eq. (2), was calculated for each bin

$$D_{\mathrm{KL}}(P||Q) = \sum_i P(i) \ln \frac{P(i)}{Q(i)} \tag{2}$$

Here, Q is the target Gaussian fit and P is the empirical histogram. This divergence measures how much information is lost if Q estimates P. If P and Q are exactly equal, the divergence equals zero. It follows that the Kullback–Leibler divergence (D_{KL}) gives a good measure to determine if one PDF (e.g., a PDF based on bin 1) is more Gaussian than another (e.g., a PDF based on bin 2). In Fig. 5, the two D_{KL} values for bins 1 and 2 (i.e., $D_{\mathrm{KL}} = O(10^{-4})$) suggest that the preponderance of empirical wave elevation samples follows a normal process.

For wave group indices of $k = 1$, 3, 6, and 9, the ensemble average of the wave elevation time series in each bin, lined up such that the maxima of the derived process occur at the same time $t = t_0$, was calculated. Specifically, the ensemble average of the time series is based on the maximum of $z_k(t_0)$, defined as $\hat{z}_k$, with the results shown in the left column of Fig. 6. Two hundred seconds of the wave elevations are shown where the time of $\hat{z}_k$ maxima, t_0, has been arbitrarily set equal to 100 s. The average is an approximation to $E[\eta(t)|z_k(t_0) = \hat{z}_k]$. The top 50 values of the derived process for each bin were also identified, and those wave groups along with the ensemble average of those top 50 realizations are given in the right column of Fig. 6.

The maxima of the derived process clearly identify wave groups. The ensemble averages show little correlation between

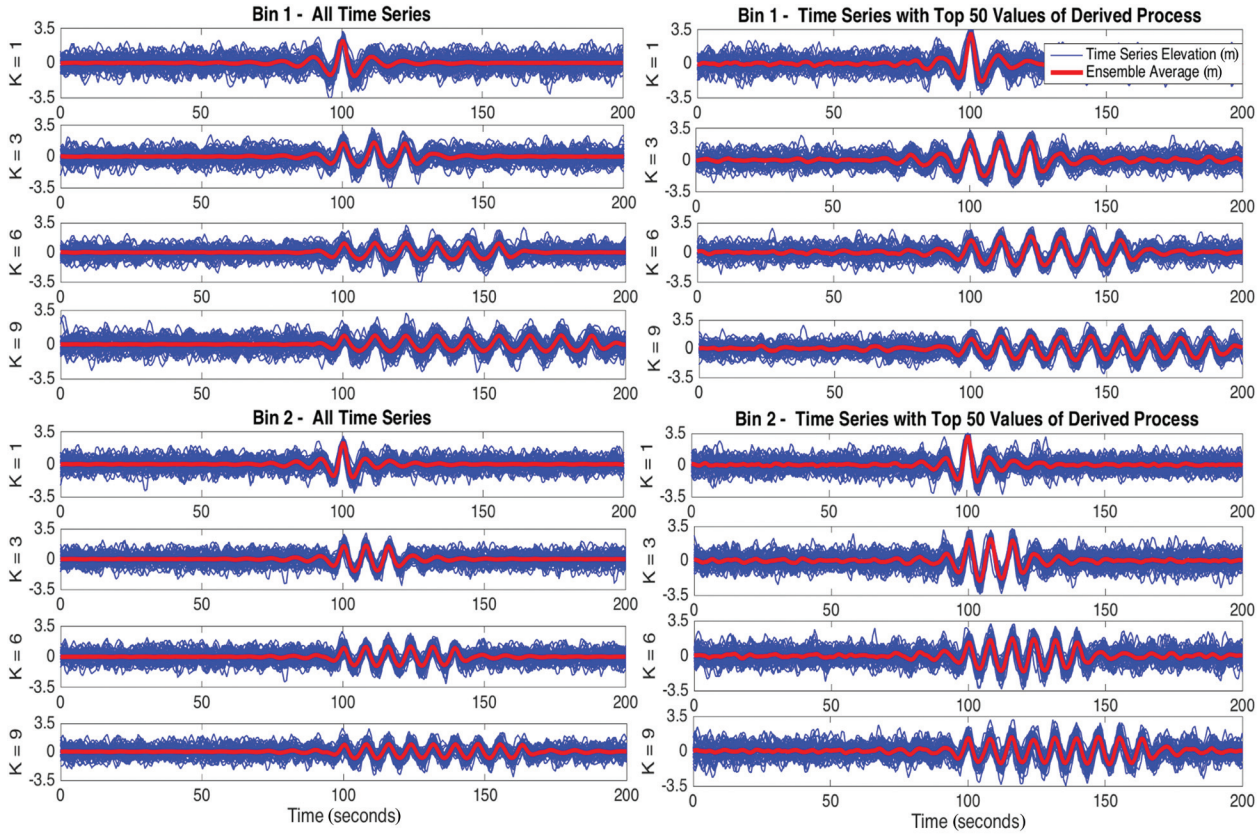

Fig. 6 Left panel: wave groups and ensemble average for all time series (in meters) with arbitrary 50 time series plotted for $k = 1, 3, 6$, and 9 for bins 1 and 2. Right panel: time series containing top 50 maxima of derived process for $k = 1, 3, 6$, and 9 and ensemble average for bins 1 and 2. t_o, time of maximum $z_k(t)$, shifted to 100 s without loss of generality.

different wave elevation time series except in the time period immediately following t_0, i.e., $t_0 < t < k\tau$. In addition, the mean group amplitude matches the trends shown in Fig. 4 where wave groups of longer runs have progressively smaller mean amplitudes. The mean wave group amplitudes for bin 1, Fig. 6 right column, compared to those in bin 2, Fig. 6 right column, for the same value of k, are noticeably larger since they represent the set of wave groups with approximate probability of exceedance (PoE) of $50/3350 = 1.49\%$ compared to a PoE of $50/2340 = 2.14\%$, for bins 1 and 2, respectively.

Time series with the maximum value of the derived process for each bin is also shown in Fig. 7. Wave groups for index $k = 1$, 3, 6, and 9 are shown along with the ensemble average of time series containing the top 50 values of the derived process. Without the superposition of the ensemble average, it would be difficult to identify the individual wave groups from a single realization. In addition, the wave group, $k = 6$ for bin 2, fails to meet the threshold criteria of the traditional wave group definition. This implies that the wave group with average amplitude, $\hat{z}_6/6$, shown in Fig. 7, would be deemed less probable if a threshold requirement of $\hat{z}_6/6$ was used.

Order Statistics and Extreme Values. Continuing in this section, order statistics and extreme value PDFs based on block maxima for the two bins are presented. The overall objective is to determine which statistical models reasonably approximate the various distributions associated with rare wave groups. It should then be possible to estimate the probability of encountering large amplitude wave groups for a given lifetime exposure to a specific sea state.

Figure 5, with the low D_{KL} values, suggests that it is reasonable to assume a normal process for the initial wave elevation. However, this modeling assumption may not necessarily be valid for

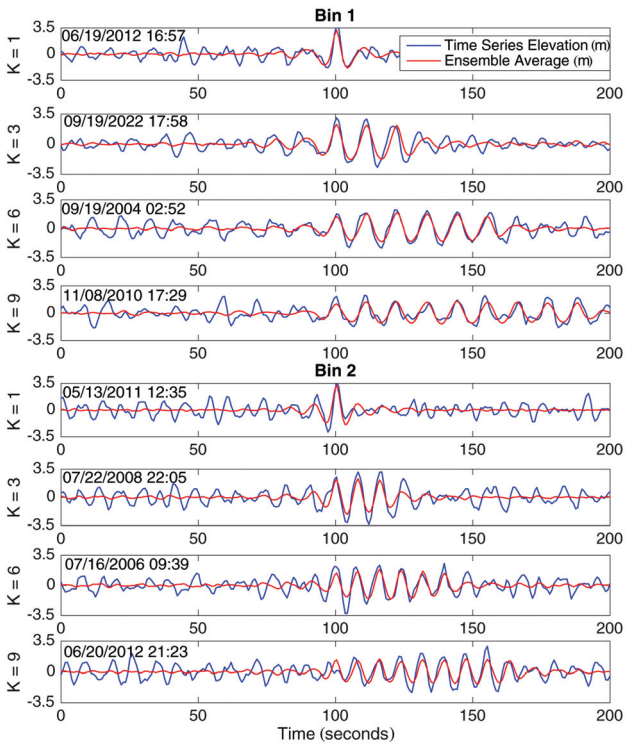

Fig. 7 Representative time series (in meters) containing maximum of derived process, for $k = 1$, 3, 6, and 9 and (top 50) ensemble average for bins 1 and 2. t_o, time of maximum $z_k(t)$, shifted to 100 s without loss of generality.

the tails of the original process histogram. While this area of research is ongoing, it is possible to show preliminary comparisons.

The genesis for the derived process $z_k(t)$ defined in Eq. (1) and the subsequent realization that maxima of $z_k(t)$ occur at times corresponding to rare wave groups was presented in Ref. [1]. Consistent with many marine processes (e.g., St. Dennis and Pierson [12]), the design loads generator methodology (Alford et al. [13], Kim et al. [14], and Kim and Troesch [1]) is based on the assumption that the seaway can be approximated as a zero-mean, Gaussian process. It follows that the derived process of Eq. (1) is the sum of normal processes, and thus a normal process itself. Given the variance of $z_k(t)$, then, it should be possible to estimate various extreme value statistics and distributions.

If the variance of a zero-mean random process, x, is defined as σ^2, then a normalized standard random variable is $\xi = x/\sigma$ and the initial standardized normal distribution (i.e., PDF) is given as $f(\xi)$

$$f(\xi) = \frac{1}{\sqrt{2\pi}} e^{-}\left(\xi^2/2\right) \tag{3}$$

The initial cumulative distribution function $F(\xi)$, or more commonly for a standardized normal PDF, $\Phi(\xi)$, becomes

$$F(\xi) = \Phi(\xi) = \int_{-\infty}^{\xi} d\zeta f(\zeta) \tag{4}$$

Following classical order statistics for iid random variables (e.g., Ochi [15]), the PDF and cumulative distribution function of extreme values are defined as $g(y_m)$ and $G(y_m)$, respectively. Here, y_m is the largest of m observations of ξ. The relationships between $g(y_m)$ and $G(y_m)$ and initial distributions $f(\xi)$ and $F(\xi)$ are

$$g(y_m) = mf(y_m)\{F(y_m)\}^{m-1} \tag{5}$$

and

$$G(y_m) = \{F(y_m)\}^m \tag{6}$$

The "most probable" extreme value is defined as the peak in the $g(y_m)$ distribution and denoted as $\widetilde{y_m}$. If both m and $\widetilde{y_m}$ are large, it can be shown as (e.g., Ochi [15])

$$1 - F(\widetilde{y_m}) = \frac{1}{m} \tag{7}$$

Note that Eq. (7) is valid only in an asymptotic sense and may be a poor approximation for the maxima associated with the relatively short 30 min time histories used in this work. Therefore, to find the value of m that best describes a normal extreme value distribution for the 30-min maxima, the following strategy was employed. Equation (5) was evaluated for a range of m values using the derived process ensemble variance. The normal-based $g(y_m)$ was then compared with the ensemble extreme value histogram using the Kullback–Leibler Divergence (i.e., Eq. (2)).

The best fit was deemed to be that value of m which resulted in a minimum value of D_{KL}. One example is Fig. 8 where the bin 1 30-min extreme value distribution for $k = 1$ is shown. The values of m and D_{KL} for this case are 1548 and 0.0484, respectively. A similar analysis was completed for all the derived processes in bins 1 and 2, $k = 1.15$ and the results are shown in Table 3.

In order to access the applicability of a normal extreme value distribution model for rare wave group statistics, several statistical metrics of interest can be determined from the data:

(1) Ensemble standard deviation of the original derived process, $z_k(t)$, defined as σ_{z_k}.
(2) Empirical histogram of the derived process 30-min maxima, $\hat{z}_k$.
(3) Mean of the empirical 30-min derived process maxima, defined as $\overline{\hat{z}_k}$.

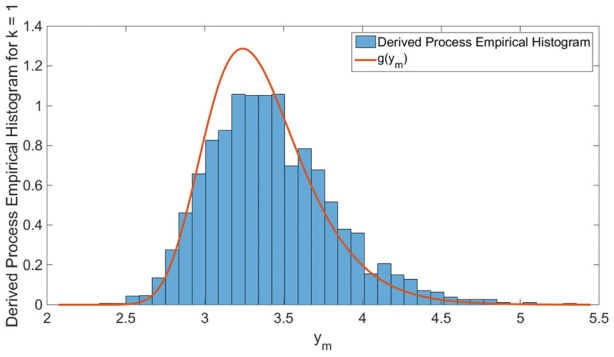

Fig. 8 Extreme value distribution for $k = 1$ bin 1 with D_{KL}: 0.0484 ($y_m = \hat{z}_1/\sigma_1$: $m = 1548$; $N = 3350$)

Table 3 Order statistics for bins 1 and 2

k	σ_{z_k}/k	$\overline{\hat{z}_k}/\sigma_{z_k}$	m_k	D_{KL_k}	$\overline{\hat{z}_k}/(\sigma_{z_k} \times \overline{y_{m_k}})$
			Bin 1		
1	0.693	3.424	1548	0.048	1.018
2	0.553	3.354	1164	0.064	1.021
3	0.478	3.297	902	0.076	1.027
4	0.428	3.241	715	0.090	1.031
5	0.392	3.198	617	0.092	1.032
6	0.363	3.162	534	0.103	1.034
7	0.340	3.132	473	0.115	1.037
8	0.321	3.102	428	0.113	1.038
9	0.305	3.072	384	0.116	1.039
10	0.291	3.047	354	0.112	1.040
11	0.279	3.021	325	0.113	1.041
12	0.267	2.994	298	0.117	1.041
13	0.258	2.974	277	0.118	1.043
14	0.249	2.955	263	0.113	1.042
15	0.241	2.935	247	0.110	1.043
			Bin 2		
1	0.703	3.478	1855	0.062	1.019
2	0.567	3.360	1251	0.044	1.017
3	0.496	3.302	999	0.050	1.019
4	0.443	3.260	840	0.056	1.022
5	0.403	3.230	755	0.061	1.022
6	0.371	3.198	668	0.065	1.024
7	0.345	3.171	602	0.072	1.025
8	0.324	3.144	538	0.074	1.028
9	0.305	3.117	491	0.074	1.028
10	0.290	3.100	464	0.074	1.029
11	0.277	3.082	437	0.075	1.029
12	0.265	3.060	406	0.075	1.029
13	0.255	3.043	382	0.075	1.030
14	0.246	3.026	359	0.076	1.031
15	0.238	3.012	343	0.078	1.031

(4) Expected extreme value, based on $g(y_{m_k})$ (Eq. (5)), defined as $\overline{y_{m_k}}$.

The derived process maxima, $\hat{z}_k$, which is based on the physical data samples, corresponds to the normal random variable $y_{m_k}\sigma_{z_k}$. In this sense, then, the mean of the empirical histogram, $\overline{\hat{z}_k}$, can be compared to the expected value $E[y_{m_k}]$ defined as $\overline{y_{m_k}}$.

Table 3 contains the following statistical information for each wave group index k in bins 1 and 2: the normalized ensemble root mean square (RMS) for $z_k(t)$, σ_{z_k}/k; the histogram mean, $\overline{\hat{z}_k}$, normalized by the ensemble RMS; the value of m corresponding to the minimum D_{KL} value; the minimum D_{KL} value; and the histogram mean normalized by the corresponding Gaussian expected extreme value.

From Table 3, we may make the following observations:

(1) Column 2, σ_{z_k}/k, follows the similar trend shown in Fig. 4. The RMS of the derived processes, normalized by the wave group index, k, decreases with increasing k.

(2) Column 3, $\overline{\hat{z}_k}/\sigma_{z_k}$, shows the effective rarity of the kth wave group in 30 mins. The reduction in $\overline{\hat{z}_k}/\sigma_{z_k}$ as k increases is explained, in part by a finite record (30 mins) and the "nature" of the derived process (z_k), including, but not limited to, its tendency to shorten the record as k increases. However, the empirical data of Table 3 exhibit a slightly faster decrease than numerical simulations. Research is ongoing to better understand how statistics of wave groups with longer runs (large k values) differ from the statistics of wave groups with shorter runs (small k values) for similar exposure times.

(3) Column 4, m_k, shows a trend similar to that of $\overline{\hat{z}_k}/\sigma_{z_k}$, in that the rare wave groups of longer runs seem to be disproportionally smaller than wave groups of shorter runs.

(4) Column 5, D_{KL_k}. There are two points to consider with the minimum D_{KL}: (i) How close are the extreme value histograms to normal extreme value PDFs? and (ii) How does minimum D_{KL} vary with wave index k?

   i. Prior to comparing the 30-min ensemble histograms with a normal extreme value PDF through use of the D_{KL} value, an estimate of the effect of finite data (e.g., $N = 3350$ samples) was made. Extensive numerical experiments were conducted using a random number generator based on $g(y_m)$, Eq. (5). Sixteen hundred (1600) normally distributed extreme value histograms, each containing 3350 samples, were constructed and compared through the Kullback–Leibler divergence (Eq. (2)) with an analytic form of $g(y_m)$. The mean D_{KL} value and the standard deviation of the D_{KL} estimates were 0.0042 and 0.0012, respectively. From the values of the minimum D_{KL} values in Table 3, we can see that the buoy data are less of a reasonable fit to the extreme value PDF of a normal process than that shown in Fig. 5, where a normal fit to the original process was quite good. However, Fig. 8 graphically illustrates that an empirical histogram with D_{KL} value of $O(0.05)$ is qualitatively captured over much of the range of the normal extreme value PDF, $g(y_m)$.

   ii. The results for both bins 1 and 2 show that the D_{KL} divergence increases with increasing wave run index, k, although the relative increase is significantly less pronounced for bin 2 than for bin 1. This may be due to the quality of the data in the respective bins or may be due to trends related to wave steepness. Research in this area is continuing.

(5) Column 6, $\overline{\hat{z}_k}/(\sigma_{z_k} \times \overline{y_{m_k}})$, shows that while the D_{KL} divergence comparison suggests a small, but consistent, deviation from a normal extreme value PDF, the mean of the empirical histogram is approximated by the expected value of an equivalent normal process to within 2–4%.

Extreme Value Estimates for Exposure Periods Larger Than 30 mins. The previous subsection, Order Statistics and Extreme Values, confirmed that estimates of rare wave group amplitudes and runs may be modeled as a normal process. However, this conclusion is based on examining maximum wave groups within 30-min blocks of time. It is reasonable then to consider how well the normal model would predict extreme wave group characteristics for longer exposure periods.

As one example, consider the following strategy to estimate the mean of the largest wave group amplitudes in 25 hrs. By assuming that the N realizations of the block maxima (e.g., $N = 3350$ for bin 1), $\hat{z}_k$, are associated with iid random variables, it is possible to combine 50 $z_k(t)$ time series, at random, into a representative 25-hr record.

While there are several ways to approximate the statistics of that 25-hr maximum, the approach used here is to determine the mean value of $\hat{z}_k$ samples whose PoE is defined as $\alpha = 1/50 = 0.02$. This is accomplished by sorting the empirical $\hat{z}_k$ values in ascending order and taking the average of the top 2% values; 67 of 3350 values for bin 1 and 49 of 2430 values for bin 2. For comparison with the analytical extreme value PDF, $g(y_m)$, the expected value of y_m whose PoE is α, $\overline{y^{\alpha}_{m_k}}$, is also calculated using the following relation:

$$\overline{y^{\alpha}_{m_k}} = \frac{1}{\alpha}\int_{y^{\alpha}_{m_k}}^{\infty} z \times g(z)\,\mathrm{d}z \tag{8}$$

Here, $y^{\alpha}_{m_k}$ is the y_m value whose PoE is α. The results are shown in Table 4. Similar to the notation used in Table 3, the histogram derived process random variables with probability α/m are denoted as $\hat{z}^{\alpha}_k$, and the normal extreme value random variables, with probability α/m, as $y^{\alpha}_{m_k}$.

Table 4 contains the following statistical quantities for both bins 1 and 2: the mean of the 30-min derived process extreme value whose PoE is α, $\overline{\hat{z}^{\alpha}_k}$, normalized by the ensemble variance; $\overline{\hat{z}^{\alpha}_k}$ normalized by k times the significant wave amplitude; and $\overline{\hat{z}^{\alpha}_k}$ normalized by $\sigma_{z_k} \times y^{\alpha}_{m_k}$. The following observations may be made, based on the results shown in Table 4:

(1) Column 2, $\overline{\hat{z}^{\alpha}_k}/\sigma_{z_k}$ (the normalized mean extreme values of derived processes, with a PoE of $\alpha = 0.02$) range from 4.4 σ to 4.7 σ. This compares with the 30-min maxima shown in Table 3 of 2.9–3.5 σ, and thus illustrates the effect of increasing exposure times by a factor of 50.

(2) Column 3, $\overline{\hat{z}^{\alpha}_k}/(kH_s/2)$, is a measure of the average wave amplitude in the kth wave group relative to the significant wave amplitude $H_s/2$. For a 25-hr exposure, rare wave groups with run lengths of 1–9 have average amplitudes that exceed the significant wave amplitude by factors ranging from 1 ($k = 9$) to 2.3 ($k = 1$).

Table 4 Order statistics for time series with derived process values: PoE 2%

k	$\overline{\hat{z}^{\alpha}_k}/\sigma_{z_k}$	$\overline{\hat{z}^{\alpha}_k}/\frac{kH_s}{2}$	$\overline{\hat{z}^{\alpha}_k}/(\sigma_{z_k} \times \overline{y^{\alpha}_{m_k}})$
		Bin 1	
1	4.661	2.349	1.013
2	4.540	1.825	1.021
3	4.621	1.606	1.048
4	4.590	1.430	1.030
5	4.630	1.319	1.068
6	4.631	1.223	1.053
7	4.608	1.140	1.051
8	4.577	1.068	1.067
9	4.584	1.016	1.060
10	4.562	0.965	1.060
11	4.539	0.919	1.064
12	4.518	0.879	1.070
13	4.508	0.846	1.065
14	4.492	0.814	1.076
15	4.489	0.788	1.062
		Bin 2	
1	4.749	2.377	1.020
2	4.565	1.843	0.996
3	4.539	1.601	1.009
4	4.532	1.429	1.018
5	4.509	1.293	0.998
6	4.486	1.184	1.035
7	4.481	1.100	1.018
8	4.490	1.034	1.014
9	4.498	0.978	1.028
10	4.527	0.935	1.053
11	4.537	0.894	1.036
12	4.525	0.854	1.070
13	4.505	0.818	1.048
14	4.481	0.784	1.067
15	4.429	0.749	1.062

(3) Column 4, $\overline{\hat{z}_k^{\alpha}}/(\sigma_{z_k} \times \overline{y_m^{\alpha}})$, compares the empirical mean extreme values, $\hat{z}_k^{\alpha}$, with corresponding values estimated using a normal extreme value distribution, $\overline{y_{m_k}^{\alpha}}$. These values indicate that empirical values, for $\alpha = 0.02$, exceed the normal theoretical values by 7% or less. The extent of the overprediction is a function of the wave group index k, ranging from less than 1% for $k = 1$ to approximately 7% for the larger values of k.

Conclusions

In this paper, time series elevation data from the Pt. Reyes buoy were examined to determine to what extent rare wave groups exist in nature. A definition of a rare wave group, defined by Kim and Troesch, as determined by a derived process similar to a moving average, i.e., Eq. (1), was used as the basis for identifying regions where rare wave groups could be found. Such a definition removes the more stringent criteria of a strict threshold crossing and considers more likely wave group events that can also cause extreme dynamic events. With the derived process, wave groups of 1–15 waves were identified. Considering two different bins with the same range of significant wave heights but with different ranges of peak modal periods gives a new understanding of the grouping of waves. A probability density function of a representative time series was calculated and compared to a Gaussian distribution in terms of its skewness, kurtosis, and Kullback–Leibler divergence. This process provided a means to estimate the size, run length, and return period of wave groups.

For the data and environmental conditions considered here, the normal distribution appeared to yield a reasonable model for the initial distribution and subsequent extreme value distribution. This analysis represents a clear and promising method for identifying rare wave groups and their application to marine platform design.

Acknowledgment

The work in this paper was supported by the Office of Naval Research (ONR) through the Stochastic Environments Leading to Design Responses in Nonlinear Seakeeping Simulations project (program manager Kelly Cooper) and the University of Michigan, through the Rackham Graduate School Merit Fellowship.

Nomenclature

N = number of samples in the empirical histogram or ensemble average
y_m = ordered set of $\hat{z}_k$ in ascending order of magnitude
z_{kj} = individual jth 30-min time series derived process of k waves
$\hat{z_{kj}}$ = value of largest derived process for k waves in jth 30-min time series
$\overline{\hat{z}_k}$ = ensemble average (mean) of derived process maxima for a 30-min time series
η = wave elevation as a function of time

References

[1] Kim, D., and Troesch, A., 2013, "Statistical Estimation of Extreme Roll Response in Short Crested Irregular Seas," Transactions of SNAME 2013, pp. 123–160.
[2] Longuet-Higgins, M., 1957, "The Statistical Analysis of a Random, Moving Surface," Philos. Trans. R. Soc. London, **249**(966), pp. 321–387.
[3] Goda, Y., 1976, "On Wave Groups," Offshore Structures (BOSS), Vol. 1, pp. 115–127.
[4] Kimura, A., 1980, "Statistical Properties of Random Wave Groups," 17th Conference on Coastal Engineering, pp. 2955–2973.
[5] Liu, Z., Elgar, S., and Guza, R., 1993, "Groups of Ocean Waves: Linear Theory, Approximations to Linear Theory, and Observations," Coastal Eng., **1992**(2), pp. 114–159.
[6] Masson, D., and Chandler, P., 1993, "Wave Groups: A Closer Look at Spectral Methods," Coastal Eng., **20**(3–4), pp. 249–275.
[7] Ochi, M. K., 1998, *Ocean Waves* (Cambridge Ocean Technology Series), Cambridge University Press, Cambridge.
[8] Tucker, M., and Pitt, E., 2001, *Waves in Ocean Engineering* (Elsevier Ocean Engineering Series), Elsevier Science, Oxford.
[9] CDIP, "Station 029," Pt. Reyes Buoy, CA, http://cdip.ucsd.edu/?nav=recent&stn=029&xitem=info.
[10] Otnes, R., and Enochson, L., 1978, *Applied Time Series Analysis: Basic Techniques*, Vol. 1, Wiley, New York.
[11] Kullback, S., and Leibler, R., 1951, "On Information and Sufficiency," Ann. Math. Stat., **22**(1), pp. 79–86.
[12] St. Dennis, M., and Pierson, W. J., 1953, "On the Motions of Ships in Confused Seas," Trans. SNAME, **61**, pp. 302–316.
[13] Alford, L., Kim, D., and Troesch, A., 2011, "Estimation of Extreme Slamming Pressures Using the Non-Uniform Fourier Phase Distributions of a Design Loads Generator," Ocean Eng., **38**(5–6), pp. 748–762.
[14] Kim, D., Engle, A., and Troesch, A., 2011, "Estimates of Long-Term Combined Wave Bending and Whipping for Two Alternative Hull Forms," Transactions of SNAME 2011, Vol. 119.
[15] Ochi, M., 1990, *Applied Probability and Stochastic Processes in Engineering and Physical Sciences*, Wiley, New York.