Generating an ensemble of mutually exclusive and exhaustive waves targeted for extreme responses

Seyffert, H. C.

DOI
10.1016/j.oceaneng.2021.110172

Publication date
2022

Document Version
Final published version

Published in
Ocean Engineering

Citation (APA)
Seyffert, H. C. (2022). Generating an ensemble of mutually exclusive and exhaustive waves targeted for extreme responses. *Ocean Engineering*, *243*, Article 110172. https://doi.org/10.1016/j.oceaneng.2021.110172

Ocean Engineering 243 (2022) 110172

Contents lists available at ScienceDirect

Ocean Engineering

journal homepage: www.elsevier.com/locate/oceaneng

Generating an ensemble of mutually exclusive and exhaustive waves targeted for extreme responses

H.C. Seyffert

Department of Maritime & Transport Technology, Delft University of Technology, The Netherlands

ARTICLE INFO

Keywords:
Joint extreme values
Rare responses
Targeted simulation
Long return-periods
Rare wave groups

ABSTRACT

To examine the reliability or performance of marine systems related to dynamic time-varying responses, time domain simulation may be required. But for long durations, brute-force simulation is clearly not feasible from a computational point of view to examine converged extreme value statistics. An additional challenge is defining a set of mutually exclusive and exhaustive wave excitation records which excite the desired extreme response(s). To address these challenges, this work develops a non-linear Design Loads Generator (NL-DLG) process which links wave profiles targeted for extreme responses generated via response-conditioning wave techniques to a probabilistic framework which examines the possible correlation or mutual exclusivity between those wave profiles. The end result is an ensemble of short targeted wave profiles which excite converged extreme value return-period statistics of a defined response that may be excited by multiple correlated processes. The method of the NL-DLG process is explained by applying it towards rare wave groups, where the result is an ensemble of short wave profiles which contain extreme occurrences of different wave groups targeted for the defined exposure duration. These generated wave excitation profiles are compared to physical data from the Pt. Reyes buoy and then numerical simulations are employed to consider rarer events.

1. Introduction

Considering extreme ocean environments is necessary for understanding the associated risk and performance of marine systems which operate in such climates. To link hydrostatic and structural codes or to consider dynamic time-varying responses the wave excitation profiles which excite extreme events must also be known. As we examine rarer events, the challenge is how to appropriately design testing regimes to evaluate system performance when there are multiple relevant inputs to consider that may in fact be correlated and whose correlation may change when considering more extreme joint occurrences. There may also be interest in examining the extreme environments themselves, particularly wave forms expected to excite extreme dynamic responses. The desire to retain the wave profiles associated with the extreme events suggests a simulation-based approach, but brute-force simulation clearly becomes infeasible for longer return periods.

To address these challenges, this paper presents a non-linear Design Loads Generator (NL-DLG) process to assemble an ensemble of wave profiles targeted for a defined response for long return-periods. As an example of such a response, consider vessel parametric roll, which may be excited by group-like behavior in the ocean excitation: specifically wave groups where the period between successive wave crests is nearly half the vessel roll natural period. Here, nonlinear time domain simulation may be crucial for assessing the risk of extreme roll responses, but as noted by Themelis and Spyrou (2007) and Kim and Troesch (2015), identifying an ensemble of mutually exclusive and exhaustive wave groups expected to lead to extreme vessel parametric roll responses is extremely challenging. Kim and Troesch further note the difficulty of defining mutually exclusivity for random time histories, and specifically of calculating the joint occurrence of different wave groups.

Motivated by such challenges, the presented NL-DLG process utilizes a response-conditioning wave technique to generate specific wave sequences targeted for extreme responses, then applies a probabilistic framework to construct an ensemble of exhaustive and mutually exclusive wave profiles targeted for a defined return-period extreme response. The NL-DLG process, extended and completely formalized from its original presentation in (Seyffert, 2018), is illustrated by constructing wave time series which contain rare wave groups and excite distributions of joint extreme rare wave group occurrences. The clear follow-up is applying these joint wave group occurrences to parametric roll responses, though for now only the problems of defining mutually exclusive targeted wave excitations and joint wave group occurrences are considered, while parametric roll excited by such wave groups is left to future work.

To determine the quality of the resulting ensemble of NL-DLG wave profiles, distributions of joint extreme rare wave group occurrences

E-mail address: H.C.Seyffert@tudelft.nl.

https://doi.org/10.1016/j.oceaneng.2021.110172
Received 25 March 2021; Received in revised form 21 October 2021; Accepted 6 November 2021
Available online 20 December 2021

resulting from the NL-DLG wave profiles are compared with empirical distributions measured from physical oceanographic data. Then, brute-force simulations are used for comparison with a longer return period.

This work is structured as follows: Section 2 discusses the relevant background literature. Section 3 introduces the physical oceanographic data which provides the comparison with the wave elevation time series generated via the NL-DLG process. Section 4 presents the method to assemble the ensemble of targeted wave profiles and resulting joint occurrences of extreme responses, which is illustrated via distributions of joint extreme wave group occurrence in Section 5. Finally, Section 6 offers some conclusions and the accompanying appendix gives an in-depth description of the probabilistic framework used within the NL-DLG process.

2. Background

The NL-DLG process generates an ensemble of exhaustive and mutually exclusive wave profiles targeted for rare wave group occurrences, where the wave ensemble statistics mimic the converged extreme value statistics of brute-force simulations for a defined return-period. This section therefore provides relevant background on wave groups (which provide the application example to describe the NL-DLG process), response-conditioning wave techniques (which are used within the NL-DLG process to target for extreme responses), and joint distributions of random variables (which are a product of the NL-DLG generated wave profiles).

2.1. Identification of wave groups

Many definitions of wave groups exist, based on, *e.g.*, Gaussian random processes, envelope threshold crossings and generally narrow-banded spectra (Longuet-Higgins, 1957; Ewing, 1973), a narrow-banded assumption with possible dependence between successive waves conditioned on waves exceeding the significant wave height, H_s (Rye, 1974; Goda, 1976), linear theory with narrow-banded spectra (Liu et al., 1993), and the theory of quasi-determination (Boccotti, 2015; Fedele, 2005a,b). But as discussed by Bassler et al., wave groups which are restricted to a successive threshold crossing requirement will not include more probable wave group sequences allowing possible down-crossings below the required threshold (Bassler et al., 2010). These groups of waves, which may satisfy the threshold requirement in a mean sense but not in total, can cause significant system responses through resonant excitation, since this definition can then include some prescribed time between successive wave peaks, which when tuned to a vessel natural period can excite parametric resonance.

Such wave group occurrences can be identified using a Gaussian derived process $\mathrm{L}_j(t)$, which is a characteristic process of the ocean excitation that indicates group-like behavior of the input, namely the occurrence of a wave group of j waves (Kim et al., 2014). This expression identifies rare wave group occurrences by identifying the maximum value of $\mathrm{L}_j(t)$ over some exposure time, which targets for j large wave peaks (no threshold requirement) separated in time by a user-defined period τ, as in Eq. (1):

$$\mathrm{L}_j(t) = \sum_{p=1}^{j} \eta(t + (p-1)\tau) \tag{1}$$

where

$\mathrm{L}_j(t)$ = Gaussian derived process indicating group-like behavior, j waves

$\eta(t)$ = wave elevation at a specific spatial coordinate

τ = defined period of interest in seconds

j = wave group index (number of waves in group)

Using Eq. (1) as a filter on a wave excitation record $\eta(t)$ identifies potential instances of wave groups of j waves separated in time by τ seconds based on return-period extreme values of $\mathrm{L}_j(t)$, denoted here as $\hat{\mathrm{L}}_j$. It has been shown using Fourier Transform theory and the Wiener–Khintchine relations that the mathematical formulation for the expected shape of a wave group conditioned on extreme values of $\mathrm{L}_j(t)$ has a closed-form expression which matches numerical simulations and physical oceanographic data (Seyffert et al., 2016). This expected wave group shape for j waves is proportional to the sum of j autocorrelation functions of the wave spectrum, separated in time by $(p-1)\tau$ seconds, $p = 1, \ldots, j$. The constant of proportionality is the value of the expected extreme value of $\mathrm{L}_i(t)$ over the exposure period, $\hat{\mathrm{L}}_j$, normalized by the derived process variance, $\sigma^2_{\mathrm{L}_j}$. Thorough reviews of wave group statistics using this definition are given in Seyffert and Troesch (2016a), Seyffert et al. (2016).

2.2. Response-conditioning wave techniques

Response-conditioning wave techniques (RCWT) can be used to construct wave profiles that excite a specific linear response value at a pre-determined time. Consider a stochastic wave elevation time series $\eta(t)$:

$$\eta(t) = \sum_{m=1}^{N} a_m cos(\omega_m t + \phi_m) \tag{2}$$

where

$a_m = \sqrt{2S(\omega_m)\Delta\omega}$

$S(\omega)$ = single-sided wave energy spectrum

ϕ_m = phase between $-\pi$ and π

If the phases ϕ are uniformly distributed between $-\pi$ and π, as $N \to \infty$ the random variable H expressed by the random process $\eta(t)$ approaches a Gaussian random variable. RCWT's condition a wave to lead to a pre-determined response value at a pre-determined time, say $t = 0$. In this case, the phases ϕ are tuned based on a transfer function of the desired response. Consider now a response of interest $\mathrm{L}_j(t)$, say, a rare wave group of j waves like defined in Eq. (1), which is excited by the wave elevation $\eta(t)$:

$$\mathrm{L}_j(t) = \sum_{m=1}^{N} a(\omega_m) M_{\mathrm{L}_j}(\omega_m) cos(\omega_m t + \phi_m + \psi_m) \tag{3}$$

where

$a(\omega_m) = \sqrt{2S(\omega_m)\Delta\omega}$

$S(\omega)$ = single-sided wave energy spectrum

$H_{\mathrm{L}_j}(\omega) = \sum_{p=1}^{j} exp(i\omega(p-1)\tau)$

$M_{\mathrm{L}_j}(\omega_m) = |H_{\mathrm{L}_j}(\omega_m)|$

ϕ_m = wave phase between $-\pi$ and π from Eq. (2)

$\psi_m = arg(H_{\mathrm{L}_j}(\omega_m))$

RCWT's focus on defining the phases ϕ such that the resulting $\mathrm{L}_j(t = 0)$ response is a rare occurrence. Many different RCWT's exist to define such waves, *e.g.*, Lindgren (1970), Tromans et al. (1991), Taylor et al. (1997), Friis-Hansen and Nielsen (1995), Torhaug (1996), Adegeest et al. (1998) and Dietz (2004) and generally assume perfect phase alignment of the frequency components in Eq. (3). The RCWT used in this paper is the Design Loads Generator (DLG), which instead of assuming perfect phase alignment to focus the frequency components estimates a distribution of phases leading to extreme responses $\mathrm{L}_j(t = 0)$. This is accomplished by using a modified Gaussian distribution as the initial guess of this phase distribution leading to extreme responses (see Alford and Troesch, 2009) and an acceptance–rejection algorithm for ensuring that the chosen phases do in fact excite a distribution of return-period extreme responses (see Kim, 2012). The efficacy of the DLG as a RCWT was positively evaluated in Seyffert et al. (2020).

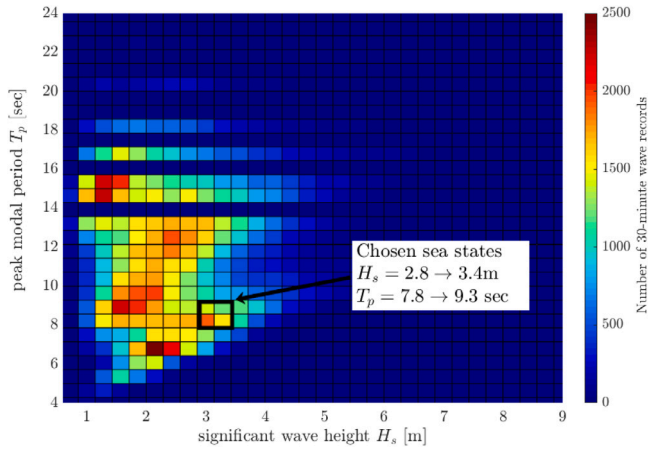

Fig. 1. Available 30-min records from Pt. Reyes buoy and chosen subset of records for analysis.

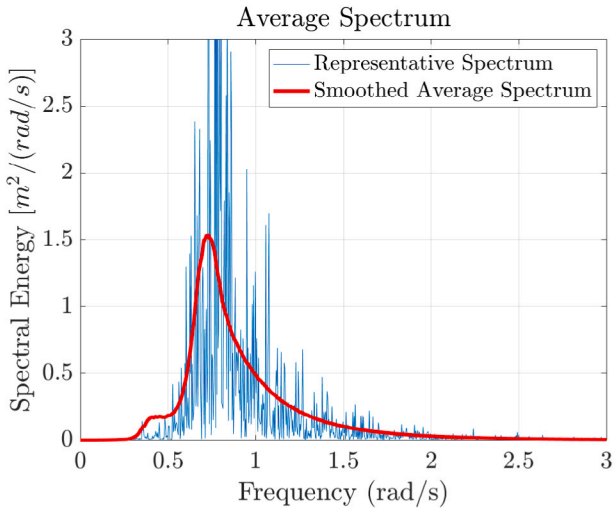

Fig. 2. Representative and smoothed average spectrum of examined Pt. Reyes buoy wave records.

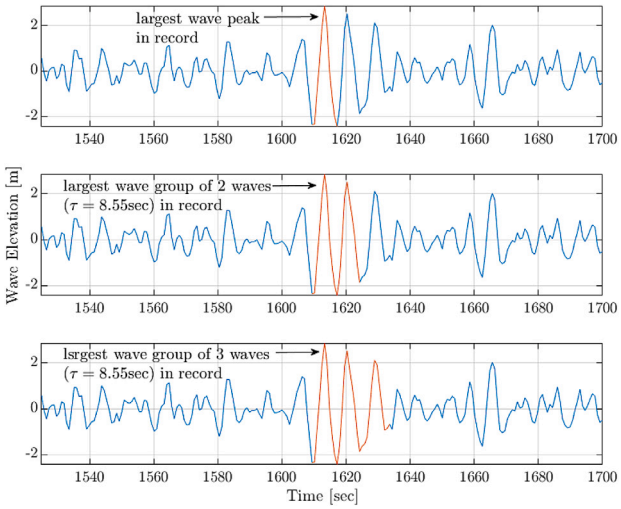

Fig. 3. Pt. Reyes buoy time series from January 23, 1997 08:52, showing clustered wave groups of 1, 2, and 3 peaks.

Table 1
Ranges of significant wave height H_s and peak modal period T_p for examined time series.

Parameter	Value
H_s Range	2.8 → 3.4 m
T_p Range	7.8 → 9.3 s
Number of 30-min time series	6052
Total record time	3026 h

2.3. Joint distributions of random variables

To determine whether the ensemble of wave profiles generated by the NL-DLG process do in fact mimic converged return-period statistics as from full-length realizations, we examine joint distributions of random variables which are excited by these generated time series. Specifically for the example of this paper, we examine joint distributions of rare wave group occurrences, with wave groups of j waves defined by Eq. (1) and with extreme wave group occurrences indicated by occurrences of $\hat{\mathrm{L}}_j$.

Many different methods exist to estimate joint distributions and joint extreme occurrences of random variables, *e.g.*, copula techniques (de Waal and van Gelder, 2005; Bastian et al., 2009; Bartoli et al., 2011; Vanem, 2020), a conditional extreme model (Heffernan and Tawn, 2004), the concepts of max-stability of asymptotic distributions (Ewans and Jonathan, 2014), multivariate extreme value modeling (Gouldby et al., 2017), statistical emulators (Jones et al., 2018), and of course environmental contour methods (Haver, 1987; Winterstein et al., 1993; Jonathan et al., 2014; Huseby et al., 2015; Vanem, 2017; Haselsteiner et al., 2017; Chai and Leira, 2018). But the desire of the NL-DLG process specifically is to assemble an ensemble of mutually exclusive and exhaustive short wave profiles targeted for rare responses which could be used for efficient simulations. Joint distributions of extreme responses are therefore a useful output and point of comparison for these generated targeted wave profiles.

3. Environmental wave data

To evaluate the NL-DLG wave profiles based on the generated distributions of joint wave groups occurrences (using the wave group definition from Eq. (1)), similar empirical distributions will be made from physical oceanographic data measured by the Pt. Reyes buoy, operated by the Coastal Data Information Program (CDIP, 0000). Detailed background about the operation of this buoy can be found in Seyffert and Troesch (2016a). Here a subset of the available data is examined, as indicated in Fig. 1 and described by Table 1, relating to 3026 h of buoy data. The smoothed average spectrum relating to these chosen buoy records is given in Fig. 2.

3.1. Joint occurrence of wave groups in physical oceanographic data

The NL-DLG process uses a RCWT to generate short wave profiles targeted for a specific response, then links these waves within a probabilistic framework which considers how those wave profiles are correlated, indicated by some joint extreme event occurrence. As an example of such a joint occurrence, consider Fig. 3, which shows a record from the Pt. Reyes buoy, captured on January 23, 1997 starting at 08:52. The top inset shows the wave elevation record in blue, with the largest wave peak in the 30-min record highlighted in red, identified by the time when $\mathrm{L}_1(t)$ from Eq. (1) is at a 30-min maximum, or when $\hat{\mathrm{L}}_1$ occurs. The middle inset shows the same wave record in blue, and highlights in red the largest wave group of 2 waves in this record, with the waves separated in time by 8.55 sec, the average peak modal period for the chosen subset of buoy records. This wave group is identified by the time of $\hat{\mathrm{L}}_2$ occurrence. The bottom inset shows the same record and when the largest wave group of 3 waves occurs, identified when $\hat{\mathrm{L}}_3$ occurs with $\tau = 8.55$ sec, highlighted in red.

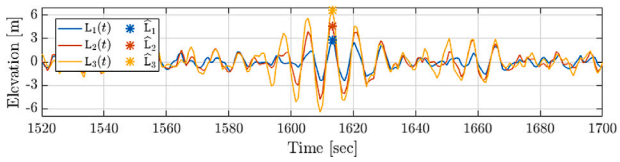

Fig. 4. $L_1(t)$, $L_2(t)$, and $L_3(t)$ with $\tau = 8.55$ sec from Pt. Reyes buoy time series January 23, 1997 08:52. Extreme values of the derived processes, $\hat{L}_1$, $\hat{L}_2$, and $\hat{L}_3$ indicate the occurrence of a rare wave group of 1, 2, and 3 peaks.

Fig. 4 further shows the time series of the derived processes $L_1(t)$, which is in fact the wave elevation record, $L_2(t)$, and $L_3(t)$ from the same Pt. Reyes buoy record from January 23, 1997 08:52, all with $\tau = 8.55$ sec. The derived process $L_j(t)$ acts as a moving-average filter on the wave elevation $\eta(t)$ based on the defined peak separation period τ. Extreme values of the derived process $\hat{L}_j$ indicate the start of a rare wave group of j waves separated in time by τ seconds.

Notice that the largest wave group of 3 waves for this 30-min record also contains the largest wave groups of 2 waves and 1 wave. This phenomenon highlights the need to consider the probability of joint wave group occurrence, as return-period extreme wave groups of different wave index (or number of wave peaks) may occur simultaneously. When considering the effects of wave groups on dynamic performance of marine systems, such as parametric rolling of ships (Shin et al., 2005), excitation of wave energy converters, or extreme pitch response of spar platforms (Seyffert and Troesch, 2016b), the probability of joint wave group occurrences may be required to examine overall system performance. The underlying wave profiles, like in Fig. 3 would therefore be useful for more in-depth non-linear time simulation tools.

The NL-DLG process generates such an ensemble of exhaustive and mutually exclusive wave profiles targeted for such extreme responses. Here, exhaustive means exhaustive over a defined range of potentially interesting excitation inputs (*e.g.*, wave groups of defined groups indices). It would indeed be useful to have an ensemble of wave profiles targeted for extreme wave group occurrences of varying group index, with a group period prescribed to a vessel natural period, like half roll, because wave groups of different wave index might all have some impact on extreme parametric responses, as noted by Kim and Troesch (2015). The need for mutually exclusive wave profiles is motivated by the fact that targeting for rare wave group occurrence might lead to repeating time series which excite extreme values of different wave group indices.

4. Methods

The aim of the NL-DLG process is to generate an ensemble of targeted mutually exclusive and exhaustive wave profiles which excite maxima of a number of surrogate processes relevant for a system response and relating to a defined return-period. Surrogate processes act as indicators of other processes of interest; *e.g.*, extrema of the derived (surrogate) process in Eq. (1) indicate occurrences of wave groups of j waves, and wave groups themselves are expected to be a surrogate process for extreme roll responses. This surrogate process is an easier way to identify group-like behavior than threshold crossings but still indicates when group-like behavior is expected in the underlying process.

The resulting NL-DLG wave profiles are called targeted because they focus specifically on the times in the wave excitation which are expected to excite extrema of the surrogate processes. The result is an ensemble of short wave profiles targeted for return-period maxima of defined surrogate processes, but with simulation time far shorter than the return-period. This is accomplished within the NL-DLG process by generating excitation time series targeted for surrogate process extrema by response-conditioning wave techniques and linking these excitation time series via a probabilistic framework which considers the possibility that extrema of different surrogate processes occur clustered together over the return-period, like in Fig. 3.

The full process is described in the Appendix and implemented in the accompanying code (Seyffert, 2021), but the steps are generally described here and illustrated in Figs. 5–6, continuing the example of considering 30-min maxima of wave groups of 1, 2, and 3 waves with separation period $\tau = 8.55$ sec. Note that the numbered steps in Figs. 5–6 relate to the steps in the Appendix.

4.1. Generation of excitation time series via response-conditioning wave techniques

Response-conditioning wave techniques, as first described in Section 2.2, are used in the NL-DLG process to generate an ensemble of wave profiles targeted for extrema of defined surrogate processes. In this example, the DLG is used as the RCWT and generates waves, called $\eta_j(t)$ which excite maxima of $L_j(t)$, that are (user-defined) 70 sec long, $t \in [-10, 60]$, with the extreme event at $t = 0$. The constructed wave profiles have the smoothed average wave energy spectrum measured from the Pt. Reyes buoy records, shown in Fig. 2, with 300 frequency components ranging from 0 to 2.45 rad/s.

Within the NL-DLG process, the DLG generates an ensemble of wave excitation records (Fig. 5 point (1)) which excite a distribution of 30-min extreme values of $L_1(t)$, $L_2(t)$, and $L_3(t)$ with $\tau = 8.55$ sec, $\hat{L}_1$, $\hat{L}_2$, and $\hat{L}_3$ (Fig. 5 point (1a)). The question is: do the DLG waves constructed for specific realizations of surrogate process exposure-period maxima also excite extreme realizations of other surrogate processes (Fig. 5 point (1b))? Fig. 3 suggests that 30-min extremes of $L_1(t)$, $L_2(t)$, and $L_3(t)$ may sometimes be clustered, though this is not expected to always be the case.

4.2. Probabilistic framework to link excitation time series

Therefore to ensure the RCWT-generated wave profiles are mutually exclusive and exhaustive, the NL-DLG process employs a probabilistic framework to determine the probability that the realizations of different surrogate process maxima cluster together over the exposure. Examining the ensemble of DLG waves, *e.g.*, $\eta_1(t)$ constructed to excite extrema of $L_1(t)$, it is possible to determine the probability that these waves also excite exposure-period maxima of $L_2(t)$ and $L_3(t)$; the same can be done for the $\eta_2(t)$ and $\eta_3(t)$ waves (Fig. 6 point (2) following notation in Appendix, Eq. (10)). By considering the possible ways that surrogate process maxima can cluster together over an exposure (the maxima configurations described in Fig. 6 point (3)) it is possible to relate the probabilities of how the surrogate process maxima cluster (Fig. 6 point (3a)) to the probabilities that an exposure fits one of the exposure-period maxima configurations (Fig. 6 points (3b)–(4)).

This is accomplished by combining probability spaces relating to the ensembles of $\eta_1(t)$, $\eta_2(t)$, and $\eta_3(t)$ wave record realizations into a single probability space representing how multiple surrogate process maxima realizations may potentially cluster over an exposure. This final probability space is based on the fact that over a single exposure, each surrogate process experiences a single maximum which may or may not occur clustered together with other surrogate process maxima.

The probabilistic framework of the NL-DLG process is used to combine information about surrogate process maxima clustering from multiple "experiments", *i.e.*, the ensembles of $\eta_1(t)$, $\eta_2(t)$, and $\eta_3(t)$. These DLG wave ensembles $\eta_i(t)$ each lead to a sample space (Eq. (13)) and collect some overlapping information; *e.g.*, the probability that extrema of $L_1(t)$ and $L_2(t)$ occur clustered is estimated from the ensemble of $\eta_1(t)$ and $\eta_2(t)$ waves (orange slices of pie charts, Fig. 6 point (2)). But these individual sample spaces do not capture the information regarding how *all* surrogate process maxima may occur clustered together over a single exposure; *e.g.*, the $\eta_1(t)$ waves by definition cannot say how extrema of $L_2(t)$ and $L_3(t)$ may occur clustered together but separate from an extrema of $L_1(t)$, over an exposure. This probability can only

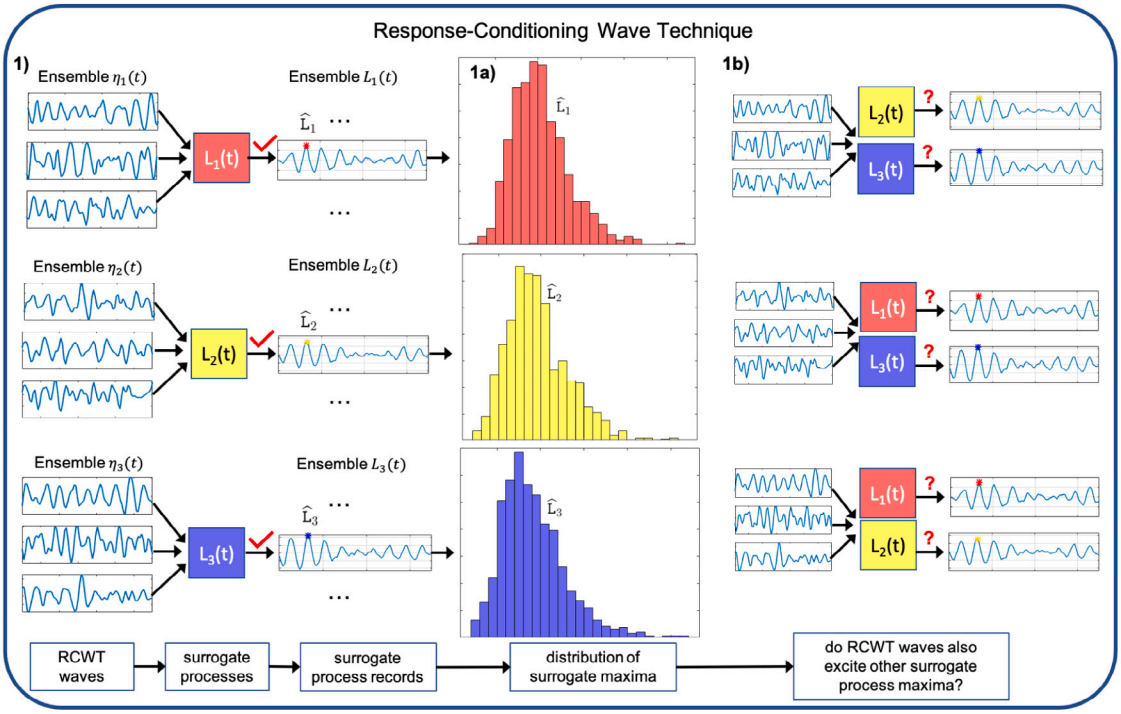

Fig. 5. Generation of ensemble of waves $\eta_1(t), \eta_2(t)$, and $\eta_3(t)$ from a RCWT (1) which excite a distribution of exposure-period maxima of the defined surrogate process (1a), and potentially also exposure-period maxima of other surrogate processes (1b).

be determined from the $\eta_2(t)$ and $\eta_3(t)$ waves (green slices of pie charts, Fig. 6 point (2)).

The probabilistic framework of the NL-DLG process therefore combines information from the individual sample spaces stemming from the $\eta_1(t)$, $\eta_2(t)$, and $\eta_3(t)$ wave record realizations which give overlapping information, but which individually cannot fully describe how all the surrogate process maxima may cluster over an exposure period. This framework results in some number, *num*T, of mutually exclusive and exhaustive wave profiles which target for surrogate process exposure-period-maxima but are significantly shorter than that return-period and are meant to mimic the statistics from *num*T full-length simulations. An algorithm to calculate the number *num*T is given in the Appendix.

5. Application to joint occurrence of rare wave groups

Using the DLG as a RCWT, the NL-DLG process assembles an ensemble of mutually exclusive and exhaustive short wave profiles targeted for specific rare wave group occurrences, based on a defined return-period and group indices of interest. These NL-DLG waves generate joint distributions of rare wave group occurrences, with each sample relating to a unique wave profile which targets for return-period statistics but is much shorter than the return-period. Such distributions of rare joint wave group occurrences are also assembled from the Pt. Reyes buoy data, and from brute-force simulations for longer return-periods for comparison.

5.1. 30-Minute wave group statistics

First the statistics relating to 30-min extreme wave groups are examined, as identified in the 30-min Pt. Reyes buoy records using Eq. (1) and as constructed by the DLG targeted for 30-min rare wave group statistics. 30-min extreme wave groups of j waves occur when $\mathrm{L}_j(t)$ attains its 30-min maximum value, denoted as $\hat{\mathrm{L}}_j$. The separation period between wave peaks is $\tau = 8.55$ sec.

From the Pt. Reyes buoy records, Eq. (1) is used to identify the largest wave group of j waves in each record. Then, the most-likely extreme 30-min value, the target extreme value or TEV, is estimated from the Pt. Reyes records by finding the Gaussian extreme value distribution with the best-fit to the empirical histogram. This best-fit is measured by the Kullback–Leibler divergence, which measures the information lost when one probability distribution is used to approximate another; the best possible fit is when the divergence is zero. In this way, it is possible to estimate the 30-min most-likely extreme wave group value, TEV, of the Pt. Reyes wave records. This TEV is used by the DLG to tune the magnitude of extreme responses excited by the generated ensemble of wave profiles.

Fig. 7 gives the empirical probability density function (pdf) and cumulative distribution function (cdf) of the 30-min maximum derived process values $\hat{\mathrm{L}}_j$ normalized by the respective standard deviation σ_{L_j} from the 1000 DLG waves $\eta_j(t)$ constructed for each surrogate process $\mathrm{L}_j(t)$ and from the 6052 Pt. Reyes buoy records for wave group indices $j = 1, 2, \ldots, 6$. Fig. 7 already indicates that the wave group statistics measured from the Pt. Reyes buoy records deviate from the expected Gaussian distributions assembled by the DLG, even more so for increasing wave group index j. This was also noted in Seyffert et al. (2016), that successive local maxima of $\mathrm{L}_j(t)$ and their arrival rates are not independent and their dependence increases with increasing wave group index j. This clustering of peaks of $\mathrm{L}_j(t)$ impacts the threshold-crossing rate (Wirsching et al., 2006) and therefore the estimation of extreme values.

Fig. 8 gives the average $\hat{\mathrm{L}}_j$ value normalized by σ_{L_j} and group index j from the Pt. Reyes buoy wave records ($\times$) and the DLG wave time series $\eta_j(t)$ generated for use within the NL-DLG process ($\circ$). This figure indicates that the relative difference in extreme values of the derived process, indicating the occurrence of a rare wave group, decreases for increasing wave group index j. Accordingly, one might expect that wave groups of higher group index j have a higher correlation than do wave groups of lower index, since the difference between $\hat{\mathrm{L}}_j/(j \times \sigma_{\mathrm{L}_j})$ diminishes for increasing j and wave sequences which might excite an extreme value $\hat{\mathrm{L}}_j$ might similarly excite extreme values of other $\hat{\mathrm{L}}_{j\pm1}$, $\hat{\mathrm{L}}_{j\pm2}$, etc. Fig. 8 therefore suggests that the dependence structure between different wave group indices will depend on those indices.

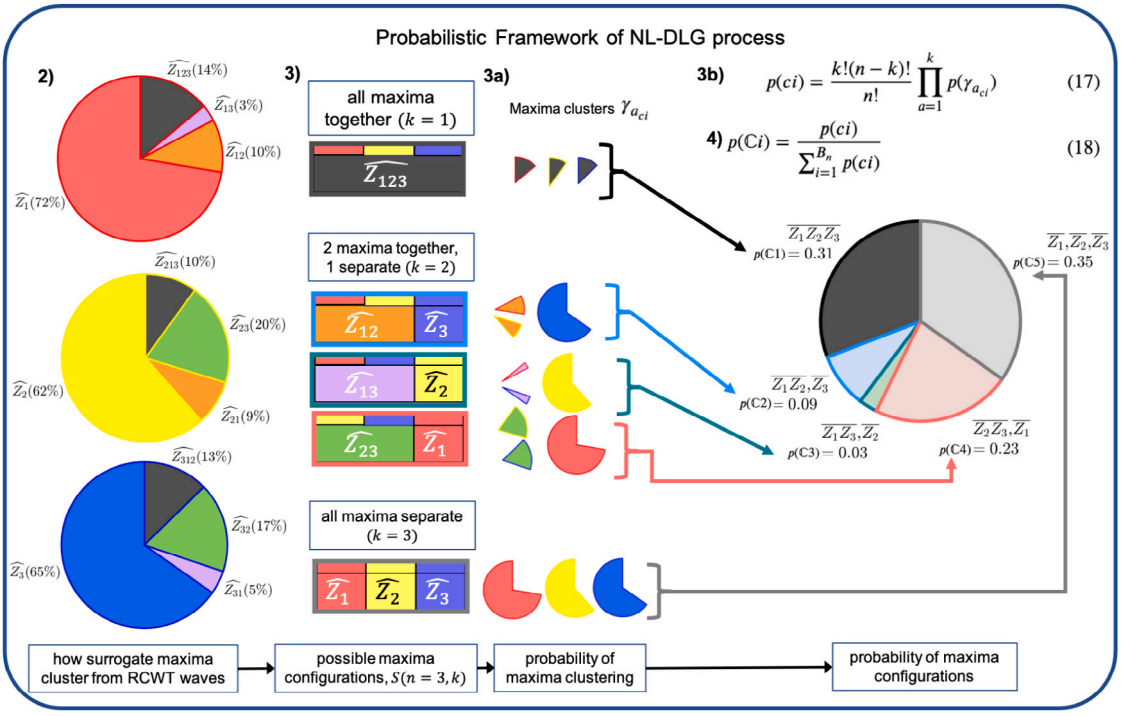

Fig. 6. Probabilistic framework within the NL-DLG process which determines how surrogate process maxima excited by $\eta_i(t)$ cluster with other surrogate maxima (2), then relating the possible maxima configurations defined for the number of surrogate processes n by the Stirling number of the second kind S(n,k) (3) with the probabilities of these surrogate maxima clustering (3a) to define the probability of experiencing these different maxima configurations over the exposure period (3b)–(4).

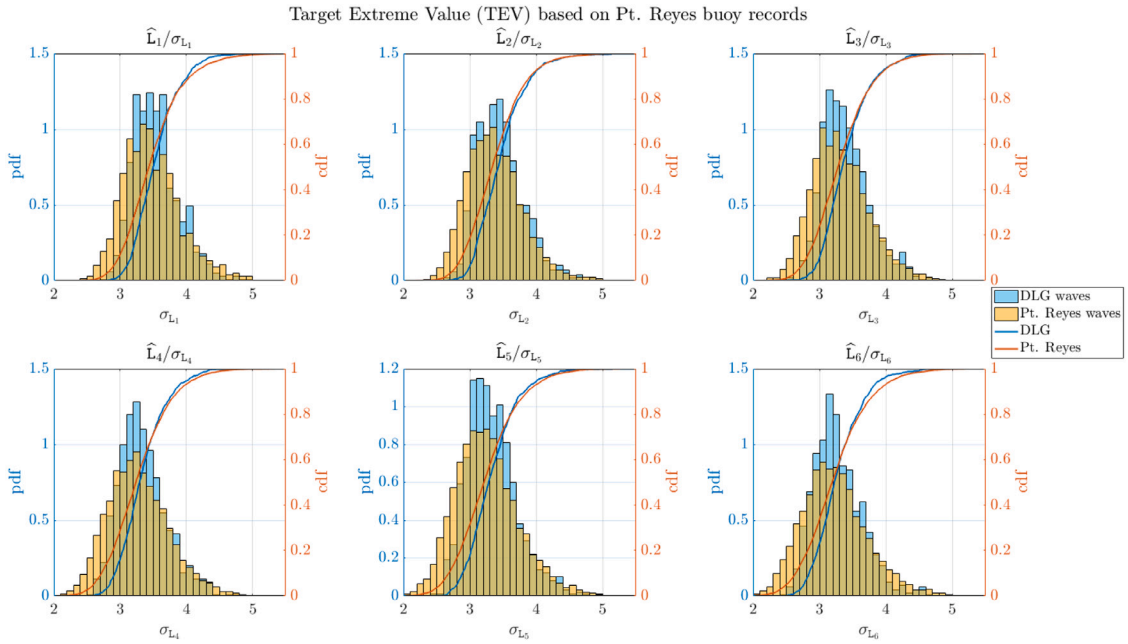

Fig. 7. Empirical pdf (left y-axis) and cdf (right y-axis) of 30-min maximum derived process values ($\hat{\mathrm{L}}_j$) normalized by the respective standard deviation (σ_{L_j}) from 1000 DLG waves $\eta_j(t)$ for each index j (with TEV based on Pt. Reyes buoy records) and 6052 Pt. Reyes buoy records.

5.1.1. Joint occurrence of 2 wave groups

Now, joint occurrences of wave groups of two different group indices j and k are considered, *e.g.*, whether the largest wave group of 2 waves and the largest wave group of 3 waves occurs during the same segment of a 30-min return-period so as to be considered a single event, like illustrated in Fig. 3. The NL-DLG process assembles 30-min joint wave group statistics first by constructing an ensemble of waves $\eta_j(t)$ from the DLG that excite $\hat{\mathrm{L}}_j$ values for $j = 1, 2, \ldots, 6$. Joint wave group statistics for wave groups of j and k waves are then examined via the NL-DLG process waves when considering only two surrogate processes, $\mathrm{L}_j(t)$ and $\mathrm{L}_k(t)$. The resulting ensemble of NL-DLG process wave realizations then gives a distribution of joint wave group statistics, where some of these extreme wave groups occur clustered together.

To construct the joint distributions of wave group occurrences from the Pt. Reyes buoy with a 30-min return period, first the largest wave

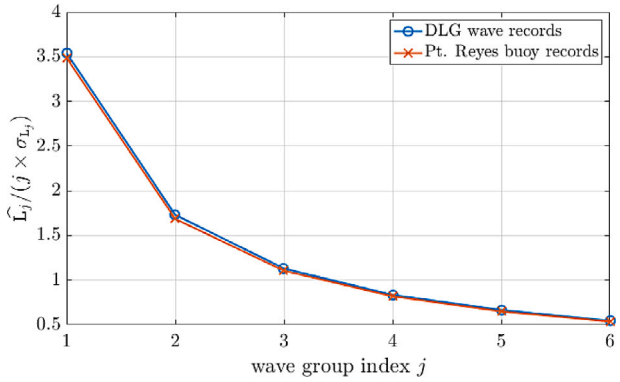

Fig. 8. Average 30-min maximum derived process value $\hat{\mathrm{L}}_j$, normalized by wave group index j and standard deviation (also, $\hat{z}_{jj}/j$, see Eq. (8)), from 1000 DLG time series $\eta_j(t)$ for each index j and 6052 Pt. Reyes buoy records.

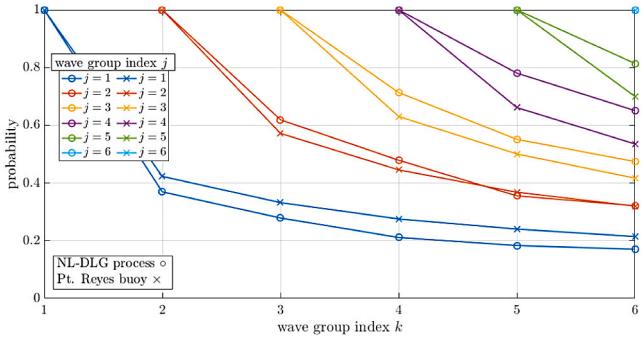

Fig. 9. Probability that 30-min $\hat{\mathrm{L}}_j$ and $\hat{\mathrm{L}}_k$ occur clustered together within record, NL-DLG process (∘) as compared to Pt. Reyes buoy records (×).

group of j waves for $j = 1, 2, \ldots, 6$ wave peaks with $\tau = 8.55$ sec between wave peaks is identified for each record using Eq. (1). For each 30-min record, a joint occurrence is recorded if the index of a wave group of k waves occurs at maximum 10 sec before or 60 sec after the largest wave group of j waves within that record,[1] for $j, k = 1, 2, \ldots, 6$ and $j \neq k$. If the largest wave groups of j and k waves for $j, k = 1, 2, \ldots, 6$ and $j \neq k$ occur outside this $t \in [-10, 60]$ sec interval, no joint occurrence is recorded. Instead, the extreme values for these realizations of wave groups of j and k waves are recorded separately.

Fig. 9 shows the probability that $\hat{\mathrm{L}}_j$ and $\hat{\mathrm{L}}_k$ occur clustered together over the record length, as estimated by the NL-DLG process (∘) and compared to the Pt. Reyes buoy records (×). For example, the darkest blue line gives the probability that $\hat{\mathrm{L}}_1$, the largest wave peak of the record, occurs together with $\hat{\mathrm{L}}_k$ for $k = 1, 2, \ldots, 6$. As the index of $\hat{\mathrm{L}}_k$ increases (increasing x-axis), the examined $\hat{\mathrm{L}}_k$ values (colored lines) are evaluated for $j = k, \ldots, 6$, as $\hat{\mathrm{L}}_k$ values for $k < j$ are examined in the previous lines.

Fig. 9 indicates that it is more likely that 30-min extreme wave groups occur clustered together when both group indices are high, as also suggested from Fig. 8, though the NL-DLG process slightly underestimates this probability as compared to the Pt. Reyes buoy records for lower group indices. The NL-DLG process wave profiles best estimate the Pt. Reyes probabilities that the 30-min extreme wave groups of j and k waves occur clustered together during the records when at least one of the group indices is small. This is unsurprising considering Fig. 8, as there is a stronger distinction between the extreme values which indicate rare wave groups of fewer waves, versus larger group indices.

Fig. 10 gives the distributions of joint 30-min $\hat{\mathrm{L}}_2/\sigma_{\mathrm{L}_2}$ and $\hat{\mathrm{L}}_3/\sigma_{\mathrm{L}_3}$ values, corresponding to the largest wave group occurrence of wave groups of 2 and 3 waves in the 30-min exposure, from the NL-DLG process (left inset) and Pt. Reyes buoy records (middle inset); the marginal 30-min distributions of $\hat{\mathrm{L}}_2/\sigma_{\mathrm{L}_2}$ and $\hat{\mathrm{L}}_3/\sigma_{\mathrm{L}_3}$ values are given in the right inset. The NL-DLG process constructs $numT = 1365$ targeted wave profiles approximating 30-min statistics of wave groups of 2 and 3 waves, so 1365 Pt. Reyes buoy time series are examined (see Algorithm 1 in the Appendix for calculating $numT$). The total simulation time associated with the 1365 NL-DLG process wave profiles is 75.5 hr; the total time associated with the 1365 Pt. Reyes buoy records is 682.5 hr (*i.e.*, $numT \times 0.5$ hr/record).

Note that each point in the distributions in Fig. 10 is associated with a unique wave profile, from either the NL-DLG process or the Pt. Reyes buoy, which excites this 30-min joint extreme wave group occurrence. These wave profiles could be used in a dynamic assessment where time domain simulations may be useful to examine the performance of a marine system. For these 30-min joint wave group statistics, the NL-DLG process assembles a quite similar distribution of joint extreme occurrences of wave groups of 2 and 3 waves with over 9 times less required simulation time relating to the time series realizations containing those joint occurrences.

The NL-DLG process and Pt. Reyes buoy waves both indicate bimodal behavior in the joint occurrence of rare wave groups of 2 and 3 waves relating to around $2.9\sigma_{\mathrm{L}_2}/2.9\sigma_{\mathrm{L}_3}$ and $3.3\sigma_{\mathrm{L}_2}/3.3\sigma_{\mathrm{L}_3}$. While not perfect, the NL-DLG process estimation of the joint rare wave group occurrences are quite reasonable when considering the reduced simulation time associated with these joint occurrences. Such a saving in required simulation time for converged statistics offers major advantages, as high-fidelity modeling tools can be directed exclusively to times when interesting responses are expected to occur (identified within the targeted NL-DLG process wave profiles) versus examining the entire return-period-length record for extreme responses (as done with the Pt. Reyes buoy records).

5.2. 30-Minute statistics based on theory

As the NL-DLG process utilizes a Gaussian RCWT to generate the wave profiles, an important question to consider is whether differences in the estimation of joint wave group occurrences stem from assumptions in the probabilistic framework of the NL-DLG process or from that Gaussian assumption. Clearly, Fig. 7 suggests that the extreme values $\hat{\mathrm{L}}_j$ indicating rare wave group occurrence from the Pt. Reyes buoy records already deviate from the expected Gaussian distribution even based on the same target extreme value, TEV. Therefore, this section compares 30-min rare wave group statistics assembled from brute-force Monte Carlo Simulations (MCS) with statistics from the NL-DLG process. Now, the TEV which defines the magnitude of the surrogate process extrema that condition the DLG waves is estimated by theory. This theoretical TEV is based on the return-period and spectral parameters of the derived Gaussian processes $\mathrm{L}_j(t)$ by Eq. (4). Eq. (4) does not depend on the bandwidth parameter of the spectrum of $\mathrm{L}_j(t)$, though it is dependent on a large number of observations (*i.e.*, a long return-period) (Ochi and Motter, 1973).

$$TEV_{\mathrm{L}_j} = \frac{\hat{\mathrm{L}}_j}{\sigma_{\mathrm{L}_j}} = \sqrt{2ln\left(\frac{T}{2\pi}\sqrt{\frac{m_{2_{\mathrm{L}_j}}}{m_{0_{\mathrm{L}_j}}}}\right)} \tag{4}$$

[1] The DLG wave records are $t = 70$ sec long, $t \in [-10, 60]$ sec, with the extreme event occurring at $t = 0$ seconds. Since the clustered extreme wave group occurrences identified by $\hat{\mathrm{L}}_j$ and $\hat{\mathrm{L}}_k$ can occur potentially clustered over each 70-sec DLG record, we do a similar comparison in the Pt. Reyes buoy records and look for an extreme occurrence of a wave group of k waves that occurs at maximum 10 sec before to 60 sec after the occurrence of an extreme wave group of j waves.

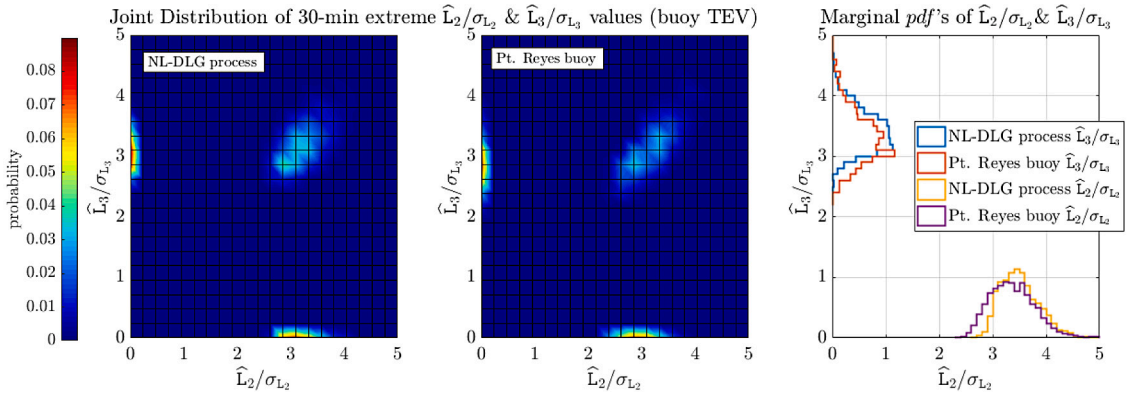

Fig. 10. Joint probability distribution for 30-min occurrence of wave groups of 2 and 3 waves, based on joint 30-min $\hat{L}_2/\sigma_{L_2}$ and $\hat{L}_3/\sigma_{L_3}$ values, from the NL-DLG process (left) and Pt. Reyes buoy records (middle), both using the same colorbar. Marginal *pdf*'s are in the right inset. Both the NL-DLG process and Pt. Reyes buoy distributions use *num*T = 1365 samples (relating to 1365 unique wave profiles).

where

TEV_{L_j} = target extreme value for wave group of j waves over return-period T

$\hat{L}_j$ = return-period extreme value of $L_j(t)$

$\sigma_{L_j} = \sqrt{m_{0_{L_j}}}$ = standard deviation of $L_j(t)$

T = return-period in sec

$m_{0_{L_j}}$ = zeroth spectral moment of $L_j(t)$

$m_{2_{L_j}}$ = second spectral moment of $L_j(t)$

Figs. 11–12 compare the empirical distributions of extreme values $\hat{L}_j$ for $j = 1, 2, \ldots, 6$ and the average normalized derived process value from 1000 DLG records for each group index based on the theoretical TEV, Eq. (4), and 6052 30-min MCS. Fig. 11 indicates that the distribution of extreme values $\hat{L}_j$ collected from the 30-min MCS gives a significantly closer match to the distribution of $\hat{L}_j$ assembled from the DLG waves based on the theoretical TEV, as expected since both these DLG and MCS waves are based on a Gaussian assumption. Fig. 12 similarly shows that the average normalized derived process values from the MCS and DLG waves are nearly identical; the magnitudes of these average normalized values are quite similar to the respective values in Fig. 8, but here there is much closer agreement between the DLG and MCS.

Similar to Fig. 9, Fig. 13 shows the probability that $\hat{L}_j$ and $\hat{L}_k$ occur clustered together over the record length, as estimated by the NL-DLG process (∘) and compared to the 30-min MCS (×). Here, the NL-DLG process overall gives better estimates of the probabilities that the rare wave groups cluster over the 30-min return-period, when compared with the 30-min MCS versus the Pt. Reyes buoy records. When compared with the Pt. Reyes buoy records, the NL-DLG process gave slightly better estimates of this probability for low wave group indices, but compared to the MCS, the NL-DLG process overall gives more accurate probability estimates for most comparisons of group indices, including larger group indices.

Finally, Fig. 14 gives the distributions of joint 30-min $\hat{L}_2/\sigma_{L_2}$ and $\hat{L}_3/\sigma_{L_3}$ values from the NL-DLG process (left inset) and 30-min MCS (middle inset) with the marginal 30-min extreme distributions in the right inset. This NL-DLG process joint distribution is based on *num*T = 1350 samples, relating to 75.2 hr of simulation time, versus 1350 30-min MCS, relating to 675 hr of simulation.

Figs. 11–14 suggest that the performance of the NL-DLG process depends on how well the RCWT realizations approximate the expected surrogate process statistics. Clearly though, there is still some small discrepancy in how the NL-DLG process estimates the mutual exclusivity of wave records, relating to step 2 of the full method in the Appendix, which can be further refined. In this case, the slightly non-Gaussian wave group statistics from the Pt. Reyes buoy records may stem from the relatively short record length and the dependence between successive wave peaks for higher wave group indices. When the RCWT realizations well approximate the marginal distributions of surrogate process extrema, as examined in this sub-section by MCS, the NL-DLG process well estimates joint distributions of rare wave group occurrences with significantly reduced simulation time relating to the resulting wave profiles.

5.3. 25-Hour statistics

Finally, a longer return-period is examined, this time comparing the joint statistics assembled by the NL-DLG process with statistics of 1000 25-hr MCS defined by the smoothed average spectrum in Fig. 2. Note that a longer return-period may not satisfy traditional stationarity requirements, *e.g.*, a 3-hr sea state. However, for Gaussian statistics, longer exposure periods can also be used to examine extreme values associated with a shorter exposure and an applied risk parameter or allowable probability of exceedance, see Ochi (1990).

Fig. 15 gives the statistics of the average normalized 25-hr maximum derived process values from the 25-hr MCS (×) and DLG waves targeted for 25-hr $\hat{L}_j$ occurrences (∘), overlaid with this comparison for the 30-min statistics from Fig. 8, comparing the 30-min DLG records with the 30-min MCS. For the 25-hr return-period, the trend of the average normalized 25-hr maximum derived process values is very similar as the trend for 30-min statistics. However, the successive difference between the normalized derived process maxima values for increasing group index is larger for the 25-hr statistics than for the 30-min statistics.

Given this larger distinction between extreme values which indicate rare wave group occurrence for 25-hr vs. 30-min statistics, Fig. 16 shows that the NL-DLG process (∘) closely estimates the probability that the 25-hr extreme wave groups of group index j and k occur clustered together over the exposure, compared with 25-hr MCS (×). For this longer return-period, there is a lower probability that the extreme wave groups of different group index occur together, as might be expected. Figs. 15–16 when directly compared to Figs. 12–13 strongly suggest that the dependence structure between extreme wave groups depends both on the group indices *and* the return-period.

Fig. 17 gives the joint 25-hr distributions of $\hat{L}_2/\sigma_{L_2}$ and $\hat{L}_3/\sigma_{L_3}$ values from the NL-DLG process (left inset) and 25-hr MCS (middle inset), and the marginal 25-hr extreme distributions (right inset). As only 1000 25-hr MCS were constructed, only 1000 NL-DLG process wave profiles are examined (out of the possible *num*T = 1119 realizations).

The total time associated with the 1000 NL-DLG process wave profiles is 74.1 hr while the total time associated with the 1000 25-hr MCS is 25,000 hr. Now the power of the NL-DLG process is truly

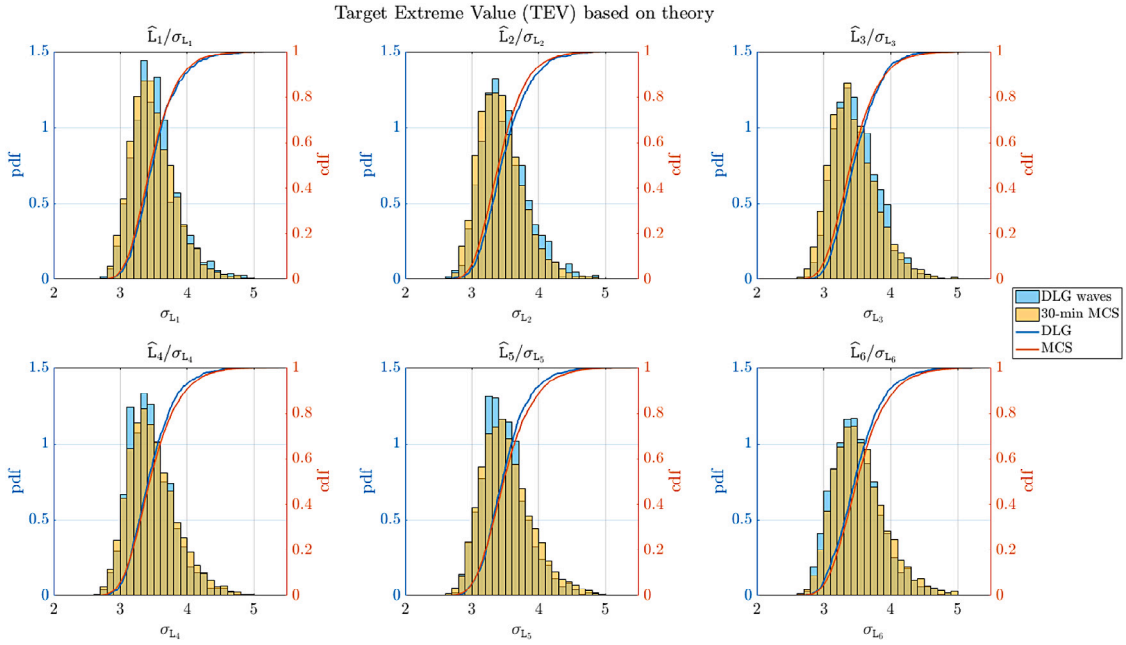

Fig. 11. Empirical pdf and cdf of 30-min maximum derived process values ($\widehat{\mathrm{L}}_j$) normalized by the respective standard deviation (σ_{L_j}) from 1000 DLG waves $\eta_j(t)$ for each index j (with TEV based on theory) and 6052 30-min MCS.

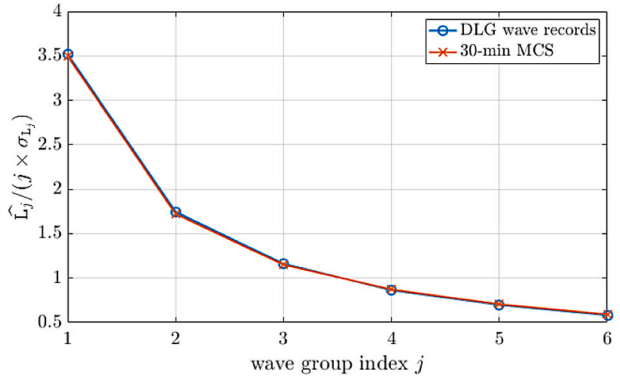

Fig. 12. Average 30-min maximum derived process value $\widehat{\mathrm{L}}_j$, normalized by wave group index j and standard deviation (also, $\hat{z}_{jj}/j$, see Eq. (8)), from 1000 DLG time series $\eta_j(t)$ for each index j (with TEV based on theory) and 6052 30-min MCS.

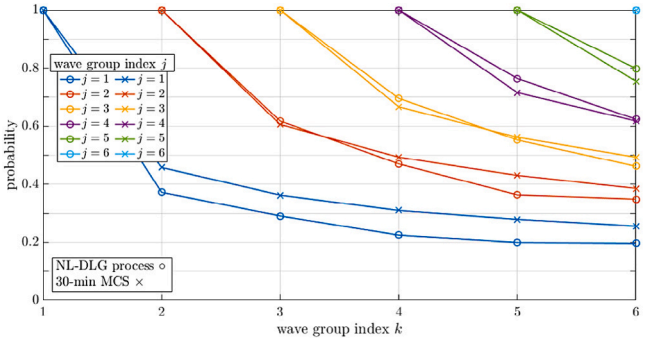

Fig. 13. Probability that 30-min $\widehat{\mathrm{L}}_j$ and $\widehat{\mathrm{L}}_k$ occur clustered together within record, NL-DLG process (∘) as compared to 6052 30-min MCS (×).

displayed, as the simulation time of the wave profiles associated with the distribution of joint extreme wave group occurrences is again directed only to the times when these joint extreme occurrences are expected, leading to a relatively low required simulation time for these 25-hr statistics. In contrast, the samples in Fig. 17 from the 25-hr MCS represent significantly more simulation time. Now, the NL-DLG process wave profiles exciting 25-hr statistics represent almost 340 times *less* simulation time than the same number of 25-hr MCS.

Fig. 17 also offers a striking comparison to Fig. 14, as they both consider the distribution of joint extreme occurrences of wave groups of 2 and 3 waves, but for different return-periods. Clearly the correlation between wave groups is dependent on the rareness of the wave groups themselves.

5.4. Computational effort of the NL-DLG process

The computational effort to run the NL-DLG process depends on the generation of the excitation profiles and the probabilistic framework linking those excitation realizations to the *num*T NL-DLG wave profiles. Here, the DLG was used for the RCWT, and the computational effort is mainly directed towards estimating the distribution of phases ϕ in Eq. (3) which excite extreme $\mathrm{L}_j(t = 0)$ values. The computational time for the DLG to estimate this phase distribution and generate the time series realizations is not dependent on the return-period, and only slightly dependent on the number of wave realizations to generate. To generate an ensemble of 1000 $\eta_j(t)$ DLG wave realizations targeted to excite $\widehat{\mathrm{L}}_j$ values takes about 2 min, regardless of the return-period. The computational time for the probabilistic framework of the NL-DLG process takes about 3 min to run. Therefore, about 15 min of computational time on a MacBook Pro 2.3 GHs Intel Core i5 generated the statistics and NL-DLG wave profiles relating to any combination of joint wave group occurrences of 1 to 6 waves. In contrast, on the same computer, it took about 21 min to run the 6052 30-min brute-force MCS.[2] In this case, the NL-DLG process's computational time is less than 70% of the computational time based on brute-force MCS.

For the 25-hr wave group statistics, the computational time of the NL-DLG process remains unchanged: 15 min for the wave group

[2] note 6052 30-min MCS were run to give a direct comparison to the 6052 available Pt. Reyes buoy records

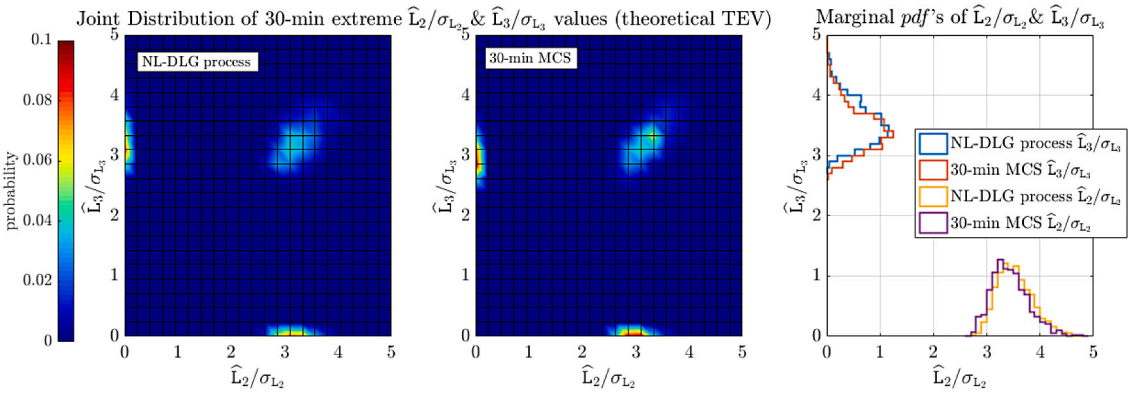

Fig. 14. Joint probability distribution for 30-min extreme occurrence of wave groups of 2 and 3 waves, based on joint 30-min $\widehat{L}_2/\sigma_{L_2}$ and $\widehat{L}_3/\sigma_{L_3}$ values, from the NL-DLG process (left) and Pt. Reyes buoy records (middle), both using the same colorbar. Marginal pdf's are in the right inset. Both the NL-DLG process and Pt. Reyes buoy distributions use $numT = 1350$ samples (relate to 1350 unique wave profiles).

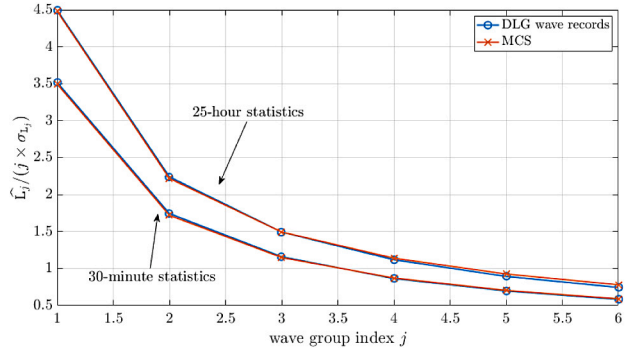

Fig. 15. Average 30-min and 25-hr maximum derived process values normalized by wave group index i ($\widehat{z}_{ii}/i$), from 1000 DLG time series $\eta_j(t)$ for each index j and 6052 30-min MCS/1000 25-hr MCS.

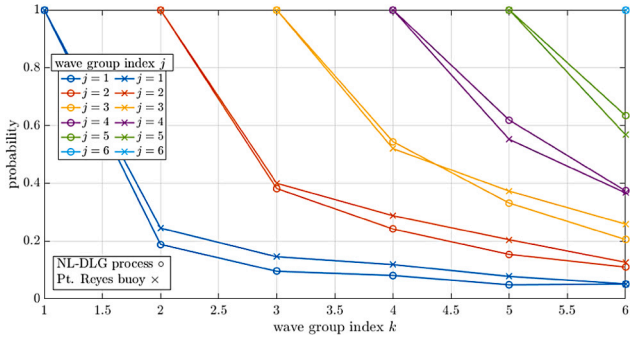

Fig. 16. Probability that 25-h $\widehat{L}_j$ and $\widehat{L}_k$ occur clustered together within record, NL-DLG process (∘) as compared to MCS (×).

statistics and generating the resulting NL-DLG wave profiles. For 1000 25-hr brute-force MCS, the computational time was about 30-min. Clearly, the NL-DLG process can efficiently simulate statistics of joint extreme occurrences with reduced computational effort from brute-force methods, and the savings increase for longer return periods. However, as noted earlier the major efficiency the NL-DLG process offers is in the reduced simulation time relating to the generated NL-DLG wave profile realizations, versus an equivalent number of full return-period length simulations.

6. Conclusions

This paper presented and formalized the NL-DLG process as a method to assemble an ensemble of mutually exclusive and exhaustive short excitation profiles which excite joint extreme values of surrogate processes interesting for design responses. The NL-DLG process was applied specifically to assemble wave profiles which contained instances of rare wave groups. The ensemble of generated wave profiles was evaluated by comparing resulting distributions of joint extreme wave group occurrences with distributions of extreme wave group occurrences measured from physical oceanographic data via the Pt. Reyes buoy.

To determine whether discrepancies in the comparisons stemmed from Gaussian assumptions or assumptions in the probabilistic framework of the NL-DLG process itself, brute-force Monte Carlo Simulations were also performed. Here, the NL-DLG process gave an improved estimate of joint wave group occurrence probabilities, as compared to the physical data from the Pt. Reyes buoy. Though there may still be ways to refine the assumptions of the probability framework of the NL-DLG process, this comparison suggests that the NL-DLG process can at least reasonably estimate joint Gaussian statistics, with even better performance and computational and simulation-time for longer return periods. The capability offered by the NL-DLG process, specifically based on the ensemble of short wave profiles targeted for specific surrogate process extrema, offers engineers a way to define an efficient testing regime targeted for extreme events and quickly estimate joint statistics of potentially correlated environmental phenomena like rare wave groups.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The author gratefully thanks the anonymous reviewers, whose thoughtful and thorough reviews contributed to a significantly improved paper.

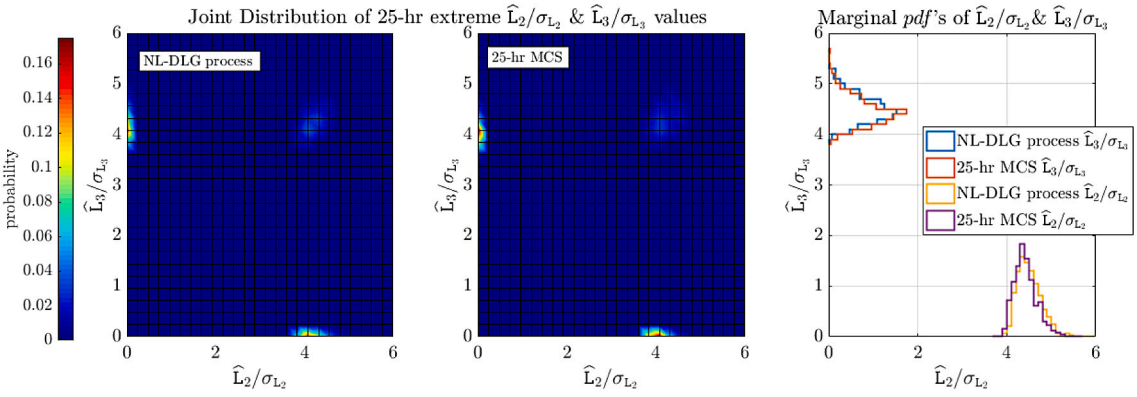

Fig. 17. Joint probability distribution for 25-hr extreme occurrence of wave groups of 2 and 3 waves, based on joint 25-hr extreme $\mathrm{L}_2/\sigma_{\mathrm{L}_3}$ and $\mathrm{L}_3/\sigma_{\mathrm{L}_3}$ values, from the NL-DLG process (left) and 25-h MCS (middle), both using the same colorbar. Marginal pdf's are in the right inset. Both the NL-DLG process and MCS distributions use $num\mathrm{T} = 1000$ samples.

Appendix

A.1. Constructing joint wave group occurrence distribution via the NL-DLG process

To generate an ensemble of mutually exclusive and exhaustive wave profiles targeted for return-period extreme events of multiple potentially correlated surrogate processes, the NL-DLG process is employed. The NL-DLG process was partially developed in Seyffert (2018) to estimate the lifetime failure probability of a complex non-linear system subject to combined, potentially correlated, stationary, non-Gaussian loading where a multi-dimension, potentially non-linear limit surface describes load combination levels which indicate instances of failure. This work formalizes the process give in Seyffert (2018) and for the first time generalizes the expressions to consider any number of possible load combinations, or surrogate processes.

The NL-DLG process constructs an ensemble of short excitation records (wave profiles) representative of the specific operational condition which lead to the most-likely exceedances of a defined limit surface (henceforth referred to as "failures") over the defined return-period. The definition of what counts as a failure is subjective, and can be exploited to the situation at hand. In Seyffert et al. (2019) the NL-DLG process was used to assemble waves expected to excite the most-likely collapse mechanism of a stiffened ship panel under combined loading; in that example failure was literal collapse. For the example in this paper, "failures" are defined as the most extreme wave group occurrences over a defined return period, so the NL-DLG process will be used to assemble an ensemble of waves which contain these most extreme wave group occurrences expected over the return period.

This ensemble of NL-DLG process wave record realizations can be interpreted as "abridged Monte Carlo Simulations" because they are meant to excite the same statistical response of the system as would an equivalent number of full-length Monte Carlo Simulations, but with major savings in simulation time, as the NL-DLG process wave focus solely on times when failures are expected to occur, even for long return-periods.

A.1.1. Identifying surrogate processes

Consider a complex system which is excited by combined, stochastic, and potentially non-Gaussian excitation. This excitation and the subsequent responses of interest may be non-linear functions of a global, Gaussian input, $\eta(t)$ (*e.g.*, wave excitation). Given that all system responses are excited by this excitation input $\eta(t)$, they likely have some level of (potentially unknown) correlation. These non-linear responses can be considered the vector $\mathbf{NL}(t)$. A limit surface G describes all combinations of the non-linear responses which result in failure, as shown in Eq. (5). System failure is the event that the load vector $\mathbf{NL}(t)$ exceeds the limit surface G within some time duration T, given by Eq. (6).

$$\begin{aligned}
&\eta(t) = \text{time series of the Gaussian input which excites the complex system}\\
&\mathtt{NL}_j(t) = j\text{th response time series resulting from a potentially non-linear}\\
&\quad\text{transformation of } \eta(t)\\
&G(\mathtt{NL}_1,\ldots,\mathtt{NL}_m) = 0 \equiv m\text{−dimension limit surface, which may be a}\\
&\quad\text{non-linear function of } m \text{ non-linear responses}
\end{aligned} \tag{5}$$

$$\begin{aligned}
&\mathbf{NL}(t) = \text{vector of } n \text{ non-linear load/ load effect time series}\\
&N_G(T) = \text{the number of up-crossings of } G \text{ by } \mathbf{NL}(t) \text{ in [0,T]}\\
&p(\mathbb{F}) = p(N_G(T) \geq 1) = \text{probability of failure}
\end{aligned} \tag{6}$$

While many methods exist to solve a problem such as presented in Eq. (6), see *e.g.*, Rice (1944), Winterstein et al. (1993), Giske et al. (2017) and Jensen (2007), here the desire is to estimate the failure probability while retaining a link with the excitation records (wave profiles) which excite such failures. To do so, the NL-DLG process first makes use of surrogate processes, similar to reduced order modeling, in that only essential physics of a system are retained, thus producing a system that exhibits relevant behavior. Surrogate processes can be defined as indicators of extreme behavior of some linear or non-linear responses of interest, as in Eq. (7). Here, the surrogate processes are defined as weighted sums of the Gaussian derived process whose extreme values indicate the occurrence of rare wave groups:

$$z_i(t) = \alpha\frac{\mathrm{L}_1(t)}{\sigma_{\mathrm{L}_1}} + \beta\frac{\mathrm{L}_2(t)}{\sigma_{\mathrm{L}_2}} + \cdots + \gamma\frac{\mathrm{L}_n(t)}{\sigma_{\mathrm{L}_n}} \tag{7}$$

where

$$\begin{aligned}
&z_i(t) \equiv i\text{th surrogate process composed of weighted, standard}\\
&\quad\text{deviation-normalized derived processes } \mathrm{L}_i(t),\ i = 1,\ldots,n\\
&\alpha,\beta,\ldots,\gamma = n \text{ weighting factors}\\
&\mathrm{L}_i(t) = \text{linear function that best indicates the extreme behavior of a}\\
&\quad\text{potentially non-linear response } \mathtt{NL}_i(t)\text{; here extreme values of } \mathrm{L}_i(t)\\
&\quad\text{from Eq.(1) indicate the occurrence of a rare wave group of } i \text{ waves}\\
&\sigma_{\mathrm{L}_i} = \text{standard deviation of } \mathrm{L}_i(t)
\end{aligned}$$

By varying the weighting factors $\alpha, \beta, \ldots, \gamma$ it is possible to emphasize some of the derived processes $\mathrm{L}_i(t)$ while de-emphasizing others. To examine the joint occurrence of rare wave groups of different group indices, simplify the surrogate processes by Eq. (8):

$$z_i(t) = \frac{\mathrm{L}_i(t)}{\sigma_{\mathrm{L}_i}} \tag{8}$$

$g(z_{i,T})$ = extreme value distribution of the surrogate process $z_i(t)$, based on the exposure period, T (9)

$\eta_i(t)$ = ensemble time series realization constructed to lead to an exposure-period-maximum of the surrogate process $z_i(t)$

$z_{ii}(t)$ = ensemble time series realization of the surrogate process $z_i(t)$ excited by $\eta_i(t)$

$\widehat{z_{ii}}$ = maximum of $z_{ii}(t)$, which is a member of the exposure-period extreme value distribution of $z_i(t)$;*i.e.*, $\widehat{z_{ii}} \in g(z_{i,T})$

$z_{ij}(t)$ = ensemble time series realization of the surrogate process $z_j(t)$ excited by $\eta_i(t)$

$\widehat{z_{ij}}$ = ensemble maximum of $z_{ij}(t)$ realization

Box I.

Therefore, each surrogate process, $z_i(t)$, $i = 1, \ldots, n$ mimics the extreme behavior of the associated response of interest (for this paper specifically: the occurrence of a rare wave group of i waves, indicated by an extreme value of $\mathrm{L}_i(t)$). Then the DLG can be used to construct ensembles of short excitation time series $\eta_i(t)$ which excite return-period extreme values of the associated surrogate process $z_i(t)$, and which are expected to excite extreme values of the associated (potentially non-linear) response of interest. Relating extreme behavior of the non-linear responses to extreme behavior of the associated surrogate processes in this way is not an unusual choice. A similar assumption is the basis of wave conditioning techniques and the critical wave episode model, in that non-linear responses are associated with (and specifically are corrections of) linear responses (Drummen et al., 2009; Torhaug, 1996).

Note that the Gaussian inputs constructed by the DLG to produce exposure-period-maxima of $z_i(t)$ may also excite extreme values of the other surrogate processes $z_j(t)$. Based on the correlation between the surrogate processes, it is possible that extrema of $z_i(t)$ and $z_j(t)$ are excited by the same Gaussian inputs. The potential overlap between the ith and the jth surrogate process maxima is accounted for by the variables in Eq. (9) given in Box I.

A.1.2. Sensitivity of surrogate process choice

The NL-DLG process is sensitive to the omission of important surrogate processes, as the resulting NL-DLG wave profiles are targeted to excite return-period extreme values of the defined surrogates. Therefore, if a surrogate process which has an important contribution to the system response is omitted, the NL-DLG process waves will similarly omit wave sequences targeted for these important events. In the same way, if a surrogate process is a poor indicator of the associated non-linear load/ load effect, the usefulness of the resulting NL-DLG process waves will suffer. On the other hand, including erroneous surrogate processes will not negatively effect the resulting NL-DLG process waves, apart from adding unnecessary simulation time to the resulting NL-DLG process waves. The probabilistic framework of the NL-DLG process will consider how the exposure-period maxima of this unnecessary surrogate cluster with the maxima of the other surrogates, and construct the resulting NL-DLG waves according to how these maxima cluster, regardless of whether these maxima are interesting to the system response. The derived Gaussian process defined by Eq. (1) has been shown theoretically and based on physical data and numerical simulations to be a good surrogate process to indicate the occurrence of rare wave groups (Seyffert and Troesch, 2016a; Seyffert et al., 2016). Similar investigations should be performed to evaluate the choice of surrogates for other applications.

A.1.3. Clustering of surrogate process maxima

Over a full exposure period, all surrogate processes experience an exposure-period-maximum. However, depending on the relation between the surrogate processes, those maxima may be excited by the same segment of the global, Gaussian system excitation $\eta(t)$ and may cluster together. For a total of n different surrogate processes that act as indicators of extreme behavior for n non-linear responses, there are multiple types of events, here called maxima clusters, that describe the potential clustering of surrogate process exposure-period-maxima, as given in Eq. (10) in Box II.

Only three types of clusters have been defined in Eq. (10) for brevity. Given n surrogate processes, similar events can be defined for surrogate process maxima in clusters of 1 to n maxima. Maxima clusters which contain more than one surrogate process maxima may come from more than one event, meaning the cluster can be excited by more than one type of excitation time series as shown in Eq. (11):

$$\begin{aligned} \overline{Z_i} &\text{ comes from only a } \widehat{Z_i} \text{ event} \\ \overline{Z_i Z_j} &\text{ comes from a } \widehat{Z_i Z_j} \text{ or } \widehat{Z_j Z_i} \text{ event} \\ \overline{Z_i Z_j Z_k} &\text{ comes from a } \widehat{Z_i Z_j Z_k},\ \widehat{Z_j Z_i Z_k},\ \text{or } \widehat{Z_k Z_i Z_j} \text{ event} \end{aligned} \tag{11}$$

Using the notation in Eq. (10)–(11), it is possible to describe the potential clustering of surrogate process maxima over an exposure period. As an example, consider a scenario when there are three responses of interest, *e.g.*, the occurrence of a rare wave group with 1, 2, or 3 waves, represented by three surrogate processes $z_1(t)$, $z_2(t)$, and $z_3(t)$. Each surrogate process experiences its maximum over the exposure, but those maxima may be clustered depending on the relationship between the surrogate processes (as indicated in Fig. 3).

There are five ways for three surrogate process maxima to cluster over a full exposure in which the order of the maxima clusters does not matter; call these possible groupings maxima configurations $\mathbb{C}i$ with $i = 1, 2, \ldots, 5$. All three maxima may occur clustered together, defined by $\mathbb{C}1 : \{\overline{Z_1 Z_2 Z_3}\}$. There are three ways that two surrogate process maxima may cluster together with the third separate, defined by $\mathbb{C}2 : \{\overline{Z_1 Z_2}, \overline{Z_3}\}$, $\mathbb{C}3 : \{\overline{Z_1 Z_3}, \overline{Z_2}\}$, and $\mathbb{C}4 : \{\overline{Z_2 Z_3}, \overline{Z_1}\}$. Or, all three maxima may occur un-clustered, defined by $\mathbb{C}5 : \{\overline{Z_1}, \overline{Z_2}, \overline{Z_3}\}$.

These maxima configurations make up the partitions of a finite set of n surrogate process maxima. Therefore, the total number of possible maxima configurations given n surrogate processes is the Bell number, B_n (Bell, 1938). With n defined surrogate processes, these B_n maxima configurations are an exhaustive definition of how n surrogate process exposure-period-maxima may cluster over an exposure, Eq. (12):

$$p(\mathbb{C}1) + p(\mathbb{C}2) + \cdots + p(\mathbb{C}B_n) = 1 \tag{12}$$

where

$\mathbb{C}1, \mathbb{C}2, \ldots, \mathbb{C}B_n \equiv$ maxima configuration $1, 2, \ldots, B_n$

$p(\mathbb{C}i)$ = probability that an exposure is described by the maxima configuration $\mathbb{C}i$ with $i = 1, 2, \ldots, B_n$

B_n = total number of possible configurations, given by the Bell number for the number of surrogate processes, n

$$= \sum_{k=0}^{n} \left\{ {n \atop k} \right\} = \sum_{k=0}^{n} \frac{1}{k!} \sum_{j=0}^{k} (-1)^{(k-j)} \binom{k}{j} j^n$$

A.1.4. Probability spaces of maxima configurations

Continuing the example in which three responses and associated surrogate processes are defined, the DLG constructs ensembles of $\eta_1(t)$,

$\overline{Z_i}$ = this cluster realization comes from 1 type of event and can be excited only by $\eta_i(t)$ excitation (10)

(1) $\widehat{Z_i}$, from $\eta_i(t)$: $\widehat{z_{ii}} \in g(z_{i,T})$, and all $\widehat{z_{ij}} < \widehat{z_{jj}}$ at PNE of $\widehat{z_{ii}}$ for $j = 1, \ldots, n$ and $j \neq i$

$\overline{Z_i Z_j}$ = this cluster realization comes from 2 types of events and can be excited by $\eta_i(t)$ or $\eta_j(t)$ excitation:

(1) $\widehat{Z_i Z_j}$, from $\eta_i(t)$: $\widehat{z_{ii}} \in g(z_{i,T})$; $\widehat{z_{ij}} > \widehat{z_{jj}}$ at PNE of $\widehat{z_{ii}}$; and all $\widehat{z_{ik}} < \widehat{z_{kk}}$ at PNE of $\widehat{z_{ii}}$ for $k = 1, \ldots, n$ and $k \neq i, j$

(2) $\widehat{Z_j Z_i}$, from $\eta_j(t)$: $\widehat{z_{jj}} \in g(z_{j,T})$; $\widehat{z_{ji}} > \widehat{z_{ii}}$ at PNE of $\widehat{z_{jj}}$; and all $\widehat{z_{jk}} < \widehat{z_{kk}}$ at PNE of $\widehat{z_{jj}}$ for $k = 1, \ldots, n$ and $k \neq i, j$

$\overline{Z_i Z_j Z_k}$ = this cluster realization comes from 3 types of events and can be excited by $\eta_i(t)$, $\eta_j(t)$, or $\eta_k(t)$ excitation:

(1) $\widehat{Z_i Z_j Z_k}$, from $\eta_i(t)$: $\widehat{z_{ii}} \in g(z_{i,T})$; $\widehat{z_{ij}} > \widehat{z_{jj}}$ at PNE of $\widehat{z_{ii}}$; $\widehat{z_{ik}} > \widehat{z_{kk}}$ at PNE of $\widehat{z_{ii}}$ & all $\widehat{z_{ip}} < \widehat{z_{pp}}$ at PNE of $\widehat{z_{ii}}$ for $p = 1, \ldots, n$ and $p \neq i, j, k$

(2) $\widehat{Z_j Z_i Z_k}$, from $\eta_j(t)$: $\widehat{z_{jj}} \in g(z_{j,T})$; $\widehat{z_{ji}} > \widehat{z_{ii}}$ at PNE of $\widehat{z_{jj}}$; $\widehat{z_{jk}} > \widehat{z_{kk}}$ at PNE of $\widehat{z_{jj}}$ & all $\widehat{z_{jp}} < \widehat{z_{pp}}$ at PNE of $\widehat{z_{jj}}$ for $p = 1, \ldots, n$ and $p \neq i, j, k$

(3) $\widehat{Z_k Z_i Z_j}$, from $\eta_k(t)$: $\widehat{z_{kk}} \in g(z_{k,T})$; $\widehat{z_{ki}} > \widehat{z_{ii}}$ at PNE of $\widehat{z_{kk}}$, $\widehat{z_{kj}} > \widehat{z_{kk}}$ at PNE of $\widehat{z_{kk}}$ & all $\widehat{z_{kp}} < \widehat{z_{pp}}$ at PNE of $\widehat{z_{kk}}$ for $p = 1, \ldots, n$ and $p \neq i, j, k$

Box II.

$\eta_2(t)$, and $\eta_3(t)$ time series. Consider that the $\eta_1(t)$ time series are constructed so that each $z_{11}(t)$ time series has a maximum $\widehat{z_{11}} \in g(z_{1,T})$. These $\eta_1(t)$ time series also excite $z_{12}(t)$ and $z_{13}(t)$ time series, whose resulting maxima may also be exposure-period-extreme values. In this way, the maxima $\widehat{z_{11}}$ can be classified into disjoint and exhaustive groups depending on whether the $\eta_1(t)$ time series which lead to $\widehat{z_{11}} \in g(z_{1,T})$ also lead to $\widehat{z_{12}} \in g(z_{2,T})$ and/ or $\widehat{z_{13}} \in g(z_{3,T})$, as described by the criteria in Eq. (10). Each $\eta_1(t)$ time series can be seen as an individual trial of an experiment with the possible outcomes: $\{\widehat{Z_1}, \widehat{Z_1 Z_2}, \widehat{Z_1 Z_3}, \widehat{Z_1 Z_2 Z_3}\}$. Then, a measurable probability space, $(\Omega_i, \mathcal{F}_i, P_i)$, for the possible event outcomes from an $\eta_i(t)$ time series is defined:

$$
\begin{aligned}
\Omega_i &= \{\widehat{Z_i}, \widehat{Z_i Z_j}, \widehat{Z_i Z_k}, \widehat{Z_i Z_j Z_k}\} \\
\mathcal{F}_i &= \{\Omega_i, \emptyset, \widehat{Z_i}, \widehat{Z_i Z_j}, \widehat{Z_i Z_k}, \widehat{Z_i Z_j Z_k}, \widehat{Z_i} \cup \widehat{Z_i Z_j}, \widehat{Z_i} \cup \widehat{Z_i Z_k}, \widehat{Z_i} \cup \widehat{Z_i Z_j Z_k}, \\
&\quad \widehat{Z_i Z_j} \cup \widehat{Z_i Z_k}, \widehat{Z_i Z_j} \cup \widehat{Z_i Z_j Z_k}, \widehat{Z_i Z_k} \cup \widehat{Z_i Z_j Z_k}, \widehat{Z_i} \cup \widehat{Z_i Z_j} \cup \widehat{Z_i Z_k}, \\
&\quad \widehat{Z_i} \cup \widehat{Z_i Z_j} \cup \widehat{Z_i Z_j Z_k}, \widehat{Z_i} \cup \widehat{Z_i Z_k} \cup \widehat{Z_i Z_j Z_k}, \widehat{Z_i Z_j} \cup \widehat{Z_i Z_k} \cup \widehat{Z_i Z_j Z_k}\} \\
P_i &= \{p(\{\widehat{Z_i}\}), p(\{\widehat{Z_i Z_j}\}), p(\{\widehat{Z_i Z_k}\}), p(\{\widehat{Z_i} \cup \widehat{Z_i Z_j}\}), \ldots\}
\end{aligned}
\tag{13}
$$

where

Ω_i = sample space of all possible outcomes from a $\eta_i(t)$ time series

$\mathcal{F}_i$ = event space, which is a σ-algebra

P_i = probability measure

$$p(\{\widehat{Z_i}\}) = \frac{\text{number of } \eta_i(t) \text{ time series which satisfy } \widehat{Z_i} \text{ condition}}{\text{number of } \eta_i(t) \text{ time series}}$$

$$p(\{\widehat{Z_i Z_j}\}) = \frac{\text{number of } \eta_i(t) \text{ time series which satisfy } \widehat{Z_i Z_j} \text{ condition}}{\text{number of } \eta_i(t) \text{ time series}}$$

$$p(\{\widehat{Z_i Z_j Z_k}\}) = \frac{\text{number of } \eta_i(t) \text{ time series which satisfy } \widehat{Z_i Z_j Z_k} \text{ condition}}{\text{number of } \eta_i(t) \text{ time series}}$$

$$p(\{\widehat{Z_i}\}) + p(\{\widehat{Z_i Z_j}\}) + p(\{\widehat{Z_i Z_k}\}) + p(\{\widehat{Z_i Z_j Z_k}\}) = 1, \text{ by definition}$$

The probability measure for any union of the events in the $(\Omega_i, \mathcal{F}_i, P_i)$ space is a simple sum because all of the events are by definition mutually exclusive. Measurable probability spaces for the possible event outcomes from the excitation time series $\eta_j(t)$, $(\Omega_j, \mathcal{F}_j, P_j)$, and $\eta_k(t)$, $(\Omega_k, \mathcal{F}_k, P_k)$, can similarly be defined. As shown in Eq. (10), the maxima cluster realizations may be excited by different excitation time series $\eta_i(t)$, $\eta_j(t)$, or $\eta_k(t)$. The events in the $(\Omega_i, \mathcal{F}_i, P_i)$ spaces are the events which can lead to the maxima clusters, as described in Eq. (11). For example, the above defined maxima configuration $\mathbb{C}4 : \{\overline{Z_2 Z_3}, \overline{Z_1}\}$ may occur as $\{\widehat{Z_2 Z_3}, \widehat{Z_1}\}$ or $\{\widehat{Z_3 Z_2}, \widehat{Z_1}\}$ where the order of the two maxima clusters does not matter. This maxima configuration simply describes the event that the exposure-period-maximum of $z_1(t)$ occurs unclustered over a full exposure, and that the exposure-period-maxima of $z_2(t)$ and $z_3(t)$ cluster together over an exposure.

A.1.5. Constructing ensemble of excitation time series from the NL-DLG process

To determine the probability that surrogate process maxima occur clustered over the exposure and construct the ensemble of NL-DLG process wave realizations, the following steps are employed: Instances of extreme values of the potentially non-linear responses of interest are approximated by defining surrogate processes which indicate extremes of those responses. A RCWT is used to generate an ensemble of excitation time series which excite a return-period extreme value distribution of the defined surrogate processes. Then, the excitation time series realizations are examined, Eq. (10)–(12) describe how these realizations can be grouped into potential maxima configurations, and the probability of experiencing each maxima configuration over an exposure period is estimated. Given the probability of these maxima configurations, time series generated by the RCWT and belonging to each specific maxima configuration type can be drawn based on their specific distribution to assemble an ensemble of excitation time series which approximate return-period statistics. Each step of the NL-DLG process is described below for any number of n non-linear responses of interest ($\mathtt{NL}_i(t)$ for $i = 1, \ldots, n$), relating to n surrogate processes ($z_i(t)$ for $i = 1, \ldots, n$) which act as indicators of extreme behavior for the n non-linear responses. Supplementary code is linked to this paper to perform these calculations given excitation time series constructed by a RCWT for the desired return period (Seyffert, 2021).

1. Construct a sufficient number of RCWT time series $\eta_i(t)$ for $i = 1, \ldots, n$. The simulation length should be based on the wave spectrum autocorrelation period.
2. Group these excitation simulation records into the maxima clusters described by Eq. (10)–(11). To determine whether an excitation record $\eta_i(t)$ excites maxima of other surrogate processes (*i.e.*, whether there is a $\widehat{Z_i}$, $\widehat{Z_i Z_j}$, or $\widehat{Z_i Z_j Z_k}$ event, following the notation from Eq. (10)–(11)), the following criterion is used: The extreme values $\widehat{z_{ii}}$ excited by $\eta_i(t)$ can be ordered based on increasing probability of non-exceedance, PNE; these $\eta_i(t)$ wave realizations can excite time series of the other surrogate processes. Then, those other surrogate process maxima, $\widehat{z_{ij}}$ and $\widehat{z_{ik}}$, are compared to the values $\widehat{z_{jj}}$ and $\widehat{z_{kk}}$, respectively, at the same PNE corresponding to the $\widehat{z_{ii}}$ value. If the values $\widehat{z_{ij}}$ and $\widehat{z_{ik}}$ conditioned on $\widehat{z_{ii}}$ occurrence at a certain PNE are larger than the respective $\widehat{z_{jj}}$ and $\widehat{z_{kk}}$ values at the same PNE, it is said that the surrogate process maxima cluster. Essentially, this criterion compares the maximum value of a surrogate process $z_j(t)$ excited by a wave record $\eta_i(t)$, meaning it is conditioned on the occurrence of the exposure-period-maximum $\widehat{z_{ii}}$ with $i \neq j$,

to the value of the extreme value distribution of $z_j(t)$ ($g(z_{j,T})$) at the same PNE to decide whether that specific maxima occurrence of $z_j(t)$ occurs clustered with a maxima occurrence of $z_i(t)$.

3. For $k = 1, \dots, n$, partition the n surrogate process maxima into k groups, in which each surrogate process maximum occurs exactly once and the order of maxima clusters does not matter. This generates the maxima configurations described previously. The number of possible partition schemes for n surrogate maxima clustered into k groups is $S(n,k)$, the Stirling number of the second kind. All partitions over $k = 1, \dots, n$ describe the maxima configurations $\mathbb{C}i$ with $i = 1, 2, \dots, B_n$ (*i.e.*, $\sum_{k=1}^{n} S(n,k) = B_n$).
For $1 < k < n$ there is more than one possible partition scheme for the k groups of n surrogate process maxima. For example, given $n = 3$ and $k = 2$, there are $S(3,2) = 3$ possible partition schemes to group $n = 3$ surrogate process maxima into $k = 2$ groups (*i.e.*, $\mathbb{C}2 : \{\overline{Z_1 Z_2}, \overline{Z_3}\}$, $\mathbb{C}3 : \{\overline{Z_1 Z_3}, \overline{Z_2}\}$, and $\mathbb{C}4 : \{\overline{Z_2 Z_3}, \overline{Z_1}\}$).
For a given maxima configuration $\mathbb{C}i$ which has k maxima clusters, call each of the k clusters, as defined by Eq. (10), $\gamma_{a_{ci}}$ where $a = 1, \dots, k$. The size of each cluster, $|\gamma_{a_{ci}}|$, relates to the number of surrogate process maxima contained within that cluster. For example, the cluster $\overline{Z_1 Z_2}$ has size 2. For a configuration $\mathbb{C}i$ which has k maxima groups:

$$\sum_{a=1}^{k} |\gamma_{a_{ci}}| = n \tag{14}$$

Note that $|\gamma_{a_{ci}}| \leq n$ for all $a = 1, \dots, k$, meaning that the size of a group cannot exceed the number of surrogate processes, n.

   (a) Determine the probability the surrogate process maxima cluster together as described by $\gamma_{a_{ci}}$:

   - If $|\gamma_{a_{ci}}| = 1$, the cluster contains a single surrogate process maximum; *e.g.*, if $\gamma_{a_{ci}} = \overline{Z_i} = \widehat{Z_i}$ where $i = 1, 2, \dots, n$, then:

$$p(\gamma_{a_{ci}}) = p(\overline{Z_i}) = p(\widehat{Z_i}) \tag{15}$$

   - If $|\gamma_{a_{ci}}| > 1$, the cluster contains more than one surrogate process maximum. The probability of experiencing the cluster, $\gamma_{a_{ci}}$, is the sum of the probability of experiencing that cluster given it is excited by *any* excitation record $\eta_i(t)$ $i = 1, \dots, n$ which could lead to that specific clustering of maxima. For example, say $\gamma_{a_{ci}} = \overline{Z_1 Z_2 Z_3}$. Then:

$$\begin{aligned} p(\gamma_{a_{ci}}) &= p(\overline{Z_1 Z_2 Z_3}) \\ &= p(\widehat{Z_1 Z_2 Z_3}) + p(\widehat{Z_2 Z_1 Z_3}) + p(\widehat{Z_3 Z_1 Z_2}) \end{aligned} \tag{16}$$

     The probability of experiencing the cluster, $\gamma_{a_{ci}} = \overline{Z_i Z_j Z_k}$, is the sum of the probabilities of experiencing that cluster of surrogate process maxima from any of the $\eta_i(t)$, $\eta_j(t)$, or $\eta_k(t)$ records.

   (b) Find the unconditioned probability of experiencing the maxima configuration ci which contains k maxima groups:

$$p(ci) = \frac{k!(n-k)!}{n!} \prod_{a=1}^{k} p(\gamma_{a_{ci}}) \tag{17}$$

     This probability is not yet conditioned on the fact that only physically-realizable maxima configuration types ($\mathbb{C}i$) are considered, hence the different script to describe the configuration, ci.

4. The probability of experiencing a maxima configuration over a full exposure must be conditioned on being a physically possible maxima configuration. This normalization is necessary because the maxima configurations for n surrogate processes are defined by drawing a sample from the sample spaces Ω_i for $i = 1, \dots, n$, as in Eq. (13). This leads to another probability space, representing each possible drawing of events composed of a single event from $\Omega_1, \Omega_2, \dots$, and then Ω_n. This probability space, though, also includes maxima configurations which are not physically realizable, *i.e.*, configurations where surrogate process exposure-period-maxima may occur more than once. Ignoring these non-physically-realizable maxima configurations requires the normalization performed in Eq. (18):

$$p(\mathbb{C}i) = \frac{p(ci)}{\sum_{i=1}^{B_n} p(ci)} \tag{18}$$

A.1.6. Generating ensemble of NL-DLG process wave profiles

Based on the maxima configuration probabilities given in Eq. (18) time series can be drawn from the original ensembles of $\eta_i(t)$, $\eta_j(t)$, and $\eta_k(t)$ realizations which satisfy the maxima configuration clusters referenced in Eq. (10). Based on the maxima configuration probabilities and the number of time series which fit each maxima cluster, excitation time series from the RCWT will be linked to create *num*T NL-DLG process waves that are meant to approximate the return-period statistics of *num*T full-length Monte Carlo Simulations. The number *num*T is limited to prevent oversampling of the distributions of the maxima clusters and can be calculated by Algorithm 1 which is also implemented in the accompanying code.

ALGORITHM 1: Calculate *num*T.

Input : n – number of surrogate processes
$\gamma_{a_{ci}}$ – possible maxima clusters defined by Eq. (10)
Output: *num*T – number of NL-DLG process wave profiles that can be assembled

```
1  K ← [1, 2, ..., n]                    /* possible number of maxima clusters */
2  for k ← K do
3      num ← [1, 2, ..., S(n,k)]          /* number of possible maxima
         configurations of n surrogate processes clustered in k
         groups; 2nd Stirling number */
4      for e ← num do
           /* go over num maxima configurations with k clusters */
5          A ← [1, 2, ..., k]  /* go over k maxima clusters, γ_aci within
             each maxima configuration */
6          for a ← A do
7              L ← count(γ_aci)  /* number of γ_aci clusters identified
                 in all RCWT waves */
8              P ← Σ_{i=1}^{B_n} p(Ci*)          /* sum of configuration
                 probabilities, only considering maxima
                 configurations which include this specific maxima
                 cluster */
9              g ← floor(L/P)  /* maximum possible number of NL-DLG
                 waves which could accommodate the required number
                 of γ_aci maxima clusters (L) based on the total
                 probability of experiencing this cluster over an
                 exposure (P) */
10         end
11     end
12 end
13 numT ← min(g)  /* numT is the minimum over the maximum possible
     number of NL-DLG waves that can accommodate the sampled γ_aci
     maxima clusters, calculated in line 9 over each loop, when
     considering all maxima clusters */
```

A.1.7. Linking ensemble of excitation time series to system failure

The ensemble of excitation time series constructed by the NL-DLG process can also be used as input to quickly estimate system failure probability given a defined limit surface, G, as in Eq. (5). In this case, the non-linear responses of interest should be chosen to emphasize n different regions of G. As described above, surrogate processes which indicate extreme responses of the non-linear responses should be defined and a RCWT used to generate ensembles of excitation time

$$p(\overline{Z_i}^{F^c}) = \frac{\text{number of } \overline{Z_i}^{F^c} \text{ events}}{\text{number } \overline{Z_i} \text{ events}} \tag{19}$$

$$p(\overline{Z_i Z_j}^{F^c}) = \frac{\text{number of } \overline{Z_i Z_j}^{F^c} \text{ events}}{\text{number of } \overline{Z_i Z_j} \text{ events}} = \frac{\text{number of } \widehat{Z_i Z_j}^{F^c} \text{ events} + \text{number of } \widehat{Z_j Z_i}^{F^c} \text{ events}}{\text{number of } \widehat{Z_i Z_j} \text{ events} + \text{number of } \widehat{Z_j Z_i} \text{ events}}$$

$$p(\overline{Z_i Z_j Z_k}^{F^c}) = \frac{\text{number of } \overline{Z_i Z_j Z_k}^{F^c} \text{ events}}{\text{number of } \overline{Z_i Z_j Z_k} \text{ events}} = \frac{\text{number of } \widehat{Z_i Z_j Z_k}^{F^c} \text{ events} + \text{number of } \widehat{Z_j Z_i Z_k}^{F^c} \text{ events} + \text{number of } \widehat{Z_k Z_i Z_j}^{F^c} \text{ events}}{\text{number of } \widehat{Z_i Z_j Z_k} \text{ events} + \text{number of } \widehat{Z_j Z_i Z_k} \text{ events} + \text{number of } \widehat{Z_k Z_i Z_j} \text{ events}}$$

where

$\widehat{Z_i}^{F}, \widehat{Z_i Z_j}^{F}, \widehat{Z_i Z_j Z_k}^{F}$ ≡ event that system fails when excited by the $\eta_i(t)$, $\eta_j(t)$, or $\eta_k(t)$ time series associated with the corresponding $\widehat{Z_i}$, $\widehat{Z_i Z_j}$, or $\widehat{Z_i Z_j Z_k}$ event

$\widehat{Z_i}^{F^c}, \widehat{Z_i Z_j}^{F^c}, \widehat{Z_i Z_j Z_k}^{F^c}$ ≡ event that system does not fail when excited by the $\eta_i(t)$, $\eta_j(t)$, or $\eta_k(t)$ time time series associated with the corresponding $\widehat{Z_i}$, $\widehat{Z_i Z_j}$, or $\widehat{Z_i Z_j Z_k}$ event

Box III.

series which lead to a distribution of extreme return-period surrogate process values. Then the surrogate process extrema are grouped into the maxima configurations, $\mathbb{C}i$ for $i = 1, \ldots, B_n$, based on the maxima clusters described in Eq. (10).

In each type of maxima configuration, the system is excited by at least one and up to n maxima clusters. Rather than determine the order and probability of multiple failures due to the multiple maxima clusters, which would be problematic if not impossible, it is easier to consider the null problem. So, given each maxima configuration type, a conditional non-failure probability can be determined using those grouped excitation records, as in Eq. (19) (see Box III on the next page). For the excitation simulations that fit into a cluster type, determine the probability of system non-failure given the system is excited by this group of simulations. This is done simply by exciting the system with that group of excitation simulations, and collecting the number of non-failures that result.

Then, find the probability of non-failure for the system given an exposure to the given maxima configuration $\mathbb{C}i$. The maxima events are defined so that any maxima clusters within a type of maxima configuration (*e.g.*, $\overline{Z_1 Z_2}$ and $\overline{Z_3}$ are clusters within the maxima configuration $\mathbb{C}2$) are, by definition, unclustered. Therefore, the probabilities of non-failure due to excitation which leads to the realization of the clusters within a maxima configuration are independent. So, the probability of non-failure given a specific maxima configuration is given by Eq. (20), starting by first defining the probability of failure given a specific maxima configuration, $p(\mathbb{F}|\mathbb{C}i)$, via cluster failure probabilites $p(\gamma^F_{a_{ci}}|\mathbb{C}i)$:

$$\begin{aligned} p(\mathbb{F}|\mathbb{C}i) &= p(\cup_{a=1}^{k} \gamma^F_{a_{\mathbb{C}i}}|\mathbb{C}i) \\ &= 1 - p\left(\mathbb{F}^c|\mathbb{C}i\right) \\ &= 1 - p\left(\left(\cup_{a=1}^{k} \gamma^F_{a_{ci}}\right)^c|\mathbb{C}i\right) \\ &= 1 - p\left(\cap_{a=1}^{k} \gamma^{F^c}_{a_{ci}}|\mathbb{C}i\right) \\ \therefore p(\mathbb{F}^c|\mathbb{C}i) &= p\left(\gamma^{F^c}_{1_{ci}}|\mathbb{C}i\right) p\left(\gamma^{F^c}_{2_{ci}}|\mathbb{C}i\right) \cdots p\left(\gamma^{F^c}_{k_{ci}}|\mathbb{C}i\right) \end{aligned} \tag{20}$$

Then, the overall system failure, $p(\mathbb{F})$, is estimated considering all possible maxima configurations, using Eqs. (18) and (20):

$$p(\mathbb{F}) = 1 - \left(\sum_{i=1}^{B_n} p(\mathbb{F}^c|\mathbb{C}i) p(\mathbb{C}i)\right) \tag{21}$$

References

Adegeest, L.J.M., Braathen, A., Løseth, R.M., 1998. Use of non-linear sea-loads simulations in design of ships. In: Practical Design of Ships and Mobile Units. pp. 53–58.

Alford, L.K., Troesch, A.W., 2009. Generating extreme ship responses using non-uniform phase distributions. Ocean Eng. (ISSN: 00298018) 36 (9–10), 641–649. http://dx.doi.org/10.1016/j.oceaneng.2009.03.002.

Bartoli, G., Mannini, C., Massai, T., 2011. Quasi-static combination of wind loads: A copula-based approach. J. Wind Eng. Ind. Aerodyn. 99 (6–7), 672–681.

Bassler, C.C., Belenky, V., Dipper, M.D., 2010. Characteristics of wave groups for the evaluation of ship response in irregular seas. In: Proceedings ASME 2010 29th OMAE Conference. No. 20241.Shanghai, China.

Bastian, K., Pahlow, M., Hundecha, Y., Schumann, A., 2009. Probability analysis of hydrological loads for the design of flood control systems using copulas. J. Hydrol. Eng. 15 (5), 360–369.

Bell, E.T., 1938. The iterated exponential integers. Annal. Math. (2) Ser. 39 (3), 539–557.

Boccotti, P., 2015. Wave Mechanics and Wave Loads on Marine Structures. Elsevier; Butterworth-Heinemann.

CDIP, 2021. Station 029 Pt. Reyes Buoy. University of California San Diego, Coastal Data Information Program.

Chai, W., Leira, B.J., 2018. Environmental contours based on inverse SORM. Mar. Struct. (ISSN: 09518339) 60 (March), 34–51. http://dx.doi.org/10.1016/j.marstruc.2018.03.007.

de Waal, D.J., van Gelder, P.H.A.J.M., 2005. Modelling of extreme wave heights and periods through copulas. Extremes (ISSN: 13861999) 8 (4), 345–356. http://dx.doi.org/10.1007/s10687-006-0006-y.

Dietz, J.S., 2004. Application of Conditional Waves as Critical Wave Episodes for Extreme Loads on Marine Structures (Ph.D. thesis). (July), Technical University of Denmark, ISBN: 8789502434.

Drummen, I., Wu, M.K., Moan, T., 2009. Numerical and experimental investigations into the application of response conditioned waves for long-term nonlinear analyses. Mar. Struct. (ISSN: 09518339) 22 (3), 576–593. http://dx.doi.org/10.1016/j.marstruc.2008.12.002.

Ewans, K., Jonathan, P., 2014. Evaluating environmental joint extremes for the offshore industry. J. Mar. Syst. 130, 124–130, URL http://arxiv.org/abs/1211.1365.

Ewing, J.A., 1973. Mean length of runs of high waves. J. Geophys. Res. 78 (12), 1933–1936.

Fedele, F., 2005a. Successive wave crests in Gaussian seas. Probab. Eng. Mech. 20, 355–363.

Fedele, F., 2005b. On wave groups in a Gaussian sea. Ocean Eng. 33, 2225–2239.

Friis-Hansen, P., Nielsen, L.P., 1995. On the new wave model for the kinematics of large ocean waves. In: Proceedings of the 14th International Conference on Offshore Mechanics and Arctic Engineering, Copenhagen, Denmark. pp. 14–24.

Giske, F.I.G., Leira, B.J., Øiseth, O., 2017. Full long-term extreme response analysis of marine structures using inverse FORM. Probab. Eng. Mech. 50 (October), 1–8. http://dx.doi.org/10.1016/j.probengmech.2017.10.007.

Goda, Y., 1976. On Wave Groups. In: BOSS'76, vol. 1, pp. 115–128.

Gouldby, B., Wyncoll, D., Panzeri, M., Franklin, M., Hunt, T., Hames, D.c, Tozer, N., Hawkes, P., Dornbusch, U., Pullen, T., 2017. Multivariate extreme value modelling of sea conditions around the coast of England. Proc. Inst. Civ. Eng. Marit. Eng. (ISSN: 1741-7597) 170 (1), 3–20. http://dx.doi.org/10.1680/jmaen.2016.16.

Haselsteiner, A.F., Ohlendorf, J.-H., Wosniok, W., Thoben, K.-D., 2017. Deriving environmental contours from highest density regions. Coast. Eng. (ISSN: 03783839) 123, 42–51. http://dx.doi.org/10.1016/j.coastaleng.2017.03.002.

Haver, S., 1987. On the joint distribution of heights and periods of sea waves. Ocean Eng. (ISSN: 00298018) 14 (5), 359–376. http://dx.doi.org/10.1016/0029-8018(87)90050-3.

Heffernan, J.E., Tawn, J.A., 2004. A conditional approach for multivariate extreme values. J. R. Stat. Soc. Ser. B Stat. Methodol. (ISSN: 13697412) 66 (3), 497–530. http://dx.doi.org/10.1111/j.1467-9868.2004.02050.x.

Huseby, A.B., Vanem, E., Natvig, B., 2015. Alternative environmental contours for structural reliability analysis. Struct. Saf. (ISSN: 01674730) 54, 32–45. http://dx.doi.org/10.1016/j.strusafe.2014.12.003.

Jensen, J.J., 2007. Efficient estimation of extreme non-linear roll motions using the first-order reliability method (FORM). J. Mar. Sci. Technol. (ISSN: 09484280) 12 (4), 191–202. http://dx.doi.org/10.1007/s00773-007-0243-z.

Jonathan, P., Ewans, K., Flynn, J., 2014. On the estimation of ocean engineering design contours. J. Offshore Mech. Arct. Eng. 136 (4), http://dx.doi.org/10.1115/1.4027645.

Jones, M., Hansen, H.F., Zeeberg, A.R., Randell, D., Jonathan, P., 2018. Uncertainty quantification in estimation of extreme environments. Coast. Eng. (ISSN: 0378-3839) 141, 36–51. http://dx.doi.org/10.1016/j.coastaleng.2018.07.002.

Kim, D.-H., 2012. Design Loads Generator: Estimation of Extreme Environmental Loadings for Ship and Offshore Applications (Ph.D. thesis). University of Michigan.

Kim, D.-H., Belenky, V., Campbell, B.L., Troesch, A.W., 2014. Statistical estimation of extreme roll in head seas. In: Proceedings of the 33rd International Conference on Ocean, Offshore and Arctic Engineering.

Kim, D.-H, Troesch, A.W., 2015. Statistical estimation of extreme roll responses in short crested irregular head seas. Trans. Soc. Nav. Archit. Mar. Eng. 121, 123–160.

Lindgren, G., 1970. Some properties of a normal process near a local maximum. Ann. Math. Stat. 41 (6), 1870–1883.

Liu, Z., Elgar, S., Guza, R.T., 1993. Groups of ocean waves: Linear theory, approximations to linear theory, and observations. J. Waterw. Port Coast. Ocean Eng. 119, 144–159.

Longuet-Higgins, M.S., 1957. The statistical analysis of a random, moving surface. Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Sci. 249, 321–387.

Ochi, M.K., 1990. Applied Probability & Stochastic Processes in Engineering & Physical Sciences. Wiley series in probability and mathematical sciences.

Ochi, M.K., Motter, L.E., 1973. Prediction of slamming characteristics and hull responses for ship design. Trans. SNAME 81, 144–176.

Rice, S.O., 1944. Mathematical analysis of random noise. Bell Syst. Tech. J. 23 (3), 282–332.

Rye, H., 1974. Wave group formation among storm waves. In: Proceedings 14th International Conference on Coastal Engineering. ASCE, Copenhagen, pp. 164–183, http://dx.doi.org/10.9753/icce.v14.

Seyffert, H.C., 2018. Extreme Design Events due to Combined, Non-Gaussian Loading (Ph.D. thesis). The University of Michigan.

Seyffert, H.C., 2021. Supplementary code and data for "generating an ensemble of mutually exclusive and exhaustive waves targeted for extreme responses". 4TU.ResearchData. Software. http://dx.doi.org/10.4121/14899818.

Seyffert, H.C., Kana, A.A., Troesch, A.W., 2020. Numerical investigation of response-conditioning wave techniques for short-term rare combined loading scenarios. Ocean Eng. (ISSN: 0029-8018) 213, http://dx.doi.org/10.1016/j.oceaneng.2020.107719.

Seyffert, H.C., Kim, D.-H., Troesch, A.W., 2016. Rare wave groups. Ocean Eng. 122, 241–252. http://dx.doi.org/10.1016/j.oceaneng.2016.05.053.

Seyffert, H.C., Troesch, A.W., 2016a. Data mining Pt. Reyes buoy for rare wave groups. J. Offshore Mech. Arct. Eng. 138, http://dx.doi.org/10.1115/OMAE201541451,

Seyffert, H.C., Troesch, A.W., 2016. Spar platform response due to rare wave groups. In: PRADS 2016 - Proceedings of the 13th International Symposium on PRActical Design of Ships and Other Floating Structures. ISBN: 9788774754732.

Seyffert, H.C., Troesch, A.W., Collette, M.D., 2019. Combined stochastic lateral and in-plane loading of a stiffened ship panel leading to collapse. Mar. Struct. 67, http://dx.doi.org/10.1016/j.marstruc.2019.04.008.

Shin, Y.S., Belenky, V., Paulling, J.R., Weems, K.M., Lin, W.M., McTaggart, K., Spyrou, K., Treakle, T.W., Levadou, M., Hutchison, B.L., Falzarano, J., Chen, H., Letizia, L., 2005. Criteria for parametric roll of large containerships in longitudinal seas. Tran. Soc. Nav. Archit. Mar. Eng. 112, 14–47.

Taylor, P.H., Jonathan, P., Harland, L.A., 1997. Time domain simulation of jack-up dynamics with the extremes of a gaussian process. J. Vib. Acoust. (ISSN: 07393717) 119 (4), 624. http://dx.doi.org/10.1115/1.2889772.

Themelis, N., Spyrou, K.J., 2007. Probabilistic assessment of ship stability. SNAME Trans. (ISSN: 00811661) 115 (September), 181–206.

Torhaug, R., 1996. Extreme Response of Nonlinear Ocean Structures: Identificaiton of Minimal Stochastic Wave Input for Time-Domain Simulation (Ph.D. thesis). Stanford University.

Tromans, P.S., Anaturk, A.R., Hagemeijer, P., 1991. A new model for the kinematics of large ocean waves - application as a design wave. In: Proceedings, 1st Internationl Offshore and Polar Engineering Conference (ISOPE). Vol. III. pp. 64–71.

Vanem, E., 2017. A comparison study on the estimation of extreme structural response from different environmental contour methods. Mar. Struct. (ISSN: 09518339) 56, 137–162. http://dx.doi.org/10.1016/j.marstruc.2017.07.002.

Vanem, E., 2020. Bivariate regional extreme value analysis for significant wave height and wave period. Appl. Ocean Res. (ISSN: 0141-1187) 101, 102266. http://dx.doi.org/10.1016/j.apor.2020.102266.

Winterstein, S.R., Ude, T.C., Cornell, C.A., Bjerager, P., Haver, S., 1993. Environmental parameters for extreme response: inverse FORM with omission factors. In: Proc. of Intl. Conf. on Structural Safety and Reliability (ICOSSAR93).

Wirsching, P., Paez, T., Ortiz, K., 2006. Random vibrations: Theory and practice.