Required test durations for converged short-term wave and impact extreme value statistics — Part 1

Ferry dataset

van Essen, Sanne M.; Scharnke, Jule; Seyffert, Harleigh C.

DOI
10.1016/j.marstruc.2023.103410

Publication date
2023

Document Version
Final published version

Published in
Marine Structures

Citation (APA)
van Essen, S. M., Scharnke, J., & Seyffert, H. C. (2023). Required test durations for converged short-term wave and impact extreme value statistics — Part 1: Ferry dataset. *Marine Structures*, *90*, Article 103410. https://doi.org/10.1016/j.marstruc.2023.103410

Marine Structures 90 (2023) 103410

Contents lists available at ScienceDirect

Marine Structures

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

Required test durations for converged short-term wave and impact extreme value statistics — Part 1: Ferry dataset

Sanne M. van Essen [a,b,*], Jule Scharnke [b], Harleigh C. Seyffert [a]

[a] Department of Maritime & Transport Technology, Delft University of Technology, Delft, The Netherlands
[b] Maritime Research Institute Netherlands (MARIN), Wageningen, The Netherlands

ARTICLE INFO

Keywords:
Convergence
Extreme value statistics
Waves
Wave impact loads
Seakeeping
Most probable maximum
Short-term variability
Seed variations
Experiments
Green water
Ferry
Design loads

ABSTRACT

For the design of maritime structures in waves, the extreme values of responses such as motions and wave impact loads are required. Waves and wave-induced responses are stochastic, so such responses should always be related to a probability. This information is not easy to obtain for strongly non-linear responses such as wave impact forces. Usually class rules or direct assessment via experiments or numerical simulations are applied to obtain extreme values for design. This brings up questions related to the convergence of extreme values: how long do we need to test in order to obtain converged statistics for the target duration? Or, vice versa: given testing data, what is the uncertainty of the associated statistics? Often the test or simulation duration is cut up in 'seeds' or 'realisations', with an exposure duration of one or three hours based on the typical duration of a steady environmental condition at sea, or the time that a ship sails a single course. The required number of seeds for converged results depends on the type of structure and response, the exposure duration, and the desired probability level. The present study provides guidelines for the convergence of most probable maximum (MPM) wave crest heights and MPM green water wave impact forces on a ferry. Long duration experiments were done to gain insight into the required number of seeds, and the effect of fitting. The present paper presents part 1 of this study; part 2 [1] presents similar results for wave-in-deck loads on a stationary deck box.

1. Introduction

1.1. Background

For the design or reliability assessment of a maritime structure, it is required to know the expected extreme values of wave-induced responses for a defined duration. There will be statistical variability to these values, due to the stochastic nature of the ocean excitation. For strongly non-linear responses such as wave impact forces (e.g., green water, slamming, air gap impacts), this variability may be relatively large. In order to obtain extreme response statistics for a defined duration or return-period with an uncertainty below a given tolerance, it is therefore important to estimate the required test duration at the start of a numerical or experimental campaign. Or, vice versa: given response data with a certain duration, it is desirable to know the uncertainty of the resulting extreme response statistics; this paper focuses on this question in particular. The more non-linear the response, the longer the required test duration for a given uncertainty and the higher the uncertainty for a given duration. Without clearly defining the required test duration for an experiment or simulation and the expected resulting response variability, there may be large uncertainty

* Corresponding author at: Department of Maritime & Transport Technology, Delft University of Technology, Delft, The Netherlands.
E-mail address: s.m.vanessen@tudelft.nl (S.M. van Essen).

https://doi.org/10.1016/j.marstruc.2023.103410
Received 30 November 2021; Received in revised form 6 February 2023; Accepted 24 February 2023
Available online 3 March 2023

in the design load for the structure, its failure probability may be over- or under-estimated, or the validation difference between experimental and numerical values may be within the uncertainty of both results.

The direct motivation for the present study is the desire to use long duration experimental data as validation material for methods to predict extreme values of wave impacts. Examples of such methods are screening approaches (e.g., [2–4]), response-conditioning approaches (e.g., [5–8]) or adaptive sampling approaches (e.g., [9–11]). In order to validate these approaches, it is necessary to obtain insight into the variability of the available experimental data. A more practical motivation to want this information could be to plan a test campaign to evaluate the design of a new structure exposed to wave impacts.

The present study and its 'sister study' [1] provide this insight for the variability of wave crests and green water impact loads on a sailing ferry, and wave-in-deck loads on a stationary deck box, respectively. The two long-duration datasets described in these publications will be used later for the validation study described above.

1.2. Short-term analysis: exposure duration and number of seed variations

In general, it can be said that the statistical (aleatoric) uncertainty of wave and response values can be reduced by increasing the test or simulation duration. Ideally a wave record should be infinite to eliminate sampling variability (see, e.g., [12]), but this is of course unfeasible. Here uncertainty is not knowing the single, true value of a quantity while the variability describes the fact that a quantity can take on different true values at different times, locations, etc. [13]. Statistical variation is not the only source of uncertainty in extreme wave impact values based on experiments or simulations; [14] provides an overview of many possible epistemic and aleatoric sources (e.g., sensor accuracy, residual waves in a basin, numerical implementation, scaling, etc.). These other sources are not considered here.

Long- and short-term variability of a response are usually considered separately. Long-term variability is related to the response variation over the many wave conditions a ship encounters in its lifetime; short-term variability is related to the variability of the response over different wave realisations of the same sea state condition. It is common practice to assume that long- and short-term variability can be decoupled. Long-term variability may first be considered via contour methods (e.g., [15–17]) or design sea state methods [18], resulting in a few sea states (see, e.g., [19]) at a given return period. Short-term variability and extreme values can then be evaluated by sufficiently long duration tests or calculations, such that the responses at an appropriate quantile of the short-term distribution can be evaluated (see Section 4.4). There is some debate about some assumptions in the contour methods (see e.g., [20–22]), but it is common practice to apply them. Here we will focus on the short-term responses based on experimental data, assuming a few critical sea states have already been selected from the scatter diagram.

Usually, the short-term duration of experiments is cut up into 'seeds' or 'realisations', each with the same exposure duration. A test with a duration of 30 h can equivalently be seen as 10 realisations ('seed variations') of 3 h exposure duration, or as 60 seeds of 30 min exposure, etc. Here we will use the term 'seed' to indicate a random realisation of a test condition. The exposure duration is usually determined at the start of a study (see Section 1.3). Section 1.4 discusses literature on the required number of seeds for converged extreme value statistics. Fitting theoretical formulations to the crest or peak distributions is a commonly applied way to reduce this variability and to extrapolate to longer durations. This is further discussed in Section 4.6. It should be noted that a high quantile of a 1 h distribution is equivalent to a lower quantile of a 3 h or 10 h distribution. There are therefore many ways to present the same results. In order to provide results that can be applied, we stuck as much as possible to the industry standard/common way that tests are performed in applied research institutes (see Sections 1.3 and 1.4). In the remainder of this paper, durations will be abbreviated: three hours as 3 h, one hour as 1 h, 30 min as 30 min, etc.

1.3. Guidelines concerning exposure duration

Some guidelines and literature are available for test durations of extreme wave impact event analysis. They are usually defined separately for offshore platforms and ships. The selected exposure duration should be related to the stationary duration of the wave conditions. For ships this is typically shorter, as the steady state duration is also determined by the time a ship sails a single course or speed. For the offshore industry, a 3 h stationary duration is commonly applied [12,23]. It is stated by [24] that 3 h is seen as the 'standard duration', but the period of stationarity can range from 30 min to 10 h, whereas [25] states that "information on sea state duration is to be provided (…) by metocean specialists". The ITTC guideline for wave modelling [26] states that normally a duration of 1 h is used for seakeeping and 3 h in offshore engineering. The ITTC guidelines for experiments on rarely occurring events [27] recommend to increase the exposure duration and increase the sea state until a minimum number of extreme events is included (minimum frequency of event occurrence set very high at 40%–60% of the wave encounters, plus 100 wave encounters as the lower limit, 200 as standard and 400 as excellent practice). Summarising, an exposure duration of 3 h is generally applied for offshore platforms and shorter durations for ships (e.g., 1 h or 100–200 wave encounters).

1.4. Guidelines concerning number of seeds

For the offshore industry, the minimum number of 3 h seeds required to derive extreme wave impact loads is set at 20–30 [28], 'a large number' [24], a 'sufficient' number [12], 16 for air-gap and impact load analysis for column-stabilised units [29], 'sufficient' or 10 for 'Gumbel fitting to the maximum values' for air-gap and wave impacts on semi-submersibles [23], or 'multiple' [30]. Large numbers of seeds in different wave conditions were analysed by [31] for slamming impacts on a large-volume cylinder, concluding that 50–60 seeds are required for convergence. Some other references concerning the number of seeds for wave-in-deck

type of impacts are provided in part 2 of this study [1]. For ships, the minimum number of (usually) 1 h seeds required to derive extreme wave impact loads is set at 20–30 [18] or 16 (indirectly derived from a criterion based on the exposure duration and target exceedance probability provided in [32]). An alternative approach suggested by [27,32,33] is to test only 1 seed in such high waves that a large number of events is obtained. This is very efficient, but it requires testing in a very severe condition that may not be representative of the operational regime loads. This alternative is therefore not further treated in the present study. Summarising, there is a preference for 10–30 seeds of 3 h for offshore platforms, or even up to 50–60. For ships, shorter exposure durations are applied with 1–30 seeds. However, the guidelines vary greatly and only a few studies were found that support these numbers with data.

2. Objectives and approach

The previous sections indicate that there is no clear consensus in literature about the number of seeds required for convergence of extreme value statistics. There is more consensus about steady-state exposure durations; therefore, we will consider 3 h, 1 h and 30 min durations. It is common practice in (applied) research wave basins to perform tests in a number of wave seeds with such durations. In the present study we focus on the most probable maximum (MPM) of wave crest heights and wave impact force peaks, Section 4.4 explains why. Part 2 of this study [1] also evaluates extreme values with a longer return period for an offshore structure. The objectives of the present study are therefore to:

- Quantify the convergence of the short-term wave crest and impact force peak MPM from experiments at exposure durations of 3 h, 1 h or 30 min.
- Evaluate how the convergence and accuracy of these predictions is affected by fitting.
- Evaluate whether there is a difference in MPM values when they are obtained based on distributions that include all peaks in a sea state or only the maximum of each seed (ensemble or block maxima).

These objectives were reached by analysing a long-duration wave impact experimental campaign. This dataset contains wave measurements and green water impacts loads on the accommodation of a sailing ferry in irregular waves, with test durations of 34 and 23 h (representing two different wave conditions). For reference, part 2 of this study in [1] discusses similar results for wave-in-deck loads on a stationary deck box, with a test duration of 300 h plus deterministic repeat tests for one wave condition. In both parts of the study, the total test duration was cut up into different combinations of exposure duration and number of seeds. The variability of the MPM values was then evaluated as a function of the number of considered seeds. The effect of fitting is also considered in both parts of the study.

The experiments are described in Section 3, and the analysis methods in Section 4. Three types of probability distribution are distinguished in the present study: based on all peaks in a single seed (DSR), all peaks in multiple seeds (DNR), or on the maximum/'extreme' value per seed (DEM). This is explained in Section 4.3. The resulting wave crest height and green water impact force peak DSR distributions are discussed in Section 5.1 and the fitted equivalents in Section 5.2. The resulting DEM distributions and the difference between the extreme values from DNR and DEM are discussed in Section 5.3. Next, Section 5.4 discusses the convergence of the MPM wave crests and impact force peaks with and without fitting when considering an increasing number of seeds. The fitted results do not converge to exactly the same extreme values as the original measured data; Sections 5.5 and 5.6 discusses the bias introduced by the fitting and extrapolation. Comparison with the results for the other dataset in part 2 of this study is examined in Section 5.7. Finally, some conclusions are drawn in Section 6 and possible future work is discussed in Section 7.

3. Experiments

All results in the present publication are provided at full scale. Free-sailing experiments were done with a model of the 190 m 'MARIN ferry 2' [34] at scale 35.986, see Fig. 1. The freeboard of the model was lowered and the draught increased compared to the original design, in order to increase the probability of green water events. These experiments were done within the project 'SCREAM' (SCReening for Extremes And Maxima in waves) of the Cooperative Research Ships (CRS). The scale, main dimensions and loading condition of this ship are provided in Appendix A. It was equipped with bilge keels and an active propeller and rudder. The tests were done in MARIN's Seakeeping and Manoeuvring Basin, measuring 170 × 40 × 5 m [35]. Appendix A provides information about the instrumentation of the model, the test set-up and the wave probes around it. The present publication mainly uses the integrated green water impact force F_X over all 40 panels on the front of the accommodation and the non-dimensionalised wave elevation η_{W21}/H_s measured at wave probe W21 (see Fig. 12). Details about the force panels are also provided in Appendix A and in [36].

The model was tested in irregular bow-quartering waves at forward speed. The experimental conditions are listed in Table 1. The 'heavy condition' was tested for a duration of almost 35 h, the 'extreme condition' for almost 24 h. These conditions were selected such that a significant number of green water and bow-flare slamming events occurred (this is also indicated in the table). The wave conditions were selected in a steep and rare but still realistic part of a world-wide scatter diagram. 4.6 kn is a realistic speed in the heavy wave condition based on the ship properties and estimated engine power (5 kn is prescribed in extreme conditions by for instance [32]). 9.7 kn may be possible to reach, but the captain will probably not be comfortable with this speed. This is why the second wave condition is considered to be extreme.

The waves were measured around the model during the tests (no undisturbed wave repeat tests were done for the long duration tests) and the model sailed at a relatively low speed. However, the possible influence of radiated and diffracted wave components on the convergence results at probe W21 is considered to be small (see Appendix A).

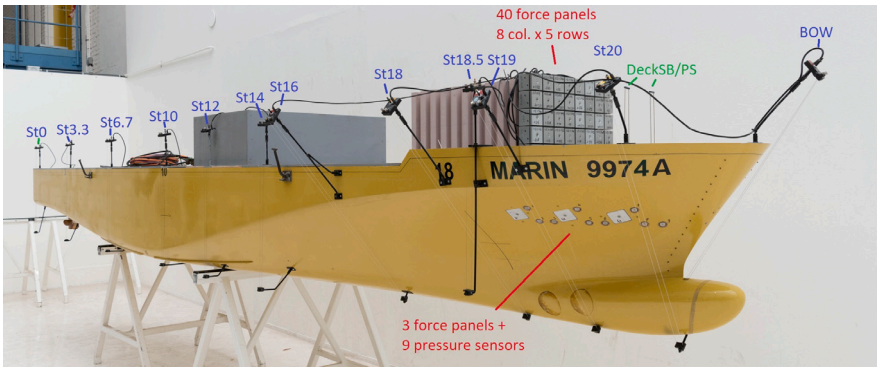


Fig. 1. Model of the MARIN ferry 2 and some of its instrumentation: relative wave elevation probes, force panels on the accommodation and force panels and local pressure sensors in the bow flare.

Table 1
Test programme and maximum measured wave crests and impact force peaks for the ferry at forward speed, where μ = heading (150 deg is bow-quartering waves), H_s = significant wave height, T_p = peak wave period, V_x = mean forward speed, V_y = mean drift speed (kn = knots) and $S = H_s/T_p^2$ = wave steepness. All waves: long-crested JONSWAP wave spectra with peak enhancement factor $\gamma = 3.3$. The number of wave crests (n_e) is the number of encountered crests by the ship in all seeds combined. The number of green water impacts (n_i) is defined similarly.

Condition	Dur. (h)	# runs	μ (deg)	H_s (m)	T_p (s)	V_x (kn)	V_y (kn)	S (m/s^2)	Wave crests		Impact peaks	
									n_e	max. (m)	n_i	max. (kN)
Heavy	34:51:16	60	150	8.1	9.4	4.6	0.7	0.092	16 113	11.1	515	4.09×10^3
Extreme	23:42:49	93	150	8.3	10.0	9.7	1.1	0.083	12 096	11.1	1001	1.06×10^4

4. Definitions and analysis methods

4.1. Concatenation of runs to generate long duration time traces

Because the basin has a finite length, the tests were performed in multiple consecutive non-repeating runs. These were joined in the analysis. At 9.7 kn one run had a duration of around 15 min full-scale, at 4.6 kn this was 35 min. After removing part of the time traces at the start and end of each run (making sure not to cut too close to a wave impact), long duration results were obtained by concatenating the runs. This led to the test durations in Table 1.

4.2. Cutting long duration time traces in realisations with different exposure durations

After the concatenation of the runs, the resulting long test durations were divided into different combinations of exposure duration and number of realisations. An exposure duration of 1 h complies with excellent practice in the ITTC guidelines (see Section 1.3): the tested wave conditions were severe and led to a high frequency of impact events, and the average number of wave encounters per hour was 460–510 in the two test conditions. As explained in Section 2, exposure durations around 3 h, 1 h and 30 min were considered. The final durations are not exactly rounded because the available total duration consisted of an integer number of runs through the basin. The available tests allow for the following number of seeds (M):

- 12 seeds × 2:54:16 h (heavy condition) and 8 seeds × 2:57:51 h (extreme condition);
- 35 seeds × 0:59:45 h (heavy condition) and 24 seeds × 0:59:17 h (extreme condition);
- 70 seeds × 0:29:52 h (heavy condition) and 48 seeds × 0:29:38 h (extreme condition).

4.3. Peaks, ensemble maxima and distributions

As mentioned in Section 3, wave elevation η_{W21} and force F_X are the most important signals in the present study. The crests in the wave elevation time traces were identified using a zero up-crossing analysis, where each peak and trough are after the corresponding zero up-crossing and before the next zero up-crossing. The force signal is not continuous; peaks were only measured during green water events. A peak over threshold analysis was therefore used to identify its peaks. An ensemble maximum is defined as the maximum peak value in a certain exposure duration. Samples, peaks and ensemble maxima in general have different probability distributions, as illustrated for theoretical Gaussian wave crests in Fig. 2. The crests/peaks in wave elevation and accommodation force are indicated by $\eta_{W21,C}$ and $F_{X,C}$, and their ensemble maxima by $\eta_{W21,E}$ and $F_{X,E}$, respectively. In order to avoid contamination of the test results by the concatenation of the runs discussed in Section 4.1, all peaks and maxima within one wave period of a concatenation were removed from the dataset. All wave results were furthermore non-dimensionalised with the significant wave

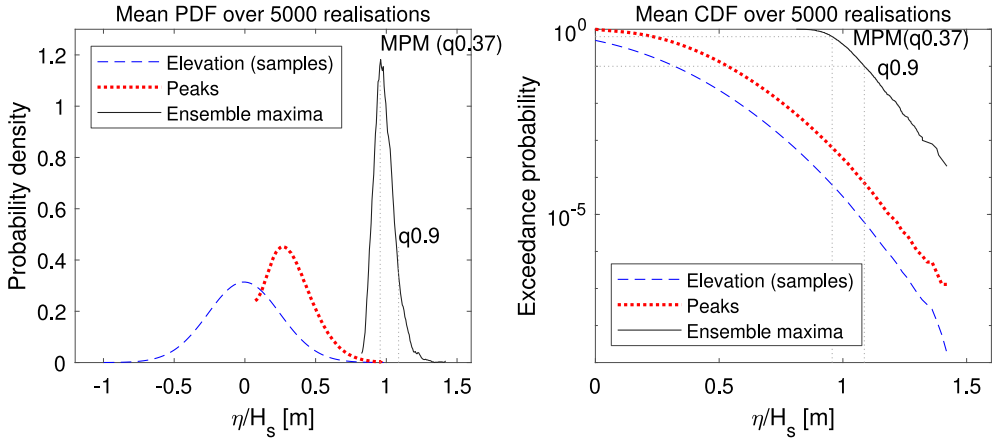


Fig. 2. Typical probability density function and inverse cumulative distribution of samples, peaks and ensemble maxima of wave and response signals — here for 5000 seeds of 3 h linear Gaussian waves with a JONSWAP spectrum and H_s 5 m, T_p 9 s, γ 3.3. The elevation and peak distributions are the mean distributions over all seeds. The MPM (36.8% quantile) and 90% quantile are also indicated.

height H_s. The wave impact forces are presented as absolute values, as there is no logical reference for them. In practice, an uncertainty requirement for structural design forces would probably be related to the maximum allowable design force on the structure.

As mentioned in Section 2, we distinguish three types of exceedance probability distributions. Firstly, a DSR distribution is based on all crests or peaks in a single seed (or realisation): $\eta_{W21,C}$ and $F_{X,C}$ in our case. The probability level in the DSR in the present study is based on the total number of wave encounters n_e for all signals. This is also done for the impact loads, even though they do not peak at every wave encounter. This is required in order to relate the number of impact peaks to a time duration — there are multiple force peaks per exposure duration so this information would be lost if the DSR probability level was plotted based on the number of impact peaks. The probability levels in the DSR are set to $1/n_e$ for the highest value and $(n_e\text{-}1)/n_e$ for the lowest value. Secondly, the DNR distribution is based on all peaks or crests over a number of seeds (or realisations). The DNR is easily derived from multiple DSRs. All peaks in the available seeds are assembled, after which a new distribution is generated. This is effectively a long duration distribution (e.g., 10 realisations of 3 h lead to a 30 h DNR). The probability level in the DNR is also based on the total number of wave encounters for all signals. Finally, a DEM distribution is based on ensemble maxima: $\eta_{W21,E}$ and $F_{X,E}$ in our case. It uses the largest value of each seed for a given exposure duration. The probability level for the DEM in the present paper is based on the number of ensemble maxima (so the number of seeds N), both for the wave crests and the impact force peaks. The probability levels in the DEM are set to $1/N$ for the highest value and $(N\text{-}1)/N$ for the lowest value. For the waves, some theoretical formulations exist for these distributions. These are described in Section 4.5.

4.4. Extreme values

As already mentioned in Section 2, short-term extreme response values used in design are often required in the form of the MPM or a quantile value q. The point at which 90% of the extreme values in the DEM are smaller is indicated by $q = 90\%$, as also shown in Fig. 2. The MPM is the extreme value that is most likely to occur, which is equal to $q = 36.8\%$ for a linear Gaussian signal (see Fig. 2). For ships, the MPM is often considered as design value (e.g., [18] and experience with commercial testing at MARIN). The MPM can be estimated from data by taking the mode of the binned or fitted DEM, or by taking the extreme value with 63.2% exceedance probability (so cumulative probability quantile $q = 36.8\%$, assuming the extreme value distribution approximates a Gaussian distribution, see e.g., [37]). The 63.2% exceedance probability is a more robust definition for datasets that are not very large, as the mode definition is very sensitive to outliers, and the fitting definition may introduce extra fitting errors. Even though we do not consider purely Gaussian signals, the 63.2% exceedance probability definition is therefore applied in the present study. This is formulated in Eq. (1), where $S_E(N)$ are the DEM ensemble maxima values of N seeds, D is their exceedance probability distribution and $\widehat{S_E}(N)$ is the MPM value interpolated from N seeds. We use linear interpolation of the empirical distribution for this. Signal S can either be the non-dimensional wave elevation or the impact force (with the ensemble maxima values $\eta_{W21,E}$ and $F_{X,E}$ defined above). As explained in the previous section, a probability of $1/N$ is used for the highest value in D and $(N\text{-}1)/N$ for the lowest value.

$$\begin{aligned} D(s) &= P\left(S_E(N) \geq s\right) \\ D\left(\widehat{S_E}(N)\right) &= P\left(S_E(N) \geq \widehat{S_E}(N)\right) = 0.632 \\ \widehat{S_E}(N) &= \phi_{0.632}\left[S_E(N)\right] \end{aligned} \tag{1}$$

For offshore structures, quantile values of 85%–95% are considered a reasonable choice for design (e.g., [24,30]). Eq. (1) can still be used in that case, with exceedance probability 0.15 or 0.05 instead of 0.632 and the result called quantile instead of MPM. This is discussed in part 2 of the present study for the deck box [1]; here we only use the MPM.

Alternatively, the MPM can be estimated using a DNR distribution. This is formulated in Eq. (2), where $n_{e,t}$ is the average number of wave encounters in the target exposure duration and P_t is the lowest available exceedance probability level for that exposure duration. $S_C(N)$ are the DNR crest/peak values of N seeds, H is their exceedance probability distribution, and $\widehat{S_C}(N)$ is the MPM value interpolated from N seeds. We use linear interpolation of the empirical distribution for this. Again, signal S can either be the non-dimensional wave elevation or the impact force (with the crest/peak values $\eta_{W21,C}$ and $F_{X,C}$ defined above), or their fitted equivalents. As explained in the previous section, a probability of $1/n_e$ is used for the highest value in H and $(n_e\text{-}1)/n_e$ for the lowest value (where n_e is the total number of wave encounters in all seeds combined).

$$\begin{aligned} P_t &= 1/n_{e,t} \\ H(s) &= P\left(S_C(N) \geq s\right) \\ H\left(\widehat{S_C}(N)\right) &= P\left(S_C(N) \geq \widehat{S_C}(N)\right) = P_t \\ \widehat{S_C}(N) &= \phi_{P_t}\left[S_C(N)\right] \end{aligned} \tag{2}$$

Again, the DNR can also be used to derive higher quantiles for offshore structures. The definition in [29] for peak-over-threshold quantiles is used for this, modified to account for the total number of wave encounters in all seeds n_e as basis for the exceedance probability. P_t in Eq. (2) is now substituted by P_q that is defined in Eq. (3) (the rest of Eq. (2) stays the same). M in this formulation is the total number of available seeds (see Section 4.2). Example: for the MPM in the heavy condition with a duration of 1 h ($q = 0.368$, $M = 35$ and $n_e = 16113$) this reduces to $P_q = 2.169 \times 10^{-3}$, which is very close to $P_t = 1/(16113/35) = 2.172 \times 10^{-3}$ from Eq. (2). The corresponding 90% quantile would be found at $P_q = 2.29 \times 10^{-4}$. This is used in part 2 of the present study for the deck box; here we only use the MPM.

$$P_q = 1 - q_{mod} = 1 - q^{M/n_e} \tag{3}$$

4.5. Theoretical wave distributions

It can be argued that the DSR of wave crest heights is often compared to theoretical distributions [38], whereas these are actually derived for the DNR. These authors therefore introduced new empirical 'Huang-Zhang' upper and lower limits for DSR distributions (for details see Appendix B). The wave crest height results in the present study are compared to these. For the DNR of the wave crest heights, some theoretical forms are available: Rayleigh, Forristall [39], Huang-Zhang and CresT [40]. Details of these distributions are provided in Appendix B. Note that there is a small difference between the DNR and the mean over all available DSR seeds at the same probability level (see also e.g., [38]).

4.6. Fitting

Measured distributions from experiments are often fitted with theoretical extreme value distributions. This is done in order to reduce the statistical variability of the single measured maxima or to extrapolate measured statistics to a longer duration. Especially when experimental results with a short exposure duration are used and/or far extrapolations, this can lead to errors. Wave impacts can be 'badly-behaving problems' [41], where the phenomena that play a role for low-magnitude responses are different than those for high-magnitude responses (e.g., plunging or dambreak versus hammerfist green water events). This makes extrapolation difficult, if not impossible. However, fitting and extrapolation of the DSR based on only a few seeds are commonly used in applied research to obtain extreme wave impact design values. The present study therefore also evaluates the benefits of fitting to the DSR and DNR.

For measured DSR/DNR crest distributions in ship and offshore design, many authors successfully apply or recommend 2- or 3-parameter Weibull fitting (e.g., [18,42–44]). We therefore used the 3-parameter Weibull distribution (Eq. (4)). The distribution was fitted to the data using least-squares, as [24] states that this often gives a better tail fit for Weibull fitting than the method of moments or maximum likelihood estimation. The influence of the higher peaks was further increased by only considering a certain percentage of the highest peaks in the fits. The applied percentage is indicated in each plot in Section 5.2.

$$P(x > X) = \exp\left(-\left(\frac{x-\theta}{\alpha}\right)^{\beta}\right) \tag{4}$$

In the remainder of this paper, we compare the following extreme values and their variability for an increasing number of seeds:

1. MPM from DEM that is based on the measured DSRs ('MPM from measured data').
2. MPM from DEM that is based on the Weibull-fitted DSRs ('MPM from fitted data').
3. MPM based on the measured DNR ('MPM from DNR').

A fourth one could be from a GEV fit to the empirical DEM. However, the objective of the present paper is to evaluate the convergence of the results by increasing the number of included seeds (see next section), which means that we start from very low number of seeds. Obtaining a MPM value from (in the limiting case) only 2 or 3 DEM values is already questionable, but fitting a GEV to only 2 or 3 data points and then obtaining the MPM was judged to be unfeasible. Options 1 to 3 also align with what is done in practice in wave basins to evaluate the MPM of marine structural responses. We therefore stuck to the three options above.

4.7. RMSE convergence metric for MPM values

The main objective of the present work is to evaluate how many seeds are required for a given extreme value accuracy. To this end, the root mean square error (RMSE) for the MPM value over a given number of seeds is introduced. The RMSE defined here expresses the difference between the MPM over a randomly selected number of seeds (1*, 2*, ..., N) and the MPM over all available seeds (M). The latter is seen as the 'true' value (some comments about this assumption can be found in Section 4.9). The indices with a star indicate a randomly picked set of N seeds out of the available M (random permutation without repeating elements). The number N is variable, and M was fixed during the performance of the experiments. By increasing the number of seeds included in the calculation N, its effect on the RMSE can be evaluated.

Each time we randomly pick N seeds leads to a different result. We will call this the number of 'seed picking realisations' W, not to be confused with the random wave seed (also called wave realisation) itself. The lowest number of available seeds M is 8 (3 h extreme condition) and the highest number is 70 (30 min heavy condition). The number of combinations, $n_{options}$, for an 'unordered sampling without replacement' seed picking routine is given in Eq. (5). Calculating the MPM over all combinations is not feasible: $N = 35$ and $M = 70$ for instance gives 1.1×10^{20} options. Instead, we randomly pick 500 times for all combinations of N and M ($W = 500$). This provides a fair indication of the RMSE for larger N: the curves of the RMSE plots in Section 5.4 became quite smooth, whereas they are very peaked for low numbers of W.

$$n_{options} = \frac{M!}{N!(M-N)!} \tag{5}$$

Finally, the RMSE was calculated over W. This is formulated in Eq. (6), where the MPM values $\widehat{S_E}(N)$ are defined as in Eq. (1) and index w refers to the seed picking realisations. We do this for all possible values of N, in order to obtain the convergence RMSE for the MPM values as a function of N. The seed picking routine is 'unordered sampling without replacement': we draw N samples from the set of M, such that ordering does not matter and repetition is not allowed. We used $W = 500$ for our analysis. For example, 70 seeds are available for the 30 min exposure duration in the heavy condition (so $M = 70$). We randomly pick 10 seeds out of these 70 ($N = 10$), and repeat this 500 times ($W = 500$). We then calculate the RMSE over these 500 seed picking realisations, for the MPM over the 10 seeds compared to the MPM value over the fixed 70 seeds. This can be done for any value of N between 1 and M.

$$R(N) = \sqrt{\frac{\sum_{w=1}^{W}\left[\widehat{S_{E,w}}(N) - \widehat{S_E}(M)\right]^2}{W}} = \sqrt{\frac{\sum_{w=1}^{500}\left[\widehat{S_{E,w}}(N) - \widehat{S_E}(M)\right]^2}{500}} \tag{6}$$

The procedure above was used to evaluate the convergence of the extreme values in the measured original data (option 1 in Section 4.6). For the DSR-fitted data (option 2 in Section 4.6), there are two different things to evaluate: the convergence over an increasing number of seeds, and the accuracy of the fitted data compared to the measured data (the bias). The bias is discussed in Section 4.8. The convergence of the fitted data was evaluated in the same way as for the measured data. This was done using Eq. (6), but with S_E in this equation and in the MPM definition of Eq. (1) substituted by the ensemble maxima in the fit (option 2 in Section 4.6) at the same probability levels S_E^f. This results in Eq. (7).

$$R^f(N) = \sqrt{\frac{\sum_{w=1}^{W}\left[\widehat{S^f_{E,w}}(N) - \widehat{S^f_E}(M)\right]^2}{W}} = \sqrt{\frac{\sum_{w=1}^{500}\left[\widehat{S^f_{E,w}}(N) - \widehat{S^f_E}(M)\right]^2}{500}} \tag{7}$$

The RMSE values were calculated for the measured wave and impact data and for the fitted wave and impact data, for exposure durations 3 h, 1 h and 30 min and both wave conditions. The results can be compared to an arbitrary convergence criterion, in order to determine how many seeds are required. This value will depend on the application; it could be a relative uncertainty criterion (e.g., 2% of H_s for the wave crests), an absolute uncertainty criterion (e.g., 200 kN wave impact force for design of a structure on deck), a criterion based on the derivative of the RMSE with respect to the number of seeds (the slope of the lines in Section 5.4) or a combination.

4.8. RMSE bias metric for MPM values

To evaluate the bias of the fits compared to the measured values, a similar formulation was used. This bias RMSE was again defined based on the MPM defined in Eq. (1), and W is again 500. S_E^f here indicates the ensemble maxima in the fit (option 2 in Section 4.6), and S_E again the ensemble maxima in the measured signal. This RMSE value for the fitted data does not converge to zero for larger N, but to a constant value: the bias introduced by fitting the DSRs.

$$B^f(N) = \sqrt{\frac{\sum_{w=1}^{W}\left[\widehat{S^f_{E,w}}(N) - \widehat{S_E}(M)\right]^2}{W}} = \sqrt{\frac{\sum_{w=1}^{500}\left[\widehat{S^f_{E,w}}(N) - \widehat{S_E}(M)\right]^2}{500}} \tag{8}$$

4.9. Assumptions

It is assumed that the MPM value over the maximum number of available seeds M is converged and the 'truth'. This is not necessarily true of course, but it is the best estimate that we have based on the available data.

5. Results and discussion

5.1. Wave crest and impact force peak distributions (DSR and DNR)

The DSRs and DNR of the non-dimensional wave crests measured at probe W21 were derived using the procedure explained in Sections 4.1 to 4.3 and 4.5. The DSRs are shown in Fig. 3 for 3 h, 1 h and 30 min exposure duration (in grey). The number of DSR increases with decreasing exposure duration – towards the bottom of the plots – as more realisations are available in the experiments. This was explained in Section 4.2. The DNR is also in these figures (in red). The DNR is the same for all exposure durations, but different for the two wave conditions. The measured distributions are compared to the Huang-Zhang DSR upper and lower limits for 3 h exposure and to the Rayleigh, Forristall, Huang-Zhang and CresT DNR. These theoretical distributions are specified in Appendix B.1, and the comparison of the measured data with these theoretical distributions is discussed in Appendix B.2.

The variation in DSR over all seeds decreases with increasing exposure duration; more crests are measured per seed, so the variability between the seeds decreases. As expected, the plots also show that the ensemble maximum wave crest heights (the largest value in each seed) increase with increasing exposure duration. This is discussed in detail in Section 5.3. This is a direct consequence of the stochasticity of waves. It confirms again that the exposure duration should be selected based on the expected steady state duration at sea: this will deliver realistic extreme values.

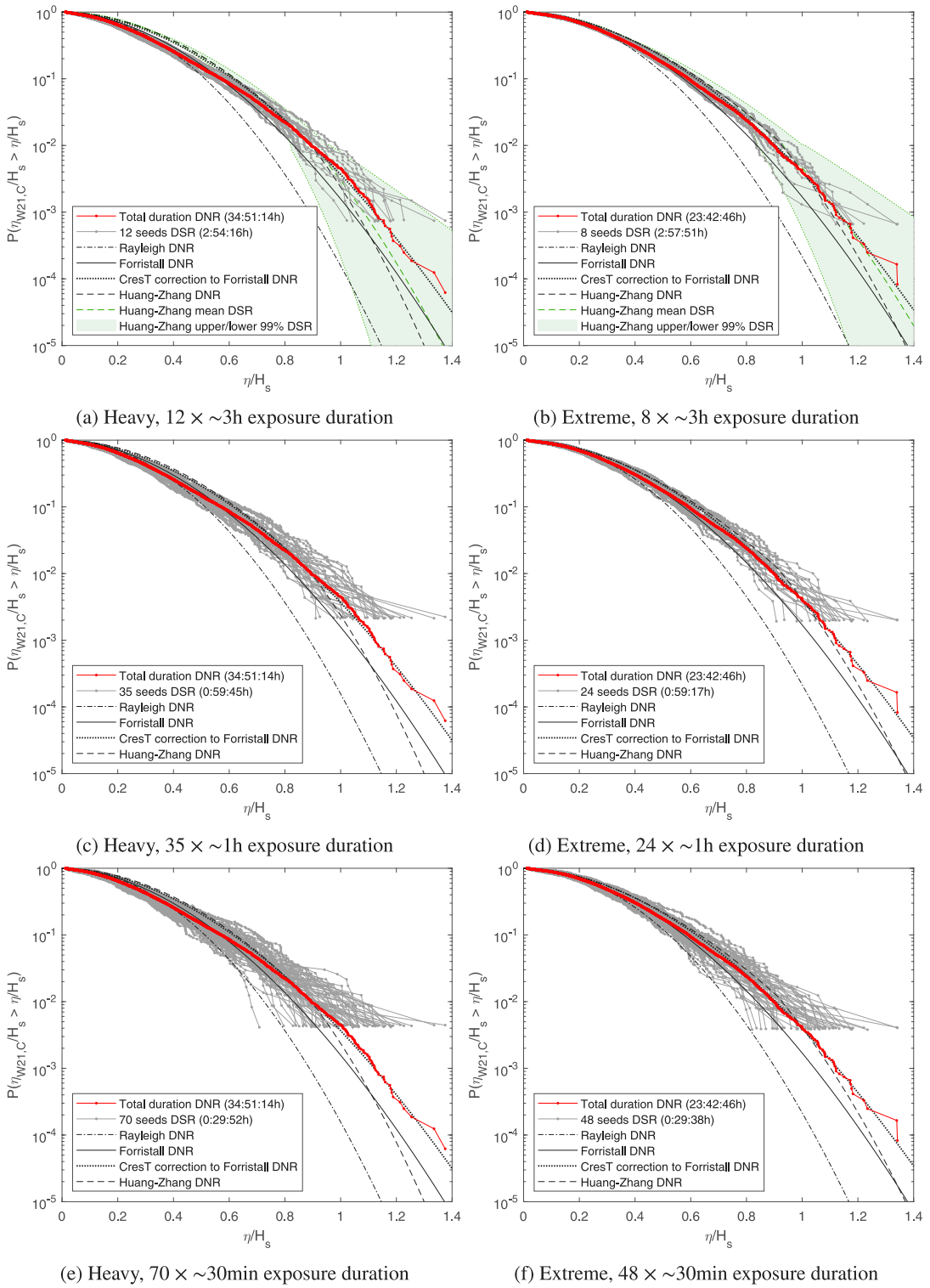
Similar DSRs and DNR are shown in Fig. 4 for the green water impact force peaks. No theoretical distributions are available for these distributions. However, similar trends are visible as for the wave crest distributions: an increasing exposure duration decreases the variability between the seeds, and increases the ensemble maximum peak values (again, see Section 5.3). The variability in the force peaks is a lot larger than in the wave crests however. This can be explained by the fact that variability in green water impact forces is not only influenced by the variability in the incoming waves, but also by that in the vessel speed, vessel motions, water flowing over the deck, etc.

Comparing Figs. 3 and 4 shows that the non-dimensional wave crest distributions of the two wave conditions are very similar, but the impact force peak distributions are very different. This is the case for the dimensional wave crests as well; Table 1 shows that the H_s values of the two conditions are similar (8.1 and 8.3 m, respectively). The extreme condition leads to significantly higher loads than the heavy condition (note that the scale of the x-axis is the same for the non-dimensional crest heights in the two wave conditions, and different for the impact forces). This may partly be due to a slightly different wave steepness, but the biggest difference is the forward speed of the vessel. In the heavy condition it sails at 5 knots, and in the extreme condition at 10 knots. This caused a big difference in impact load magnitude.

5.2. Fitting of the crest and peak distributions

Section 4.6 explained the 3-parameter Weibull fitting procedure we apply to the DSRs. For reference, we first fit the same distribution type to the DNR distributions. This provides an idea of the suitability of the Weibull fit over a large range of crest heights. The result is shown in Fig. 5 for the non-dimensional wave crests and the green water impact force peaks. For both, the 30% highest peaks were used in the fit. The figure shows that Weibull fits the wave crest data very well, up to some limit. This may be due to the influence of wave breaking for the highest and steepest crests. Based on this, Weibull is expected to give a slight overestimation of the wave crest heights when extrapolated to lower probabilities. For the impact force peaks, the fit seems very good in the extreme condition. The fit quality for the heavy condition looks more similar to that for the wave crests: quite good up to some limit force, which may be associated with the onset of wave breaking (or another type of green water event). The green water impacts in the heavy condition seem more correlated to the wave crest height than in the extreme condition. This may be because forward speed plays a large role in the extreme condition. This seems to lead to a more Weibull-following behaviour of the force peaks. Based on this, Weibull is expected to give a slight overestimation of the green water impact forces when extrapolated to lower probabilities for the heavy condition, and a good estimate for the extreme condition.

Fitting the DSRs can in theory be used to reduce the variability of the measured maxima and to extrapolate to longer exposure durations. Each DSR was therefore fitted using the procedure explained in Section 4.6. As explained in that section, this was done because this is a common way to extrapolate commercial tests in only a few seeds to higher quantiles in applied research institutes. For the waves, the 30% highest crests were used for fitting. Fewer events are available per seed for the green water impacts, so the 90% highest force peaks were used for fitting. The results for the non-dimensional waves and green water impacts in the heavy condition for 3 h exposure duration are shown in Fig. 6. Similar results for the other exposure durations and the extreme condition are shown in Appendix C. The number of impact peaks in 30 min was too low for a fit, so this duration was omitted for the impacts. These plots seem to confirm that the variability between the seeds reduces when proper fitting is applied to the DSRs. This would be good news: compared to the unfitted results, it may be possible to get convergence at a lower number of seeds. However, there are a few remarks. Firstly, the plots show that while the fits may in some cases decrease the variability, they may also increase the mean deviation from the DNR. Secondly, whether or not the fits converge quicker needs to be quantified, which is done in Section 5.4. Finally, the quality of the fit plays a large role. This quality is higher for fits based on measured data with a longer exposure duration. This is quantified in Section 5.5. It should also be kept in mind that we use one fitting procedure, where many are possible. Another choice of fit distribution type, fitting procedure or included percentage of peaks may lead to different results. However, the present procedure complies with common practice (as explained in Section 4.6), so it provides a fair indication of what can be expected.



(a) Heavy, 12 × ∼3h exposure duration

(b) Extreme, 8 × ∼3h exposure duration

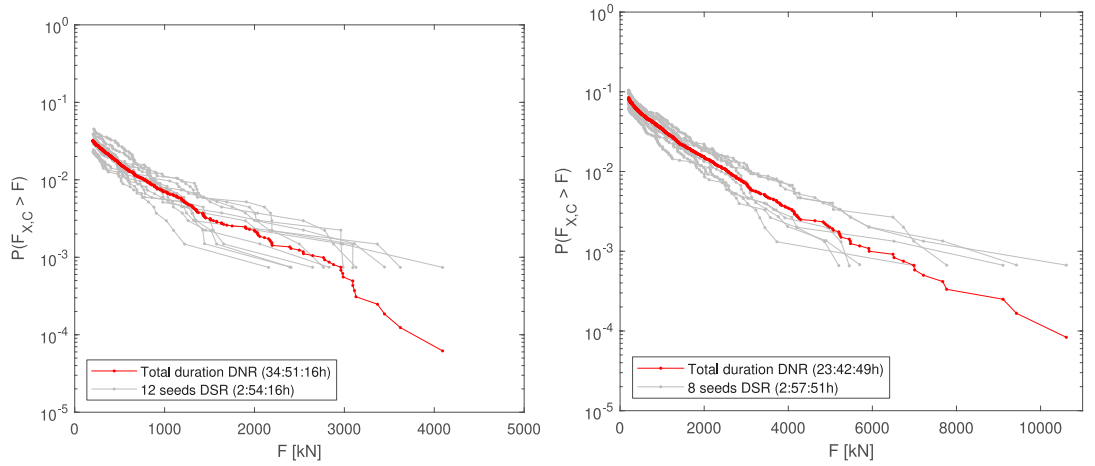
(c) Heavy, 35 × ∼1h exposure duration

(d) Extreme, 24 × ∼1h exposure duration

(e) Heavy, 70 × ∼30min exposure duration

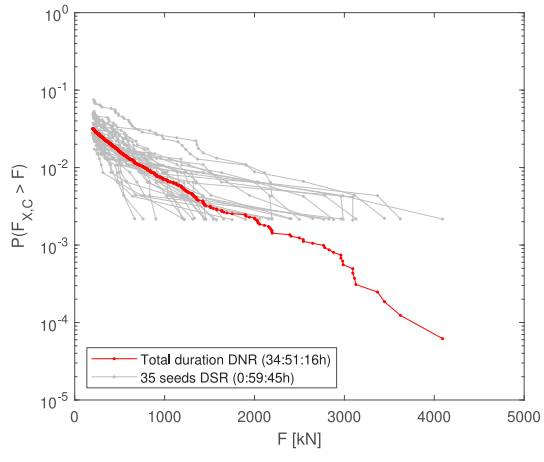
(f) Extreme, 48 × ∼30min exposure duration

Fig. 3. DSR and DNR distributions of $\eta_{W21,C}/H_s$, in the heavy and extreme condition for different exposure durations and number of seeds (including theoretical crest height distributions).

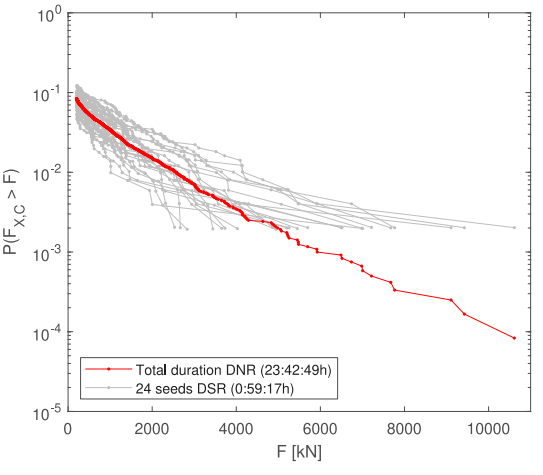


(a) Heavy, 12 × ∼3h exposure duration

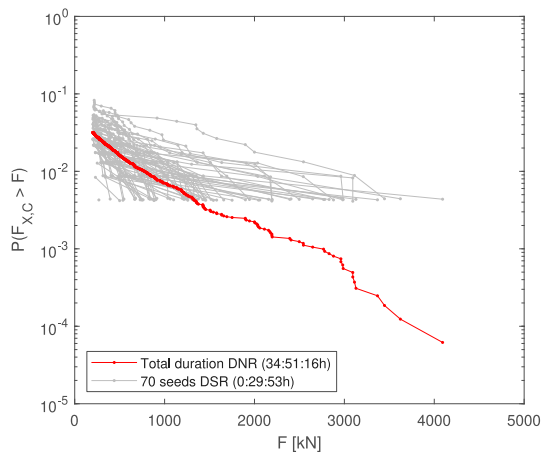
(b) Extreme, 8 × ∼3h exposure duration



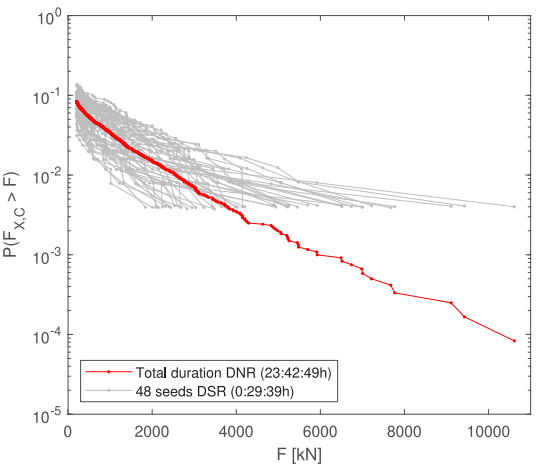
(c) Heavy, 35 × ∼1h exposure duration



(d) Extreme, 24 × ∼1h exposure duration



(e) Heavy, 70 × ∼30min exposure duration



(f) Extreme, 48 × ∼30min exposure duration

Fig. 4. DSR and DNR distributions of $F_{X,C}$, in the heavy and extreme condition for different exposure durations and number of seeds. No theoretical distributions are available for these measurements.

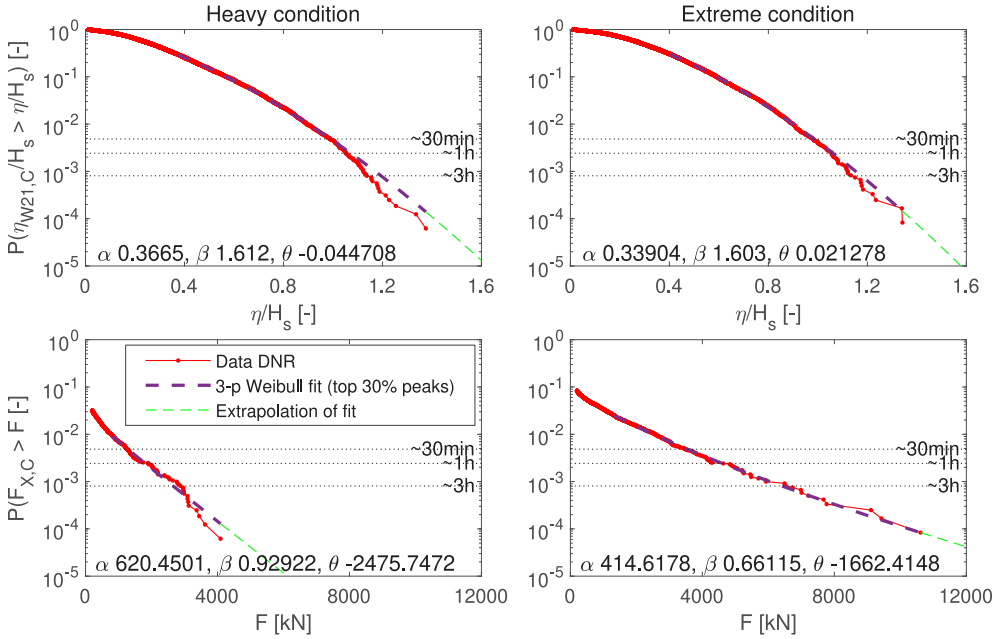
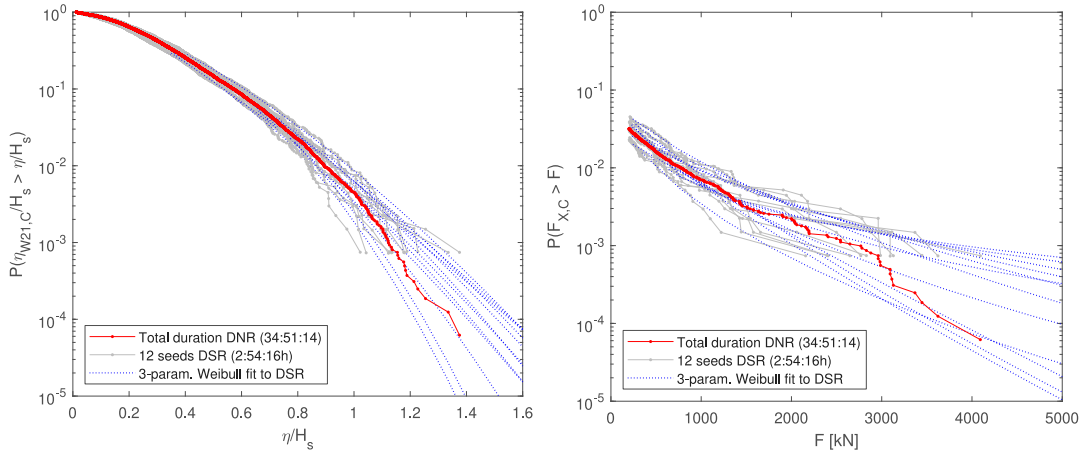


Fig. 5. DNR least-squares fitting result for $\eta_{W21,C}/H_s$ (top) and for $F_{X,C}$ (bottom) with 3-parameter Weibull, considering 30% of the highest crests or peaks. The approximate 30 min, 1 h and 3 h probability levels are also indicated (based on wave peak encounter period).



(a) Heavy, 12 × ~3h exposure duration (b) Heavy, 12 × ~3h exposure duration

Fig. 6. DSR and DNR of $\eta_{W21,C}$ (left) and $F_{X,C}$ (right), in the heavy condition for 3 h exposure duration. Including 3-parameter Weibull fits to top 30% crests (waves) or top 90% peaks (forces) in the DSR distributions.

5.3. Deriving design loads based on DEM versus DNR

This section examines MPM values based on the DEM and the DNR, to see whether the DNR can be used for such an extreme value estimate. A possible limitation of using the DNR for this purpose is the level of independence of the peaks. The MPM values for the wave crest heights and impact force peaks were estimated using the procedures in Sections 4.3, 4.5 and 4.6 and following estimation options 1, 2 and 3 listed in Section 4.6: from the DEM based on the empirical DSRs ('MPM from measured data', Eq. (1)), from the DEM based on the Weibull-fitted DSRs ('MPM from fitted data', Eq. (1)) and from the measured/empirical DNR ('MPM from DNR', Eq. (2)). Comparing these results will show the consistency of the methods and provide an indication of the extreme values for our test case. The present section shows the MPM results for the maximum number of available seeds (M). The next section discusses how the results converge to these values for lower numbers of seeds.

Fig. 7 shows the DEM based on the measured DSRs for the waves and impacts, for both test conditions and three exposure durations. The number of points in 3 h is lower than in 1 h or 30 min, because fewer seeds are available. The figures show that both

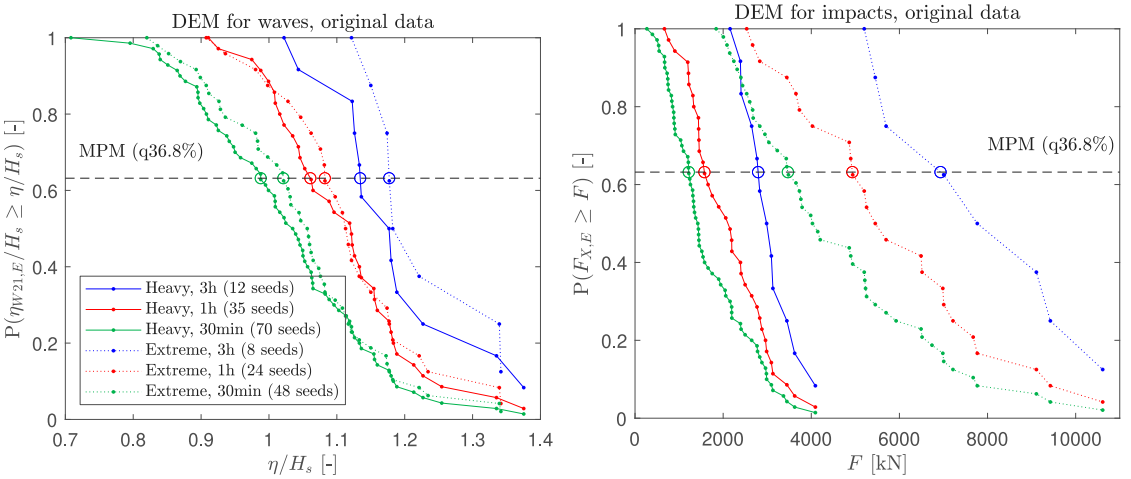


Fig. 7. DEM based on empirical DSRs of $\eta_{W21,E}/H_s$ (left) and $F_{X,E}$ (right) for both test conditions and all three exposure durations, incl. approximation of MPM.

Table 2
MPM values based on the DNR following Eq. (2), versus MPM values based on the DEM following Eq. (1). The DEM here is based either on the measured DSRs ('DEM_m') or on the fitted DSRs ('DEM_f'). This corresponds to options 1 to 3 (in 3,1,2 order) from Section 4.6.

	MPM of η/H_s [–]			MPM of F_X [kN]		
Based on	DNR	DEM_m	DEM_f	DNR	DEM_m	DEM_f
Heavy, 3 h	1.15	1.13	1.22	2.9×10^3	2.8×10^3	3.0×10^3
Heavy, 1 h	1.06	1.06	1.10	1.9×10^3	1.6×10^3	1.8×10^3
Heavy, 30 min	1.00	0.99	1.01	1.3×10^3	1.2×10^3	–
Extreme, 3 h	1.16	1.18	1.23	6.8×10^3	6.9×10^3	6.1×10^3
Extreme, 1 h	1.06	1.08	1.10	4.9×10^3	4.9×10^3	4.4×10^3
Extreme, 30 min	0.99	1.02	1.02	3.7×10^3	3.5×10^3	–

the wave and the impact force ensemble maxima increase with exposure duration (as expected from a stochastic signal). As also observed based on the DSRs in the previous sections, the wave ensemble maxima are relatively similar in the two test conditions, whereas the impact forces differ significantly. This is probably due to the higher vessel speed in the extreme test condition. The obtained MPM values are seen as the 'true' converged values in the next section, which derives the convergence with an increasing number of seeds with respect to this true value. Table 2 shows that the MPM values from the measured DEM in Fig. 7 are close to the DNR values from Fig. 5 at the same exposure duration, even though all peaks are included in the DNR, and only the extreme value for each seed in the DEM. The linear DEM interpolation in Fig. 7 is relatively coarse, so this may have introduced small differences as well. This indicates that the choice to include all peaks or only the ensemble maximum peaks does not matter much for the extreme value prediction in the present case, even though the impact frequency is high.

The table also includes the results based on the DEM from fitted DSRs (option 2 from Section 4.6, see Section 5.2), which are derived from similar (omitted) plots as Fig. 7. Comparison between the MPM from the measured and fitted versions of the DEM show that fitting introduces a bias. This is further examined in Section 5.5.

5.4. MPM convergence for original and fitted data

The distribution plots in the previous sections provide qualitative insights in the variability and maxima of wave crests and green water impact force peaks as a function of the number of seeds. In order to determine how many seeds are needed at a certain exposure duration, quantification is required. This is evaluated using the RMSE convergence metric $R(N)$ defined in Section 4.7, calculated for an increasing number of seeds N. The results for the non-dimensional wave crests and impact force peaks are plotted in Fig. 8. The results on the left are based on the original measured data; the results on the right on the fitted data (option 1 and 2 in Section 4.6, respectively). Note that the 30 min results for the fitted impact force peaks are missing, as the number of peaks per seed was insufficient for a fit. In these plots, the two wave conditions are directly compared in a plot per type of data (original or fitted). In Appendix D, the same plots are reorganised such that the two types of data (original or fitted) are directly compared in a plot per wave condition.

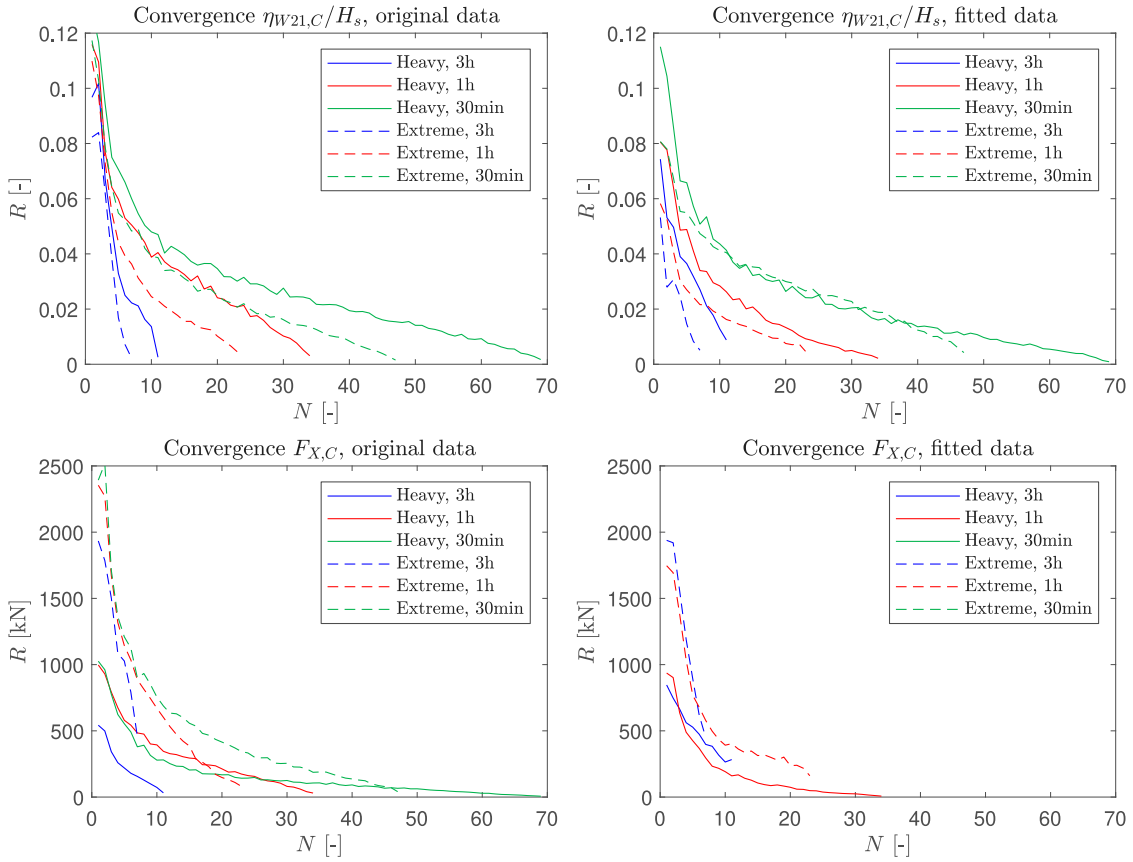


Fig. 8. Convergence metric $R(N)$ and $R^f(N)$ from Eqs. (6) and (7) over 500 seed picking realisations for waves and impacts, original and fitted data. The superscript f is omitted in the y-axis label for the fitted data.

These plots provide an indication of the accuracy that can be expected for the wave crest or impact peak MPM values in different exposure durations and wave conditions, for an increasing number of seeds. They can also be used to derive the required number of seeds in order to obtain extreme wave crest and impact peak results with an arbitrary RMSE uncertainty criterion (see Section 4.7) for the present test case. The following can be observed from Fig. 8:

- For longer exposure durations, convergence is reached with fewer seeds. However, consider that the total required test time may not be shorter (e.g., 6×3 h $> 15 \times 30$ min).
- The MPM values in the heavy and extreme test condition converge at a similar (but not identical) rate. The small differences are probably related to the vessel speed and the details of the waves: steepness, timing, breaking, etc.
- The wave crests and the impact force peak MPM also convergence in a similar and monotonic way; there are no indications in the impact force plots that suggest different loading processes. The present case is a relatively well-behaving problem; for more badly behaving impact problems where different loading processes play a role this may be different.
- The influence of fitting on the convergence is small (see Figs. 15 and 16 in Appendix D). For many combinations of wave condition and duration the process of converging after fitting is slightly slower than before (contrary to expectations), for others it is slightly faster. The effects are small, and we cannot conclude that fitting speeds up or slows down the convergence in general.
- Using only 1 or 2 seeds (as is common practice for ship testing, see Section 1.4) leads to quite inaccurate MPM results. If for instance only 1 seed with a duration of 1 h is tested, the MPM uncertainty expressed as RMSE is around 11% of the H_s for the wave crests, and between 500–2500 kN for the impact forces (5%–24% of the max. measured force during the test campaign).

It might be expected that the number of required seeds is lower for more linear problems and higher for more non-linear or 'badly-behaving' problems (where other phenomena play a role for low and high peaks). For lower sea states it is hard to say what is expected; the waves and impact events are expected to be more linear and well-behaved, but there will also be fewer impact peaks per seed.

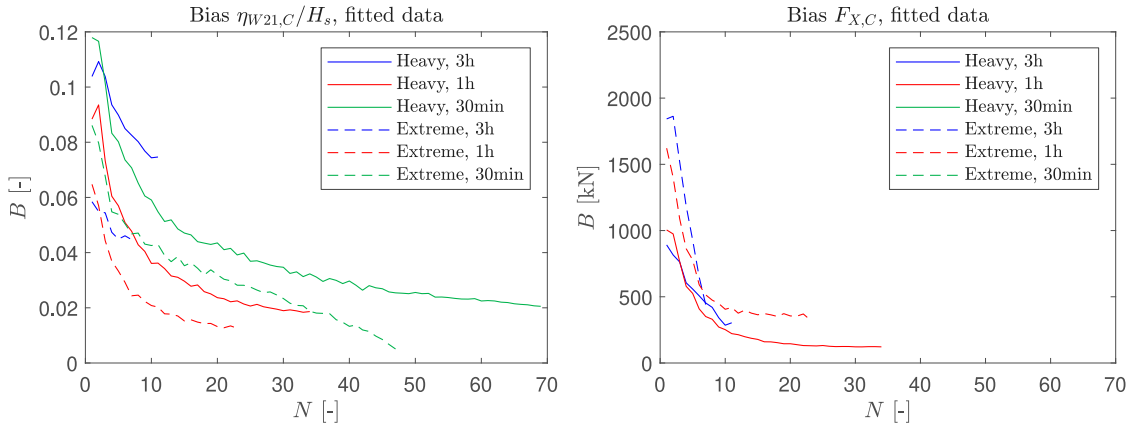


Fig. 9. Bias metric $B^f(N)$ from Eq. (8) over 500 seed picking realisations for waves and impacts, fitted data. The superscript f is omitted in the y-axis label.

5.5. MPM bias due to fitting

Fig. 8 showed the convergence of the RMSE following Eqs. (6) and (7) for the original experiments and the fits, which is a good measure for the convergence of the data. However, Fig. 6 and Appendix C show that the fits may also introduce a bias in the MPM with respect to the long duration data. This was evaluated using bias metric $B^f(N)$ defined in Eq. (8). Results are shown in Fig. 9. These plots show a bias due to fitting that does not converge to zero for large N. The bias for the maximum number of available seeds (M) is similar to the difference between the MPM values based on the fitted and measured versions of the DEM in Table 2. The fitted distributions in Section 5.2 and Appendix C show that the bias for the wave crests is almost always conservative (towards larger wave crests). This is not visible in Fig. 9 due to the positive definition of RMSE bias metric. The conservatism for large crest heights is probably related to the steepness of the wave conditions; wave breaking or other dissipating phenomena reduce the highest wave crest heights, which is not considered sufficiently in the fits based on lower heights. Other wave and wave impact problems do not necessarily show a similar conservatism, see for instance the explanation of badly-behaving problems in Section 4.6. The figure shows that the RMSE bias in MPM due to the fitting goes roughly to ~0.01 to 0.07 for the non-dimensional wave crest heights (1%–7% of H_s), and to very small values for the maximum measured impact forces. This means that the force fits match the long duration quite well.

Similar as for the convergence results, Figs. 17 and 18 in Appendix D show re-organised plots based on the same data, where the fit bias results are plotted together with the fit convergence results per wave condition. These plots show that the RMSE convergence metric and the RMSE bias metric for the fitted result are the same order of magnitude. This could indicate that the considered fits are well-balanced. It also indicates that the convergence may be slightly quicker (or slower) than for the original data in some cases, but the fitted results do not converge to the same value as the original data. They converge to the bias value.

These results depend strongly on the choice of theoretical fit distribution, the fitting method and the number of crests or peaks to include. The present results provide an indication of the results for common practice choices (see Section 4.6). Evaluating the influence of these parameters in detail could be a follow-up study (see Section 7).

5.6. Estimated MPM bias due to fitting and extrapolation

The bias metric $B^f(M)$ at the maximum number of available seeds M (to make sure the bias results are converged) can also be used to evaluate the quality of extrapolation using the fits. For experiments with ships this is a relatively common procedure: tests with a duration of for instance 1 h are fitted and extrapolated in order to obtain 1.5 or 2 h MPM values for design.

In order to evaluate the bias considering extrapolation of the fits, we need a modified bias metric. The reference 'true' value MPM over all available DSR in the original metric in Eq. (8) is not directly available for extrapolation. A modified estimated bias metric $B_e^f(N)$ is therefore defined in Eq. (9), only for extrapolation. The MPM from the DNR was selected here as 'true' value instead of the MPM based on the DEM, because this also allows for a bias estimation of extrapolated fits. Section 5.3 showed that the MPM derived from the DNR is not identical to that from the DEM for this test case, but it is a good approximation. The bias increases quickly with extrapolation 'distance', and it generally becomes an order higher than the difference between the MPM value from the DNR versus de DEM. The exposure duration exceedance probability level P_t can be selected based on the average number of wave encounters in 30 min, 1 h or 3 h in order to evaluate extrapolation.

$$B_e^f(N) = \sqrt{\frac{\sum_{w=1}^{500}\left[\widehat{S_{E,w}^f}(N) - DNR(P_t)\right]^2}{500}} \tag{9}$$

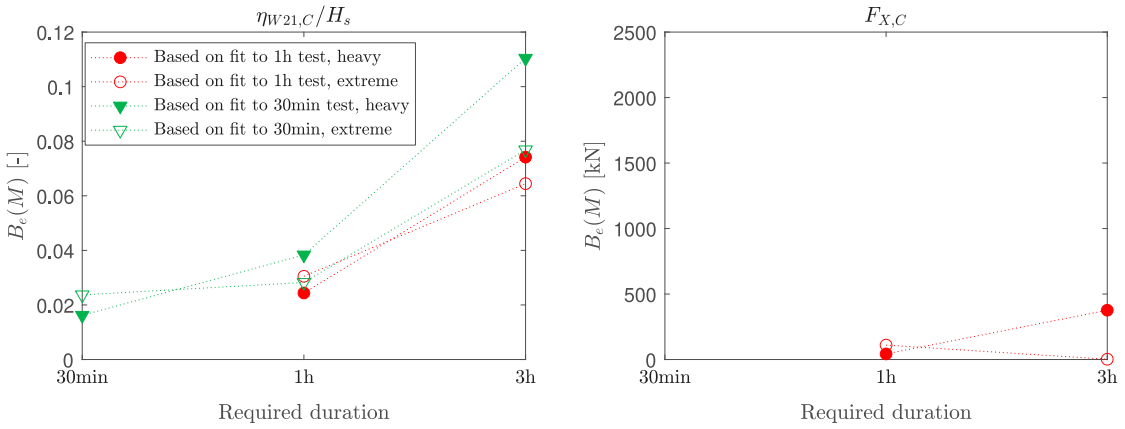


Fig. 10. Estimated converged bias metric $B(M)$ from Eq. (9), for extrapolated results using fitting from short tests to longer required durations (using the fitting procedure in Section 4.6). Wave crests (left) and impact force peaks (right). Legend is the same for both subfigures and superscript f is omitted in the y-axis label.

The resulting converged estimated bias metric $B_e^f(M)$ is visualised in Fig. 10 for both the wave crests and the impact force peaks. The bias in dataset 'based on fit to 1 h test' at 1 h duration on the x-axis in Fig. 10 is equal to the 1 h converged bias for the same wave condition in Fig. 9. This figure shows: the further the extrapolation and the shorter the basis exposure duration, the worse the results. For instance, the wave crest bias from the fitted 30 min duration in the heavy condition is around 0.02 (2% H_s), and extrapolation leads to a bias around 0.04 at 1 h and 0.11 at 3 h exposure durations (green triangles). Similar fitting trends are valid for the extreme wave condition. It is also observed that fitting and extrapolation lead to quite accurate results for the impact data in this case. This is also visible in Appendix C. This probably has two causes. Firstly, wave breaking introduces another regime for the highest wave crests and reduces the extrapolation accuracy for the wave crests. This is not as clearly visible in the impact peak distributions. Secondly, due to the vessel freeboard, there is a 'threshold' for the impacts: only the highest wave crests lead to impacts, which may all be in the same wave regime. A solution for the waves could be to use an even lower percentage of the highest crests, but this is not further evaluated here. Again, these results depend on the choice of theoretical fit distribution. Results for a common procedure are provided here.

5.7. Comparison of the ferry and deck box results

Similar analyses and results are presented for 300 h of wave-in-deck forces on a stationary deck box in [1] (part 2 of the present study). The case studies were similar: wave elevations and wave impact forces were measured using similar instrumentation in long duration experiments. There were also some important differences:

1. There were on average only 2–3 wave-in-deck impacts on the deck box per 3 h seed; this was 44 (heavy) or 127 (extreme) for the green water impacts on the ferry (Table 3). Obviously, this makes a difference for the convergence of the impact forces; this is not only related to exposure duration, but also to the number of events in that duration. For this reason, the deck box impact results were presented for longer exposure durations: 60 h, 30 h and 15 h. The wave crests were evaluated for 3 h and 1 h in both case studies.
2. The steepness of the wave conditions was different (Table 3). Usually it is assumed that some wave crests start breaking around steepness values of 0.06–0.08 m/s^2. The deck box conditions were in the transition area from non-breaking to breaking waves, whereas the ferry conditions are above the breaking limit. There may therefore be many breaking crests in one deck box seed, and few in another. The waves for the ferry are above the breaking limit and breaking was frequently observed in all seeds. This is supported by the variation between DSRs in Fig. 11: there is less variation between the DSRs for the ferry than for the deck box. For the ferry, most distributions clearly show the downward trend expected for breaking waves in the tail of the distribution (observed by e.g., [40,45,46]). For the deck box data this is missing for some seeds and present for others. This could introduce more variability and slower convergence of waves with a steepness around the start of breaking.
3. The wave crests directly hit the deck box, whereas they first ran over deck before impacting the ferry accommodation. Direct impacts increase the influence of complicated impact properties such as the relative wave steepness with respect to the structure and air entrapments (see e.g., [47,48]). Steep, almost breaking wave crests close to the structure may lead to another load regime for the deck box than the other waves, in line with a badly-behaving problem. This is less likely for the green water loads, which makes them easier to fit and extrapolate.
4. For the measurement of the crest heights different types of instrumentation were used in the two test campaigns. For the ferry test campaigns acoustic type wave probes were used. These types of wave probes are typically used for wave measurements at forward speed. For the deck box wave measurements resistant type wave probes were used. The two types of wave probes are based on different measurement techniques. Differences in measured crest heights can be observed especially in very

Table 3
Wave conditions for the ferry and deck box. All: irregular, long-crested, JONSWAP, $\gamma = 3.3$. With: $S = H_s/T_p^2$ = wave steepness, λ_p = peak wave length, x = distance to the wave generator, NS = North Sea.

Campaign	Condition	Duration (h)	H_s (m)	T_p (s)	S (m/s^2)	λ_p (m)	x (m)	x/λ_p	# wave crests	# impacts
P1, Ferry	Heavy	34:51:16	8.1	9.4	0.092	138	360–5700	2.6–41.7	16 113	515
P1, Ferry	Extreme	23:42:49	8.3	10.0	0.083	156	360–5700	2.3–36.9	12 096	1001
P2, Deck box	10,000 yr NS	300:00:00	17.0	15.9	0.067	395	1500	3.8	79 859	254

Table 4
Number of seeds required for convergence of the wave crest MPM (original data) following the analysis in Section 4.7, for convergence criterion $R(N) = 3\%\ H_s$ and the experimental campaigns in part 1 and 2 of this study.

Exposure duration	Required N (P1, Heavy)	Required N (P1, Extreme)	Required N (P2, 10,000 yr NS)
30 min	23	15	–
1 h	17	8	22
3 h	6	5	14

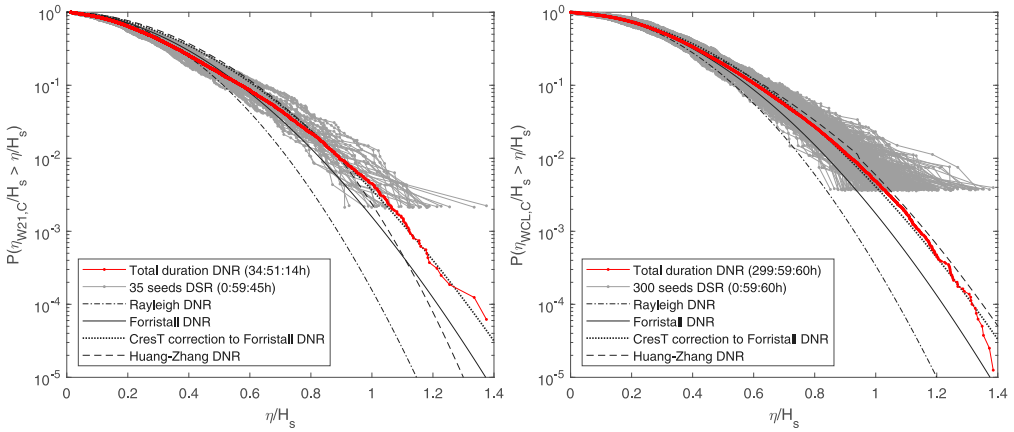
steep and breaking waves, as presented in [49]. The type of measurement probe might therefore also have an effect on the variability of the crests.

5. The deck box tests were done at zero speed and the ferry tests at forward speed. In theory this should not matter, but on average the ferry was farther from the wave generator than the deck box (Table 3). As wave energy can only be inputted at the generators, this may have some influence (especially for steep near-breaking waves, see e.g., [50]). The distance to the wavemaker x divided by the wave length λ_p is much lower for the deck box than for the average ferry distance (Table 3). It was shown by [51] that higher-order non-linear wave effects converge slowly with distance to the wave generator (even for identical wave spectra at different locations), and that the wave non-linearity is converged around $x/\lambda_p = 20$–30. The non-linearity of the waves is therefore probably more converged at the average ferry distance than at the deck box distance. This may cause additional variability in the deck box wave crests.
6. The tests were done in different basins, with different wave generators and instrumentation. Both are position-controlled flap-type wave generators, controlled with similar software. The water depths of both basins can be considered deep at the tested scales. This is therefore not expected to introduce differences in the results.

For the purpose of comparison of the non-dimensional wave crest MPM convergence in the two data sets, the convergence criterion for the RMSE in Section 4.7 is set at 0.03 (so at 3% of the H_s values) for both test cases. Based on this criterion, the required number of seeds can be derived from the convergence plots (Fig. 8 and a similar plot in the deck box paper). The results are presented in Table 4 for the two data sets. This table summarises the results from Section 5.4 and the similar section in the deck box paper. The deck box data were not analysed for 30 min, as it is not common practice to apply this duration in the offshore industry. The table shows that the waves in both case studies converge at a similar rate, although the deck box waves seem slightly slower. This can probably be attributed to points 2, 3, 4 and 5 above. Because the deck box air-gap impacts are much rarer than the ferry green water impacts (point 1 above) and because there is no obvious reference force to relate the absolute forces in both cases, these results cannot be directly compared.

6. Conclusions

The first objective of the present study was to quantify the convergence of the short-term wave crest and impact force peak MPM from experiments at exposure durations of 3 h, 1 h and 30 min. Convergence plots were made, showing the RMSE in MPM values as a function of the number of seeds (or realisations). These can be used to obtain an idea of the expected error in the extreme values when a certain number of seeds is tested in an experiment, or vice versa, to determine how many seeds are required in order to obtain a target uncertainty. This was evaluated for two test cases, green water impacts on a sailing ferry in the present study and wave-in-deck impacts on a stationary deck box in [1]. The required number of seeds for impact force peak MPM convergence cannot be directly compared, because the deck box impacts were much rarer than the green water impacts on the ferry (on average 1 versus 15–42 impacts per hour). Taking an RMSE criterion of 3% H_s for the wave crest MPM values results in 5–14 required 3 h seeds for both cases, or 8–22 required 1 h seeds. This is based on two test campaigns with in total three wave conditions in relatively steep waves. The difference between the wave crest MPM convergence of the two test campaigns is larger than between the two wave conditions in one campaign. This can have several possible causes, including differences in wave steepness, wave measurement equipment, average distance to the wave generator, or possibly other differences between the basins. The convergence of the green water force MPM values on the ferry presented here is representative for structures with a relatively high impact frequency, whereas the deck box results are more representative for structures with a low impact frequency. Using only 1 or 2 seeds leads to quite inaccurate MPM results: for 1 seed with a duration of 1 h, the RMSE uncertainty in the MPM is ~11% of the H_s for the wave crests, and ~5%–24% of the max. measured green water force on the ferry accommodation.



(a) Ferry, $S = 0.092$ (Fig. 3) (b) Deck box, $S = 0.067$ ([1])

Fig. 11. DSR and DNR distributions of wave crest height around 1 h exposure duration for the ferry (heavy wave condition) and deck box (repeat 1 of the selected wave condition).

The second objective of the study was to evaluate how extreme value convergence is influenced by fitting and extrapolation. Using a 3-parameter Weibull fit, it was concluded that fitting does not significantly affect the number of required seeds for convergence of the wave crests or impact force peaks. For the MPM wave crests, fitting introduces an RMSE bias of 2%–3% of the H_s for exposure durations between 30 min and 3 h. Extrapolation increases this and the further the extrapolation, the worse the results for the MPM wave crests. This is probably because breaking of the highest wave crests dissipates energy, which is not considered well by the fits. For the MPM wave impact force peaks, fitting only introduces a very small RMSE bias (0.1%–10% of the max. measured forces) for exposure durations between 30 min and 3 h. Extrapolation sometimes increases and sometimes decreases these biases. These results depend strongly on the choice of theoretical fit distribution, the fitting method and the number of crests or peaks to include. The present results provide an indication of the results for common practice choices. Evaluating the influence of these parameters in detail could be a follow-up study (see Section 7).

The final objective was to evaluate whether there is a difference in extreme value results when they are obtained based on DNR or DEM distributions. It was concluded that this difference is very small for the present ferry wave impact dataset, even though the impact frequency is high.

7. Future work

As explained in Section 1, the present results will be used as reference for an extreme value prediction method validation study. Some other things would be interesting to evaluate based on the present results. Firstly, the applied 3-parameter Weibull/least-squares fitting method was based on literature and best practice at MARIN. The best fitting method is probably case-specific, but studying its influence may further improve results. Secondly, it was observed that the difference between the wave crest MPM convergence of the two test campaigns is larger than between the two test conditions in one campaign. This can have several possible causes (as detailed in Section 5.7 and summarised in the conclusions). Further research could shed more light on the most likely cause. Thirdly, this study (part 1 and 2) is based on long duration experimental campaigns for only two different structures and in total three test conditions. The cases are quite different (green water in a sailing ship and air-gap impacts on a stationary deck box). The resulting number of required seeds can therefore be used as order of magnitude for extreme values of wave impacts on other marine structures. However, performing similar tests and analyses for more wave conditions may provide more clarity on the effects of wave steepness, non-linear wave-wave interaction, wave breaking and test specifics (instrumentation, basin, wave generation). Repeating similar long duration wave impact experiments with other marine structures would also be valuable for less structure dependence. Finally, it would be interesting to see how the present results compare to required durations for convergence of other non-linear and rare response phenomena (such as broaching, parametric roll, slamming, etc.)

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.

Data availability

The authors do not have permission to share data.

Table 5
Properties and loading condition of the ferry, where L_{pp} = the ship length between perpendiculars, B = its width, T = its even keel draught, Z_{bow} = the freeboard at the bow, LCG, VCG = the vertical and longitudinal location of the centre of gravity (from APP and keel), k_{xx}, k_{yy}, k_{zz} = the radii of inertia, GM = the metacentric height and T_ϕ = the ship natural roll period. All values except the main dimensions were measured.

Scale (–)	L_{pp} (m)	B (m)	T (m)	Z_{bow} (m)	LCG (m)	VCG (m)	k_{xx} (m)	k_{yy} (m)	k_{zz} (m)	GM (m)	T_ϕ (s)
35.986	190.0	30.0	9.0	10.5	87.7	14.3	10.6	47.3	47.6	0.98	24.7

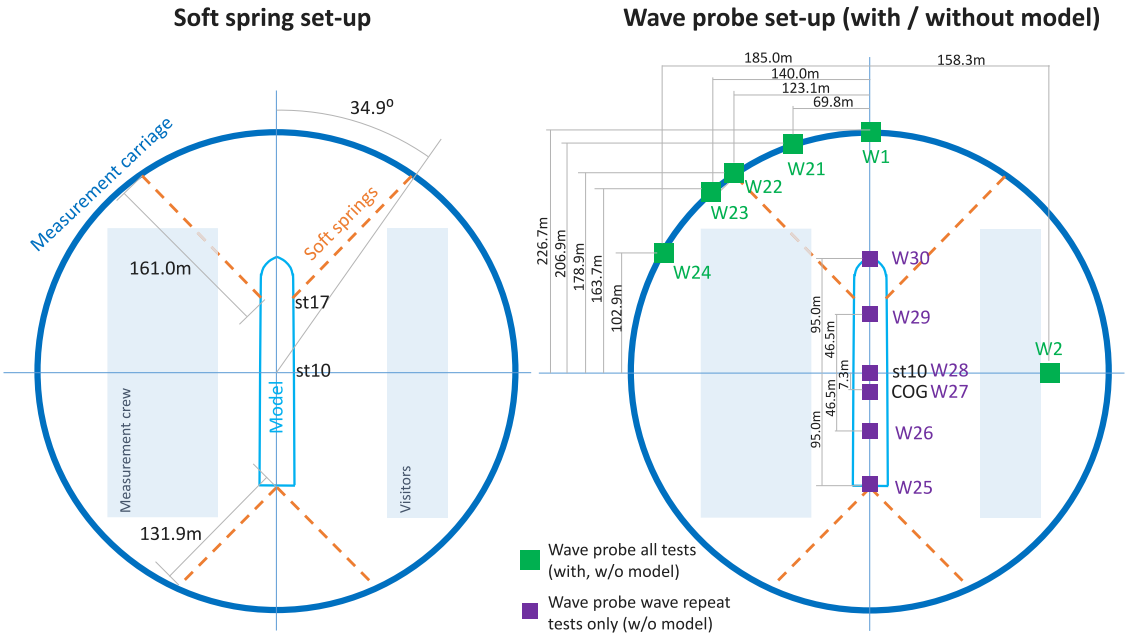


Fig. 12. Soft spring set-up for the ferry (left) and the corresponding wave probe set-up (right, with and without model).

Acknowledgements

This publication is part of the project "Multi-fidelity Probabilistic Design Framework for Complex Marine Structures" (project number TWM.BL.019.007) of the research programme "Topsector Water & Maritime: the Blue route" which is (partly) financed by the Dutch Research Council (NWO). The authors thank the Cooperative Research Ships (CRS) for the use of the ferry data (more information: https://www.crships.org/).

Appendix A. Extra experimental information

The main dimensions and loading condition of the ferry are provided in Table 5. Other details about the experiments are provided below.

Test set-up. The ferry was tested in a free-sailing set-up with an active propeller, rudder and autopilot steering on heading. A constant speed was required for the measurement carriage and the attached wave probes, for later deterministic numerical reproduction of the wave conditions. The carriage therefore moved at a constant x- and y-speed instead of following the model in a 'normal' set-up. This required addition of some soft springs that reduced the low-frequency surge, sway and yaw motions, to keep the model under the carriage. The y-speed of the carriage was derived from the average drift speed during free-sailing trial runs at the same speed in the same wave condition. Fig. 12 on the left shows the test set-up. The natural surge, sway and yaw periods of this set-up were 54, 67 and 34 s at speed, respectively (derived from decay experiments). The springs only limited the low-frequency horizontal motions, and kept the horizontal wave–frequency motions and all other degrees of freedom and as free as possible. A constant propeller rotation rate was applied for each test, which was estimated based on the trial runs in the same condition.

Instrumentation. The model was instrumented to measure the six degree of freedom ship motions (optical tracking system), rigid-body accelerations at 5 locations and relative wave elevation at 12 locations along the weather side of the ship at 2 locations on the fore deck (resistance-type probes). Fig. 1 shows the names of the sensors. Green water wave impact forces were measured at 40 panels on the front of the accommodation. More information about these panels is given below. Slamming forces were measured using 3 force panels and 9 pressure sensors in the bow flare. The incoming waves around the ship were also measured at 6 probes, attached to the measurement carriage (see Fig. 12 on the right). Two of the incoming wave probes (W1 and W2) were acoustic probes,

all others were resistance-type probes (including W21). The undisturbed waves without model were measured in deterministic repeat runs for 8 selected runs with a large green water impact. All probes in green in Fig. 12 were present for all tests (with or without the model); the purple probes were added only for some undisturbed wave repeat runs without model. The ship motions were measured at a full-scale sample frequency of 16.7 Hz, incoming waves at 16.7 or 33.3 Hz, relative wave elevation at 33.3 Hz and forces and pressures at 800.2 Hz.

The ferry accommodation force panels were placed in 5 rows of each 8 panels, at 13.0 m behind the forward perpendicular (FPP). Each force panel had a size of 1.8 × 1.8 m full-scale. Previous research showed that the rise times of green water impacts can be around 0.10–0.35 s (full-scale), and that hydro-elasticity plays a small role for such impacts [52]. This indicates that the interaction between the loads and the deformation of the structure is low, so they can be decoupled. The loads on the selected deck structures were therefore conduced with stiff force measurement panels mounted on a stiff structure. The natural frequency of the panels was around 2100 rad/s full-scale in the main force direction. This is orders faster than the typical rise times, so any possible resonance was small, and could be filtered in the analysis. No correction was applied for the inertial loads on the panels due to the global ship motions, because these forces were very small (max. 2.7 kN on one panel, compared to impact forces in the order of 10^3 kN).

Radiated and diffracted wave components. The waves used in the present publication were measured at probe W21 at portside front of the model. The model sailed at a relatively low speed. This means that there may be some influence of radiated and diffracted waves from the model on these results (Brard number $\tau = V_s\omega/g$ ~0.18–0.20, energy may propagate forward for $\tau < 0.25$). However, for a quite similar case (containership of 230 m, scale ~38 at 5 kn speed in waves with T_p 9.7 s, so $\tau = 0.17$), Rankine-source diffraction calculations were performed in the preparation of [50,53] to evaluate this. This was done in the same test set-up and basin as the ferry tests. For the containership the results at probe W1 showed radiated and diffracted wave elevations around 3% of the incoming wave elevation. The present model is slightly smaller and more slender (ferry of 190 m, scale ~36 at 5 and 10 kn speed in waves with T_p 9.4–10.0). The considered probe was W21 instead of W1, but these are close. A similar or slightly lower influence of radiated an diffracted wave components can therefore be expected. There is no reason to assume that these small components converge slower or faster than the incoming wave components, so this will not significantly influence the convergence results in the present paper.

Appendix B. Theoretical wave crest distributions

B.1. Formulations

The Rayleigh DNR wave crest distribution for Gaussian narrow-banded wave spectra is provided in Eq. (10), where H_s is the significant wave height and η is the wave crest height.

$$P_{Ray}(\eta_c > \eta) = \exp\left(-\frac{8\eta^2}{H_s^2}\right) \tag{10}$$

The Forristall DNR wave crest distribution [39] for second-order waves is provided in Eq. (11a). It is based on a 2-parameter Weibull distribution, with α and β parameters that are based on fitting to second-order wave simulations. These parameters are provided for long-crested waves in Eqs. (11b) and (11c), where S_1 is a wave steepness parameter (Eq. (12)) and U_r is the Ursell number (Eq. (13), characterising the effect of water depth on wave non-linearity). T_1 in these formulations is the mean wave period, d the water depth and k_1 the corresponding wave number.

$$P_{For}(\eta_c > \eta) = \exp\left(-\left(\frac{\eta}{\alpha H_s}\right)^{\beta}\right) \tag{11a}$$

$$\alpha = 0.3536 + 0.2892S_1 + 0.1060U_r \tag{11b}$$

$$\beta = 2 - 2.1597S_1 + 0.0968U_r^2 \tag{11c}$$

$$S_1 = \frac{2\pi H_s}{gT_1^2} \tag{12}$$

$$U_r = \frac{H_s}{k_1^2 d^3} \tag{13}$$

A rough indication of the expected DNR based on analysis in the CresT project [40] is also included in the present paper: it was argued that a correction of 4%–8% above the Forristall distribution should be applied for crest heights in waves with low directional spreading. As these percentages are expected to increase with decreasing spreading (we have long-crested waves) and with wave steepness (we have relatively steep waves), 8% was applied. This theoretical distribution is called 'CresT correction to Forristall DNR' in the present paper.

The Huang-Zhang theoretical wave crest height distributions were based on a Weibull fit to 1000 seeds of 3 h non-linear wave simulations. There are two versions of the distribution: for DSR limits (mean and upper and lower 99% bands) and for DNR [38]. For reference, DSR is called 'PDSR' in their publication, and DNR is called 'PDER' (the present publication sticks to DSR and DNR). The basic formulation for all these distributions is the Weibull formulation (provided in Eq. (14a)). The same definitions for wave

Table 6
Coefficients Huang-Zhang distributions [38].

Type	DSR, mean	DSR, upp99%	DSR, low99%	DSR, mean	DSR, upp99%	DSR, low99%	DNR	DNR
P	$<10^{-2}$	$<10^{-2}$	$<10^{-2}$	$>10^{-2}$	$>10^{-2}$	$>10^{-2}$	$<10^{-2}$	$>10^{-2}$
a_0	0.2894	0.1334	0.3752	0.3712	0.3516	0.3708	0.3242	0.3733
a_1	12.3011	13.0432	9.1269	1.0087	2.3892	0.9835	11.7467	0.9398
a_2	−662.632	−751.0935	−446.8393	−43.0667	−105.4167	−39.9404	−652.147	−40.0095
a_3	12 153.3466	14 727.6571	7837.572	567.5292	1557.3914	598.3819	12 308.3001	512.0601
a_4	−68 031.8045	−87 711.081	−42 682.1177	−1173.1204	−5807.4512	−2144.492	−70 529.2504	−849.0734
a_5	0.3779	0.084	0.8351	0.1276	0.0791	0.2427	0.3785	0.1294
a_6	−3.7904	−5.2045	−5.7942	0.3115	0.355	−0.5379	−3.8837	0.2882
b_0	1.5277	0.7965	2.4637	2.006	1.62	2.212	1.7321	2.0411
b_1	67.2118	44.9229	82.5688	4.841	19.72	10.951	72.7179	3.6068
b_2	−3683.1338	−2525.6956	−4300.5362	−321.181	−909.665	−637.673	−4093.6916	−272.2806
b_3	63 759.0846	47 184.1324	68 857.4691	846.332	10 465.556	6707.056	72 132.2957	−58.9649
b_4	−336 712.3631	−270 084.1075	−337 914.7489	27 223.189	−22 952.443	−10 244.141	−386 504.7502	32 786.2976
b_5	−8.1382	−13.316	2.5661	0.832	−3.726	4.505	−9.9594	0.9003

steepness parameter and the Ursell number are still used. The values of coefficients a_0 to a_6 and b_0 to b_5 were based on non-linear wave simulations. They were defined separately for probabilities below and above 10^{-2}, see Table 6. No smoothing between the regimes was applied. NB. A small mistake in the original publications was corrected: the a_4 value in Eq. (14b) was given to the power of four. This power was removed in order to arrive at realistic distributions.

$$P_{HZ}(\eta_c > \eta) = \exp\left(-\left(\frac{\eta}{\alpha' H_s}\right)^{\beta'}\right) \tag{14a}$$

$$\alpha' = a_0 + a_1 S_1 + a_2 S_1^2 + a_3 S_1^3 + a_4 S_1^4 + a_5 U_r + a_6 U_r^2 \tag{14b}$$

$$\beta' = b_0 + b_1 S_1 + b_2 S_1^2 + b_3 S_1^3 + b_4 S_1^4 + b_5 U_r^2 \tag{14c}$$

B.2. Comparison of measurements with theoretical wave distributions

As explained in Section 3, the waves for the ferry tests were measured with the model present, and there may be a small influence of radiated and diffracted waves from the model on the measurements at wave probe 21. In addition, the wave conditions were quite steep and some waves dissipate energy by breaking. As the wave generators can only input energy at the basin sides, this may lead to some spatial non-uniformity of the wave conditions over the basin length. However, by comparing these measurements to the theoretical wave crest distributions, we can still get a rough idea of their non-linearity.

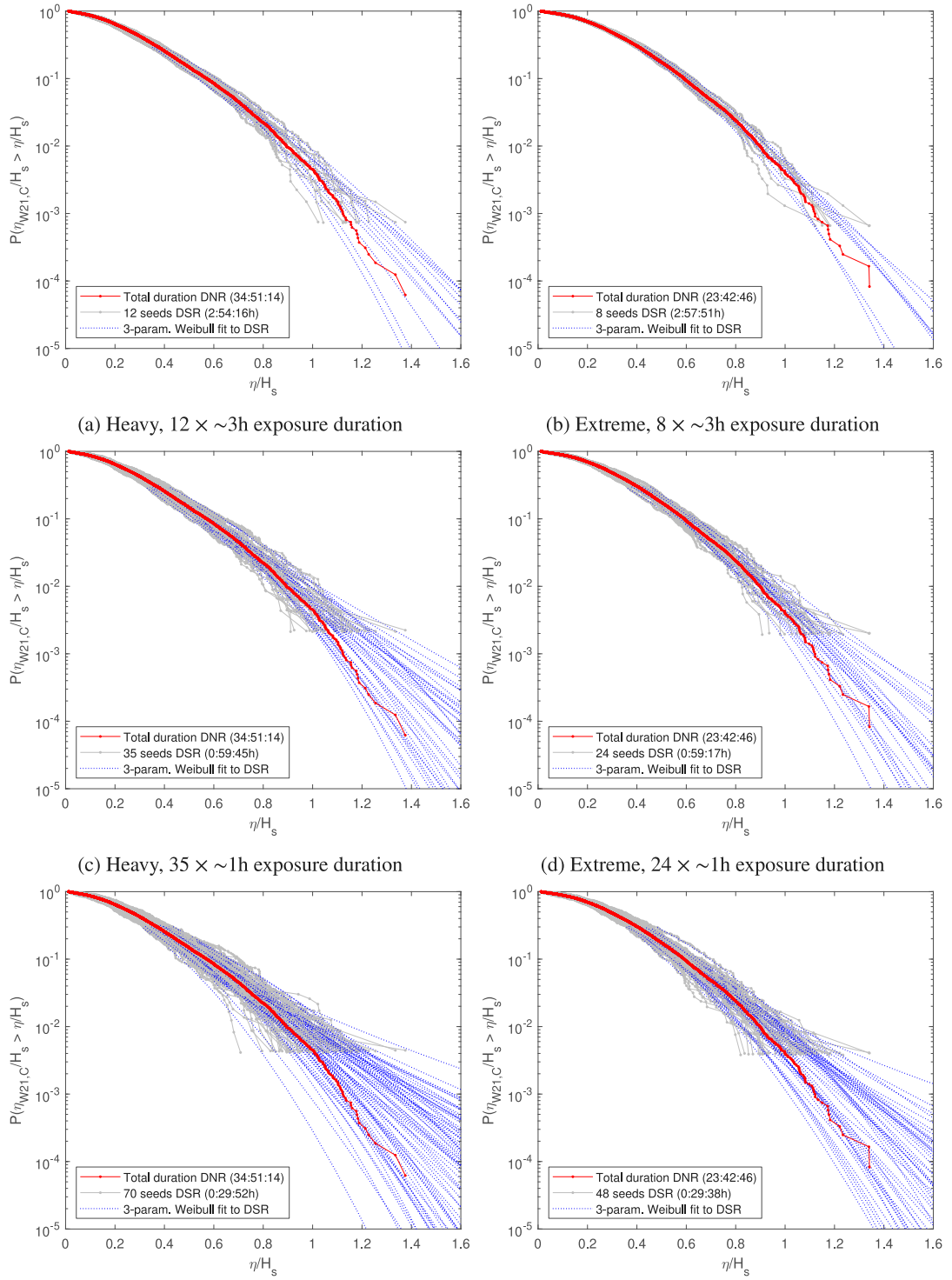
The measured DNR in Fig. 3 is above the second-order Forristall crest distribution, as expected for long-crested waves (see e.g., [40,45,46]). Measurements in a basin at Imperial College discussed in the first publication indicated that crest heights are expected to exceed Forristall for probabilities below around 10^{-3}. The present measurements show that this already happens at higher probabilities, around 10^{-1}. This may be because the wave conditions are relatively steep. The measured DNR is quite close to the Huang-Zhang DNR distribution for the probability range between $10^{-2} - 5 \cdot 10^{-2}$. For higher probabilities than $5 \cdot 10^{-2}$ (lower crests), the measurements are slightly lower than the Huang-Zhang distributions. Here Forristall provides a better match. For lower probabilities than 10^{-2} (higher crests), the measured crests are significantly higher than Huang-Zhang. The tested wave conditions were quite steep ($H_s/T_p^2 \sim 0.08$–0.09), and breaking wave events were observed. This probably causes the energy dissipation and drop of the higher crests below the distribution expected based on the lower crests. However, the drop is not as significant as predicted by Huang-Zhang. Interesting to see is that the simple 8% correction to Forristall from the CresT JIP leads to results that are very close to the measurements for lower probabilities ($<10^{-2}$).

Similar things were observed for the measured DSRs in Fig. 3: most of them are above Forristall. Of course there is a lot of scatter in the peak values between the seeds, but their mean also seems to approximately follow Huang-Zhang. The width of the 99% DSR interval from Huang-Zhang for 3 h exposure seems to comply quite well with the measurements. Based on the present results, this new empirical estimate of the upper and lower limits between which the crest distribution of a single seed should lie can be helpful in new test campaigns.

Finally, it can be concluded that the proposed 99% intervals for the individual seed distributions of wave crests proposed by [38] match the present data relatively well; they can be used to roughly estimate the uncertainty of a single seed measured wave crest distribution at different probability levels.

Appendix C. Weibull fit plots

Fig. 13 provides all Weibull fit plots for the wave crests and Fig. 14 provides the same plots for the green water impact force peaks. The number of impact peaks in 30 min exposure duration was too low for a fit, so this duration was omitted.



(a) Heavy, 12 × ∼3h exposure duration

(b) Extreme, 8 × ∼3h exposure duration

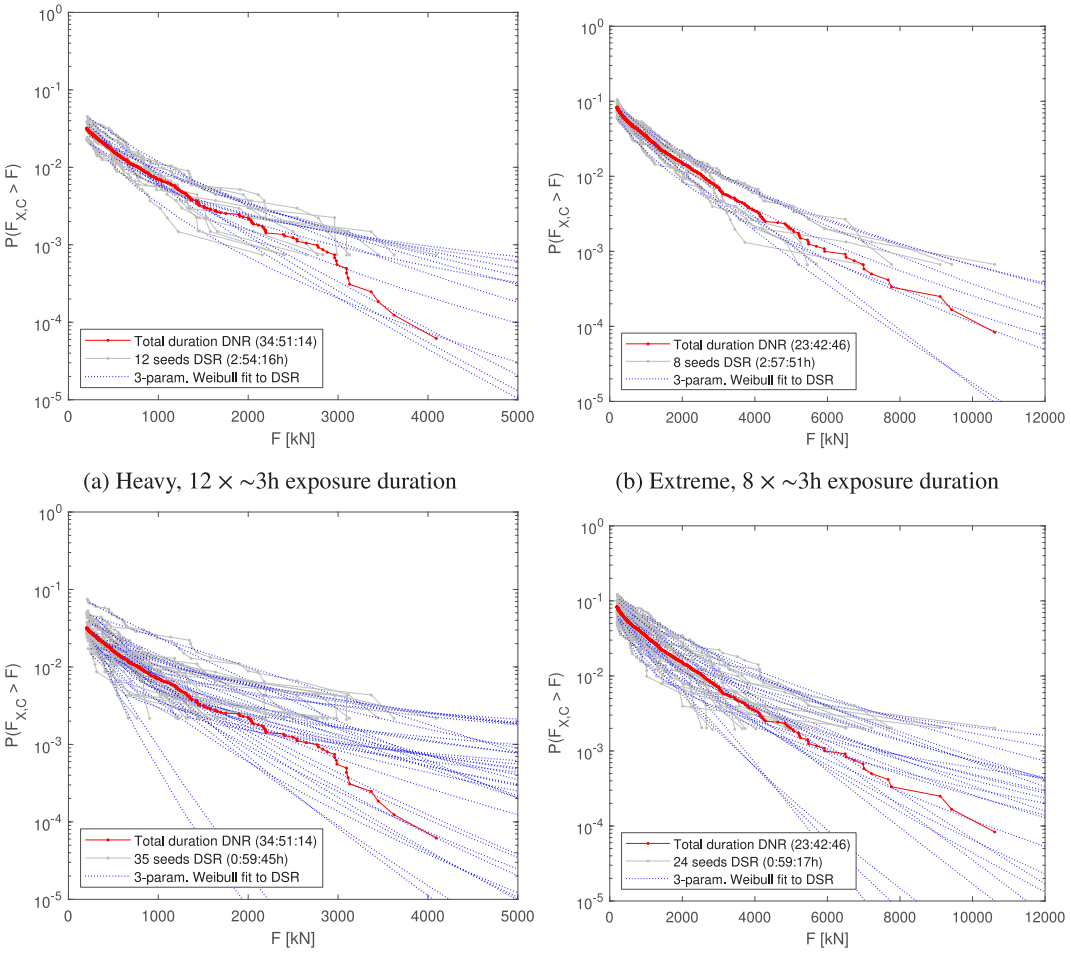
(c) Heavy, 35 × ∼1h exposure duration

(d) Extreme, 24 × ∼1h exposure duration

(e) Heavy, 70 × ∼30min exposure duration

(f) Extreme, 48 × ∼30min exposure duration

Fig. 13. DSR and DNR distributions of $\eta_{W21,C}$, in the heavy and extreme condition for different exposure durations and number of seeds. Including 3-parameter Weibull fits to top 30% crests in the DSR distributions.



(a) Heavy, 12 × ∼3h exposure duration

(b) Extreme, 8 × ∼3h exposure duration

(c) Heavy, 35 × ∼1h exposure duration

(d) Extreme, 24 × ∼1h exposure duration

Fig. 14. DSR and DNR distributions of $F_{X,C}$, in the heavy and extreme condition for different exposure durations and number of seeds. Including 3-parameter Weibull fits to top 90% peaks in the DSR distributions. The 30 min exposure duration is omitted, as the number of wave impact peaks per seed was too low for a good fit in many of the seeds.

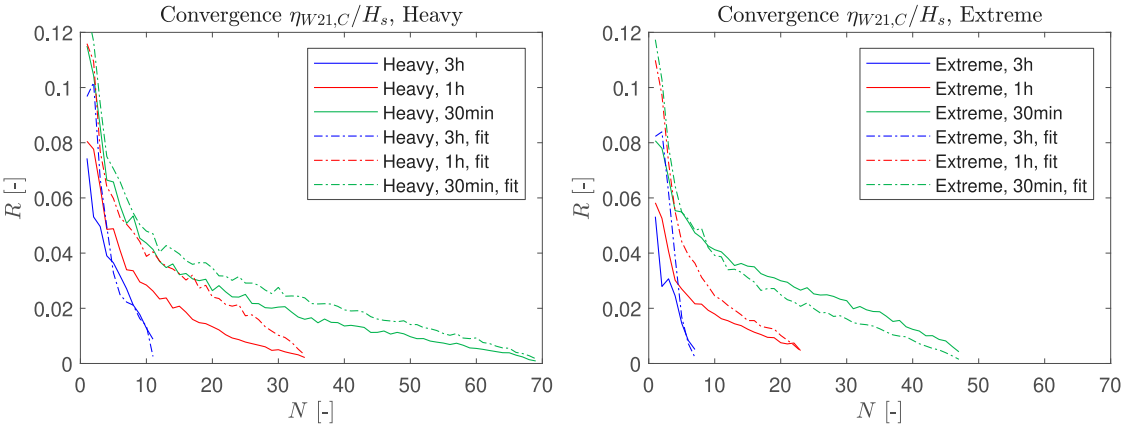


Fig. 15. Re-ordered convergence plots for the non-dimensional wave crests, per wave condition. The superscript f is omitted in the y-axis label for the fitted convergence metric.

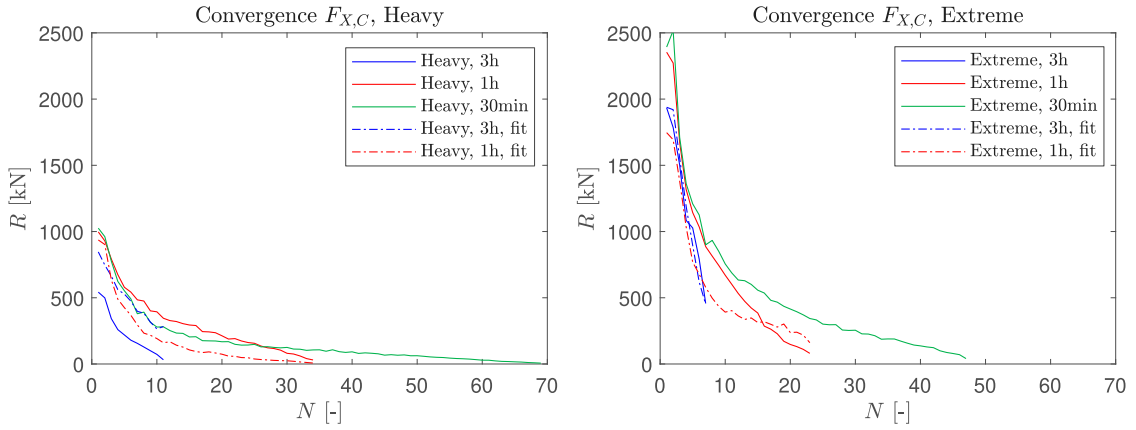


Fig. 16. Re-ordered convergence plots for the green water force peaks, per wave condition (30 min fits not available). The superscript f is omitted in the y-axis label for the fitted convergence metric.

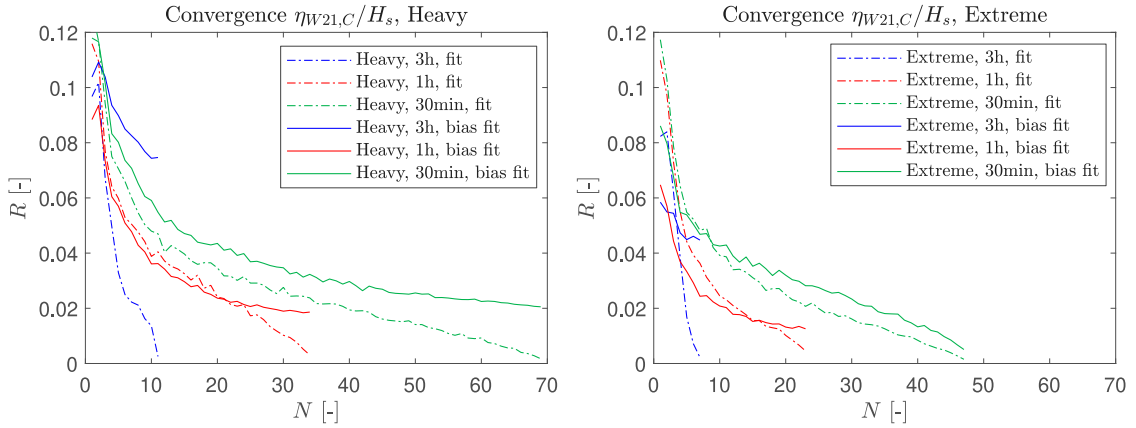


Fig. 17. Re-ordered convergence and bias plots for the fitted non-dimensional wave crests, per wave condition. The superscript f is omitted in the y-axis label for both the convergence and bias metric.

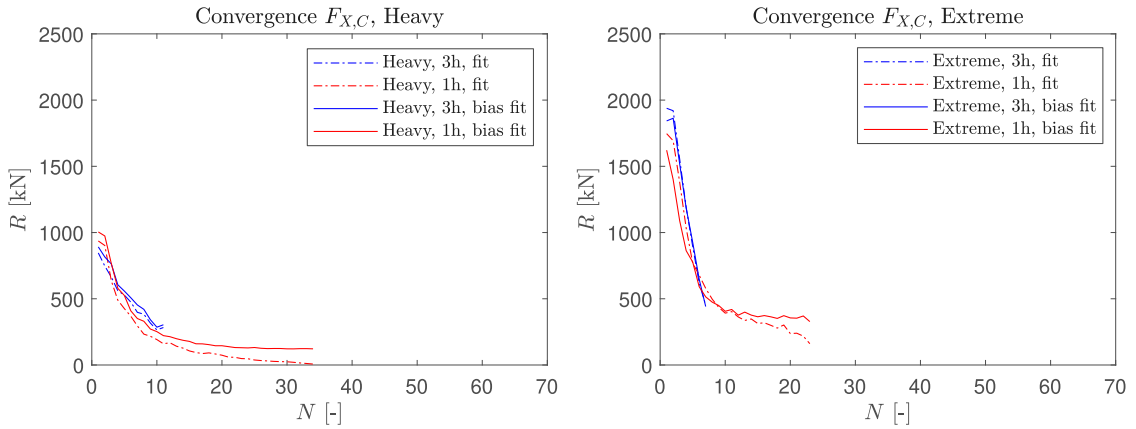


Fig. 18. Re-ordered convergence and bias plots for the fitted green water force peaks, per wave condition (30 min fits not available). The superscript f is omitted in the y-axis label for both the convergence and bias metric.

Appendix D. Re-ordered convergence and bias plots

In Section 5.4, the convergence plots are arranged to show both wave conditions in the same plot, separately for the original and fitted data. In Figs. 15 and 16, the same convergence data is shown, but now organised per wave condition. This enables direct comparison of the original and fitted data. For this bias this is not necessary, as it is only defined for the fitted data. Similarly, Figs. 17 and 18 show the results for the bias and convergence of the fits in the same plot, per wave condition.

References

[1] Scharnke J, van Essen SM, Seyffert HC. Required test durations for converged short-term wave and impact extreme value statistics - Part 2: deck box dataset. Mar Struct 2023;(103411). http://dx.doi.org/10.1016/j.marstruc.2023.103411.

[2] Bunnik THJ, Stansberg CT, Pákozdi C, Fouques S, Somers L. Useful indicators for screening of sea states for wave impacts on fixed and floating platforms. In: 37th Int. conf. ocean offshore arct. eng. (OMAE). Madrid, Spain: ASME; 2018, http://dx.doi.org/10.1115/OMAE2018-78544.

[3] Bunnik THJ, Scharnke J, de Ridder EJ. Efficient indicators for screening of random waves for wave impacts on a jacket platform and a fixed offshore wind turbine. In: 38th Int. conf. ocean offshore arct. eng. (OMAE). Glasgow, UK: ASME; 2019, http://dx.doi.org/10.1115/OMAE2019-95481.

[4] van Essen SM, Monroy C, Shen Z, Helder J, Kim D-H, Seng S, Ge Z. Screening wave conditions for the occurrence of green water events on sailing ships. Ocean Eng 2021;234(109218). http://dx.doi.org/10.1016/j.oceaneng.2021.109218.

[5] Tromans P, Anaturk A, Hagemeijer P. A new model for the kinematics of large ocean waves - application as a design wave. In: 1st ISOPE conf., vol. III. Tsukuba City, Japan: Int. Soc. of Offshore and Polar Eng.; 1991, p. 64–71.

[6] Friis-Hansen P, Nielsen L. On the new wave model for the kinematics of large ocean waves. In: 14th Int. conf. offshore mech. and arct. eng., Copenhagen, Denmark. 1995, p. 14–24.

[7] Dietz J. Application of conditional waves as critical wave episodes on marine structures (Ph.D. thesis), Technical University of Denmark; 2004.

[8] Kim D-H. Design loads generator: estimation of extreme environmental loadings for ship and offshore applications (Ph.D. thesis), Ann Arbor, USA: University of Michigan; 2012.

[9] Gramstad O, Agrell C, Bitner-Gregersen E, Guo B, Ruth E, Vanem E. Sequential sampling method using Gaussian process regression for estimating extreme structural response. Mar Struct 2020;72(102780). http://dx.doi.org/10.1016/j.marstruc.2020.102780.

[10] Sapsis TP. Statistics of extreme events in fluid flows and waves. Annu Rev Fluid Mech 2021;53:85–111. http://dx.doi.org/10.1146/annurev-fluid-030420-032810.

[11] Guth S, Sapsis T. Wave episode based Gaussian process regression for extreme event statistics in ship dynamics: Between the Scylla of Karhunen–Loeve convergence and the Charybdis of transient features. Ocean Eng 2022;266(112633). http://dx.doi.org/10.1016/j.oceaneng.2022.112633.

[12] Bitner-Gregersen EM, Dong S, Fu T, Ma N, Maisondieu C, Miyake R, Rychlik I. Sea state conditions for marine structures' analysis and model tests. Ocean Eng 2016;119:309–22. http://dx.doi.org/10.1016/j.oceaneng.2016.03.024.

[13] Begg SH, Welsh MB, Bratvold RB. Uncertainty vs. Variability: What's the difference and why is it important? In: SPE hydrocarbon economics and evaluation symp.. 2014, http://dx.doi.org/10.2118/169850-MS.

[14] Qiu W, Sales Junior J, Lee D, Lie H, Magarovskii V, Mikami T, Rousset J-M, Sphaier S, Tao L, Wang X. Uncertainties related to predictions of loads and responses for ocean and offshore structures. Ocean Eng 2014;86:58–67. http://dx.doi.org/10.1016/j.oceaneng.2014.02.031.

[15] Winterstein S, Ude T, Cornell C, Bjerager P, Haver S. Environmental parameters for extreme response: inverse FORM with omission factors. In: Proc. of int. conf. on structural safety and reliability (ICOSSAR93). 1993.

[16] Haver S, Winterstein SR. Environmental contour lines: a method for estimating long term extremes by a short term analysis. In: SNAME annu. meet.. Houston, USA: Society of Naval Architects and Marine Engineers; 2008.

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

[18] DNV. Det norske veritas, class guideline DNVGL-CG-0130: Wave loads. Oslo, Norway: Det Norske Veritas; 2018, URL; https://rules.dnv.com/docs/pdf/DNV/CG/2018-01/DNVGL-CG-0130.pdf.

[19] Baarholm GS, Haver S, Økland OD. Combining contours of significant wave height and peak period with platform response distributions for predicting design response. Mar Struct 2010;23:147–63. http://dx.doi.org/10.1016/j.marstruc.2010.03.001.

[20] Derbanne Q, de Hauteclocque G, Dumont M. How to account for short-term and long-term variability in the prediction of the 100 years response? In: 36th Int. conf. ocean offshore arct. eng. (OMAE). Trondheim, Norway: ASME; 2017, http://dx.doi.org/10.1115/OMAE2017-61701.

[21] Seyffert H, Kana A. Response-based reliability contours for complex marine systems considering short and long-term variability. Appl Ocean Res 2020;103:102332. http://dx.doi.org/10.1016/j.apor.2020.102332.

[22] Derbanne Q, de Hauteclocque G. A new approach for environmental contour and multivariate de-clustering. In: 38th Int. conf. ocean offshore arct. eng. (OMAE). Glasgow, UK: ASME; 2019, http://dx.doi.org/10.1115/OMAE2019-95993.

[23] ABS. Guidance notes on air gap and wave impact analysis for semisubmersibles. Spring, USA: American Bureau of Shipping; 2020, URL; https://ww2.eagle.org/content/dam/eagle/rules-and-guides/current/offshore/249-gn-airgapanalysis-semisubmersibles-2018/air-gap-analysis-gn-may20.pdf.

[24] DNV. Recommended practice DNV-RP-C205: Environmental conditions and environmental loads. Oslo, Norway: Det Norske Veritas; 2019, URL; https://www.dnv.com/oilgas/download/dnv-rp-c205-environmental-conditions-and-environmental-loads.html.

[25] BV. Rule note NR445: Rules for the classification of offshore units (Part D). Paris, France: Bureau Veritas; 2016, URL; https://erules.veristar.com/dy/data/bv/pdf/445-NR_PartD_2016-12.pdf.

[26] ITTC. Recommended practice 7.5-02-07-01.2: laboratory modelling of waves: Regular, irregular and extreme events. In: International towing tank conference. 2017, URL; https://www.ittc.info/media/8095/75-02-07-012.pdf.

[27] ITTC. Recommended practice 7.5-02-07-02.3: Experiments on rarely occurring events. In: International towing tank conference. 2017, URL; https://www.ittc.info/media/8105/75-02-07-023.pdf.

[28] Naess A, Gaidai O, Teigen P. Extreme response prediction for nonlinear floating offshore structures by Monte Carlo simulation. Appl Ocean Res 2007;29(4):221–30. http://dx.doi.org/10.1016/j.apor.2007.12.001.

[29] DNV. Offshore technical guidance, horizontal wave impact loads for column stabilised units, DNVGL-OTG-14. Oslo, Norway: Det Norske Veritas; 2019, URL; https://www.dnv.com/maritime/Offshore/technical-guidance-otg.html.

[30] NORSOK. Standard N-003 Actions and action effects. Standards Norway; 2007, URL; https://www.standard.no/en/sectors/energi-og-klima/petroleum/norsok-standard-categories/n-structural/n-0032/.

[31] Lian G. Slamming loads on large volume structures from breaking waves (Ph.D. thesis), Stavanger, Norway: University of Stavanger; 2018.

[32] BV. Rule note NR583: Whipping and springing assessment. Paris, France: Bureau Veritas; 2015, URL; https://erules.veristar.com/dy/data/bv/pdf/583-NR_2015-07.pdf.

[33] NORDFORSK. Assessment of ship performance in a seaway: results of a Nordic co-operative project on seakeeping performance of ships. Copenhagen, Denmark: Nordic Co-operative Organisation For Applied Research c/o Technoconsult; 1987.

[34] Ferrari V, Gornicz T, Kisjes A, Quadvlieg FHHA. Influence of skeg on ship manoeuvrability at high and low speeds. In: 14th PRADS conf. (2019). Lecture notes civil eng., 63, 2021, p. 384–403. http://dx.doi.org/10.1007/978-981-15-4624-2_23.
[35] Dallinga RP. The new seakeeping and manoeuvring basin of MARIN. In: Workshop on waves. Kyoto, Japan; 1999.
[36] van de Bunt E, Dekker J, Scharnke J, Jaouën F. Applying force panels for wave impact measurements. Ocean Eng 2021;232(108857). http://dx.doi.org/10.1016/j.oceaneng.2021.108857.
[37] Ochi M. Applied probability and stochastic processes in engineering and physical sciences. Wiley series in probability and mathematical sciences, Singapore: John Wiley & Sons; 1990.
[38] Huang Z, Zhang Y. Semi-empirical single realization and ensemble crest distributions of long-crest nonlinear waves. In: 37th Int. conf. ocean offshore arct. eng. (OMAE). Madrid, Spain: ASME; 2018, http://dx.doi.org/10.1115/OMAE2018-78192.
[39] Forristall GZ. Wave crest distributions: Observations and second-order theory. Phys Oceanogr 2000;30(8):1931–43. http://dx.doi.org/10.1175/1520-0485(2000)030<1931:WCDOAS>2.0.CO;2.
[40] Buchner B, Forristall GZ, Ewans K, Christou M, Hennig J. New insights in extreme crest height distributions (a summary of the 'CresT' JIP). In: 30th Int. conf. ocean offshore arct. eng. (OMAE). Rotterdam, The Netherlands: ASME; 2011, http://dx.doi.org/10.1115/OMAE2011-49846.
[41] Haver S. Freak waves: a suggested definition and possible consequences for marine structures. In: Work. rogue waves. Brest, France: Ifremer; 2004, URL; http://www.ifremer.fr/web-com/stw2004/rw/fullpapers/haver.pdf.
[42] ITTC. Recommended practice 7.5-02-07-02.6: Global loads seakeeping procedure. In: International towing tank conference. 2017, URL; https://www.ittc.info/media/8109/75-02-07-026.pdf.
[43] Haver S. Lecture notes: Metocean modelling and prediction of extremes. Stavanger, Norway: Haver & havet, University in Stavanger, NTNU; 2017, p. 1–256.
[44] Mauro F, Braidotti L, la Monaca U, Nabergoj R. Extreme loads determination on complex slender structures. Int Shipbuild Prog 2019;66:57–76. http://dx.doi.org/10.3233/ISP-180256.
[45] Latheef M, Swan C. A laboratory study of wave crest statistics and the role of directional spreading. Proc R Soc A Math Phys Eng Sci 2013;469(2152). http://dx.doi.org/10.1098/rspa.2012.0696.
[46] Hennig J, Scharnke J, Swan C, Hagen Ø, Ewans K, Tromans P, Forristall G. Effect of short-crestedness on extreme wave impact - summary of findings from the JIP 'ShorTCresT'. In: 34th Int. conf. ocean offshore arct. eng. (OMAE). St John's, Newfoundland, Canada; 2015, http://dx.doi.org/10.1115/OMAE2015-41167.
[47] Bogaert H. An experimental investigation of sloshing impact physics in membrane LNG tanks on floating structures (Ph.D. thesis), Delft, The Netherlands: Delft University of Technology; 2018, http://dx.doi.org/10.4233/uuid:96870b88-e07b-4ec2-8bd4-ef2cd3713568.
[48] Scharnke J. Elementary loading processes and scale effects involved in wave-in-deck type of loading - a summary of the BreaKin JIP. In: 38th Int. conf. ocean offshore arct. eng. (OMAE). Glasgow,UK; 2019, http://dx.doi.org/10.1115/OMAE2019-95004.
[49] Hennig J, Scharnke J, Schmittner C, van den Berg J. ShorTCresT: Directional wave measurements at MARIN. In: 34th Int. conf. ocean offshore arct. eng. (OMAE). St John's, Newfoundland, Canada; 2015, http://dx.doi.org/10.1115/OMAE2015-41169.
[50] van Essen SM. Influence of wave variability on ship response during deterministically repeated seakeeping tests at forward speed. In: 14th Conf. practical design of ships other floating struc (PRADS2019), Yokohama Japan. Lecture notes civil eng., 63, Singapore: Springer Nature; 2021, p. 899–925. http://dx.doi.org/10.1007/978-981-15-4624-2_54.
[51] Canard M, Ducrozet G, Bouscasse B. Varying ocean wave statistics emerging from a single energy spectrum in an experimental wave tank. Ocean Eng 2022;246(110375). http://dx.doi.org/10.1016/j.oceaneng.2021.110375.
[52] Buchner B. Green water on ship-type offshore structures (Ph.D. thesis), Delft, The Netherlands: Delft University of Technology; 2002.
[53] van Essen SM. Variability in encountered waves during deterministically repeated seakeeping tests at forward speed. In: 38th Int. conf. ocean offshore arct. eng. (OMAE). Glasgow, UK: ASME; 2019, http://dx.doi.org/10.1115/OMAE2019-95065.