scientific reports

OPEN

Novel methods for reliability study of multi-dimensional non-linear dynamic systems

Oleg Gaidai[1], Jingxiang Xu[1]✉, Ping Yan[1], Yihan Xing[2], Kelin Wang[1] & Zirui Liu[1]

This research presents two unique techniques for engineering system reliability analysis of multi-dimensional non-linear dynamic structures. First, the structural reliability technique works best for multi-dimensional structural responses that have been either numerically simulated or measured over a long enough length to produce an ergodic time series. Second, a novel extreme value prediction method that can be used in various engineering applications is proposed. In contrast to those currently used in engineering reliability methodologies, the novel method is easy to use, and even a limited amount of data can still be used to obtain robust system failure estimates. As demonstrated in this work, proposed methods also provide accurate confidence bands for system failure levels in the case of real-life measured structural response. Additionally, traditional reliability approaches that deal with time series do not have the benefit of being able to handle a system's high dimensionality and cross-correlation across several dimensions readily. Container ship that experiences significant deck panel pressures and high roll angles when travelling in bad weather was selected as the example for this study. The main concern for ship transportation is the potential loss of cargo owing to violent movements. Simulating such a situation is difficult since waves and ship motions are non-stationary and complicatedly non-linear. Extreme movements greatly enhance the role of nonlinearities, activating effects of second and higher order. Furthermore, laboratory testing may also be called into doubt due to the scale and the choice of the sea state. Therefore, data collected from actual ships during difficult weather journeys offer a unique perspective on the statistics of ship movements. This work aims to benchmark state-of-the-art methods, making it possible to extract necessary information about the extreme response from available on-board measured time histories. Both suggested methods can be used in combination, making them attractive and ready to use for engineers. Methods proposed in this paper open up possibilities to predict simply yet efficiently system failure probability for non-linear multi-dimensional dynamic structure.

Generally, it is quite challenging to calculate realistic structural system reliability by using conventional theoretical reliability methods[1–7]. The latter is often caused by many system freedom degrees and random variables that control dynamic systems. Theoretically, it is possible to estimate complicated structural systems' reliability straightforwardly by using sufficient data or direct Monte Carlo simulations[8–13]. The experimental or computational costs may be prohibitive for many complicated dynamic systems. Authors have developed a unique reliability approach for structural systems to lower measurement or computing costs as a result of the latter argument.

This work focuses on the general extreme value theory-based approach, in which it is anticipated that neither the physical dynamics of the water waves nor any other dynamic or environmental system would significantly contribute to the emergence of uncommon occurrences[14–20]. Although beyond the scope of our work, there have been some successful attempts to study extreme events in water waves (often called rogue or freak waves) with distributions that are uniquely determined by the dynamics of the physical system. For instance[20–24], have shown that water waves departing from linear theory will modify their distribution from a Rayleigh type[25–30], to a distribution dependent on the square root of the wave steepness. Similarly[31–35], have shown that a Rayleigh distribution modified by a polynomial function of the ratio between height and water depth controls extreme events in Hurricane data. In addition, spectrum bandwidth seems to have different effects in extreme wave distribution depending on whether they are in deep waters[36–38]. Furthermore, ocean processes such as shoaling or wave-current systems that drive wave trains out of equilibrium have been experimental[39–42] associated with increasing the occurrence of extreme waves by order of magnitude. However, it has been recently found that no

[1]Shanghai Ocean University, Shanghai, China. [2]University of Stavanger, Stavanger, Norway. ✉email: jxxu@shou.edu.cn

established theoretical distribution to date, neither universal as Gumbel nor based on physical principles, can describe extreme wave statistics in a wide range of conditions[43–49].

Figure 1 sketches the flow chart for the methodology suggested in this paper. For the definition of synthetic **R** vector and more details see next Section. Note that methods introduced by authors here do not rely on Gumbel (or any other type) distribution type assumption. Instead, Gumbel-based extrapolation was used just for comparison. By applying the new approach to a collection of data from a real-world on-board measurement experiment aboard a container ship, this part aims to demonstrate the previously stated methodology's effectiveness. Ship dynamics is a well-known example of a highly non-linear, multi-dimensional, and cross-correlated dynamic system that is difficult to analyse. Furthermore, system reliability research is crucial for container ships traversing the Atlantic Ocean in actual, occasionally severe weather. Typically, it is considered that ocean waves constitute an ergodic random process (stationary and homogenous) within 3 h storm.

Methods

Consider an offshore MDOF structure that is subjected to ergodic environmental loadings, such as those caused by the wind and waves in the area. The other option is to see a process dependent on external factors, whose fluctuation over time may be modelled as an independent ergodic process. Let $X_1, \ldots, X_{N_X}$ be consequent in time local maxima of the process $X(t)$ at monotonously increasing discrete time instants $t_1^X < \cdots < t_{N_X}^X$ in $(0, T)$. The analogous definition follows for other MDOF response components $Y(t), Z(t), \ldots$ with $Y_1, \ldots, Y_{N_Y}; Z_1, \ldots, Z_{N_Z}$ and so on. For simplicity, all $R(t)$ components, and therefore its maxima are assumed to be non-negative. The aim is to estimate the system failure probability

$$1 - P = \text{Prob}\left(X_T^{\max} > \eta_X \cup Y_T^{\max} > \eta_Y \cup Z_T^{\max} > \eta_Z \cup \ldots\right) \tag{1}$$

with

$$P = \int\limits_{(0,0,0,\ldots)}^{(\eta_X, \eta_Y, \eta_Z, \ldots)} \int\int p_{X_T^{\max}, Y_T^{\max}, Z_T^{\max}, \ldots}\left(X_T^{\max}, Y_T^{\max}, Z_T^{\max}, \ldots\right) \\ dX_T^{\max} dY_{N_Y}^{\max} dZ_{N_z}^{\max} \ldots \tag{2}$$

being the probability of non-exceedance for response components $\eta_X, \eta_Y, \eta_Z, \ldots$ critical values; $\cup$ denotes logical unity operation «or»; and $p_{X_T^{\max}, Y_T^{\max}, Z_T^{\max}, \ldots}$ being joint probability density of the global maxima over the entire time span $(0, T)$.

In practice, however, it is not feasible to estimate the latter joint probability distribution directly $p_{X_T^{\max}, Y_T^{\max}, Z_T^{\max}, \ldots}$ due to its high dimensionality and available data set limitations. In other words, the time instant when either $X(t)$ exceeds η_X, or $Y(t)$ exceeds η_Y, or $Z(t)$ exceeds η_Z, and so on, the system is regarded as immediately as failed. Fixed failure levels $\eta_X, \eta_Y, \eta_Z, \ldots$ are of course individual for each unidimensional response component of $R(t)$. $X_{N_X}^{\max} = \max\{X_j; j = 1, \ldots, N_X\} = X_T^{\max}, Y_{N_Y}^{\max} = \max\{Y_j; j = 1, \ldots, N_Y\} = Y_T^{\max}$, $Z_{N_z}^{\max} = \max\{Z_j; j = 1, \ldots, N_z\} = Z_T^{\max}$, and so on.

Next, the local maxima time instants $\left[t_1^X < \cdots < t_{N_X}^X; t_1^Y < \cdots < t_{N_Y}^Y; t_1^Z < \cdots < t_{N_z}^Z\right]$ in monotonously non-decreasing order are sorted into one single merged time vector $t_1 \leq \cdots \leq t_N$. Note that $t_N = \max\{t_{N_X}^X, t_{N_Y}^Y, t_{N_z}^Z, \ldots\}$, $N = N_X + N_Y + N_Z + \ldots$. In this case t_j represents local maxima of one of MDOF bio-system response components either $X(t)$ or $Y(t)$, or $Z(t)$ and so on. That means that having $R(t)$ time record, one needs continuously and simultaneously screen for unidimensional response component local maxima and record its exceedance of the MDOF limit vector $(\eta_X, \eta_Y, \eta_Z, \ldots)$ in any of its components $X, Y, Z, \ldots$. The local unidimensional response component maxima are merged into one temporal non-decreasing vector $\overrightarrow{R} = (R_1, R_2, \ldots, R_N)$ in accordance with the merged time vector $t_1 \leq \cdots \leq t_N$. That is to say, each local maxima R_j is the actual encountered local maxima corresponding to either $X(t)$ or $Y(t)$, or $Z(t)$ and so on. Finally, the unified limit vector $(\eta_1, \ldots, \eta_N)$ is introduced with each component η_j is either η_X, η_Y or η_Z and so on, depending on which of $X(t)$ or $Y(t)$, or $Z(t)$ etc., corresponding to the current local maxima with the running index j.

Next, a scaling parameter $0 < \lambda \leq 1$ is introduced to artificially simultaneously decrease limit values for all response components, namely the new MDOF limit vector $(\eta_X^\lambda, \eta_Y^\lambda, \eta_z^\lambda, \ldots)$ with $\eta_X^\lambda \equiv \lambda \cdot \eta_X, \equiv \lambda \cdot \eta_Y, \eta_z^\lambda \equiv \lambda \cdot \eta_Z$, … is introduced. The unified limit vector $(\eta_1^\lambda, \ldots, \eta_N^\lambda)$ is introduced with each component η_j^λ is either $\eta_X^\lambda, \eta_Y^\lambda$ or η_z^λ and so on. The latter automatically defines probability $P(\lambda)$ as a function of λ, note that $P \equiv P(1)$ from Eq. (1). Non-exceedance probability $P(\lambda)$ can be now estimated as follows:
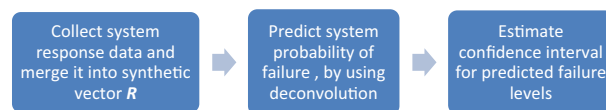

Figure 1. Flow chart for described methodology.

$$P(\lambda) = \text{Prob}\left\{R_N \leq \eta_N^\lambda, \ldots, R_1 \leq \eta_1^\lambda\right\}$$

$$= \text{Prob}\left\{R_N \leq \eta_N^\lambda \mid R_{N-1} \leq \eta_{N-1}^\lambda, \ldots, R_1 \leq \eta_1^\lambda\right\} \cdot \text{Prob}\left\{R_{N-1} \leq \eta_{N-1}^\lambda, \ldots, R_1 \leq \eta_1^\lambda\right\} \quad (3)$$

$$= \prod_{j=2}^{N} \text{Prob}\left\{R_j \leq \eta_j^\lambda \mid R_{j-1} \leq \eta_{1j-}^\lambda, \ldots, R_1 \leq \eta_1^\lambda\right\} \cdot \text{Prob}\left(R_1 \leq \eta_1^\lambda\right)$$

In practice, the dependency between neighbouring R_j is not always negligible; thus, the following one-step (called here conditioning level $k = 1$) memory approximation is now introduced

$$\text{Prob}\left\{R_j \leq \eta_j^\lambda \mid R_{j-1} \leq \eta_{j-1}^\lambda, \ldots, R_1 \leq \eta_1^\lambda\right\} \approx \text{Prob}\left\{R_j \leq \eta_j^\lambda \mid R_{j-1} \leq \eta_{j-1}^\lambda\right\} \quad (4)$$

for $2 \leq j \leq N$ (called here conditioning level $k = 2$). The approximation introduced by Eq. (4) can be further expressed as

$$\text{Prob}\left\{R_j \leq \eta_j^\lambda \mid R_{j-1} \leq \eta_{j-1}^\lambda, \ldots, R_1 \leq \eta_1^\lambda\right\} \approx \text{Prob}\left\{R_j \leq \eta_j^\lambda \mid R_{j-1} \leq \eta_{j-1}^\lambda, R_{j-2} \leq \eta_{j-2}^\lambda\right\} \quad (5)$$

where $3 \leq j \leq N$ (will be called conditioning level $k = 3$), and so on. The motivation is to monitor each independent failure that happened locally first in time, thus avoiding cascading local inter-correlated exceedances.

Equation (5) presents subsequent refinements of the statistical independence assumption. The latter type of approximation captures the statistical dependence effect between neighbouring maxima with increased accuracy. Since the original MDOF bio-process $\boldsymbol{R}(t)$ was assumed ergodic and therefore stationary, the probability $p_k(\lambda) := \text{Prob}\{R_j > \eta_j^\lambda \mid R_{j-1} \leq \eta_{j-1}^\lambda, R_{j-k+1} \leq \eta_{j-k+1}^\lambda\}$ for $j \geq k$ will be independent of j but only dependent on conditioning level k. Thus non-exceedance probability can be approximated as in the Naess–Gaidai method[43–49], where

$$P_k(\lambda) \approx \exp\left(-N \cdot p_k(\lambda)\right), \quad k \geq 1 \quad (6)$$

Note that Eq. (6) follows from Eq. (1) by neglecting $\text{Prob}(R_1 \leq \eta_1^\lambda) \approx 1$, as the design failure probability is usually very small. Further, it is assumed $N \gg k$.

Note that Eq. (5) is similar to the well-known mean up-crossing rate equation for the probability of exceedance[50,51]. There is obvious convergence with respect to the conditioning parameter k

$$P = \lim_{k \to \infty} P_k(1); \quad p(\lambda) = \lim_{k \to \infty} p_k(\lambda) \quad (7)$$

Note that Eq. (6) for $k = 1$ turns into the quite well-known non-exceedance probability relationship with the mean up-crossing rate function

$$P(\lambda) \approx \exp\left(-\nu^+(\lambda)\, T\right); \quad \nu^+(\lambda) = \int_0^\infty \zeta p_{R\dot{R}}(\lambda, \zeta) d\zeta \quad (8)$$

where $\nu^+(\lambda)$ is the mean up-crossing rate of the response level λ for the above assembled non-dimensional vector $R(t)$ assembled from scaled MDOF bio-system response $\left(\frac{X}{\eta_X}, \frac{Y}{\eta_Y}, \frac{Z}{\eta_Z}, \ldots\right)$. In the above, the stationarity assumption has been used. The proposed methodology can also treat the non-stationary case. An illustration of how the methodology can be used to treat non-stationary cases is provided. Consider a scattered diagram of $m = 1, .., M$ environmental states, each short-term bio-environmental state having a probability q_m, so that $\sum_{m=1}^{M} q_m = 1$. The corresponding long-term equation is then

$$p_k(\lambda) \equiv \sum_{m=1}^{M} p_k(\lambda, m) q_m \quad (9)$$

with $p_k(\lambda, m)$ being the same function as in Eq. (7) but corresponding to a specific short-term environmental state with the number m. Next, by plotting $\ln\left\{\ln\left(p_k(\lambda)\right) - d_k\right\}$ versus $\ln(a_k\lambda + b_k)$, often nearly perfectly linear tail behaviour is observed.

It is useful to do the optimisation on the logarithmic level by minimising the following error function F with respect to the four parameters a_k, b_k, c_k, p_k, q_k

$$F\left(a_k, b_k, c_k, p_k, q_k\right) = \int_{\lambda_0}^{\lambda_1} \omega(\lambda)\left\{\ln\left(p_k(\lambda)\right) - d_k + (a_k\lambda + b_k)^{c_k}\right\}^2 d\lambda, \quad \lambda \geq \lambda_0 \quad (10)$$

with λ_1 being a suitable distribution tail cut-off value, namely the largest wave height value, where the confidence interval width is still acceptable. Optimal values of the parameters a_k, b_k, c_k, p_k, q_k may also be determined using a sequential quadratic programming (SQP) method incorporated in the NAG Numerical Library[52]. Weight function ω can be defined as $\omega(\lambda) = \left\{\ln\text{CI}^+(\lambda) - \ln\text{CI}^-(\lambda)\right\}^{-2}$ with $\left(\text{CI}^-(\lambda), \text{CI}^+(\lambda)\right)$ being a confidence interval (CI), empirically estimated from the simulated or measured dataset[43–49]. When the parameter $c = \lim_{k \to \infty} c_k$ is equal to 1 or close to it, the distribution is close to the Gumbel distribution.

For any general ergodic wave height or wind speed process, the sequence of conditional exceedances over a threshold λ can be assumed to constitute a Poisson process. However, in general, non-homogeneous one. Thus, for levels of λ approaching 1, the approximate limits of a p-% confidence interval (CI) of $p_k(\lambda)$ can be given as follows

$$\text{CI}^{\pm}(\lambda) = p_k(\lambda)\left(1 \pm \frac{f(p)}{\sqrt{(N-k+1)p_k(\lambda)}}\right). \tag{11}$$

with $f(p)$ being estimated from the inverse normal distribution, for example, $f(90\%) = 1.65$, $f(95\%) = 1.96$. with N being the total number of local maxima assembled in the analysed vector $\overrightarrow{R}$. Next, a novel extrapolation method is briefly introduced. Accurate extreme value prediction is a common and challenging engineering reliability task, especially when available data is scarce. Therefore, developing novel, efficient and accurate extrapolation techniques are of great practical importance. Let one consider a stationary stochastic process $X(t)$, either simulated or measured over a specific time span $0 \le t \le T$, and which is represented as a sum of two independent stationary processes $X_1(t)$ and $X_2(t)$, namely

$$X(t) = X_1(t) + X_2(t) \tag{12}$$

Note that this paper aims at a general methodology applicable to extreme value predictions for a wide range of loads and responses for various vessels and offshore structures. For the process of interest $X(t)$ one may obtain marginal PDF (probability density function) p_X by two distinctive ways:

(A) By directly extracting p_X^A from the available data set, i.e. time series $X(t)$,
(B) By separately extracting PDFs from the process components $X_1(t)$ and $X_2(t)$, namely p_{X_1} and p_{X_2}, then applying convolution $p_X^B = \text{conv}(p_{X_1}, p_{X_2})$.

Both p_X^A and p_X^B are being approximations of the target PDF p_X. Approach (A) is more straightforward to use, however (B) would provide a more accurate estimate of the target PDF p_X. An advantage of using convolution in case (B) is based on the fact that convolution enables extrapolation of the directly extracted empirical PDF p_X^A, without pre-assuming any specific extrapolation functional class, e.g. generalised extreme value distributions (GEV) needed to extrapolate distribution tail towards design low probability level of interest. Note that most existing extrapolation methods, widely adopted in engineering practice, rely on assuming certain extrapolation functional classes, e.g.[43-49]. To name some of those most popular existing methods: Pareto based distribution peak over the threshold (POT)[50].

The two independent component representation given by Eq. (12) is seldom available; therefore, one may look for artificial ways to estimate p_{X_1} and p_{X_2}, or in the simplest case, find two identically distributed process components $X_1(t)$ and $X_2(t)$ with $p_{X_1} = p_{X_2}$. This paper is focused on the latter alternative, i.e. case when processes $X_1(t)$ and $X_2(t)$ are equally distributed. Therefore the current study goal would be, given directly estimated distribution p_X as in option (A), to find component distribution p_{X_1} such that

$$p_X = \text{conv}(p_{X_1}, p_{X_1}) \tag{13}$$

thus restricting this study only to a deconvolution case. In order to exemplify the latter idea regarding how to estimate the unknown distribution robustly p_{X_1}, and subsequently to improve (say extrapolate) the given empirical distribution p_X.

Accurate extreme value prediction becomes extremely challenging whenever there is a scarcity of data in engineering. Hence, a novel, efficient and accurate extrapolation method must be developed. Such a method better facilitates better design and reliability development.

In a case where the stationary stochastic process $X(t)$ which is either measured or simulated in a time span of $0 \le t \le T$, and is then denoted as a sum of two independent stationary processes $X_1(t)$ and $X_2(t)$.

$$X(t) = X_1(t) + X_2(t) \tag{14}$$

Noteworthy with a general methodology, this paper hopes to better predict the extreme responses in engineering, with its primary focus on mechanical engineering, e.g. offshore jacket platform dynamics. Note that the method described here assumes a stationary dynamic system. In the case of non-stationary processes (long-term analysis), there will be a need to have stationary sub-parts in it (e.g. short-term analysis with 3-h sea states, taken from the scattered diagram).

The marginal PDF (probability density function) p_X can be acquired in two different ways for the process of interest $X(t)$:

(A) Using the available data set, i.e., time series $X(t)$, to directly extract p_X^A.
(B) Using the process components $X_1(t)$ and $X_2(t)$ to individually extract their PDFs from, namely p_{X_1} and p_{X_2}, and then utilising convolution $p_X^B = \text{conv}(p_{X_1}, p_{X_2})$.

In both processes, p_X^A and p_X^B are the approximations of the target PDF p_X. The initial approach (A), is more straightforward; however, the second approach (B), gives a better and more accurate estimate of the target PDF p_X.

The convolution method in approach (B) is also advantageous since it facilitates the direct extraction of the empirical PDF p_X^A, without a presumption of any extrapolation functional class.

However, the two independent components from Eq. (10) are often unknown and thus p_{X_1} and p_{X_2} must be estimated. If not, in a more straightforward scenario, both the identically distributed process components $X_1(t)$ and $X_2(t)$ with $p_{X_1} = p_{X_2}$ must be estimated. The latter-mentioned method will be examined in this paper, considering that both the process of $X_1(t)$ and $X_2(t)$ are equally distributed. Thus, using the directly estimated distribution p_X like approach A), the component distribution p_{X_1} is derived

$$p_X = \text{conv}(p_{X_1}, p_{X_1}) \tag{15}$$

This limits this case to only a deconvolution scenario. The convolution of two vectors is characterised by the overlapping area of both the vectors, $\boldsymbol{u}$ and $\boldsymbol{v}$. Thus, convolution is algebraically similar to the multiplication of polynomials whose coefficients are the elements of $\boldsymbol{u}$ and $\boldsymbol{v}$. Let $m = \text{length}(\boldsymbol{u})$ and $n = \text{length}(\boldsymbol{v})$. Then $\boldsymbol{w}$ is the vector of length $m + n - 1$, whose k th element is

$$w(k) = \sum_{j=1}^{m} u(j)v(k - j + 1) \tag{16}$$

The sum is over all the values of j that lead to legal subscripts for $u(j)$ and $v(k - j + 1)$, specifically $j = \max(1, k + 1 - n) : 1 : \min(k, m)$. When $m = n$, as will be the main case in this paper, the latter yields

$$
\begin{aligned}
w(1) &= u(1) \cdot v(1) \\
w(2) &= u(1) \cdot v(2) + u(2) \cdot v(1) \\
w(3) &= u(1) \cdot v(3) + u(2) \cdot v(2) + u(3) \cdot v(1) \\
&\cdots \\
w(n) &= u(1) \cdot v(n) + u(2) \cdot v(n - 1) + \cdots + u(n) \cdot v(1) \\
&\cdots \\
w(2n - 1) &= u(n) \cdot v(n)
\end{aligned}
\tag{17}
$$

From Eq. (13), $\boldsymbol{u} = \boldsymbol{v} = (u(1), ..., u(n))$, reduced parts of $\boldsymbol{w}$-components $w(n + 1), ..., w(2n - 1)$ are obtained, when the index increases from $n + 1$ to $2n - 1$. The latter extends vector $\boldsymbol{w}$ into the support domain, double the initial distribution support domain. In short, the distribution support length is doubled, $(2n - 1) \cdot \Delta x \approx 2n \cdot \Delta x = 2X_L$. When comparing with the initial distribution support length $n \cdot \Delta x = X_L$. Δx, in this case, is the constant length of each discrete distribution bin. In short, the convolution convects the distribution tail properties further down the tail.

The representation of the empirical target distribution p_X is $\boldsymbol{w} = (w(1), ..., w(n))$ where n is the length of distribution support $[0, X_L]$. In this paper, only the one-sided positive random variables, $X \geq 0$, are considered to minimise complexities. Moreover, only the deconvolution cases considered as Eq. (13) will be $\boldsymbol{u} = \boldsymbol{v}$. Judging from Eq. (13), the vectors $\boldsymbol{w}$ and $\boldsymbol{u}$, has a corresponding distribution p_X and p_{X_1}, respectively. The unknown components $\boldsymbol{u} = \boldsymbol{v} = (u(1), ..., u(n))$, can be found from the given $\boldsymbol{w} = (w(1), ..., w(n))$ in Eq. (13). It starts from the first component $u(1) = \sqrt{w(1)}$, then to the second component $u(2) = \frac{w(2)}{2u(1)}$, and until n component $u(n)$.

Through this method, a simple linear extrapolation of self-deconvoluted vector $(u(1), ..., u(n))$ towards $(u(n + 1), ..., u(2n - 1))$ is achieved. In short, p_{X_1} has its tail extrapolated linearly within the following range $(X_L, 2X_L)$. Thus now, the p_{X_1} is known as a deconvoluted distribution and, in its discrete form, is characterised by a projected vector $\boldsymbol{u}$. Based on Eq. (13), the vector $\boldsymbol{w}$ is extended and extrapolated doubling the length of the initial distribution support domain. In short, the p_X distribution support length is doubled, $(2n - 1) \cdot \Delta x \approx 2n \cdot \Delta x = 2X_L$, in comparison with the initial distribution support length $n \cdot \Delta x = X_L$.

A smoothing tail procedure helps smooth the tails as the obtained measurements or Monte Carlo simulations are not smooth enough. Using the p_X tail interpolation, the original distribution $p_X(x)$ tail has been introduced since a CDF distribution regularly has high tail values x. Furthermore, the Naess–Gaidai (NG) method was implemented since the tails become similar to $\exp\{-(ax + b)^c + d\}$ with a, b, c, d at $x \geq x_0$ Where a, b, c, d are suitable constants. Similarly, the tail's linear extrapolation of p_{X_1} is the preferred unbiased option. Other non-linear extrapolation approaches can be similarly used in the proposed method, but they typically introduce different assumptions and biases.

Distribution $p_X(x) \equiv p_X$ tail interpolation was performed, as CDF distribution tail is generally quite regular for high tail values x. At the same time, other non-linear extrapolation approaches can easily plug into the proposed method, but then certain assumptions and biases would be introduced. The NG extrapolation method has been used; see "Introduction" section.

In the following, numerical results are presented based on the proposed deconvolution extrapolation method outlined in the previous Section. As discussed in the previous Section, the deconvolution extrapolation technique does not pre-assume any specific extrapolation functional class needed to extrapolate the distribution tail.

Since in most reliability analysis engineering applications, it is more important to estimate the probability of exceedance, i.e. 1-CDF where CDF stands for cumulative density function, rather than the marginal PDF, subsequently in this paper notation f_X will stand for the probability of exceedance 1-CDF, analogous to the marginal probability density function PDF p_X in the previous section. However, the proposed methodology may be suitable for any sufficiently regular monotonously decreasing either concave or convex function tail.

To validate the above-suggested extrapolation methodology, the «shorter» version of the original data set has been used for extrapolation for the sake of comparison with predictions based on the entire «longer» data set. Therefore, this work aims to demonstrate that the recommended extrapolation approach is at least a few orders of magnitude efficient.

The description above shows that an iterative technique may be used, whereas a marginal PDF can be created using 1-CDF and then generate a new artificial smoother CDF using integration. The latter can significantly facilitate extrapolation if there are distribution tail irregularities due to the scarcity of the underlying data set.

Next, the procedure of discrete convolution, or rather de-convolution (as the purpose was to find a deconvoluted 1-CDF distribution f_{X_1}, given the empirical distribution f_X outlined in the previous section, is based on sequential solving of Eq. (4). Since the resulting deconvoluted values $\boldsymbol{u} = (u(1), ..., u(n))$ are typically following a monotonously decreasing pattern (the same was assumed for the empirical parent distribution f_X), it appears that some last values of resulting vector $\boldsymbol{u}$, say $(u(n - L), ..., u(n))$ for some $L < n$ may become negative. Because positive numbers may only represent distributions, the latter is a numerical mistake and cannot be accepted. The following scaling method has been proposed to address that numerical difficulty. The lowest positive value f_L of the given distribution tail of f_X is taken as a pivot value. The scaling then is simply a linear transformation along the vertical y-axis of the distribution on the decimal logarithmic scale

$$g_X = \mu \left(\log_{10}(f_X) - \log_{10}(f_L) \right) + \log_{10}(f_L) \tag{18}$$

with $g_X(x)$ being scaled $\log_{10}$ version of the empirical base distribution f_X, with the reference level f_L being intact. The scaling coefficient μ is conveniently chosen to avoid the occurrence of negative components in the resulting f_{X_1}. For both numerical examples studied in this paper, $\mu = 1/3$ served that purpose well. Then when f_{X_1} was found, and back convolution $\widetilde{f_X} = \text{conv}(f_{X_1}, f_{X_1})$ as in Eq. (13) was done, the inverse scaling with μ^{-1} was performed to restore the original scale, with $\widetilde{f_X}$ being extrapolated version of f_X.

Results

A Post-Panamax cargo ship named MSC Napoli capsized in January 2007. MSC Napoli broke in two places: amidships, against a pillar bulkhead and in the engine room. Another Post-Panamax cargo ship, MOL Comfort, broke in June 2013[53,54]. Even though these two ships may not have been constructed and authorised in accordance with best practices, giving them less collapse strength than other ships of comparable sizes, both ships capsized owing to the overloading of the hull girders. Such catastrophic events require extensive investigations as they shock the business, particularly the container ship sector, in these two incidents[39–42].

This section shows how the technique mentioned above is used in practice. Motion sensors were installed on the 2800TEU Panamax container ship during its transatlantic trips. The mid and aft ship panel stresses and the ship roll angle were selected as components X, Y, and Z to form an example of a three-dimensional (3D) dynamic system.

Unidimensional extreme response values were chosen as crucial thresholds, which led to vascular failure. These values generally equate to a 25-year return time. Namely $\eta_X = 140\,\text{MPa}, \eta_Y = 110\,\text{MPa}, \eta_Z = 28$ degrees; see[45–49].

Compared to contemporary Post-Panamax container ships, this specific Panamax container ship from the late 1990s has a small bow flare angle. Hence the amount of whipping on this ship will be quite minor. Routing, a human aspect, will control the excessive response. Figure 2 presents TEU2800 mid-ship on-board strain sensors placement along with observed crack positions. Similarly, sensors were placed aft of the vessel, resulting in measured stresses in the longitudinal direction on a flat bar below the upper deck. Sensor placement was done according to DNV container vessel rules and regulations[55–59]. In order to unify all three measured time series X, Y, Z, the following scaling was performed

$$X \rightarrow \frac{X}{\eta_X}, \quad Y \rightarrow \frac{Y}{\eta_Y}, \quad Z \rightarrow \frac{Z}{\eta_Z} \tag{19}$$

making all three responses non-dimensional and having the same failure limit equal to 1. Next, all local maxima from three measured time series were merged into one single time series by keeping them in time non-decreasing order: $\overrightarrow{R} = (\max\{X_1, Y_1, Z_1\}, ..., \max\{X_N, Y_N, Z_N\})$ with each set $\max\{X_j, Y_j, Z_j\}$ being sorted according to temporally non-decreasing occurrence times of these local maxima. Figure 4 left presents an example of a non-dimensional assembled vector $\overrightarrow{R}$, consisting of assembled local maxima of TEU2800 mid and aft stresses along with vessel roll angle; $\lambda > 0.2$ cut-on limit was used for illustrative purposes, as lower values $\lambda \geq 0$ are obviously are not relevant for the failure probability distribution tail extrapolation towards the target $\lambda = 1$. Note that vector $\overrightarrow{R}$ does not have a physical meaning on its own, as it was assembled of different response components with different units of measurement (MPa and angular degrees in this case). Index j is just a running index of local maxima encountered in a non-decreasing time sequence. The «shorter» data record has been generated by taking each tenth data point from the «longer» deck panel stress data record. Therefore, the «shorter» data record had an equivalent time length of only one year.

Figure 3 on the left presents the «shorter» data record f_{X_1} tail, obtained by deconvolution as in Eq. (13), and subsequently linearly extrapolated in the terminal tail section to cover the X_1 range matching the «longer» data record. Figure 3 on the right presents the final unscaled results of the proposed in this paper technique, namely the «shorter» decimal log scale f_X tail, extrapolated by deconvolution, along with «longer» data distribution tail and NG extrapolation.
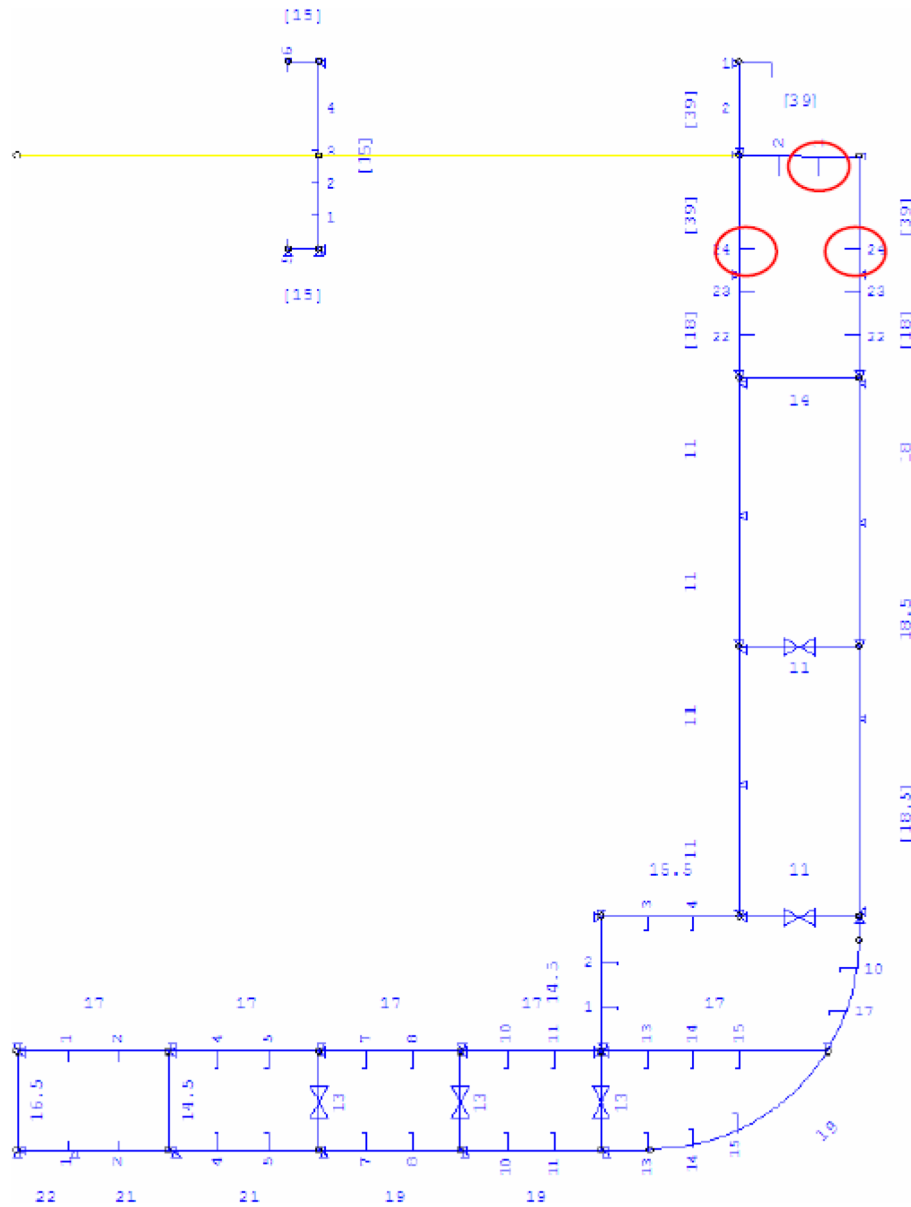
Figure 2. Layout of the mid-ship cross-section with measurement position in the upper deck and crack positions.

It is seen from Fig. 3 on the right that the proposed method performs quite well, being based on the «shorter» data set and delivering distribution quite close to the one based on the «longer» data set.

Figure 4 right presents extrapolation according to Eq. (9) towards failure state with 25 year return period, which is 1, and somewhat beyond, $\lambda = 0.4$ cut-on value was used. Dotted lines indicate extrapolated 95% confidence interval according to Eq. (10). According to Eq. (5) $p(\lambda)$ is directly related to the target failure probability $1 - P$ from Eq. (1). Therefore, in agreement with Eq. (5) system failure probability $1 - P \approx 1 - P_k(1)$ can be estimated. Note that in Eq. (5) N corresponds to the total number of local maxima in the unified response vector $\overrightarrow{R}$. Conditioning parameter $k = 6$ was found to be sufficient, due to convergence occurrence with respect to k, see Eq. (6). Figure 4 exhibits quite narrow 95% CI; the latter is due to a substantial amount of data used in this study namely over 70 trans-Atlantic voyages of the same vessel.

While being novel, the methodology described above has a clear advantage of utilising available measured data set quite efficiently, due to its ability to treat system multi-dimensionality and perform accurate extrapolation based on a relatively limited data set.
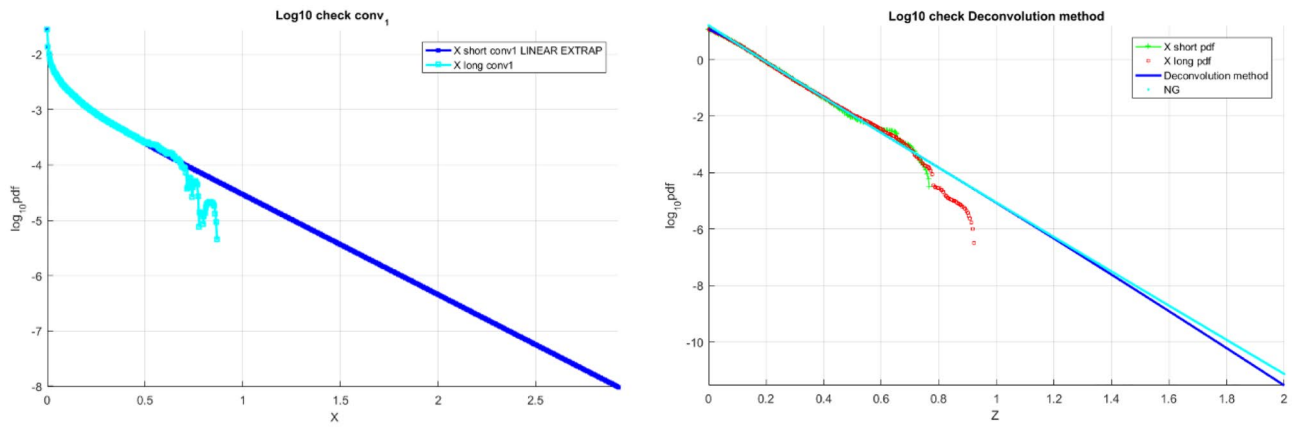
Figure 3. TEU vessel data. Left: scaled f_{X_1} tail on the decimal log scale for the «shorter» data (cyan), linearly extrapolated (dark blue). Right: unscaled raw «shorter» data (green) f_X tail on the decimal log scale, extrapolated by the deconvolution method (dark blue), along with «longer» raw data (red) and NG extrapolation (cyan).
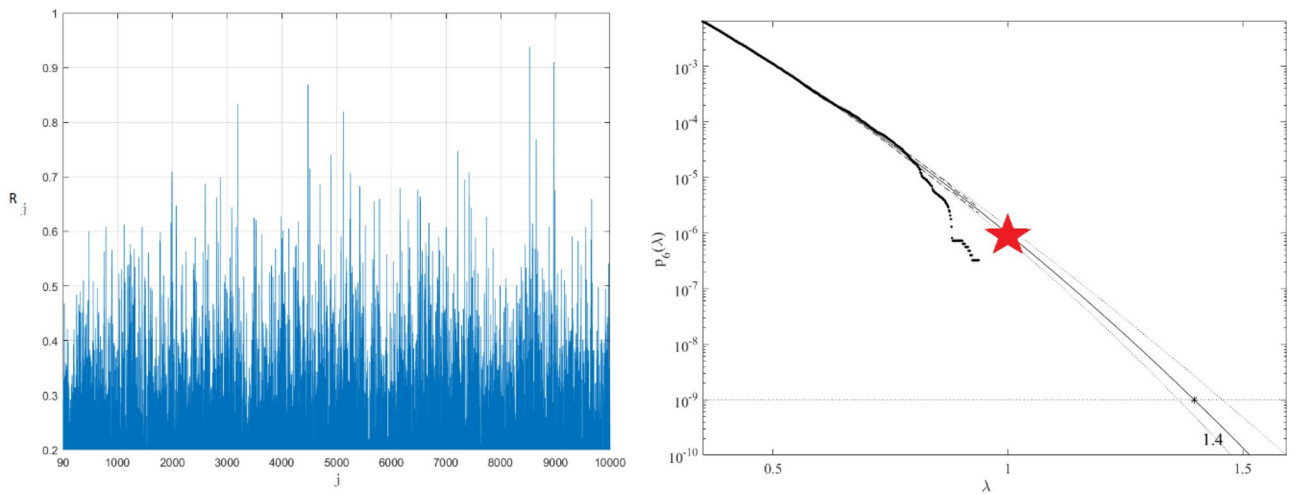

Figure 4. Left: Example of non-dimensional assembled 3D vector $\overrightarrow{R}$. Right: Extrapolation of $p_k(\lambda)$ towards critical level (indicated by a star) and beyond, $k = 6$. Extrapolated 95% CI indicated by dotted lines.

Conclusions

Traditional reliability techniques with time series do not have the benefit of effectively dealing with highly dimensional systems and cross-correlation between various system responses. The methodology's primary asset is the capacity to analyse the reliability of high dimensional non-linear dynamic systems.

This study examined ship dynamic reaction time series collected aboard a TEU2800 Panamax container ship over more than 70 transatlantic journeys between 2007 and 2010. The vessel reliability as a multi-dimensional system used novel reliability methodologies in real-time. The suggested method's theoretical justification is explained in depth. It should be noted that while using direct measurement or Monte Carlo simulation to analyse the reliability of dynamic systems is appealing, the complexity and high dimensionality of dynamic systems necessitate the development of novel, accurate, and robust techniques that can handle the available data while utilising it as effectively as possible.

The approaches discussed in this study have already been shown effective when applied to a wide variety of simulation models, but only for one-dimensional system responses. In general, highly precise predictions were made. This work focused on a general-purpose, reliable, and user-friendly multi-dimensional reliability approach. The proposed strategy yielded an excellent confidence interval, as demonstrated. As a result, the recommended technique might be helpful in a range of non-linear dynamic systems reliability investigations. Time series responses can be measured and numerically simulated and studied.

In contrast to prior reliability techniques, the new technique does not call for restarting Monte Carlo-type numerical simulation every time the system fails. As this paper's illustration of measured structural reaction shows, it is also feasible to accurately anticipate the likelihood of system collapse.

To sum up, the recommended technique may be used in various engineering fields. By no means does the given naval architecture example restrict the potential applications of a new methodology.

Data availability

References

1. Rice, S. O. Mathematical analysis of random noise. *Bell Syst. Tech. J.* **23**, 282–332 (1944).
2. Christou, M. & Ewans, K. Field measurements of rogue water waves. *J. Phys. Oceanogr.* **9**, 2317–2335 (2014).
3. Doeleman, M. W. *Rogue Waves in the Dutch North Sea*. Master's Thesis, TU Delft (2021).
4. Ducrozet, G., Abdolahpour, M., Nelli, F. & Toffoli, A. Predicting the occurrence of rogue waves in the presence of opposing currents with a high-order spectral method. *Phys. Rev. Fluids* **6**, 064803 (2021).
5. Forristall, G. On the distributions of wave heights in a storm. *J. Geophys. Res.* **83**, 2353–2358 (1978).
6. Glukhovskii, B. *Investigation of Sea Wind Waves (in Russian)* (Gidrometeoizdat, 1966).
7. Haring, R., Osborne, A. & Spencer, L. Extreme wave parameters based on continental shelf storm wave records. in Proceedings of 15th Conference on Coastal Engineering, Honolulu, HI, 151–170 (1976).
8. Jahns, H. & Wheeler, J. Long-term wave probabilities based on hindcasting of severe storms. *J. Petrol. Technol.* **25**, 473–486 (1973).
9. Karmpadakis, I., Swan, C. & Christou, M. Assessment of wave height distributions using an extensive field database. *Coast. Eng.* **157**, 103630 (2020).
10. Karmpadakis, I., Swan, C. & Christou, M. A new wave height distribution for intermediate and shallow water depths. *Coast. Eng.* **175**, 104130 (2022).
11. Kimmoun, O., Hsu, H. C., Hoffmann, N. & Chabchoub, A. Experiments on uni-directional and non-linear wave group shoaling. *Ocean Dyn.* **71**, 1105–1112 (2021).
12. Kinsman, B. *Surface Waves at Short Fetches and Low Wind Speed—A Field Study*. Chesapeake Bay. Inst., Technical Report (1960).
13. Li, Y., Draycott, S., Adcock, T. A. & Van Den Bremer, T. Surface wavepackets subject to an abrupt depth change. Part 2: Experimental analysis. *J. Fluid Mech.* **915**, A72 (2021).
14. Li, Y. *et al.* Why rogue waves occur atop abrupt depth transitions. *J. Fluid Mech.* **919**, R5 (2021).
15. Li, Y., Zheng, Y., Lin, Z., Adcock, T. A. & Van Den Bremer, T. Surface wavepackets subject to an abrupt depth change. Part 1: Second-order theory. *J. Fluid Mech.* **915**, A71 (2021).
16. Longuet-Higgins, M. On the statistical distribution of the heights of sea waves. *J. Mar. Res.* **11**, 245–265 (1952).
17. Longuet-Higgins, M. S. On the distribution of the heights of sea waves: Some effects of nonlinearity and finite band width. *J. Geophys. Res.* **85**, 1519–1523 (1980).
18. Majda, A., Moore, M. & Qi, D. Statistical dynamical model to predict extreme events and anomalous features in shallow water waves with abrupt depth change. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 3982–3987 (2019).
19. Mendes, S. & Kasparian, J. Saturation of rogue wave amplification over steep shoals. *Phys. Rev. E* **106**, 065101 (2022).
20. Mendes, S. & Scotti, A. The rayleigh-haring-tayfun distribution of wave heights in deep water. *Appl. Ocean Res.* **113**, 102739 (2021).
21. Mendes, S., Scotti, A., Brunetti, M. & Kasparian, J. Non-homogeneous model of rogue wave probability evolution over a shoal. *J. Fluid Mech.* **939**, A25 (2022).
22. Mendes, S., Scotti, A. & Stansell, P. On the physical constraints for the exceeding probability of deep water rogue waves. *Appl. Ocean Res.* **108**, 102402 (2021).
23. Miles, J. On the generation of surface waves by shear flows. *J. Fluid Mech.* **3**, 185–204 (1957).
24. Moore, N., Bolles, C., Majda, A. & Qi, D. Anomalous waves triggered by abrupt depth changes: Laboratory experiments and truncated kdv statistical mechanics. *J. Nonlinear Sci.* **30**, 3235–3263 (2020).
25. Phillips, O. On the generation of waves by turbulent wind. *J. Fluid Mech.* **2**, 417–445 (1957).
26. Phillips, O. The equilibrium range in the spectrum of wind-generated waves. *J. Fluid Mech.* **4**, 426–434 (1958).
27. Phillips, O. Spectral and statistical properties of the equilibrium range in wind-generated gravity waves. *J. Fluid Mech.* **156**, 505–531 (1985).
28. Pierson, W. J. & Marks, W. The power spectrum analysis of ocean-wave records. *Trans. Am. Geophys. Union* **33**, 834–844 (1952).
29. Pierson, W. J. & Moskowitz, L. A proposed spectral form for fully developed wind seas based on the similarity theory of S. A. Kitaigorodskii. *J. Geophys. Res.* **69**, 5181–5190 (1964).
30. Stansell, P. Distribution of freak wave heights measured in the north sea. *Appl. Ocean Res.* **26**, 35–48 (2004).
31. Tayfun, M. A. Narrow-band non-linear sea waves. *J. Geophys. Res.* **85**, 1548–1552 (1980).
32. Tayfun, M. A. & Fedele, F. Wave-height distributions and non-linear effects. *Ocean Eng.* **34**, 1631–1649 (2007).
33. Toffoli, A. *et al.* Rogue waves in opposing currents: An experimental study on deterministic and stochastic wave trains. *J. Fluid Mech.* **769**, 277–297 (2015).
34. Trulsen, K., Raustøl, A., Jorde, S. & Rye, L. Extreme wave statistics of long-crested irregular waves over a shoal. *J. Fluid Mech.* **882**, R2 (2020).
35. Trulsen, K., Zeng, H. & Gramstad, O. Laboratory evidence of freak waves provoked by non-uniform bathymetry. *Phys. Fluids* **24**, 097101 (2012).
36. Madsen, H. O., Krenk, S. & Lind, N. C. *Methods of Structural Safety* (Prentice-Hall Inc, 1986).
37. Ditlevsen, O. & Madsen, H. O. *Structural Reliability Methods* (Wiley, 1996).
38. Melchers, R. E. *Structural Reliability Analysis and Prediction* (Wiley, 1999).
39. Falzarano, J., Su, Z. & Jamnongpipatkul, A. Application of stochastic dynamical system to non-linear ship rolling problems. in *Proceedings of the 11th International Conference on the Stability of Ships and Ocean Vehicles, Athens, Greece* (2012).
40. Su, Z. *Non-linear Response and Stability Analysis of Vessel Rolling Motion in Random Waves Using Stochastic Dynamical Systems* (Texas University, 2012).
41. Ellermann, K. Non-linear dynamics of offshore systems in random seas. in *IUTAM Symposium on Fluid-Structure Interaction in Ocean Engineering*, 45–56 (Springer, 2008).
42. Andersen, I. M. V. & Jensen, J. J. Measurements in a container ship of wave induced hull girder stresses in excess of design values. *Mar. Struct.* **37**, 54–85. https://doi.org/10.1016/j.marstruc.2014.02.006 (2014).
43. Gaidai, O. et al. Offshore renewable energy site correlated wind-wave statistics. *Probab. Eng. Mech.* **68**, 103207. https://doi.org/10.1016/j.probengmech.2022.103207 (2022).
44. Gaidai, O. et al. Bivariate statistics of wind farm support vessel motions while docking. *Ships Offshore Struct.* **16**(2), 135–143 (2020).
45. Gaidai, O., Storhaug, G. & Naess, A. Extreme large cargo ship panel stresses by bivariate ACER2D method. *Ocean Eng.* **123**, 432–439 (2016).
46. Gaidai, O., Storhaug, G. & Naess, A. Statistics of extreme hydro elastic response for large ships. *Marine Struct.* **61**, 142–154 (2018).
47. Gaidai, O., Storhaug, G. & Naess, A. "Extreme value statistics of whipping response for large ships" PRADS proceedings. *Pract. Des. Ships Other Float. Struct.* **2**, 1210–1221 (2010).

48. Gaidai, O., Storhaug, G. & Naess, A. "Extreme value statistics of ship rolling" PRADS proceedings. *Pract. Des. Ships Other Float. Struct.* **2**, 457–466 (2010).

49. Gao, H., Gaidai, O., Naess, A., Storhaug, G. & Xu, X. Improving container ship panel stress prediction, based on another highly correlated panel stress measurement. *Mar. Struct.* **64**, 138–145 (2018).

50. Gaidai, O., Cao, Y. & Loginov, S. Global cardiovascular diseases death rate prediction. In *Current Problems in Cardiology*, https://doi.org/10.1016/j.cpcardiol.2023.101622 (2023).

51. Sun, J. *et al.* Extreme riser experimental loads caused by sea currents in the Gulf of Eilat. *Probab. Eng. Mech.* **68**, 103243. https://doi.org/10.1016/j.probengmech.2022.103243 (2022).

52. Numerical Algorithms Group. *NAG Toolbox for Matlab* (NAG Ltd, 2010).

53. ClassNK. Investigation report on structural safety of large container ships. *The Investigative Panel on Large Container Ship Safety* (2014). http://www.classnk.or.jp/hp/pdf/news/Investigation_Report_on_Structural_Safety_of_Large_Container_Ships_EN_ClassNK.pdf.

54. MAIB. *Report on the investigation of the structural failure of MSC Napoli English Channel on 18th January 2007, Marine Accident Investigation Branch (MAIB), Carlton House, Carlton Place, Southampton, UK, SO15 2DZ, Report No. 9/2008* (2008). https://www.gov.uk/maib-reports.

55. DNV. *Hull Monitoring Systems. DNV Rules for Classification of Ships*, Pt. 6, Ch. 11 (2005).

56. DNV. *Hull Structural Design, Ships with Length 100 Meters and Above, DNV Rules for Classification of Ships*, Pt. 3, Ch. 1 (2009).

57. DNV. *Fatigue Assessment of Ship Structures. DNV Classification Note CG-0129* (2015).

58. DNV. *Fatigue Assessment of Ship Structures. DNV GL Class Guideline DNVGL-CG-0129 Edition* (2018).

59. DNV. *Rules for Classification: Ships—DNVGL-RU-SHIP*, Pt. 3 Ch. 9, Fatigue, Edition (2018).

Author contributions

O.G.—theory and innovation. P.Y.—conceptualization. J.X.—management, corresponding author. Y.X.—validation and data analysis. K.W.—data preparation. Z.L.—editing.

Competing interests

The authors declare no competing interests.

Additional information

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

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

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