scientific reports

OPEN

Novel methods for coupled prediction of extreme wind speeds and wave heights

Oleg Gaidai[1], Yihan Xing[2]✉ & Xiaosen Xu[3]

Two novel methods are being outlined that, when combined, can be used for spatiotemporal analysis of wind speeds and wave heights, thus contributing to global climate studies. First, the authors provide a unique reliability approach that is especially suited for multi-dimensional structural and environmental dynamic system responses that have been numerically simulated or observed over a substantial time range, yielding representative ergodic time series. Next, this work introduces a novel deconvolution extrapolation technique applicable to a wide range of environmental and engineering applications. Classical reliability approaches cannot cope with dynamic systems with high dimensionality and responses with complicated cross-correlation. The combined study of wind speed and wave height is notoriously difficult, since they comprise a very complex, multi-dimensional, non-linear environmental system. Additionally, global warming is a significant element influencing ocean waves throughout the years. Furthermore, the environmental system reliability method is crucial for structures working in any particular region of interest and facing actual and often harsh weather conditions. This research demonstrates the effectiveness of our approach by applying it to the concurrent prediction of wind speeds and wave heights from NOAA buoys in the North Pacific. This study aims to evaluate the state-of-the-art approach that extracts essential information about the extreme responses from observed time histories.

It is challenging to accurately assess multi-dimensional dynamic system reliability using conventional theoretical reliability methods[1–4]. Estimating the reliability of a complicated dynamic system is easy if sufficient system response data are available or if direct Monte Carlo simulations are performed[5]. However, experimental or computational costs may be expensive for many engineering dynamic systems with complicated dynamics. Motivated by this, the authors developed a new reliability method for dynamic systems that require significantly less measurement and computing costs.

It has been a challenge to how exactly one predicts the shape and characteristics of waves through wind speed variables, even though this has been started and partially solved in the works of[6–11]. Our work, however, deals with the general approach of extreme value theory in which no physical dynamics of the water waves are expected to play a significant role in driving the appearance of rare events, i.e. there is an expected universality of extreme events in a variety of different physical and natural systems[12–17]. Although beyond the scope of our work, there Some successful attempts have been to study extreme events in water waves (often called rogue or freak waves) with distributions uniquely determined by the dynamics of the physical system. For instance[18,19], have shown that water waves departing from linear theory will modify their distribution from a Rayleigh type[20–22] to a distribution dependent on the square root of the wave steepness. Similarly[23,24], 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 types of effects in extreme wave distribution depending on whether they are in deep[25–28] or shallow water[13]. Furthermore, ocean processes such as shoaling or wave-current systems that drive wave trains out of equilibrium have been experimentally[22,25,29–37] associated with increasing the occurrence of extreme waves by order of magnitude. However, it has been recently found that no 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[6–37].

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.

[1]Shanghai Ocean University, Shanghai, China. [2]University of Stavanger, Stavanger, Norway. [3]Jiangsu University of Science and Technology, Zhenjiang, China. ✉email: yihan.xing@uis.no

nature portfolio

1

Method

Consider multi-degree of freedom (MDOF) jointly stationary dynamic environmental system with representative response vector process $\boldsymbol{R}(t) = (X(t), Y(t), Z(t), \ldots)$, that has been either measured or simulated over a sufficiently long time span $(0, T)$. Unidimensional global maxima over entire time span $(0, T)$ is denoted as $X_T^{\max} = \max\limits_{0<t\leq T} X(t), Y_T^{\max} = \max\limits_{0\leq t\leq T} Y(t), Z_T^{\max} = \max\limits_{0\leq t\leq T} Z(t), \ldots$.

By sufficiently long time span T authors mean large enough value of T with respect to the environmental dynamic system auto-correlation time. Let $X_1, \ldots, X_{N_X}$ be consequent temporal local maxima of the environmental process $X(t)$ at discrete monotonously temporally increasing times $t_1^X < \ldots < t_{N_X}^X$ within $(0, T)$. Similar definition follows for other MDOF environmental 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 $\boldsymbol{R}(t)$ components, maxima are assumed to be non-negative. The target now is to accurately estimate environmental system failure probability, namely its probability of exceedance

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

where $P = \iiint_{(0,0,0,\ldots)}^{(\eta_X, \eta_Y, \eta_Z, \ldots)} p_{X_T^{\max}, Y_T^{\max}, Z_T^{\max}, \ldots}(X_T^{\max}, Y_T^{\max}, Z_T^{\max}, \ldots) dX_T^{\max} dY_{N_Y}^{\max} dZ_{N_Z}^{\max} \ldots$ being probability of non-exceedance for critical values of environmental response components $\eta_X, \eta_Y, \eta_Z, \ldots$; $\cup$ denotes logical unity operation «or»; $p_{X_T^{\max}, Y_T^{\max}, Z_T^{\max}, \ldots}$ being joint probability density of the global maxima over the entire time span $(0, T)$. In practice, however, it is not always feasible to estimate directly $p_{X_T^{\max}, Y_T^{\max}, Z_T^{\max}, \ldots}$ joint probability distribution due to its high dimensionality and limitations of the available data set. The moment when either $X(t)$ exceeds η_X, or $Y(t)$ exceeds η_Y, or $Z(t)$ exceeds η_Z, and so on, the environmental dynamic system being regarded as failed. Fixed failure levels $\eta_X, \eta_Y, \eta_Z, \ldots$ being individual for each unidimensional response component of $\boldsymbol{R}(t)$. $X_{N_X}^{\max} = \max\{X_j ; j = 1, \ldots, N_X\} = X_T^{\max}$, $Y_{N_Y}^{\max} = \max\{Y_j ; j = 1, \ldots, N_Y\} = Y_T^{\max}$, $Z_{N_Z}^{\max} = \max\{Z_j ; j = 1, \ldots, N_Z\} = Z_T^{\max}$, and so on.

Let sort local maxima time instants $\left[t_1^X < \ldots < t_{N_X}^X; t_1^Y < \ldots < t_{N_Y}^Y; t_1^Z < \ldots < t_{N_Z}^Z\right]$ recorded in temporally non-decreasing order into one single synthetic merged time vector $t_1 \leq \ldots \leq t_N$. Note that then $t_N = \max\{t_{N_X}^X, t_{N_Y}^Y, t_{N_Z}^Z, \ldots\}, N = N_X + N_Y + N_Z + \ldots$. In this case t_j represents local maxima of one of MDOF environmental dynamic structural response components either $X(t)$ or $Y(t)$, or $Z(t)$ and so on. The latter means that having $\boldsymbol{R}(t)$ time record, one only needs simultaneously and continuously screen for the system unidimensional response component local maxima and record its exceedance of MDOF limit vector $(\eta_X, \eta_Y, \eta_Z, \ldots)$ in any of dynamic system components $X, Y, Z, \ldots$. Local unidimensional response component maxima then merged into one synthetic temporal non-decreasing vector $\vec{R} = (R_1, R_2, \ldots, R_N)$ in accordance with merged time vector $t_1 \leq \ldots \leq t_N$. That is to say each local maxima R_j is in fact 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 which of $X(t)$ or $Y(t)$, or $Z(t)$ etc. corresponding to the current local maxima with running index j.

Scaling parameter $0 < \lambda \leq 1$ being now introduced to artificially simultaneously decrease limit values for all system response components, i.e. new MDOF limit vector $(\eta_X^\lambda, \eta_Y^\lambda, \eta_z^\lambda, \ldots)$ with $\eta_X^\lambda \equiv \lambda \cdot \eta_X, \equiv \lambda \cdot \eta_Y, \eta_z^\lambda \equiv \lambda \cdot \eta_Z, \ldots$ being introduced. The unified limit vector $(\eta_1^\lambda, \ldots, \eta_N^\lambda)$ being introduced with each component η_j^λ being equal to either $\eta_X^\lambda, \eta_Y^\lambda$ or η_z^λ and so on The latter automatically defines the target 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

$$\begin{aligned} P(\lambda) &= \text{Prob}\left\{R_N \leq \eta_N^\lambda, \ldots, R_1 \leq \eta_1^\lambda\right\} \\ &= \text{Prob}\{R_N \leq \eta_N^\lambda | R_{N-1} \leq \eta_{N-1}^\lambda, \ldots, R_1 \leq \eta_1^\lambda\} \cdot \text{Prob}\left\{R_{N-1} \leq \eta_{N-1}^\lambda, \ldots, R_1 \leq \eta_1^\lambda\right\} \\ &= \prod_{j=2}^N \text{Prob}\{R_j \leq \eta_j^\lambda | R_{j-1} \leq \eta_{1j-}^\lambda, \ldots, R_1 \leq \eta_1^\lambda\} \cdot \text{Prob}\left(R_1 \leq \eta_1^\lambda\right) \end{aligned} \tag{2}$$

In practice the dependence between neighbouring R_j is not always negligible, thus following one-step (will be called conditioning level $k = 1$) memory approximation being introduced

$$\text{Prob}\{R_j \leq \eta_j^\lambda | R_{j-1} \leq \eta_{j-1}^\lambda, \ldots, R_1 \leq \eta_1^\lambda\} \approx \text{Prob}\{R_j \leq \eta_j^\lambda | R_{j-1} \leq \eta_{j-1}^\lambda\} \tag{3}$$

for $2 \leq j \leq N$ (will be called conditioning level $k = 2$). Approximation introduced by Eq. (3) may be further expressed as

$$\text{Prob}\{R_j \leq \eta_j^\lambda | R_{j-1} \leq \eta_{j-1}^\lambda, \ldots, R_1 \leq \eta_1^\lambda\} \approx \text{Prob}\{R_j \leq \eta_j^\lambda | R_{j-1} \leq \eta_{j-1}^\lambda, R_{j-2} \leq \eta_{j-2}^\lambda\} \tag{4}$$

where $3 \leq j \leq N$ (called here conditioning level $k = 3$), and so on[38]. Equation (4) presents series of subsequent refinements of statistical independence assumption. The latter type of approximations models statistical dependence effect between neighbouring maxima with an increased accuracy. Since original MDOF dynamic process $\boldsymbol{R}(t)$ was assumed ergodic and therefore stationary, probability $p_k(\lambda) := \text{Prob}\{R_j > \eta_j^\lambda | R_{j-1} \leq \eta_{j-1}^\lambda, R_{j-k+1} \leq \eta_{j-k+1}^\lambda\}$ for $j \geq k$ is independent of j but only dependent on conditioning level k. Thus non-exceedance probability may be now approximated

$$P_k(\lambda) \approx \exp\left(-N \cdot p_k(\lambda)\right), k \geq 1. \tag{5}$$

Note that Eq. (5) follows from Eq. (1) by neglecting $\mathrm{Prob}\left(R_1 \leq \eta_1^\lambda\right) \approx 1$, as design failure probability must of a small order of magnitude, and $N \gg k$. Equation (5) is similar to mean up-crossing rate equation for probability of exceedance[39–51]. There is convergence with respect to the conditioning parameter k

$$P = \lim_{k \to \infty} P_k(1); \quad p(\lambda) = \lim_{k \to \infty} p_k(\lambda) \tag{6}$$

Note that Eq. (5) for $k = 1$ turns into classic non-exceedance probability relationship with corresponding mean up-crossing rate function

$$P(\lambda) \approx \exp\left(-\nu^+(\lambda)\, T\right); \quad \nu^+(\lambda) = \int_0^\infty \zeta p_{R\dot{R}}(\lambda, \zeta)d\zeta \tag{7}$$

where $\nu^+(\lambda)$ denotes the mean up-crossing rate of the response level λ for the above assembled non-dimensional vector $R(t)$ assembled from scaled MDOF system response $\left(\frac{X}{\eta_X}, \frac{Y}{\eta_Y}, \frac{Z}{\eta_Z}, \ldots\right)$. The mean up-crossing rate is given by the Rice's formula given in Eq. (7) with $p_{R\dot{R}}$ being joint probability density for $(R, \dot{R})$ with $\dot{R}$ being the time derivative $R'(t)$ [1]. Equation (7) relies on the Poisson assumption, that is up-crossing events of high λ levels (in this paper it is $\lambda \geq 1$) may be assumed to be independent.

The proposed methodology may be used for nonstationary cases as well. Given environmental scattered diagram of $m = 1, .., M$ sea states, each environmental short-term sea state having probability q_m, so that $\sum_{m=1}^{M} q_m = 1$. Next, let one introduce the long-term equation

$$p_k(\lambda) \equiv \sum_{m=1}^{M} p_k(\lambda, m) q_m \tag{8}$$

with $p_k(\lambda, m)$ being the same function as in Eq. (6), but corresponding to a specific short-term sea state with number m. The above introduced $p_k(\lambda)$ as functions being often regular in the tail, specifically for values of λ approaching and exceeding 1. More specifically, for $\lambda \geq \lambda_0$, the probability distribution tail behaves similar to $\exp\left\{-(a\lambda + b)^c + d\right\}$ with a, b, c, d being suitably fitted constants for suitable tail cut-on λ_0 value. Therefore, one can write

$$p_k(\lambda) \approx \exp\left\{-(a_k\lambda + b_k)^{c_k} + d_k\right\}, \lambda \geq \lambda_0 \tag{9}$$

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 being observed. Optimal values of the parameters a_k, b_k, c_k, p_k, q_k may be determined using a sequential quadratic programming (SQP) method incorporated in the NAG Numerical Library[52].

Let us consider stationary stochastic process $X(t)$, being either simulated or measured over a certain time span $0 \leq t \leq T$, and which is being represented as a sum of two independent identically distributed stationary environmental dynamic processes $X_1(t)$ and $X_2(t)$, namely

$$X(t) = X_1(t) + X_2(t) \tag{10}$$

Denoting probability density function (PDF) of $X(t)$ as p_X and PDF of $X_1(t)$ and $X_2(t)$ as p_{X_1}, following convolution equation holds

$$p_X = \mathrm{conv}\left(p_{X_1}, p_{X_1}\right) \tag{11}$$

In order to exemplify the latter idea regarding how to estimate unknown probability distribution robustly p_{X_1}, subsequently improving given empirical distribution p_X. In the next, authors briefly discuss common knowledge regarding the discrete convolution of two vectors. The convolution of two vectors, u and v, representing an area of overlap of vector components, as v slides across u. Algebraically, convolution being the same operation as multiplying polynomials whose coefficients are the elements of, u and v. Let $m = \mathrm{length}(u)$ and $n = \mathrm{length}(v)$. Then w being the vector of length $m + n - 1$, with k-th element being

$$w(k) = \sum_{j=1}^{m} u(j)v(k - j + 1) \tag{12}$$

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 study

$$w(1) = u(1) \cdot v(1)$$
$$w(2) = u(1) \cdot v(2) + u(2) \cdot v(1)$$
$$w(3) = u(1) \cdot v(3) + u(2) \cdot v(2) + u(3) \cdot v(1)$$
$$\ldots$$
$$w(n) = u(1) \cdot v(n) + u(2) \cdot v(n-1) + \cdots + u(n) \cdot v(1) \qquad (13)$$
$$\ldots$$
$$w(2n-1) = u(n) \cdot v(n)$$

From Eq. (12) one can also observe that having found $u = v = (u(1), .., u(n))$, one can gradually obtain w-components $w(n+1), \ldots, w(2n-1)$, as index increases from $n+1$ to $2n-1$. The latter clearly would extend vector w into support domain that is twice longer than the original distribution support domain, i.e. doubling the p_X distribution support length $(2n-1) \cdot \Delta x \approx 2n \cdot \Delta x = 2X_L$, as compared to the original probability distribution support length $n \cdot \Delta x = X_L$ with Δx being constant length of each descrete distribution bin. Note that $w = (w(1), \ldots, w(n))$ being discrete representation of the target empirical distribution p_X, and n representing length of probability distribution support $[0, X_L]$, for simplicity in this paper one is limited to the case of one-sided positive valued random variables, namely $X \geq 0$. As will be further discussed, a simple linear extrapolation of self deconvoluted vector $(u(1), \ldots, u(n))$ towards $(u(n+1), \ldots, u(2n-1))$ is suggested, with p_{X_1} having its tail linearly extrapolated within the range $(X_L, 2X_L)$. Using Eq. (11) original vector w will be extended and extrapolated into support domain being twice longer than original distribution support domain, namely doubling p_X distribution support length $(2n-1) \cdot \Delta x \approx 2n \cdot \Delta x = 2X_L$, as compared to original probability distribution support length $n \cdot \Delta x = X_L$. To smoothen original distribution $p_X(x)$ tail, authors have performed distribution p_X tail interpolation for high tail values x, using Naess-Gaidai (NG) extrapolation method[53]. As has been discussed above, the proposed deconvolution extrapolation technique has an advantage of not presuming any specific extrapolation functional class needed to extrapolate distribution tail.

To validate the above-suggested extrapolation methodology, the «shorter» version of the original data set was used for extrapolation to compare with predictions based on a full «longer» data set.

Results

This section aims to demonstrate the efficiency of the previously described methodology by applying the new method to the National Oceanic and Atmospheric Administration buoy (NOAA) ten-minute average wind speed values and significant wave heights (calculated as the average of the highest one-third of all wave heights during the 20-min sampling period) in the North Pacific region near the Hawaiian Islands. For this study, the NOAA wind and wave measurement location Station 51003 (LLNR 28005.7)—WESTERN HAWAII—205 NM SW of Honolulu was selected, and its measured wind speed and wave height values were set as two environmental system components (dimensions) X, Y, thereby serving as an example of a two-dimensional (2D) environmental system. Unidimensional extreme response values were chosen as the whole analysed data set maximum wind speeds and wave heights respectively during observation period between years 2009–2014 for the chosen buoy in situ measurement location.

Figure 1 presents National Oceanic and Atmospheric Administration[54], buoy locations in North Pacific, blue circle indicates area of interest. Figure 2 shows data buoy containing sensors used to monitor and collect atmospheric and oceanographic conditions, collected data is then converted into an electronic signal and transmitted to shore or logged in the onboard data unit. This 3-m foam buoy has following characteristics:

Site elevation: sea level.
Air temp height: 3.4 m above site elevation.
Anemometer height: 3.8 m above site elevation.
Barometer elevation: 2.4 m above mean sea level.
Sea temp depth: 2 m below water line.
Water depth: 1987 m.
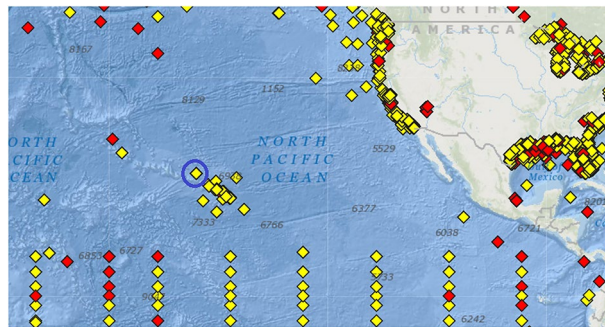
Figure 1. Wind and wave measurements location according to National Oceanic and Atmospheric Administration, blue circle indicates measurement station of interest. Figure taken from National Oceanic and Atmospheric Administration, https://www.ndbc.noaa.gov.

Figure 2. National Oceanic and Atmospheric Administration buoy[54].
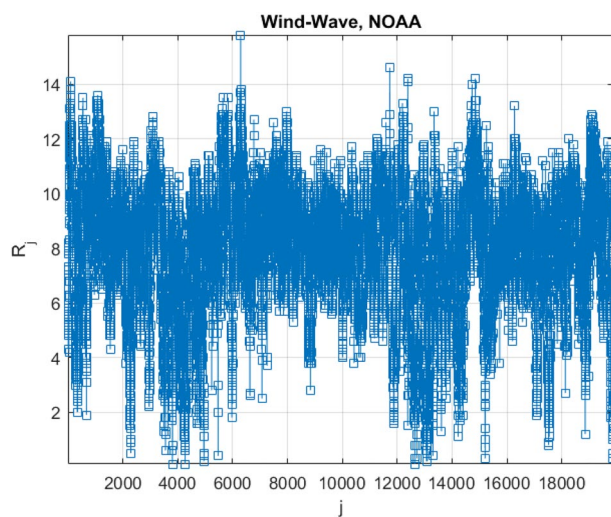
**Figure 3.** Scaled non-dimensional assembled 2D vector $\vec{R}$.

Watch circle radius: 5004 yards.

In order to unify all two measured time series X, Y the following scaling was performed:

$$X \rightarrow \frac{X}{\eta_X}, Y \rightarrow \frac{Y}{\eta_Y} \tag{14}$$

making all two responses non-dimensional and having the same failure limit equal to 1. Next, all local maxima from two measured time series were merged into one single time series by keeping them in time non-decreasing order: $\vec{R} = (\max\{X_1, Y_1\}, \ldots, \max\{X_N, Y_N\})$ with the whole vector $\vec{R}$ being sorted in temporally non-decreasing order of occurrence of these local maxima.

Figure 3 presents example of non-dimensional assembled vector $\vec{R}$, consisting of assembled local maxima of raw daily largest highest daily windcast speed data. The failure probability distribution tail extrapolation was performed towards 100 years return period. Synthetic vector $\vec{R}$ does not have physical meaning on its own, as it assembled of completely different response components. Index j is just a running index of local maxima encountered in non-decreasing time sequence. Index j being running index of local maxima encountered in non-decreasing time sequence. The «shorter» data record was generated by taking each tenth data point from the «longer» data record.
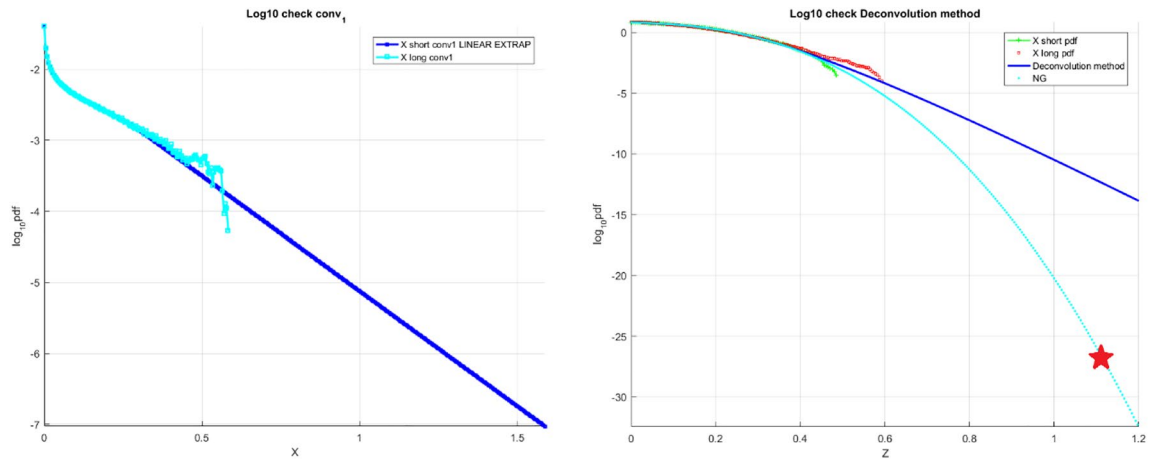
Figure 4. Combined wind speed and wave height data. Left: scaled f_{X_1} tail on the decimal log scale for the «shorter» data (cyan), linearly extrapolated (dark blue). Right: unscaled raw «shorter» data (green) f_X tail on the decimal log scale, extrapolated by deconvolution method (dark blue), along with «longer» raw data (red) and NG extrapolation (cyan). Star indicates 100 years return period prediction.

Figure 4 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 4 on the right presents 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.

It is seen that NG prediction does not agree with deconvolution approach, leading to non-conservative prediction, indicated by star in Fig. 4 on the right. It is seen from Fig. 4 on the right that the proposed method performs quite well, being based on the «shorter» data set, and delivering probability distribution close to the one based on the «longer» data set. Figure 4 on the right presents extrapolation according to Eq. (9) as well novel deconvolution method towards failure state with 100 year return period, which is 1, and somewhat beyond, $\lambda = 0.05$ 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 $1 - P \approx 1 - P_k(1)$ may be now estimated. Note that in Eq. (5) N corresponds to a total number of local maxima in the unified response vector $\vec{R}$. Conditioning parameter $k = 6$ was found to be sufficient due to occurrence of convergence with respect to k, see Eq. (6). The generalization potential of the proposed methods is extension to non-stationary systems with an underlying trend. If the underlying trend would be known, which is rarely the case, it may be subtracted to bring measurements to stationary state. Otherwise proper trend analysis would be needed in combination with advocated here methods. The main model assumptions subsequent limitation is the system quasi-stationarity as mentioned before.

Conclusions

The key advantage of introduced methodology is its ability to assess reliability of non-linear dynamic systems with high dimensionality.

This paper studied wind speeds and wave heights measured by National Oceanic and Atmospheric Administration buoys in North Pacific region, during years 2009–2014. Novel environmental reliability method has been applied to predict occurrence of extreme waves within time horizon of 100 years. Theoretical reasoning behind the proposed method has been given in detail.

Methods introduced in this paper, have been previously validated by application to range of dynamic environmental systems, but for only one-dimensional system responses. This study has aimed at further development of robust and simple general purpose multi-dimensional reliability method. Both measured and numerically simulated time series responses can be analyzed. In case of measured system response, as illustrated in this paper, an accurate prediction of environmental system failure probability is possible.

Finally, the suggested methodology can be used in wide range of modern engineering areas of applications. The presented environmental example does not limit applicability area of new method by any means.

Data availability

References

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

2. Madsen, H. O., Krenk, S. & Lind, N. C. *Methods of Structural Safety* (Prentice-Hall Inc, 1986).
3. Ditlevsen, O. & Madsen, H. O. *Structural Reliability Methods* (Wiley, 1996).
4. Melchers, R. E. *Structural Reliability Analysis and Prediction* (Wiley, 1999).
5. Naess, A. & Gaidai, O. Estimation of extreme values from sampled time series. *Struct. Saf.* **31**(4), 325–334 (2009).
6. Miles, J. On the generation of surface waves by shear flows. *J. Fluid Mech.* **3**, 185–204 (1957).
7. Phillips, O. On the generation of waves by turbulent wind. *J. Fluid Mech.* **2**, 417–445 (1957).
8. Phillips, O. The equilibrium range in the spectrum of wind-generated waves. *J. Fluid Mech.* **4**, 426–434 (1958).
9. Phillips, O. Spectral and statistical properties of the equilibrium range in wind-generated gravity waves. *J. Fluid Mech.* **156**, 505–531 (1985).
10. Pierson, W. J. & Marks, W. The power spectrum analysis of ocean-wave records. *Trans. Am. Geophys. Union* **33**, 834–844 (1952).
11. 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).
12. Christou, M. & Ewans, K. Field measurements of rogue water waves. *J. Phys. Oceanogr.* **9**, 2317–2335 (2014).
13. Doeleman, M. W. Rogue waves in the Dutch North Sea. Master's thesis, TU Delft (2021).
14. 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).
15. Forristall, G. On the distributions of wave heights in a storm. *J. Geophys. Res.* **83**, 2353–2358 (1978).
16. Gaidai, O. *et al.* Novel methods for wind speeds prediction across multiple locations. *Sci. Rep.* **12**, 1–9 (2022).
17. Glukhovskii, B. *Investigation of Sea Wind Waves (in russian)* (Gidrometeoizdat, 1966).
18. Tayfun, M. A. Narrow-band non-linear sea waves. *J. Geophys. Res.* **85**, 1548–1552 (1980).
19. Tayfun, M. A. & Fedele, F. Wave-height distributions and non-linear effects. *Ocean Eng.* **34**, 1631–1649 (2007).
20. Karmpadakis, I., Swan, C. & Christou, M. Assessment of wave height distributions using an extensive field database. *Coast. Eng.* **157**, 103630 (2020).
21. Karmpadakis, I., Swan, C. & Christou, M. A new wave height distribution for intermediate and shallow water depths. *Coast. Eng.* **175**, 104130 (2022).
22. 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).
23. Haring, R., Osborne, A. & Spencer, L. Extreme wave parameters based on continental shelf storm wave records. In *Proc. 15th Int. Conf. on Coastal Engineering*, 151–170 (1976).
24. Jahns, H. & Wheeler, J. Long-term wave probabilities based on hindcasting of severe storms. *J. Pet. Technol.* **25**, 473–486 (1973).
25. Longuet-Higgins, M. On the statistical distribution of the heights of sea waves. *J. Mar. Res.* **11**, 245–265 (1952).
26. 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).
27. Zhang, J., Benoit, M., Kimmoun, O., Chabchoub, A. & Hsu, H. C. Statistics of extreme waves in coastal waters: Large scale experiments and advanced numerical simulations. *Fluids* **4**, 99 (2019).
28. Zhang, H., Reynolds, R. & Bates, J. *Blended and Gridded High Resolution Global Sea Surface Wind Speed and Climatology from Multiple Satellites: 1987 - Present* (American Meteorological Society, 2006).
29. Kinsman, B. Surface waves at short fetches and low wind speed—a field study. Chesapeake Bay. Inst.,Tech. Rep (1960).
30. 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).
31. Li, Y. *et al.* Why rogue waves occur atop abrupt depth transitions. *J. Fluid Mech.* **919**, R5 (2021).
32. 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).
33. 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).
34. Mendes, S., Kasparian, J., 2022. Saturation of rogue wave amplification over steep shoals. Physical Review E - accepted for publication (https://journals.aps.org/pre/accepted/a2077KaaI041ec0626a21bd0ec4388c81ed896128). https://arxiv.org/abs/physics.flu-dyn/2207.13869.
35. Mendes, S. & Scotti, A. The rayleigh-haring-tayfun distribution of wave heights in deep water. *Appl. Ocean Res.* **113**, 102739 (2021).
36. 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).
37. 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).
38. Naess, A., Leira, B. J. & Batsevych, O. System reliability analysis by enhanced Monte Carlo simulation. *Struct. Saf.* **31**, 349–355 (2009).
39. Xing, Y., Gaidai, O., Ma, Y., Naess, A. & Wang, F. A novel design approach for estimation of extreme responses of a subsea shuttle tanker hovering in ocean current considering aft thruster failure. *Appl. Ocean Res.* https://doi.org/10.1016/j.apor.2022.103179 (2022).
40. Gaidai, O. *et al.* Offshore renewable energy site correlated wind-wave statistics. *Probab. Eng. Mech.* https://doi.org/10.1016/j.probengmech.2022.103207 (2022).
41. Sun, J. *et al.* Extreme riser experimental loads caused by sea currents in the Gulf of Eilat. *Probab. Eng. Mech.* https://doi.org/10.1016/j.probengmech.2022.103243 (2022).
42. Gaidai, O. & Naess, A. Extreme response statistics for drag dominated offshore structures. *Probab. Eng. Mech.* **23**, 180–187 (2008).
43. Xu, X. *et al.* Bivariate statistics of floating offshore wind turbine dynamic response under operational conditions. *Ocean Eng.* https://doi.org/10.1016/j.oceaneng.2022.111657 (2022).
44. Gaidai, O. *et al.* Improving extreme anchor tension prediction of a 10-MW floating semi-submersible type wind turbine, using highly correlated surge motion record. *Front. Mech. Eng.* https://doi.org/10.3389/fmech.2022.888497 (2022).
45. Gaidai, O., Xing, Y. & Xu, X. COVID-19 epidemic forecast in USA East coast by novel reliability approach. *Res. Square* https://doi.org/10.21203/rs.3.rs-1573862/v1 (2022).
46. Xu, X. *et al.* A novel multi-dimensional reliability approach for floating wind turbines under power production conditions. *Front. Mar. Sci.* https://doi.org/10.3389/fmars.2022.970081 (2022).
47. Gaidai, O., Xing, Y. & Balakrishna, R. Improving extreme response prediction of a subsea shuttle tanker hovering in ocean current using an alternative highly correlated response signal. *Results Eng.* https://doi.org/10.1016/j.rineng.2022.100593 (2022).
48. Cheng, Y., Gaidai, O., Yurchenko, D., Xu, X. & Gao, S. Study on the dynamics of a payload influence in the polar ship. In *The 32nd International Ocean and Polar Engineering Conference*, Paper Number: ISOPE-I-22-342 (2022).
49. Gaidai, O., Storhaug, G., Wang, F., Yan, P., Naess, A., Wu, Y., Xing, Y. & Sun, J. On-board trend analysis for cargo vessel hull monitoring systems. In *The 32nd International Ocean and Polar Engineering Conference*, Paper Number:ISOPE-I-22-541 (2022).
50. Gaidai, O. *et al.* Offshore renewable energy site correlated wind-wave statistics. *Probab. Eng. Mech.* https://doi.org/10.1016/j.probengmech.2022.103207 (2022).
51. 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).

52. Numerical Algorithms Group. *NAG Toolbox for Matlab* (NAG Ltd, 2010).
53. Naess, A. & Moan, T. *Stochastic Dynamics of Marine Structures* (Cambridge University Press, 2013).
54. National Oceanic and Atmospheric Administration, https://www.ndbc.noaa.gov

Author contributions

All authors contributed equally.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Y.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.