Novel methods for wind speeds prediction across multiple locations

Oleg Gaidai1, Jingxiang Xu1, Ping Yan1, Yihan Xing2, Yu Wu3 & Fuxi Zhang1

This article provides two unique methodologies that may be coupled to study the dependability of multidimensional nonlinear dynamic systems. First, the structural reliability approach is well suited for multidimensional environmental and structural reactions and is either measured or numerically simulated over sufficient time, yielding lengthy ergodic time series. Second, a unique approach to predicting extreme values has technical and environmental implications. In the event of measurable environmental loads, it is also feasible to calculate the probability of system failure, as shown in this research. In addition, traditional probability approaches for time series cannot cope effectively with the system's high dimensionality and cross-correlation across dimensions. It is common knowledge that wind speeds represent a complex, nonlinear, multidimensional, and cross-correlated dynamic environmental system that is always difficult to analyze. Additionally, global warming is a significant element influencing ocean waves throughout time. This section aims to demonstrate the efficacy of the previously mentioned technique by applying a novel method to the Norwegian offshore data set for the greatest daily wind cast speeds in the vicinity of the Landvik wind station. This study aims to evaluate the state-of-the-art approach for extracting essential information about the extreme reaction from observed time histories. The approach provided in this research enables the simple and efficient prediction of failure probability for the whole nonlinear multidimensional dynamic system.

Methods

It is challenging to estimate structural system reliability using classic engineering methods1-2. The latter is usually due to a large number of degrees of system freedom and random variables governing dynamic systems. The reliability of complex structural systems can be straightforwardly estimated either by having enough measurements or by direct numerical Monte Carlo simulations3-11. However, experimental and computational often are unaffordable for many complicated environmental systems. Authors have introduced a novel reliability method for structural systems aiming to reduce measurement or computational costs.

Ocean waves are usually assumed to be a random ergodic process (stationary and homogenous). Consider MDOF (multi-degree of freedom) offshore structure subjected to ergodic environmental loadings, for example, originating from surrounding waves and wind. MDOF structural response vector process \( \mathbf{R}(t) = (X(t), Y(t), Z(t), \ldots) \), that has been either measured or simulated over a long enough time interval \((0, T)\). Unidimensional response process maxima over the entire time span \((0, T)\) denoted as \(X_{T}^{\max} = \max_{0 \leq t \leq T} X(t)\), \(Y_{T}^{\max} = \max_{0 \leq t \leq T} Y(t), Z_{T}^{\max} = \max_{0 \leq t \leq T} Z(t), \ldots\)

Let \(X_{1}, \ldots, X_{N_{X}}\) represent local maxima of the time-dependent process \(X(t)\) at discrete monotonically rising time moments \(t_{1}^{X} < \cdots < t_{N_{X}}^{X}\) in \((0, T)\). Similar definitions may be provided for the additional MDOF response.

1 Shanghai Engineering Research Center of Marine Renewable Energy, College of Engineering Science and Technology, Shanghai Ocean University, Shanghai, China. 2 University of Stavanger, Stavanger, Norway. *Email: jxxu@shou.edu.cn
components \(Y(t), Z(t), \ldots \) using \(Y_1, \ldots, Y_{N_Y}; Z_1, \ldots, Z_{N_Z}, \ldots\). All MDOF \(R(t)\) components and their unidimensional maxima are considered to be bigger than zero in this investigation\textsuperscript{12}.

The objective is to determine the MDOF system failure probability, also known as the exceedance probability

\[
P = \text{Prob}(\max_{T \in \mathcal{T}} Y > \eta_Y, \max_{T \in \mathcal{T}} Z > \eta_Z, \ldots)
\]

with \(P = \int_{\mathcal{T}} \mathcal{P}_{\max_{T \in \mathcal{T}} Y > \eta_Y, \max_{T \in \mathcal{T}} Z > \eta_Z, \ldots} dY_{\max} dZ_{\max} \ldots\) being the non-exceedance probability for critical unidimensional response values \(\eta_Y, \eta_Z, \ldots\); with \(\cup\) denoting logical unity operation; \(\mathcal{P}\) denoting the global maxima joint probability density over the time period \((0, T]\). In reality, it is difficult to directly estimate the joint probability distribution due to its large dimensionality and the data collection limits.

The moment in time when either \(X(t)\) exceeds \(\eta_X\), or \(Y(t)\) exceeds \(\eta_Y\), or \(Z(t)\) exceeds \(\eta_Z\), and so on, the dynamic system is regarded as efficiently failed. Pre-defined failure levels \(\eta_X, \eta_Y, \eta_Z, \ldots\) are individual for every unidimensional response component of \(R(t)\). \(X_{\max} = \max \{X_j; j = 1, \ldots, N_X\} = X_{T}^T\) and \(Y_{\max} = \max \{Y_j; j = 1, \ldots, N_Y\} = Y_{T}^T\), \(Z_{\max} = \max \{Z_j; j = 1, \ldots, N_Z\} = Z_{T}^T\), \ldots, see Fig. 1.

The next step is to sort time instants unidimensional response component maxima \(t_{X_1} < \cdots < t_{X_N}; t_{Y_1} < \cdots < t_{Y_N}; t_{Z_1} < \cdots < t_{Z_N}\) in non-decreasing temporal order into one single merged time series vector \(t_1 \leq \cdots \leq t_N\), with \(N = N_X + N_Y + N_Z + \ldots\). With \(t_j\) representing unidimensional response component maxima of one of the MDOF structural response components. So having \(R(t)\), one must concurrently and continuously screen for each unidimensional response component temporal maxima, noting its exceedances of MDOF limit vector \((\eta_X, \eta_Y, \eta_Z, \ldots)\) in any of its unidimensional response components \(X, Y, Z, \ldots\). Each unidimensional response component maxima are then combined into a single temporal non-decreasing vector \(R = (R_1, R_2, \ldots, R_N)\) based on the merged time vector \(t_1 \leq \cdots \leq t_N\). Finally, the unified MDOF limit vector \((\eta_1, \ldots, \eta_N)\) is now introduced with each component \(\eta_j\) being either \(\eta_X, \eta_Y, \eta_Z, \ldots\), depending on which of \(X(t)\) or \(Y(t)\), or \(Z(t)\) … corresponds to the current unidimensional response component maxima with the running index \(j\), see\textsuperscript{12–16}.

Scaling parameter \(0 < \lambda \leq 1\) is now added to artificially limit lower limit values for all unidimensional response components, namely the new MDOF limit vector \((\eta_1^\lambda, \eta_2^\lambda, \ldots)\) with \(\eta_j^\lambda = \lambda \cdot \eta_j\), \(\lambda \cdot \eta_1, \lambda \cdot \eta_2, \ldots\) is now introduced, see\textsuperscript{12}. The unified MDOF limit vector \((\eta_1^\lambda, \ldots, \eta_N^\lambda)\) has each component \(\eta_j^\lambda\) as either \(\eta_1^\lambda, \eta_2^\lambda, \ldots\).

The latter specifies the probability distribution function \(P(\lambda)\) as a function of artificial parameter \(\lambda\), note that \(P = P(1)\) from Eq. (1). The non-exceedance probability function \(P(\lambda)\) may be expressed as follows

\[
P(\lambda) = \text{Prob}\{R_N \leq \eta_1^\lambda, \ldots, R_1 \leq \eta_1^\lambda\} = \text{Prob}\{R_N \leq \eta_1^\lambda | R_{N-1} \leq \eta_1^\lambda, \ldots, R_1 \leq \eta_1^\lambda\} \cdot \text{Prob}\{R_{N-1} \leq \eta_1^\lambda, \ldots, R_1 \leq \eta_1^\lambda\}
\]

\[
= \prod_{j=2}^{N} \text{Prob}\{R_j \leq \eta_1^\lambda | R_{j-1} \leq \eta_1^\lambda, \ldots, R_1 \leq \eta_1^\lambda\} \cdot \text{Prob}\{R_1 \leq \eta_1^\lambda\}
\]

The cascade of approximations based on conditioning is described. In practice, dependencies between adjacent local maxima \(R_i\) are not insignificant; hence the following one step (defined as conditioning level \(k = 1\)) memory approximation may be implemented.
\( \text{Prob}[R_j \leq \eta_j^d | R_{j-1} \leq \eta_{j-1}^d, \ldots, R_1 \leq \eta_1^d] \approx \text{Prob}[R_j \leq \eta_j^d | R_{j-1} \leq \eta_{j-1}^d] \) for \( 2 \leq j \leq N \) (defined as conditioning level \( k = 2 \)). Then Eq. (3) can be further developed as

\[ \text{Prob}[R_j \leq \eta_j^d | R_{j-1} \leq \eta_{j-1}^d, \ldots, R_1 \leq \eta_1^d] \approx \text{Prob}[R_j \leq \eta_j^d | R_{j-1} \leq \eta_{j-1}^d, R_{j-2} \leq \eta_{j-2}^d] \]

with \( 3 \leq j \leq N \) (defined as level \( k = 3 \)). Equation (4) presents subsequent refinements of local consecutive maxima independence assumption. The latter reflects the statistical dependence between neighbouring local maxima with ever-increasing accuracy. As the MDOF response process \( R(t) \) was assumed ergodic, the probability \( P_k(\lambda) = \text{Prob}[R_j > \eta_j^d | R_{j-1} \leq \eta_{j-1}^d, R_{j-k+1} \leq \eta_{j-k+1}^d] \) for \( j \geq k \) is independent of \( j \) and entirely on the conditioning level \( k \)

\[ P_k(\lambda) \approx \exp(-N \cdot p_k(\lambda)), \quad k \geq 1. \]

that Eq. (5) is similar to the mean up-crossing rate formula for the probability of exceedance\(^4\). There is gradual convergence with respect to the conditioning parameter \( k \)

\[ P = \lim_{k \to \infty} P_k(1); \quad \mu(\lambda) = \lim_{k \to \infty} p_k(\lambda) \]

Note that Eq. (5) \( k = 1 \) case corresponds to the non-exceedance probability relationship with the mean up-crossing rate function

\[ P(\lambda) \approx \exp(-\mu(\lambda) T); \quad \mu(\lambda) = \int_0^{\infty} \xi p_{RR}(\lambda, \xi) d\xi \]

with \( \mu(\lambda) \) being the mean up-crossing rate of the response level \( \lambda \) for the above-described non-dimensional vector \( R(t) \) assembled from the scaled MDOF system response \( \left( X_1, Y_1, Z_1, \ldots \right) \). The mean up-crossing rate from Rice’s formula given in Eq. (7) with \( p_{RR} \) being the joint probability density for \( (R, \dot{R}) \) with \( \dot{R} \) being time derivative \( R(t) \), see\(^4\). Equation (7) relies on the Poisson assumption, that is, up-crossing events of high \( \lambda \) levels (in this paper it is \( \lambda \geq 1 \)) can be assumed to be independent.

In the preceding, the assumption of ergodicity was used. For non-stationary cases, the illustration will be done for offshore engineering and naval architecture areas of engineering applications. Given, for example, the scattered diagram of \( m = 1, \ldots, M \) environmental states, each short-term environmental state has the probability \( q_m \), so that \( \sum_{m=1}^{M} q_m = 1 \). Next, the long-term equation can be introduced

\[ p_k(\lambda) \equiv \sum_{m=1}^{M} p_k(\lambda, m) q_m \]

with \( p_k(\lambda, m) \) being the similar function as in Eq. (6), but corresponding to a short-term environmental state with the number \( m \).

Next, the Naess-Gaidai (NG) approach\(^11\) is presented briefly. The above-described \( p_k(\lambda) \) functions are often tail-regular, i.e. for extreme values of \( \lambda \) approaching extreme level 1. More specifically, for \( \lambda \geq \lambda_0 \), the distribution tail behaves like \( \exp\left\{ -\left( a\lambda + b \right)^c + d \right\} \) with \( a, b, c, d \) being fitted constants for appropriate tail cut-off \( \lambda_0 \) value.

NG method assumes

\[ p_k(\lambda) \approx \exp\left\{ -(a_k\lambda + b_k)^c + d_k \right\}, \lambda \geq \lambda_0 \]

When \( \ln\left( \ln\left( p_k(\lambda) \right) - d_k \right) \) is plotted against \( \ln(ak\lambda + b_k) \), almost linear tail behaviour is often seen. The logarithmic level was optimized by minimizing the error function \( F \) with regard to its four parameters \( a_k, b_k, c_k, p_k, q_k \)

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

with \( \lambda_0 \) being an appropriately selected distribution tail cut-off number for which the breadth of the confidence interval (CI) is acceptable. The optimal parameters \( a_k, b_k, c_k, p_k, q_k \) may be calculated using the sequential quadratic programming (SQP) approach implemented in NAG Numerical Library\(^19\).

**Novel extrapolation method.** Accurate prediction of extreme values is a typical and demanding engineering dependability challenge, particularly when data is insufficient. Consequently, the development of unique, efficient, and precise extrapolation methods is of enormous practical value.

Consider a stationary stochastic process \( X(t) \), either simulated or measured over a given time interval \( 0 \leq t \leq T \), and which is represented as a sum of two independent stationary processes \( X_1(t) \) and \( X_2(t) \), namely

\[ X(t) = X_1(t) + X_2(t) \]

Note that this study seeks to provide a generic technique applicable to predicting extreme values for a broad range of loads and reactions for various vessels and offshore constructions. For the process of interest \( X(t) \), marginal PDF (probability density function) \( p_X \) may be obtained in two ways:
A. By directly extracting $p^A_X$ from the provided 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 using convolution $p^B_X = \text{conv}(p_{X_1}, p_{X_2})$.

Both $p^A_X$ and $p^B_X$ are approximations of the PDF goal, $p^A_X$ is simpler to implement, however B) would yield a more precise estimate of the desired PDF, $p_X$. Convolution permits extrapolation of the directly derived empirical PDFs, $p^A_X$, without supposing any particular extrapolation functional class, such as generalized extreme value distributions (GEV), required to extrapolate distribution tail towards the design low probability level of interest. The majority of current extrapolation techniques, which are frequently used in engineering practice, depending on the assumption of certain extrapolation functional classes, e.g.3–9.

The two independent component representation provided by Eq. (11) is seldom accessible; consequently, one may seek artificial methods to estimate, $p_X$, and $p_{X_i}$, or in the most straightforward situation, discover two identically distributed process components, $X_1(t)$ and $X_2(t)$ with $p_{X_1} = p_{X_2}$. This work focuses on the latter possibility, in which processes, $X_1(t)$ and $X_2(t)$ are dispersed equally. Given the directly estimated distribution, $p_X$, as in option A, the present study’s objective would be to identify the component distribution, $p_{X_i}$, such that

$$p_X = \text{conv}(p_{X_1}, p_{X_2})$$

(12)

therefore, limiting the scope of this research to a deconvolution situation. To illustrate the latter concept of how to estimate the unknown distribution, $p_{X_i}$, reliably and then enhance (or extrapolate) the supplied empirical distribution, $p_X$.

**Discrete convolution.** This section provides a concise overview of commonly held knowledge surrounding the discrete convolution of two vectors. The convolution of two vectors, $u$ and $v$, illustrates the vector component overlap as $v$ glides over $u$. Convolution corresponds algebraically to multiplying polynomials whose coefficients are the elements of $u$ and $v$.

Let $m = \text{length}(u)$ and $n = \text{length}(v)$. Then $w$ is the vector of length $m + n - 1$, whose $k$-th element is

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

(13)

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

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

$$w(2n-1) = u(n) \cdot v(n)$$

(14)

Having obtained $u = v = (u(1), \ldots, u(n))$, one may progressively acquire $w$-components $w(n+1), \ldots, w(2n-1)$, as the index ranges from $n+1$ to $2n-1$. Clearly, the latter would extend vector $w$ into a support domain that is twice as long as the original distribution support domain, i.e. doubling the $p_X$ distribution support length, $(2n-1) \cdot \Delta x \approx 2n \cdot \Delta x = 2X_L$, compared to the original distribution support length $n \cdot \Delta x = X_L$, where $\Delta x$ is the constant length of each discrete distribution bin. In other words, convolution may convect distribution tail attributes farther “downstream” or deeper inside the tail.

In Eq. (14), values of $u$ and $v$ outside the length of the vector are constrained to be zero; however, this is not the case in practice. Consequently, later in this research, vectors $u$ and $v$ will be extrapolated linearly on the logarithmic scale; see the next section, in which $u$ the probability distribution function will represent $u$ and $v$ by $f_X$, and $w$ by $f_{X_L}$.

Note that $w = (w(1), \ldots, w(n))$ is a discrete representation of the empirical target distribution, $p_X$, from “Results” section, with $n$ denoting the length of distribution support $[0, X_L]$. For the sake of simplicity, this work is confined to the situation of one-sided positive valued random variables, i.e. $X \geq 0$.

In the remainder of this study, only deconvolution scenarios, i.e. $u = v$ in Eq. (14), will be examined. According to Eq. (12), the distributions, $p_X$ and, $p_{X_2}$, will correspond to the vectors $w$ and $u$, respectively.

From Eq. (15) is evident that given $w = (w(1), \ldots, w(n))$ one can successively locate unknown components $u = v = (u(1), \ldots, u(n))$, beginning with the first component $u(1) = \sqrt{w(1)}$, then the second $u(2) = \frac{w(2)}{2w(1)}$, and so on until $u(n)$.

As will be addressed further in this work, the authors propose simple linear extrapolation of self-deconvoluted vector $(u(1), \ldots, u(n))$ towards $(u(n+1), \ldots, u(2n-1))$. In other words $p_X$ will have its tail linearly extrapolated in the range $(X_L, 2X_L)$. Note that $p_{X_2}$ can be called the deconvoluted distribution, that in discrete form is represented by the estimated vector $u$. Using Eq. (13) the original vector $w$ will be clearly extended and extrapolated into a support domain that is twice longer than the original distribution support domain, i.e. doubling the $p_X$ distribution support length $(2n-1) \cdot \Delta x \approx 2n \cdot \Delta x = 2X_L$, relative to the original distribution support length $n \cdot \Delta x = X_L$. To smooth out the tail of the original distribution, $p_X$, the authors conducted, $p_X$ tail interpolation, since the tail of the CDF distribution, is often rather regular for large tail values $x$. In particular, the
NG extrapolation approach has been used; see “Methods” section. In addition, various nonlinear extrapolation techniques may be readily included in the suggested method; however, this would involve certain assumptions and biases.

Results
This section provides numerical findings based on the deconvolution extrapolation approach presented. As stated in the Introduction, the deconvolution extrapolation method does not presuppose a particular functional class to extrapolate the distribution tail.

Numerical results for exceedance probability distribution tail. Since in the majority of reliability analysis engineering applications, it is more important to estimate the probability of exceedance, i.e. 1CDF where CDF stands for cumulative density function than the marginal PDF, \( f_X \) will represent the probability of exceedance 1—CDF in this paper, analogous to the marginal probability density function PDF, \( p_X \) in “Results” section. However, the suggested strategy may be applicable to any monotonically declining concave or convex function tail that is sufficiently regular.

The “shorter” version of the original data set was extrapolated and compared to predictions based on the “longer” version to verify the recommended extrapolation strategy. Therefore, this study aims to demonstrate the efficacy of the recommended extrapolation approach by at least a few orders of magnitude.

Based on the preceding reasoning, one may implement an iterative approach, whereas in the marginal PDF, one can utilize 1-CDF and then construct a new artificially smoother CDF by integration. The latter may substantially improve the extrapolation process if distribution tail anomalies are present owing to a lack of underlying data.

Next, the process of discrete convolution, or rather de-convolution (since the objective was to determine the deconvoluted 1—CDF distribution, \( f_{X1} \), given the empirical distribution, \( f_X \) is based on the sequential solution of Eq. (4). Since the deconvoluted values \( u = (u(1), \ldots, u(n)) \) normally follow a monotonically diminishing trend (the same was expected for the empirical parent distribution \( f_X \), it seems that certain final values of the resultant vector \( u \), say, \((u(n - L), \ldots, u(n))\) for some \( L < n \). The latter is an unacceptable numerical inaccuracy since only positive numbers may represent distributions. To address this numerical difficulty, the following scaling method has been implemented12,16. As the pivot value, the lowest positive value \( f_L \) of the specified distribution tail \( f_X \) is used. The scaling is thus only a linear transition along the vertical y-axis of the logarithmic decimal scale.

\[
g_X = \mu \left( \log_{10} (f_X) - \log_{10} (f_L) \right) + \log_{10} (f_L) \tag{15}\]

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 \( \mu \) is conveniently chosen to avoid the occurrence of negative components in the resulting \( f_{X1} \). For both numerical examples studied in this paper, \( \mu = 1/3 \) served that purpose well. Then when \( f_{X1} \) was found, and back convolution \( f_X = \text{conv}(f_{X1}, f_{X1}) \) as in Eq. (13) was done, the inverse scaling with \( \mu^{-1} \) was performed to restore the original scale, with \( f_X \) being extrapolated version of \( f_X \).

Synthetic example. This subsection studies the 3.65-day maximum wind speed process \( X(t) \) simulated during the given time interval \([0, T]\). The underlying normalized non-dimensional stochastic processes \( U(t) \) has been modelled as a stationary Gaussian process with zero mean value and a standard deviation equal to one. Therefore it was assumed that the \( U(t) \) mean zero up-crossing rates satisfies the equality \( v_{up}(0) = 10^{-n}/T \), with \( T = 1 \) year, the latter assumption is common in offshore wind engineering18.

For this numerical example, the “shorter” data record had been chosen to contain \( 10^4 \) data points, which is equivalent to 100 years record, since wind speed maxima process \( X(t) \) has 365/3.65=10^2 data points per year; the full (the latter in this study referred to as “longer”) data set had been chosen to contain \( 10^4 \) data points (thus 100 times longer than the “shorter” data record) are equivalent to 10,000 years.

The underlying wind speed process \( U(t) \) results in 3.65 days maximum analytical CDF distribution \( F_X^{\text{MD}}(x) = \exp \left( -q \exp \left( -\frac{x^2}{T} \right) \right) \) for the wind speed maxima process \( X(t) \), with \( q = 10 \). Note that under the above assumptions, the non-dimensional X-reference value \( x^{100\text{yr}} = 4.80 \), i.e. the wind speed with a return period of 100 years.

Note that artificial horizontal axis shift \( x \to x - x_{\text{shift}} \) with \( x_{\text{shift}} = 1.5 \) has been done in order to operate only with the distribution tail; in other words \( x_{\text{shift}} = x_0 \) from Eq. (7) has been chosen as the cut-on tail marker. The latter makes an analytical estimate for the \( x^{100\text{yr}} = 4.80 - x_{\text{shift}} = 3.3 \).

Figure 2 left presents synthetic wind speed data scaled results, Fig. 2 right presents the final unscaled results of the proposed in this paper technique, namely the “shorter” decimale log scale \( f_X \), extrapolated by deconvolution, along with “longer” data distribution tail and analytic distribution.

Finally, 100 independent numerical Monte Carlo (MC) simulations are typically used to evaluate the performance of models in a robust manner1.
Environmental example. It is generally known that the dynamics of wind speed is a highly nonlinear, multidimensional, and cross-correlated dynamic system that is always difficult to analyze. Moreover, the system dependability approach is crucial for buildings working in any particular region of interest and encountering genuine and sometimes harsh conditions. This section aims to demonstrate the efficacy of the previously mentioned technique by applying a novel method to the Norwegian maximum daily wind cast speeds data set in the vicinity of the Landvik wind station in southern Norway.

As five wind measurement sites, namely Landvik, Kjevik, Lindesnes Fyr, Lista Fyr, and Oksy Fyr, were selected for this research, their respective measured daily peak daily wind cast speeds were established as five environmental system components (dimensions) $X, Y, Z, \ldots$ therefore illustrating a five-dimensional (5D) environmental system.

During the 2010–2020 observation period, the extreme unidimensional response values for each of the five wind measurement sites were determined by doubling the maximum daily highest daily wind cast speed.

Figure 3 presents wind measurement locations according to the Norwegian Meteorological Institute\textsuperscript{21}, a blue circle indicates the area of interest, https://seklima.met.no/.

Figure 2. Synthetic wind speed data. Left: Scaled $f_X$ tail on the decimal log scale for the «shorter» data (cyan squares) «longer» data (blue squares). Right: Unscaled «shorter» decimal log scale $f_X$ tail, raw (red squares) and fitted (solid blue line, along with «longer» data (green line) and analytic (solid red line).
In order to unify all five measured time series $X, Y, Z, \ldots$, the following scaling was performed: making all five responses non-dimensional and having the same failure limit equal to 1. Next, all local maxima from five measured time series were merged into one single time series by keeping them in time non-decreasing order: $\vec{R} = (\max\{X_1, Y_1, Z_1, \ldots\}, \ldots, \max\{X_N, Y_N, Z_N, \ldots\})$ with the whole vector $\vec{R}$ being sorted according to non-decreasing times of occurrence of these local maxima.

Figure 5 presents an example of a non-dimensional assembled vector $\vec{R}$, which is comprised of assembled local maxima of raw daily greatest highest daily wind cast speed data. The failure probability distribution tail was extrapolated towards a return time of 100 years. Note that vector $\vec{R}$ has no physical significance on its own, since it is composed of wholly distinct response components. Index $j$ is only an index of local maxima encountered in a non-decreasing time sequence. Index $j$ is only an index of local maxima encountered in a non-decreasing time sequence. The «shorter» data record was constructed by extracting every tenth data point from the «longer» data record of wind speed.

Figure 6 on the left shows the «shorter» data record, $f_{X_1}$ tail, derived by deconvolution as in Eq. (13), and then linearly extrapolated in the terminal tail section to encompass the $X_1$ range corresponding to the «longer» data record. Figure 6 on the right displays the unscaled final outcomes of the approach suggested in this research, namely the «shorter» decimal log scale, $f_{X_1}$ tail, extrapolated by deconvolution, together with the «longer» data distribution tail and NG extrapolation.

It is seen that NG prediction matches the deconvolution method pretty well. Figure 6 on the right demonstrates that the suggested technique works pretty well, is based on the "shorter" data set and produces distributions that are relatively similar to those derived from the "longer" data set. Figure 6 on the right illustrates extrapolation according to Eq. (9) as well as a unique deconvolution approach from Sect. 0 towards a failure state with a 100-year return time, which is 1 and a bit beyond; $\lambda = 0.05$ was chosen as the cut-on value. The dashed lines represent the extended 95% confidence interval according to Eq. (10). According to Eq. (5), $p(\lambda)$ is exactly proportional to the goal failure probability $1-P(1)$. In accordance with Eq. (5), the system failure probability $1 - P \approx 1 - P_k(1)$ may be computed. Note that in Eq. (5), $N$ is the total number of local maxima in the response vector $\vec{R}$. Due to convergence with respect to $k$ (see Eq. 6), it was determined that $k = 5$ was an adequate conditioning value (6).

Note that, although innovative, the above-described technique has the evident benefit of using existing measured data sets rather effectively, owing to its capacity to handle system multidimensionality and conduct correct extrapolation using relatively small data sets.

---

**Figure 4.** Daily largest highest daily wind cast speed at Landvik location during the years 2010 – 2020, Norwegian Meteorological Institute.

**Figure 5.** Left: unscaled raw daily largest highest daily wind cast speed data [m/sec], Right: scaled non-dimensional assembled 5D vector $\vec{R}$. 
Conclusions

Classical dependability approaches for time series cannot properly handle systems with high dimensionality and cross-correlation between various system responses. The presented approach’s primary advantage is its ability to examine the dependability of high-dimensional nonlinear dynamic systems.

This article examined the Norwegian greatest daily wind cast speeds obtained at two distinct wind speed measurement locations near the Landvik wind station from 2012 to 2022.

The techniques presented in this study have been previously confirmed by application to a variety of simulation models, albeit only for one-dimensional system responses, and highly accurate predictions have been achieved in general. This research attempted to establish a multidimensional dependability approach that was simple, robust, and applicable to a variety of applications.

It is possible to analyze both observed and numerically generated time series responses. It is shown that the suggested strategy generated an acceptable confidence interval. Consequently, the proposed technique may become useful for various nonlinear dynamic system dependability investigations.

In the case of Monte Carlo-type numerical simulation, unlike existing dependability approaches, the novel method does not need to restart numerical simulation each time the system fails. It is possible to analyze both observed and numerically generated time series responses.

Finally, the recommended technique applies to a vast array of technical applications. The offered example of naval architecture does not restrict the scope of application for the proposed technique.

Data availability

Meteorological data used in this paper is publicly available at the Norwegian Meteorological Institute, https://seklima.met.no/.

Received: 27 July 2022; Accepted: 9 November 2022
Published online: 15 November 2022

References

1. Bokde, N., Yaseen, Z. & Andersen, G. ForecastTB—an R package as a test-bench for time series forecasting—application of wind speed and solar radiation modeling. Energies 13(10), 2578. https://doi.org/10.3390/en13102578 (2020).
2. Ditlevsen, O. & Madsen, H. O. Structural Reliability Methods (Wiley, 1996).
3. 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).
4. Gaidai, O., Xing, Y. & Xu, X. COVID-19 epidemic forecast in USA East coast by novel reliability approach. Res. Sq. https://doi.org/10.21203/rs.3.rs-1573862/v1 (2022).
5. Gaidai, O., Storhaug, G. & Naess, A. Extreme large cargo ship panel stresses by bivariate ACER2D method. Ocean Eng. 123, 432–439 (2016).
6. Gaidai, O. et al. Bivariate statistics of wind farm support vessel motions while docking. Ships Offshore Struct. 16(2), 135–143 (2020).
7. Gaidai, O., Storhaug, G. & Naess, A. Statistics of extreme hydro elastic response for large ships. Mar. Struct. 61, 142–154 (2018).
8. Gaidai, O., Storhaug, G. & Naess, A. Extreme value statistics of whipping response for large ships. In PRADS Proceedings. Practical Design of Ships and Other Floating Structures, vol 2, 1210–1221 (2010).
9. Gaidai, O., Storhaug, G. & Naess, A. Extreme value statistics of ship rolling. In PRADS Proceedings. Practical Design of Ships and Other Floating Structures, vol 2, 457–466 (2010).
10. 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).
11. Naess, A. & Gaidai, O. Estimation of extreme values from sampled time series. Struct. Saf. 31(4), 325–334 (2009).
12. Gaidai, O., Yan, P. & Xing, Y. A novel method for prediction of extreme wind speeds across parts of Southern Norway. *Front. Environ. Sci.* https://doi.org/10.3389/fenvs.2022.997216 (2022).

13. 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).

14. 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).

15. 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).

16. Gaidai, O., Yan, P. & Xing, Y. Prediction of extreme cargo ship panel stresses by using deconvolution. *Front. Mech. Eng.* https://doi.org/10.3389/fmech.2022.992177 (2022).

17. Naess, A., Leira, B. J. & Batsevych, O. Reliability analysis of large structural systems. *Probab. Eng. Mech.* 28, 164–168 (2012).

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

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

20. Naess, A. & Moan, T. *Stochastic Dynamics of Marine Structures* (Cambridge University Press, 2013).

21. Norwegian Meteorological Institute https://seklima.met.no/.

**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 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.

© The Author(s) 2022