



Estimating space–time wave statistics using a sequential sampling method and Gaussian process regression

Tianning Tang^{*}, Thomas A.A. Adcock

Department of Engineering Science, University of Oxford, Oxford, UK

ARTICLE INFO

Keywords:

Rogue wave
Wave statistics
Data-driven methods

ABSTRACT

The traditional Monte Carlo sampling of waves requires generating tens of thousands of random waves to achieve stable statistics in the tail of the distribution, which is computationally costly for numerical simulations and time-consuming and often impractical for experiments. To improve the sampling efficiency, we present a sequential sampling method, which predicts the wave statistics based on the nonlinear response of deterministic wave groups. This data-driven method starts with parameterising the linear random wave field with a series of Gaussian wave groups. The nonlinear system response from the water wave system is then approximated by observations through the nonlinear evolution of wave groups with carefully designed initial conditions. A sequential sampling method is introduced during this sampling procedure, which determines the next best sampling point based on the previous observations and the distance metric. We examine the performance of the proposed method with Monte Carlo simulation of random wave fields as well as a grid search sampling strategy. The results show the proposed method can achieve computational cost savings over several orders of magnitude for numerical simulations. The results also suggest such a sequential sampling method can be used to determine the test matrix of the wave experiments with better coverage over the parameter space.

1. Introduction

In ocean engineering practice, a fast and accurate prediction tool for wave statistics is important to the design of offshore infrastructure. For example, the design of large offshore platforms requires a sufficient gap between the deck and the mean water level to allow waves to pass underneath. The design of such large offshore platforms also needs to consider the wave statistics over an area in the ocean rather than just point statistics as these may under-predict the maximum crest height over 20% for certain conditions (Forristall, 2006).

Although modelling the magnitude of extreme waves over an area is critical in ocean engineering, currently available methods still require a trade off between speed, cost and accuracy. Various statistical models have been proposed for predicting space–time wave statistics. These models offer the fastest and almost instantaneous predictions with different orders of accuracy (see details in Forristall, 2006). Validation of the available models has been carried out against experimental and field data (see details in Fedele et al., 2013; Fedele, 2012; Forristall, 2011; Benetazzo et al., 2015; Fedele et al., 2017; Forristall, 2015). However, the application of these theoretical models are still restricted by different assumptions on the sea states during the model derivation (Karmadakis et al., 2020). For example, most of these theoretical models are limited to uni-modal sea states, whereas wave statistics

in multi-modal sea states are still not fully understood (Toffoli et al., 2008; Bitner-Gregersen and Toffoli, 2014; Toffoli et al., 2011), although it is known that other physics is involved (McAllister et al., 2017). As such, we are still waiting for a unified space–time wave statistics model, which is suitable for multi-modal sea states with a wide range of characteristics.

To obtain more accurate space–time wave statistics for certain combinations of sea state parameters, experiments and numerical simulations can be performed. Carefully designed experiments can help investigate the effect of higher order nonlinear physics, which leads to more rogue waves than expected by linear theory (Onorato et al., 2004; Latheef and Swan, 2013; Onorato et al., 2009b). Performing experiments is however costly and time consuming, as the waves are generated with a random phase. During such experiments, most waves are well captured by basic models with only the occasional wave displaying the unusual physics which is of interest. Apart from experiments, numerical simulations can also estimate the wave statistics by solving the governing equations. Klahn et al. applied fully nonlinear simulations for directionally spread random waves to explore the statistical properties of surface elevation, velocities, accelerations and forces with Monte Carlo sampling (Klahn et al., 2021a,b) (see

^{*} Corresponding author.

E-mail address: tianning.tang@eng.ox.ac.uk (T. Tang).

also other relevant work Toffoli et al., 2009, 2010b; Xiao et al., 2013; Tang and Adcock, 2021). These numerical simulations are important for understanding the statistical behaviour of nonlinear water wave systems, but are usually constrained by the available computational resources, especially when Monte Carlo sampling is considered. For example, a numerical model with a full wave breaking model would usually be too computationally intensive to perform random time series simulations (Wang et al., 2016; Lubin and Glockner, 2015). Deterministic wave groups are used as an computational cheap alternative, as the evolution of these wave groups captures the nonlinear physics when compared to the natural occurring extreme waves (Adcock et al., 2012). The shape of these wave groups can be linked back to the averaged shape of the extreme waves through the NewWave or quasi-determinism theory (Lindgren, 1970; Boccotti, 1983; Tromans et al., 1991).

Apart from the statistical models and direct numerical simulations of random time series, recent studies applied data-centric methods for predicting wave statistics (Mohamad et al., 2016; Tang and Adcock, 2022) (see also other data-centric studies in ocean engineering applications Farazmand and Sapsis, 2017; Dematteis et al., 2018). In these studies, a probabilistic decomposition framework is adopted to separate the nonlinear extreme responses from the linear ones. These nonlinear responses can be estimated from the spatial-temporal evolution of wave groups instead of nonlinear random time series simulations, which allows a better trade off between the computational effort and the accuracy of the model. However, estimating nonlinear responses from wave group simulations is still computationally intensive as the spatial-temporal response needs to be collected for all the combinations of the unstable wave group shapes, which counts up to over 18,000 nonlinear wave group simulations in previous studies (Tang and Adcock, 2022). Although such a data-centric framework can potentially reveal new physics, the initial computational cost of sampling the nonlinear spatial-temporal response of deterministic events still limits its applications, especially for small scale scientific research purposes. Additionally, the high initial computation cost also limits the versatility of such a data-centric framework for incorporating new physics with advanced numerical model.

In this study, we examine a sequential design sampling strategy, which aims at reducing the initial computational cost for the probabilistic decomposition framework and providing better statistical insights into the nonlinear spatial responses of deterministic wave groups. We followed the sequential sampling method from the previous studies (Mohamad and Sapsis, 2018; Sapsis, 2020), and modified this method for sampling the nonlinear response from deterministic wave groups with proper nonlinear spatial profile corrections. The proposed new method starts with parameterising the linear random wave fields with superposition of Gaussian wave groups, as suggested from previous studies (Farazmand and Sapsis, 2017; Tang and Adcock, 2022). We introduced the stable and unstable parameter-to-observation map to connect the observed wave group parameters with the full spatial-temporal response of wave groups. To estimate the full spatial response between observation points for the unstable parameter-to-observation map, we propose to approximate the nonlinear spatial profile of wave groups with linear envelope corrections. The accuracy of this approximation is examined with nonlinear wave group simulations. The sequential sampling method is then applied to the unstable parameter-to-observation with metric functions. The proposed method reduces the total amount of nonlinear numerical simulations of wave groups significantly. We finally combine the stable and unstable parameter-to-observation map to estimate the space-time wave statistics for directional spread deep water wave systems.

We structure this paper as follows. First, we give a general description of the framework for predicting wave statistics in Sections 2.1 and 2.2. In this framework, we first approximate random wave fields with wave group parameters, which is introduced in detail in Section 2.3. We then use a parameter-to-observation map as a convenient proxy to

connect the input wave group parameters with the nonlinear response from the water wave system in Section 2.4. The sequential sampling method is introduced in Section 2.4.2 for estimating the parameter-to-observation map with minimal runs of wave group simulations. We introduce the numerical details used in this study in Section 3. Finally we compare the estimated space-time wave statistics against nonlinear Monte Carlo simulations with random waves in Section 4.

2. Method

2.1. Summary of the framework

In this study, we utilise a sequential sampling framework to predict the crest statistics for directionally spread water waves in deep water. Traditionally, the estimation of nonlinear wave statistics requires nonlinear Monte-Carlo simulations of random wave time series, which are computationally intensive. However, the proposed framework allows accurate estimation of nonlinear wave statistics with a few carefully designed wave group simulations, which significantly reduces the computational cost and brings a new parameterisation method for random wave fields. We summarise the framework with the following steps:

Step 1. Parameterisation of random wave fields. For a directionally spread water wave system, many sea-state parameters can affect nonlinear physics. For a system with uni-modal sea states, these parameters can be wave steepness, spectral parameters of the initial wave spectrum, the water depth profile and directional spreading functions (Onorato et al., 2009b,a; Toffoli et al., 2010a). For a system with multi-modal sea states, in addition to crossing angles and number of modals, the parameters for uni-modal sea states need to be considered for each modal (Toffoli et al., 2008, 2011).

In Section 2.3, we introduce a parameterisation algorithm to reduce the dimensions of such a complex system. This parameterisation algorithm approximates the spatial envelope profile of random waves with superposition of deterministic wave group envelopes. This parameterisation effectively describes a random wave field with location, amplitude and spatial length scale of a series of elementary wave groups, which greatly reduces the number of input parameters.

Step 2. Sequential experimental design. The accurate prediction of nonlinear wave statistics requires parameter-to-observation maps to connect the wave field characteristics with the final probability distribution. We introduce the parameter-to-observation maps in Section 2.3 with additional (relative to the literature) parameters needed for directionally spread waves (see Section 2.4.1). Sequential experimental design is then applied to sample the parameter-to-observation maps with Gaussian process regression as the learning algorithm (see Section 2.4.2). Two metric functions are used for fast and accurate converging of the predictions with fewer samples required.

2.2. Problem setup

For a general nonlinear system \mathcal{N} , space and temporal response of a state variable u is governed by a nonlinear stochastic partial differential equation:

$$\frac{\partial u(x, y, t)}{\partial t} = \mathcal{N}(x, y, t, u; \kappa(\xi)), \quad \xi \in \Omega \quad (1)$$

where x, y are the horizontal spatial variables, t is the time, ξ is the sample space parameter and Ω is the sample space of the nonlinear system. We parameterise the sample space parameter with a new random variable $\kappa : \Omega \rightarrow U$ with a known distribution \hat{p}_κ .

For a fixed sample space parameter $\xi \in \Omega$ the state variable u of our nonlinear system is a deterministic function in space and time. In this study, we aim to predict the probability density function of a quantity of interest q , which can be described by a continuous nonlinear parameter-to-observation map $\hat{T}(\kappa)$:

$$q = \hat{T}(\kappa) \triangleq F(u), \quad (2)$$

where the quantity of interest q itself is an arbitrary function F of the state variable u of a nonlinear system.

The true statistics of the random quantity of interest q induced by parameter-to-observation map $\hat{T}(\kappa)$ are:

$$\hat{\rho}(s) = \frac{d}{ds} \hat{F}(s) = \frac{d}{ds} \mathbb{P}(\hat{T}(\kappa) \leq s) = \frac{d}{ds} \int_{\mathbb{A}(s)} \rho_{\kappa}(\kappa) d\kappa, \quad (3)$$

where $\mathbb{A} = \{\kappa \in U : \hat{T}(\kappa) \leq s\}$, s is the threshold parameter, $\hat{F}(s)$ is the cumulative distribution function of the quantity of interest q , and \mathbb{P} denotes the probability function. However, in this study, we examine a nonlinear system with an unknown parameter-to-observation map \hat{T} , which, in general, is expensive to evaluate. For the two-dimensional water wave system used here evaluating the true system is computationally demanding but feasible. However, for other ocean engineering problems deriving the full parameter-to-observation will not be possible. We describe our water wave system with the state variable u to be the slowly varying complex envelope, and the magnitude of which can be calculated from the surface elevation (η) as:

$$|u| = \sqrt{\eta^2 + \eta_H^2}, \quad (4)$$

where η_H is the Hilbert transform of the surface elevation. For such a nonlinear wave system in deep water x is the horizontal spatial variable in the mean wave direction, y is the horizontal spatial variable in the transverse direction. In this study, we use the Modified Nonlinear Schrodinger Equation (MNLS) (Trulsen et al., 2000) to model the state variable u . We also use the direct nonlinear Monte-Carlo random wave simulations as the ground truth to verify our model (see details in Section 3).

For a nonlinear system, the state variable u depends on sample space parameters ξ , which are, for a two dimensional water wave system, the sea state parameters (e.g. wave spectral parameters, significant wave height, directional spreading angle etc.). Because of the complexity of using sea state parameters to characterise the random wave fields, in this study, we decompose the random wave fields into individual wave groups following Farazmand and Sapsis (2017). Hence, the underlying wave spectrum and other sea state parameters can be generalised with three wave group parameters: the amplitude of the wave group A , length scale parameter in x and y direction L_x and L_y (Tang and Adcock, 2022). The details of this parameterisation process of random wave fields are presented in Section 2.3.

For engineering purposes, we aim to predict the probability density function of a normalised envelope height at a single Eulerian point:

$$q_p = |u|/H_s, \quad (5)$$

where H_s is the significant wave height of the sea state. In this study, we only examine the free wave component of the wave field as the contributions from bound harmonics can be estimated from the free wave profiles with little computational cost (Dalzell, 1999). We are also interested in the wave statistics over an area, as this is more relevant to large offshore infrastructure. In this study, we consider the new quantity of interest to be the spatial maximum envelope as:

$$q_a = \max_{x,y \in S} |u(x,y)|/H_s, \quad (6)$$

where S is the spatial sampling domain ($x \in [0, \ell_x]$, $y \in [0, \ell_y]$), and ℓ_x and ℓ_y are the length of the sampling domain in x and y directions respectively.

2.3. Parameterisation of random wave fields

For a two dimensional water wave system, there are many sea-state parameters, which control the response of the system. Directly modelling such a high dimensional system is challenging. Following the idea of Farazmand and Sapsis (2017) and Tang and Adcock (2022), we

parameterise the random wave field with a superposition of Gaussian elementary wave groups as:

$$G(x,y) = \sum_{n=1}^N g_n(x,y), \quad (7)$$

where $G(x,y)$ is the superpositioned wave field, and N is the number of wave groups with a Gaussian profile $g_n(x,y)$ as:

$$g_n(x,y) = A_n \exp \left[-\frac{(x-x_n^c)^2}{(L_n^x)^2} \right] \exp \left[-\frac{(y-y_n^c)^2}{(L_n^y)^2} \right], \quad (8)$$

where A_n is the envelope height of n th wave group with the wave group centre at (x_n^c, y_n^c) and the L_n^x and L_n^y are the two length scale parameters measuring the width of the group in the longitudinal direction x and in the transverse direction y respectively.

With this parameterisation method, we effectively reduce the number of input parameters from various sea-state parameters to the distribution of three wave group parameters $\hat{\rho}_{A,L_x,L_y}$, which can reflect all the changes in the sea state parameters. For example, the ratio between L_x and L_y is an alternative measure of the directional spread of the wave field. This ratio tends to zero in the unidirectional limit. This parameterisation of random wave fields greatly reduces the complexity of the parameter-to-observation map introduced in Section 2.4.

2.4. Parameter-to-observation map

For a two dimensional water wave system, although it is possible to follow the Eq. (2) directly and use a single parameter-to-observation map to connect the wave group parameters with the overall probability distribution. However, a wide parameter space is required to make accurate predictions. This dramatically increases the difficulty in accurately predicting the parameter-to-observation map. In this study, following the idea of probabilistic decomposition method (Mohamad and Sapsis, 2016), we proposed to separate the parameter space into two regions based on the instability boundary (see details of this separation in Appendix). Two different parameter-to-observation maps are used to predict the total probability density function:

$$\hat{\rho}(s) = \frac{d}{ds} \mathbb{P}(\hat{T}_l(\kappa) \leq s)(1-p_r) + \frac{d}{ds} \mathbb{P}(\hat{T}_r(\kappa) \leq s)p_r, \quad (9)$$

where $\hat{T}_l(\kappa)$ is the parameter-to-observation map for the stable region and $\hat{T}_r(\kappa)$ is the parameter-to-observation map for the unstable region. We connect these two parameter-to-observation maps with the probability of rare events p_r , which can be evaluated as:

$$p_r = \iiint_{R_e} \hat{\rho}_{A,L_x,L_y}(A, L_x, L_y | R_e) dA dL_x dL_y, \quad (10)$$

where R_e specifies the instability region. For space-time wave statistics, we follow Tang and Adcock (2022) and use a varied $p_r(q)$ value to be the cumulative distribution of the envelope amplitude within the instability region as:

$$p_{r,v}(q) = \int_0^q \hat{\rho}_A(A | R_e) dA. \quad (11)$$

For the stable parameter-to-observation map $\hat{T}_l(\kappa)$, the wave behaviour is well captured by linear physics and we use direct numerical simulations of the random wave field with linear theory. This inexpensive simulation also determines the conditional statistics of rare events $\hat{\rho}_{A,L_x,L_y}$ with the parameterisation process in Section 2.3. The details of the linear numerical simulation are given in Section 3.1

We use the nonlinear simulation of the elementary events within the dynamical system to obtain the parameter-to-observation map $\hat{T}_r(\kappa)$ for ‘unstable’ cases. For a two dimensional wave system, the elementary events are wave groups (Adcock and Yan, 2010). In this study, we follow Tang and Adcock (2022) and consider a Gaussian profiled wave group at linear focus as:

$$u_0(x,y) = A \exp \left[-\frac{x^2}{L_x^2} - \frac{y^2}{L_y^2} \right], \quad (12)$$

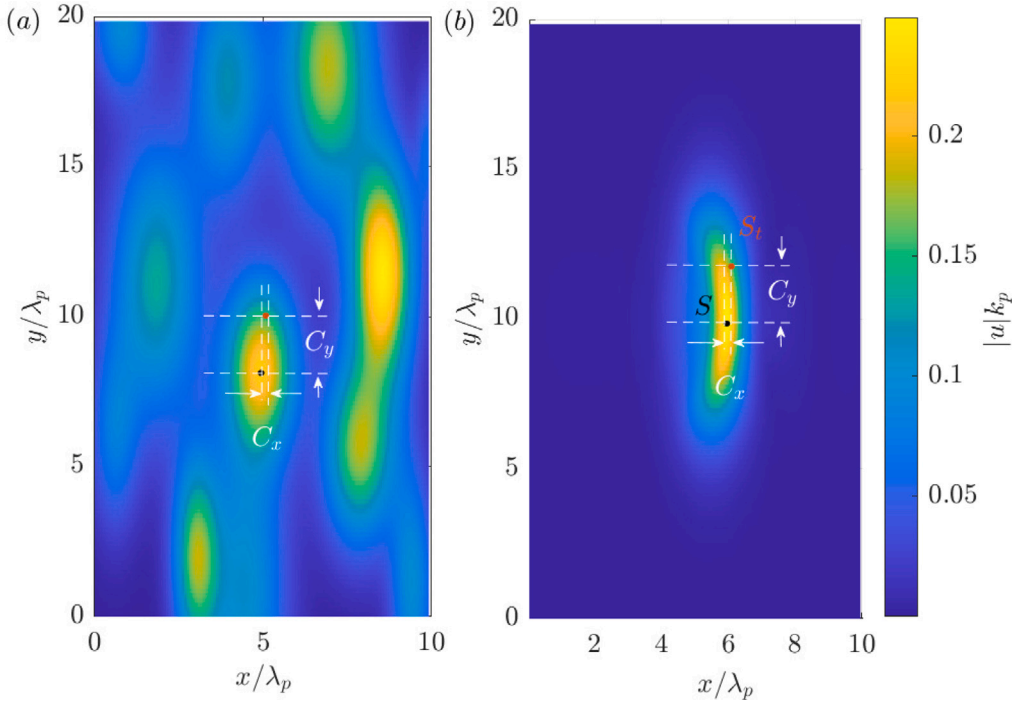


Fig. 1. Illustration of the distance parameters C_x and C_y for obtaining the sampling domain S and S_i for point measurements. (a) : a sample reconstructed wave envelope field with decomposed elementary Gaussian wave groups $g_n(x, y)$ and the identified primary wave group has a parameter of $Ak_p = 0.21$, $L_x k_p = 6.3$ and $L_y k_p = 22$. (b) : the corresponding nonlinear wave group simulations with a linear focus profile same as the identified primary wave group in (a). Envelope height is normalised by peak wavenumber k_p and spatial positions x and y are normalised by peak wavelength λ_p . The initial Gaussian spectrum has a spectral bandwidth in x and y direction of $\sigma_x = \sigma_y = 0.1$. The steepness of the random wave field is $H_s k_p = 0.25$.

where A is the envelope height, L_x and L_y are the two length scale parameters characterising the shape of the group in the x and y directions respectively.

To obtain the unstable parameter-to-observation map $\hat{T}_r(\kappa)$, we track and collect the nonlinear response of the water wave system during the evolution of deterministic wave groups. As we are primarily interested in the instability triggered during the nonlinear evolution of wave groups, we focus on the nonlinear changes of wave envelope profile close to the nonlinear focus point. In practice, following Tang and Adcock (2022), we refer to the spatial wave group profile within the time of $(t_0 \in [-2T_p, 2T_p])$, after centring the nonlinear focus at $t = 0$. The unstable parameter-to-observation map $\hat{T}_r(\kappa)$ are calculated based on the nonlinear envelope profile at every time step during this period, and can be obtained at any instantaneous time t_0 as:

$$\hat{T}_r(A, L_x, L_y) = \max_{(x,y) \in S} |u(x, y, t_0)| / H_s \times C_d, \quad (13)$$

where $|u(x, y, t_0)|$ can be obtained through nonlinear wave group simulations detailed in Section 3.2.1, S is the sampling domain, and the C_d is a distance correction coefficient for point measurements, which is explained in detail in Section 2.4.1.

2.4.1. Distance correction

For unidirectional waves, Eulerian point measurements are guaranteed to capture the envelope peak. However, real ocean waves are directionally spread, and such wave fields, point measurements are likely to miss the envelope peak (Forristall, 2006). This behaviour leads to a dramatic overestimation of wave statistics with the unstable parameter-to-observation map $\hat{T}_r(\kappa)$ if only the spatial maximum responses from the deterministic events are considered. Therefore we apply a ‘correction’ to account for this in our analysis. We demonstrate this behaviour in Fig. 1, where the sampling point misses the ‘true’ envelope peak.

We ‘correct’ our model in the following way. For Eulerian point measurements, as only the envelope amplitude at the centre point of

the computational domain is collected at any time instance, the primary wave group is defined as the wave group with its centre point closest to the sampling point. We denote the spatial distance between these two points as C_x and C_y in the x and y direction respectively. For space-time wave statistics, we are interested in the spatial maximum envelope height within the sampling domain S : $x \in [-\ell_x/2, \ell_x/2]$ and $y \in [-\ell_y/2, \ell_y/2]$, where ℓ_x and ℓ_y are the sampling length in x and y direction respectively. The primary wave group is defined as the wave group with its centre point closest to the location of the spatial maximum envelope height within the sampling domain.

Determining the impact from the distance parameter is straightforward if the spatial information of the deterministic events is available during the sampling of the parameter-to-observation map. For each unstable wave group, the associated distance parameters C_x and C_y are obtainable based on the wave group centre positions $C_x = x^c$, $C_y = y^c$. To account for the distance effect, we move the sampling domain for wave group simulations away from the centre of the domain by C_x and C_y . The new sampling domain S_i : $x \in [-\ell_x/2 + C_x, \ell_x/2 + C_x]$ and $y \in [-\ell_y/2 + C_y, \ell_y/2 + C_y]$ compensates for the error induced by the distance parameter. We demonstrate this shift of the sampling domain in Fig. 1.

The distance parameter correction can be calculated as the ratio of the spatial maximum envelope amplitude between the shifted sampling domain S_i and the perfectly centred sampling domain S (i.e. the distance between the sampling domain and centre of the wave group is zero):

$$C_{d,target} = \text{mean}_{t \in T} \frac{\max_{(x,y) \in S_i} |u(x, y, t)|}{\max_{(x,y) \in S} |u(x, y, t)|}, \quad (14)$$

where T is the time response of range used for calculating the unstable parameter-to-observation map $([-2T_p, 2T_p])$.

However, as the sequential sampling method is applied to the unstable parameter-to-observation map, the spatial information of the wave groups with shifted sampling domain $(|u(x, y)|, (x, y) \in S_i)$ are only

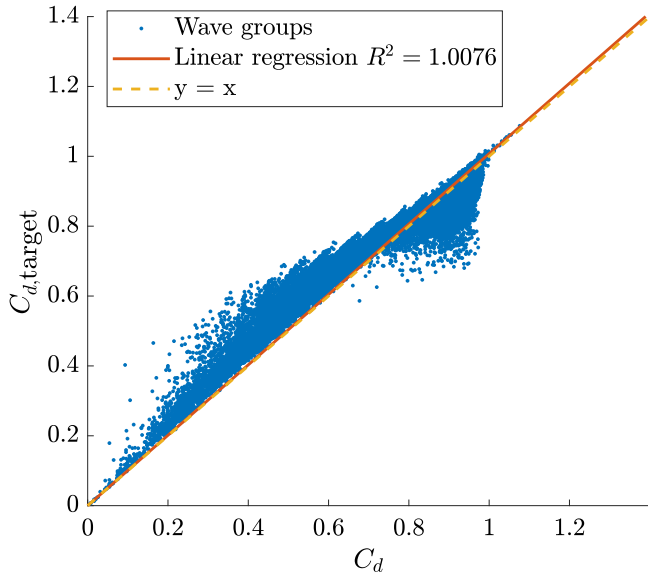


Fig. 2. Comparison between the target distance correction coefficient $C_{d,target}$ based on nonlinear simulation against linear approximation value C_d for all the wave groups within the instability region. The red line represents the linear regression results and yellow line indicates the perfect approximation line $y = x$.

available at the observation points. Hence, the correction coefficient based on nonlinear simulation cannot be determined directly within the sequential sampling framework. In this study, we propose a linear approximated distance parameter correction coefficient for each wave group within the instability region:

$$C_d = \text{mean}_{t \in T} \frac{\max_{(x,y) \in S_t} g_n(x, y, t)}{\max_{(x,y) \in S} g_n(x, y, t)}. \quad (15)$$

In Fig. 2, we demonstrate the accuracy of this linear approximation by plotting the target distance correction coefficient $C_{d,target}$ based on nonlinear simulation against our proposed linear approximation value C_d for all the wave groups within the instability region. For most cases, the linear approximation predicts the distance correction coefficient accurately. For relatively large distances (i.e. small distance correction coefficient), the linear approximation tends to slightly under predict the distance effect, and for small distances, the linear approximation tends to slightly over predict the distance effect. However, for the majority of the wave groups within the instability region, the general distance effect is well captured by our linear approximated coefficient C_d . This also agrees well with the linear regression results in Fig. 2, where the coefficient of determination of linear fitting is quite close to 1. The coefficient of determination R^2 is given as:

$$R^2 = 1 - \frac{\sum_{m=1}^n (C_{d,target,m} - C_{d,m})^2}{\sum_{m=1}^n (C_{d,target,m})^2}, \quad (16)$$

where n is the total number of wave groups within the instability region.

We further explore the over prediction from the linear approximation for relatively small distances and the under prediction for relative large distances in Fig. 3. In Fig. 3(a), we categorise the ratio of the distance correction factor between the linear approximation and nonlinear simulation by the wave steepness at focus. The ratio is close to unity for low wave steepness, but the ratio is scattered over a wide range for large wave group steepness. This indicates that the error introduced by the linear approximation is closely related to the nonlinear modifications on the shape of wave groups (see the comparison of wave group profile at focus in Fig. 1).

We further categorise the distance correction value for all the unstable wave groups in Case 1 by the distance in y direction in Fig. 3(b). The linear simulation tends to over predict the distance correction factor for small y distances and tends to under prediction the distance correction factor at large distances, which agrees well with the previous results in Fig. 2. This trend is also true for individual wave groups in Fig. 3(c), where the distance correction factor is presented for a typical wave group profile over different distances in y direction.

2.4.2. Sequential experimental design for the unstable parameter to observation map

Estimating the unstable parameter-to-observation map $\hat{T}_r(\kappa)$ with nonlinear simulations is computationally intensive. For example, in this study, we used the densely sampled $\hat{T}_r(\kappa)$ as the reference case. To estimate such a high dimensional observation map with a grid sampling method produces a total of 27900 cases of wave groups with different combinations of parameters. This is simulated with the MNLS (see details in Section 3.2.1). The high computational cost limits the utilisation of other computational intensive nonlinear models with extra physics (i.e. wave breaking, wave-current interaction). In this study, we propose to use the sequential experimental design method to greatly reduce the number of observations required to obtain the unstable parameter-to-observation map $\hat{T}_r(\kappa)$ in Section 2.4.2.

We recall that based on Eq. (3), the probability density function associated with the unstable parameter-to-observation map ($\hat{\rho}_r(q)$) can be rewritten as:

$$\hat{\rho}_r(q) = \frac{d}{dq} \int_{\mathbb{A}(q)} \hat{\rho}_\kappa(\kappa) d\kappa = \frac{d}{dq} \int_{\mathbb{A}(q)} \rho_{A,L_x,L_y}(A, L_x, L_y) dA dL_x dL_y, \quad (17)$$

where $\mathbb{A} = \{A, L_x, L_y \in R_e : \hat{T}_r(A, L_x, L_y) \leq q\}$.

The key objective of the sequential experimental design method is to estimate the probability density function $\hat{\rho}_r(q)$ without densely sampling the unstable parameter-to-observation map $\hat{T}_r(A, L_x, L_y)$. To achieve this, instead of grid search within the instability region R_e , we will determine the next ‘best’ sampling parameter sets ($\kappa^* : A^*, L_x^*, L_y^*$) based on a learning algorithm prediction from the previous dataset:

$$D_{n-1} = \{(A_i, L_{x,i}, L_{y,i}, \hat{T}_r(A_i, L_{x,i}, L_{y,i}))\}_{i=1}^{n-1}, \quad (18)$$

where D_{n-1} is the dataset from previous $n - 1$ observations at $(A_i, L_{x,i}, L_{y,i})$ and the corresponding parameter-to-observation map $\hat{T}_r(A_i, L_{x,i}, L_{y,i})$.

In this study, Gaussian process regression (Seeger, 2004) is used as the learning algorithm for estimating the parameter-to-observation map $\hat{T}_r(\kappa)$. We use Gaussian process regression in this study as this machine learning model can also return the confidence interval of the predictions, which is critical for determining the next best sampling point during the later calculations. Based on the previous dataset D_{n-1} , the Gaussian process regression predicts the probability distribution:

$$\rho_{r,n-1}(q) = \frac{d}{dq} \int_{\mathbb{A}_{n-1}(q)} \rho_{A,L_x,L_y}(A, L_x, L_y) dA dL_x dL_y, \quad (19)$$

where $\mathbb{A}_{n-1} = \{A, L_x, L_y \in R_e : \hat{T}_{r,n-1}(A, L_x, L_y) \leq q\}$, and the $\hat{T}_{r,n-1}$ is the unstable parameter-to-observation map predicted based on the previous dataset D_{n-1} .

We now consider the current prediction step with extra observation sampling parameter set ($\kappa^* : A^*, L_x^*, L_y^*$). The learning algorithm predicts the confidence interval of the probability distribution function as:

$$\hat{\rho}_n^\pm(q; A^*, L_x^*, L_y^*) = \frac{d}{dq} \int_{\mathbb{A}_n^\pm(q; \kappa^*)} \rho_{A,L_x,L_y}(A, L_x, L_y) dA dL_x dL_y, \quad (20)$$

where $\mathbb{A}_n^\pm(q; \kappa^*) = \{\kappa \in R_e : \hat{T}_{r,n}(\kappa; \kappa^*) \pm \alpha \tilde{\sigma}_n(\kappa; \kappa^*) \leq q\}$. The $\tilde{\sigma}_n$ is the confidence interval based on the standard deviation from the

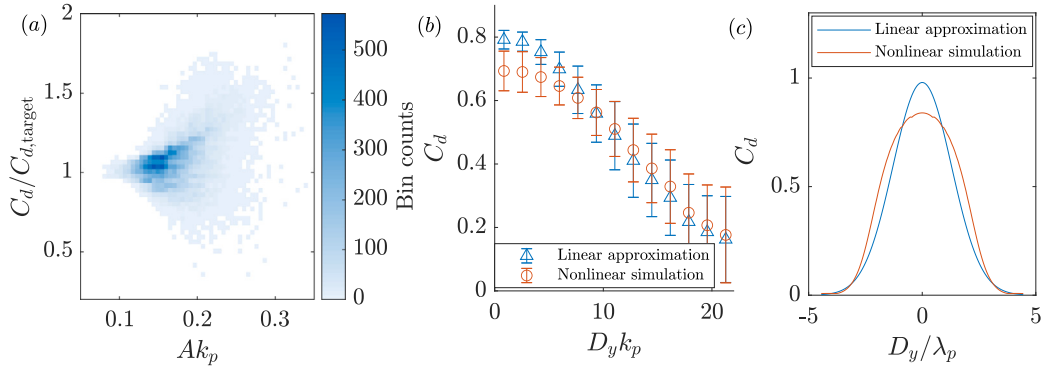


Fig. 3. (a) Binned scatter plot for the ratio between target distance correction coefficient $C_{d,target}$ and linear approximation value C_d for different wave group steepness value (Ak_p). (b) Distance correction factor for nonlinear simulation ($C_{d,target}$) and linear wave group profile (C_d) for Case 1 against distance in y direction. (c) Distance correction factor for nonlinear simulation ($C_{d,target}$) and linear wave group profile (C_d) for a typical wave group with $Ak_p = 0.28$, $L_x k_p = 5.93$ and $L_y k_p = 49$. Parameters are normalised by peak wavenumber k_p and spatial positions and peak wavelength λ_p .

Gaussian process regression, and we choose α to be a constant of 1.96, which gives the 95% of confidence interval. Unfortunately, direct computation of $\tilde{T}_{r,n}(\kappa; \kappa^*)$ requires evaluation of the unstable parameter-to-observation map $\hat{T}_r(\kappa)$ multiple times, as the value depends on the choose of the extra observation point κ^* . In this study, we approximate the value of $\tilde{T}_{r,n}(\kappa; \kappa^*)$ with the predictions previous dataset $\hat{T}_{r,n-1}(A, L_x, L_y)$. Although this approximation will not be accurate at the first few observations, the accuracy of this substitution increases as the predictions converge to the actual unstable parameter-to-observation map.

With the confidence interval of the predictions $\hat{\rho}_n^+(q; A^*, L_x^*, L_y^*)$ obtained as a function of the next κ^* , the next ‘best’ sampling parameter sets, metric functions are required to quantify the accuracy of the predictions. In this study, we proposed two metric functions for determining the next ‘best’ sampling parameter sets ($\kappa^* : A^*, L_x^*, L_y^*$). Both metric functions are essential to the accurate estimation of the final probability density function $\hat{\rho}_r(q)$ as they emphasise the model error based on different criteria.

Background statistical metric function. The first metric function gives emphasis on the main probability mass $\rho_{A,L_x,L_y}(A, L_x, L_y)$ as only the unstable parameter-to-observation map $\hat{T}_r(A, L_x, L_y)$ with non-zero $\rho_{A,L_x,L_y}(A, L_x, L_y)$ in the parameter space contributes to the final distribution. We present the iso-surfaces of probability density of the background statistics $\rho_{A,L_x,L_y}(A, L_x, L_y)$ in Fig. 4. We obtain this distribution with the parameterisation of random wave fields detailed in Section 2.3.

From Fig. 4, the distribution of the wave group parameters is actually centred in the parameter space based on the linear simulations. More importantly, a large number of the combinations of wave group parameters are very unlikely for the current sea state. In this study, apart from the instability region detailed in Appendix, we do not introduce extra physical parameters for further narrowing of the parameter space. However, physical limits can be used to further decrease the sampling size. For example, waves will break after reaching certain breaking criteria, which limits the maximum amplitude of the envelope (Babanin et al., 2007).

Based on the background statistics $\rho_{A,L_x,L_y}(A, L_x, L_y)$, we followed Mohamad and Sapsis (2018) and propose the first metric function Q^1 as:

$$Q^1(\hat{\rho}_n^+(\cdot; \kappa), \hat{\rho}_n^-(\cdot; \kappa)) \triangleq \frac{1}{2} \int |\hat{\rho}_n^-(s; \kappa) - \hat{\rho}_n^+(s; \kappa)|^2 ds. \quad (21)$$

This metric function limits the sampling combinations of wave group parameters as the contribution to the metric function Q^1 vanishes if the combinations of wave group parameters are associated with zero value of $\rho_{A,L_x,L_y}(A, L_x, L_y)$. This metric function essentially allows

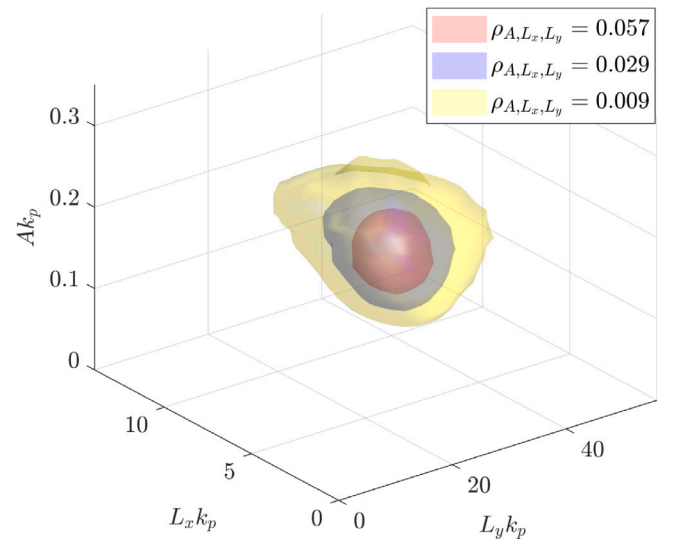


Fig. 4. Iso-surfaces of probability density of the background statistics $\rho_{A,L_x,L_y}(A, L_x, L_y)$. The initial Gaussian spectrum has a spectral bandwidth in x and y direction of $\sigma_x = \sigma_y = 0.1$. The steepness of the random wave field is $H_s k_p = 0.25$. Parameters are normalised by the peak wavenumber k_p .

us to skip these wave groups which are improbable given the sea state, and the accuracy of the learning algorithm predictions is biased towards capturing the nonlinear behaviour of the majority of the wave groups for the current sea state. Hence, unlike the uniform weighted metric function, the next best point of sampling is always selected with a non-zero from the background statistics $\rho_{A,L_x,L_y}(A, L_x, L_y)$ with the metric function Q^1 . We note that, in the present application, our understanding of the likely shapes of extreme wave-groups could also have been used rather than a purely data driven method. However, this would not be the case for other ocean engineering challenges where this method would be more valuable than in the present example application.

In Fig. 5, we present the sampling positions with the sequential experimental design using uniform weighted metric function and the biased metric function Q^1 . The number of observations required to achieve the same level of accuracy with the sequential sampling method is significantly decreased with the metric function Q^1 . This is mainly because the size of the sampling domain is greatly reduced as we are only interested in the unstable parameter-to-observation map within the yellow iso-surface. We are not trying to capture every detail at the corners of the domain, as wave groups with this shape

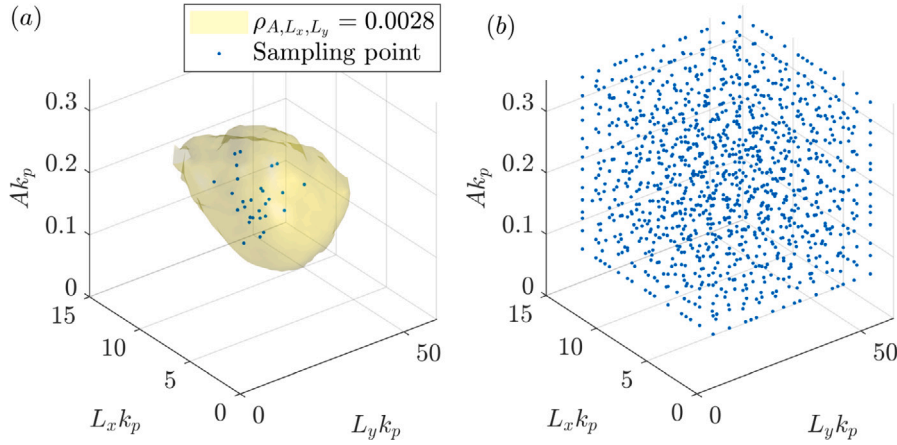


Fig. 5. Sampling positions with the sequential experimental design using (a) the biased metric function Q^1 , (b) uniformed weighted metric function within the parameter space in Eq. (33) for Case 2 in Table 1. Yellow iso-surfaces present the probability density of the background statistics $\rho_{A,L_x,L_y}(A, L_x, L_y)$ at a constant value of 0.0028. Parameters are normalised by the peak wavenumber k_p .

parameter are improbable given the current sea state. The proposed metric function Q^1 successively limits the sampling range and greatly reduces the number of observations required for inferring the unstable parameter-to-observation map.

A disadvantage of using the biased metric function Q^1 also prevents the predicted nonlinear parameter-to-observation map from being directly reused for different sea states. This is due to the weight function $\rho_{A,L_x,L_y}(A, L_x, L_y)$ being unique to each sea state. Given the fact that the metric function Q^1 can reduce by orders of magnitude in the sampling size when compared to the uniformed weighted metric function. This metric function Q^1 is still advantageous in terms of computational savings. Additionally, previously sampled parameter-to-observation map values can still be used as part of a new database, which greatly reduces the number of new sampling points required for a new sea state.

Extreme value metric function. For ocean engineering practice, extreme wave statistics are important as large waves are one of the critical design criteria. For example, the ‘wave-in-deck’ problem for offshore platforms requires platforms to be sufficiently high for the wave to pass underneath the deck. The previously introduced metric function Q^1 is unfortunately focused on the background statistics instead of the extreme statistics. Although the metric function Q^1 can capture the main body of the distribution with a few samples, the lack of focus on the extreme values still introduces large statistical errors in the tail of the distribution. We present this limitation of metric function Q^1 in Fig. 6.

From Fig. 6(a), the sequential design method with the metric function Q^1 can capture most of the statistical behaviour with the common wave group parameters combinations in the middle of the distribution range. However, from Fig. 6(b), large discrepancy can be observed at the tail of the distribution with metric function Q^1 . This is mainly because the most common wave group parameter combinations are not necessarily associated with an extreme response. Instead, areas with significant probability in background statistics $\rho_{A,L_x,L_y}(A, L_x, L_y)$ are not associated with areas with a large nonlinear response (i.e. large unstable parameter-to-observation map $\hat{T}_r(\kappa)$). Therefore, in the study we use another metric function Q^2 as the logarithmic transformation of the previous metric function (see also Mohamad and Sapsis, 2018):

$$Q^2(\bar{\rho}_n^+(\cdot; \kappa), \bar{\rho}_n^-(\cdot; \kappa)) \triangleq \frac{1}{2} \int |\log(\bar{\rho}_n^-(s; \kappa)) - \log(\bar{\rho}_n^+(s; \kappa))|^2 ds. \quad (22)$$

In this study, both metric functions are used to capture both the probability associated with the majority of the background statistics and the extreme wave statistics at the low probability region. The next best sampling point can be determined by minimising the metric function as:

$$\kappa_n = \arg \min_{\kappa} Q(\bar{\rho}^+(\cdot; \kappa), \bar{\rho}^-(\cdot; \kappa)). \quad (23)$$

We observe the desired unstable parameter-to-observation map $\hat{T}_r(A, L_x, L_y)$ at the next best sampling point by running nonlinear wave group simulation (see details in Section 3.2.1). We repeat this process until the desired level of accuracy is achieved. We define the relative accuracy level E_r based on the uncertainty from the predictions of the learning algorithm for metric function 1 as:

$$E_r^1 = \frac{\int |(\bar{\rho}_n^-(s; \kappa)) - (\bar{\rho}_n^+(s; \kappa))| ds}{\int |\bar{\rho}_n(s; \kappa)| ds}, \quad (24)$$

and the relative accuracy level for metric function 2 can be determined as:

$$E_r^2 = \frac{\int |\log(\bar{\rho}_n^-(s; \kappa)) - \log(\bar{\rho}_n^+(s; \kappa))| ds}{\int |\log(\bar{\rho}_n(s; \kappa))| ds}. \quad (25)$$

In this study, we first follow the sequential experimental design with metric function 1 to capture the statistical behaviour of non-extreme statistics. We evaluate the convergence with E_r^1 and switch to metric function 2 to capture the statistical behaviour of the extremes.

3. Numerical simulation

In this study, we perform three types of numerical simulations to fully examine the accuracy and the efficiency of this sequential sampling method for deep water waves. We use linear simulations of random wave fields to obtain the stable parameter-to-observation map $\hat{T}_l(\kappa)$ and the distributions of wave field parameters ($\hat{\rho}_\kappa$). We also simulate the nonlinear evolution of wave groups with MNLS to obtain both the densely sampled unstable parameter-to-observation map $\hat{T}_r(\kappa)$ and the estimated unstable parameter-to-observation map with sequential sampled points. Finally, the direct Monte-Carlo nonlinear simulations random wave simulations are performed with MNLS to verify the performance of the sequential sampling method. We summarise the simulations in Table 1.

We simulate both linear and nonlinear random waves with the setup used by Dysthe et al. (2003). We use a uniformly distributed computational grid with $N_x = N_y = 128$ where N_x and N_y are the number of grid points in x and y direction. Periodic boundary conditions are setup in both horizontal directions. In this study, the peak wavelength for all the generated waves is $\lambda_p = 2\pi$ m, which correspond to a peak period of $T_p = 2$ s. The length of the computation domain is 20 wavelengths in y direction and 10 wavelengths in x direction, which is achieved with the wave vector plane of $\Delta K_x = 1/(10\lambda_p)$ and $\Delta K_y = 1/(20\lambda_p)$. This means that the computational domain has a length in x direction $D_x = 2\pi/\Delta K_x$ and length in y direction $D_y = 2\pi/\Delta K_y$ with grid points: $x_j = D_x j/N_x, j = 0, 1, \dots, N_x - 1$ and $y_l = D_y l/N_y, l = 0, 1, \dots, N_y - 1$.

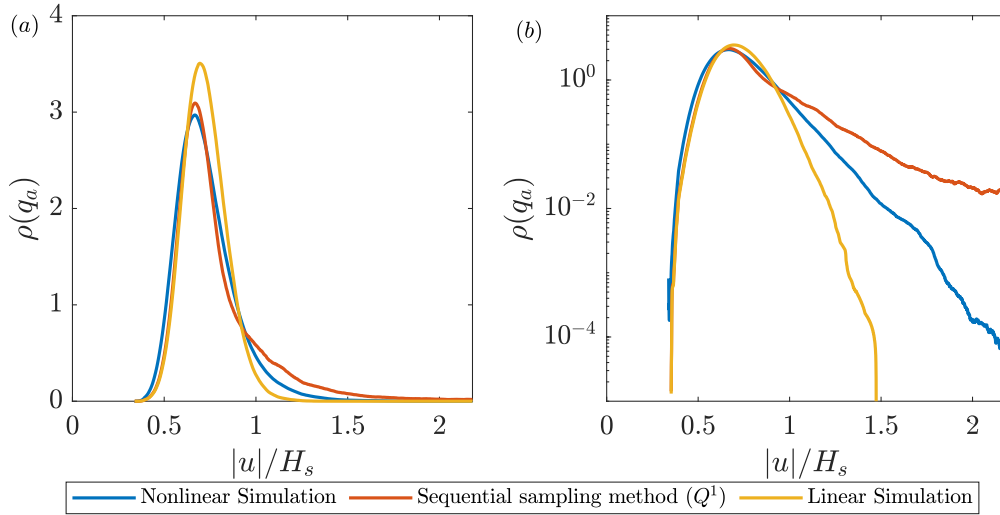


Fig. 6. Performance of sequential experimental design at iteration 10 (total of 15 sampling points) using only metric function Q^1 for space-time wave statistics (Case 2 in Table 1): (a) in normal scale, (b) in logarithmic scale.

Table 1

Summary of simulations in this study, where H_s is the significant wave height, k_p is the peak wavenumber, T_p is the peak period, σ_x and σ_y are the spectral bandwidth in x and y direction and λ_p peak wavelength.

Case	H_s, k_p	T_p [s]	σ_x	σ_y	Quantity of interest	Sampling domain (S)
1	0.25	2	0.1	0.1	$ u /H_s$	Centre point
2	0.25	2	0.1	0.1	$\max_{x,y \in S} u(x,y) /H_s$	$x \in [0, 10\lambda_p], y \in [0, 10\lambda_p]$

We start our simulation with a Gaussian spectrum given by:

$$\psi_{mn} = \frac{\epsilon}{\sqrt{2\pi\sigma_x\sigma_y}} \exp \left[-\frac{(m\Delta K_x)^2}{4\sigma_x^2} - \frac{(n\Delta K_y)^2}{4\sigma_y^2} \right], \quad (26)$$

where the $\sigma_x = \sigma_y = 0.1$ is the spectral bandwidth in x and y direction respectively. The parameter ϵ is the wave steepness scaling factor, which is tuned to match the target significant wave height of 0.25 m for the initial conditions for linear and nonlinear random simulations. The Fourier components \hat{B}_{mn} can be calculated with a random phase as:

$$\hat{B}_{mn} = \sqrt{\Delta K_x \Delta K_y} \psi_{mn} \exp i\theta_{mn}, \quad (27)$$

where θ_{mn} is the added random phase, which has a uniform distribution in $[0, 2\pi]$.

3.1. Linear simulations

For linear simulations, the amplitude can be calculated as:

$$u_0(x_j, y_l, t) = \sum_{m,n} \hat{B}_{mn} \exp(i(m\Delta K_x x_j + n\Delta K_y y_l - \omega_{mn}t)), \quad (28)$$

where ω_{mn} is the angular frequency of wave components. We simulate 40 periods forwards for each case and a total of 8000 cases with different random seeds are generated.

3.2. Non-linear simulations

In this study, we use the MNLS presented in Trulsen et al. (2000) as numerical solutions for the two dimensional water wave system. The MNLS model we used herein retains up to the third order of nonlinearity with free wave components with bandwidth constraints. Numerical models containing more physics such as fully-nonlinear code (Engsig-Karup et al., 2009), and Higher-Order Spectral Method (HOSM) (Dommermuth and Yue, 1987; West et al., 1987). However, simulation time is critical in this study, as the Monte Carlo sampled wave statistics is used as the reference. This requires a large number of

simulations (see Table 3 for details) to obtain stable statistics at the tail of the distribution. Additionally, previous studies show relatively accurate wave statistics can be obtained with MNLS. For example, Toffoli et al. applied the MNLS and HOSM for directionally spread waves, and both methods show good quantitative agreement with experimental results (Toffoli et al., 2010b) (see also detailed comparisons in Adcock and Taylor, 2016; Barratt et al., 2021). We note that the contributions from second order bound harmonics are not considered in this study, as the contributions from these second order bound harmonics can be estimated from the free wave components straightforwardly (Dalzell, 1999).

The hybrid scheme proposed by Trulsen et al. (2000), which is solved spectrally is given by:

$$\begin{aligned} \frac{\partial u}{\partial t} + L(\partial_x, \partial_y)u + \frac{i\omega_0 k_p^2}{2}|u|^2 u + \frac{3\omega_0 k_p}{2}|u|^2 \frac{\partial u}{\partial x} \\ + \frac{\omega_0 k_p}{4} u^2 \frac{\partial u^*}{\partial x} + ik_p \frac{\partial \phi}{\partial z} \Big|_{z=0} u = 0, \end{aligned} \quad (29)$$

where ω_0 is the peak angular frequency, k_p is the peak wavenumber, $*$ represents the complex conjugate, z is the vertical position, and the linear operator $L(\partial_x, \partial_y)$ is given as:

$$L(\partial_x, \partial_y) = i \left\{ \left[(1 - i\partial_x)^2 - \partial_y^2 \right]^{1/4} - 1 \right\}, \quad (30)$$

and potential of the return current ϕ is given by

$$\frac{\partial \phi}{\partial z} \Big|_{z=0} = \frac{\omega_0}{2} \frac{\partial |u|^2}{\partial x}, \quad (31)$$

and the potential satisfies Laplace's equation in the fluid. We calculate surface elevation from the complex envelope u using the following:

$$\eta = \text{Real}(u \exp(i(k_p x - \omega_0 t))). \quad (32)$$

For all the numerical simulations with MNLS, a fixed time step of 0.04 s (50 steps per period) is used.

3.2.1. Focused wave group

To obtain the parameter-to-observation map, we simulate the non-linear focusing of deterministic wave groups, which has been adopted

in various previous studies (Lo and Mei, 1985; Baldock et al., 1996; Gibbs and Taylor, 2005; Gibson and Swan, 2006). The initial condition of the wave group simulation is an envelope profile, which will perfectly focus after 15 periods in linear simulations. The focused wave profile is the same as the Gaussian wave group profile described in Eq. (12) during the parameterisation processes. We also capture the de-focusing process by simulating the evolution of the complex envelope for another 15 periods, which essentially captures the same ‘fast’ nonlinear physics as the extreme event in the random wave simulations.

In this study, although we are using the sequential design method for nonlinear wave group simulations (i.e. the parameters of the wave group are determined from previous learning model predictions), we still need to densely sample the parameter space as the reference ground ‘truth’. According to the distribution of wave group parameters obtained during the parameterisation process, we consider the following parameter space:

$$L_x k_p \in [1, 14], \quad L_y k_p \in [4, 60] \quad \text{and} \quad Ak_p \in [0.07, 0.38]. \quad (33)$$

A total of 30 bins are used for both L_x and L_y , and 31 uniformly distributed bins are used for A . This results in a total of 27900 cases of wave groups with different combinations of parameters that need to be simulated if we densely sample the parameter space.

We use a fixed 256 points in x direction and 128 points in y direction for the nonlinear wave group simulations with periodic boundary conditions at the edges of the computational domain. The physical domain size is adjusted based on the size of the wave group from 4 wavelengths to 40 wavelengths in x direction and from 4 wavelengths to 20 wavelengths in y direction linearly. A fixed time step of 0.04 s (50 steps per period) is used for wave group simulations.

3.2.2. Random wave fields

We evolve the initial envelope with the Fourier components given in Eq. (27) using the MNLS to examine the nonlinear statistics of the random wave fields. We applied the same numerical setup for nonlinear random wave simulations as the linear random wave fields. A total of 8000 realisations are examined to obtain a stable probability density function as the reference case.

4. Results

4.1. Performance of the sequential sampling method

In this study, we start with a typical problem in ocean engineering: estimating wave statistics for directionally spread waves measured at a single Eulerian point. We further investigate the accuracy of the sequential design method for space–time wave statistics, as the extreme value of surface elevation over an area. In this study, we compare the performance of a sequential sampling method with the Monte-Carlo nonlinear simulations as the ground truth.

In Fig. 7(a), we present a probabilistic distribution function of the normalised envelope measured at a single Eulerian point. The nonlinear simulation deviates from the linear predictions due to the inherent third order nonlinearity of the water wave system. This deviation is more significant at the tail of the distribution, which indicates more larger waves are expected because of correlations between wave components driven by non-linear physics. This heavy tailed statistical behaviour has been reported by various studies with analytical predictions (Fedele, 2015; Janssen and Janssen, 2019), numerical simulations (Xiao et al., 2013; Klahn et al., 2021b) and experiments (Onorato et al., 2009a; Latheef and Swan, 2013). This deviation is well predicted by the sequential sampling method proposed in this study, especially at the tail of the distribution. The sequential sampling method over predicts the probability of medium sized waves for point measurements. One reason for this over prediction is due to the weakly nonlinearity of the water wave system. From Fig. 7(a) the envelope distribution for nonlinear

simulation is still well predicted by the linear approximation. This leads to the value of the unstable parameter-to-observation map being rather small, which leads to less importance in the metric function calculation. Additionally, two metric functions proposed in this study focus on either the mass distribution (Q^1) or the situation with large unstable parameter-to-observation map values in logarithmic scale (Q^2), which are not targeted at the probability of medium sized waves. An extra metric function could improve the prediction results but due to the limited importance of capturing the probability of medium sized waves very accurately in engineering practice, we choose to keep the reduced order model simple and focussed on the extremes.

In Fig. 7(a), we present probabilistic function of normalised envelope measured at over an area. For space–time wave statistics, the nonlinear physics leads to a large deviation from the linear wave theory, especially at the tail of the distribution. This increased nonlinear amplification of the wave envelope agrees well with previous studies. This larger deviation is a combined effect of the fact that the point measurements tend to miss the peak of the envelope (see details in Fig. 1) and also the fact that space–time wave statistics simply include more wave groups during the sampling procedure. The proposed sequential sampling method can capture these two additional effects well and can predict the nonlinear space–time wave statistics accurately even at the tail of the distribution, which shows potential for engineering applications.

4.2. Computational time and the ‘curse of dimensionality’

In this subsection, we aim to explain why the sequential sampling framework proposed in this study has advantages in saving the computational time and its potential for nonlinear ocean engineering problems. An example might be loads on offshore infrastructure where both loads from unbroken waves and breaking waves are important.

We note that the calculation of computational time presented in this section is simplified. Different numerical models can result in significantly different computational time and the optimisation of the numerical code, spatial–temporal discretisation, domain size, and governing equations might also affect the computational time. However, the general conclusion should hold as we are looking at differences over orders of magnitude.

We summarise the computation costs for the current study in Table 2 and the number of cases required for different statistical approaches in Table 3. We note that the exact number of cases depends also on the nonlinearity of the system at certain sea states and the accuracy requirements on the final predictions. Therefore, the number of cases presented in Table 3 could be different in other situations, and we use these numbers mainly for demonstration purposes. To achieve a stable statistical behaviour, the nonlinear random time series is required for different random amplitudes and phases, and for the Monte Carlo type of approach, all the nonlinear cases need to be re-simulated for different sea states. The linear random time series are also needed with different random seeds during the parameterisation process. In this study, the sequential sampling method can significantly reduce the number of wave group simulations required initially, but extra simulations of wave groups are necessary when the sea state changes.

We now estimate the total simulation time required by different approaches by ignoring all the other computational cost (e.g. post-processing, data storage and loading), as these computational costs are usually correlated with the simulation time. We present the total computational time for different sampling methods in Fig. 8(a). We refer to the method previously presented in Tang and Adcock (2022) and Mohamad and Sapsis (2016) as the grid search method with parameterisation, as in these studies the random wave fields are parameterised, but the nonlinear responses from wave group simulations are sampled with the grid search method. We also estimate the computational cost for a different number of sea states. This is more relevant to engineering

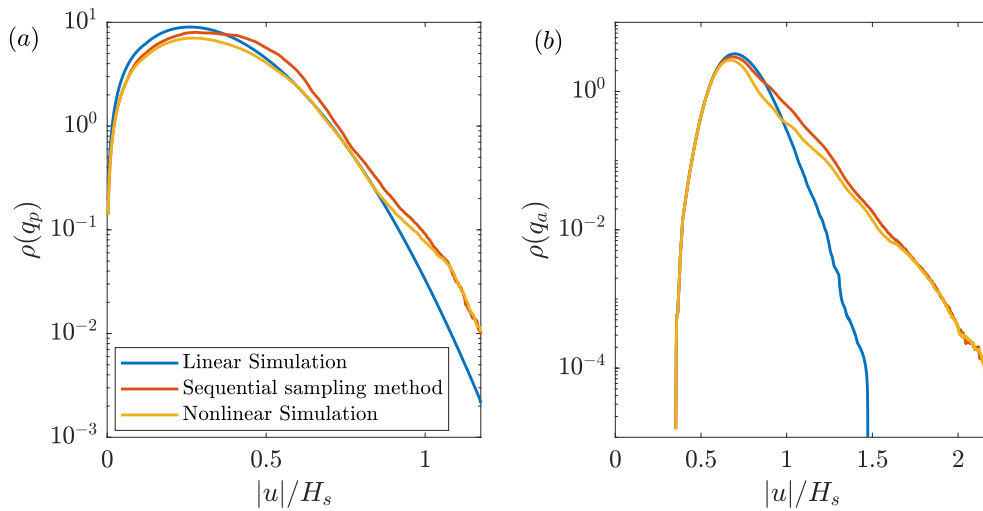


Fig. 7. Comparison of probabilistic distribution function of (a): time domain wave statistics (Case 1 in Table 1) and (b): normalised spatial maximum envelope height distribution within an squared area with a side length of $10\lambda_p$ (Case 2 in Table 1). We present the sequential design method predictions (yellow line), nonlinear simulation results (red line) and linear simulation results (blue line). Envelope height is normalised by the significant wave height (H_s). An averaging smoothing with a window of 40 bins is applied for clearer visualisation at the tail of the distribution.

Table 2

Computational time (single threaded) on a regular desktop (CPU Intel i7-6700K) for different types of simulations. Detailed numerical setups for different types of simulation are presented in Section 3.

Type of simulation	Numerical method	Computational time per case (s)
Nonlinear wave group	MNLS	21.4
Nonlinear random waves	MNLS	103
Linear random waves	Fast Fourier transform	4.89

Table 3

The number of simulation cases required for different statistical approaches. We present the initial number of simulations in this table as well as the additional simulations required if different sea states are considered in the bracket.

Type of method	Wave group	Linear random	Nonlinear random
Monte Carlo	0 (0)	0 (0)	5000 (5000)
Grid search with parameterisation	18000 (0)	8000 (8000)	0 (0)
Sequential sampling with parameterisation	57(57)	8000 (8000)	0 (0)

practice, where the nonlinear responses from hundreds of extreme sea states are required for accurate estimation of the probability of extreme waves during the life-time of offshore structures. The computational costs for both grid search sampling and the sequential sampling method are normalised against the Monte Carlo nonlinear simulation of random time series. For a relatively small number of sea states, the grid search method with parameterisation still requires a significant amount of computational resources. However, the grid search method with parameterisation has the edge as the number of sea states increases. This is primarily because linear simulations are only required once the wave group database has been established. The sequential sampling method presented herein has a fraction of the computational cost for a small number of sea states, which shows its advantages in small scale scientific research and experimental applications.

We now consider the sampling difficulty if the number of input parameters increases from 3 (A, L_x, L_y for this study) to 4. The extra parameter could be used to describe extra physics (i.e. wave breaking) and would give a better statistical model. We assume the computational cost for modelling the new physics is proportionally increased for all three types of simulations. We present the total computational time for different sampling methods for 4 inputs in Fig. 8(b). The grid search with parameterisation clearly struggles for modelling a small number of sea states efficiently and requires over 25 times of computational resources when compared to Monte Carlo simulations. This is primarily due to the number of wave group simulations which increases from 18,000 to 540,000 if there are 30 sampling points for the new input

parameter. The logarithmic increase in the sampling difficulty, when the number of input parameters increases linearly is often described as the curse of dimensionality. The ‘dimension’ we refer to herein is the dimension in input parameter space, which is different from the physical spatial–temporal dimension in water wave systems (Trunk, 1979).

For the sequential sampling method proposed in this study, we can simply repeat the process for the new input parameter, which results in an increase for wave group simulations from 57 to 1710. This is a smaller computational increase than the grid search method, and the overall computational cost is still at a fraction when compared to Monte Carlo simulations. We note that the sequential sampling method cannot completely avoid the ‘curse of dimensionality’, but the ability to incorporate more inputs parameters shows huge potential in modelling more complex systems with extra physics. Additionally, a new parameterisation method also can be developed based on the current version proposed in this paper for a better generalisation of the wave behaviour. As such, we believe that the sequential sampling method proposed herein is a plausible approach to mitigate the risk from the ‘curse of dimensionality’ for ocean engineering problems.

5. Discussion and conclusions

This paper applies the sequential experimental design method to predict space–time wave statistics in a directional spread water wave system. In the proposed model, the spatial profile of a random wave

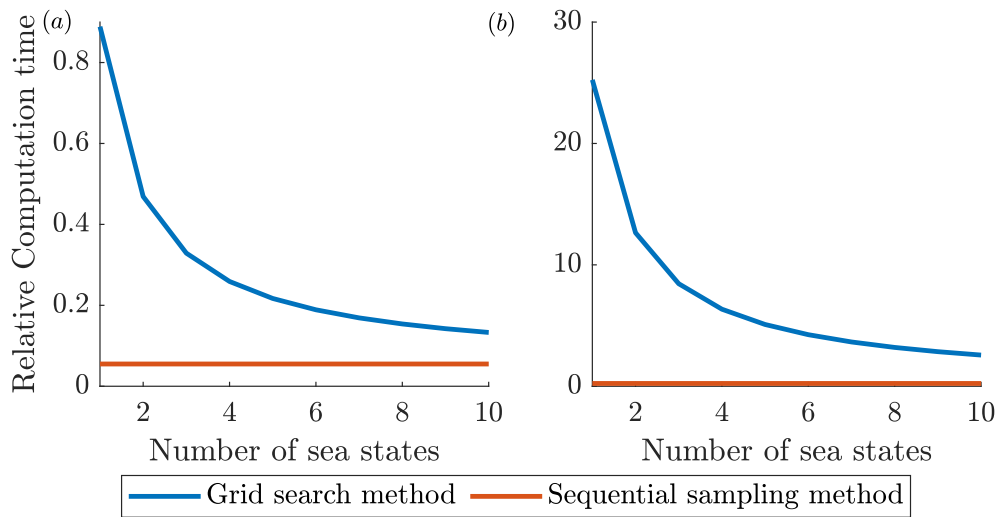


Fig. 8. Comparison of the relative computational time of (a): for current study with three input parameters after the parameterisation of the random wave field, (b): for four input parameters for the nonlinear system. We present for both the grid research sampling method (blue line) and the sequential sampling method used in this study (red line). We calculate the relative computational time through normalising the estimated computational time by the computation time required for nonlinear Monte Carlo simulation of random wave fields with the same numerical method.

field is approximated by the superposition of Gaussian wave groups. Stable and unstable parameter-to-observation maps are introduced to bridge the distribution of these wave group parameters to the final space-time crest statistics. We estimate the unstable parameter-to-observation maps by running nonlinear simulations of deterministic wave groups. However, the traditional grid search method requires sampling every combination of input parameters within the parameter space for an accurate prediction, which is computationally costly and time-consuming. In this study, we introduce the sequential experimental design method for sampling unstable parameter-to-observation map, which effectively reduces the number of samples from tens of thousands of observations to less than one hundred.

In traditional sampling methods, the next sample position is predefined. However, the proposed sequential experimental design method adjusts the next sampling point based on the previous observations. This method applies a Gaussian process learning algorithm to predict the system response based on the previous samples before the new observation is made. Based on the prediction results, the next best sampling point can be determined by minimising the metric functions. In this study, two metric functions are used and tested to achieve faster convergence and a more accurate estimation of parameter-to-observation map with fewer samples. We compare the performance of the sequential experimental design method with direct Monte Carlo simulations and the proposed sampling method can accurately estimate the wave statistics with several orders fewer sample points.

Fewer sampling points are required for estimating the wave statistics therefore saving significant amounts of computation as far fewer wave group simulations are required. This advantage of saving computational resources would be magnified when considering more complex numerical models. These computational intensive models allow other effects to be included when predicting the wave statistics. For example, wave breaking and wave current interaction effects can be fully analysed only using direct numerical modelling techniques. As it is computationally impractical to achieve stable wave statistics with Monte Carlo simulations with such a computational intensive code, a sequential sampling method would make such an investigation possible with a computationally affordable number of wave group simulations. Furthermore, such a sequential design method can also be applied during the parameter scoping for wave experiments. Based on the predictions from the learning algorithm, the test matrix can be updated after each individual experimental run. The next best sampling point provides the optimal combination of input parameters to achieve

the largest information gain. This allows a better coverage over the parameter space for valuable and expensive experimental results.

We would also like to discuss the potential of such methods. In this study, we aim at proposing a data-centric framework that could model the statistical behaviour of a nonlinear system (i.e. the two dimensional water wave system herein) with fewer observations. Such a method is valuable because observations from numerical simulations, field data or experiments are valuable and usually ‘expensive’. Furthermore, the data-centric method proposed here has huge potential for modelling more complex systems. This is not only due to the difficulty during the observation process (i.e. experiments or direct numerical simulations) but also because extra parameters would trigger the ‘curse of dimensionality’ (see Section 4.2). The method presented in this study shows a plausible workaround for mitigating such difficulties for modelling complex systems.

The two dimensional water wave system examined herein is a complex system. Even with a simplified numerical model (i.e. MNLS herein) and nondimensionalisation, there are still many parameters which could affect the wave statistics in a uni-modal sea state, such as wave steepness, spectral parameters of the initial wave spectrum and directional spreading functions. The number of sea state parameters could be double or tripled when multi-modal sea states are considered (Toffoli et al., 2008; Bitner-Gregersen and Toffoli, 2014; Toffoli et al., 2011). Additionally, with deterministic wave groups, the statistical response from the water wave system also involves a spatial-temporal envelope profile in three physical dimensions. As such, it is almost impossible to set up a unified statistical model to include all these parameters and provide accurate predictions based on observations covering the entire parameter space. In this study, we present several practical methods for estimating the nonlinear response of the system with far fewer observations required. This is achieved with both the distance correction factor, which approximates the nonlinear spatial response of the wave envelope from the measurements at the envelope peak, and the sequential sampling method, which estimates the unstable parameter-to-observation map with several orders fewer sample points. We examine the performance of the proposed methods with nonlinear Monte Carlo simulation of random waves, and the accurate prediction of wave statistics from the proposed method indicates this is a small but solid step towards a better statistical model for a nonlinear water wave system.

CRedit authorship contribution statement

Tianning Tang: Developed the data processing methods, Implemented the data processing methods, Analysed the data, Contributed to writing the paper. **Thomas A.A. Adcock:** Developed the data processing methods, Analysed the data, Contributed to writing the paper, Led the study.

Declaration of competing interest

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

Acknowledgements

This work was funded by an COVID extension to EP/R007632/1 and is now funded by EP/V050079/1. The former was a joint EPSRC/NERC grant. The latter is an EPSRC, United Kingdom grant.

Appendix. Instability region boundary

In this study, we follow Tang and Adcock (2022) and define an instability region (R_e) to separate the parameter space as:

$$R_e = \left\{ (A, L_x, L_y) \mid \frac{A_{\max}}{A} > 1.05 \right\}, \quad (\text{A.1})$$

where (A, L_x, L_y) are wave group parameters and the spatial and temporal maximum amplitude A_{\max} can be obtained through nonlinear wave group simulations as:

$$A_{\max} = \max_{x,y,t} |u(x, y, t)|. \quad (\text{A.2})$$

In this study, the boundary of the instability region can be obtained as a special case during the sequential sampling process for the unstable parameter-to-observation map $\hat{T}_r(\kappa)$ as:

$$A_{\max} = \max_{(x,y) \in S} |u(x, y, t_0)| \Big|_{t_0=0}, \quad (\text{A.3})$$

after centring the nonlinear focus time at $t = 0$.

References

- Adcock, T.A.A., Gibbs, R.H., Taylor, P.H., 2012. The nonlinear evolution and approximate scaling of directionally spread wave groups on deep water. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* 468 (2145), 2704–2721.
- Adcock, T.A.A., Taylor, P.H., 2016. Fast and local non-linear evolution of steep wave-groups on deep water: A comparison of approximate models to fully non-linear simulations. *Phys. Fluids* 28 (1), 16601.
- Adcock, T.A.A., Yan, S., 2010. The focusing of uni-directional Gaussian wave-groups in finite depth: An approximate NLSE based approach. In: 29th Int. Ocean Offshore Arct. Eng. Conf., vol. 49125, Shanghai, China, pp. 569–576.
- Babanin, A.V., Chalikov, D., Young, I.R., Savelyev, I., 2007. Predicting the breaking onset of surface water waves. *Geophys. Res. Lett.* 34 (7).
- Baldock, T.E., Swan, C., Taylor, P.H., 1996. A laboratory study of nonlinear surface waves on water. *Philos. Trans. Royal Soc. A Math. Phys. Eng. Sci.* 354 (1707), 649–676.
- Barratt, D., van den Bremer, T.S., Adcock, T.A.A., 2021. MNLS simulations of surface wave groups with directional spreading in deep and finite depth waters. *J. Ocean Eng. Mar. Energy*. 1–15.
- Benetazzo, A., Barbariol, F., Bergamasco, F., Torsello, A., Carniel, S., Sclavo, M., 2015. Observation of extreme sea waves in a space-time ensemble. *J. Phys. Oceanogr.* 45 (9), 2261–2275.
- Bitner-Gregersen, E.M., Toffoli, A., 2014. Occurrence of rogue sea states and consequences for marine structures. *Ocean Dyn.* 64 (10), 1457–1468.
- Boccotti, P., 1983. Some new results on statistical properties of wind waves. *Appl. Ocean Res.* 5 (3), 134–140.
- Dalzell, J.F., 1999. A note on finite depth second-order wave-wave interactions. *Appl. Ocean Res.* 21 (3), 105–111.
- Dematteis, G., Grafke, T., Vanden-Eijnden, E., 2018. Rogue waves and large deviations in deep sea. *Proc. Natl. Acad. Sci. USA* 115 (5), 855–860.
- Dommermuth, D.G., Yue, D.K.P., 1987. A high-order spectral method for the study of nonlinear gravity waves. *J. Fluid Mech.* 184, 267–288.
- Dysthe, K.B., Trulsen, K., Krogstad, H.E., Socquet-Juglard, H., 2003. Evolution of a narrow-band spectrum of random surface gravity waves. *J. Fluid Mech.* 478, 1–10.
- Engsig-Karup, A.P., Bingham, H.B., Lindberg, O., 2009. An efficient flexible-order model for 3D nonlinear water waves. *J. Comput. Phys.* 228 (6), 2100–2118.
- Farazmand, M., Sapsis, T.P., 2017. Reduced-order prediction of rogue waves in two-dimensional deep-water waves. *J. Comput. Phys.* 340, 418–434.
- Fedele, F., 2012. Space-time extremes in short-crested storm seas. *J. Phys. Oceanogr.* 42 (9), 1601–1615.
- Fedele, F., 2015. On the kurtosis of deep-water gravity waves. *J. Fluid Mech.* 782, 25–36.
- Fedele, F., Benetazzo, A., Gallego, G., Shih, P.C., Yezzi, A., Barbariol, F., Ardhuin, F., 2013. Space-time measurements of oceanic sea states. *Ocean Model.* 70, 103–115.
- Fedele, F., Lugni, C., Chawla, A., 2017. The sinking of the El Faro: Predicting real world rogue waves during Hurricane Joaquin. *Sci. Rep.* 7 (1), 1–15.
- Forristall, G.Z., 2006. Maximum crest heights over an area and the air gap problem. Hamburg, Germany, In: 25th Int. Ocean Offshore Arct. Eng. Conf., vol. 47489, pp. 11–15.
- Forristall, G.Z., 2011. Maximum crest heights under a model TLP deck. Rotterdam, The Netherlands, In: 30th Int. Ocean Offshore Arct. Eng. Conf., vol. 44342, pp. 571–577.
- Forristall, G.Z., 2015. Maximum crest heights over an area: Laboratory measurements compared to theory. Rotterdam, The Netherlands, In: 30th Int. Ocean Offshore Arct. Eng. Conf., vol. 56499, American Society of Mechanical Engineers.
- Gibbs, R.H., Taylor, P.H., 2005. Formation of walls of water in ‘fully’ nonlinear simulations. *Appl. Ocean Res.* 27 (3), 142–157.
- Gibson, R.S., Swan, C., 2006. The evolution of large ocean waves: The role of local and rapid spectral changes. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* 463 (2077), 21–48.
- Janssen, P.A.E.M., Janssen, A.J.E.M., 2019. Asymptotics for the long-time evolution of kurtosis of narrow-band ocean waves. *J. Fluid Mech.* 859, 790–818.
- Karmpadakis, I., Swan, C., Christou, M., 2020. Assessment of wave height distributions using an extensive field database. *Coast. Eng.* 157, 103630.
- Klahn, M., Madsen, P.A., Fuhrman, D.R., 2021a. On the statistical properties of inertia and drag forces in nonlinear multi-directional irregular water waves. *J. Fluid Mech.* 916.
- Klahn, M., Madsen, P.A., Fuhrman, D.R., 2021b. On the statistical properties of surface elevation, velocities and accelerations in multi-directional irregular water waves. *J. Fluid Mech.* 910.
- Latheef, M., Swan, C., 2013. A laboratory study of wave crest statistics and the role of directional spreading. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* 469.
- Lindgren, G., 1970. Some properties of a normal process near a local maximum. *Ann. Math. Stat.* 41 (6), 1870–1883.
- Lo, E., Mei, C.C., 1985. A numerical study of water-wave modulation based on a higher-order nonlinear Schrödinger equation. *J. Fluid Mech.* 150, 395–416.
- Lubin, P., Glockner, S., 2015. Numerical simulations of three-dimensional plunging breaking waves: Generation and evolution of aerated vortex filaments. *J. Fluid Mech.* 767, 364–393.
- McAllister, M.L., Adcock, T.A.A., Taylor, P.H., van den Bremer, T.S., 2017. The set-down and set-up of directionally spread and crossing surface gravity wave groups. *J. Fluid Mech.* 835, 131–169.
- Mohamad, M.A., Cousins, W., Sapsis, T.P., 2016. A probabilistic decomposition synthesis method for the quantification of rare events due to internal instabilities. *J. Comput. Phys.* 322, 288–308.
- Mohamad, M.A., Sapsis, T.P., 2016. Probabilistic response and rare events in Mathieu’s equation under correlated parametric excitation. *Ocean Eng.* 120, 289–297.
- Mohamad, M.A., Sapsis, T.P., 2018. Sequential sampling strategy for extreme event statistics in nonlinear dynamical systems. *Proc. Natl. Acad. Sci. USA* 115 (44), 11138–11143.
- Onorato, M., Cavaleri, L., Fouques, S., Gramstad, O., Janssen, P.A.E.M., Monbaliu, J., Osborne, A.R., Pakozdi, C., Serio, M., Stansberg, C.T., Toffoli, A., Trulsen, K., 2009a. Statistical properties of mechanically generated surface gravity waves: A laboratory experiment in a three-dimensional wave basin. *J. Fluid Mech.* 627, 235.
- Onorato, M., Osborne, A.R., Serio, M., Cavaleri, L., Brandini, C., Stansberg, C.T., 2004. Observation of strongly non-Gaussian statistics for random sea surface gravity waves in wave flume experiments. *Phys. Rev. E* 70 (6), 067302.
- Onorato, M., Waseda, T., Toffoli, A., Cavaleri, L., Gramstad, O., Janssen, P.A.E.M., Kinoshita, T., Monbaliu, J., Mori, N., Osborne, A.R., et al., 2009b. Statistical properties of directional ocean waves: The role of the modulational instability in the formation of extreme events. *Phys. Rev. Lett.* 102 (11), 114502.
- Sapsis, T.P., 2020. Output-weighted optimal sampling for Bayesian regression and rare event statistics using few samples. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* 476 (2234), 20190834.
- Seeger, M., 2004. Gaussian processes for machine learning. *Int. J. Neural Syst.* 14 (02), 69–106.
- Tang, T., Adcock, T.A.A., 2021. The influence of finite depth on the evolution of extreme wave statistics in numerical wave tanks. *Coast. Eng.* 166, 103870.
- Tang, T., Adcock, T.A.A., 2022. A reduced order model for space-time wave statistics using probabilistic decomposition-synthesis method. *Ocean Eng.* (submitted for publication).

- Toffoli, A., Babanin, A., Onorato, M., Waseda, T., 2010a. Maximum steepness of oceanic waves: Field and laboratory experiments. *Geophys. Res. Lett.* 37 (5).
- Toffoli, A., Benoit, M., Onorato, M., Bitner-Gregersen, E.M., 2009. The effect of third-order nonlinearity on statistical properties of random directional waves in finite depth. *Nonlinear Process Geophys.* 16 (1), 131–139.
- Toffoli, A., Bitner-Gregersen, E.M., Osborne, A.R., Serio, M., Monbaliu, J., Onorato, M., 2011. Extreme waves in random crossing seas: Laboratory experiments and numerical simulations. *Geophys. Res. Lett.* 38 (6).
- Toffoli, A., Gramstad, O., Trulsen, K., Monbaliu, J., Bitner-Gregersen, E., Onorato, M., 2010b. Evolution of weakly nonlinear random directional waves: Laboratory experiments and numerical simulations. *J. Fluid Mech.* 664, 313.
- Toffoli, A., Onorato, M., Osborne, A.R., Monbaliu, J., 2008. Non-Gaussian properties of shallow water waves in crossing seas. In: *Extreme Ocean Waves*. Springer, pp. 53–69.
- Tromans, P.S., Anatrak, A.R., Hagemeijer, P., 1991. New model for the kinematics of large ocean waves application as a design wave. *Proc. First Int. Offshore Polar Eng. Conf.* 8, 64–71.
- Trulsen, K., Kliakhandler, I., Dysthe, K.B., Velarde, M.G., 2000. On weakly nonlinear modulation of waves on deep water. *Phys. Fluids* 12 (10), 2432–2437.
- Trunk, G.V., 1979. A problem of dimensionality: A simple example. *IEEE Trans. Pattern Anal. Mach. Intell.* (3), 306–307.
- Wang, Z., Yang, J., Stern, F., 2016. High-fidelity simulations of bubble, droplet and spray formation in breaking waves. *J. Fluid Mech.* 792, 307–327.
- West, B.J., Brueckner, K.A., Janda, R.S., Milder, D.M., Milton, R.L., 1987. A new numerical method for surface hydrodynamics. *J. Geophys. Res. Oceans* 92 (C11), 11803–11824.
- Xiao, W., Liu, Y., Wu, G., Yue, D.K.P., 2013. Rogue wave occurrence and dynamics by direct simulations of nonlinear wave-field evolution. *J. Fluid Mech.* 720, 357–392.