# Statistical modeling of fully nonlinear hydrodynamic loads on offshore wind turbine foundations using wave episodes and targeted CFD simulations through active sampling

Stephen Guth[1], Eirini Katsidoniotaki[2,3], and Themistoklis P. Sapsis[1]

[1]Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02138, USA
[2]Department of Electrical Engineering, Uppsala University, SE 75237, Sweden
[3]Centre of Natural Hazards and Disaster Science (CNDS), SE 75236, Sweden

February 2, 2023

**Abstract**

For many design applications in offshore engineering, including offshore wind turbine foundations, engineers need accurate statistics for kinematic and dynamic quantities, such as hydrodynamic forces, whose statistics depend on the stochastic sea surface elevation. Nonlinear phenomena in the wave–structure interaction require high-fidelity simulations to be analyzed accurately. However, accurate quantification of statistics requires a massive number of simulations, and the computational cost is prohibitively expensive. To avoid that cost, this study presents a machine learning framework to develop a reliable surrogate model that minimizes the need for computationally expensive numerical simulations, which is implemented for the monopile foundation of an offshore wind turbine. This framework consists of two parts. The first focuses on dimensionality reduction of stochastic irregular wave episodes and the resulting hydrodynamic force time series. The second of the framework focuses on the development of a Gaussian process regression surrogate model which learns a mapping between the wave episode and the force-on-structure. This surrogate uses a Bayesian active learning method that sequentially samples the wave episodes likely to contribute to the accurate prediction of extreme hydrodynamic forces in order to design subsequent CFD numerical simulations. Additionally, the study implements a spectrum transfer technique to combine CFD results from quiescent and extreme waves. The principal advantage of this framework is that the trained surrogate model is orders of magnitude faster to evaluate than the classical modeling methods, while built-in uncertainty quantification capabilities allows for efficient sampling of the parameter using with the CFD tools traditionally employed.

Keywords: Heavy tails and extreme events; Offshore structures; Wave episodes; Reduced-order modeling; Optimal Experimental Design; Active Sampling.

# 1 Introduction

Renewable energy technologies are critical to secure a reliable energy future while mitigating the detrimental effects of climate change [2,18]. Of these technologies, offshore wind turbines are expected to play a key role in the global sustainable energy transition [15, 19, 29]. Over the last decade, the offshore wind turbine technology has experienced a remarkable growth [31], however, there is the need for targeted research to overcome challenges to the further offshore wind establishment, such as the relatively high cost of energy. At shallow and medium water depths, most offshore wind turbine applications use fixed-bottom foundations [7, 32], which include structures with monopile, tripod and jacket type foundations. Across their working lifetime, these structures are exposed to stochastic and extreme sea states which cause severe loads that impact the fatigue life [12, 14, 25, 38, 47, 58]. To implement current best practices, such as designing for a reliable operation lifetime of 25 years [45, 46], engineers require prior knowledge of the long-term load statistics on critical structural components (e.g. loads on the monopile support structure). Accurate analysis of the ultimate limit state (ULS) and fatigue limit state (FLS) would allow for less conservative designs and thus reduced costs.

A particular difficulty in the design of offshore wind turbines is their exposure to stochastic and harsh offshore environmental conditions. In the design stage, physical experiments and/or computational simulations are conducted in order to provide engineers with critical insights into the performance of the offshore wind turbine structures [48]. Through modeling approaches, engineers study three main load categories: structural dynamics, hydrodynamics, and aerodynamics [48]. In this study, we focus on the hydrodynamic loads on the monopile foundation of an offshore wind turbine. In the literature, several studies are available which utilize high-fidelity computational fluid dynamics (CFD) simulations to examine the monopile-type foundation in breaking waves [3, 33, 68], analyse the slamming effects [23, 49], investigate the flow and scour around the monopile [8, 61] and identify the second order hydrodynamic loads [43, 64]. Apart from the high-fidelity simulations, there are studies that implement lower fidelity models to estimate the critical loads [44, 63]. To facilitate the efficient and reliable design of an offshore wind turbine structure, engineers usually need to perform studies for sensitivity analysis, optimization, uncertainty quantification, statistics reconstruction for the ULS and FLS definition. However, these studies require a massive number of simulation runs. The high-fidelity simulations demand prohibitive computational cost and thus are infeasible in practice. Although lower fidelity numerical simulations are faster, they cannot capture high nonlinear phenomena (e.g. slamming) which are critical for the reliable design of the offshore wind turbine foundations.

Scientists have addressed the question of experimental design for these studies–what waves to simulate–with a number of methods, including stochastic wavegroups [11, 50], critical wavegroups [4–6, 57], equivalent waves [36], reduced order wavegroups [16, 20, 42], and Karhunen-Loève (KL) wave episodes [26]. At the same time, generalizing from simulation or experimental results to the recovery of steady state statistics requires the application of careful data-driven surrogate modeling [21, 53, 62]. Surrogate modeling, such as with Gaussian Process Regression (GPR) [52, 59], additionally allows the deployment of techniques from the optimal experimental design literature [13, 30, 51], where initial simulation or experimental results are used to improve subsequent designs [9, 10, 42, 54, 55]. While optimal experiment design techniques are not yet widely used in offshore wind applications, some recent studies have focused on developing reduced order models (surrogate models) for studying the support structure characteristics [41] and dynamic response [60, 69] of offshore wind turbines.

In this work, we construct a surrogate model that is able to evaluate the hydrodynamic loads on the monopile foundation of an offshore wind turbine at selected sea states. The surrogate

model is constructed using the GPR machine learning technique. The model learns to map the wave episode–hydrodynamic force relationship from CFD simulation data. In each simulation, the monopile foundation interacts with an irregular wave episode and the hydrodynamic force time series is recorded. However, to extend the GPR technique designed to map scalar quantities into one useful for time series, we emphasize dimension reduction techniques for the wave and force time series: a KL procedure for the prescribed wave episodes, which simultaneously allows for a low dimensional parametrization and acceptable statistical recovery, and Principle Component Analysis (PCA) for the simulated forces, for which no statistics are known a priori. Our surrogate model is created by leveraging simulation data from multiple sea states: one quiescent and one extreme. The irregular wave-structure interaction simulations are implemented in the open source CFD code OpenFOAM, which is able to capture strong nonlinear hydrodynamic effects (e.g. wave breaking and slamming loads).

We structure this paper as follows. In section 2, we briefly recapitulate the design of KL wave episodes first described in [26]. In section 3, we describe the techniques we use to build the surrogate model: GPR, active sampling, and spectrum transfer. In section 4, we describe our CFD simulation design in the OpenFOAM software, and describe some of the challenges we faced adapting this approach to nonlinear waves. In particular, we describe both how we account for nonlinear dynamics when the KL wave episodes have Gaussian statistics, as well our approach to generating steady state validation data. Finally, in section 5, we present results for a quiescent sea state (significant wave height $H_s = 5$ m) and an extreme sea state ($H_s = 13$ m). Additionally, we include Appendix A with details about the CFD simulations setup and the characteristics of examined wave episodes.

# 2 Irregular Wave Episodes

Our goal is to estimate the hydrodynamic load statistics for an offshore wind turbine monopile foundation subjected to a selected sea state. Traditionally, a sea state (i.e., steady state) is approximated as an irregular wave record with duration from 30 minutes to 3 hours. To avoid the enormous computational cost involved with CFD simulations of this duration, we instead perform a number of shorter simulations with carefully selected irregular wave episodes, which are defined from the wave spectrum of the examined sea state. We initially focus on the selection of the irregular wave episodes. However, we cannot create a GPR surrogate model to learn the relationship between the wave episode time series and hydrodynamic force time series. Therefore, we need to represent the time series in a reduced order form.

We follow Guth and Sapsis [26] for constructing irregular wave episodes using a KL basis and a parametrizing coefficient vector. The choice of basis corresponds to the choice of sea state, while the choice of coefficients corresponds to a particular sample from that sea state. In the remainder of this section, we will briefly recapitulate this method.

## 2.1 Wave episode representation

The sea surface elevation $x(\xi, t)$ is a stochastic process, which we assume to be zero mean, statistically stationary, and described by Gaussian statistics. Additionally, we assume that waves are long-crested (unidirectional). We use the JONSWAP spectrum [27] to describe the time series $x(\xi = 0, t)$ at a fixed spatial location:

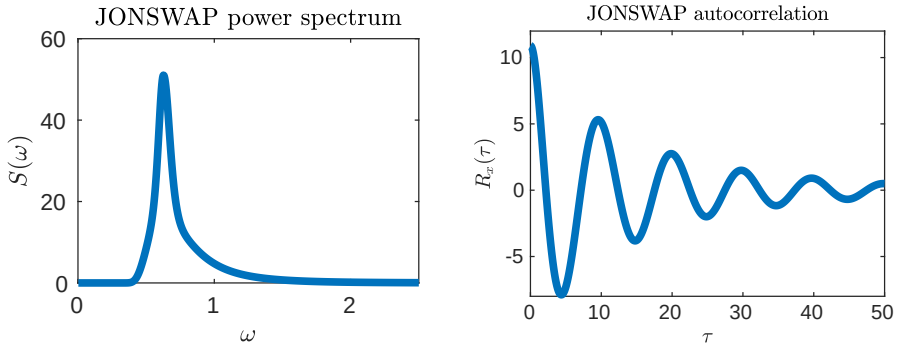


Figure 1: a) One-sided power spectral density for a JONSWAP spectrum with $H_s = 13$ meters. b) Corresponding JONSWAP time autocorrelation function.

$$S_J(\omega) = \frac{\alpha g^2}{\omega^5} \exp\left[-\frac{5}{4}\left(\frac{\omega_p}{\omega}\right)^4\right] \gamma^r \tag{1}$$
$$r = \exp\left[-\frac{(\omega-\omega_p)^2}{2\sigma^2\omega_p^2}\right]$$

where constants are given by

$$\alpha = 0.076\left(\frac{U_{10}^2}{Fg}\right)^{0.22}, \qquad \omega_p = 22\left(\frac{g^2}{U_{10}F}\right)^{\frac{1}{3}}, \quad \gamma = 3.0, \quad \sigma = \begin{cases} 0.07 & \omega \le \omega_p \\ 0.09 & \omega > \omega_p \end{cases},$$

and $F$ is the wind fetch, $g$ is the acceleration due to gravity, and $U_{10}$ is the wind speed at 10 m elevation. In this work, we consider two sea states, i.e., in both sea states the peak period is $T_p = \frac{2\pi}{\omega_p} = 8$ s, while the significant wave height $H_s = 5$ m in the mild sea state and $H_s = 13$ m in the extreme sea state. The option to examine a mild and an extreme sea state is taken to emphasize the nonlinear effects associated with high and steep waves.

Each irregular wave episode requires a distinct realization of this random process on a finite interval $[0, T]$, [4, 6, 57]. In the random phase method, the frequency space is discretized between $[\omega_{\min}, \omega_{\max}]$ (with spacing $\delta_\omega$), and the recovered signal is given by

$$x(t) = \sum_{i=1} a_i \cos(\omega_i t + \phi_i), \tag{2}$$

where $\omega_i$ is sampled uniformly between $[\omega_{i-1}, \omega_i]$, $\phi_i$ is a random phase drawn uniformly on $[0, 2\pi]$, and $a_i$ is coefficient given by

$$a_i = \sqrt{2S_J(\omega_i)\delta_\omega}. \tag{3}$$

However, the random phase model does not provide the low dimensional parametrization necessary to apply ML techniques. Instead, we follow Guth and Sapsis [26] in building a basis set corresponding to a particular sea state and interval. We present the KL theorem [34, 40]:

**Theorem 2.1** (Karhunen Loève)**.** *Consider the stochastic process $x(t)$ which is zero mean and square integrable on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Define the covariance function*

$$K_{xx}(s,t) = \mathbb{E}\left[x(s)x(t)\right], \tag{4}$$

*with corresponding integral operator over the interval $[0,T]$,*

$$T_{K_x}\phi(t) = \int_0^T K_x(t,s)\phi(t)ds, \quad t \in [0,T]. \tag{5}$$

*Then by Mercer's Theorem for every interval $[0,T]$ the operator $T_{K_x}$ has an orthonormal basis of eigenvectors $\{\hat{e}_{i,T}(t)\}$ and corresponding eigenvalues $\{\lambda_i\}$. Moreover, the coefficients*

$$\alpha_i = \int_0^T x(t)\hat{e}_{i,T}(t)dt \tag{6}$$

*are centered orthogonal random variables:*

$$\mathbb{E}[\alpha_i\alpha_j] = 0 \quad \textit{for} \quad i \neq j \quad \textit{and} \quad \textit{Var}(\alpha_i) = \mathbb{E}[Z_i^2] = \lambda_i. \tag{7}$$

*Furthermore, we can expand the random process $x(t)$ as*

$$x(t) = \sum_{i=1}^{\infty} \alpha_i \hat{e}_{i,T}(t), \quad t \in [0,T]. \tag{8}$$

In summary, the eigenvectors of the spatial covariance matrix of the sea surface form an orthonormal basis. The decomposition of $x(t)$ onto this basis produces a set of centered, orthogonal (in the random sense) coefficients. In particular, we can change back and forth between the function representation $x(t), t \in [0,T]$ and the coefficients representation $\alpha_i, i = 1, 2, ....$ In the literature, this procedure is also known as principle component analysis (PCA), proper orthogonal decomposition (POD) [22]. For this work, we will always refer to this general technique as KL.

We truncate the KL expansion in equation (8) in a finite number of modes, $n$. In this way, we may represent the stochastic process on the interval $[0,T]$ as an $n$-dimensional vector $\boldsymbol{\alpha}$ of KL coefficients, each component of which is an orthogonal random variable with variance $\lambda_i$. Each irregular wave episode corresponds to a distinct choice of coefficients.

Finally, we note that we stochastically extend the wave episode outside of the interval $[0,T]$ by use of a Gaussian process extrapolation technique [26]. This is important for physical realization of the sea surface elevation over an extended domain, and for the initialization phase of the OpenFOAM simulation.

## 2.2 Wave episode generation

Applying the KL expansion, we represent each wave episode as a truncated series of KL mode coefficients, the vector $\boldsymbol{\alpha}$. The truncation order, $n$, determines both what fraction of energy included in the wave episode is captured by the KL reduced order wave, as well as the dimensionality of the reduced wave episode. For this work, we chose a truncation order of $n = 3$ to balance fidelity against dimensionality, and in section 3.2 we describe our method for choosing the vector $\boldsymbol{\alpha}$.

We follow the same technique described in Guth and Sapsis [26] to convert from the coefficient vector $\boldsymbol{\alpha}$ to the OpenFOAM input. As we describe in section A.3 of Appendix A, the OpenFOAM

wave maker requires a series of linear waves with prescribed amplitude, period, and phase. These prescribed wave components are fixed on the interval $[20, 52]$ s, where $T = 32$ s is the KL interval length and $T_{\text{pre}} = 20$ s is the numerical initialization time associated with the ramp time in OpenFOAM simulations.

Then, we extrapolate the signal onto the interval $[0, 20]$ using the stochastic prelude technique described in Guth and Sapsis [26] that allows for statistically-consistent extrapolation. Specifically, this technique generates encounter conditions, which are statistically consistent with the sea spectrum but also transition smoothly to the prescribed wave episode. Finally, we perform a discrete Fourier transform (DFT) on this extended time series, and retain the $N = 90$ Fourier components with greatest magnitude. This DFT truncation order was chosen to balance simulation costs against reconstruction fidelity.

To summarize, OpenFOAM simulates the wave episodes for 52 s, where the first 20 s are the ramp time and the rest is the real simulation time. We note that during the later data analysis, we drop the prelude and shift the time scale forward by $T_{\text{pre}} = 20$ s so that our data records always correspond to the interval $[0, 32]$ s.

## 2.3 Basic wave parameters

The wave steepness, $s_w$, for the irregular waves is defined as [3]:

$$s_w = \frac{2\pi H_s}{g T_p^2}, \tag{9}$$

where $H_s$ is the significant wave height and $T_p$ is the peak wave period. In this study, each wave episode has duration 32 s which translates to approximately 4 wave cycles. To estimate the wave steepness for each wave episode, we consider the maximum wave height in the wave record.

The Keulegan-Carpenter (KC) number describes the relative importance of drag forces over the inertia forces exerted on the monopile foundation under the irregular wave episode, and it is defined as:

$$KC = \frac{2\pi\eta}{D}, \tag{10}$$

where $\eta$ is the maximum wave elevation at the position of the monopile foundation, which is obtained from the empty wave tank simulations (without the presence of the structure), and $D$ is the monopile foundation diameter. The KC number indicates the flow separation around the structure, i.e., for high KC number the drag force dominates resulting to flow separation, vortex shedding and increased loading on the structure. In our study, the wave episodes obtained from the quiescent sea state have KC $< 4.5$ (except the wave episode 451, KC $= 7.3$). For the extreme sea state wave episodes, $6.4 <$ KC $< 12.3$. In general, the critical region starts for KC $> 7$, since wave breaking and slamming effects start to occur [39]. For each wave episode in this study, the wave steepness and the Keulegan-Carpenter number are listed in the Appendix A (Table 2, Table 3,Table 4).

## 2.4 Breaking wave criterion

Before performing the CFD simulations, we want to have a priori-knowledge whether the analytical irregula wave episodes appear breaking wave behavior. They are assessed in relation to the breaking wave criterion [24], which is defined by:

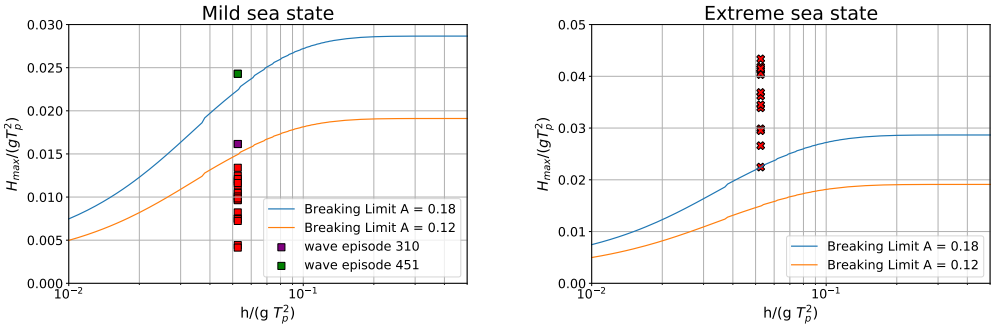


Figure 2: The analytical wave episodes are located based on their wave breaking limit. **Left:** At the quiescent sea state, the majority of the wave episodes are below the breaking region (orange), except the wave episode 451 which is exceeds the maximum breaking limit (blue). The wave episode 310 is probable to break as it is above the minimum breaking limit. **Right:** In the extreme sea state, all the wave episodes are above the maximum breaking limit (blue), therefore, breaking wave phenomena is expected to occur.

$$\frac{H_b}{gT_p^2} = A\left(1 - \exp\left(-1.5\pi\frac{d}{L_p}\right)\right)\frac{\tanh(kd)}{2\pi}, \tag{11}$$

where $H_b$ is the breaking wave height, $T_p$ is the peak wave period, $L_p$ is the wave length, $k$ is the wave number, and $d$ is the water depth. For irregular waves, $0.12 < \mathrm{A} < 0.18$. The wave number, $k$ is estimated from the dispersion relation:

$$\omega_p^2 = k \cdot g \cdot \tanh(k \cdot d). \tag{12}$$

To employ the breaking wave criterion for the examined wave episodes, $T_p = 8$ s, $d = 33$ m, and the breaking wave height, $H_s$, is equal to the maximum wave height in the 32 s duration irregular wave episode. The values are summarized in Appendix A – Table 2, Table 3, Table 4. Figure 2 presents the wave breaking assessment of the analytical wave episodes. The left plot refers to the wave episodes obtained from the quiescent sea state. Most of the waves are below the breaking limit. Wave episodes 310 and 451 constitute the exception, with the former being just above the lower breaking limit and the latter being above the upper breaking limit thus it is very likely to break. The right plot refers to the wave episodes obtained from the extreme sea state, and they are all above the upper breaking limit, therefore, they are expected to break.

## 2.5 Hydrodynamic force representation

In OpenFOAM, while the structure is modeled to interact with the irregular wave episode, a number of kinematic and dynamic quantities are recorded. We choose the total *hydrodynamic force* in the $x$-direction as the primary quantity of interest. An important obstacle we encounter is that the total force is a *time series*. That is to say, we cannot describe the force-on-structure across the whole interval $[20, 52]$ s with a single scalar quantity. In order to fit the output into the optimal

experimental design framework, we first represent the time series as a low dimensional vector $\mathbf{q}$, and each component, $q_i$, of the vector is treated separately. Specifically, we employ the following reduced order representation for the force:

$$F_x(t) = \sum_{i=1}^{n_{out}} q_i \hat{\mu}_{i,T}(t), \quad t \in [0, T]. \tag{13}$$

where the $q_i$ are the reduced order coefficients and the $\hat{\mu}_{i,T}(t)$ are the force modes. We represent each $q_i$ as a separate output component. We use $n_{out} = 12$ modes to describe the force, chosen to recover more than 99% of the signal energy. For this problem, we compute the output modes by applying a separate KL expansion to the training data ( [26]), a set of force-on-structure time series associated with the particular wave episodes we simulated.

# 3 Surrogate Modeling with Gaussian Process Regression

## 3.1 Gaussian process regression

Here we follow the Gaussian process surrogate construction, i.e. we formulate a Gaussian process surrogate model to estimate the unknown function. For an introduction to Gaussian process theory, see Rasmussen and Williams [52]. Briefly, the Gaussian process constructs a normal distribution for every choice of $\boldsymbol{\alpha}$ given by

$$p(q_i \mid \boldsymbol{\alpha}) \sim \mathcal{N}\left(\mu(\boldsymbol{\alpha}),\ \sigma(\boldsymbol{\alpha})\right), \tag{14}$$

where,

$$\mu(\boldsymbol{\alpha}) = K_*^\intercal (K + \sigma_n^2 I)^{-1} Q \tag{15}$$

$$\sigma^2(\boldsymbol{\alpha}) = K_{**} - K_*^\intercal (K + \sigma_n^2 I)^{-1} K_*, \tag{16}$$

$\mu(\vec{\alpha})$ is the posterior mean, $\sigma^2(\vec{\alpha})$ is the posterior variance, $K$, $K_*$, and $K_{**}$ are kernel matrics, and $Q$ is the observed output data vector.

For our application, we use a squared exponential kernel with automatic relevance determination (ARD), given by

$$k_{SE}^{ARD}(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2) = \sigma_k^2 \exp\left(-\frac{1}{2}\sum_i^{n_\alpha} \frac{(\alpha_{1,i} - \alpha_{2,i})^2}{l_i^2}\right) = \sigma^2 \exp\left(-\frac{1}{2}(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2)^\intercal M (\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2)\right), \tag{17}$$

where $M = \mathrm{diag}(l_i)$. The kernel hyperparameter $\sigma_k$ is a measure of variability due to choice of $\boldsymbol{\alpha}$, and the $l_i$ are a measure of the length scale associated with input dimension. The final hyperparameter, not associated with the kernel but part of the GPR scheme, is the aleatoric uncertainty $\sigma_n^2$ (also referred as irreducible uncertainty), sometimes interpreted as a regularization parameter as in ridge regression.

In equation (13), we give a representation of the output time series $F_x(t)$ as weighted sum of basis vectors. We use this representation to build a separate surrogate for each output mode as a function of the wave episode vector $\boldsymbol{\alpha}$. That is, for each $i \in [1, n_{\text{out}}]$, we build a distinct Gaussian process to estimate a posterior distribution

$$q_i(\boldsymbol{\alpha}) \sim \mathcal{N}(\mu_i(\boldsymbol{\alpha}), \sigma_i(\boldsymbol{\alpha})). \tag{18}$$

This collection of independent scalar surrogates allows us to model a low dimensional projection of the the full time series $F_x(t|\boldsymbol{\alpha})$. We can sample from the combined surrogate by sampling each $q_i$ individually from the Gaussian process posteriors, and then combining the coefficients using equation (13)

### 3.2 Active sampling of training data

In order to build the surrogate models, we require a set of training data–representative wave episodes, and the associated force-on-structure time series calculated with OpenFOAM. The training data should be representative of both quiescent and extreme wave episodes. We build our training set through two steps.

First, we uniformly sample wave episodes from the space of possible wave episodes. We perform this step by using Latin Hypercube sampling from a hyper box with side length $D = 2z^*\sqrt{\lambda_i}$, where $\sqrt{\lambda_i}$ is the KL length scale associated with the $i^{\text{th}}$ component of $\boldsymbol{\alpha}$, and $z^* = 3$ is a control parameter that balances between extreme and quiescent waves. We choose Latin hypercube sampling in order to avoid the 'clumps and voids' that are associated with independent sampling. Figure 8 (Right) also shows the spectral decay associated with those modes. We indeed see a major drop after the the first 3 directly excited modes–the higher order output modes represent memory effects and nonlinear effects.

Second, we make use of the initial uniform data points in order to optimally choose subsequent training data samples. Active sampling, or Bayesian Optimal Experimental Design, is a method to use existing data, along with a surrogate model and an acquisition function, to design later experiments. The new training data can be added to refine the intermediate model in a loop [9,10,67].

For this problem, we use Likelihood Weighted Uncertainty Sampling (LW-US) an acquisition function developed in [10, 54] to preferentially sample inputs likely to contribute to the output statistical tails – rare events. The functional form of the acquisition function is given by:

$$u_{LW-US}(\boldsymbol{\alpha}) = \frac{p_A(\boldsymbol{\alpha})}{p_q(q(\boldsymbol{\alpha}))}\sigma_q^2(\boldsymbol{\alpha}) \tag{19}$$

where $p_A(\boldsymbol{\alpha})$ is the probability of the encountering the wave episode in the steady state (from the KL construction), $p_q(q(\boldsymbol{\alpha}))$ is the likelihood associated with $q(\boldsymbol{\alpha})$ from the surrogate, and $\sigma_q^2(\boldsymbol{\alpha})$ is the posterior variance associated with the surrogate.

In the standard formulation, the LW-US acquisition function is designed for scalar outputs. Following [26], we make two changes when borrowing $u_{LW-US}(\boldsymbol{\alpha})$. First, because we represent time series using a set of multiple output modes, we associate with each mode $q_j$ a separate acquisition function $u_{LW-US,j}(\boldsymbol{\alpha})$. During the active sampling loop, we iterate through these $j$ output modes in a *round robin* format, selecting one experimental design for each. Second, because there is non-negligible intrinsic noise associated with this system, we perform uncertainty regularization. For this technique, we first estimate the intrinsic noise associated with the surrogate model. For a Gaussian process, this is simply the hyperparameter $\sigma_n^2$. Then, we replace the Gaussian process posterior variance $\sigma_q^2(\boldsymbol{\alpha})$ with the epistemic variance $\left(\sigma_q^2(\boldsymbol{\alpha}) - \sigma_n^2\right)$, i.e. total variance reduced by the aleatoric variance. This regularization step improves acquisition function performance in weighted uncertainty sampling schemes when the true function has significant aleatoric variance.

Taken together, the final acquisition function we use for each output component, $q_j$, is given by,

$$u_{LW-US,j}(\boldsymbol{\alpha}) = \frac{p_{\boldsymbol{\alpha}}(\boldsymbol{\alpha})}{p_{q_j}(q_j(\boldsymbol{\alpha}))}\left(\sigma^2_{q_j}(\boldsymbol{\alpha}) - \sigma^2_{n,j}\right). \tag{20}$$

### 3.3 Spectrum transfer

One of the advantages of wave-episode sampling is that the pairs of wave-episodes and hydrodynamic-responses can be used to quantify the statistical responses of different spectra, assuming that these have the same modal period and have similar shapes [26]. For the present context assume we have input-output pairs $\boldsymbol{\alpha}_i, \boldsymbol{q}_i$ associated with significant wave height $H_{s,1}$ and we want to use this information for quantifying statistics for another significant wave height $H_{s,2}$.

We follow the spectrum transfer technique described in Guth and Sapsis [26]. First, we calculate the KL basis for each spectrum, $S_1$ and $S_2$. Second, we calculate the eigenvalue ratios

$$r_j = \sqrt{\frac{\lambda_j^{(1)}}{\lambda_j^{(2)}}}. \tag{21}$$

Finally, we adjust the $\boldsymbol{\alpha}$ from $J^{(1)}$ to $J^{(2)}$ by rescaling

$$\alpha_j^{(1)} = \alpha_j^{(2)} r_j \tag{22}$$

Through this energy-based rescaling of the KL coefficients we are able to use the same wave-episode data to build surrogate models corresponding to different sea states. We caution, however, that the initial data set, associated with waves having significant height $H_1$ has poor skill on describing forces associated with waves of $H_2$, if the latter is significantly larger. This is not a surprise given that in the low-energy spectrum only small amplitude waves are needed to capture the overall statistics.

## 4 High-Fidelity CFD Modeling

High-fidelity CFD simulations are employed to provide the hydrodynamic performance of the monopile foundation interacting with the irregular wave episodes. Complex physical flow mechanisms, e.g. fluid viscosity, wave diffraction, radiation, wave overtopping and slamming can be captured by CFD modeling based on the solution of Navier-Stokes equation. The simulations performed with the open-source software OpenFOAM and further description of the numerical setup is available in Appendix A.

#### Wave induced force

The wave induced force on the structure, $\boldsymbol{F}$, is calculated by integrating the normal pressure, $p\hat{n}$, where $\hat{n}$ is unit normal matrix, and the tangential viscous stress vector, $\boldsymbol{\tau}$, over the surface, $A$, of the structure.

$$\boldsymbol{F} = \iint_A (p\hat{n} + \boldsymbol{\tau}) dA, \tag{23}$$

In our application, the horizontal component of the force, which is known as the *inline force*, $F_x$, has the major contribution thus it is chosen as the quantity of interest in this study.

### Numerical wave calibration

In OpenFOAM simulations, to produce the *target* irregular wave episode (e.g., analytical) at a specific location in the numerical wave tank (e.g. where the monopile foundation will be placed), we follow a *wave calibration* procedure [17]. Figure 3 shows the schematic depiction of the wave calibration which is similar to the one presented in [66] and previously has been applied by the authors in [35, 37].

The target wave episode and the numerical wave obtained from the last iteration are compared through their *density wave spectrum*. Figure 4 compares four representative wave episodes. For the waves obtained from the quiescent sea state (i.e., 404 and 403), the analytical and numerical spectrum match very well–in terms of frequency and amplitude. Conversely, the two spectrum corresponding to extreme waves (i.e., 503 and 504) present significant divergence. In general, due to complex nonlinear hydrodynamic phenomena related to high and steep waves, the numerical wave episodes 501-514 cannot match the target wave profile. In high KC numbers (see section 2.3), the real wave propagation breaks down the theoretical approach. In practice, the nonlinear interaction between the wave components leads to breaking wave phenomena, which may contribute to changing wave profile, and thus difficulty to match with the analytical solution anymore.

In Figure 2, we compare the analytical wave episodes to the two estimates of the breaking limit – defined from the breaking wave criterion (section 2.4). We observe that, for the quiescent sea state ($H_s = 5$ m), all but one of the waves are firmly below the breaking limit. However, for the extreme sea state ($H_s = 13$ m), nearly all of the waves are above the breaking limit. In Figure 5 and Figure 6, we provide an OpenFOAM visualization of the breaking wave evolution and its interaction with the monopile foundation, respectively.

### Ground truth solution

To evaluate the accuracy of the GPR surrogate model, we compare the model prediction with a reference solution–i.e., the ground truth. For the purpose of this study, the ground truth is the statistics (i.e., pdf) of the hydrodynamic force-on-structure during the steady state realization (i.e., the full sea state realization). We calculate the ground truth force statistics for the quiescent ($H_s = 5$ m) and extreme sea state ($H_s = 13$ m) by simulating the wave–structure interaction across 30 min and 60 min, respectively.

In CFD simulations, because of the computationally expensive cost as explained in section 5.2, the wave–structure interaction is usually modeled for a short time period (a few minutes). In the case of steady state realization, in order to accelerate the numerical simulation process and avoid possible numerical instabilities due to running such a long simulation, we split the 30 min and 60 min sea state into 12 and 24 intervals, respectively. Each interval constitutes a 180 s irregular wave episode. However, from each interval we discard the first 20 s which is the ramp time that the wave train needs to reach the desired profile (transient), and thus 160 s of wave-structure interaction are retained from each interval, as explained in section 2.2. The intervals are then merged to reconstruct a representative long time-record for the steady state of the particular sea state. Our application studies the interaction of the irregular wave episode with the monopile foundation of an offshore wind turbine. For this application the dynamic memory effects are not so critical as the monopile is a fixed structure. That is to say, the reconstruction of the full sea state by merging short-time intervals does not affect significantly the solution. Instead, for a floating structure whose current position is dependent on the previous dynamics, more care must be taken when reconstructing long time statistics from shorter interval simulations.

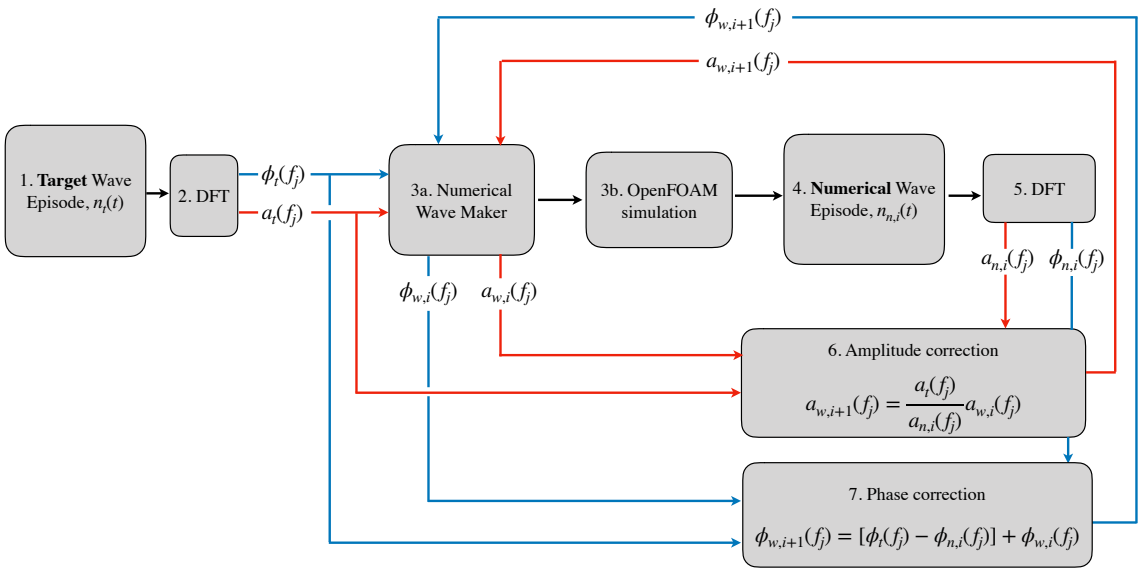


Figure 3: An overview of the numerical wave calibration procedure. 1) Definition of the *target* irregular wave, $n_t(t)$, 2) which is analyzed into *sinusoidal wave components* via DFT. For each frequency, $f_j$, the corresponding amplitude, $a_t(f_j)$, and phase, $\phi_t(f_j)$, are obtained. 3a) To numerically reproduce the target wave episode, the OpenFOAM *numerical wave maker* should receive the amplitude and phase for each frequency component. In the first iteration, the components of the target wave are the inputs to the numerical wave maker, however, these values are corrected at every iteration, i.e., $a_{w,i}(f_j) \rightarrow a_{w,i+1}(f_j)$ and $\phi_{w,i}(f_j) \rightarrow \phi_{w,i+1}(f_j)$. 3b) The OpenFOAM simulation runs and 4) the *numerical* wave episode, $n_n(t)$, is recorded at a selected location in the numerical wave tank. 5) Similarly to target wave, the numerical wave is analyzed into sinusoidal wave components via DFT. For each frequency, $f_j$, the numerical amplitude, $a_{n,i}(f_j)$, and phase, $\phi_{n,i}(f_j)$, are obtained and compared to the corresponding values from the target wave. 6) An *amplitude correction factor*, which is the ratio of the target to numerical wave amplitude, is applied to the new wave maker inputs. 7) A *phase correction factor*, which is the difference between target and numerical phase, is also applied to the new wave maker input. Steps (3)–(7) are repeated either until numerical and target waves are converged or further improvement is not succeeded.

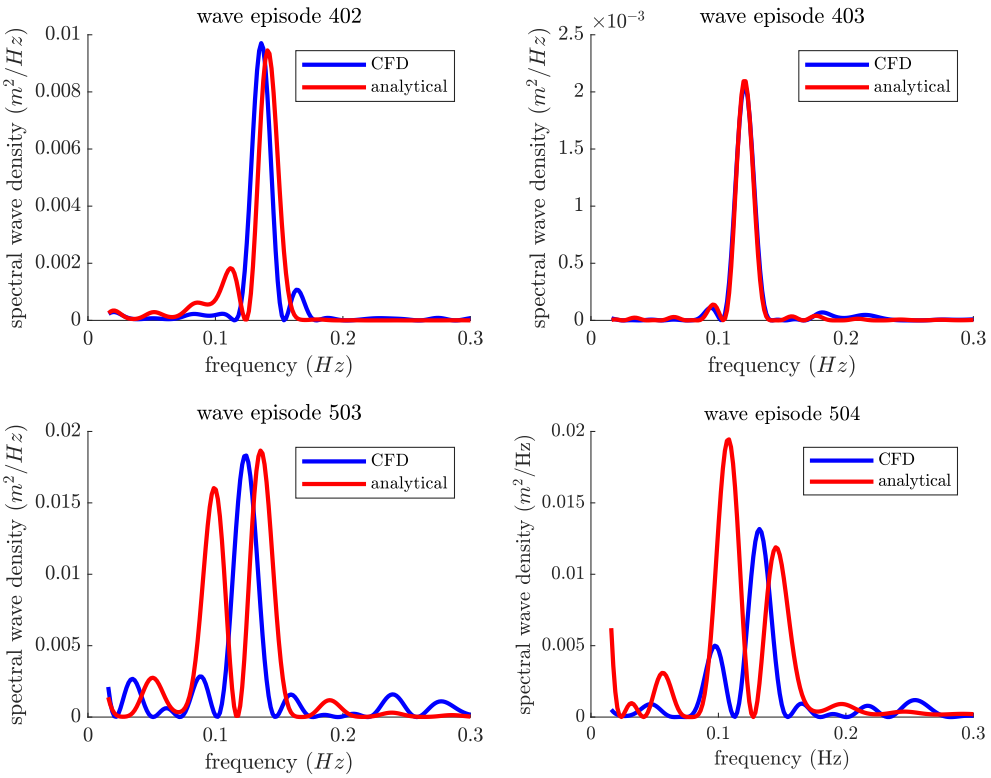


Figure 4: Numerical (CFD) and analytical wave spectral density over frequencies for the sea states with significant wave height, $H_s = 5$ m (wave episodes 402 and 403), and $H_s = 13$ m (wave episodes 503 and 504).

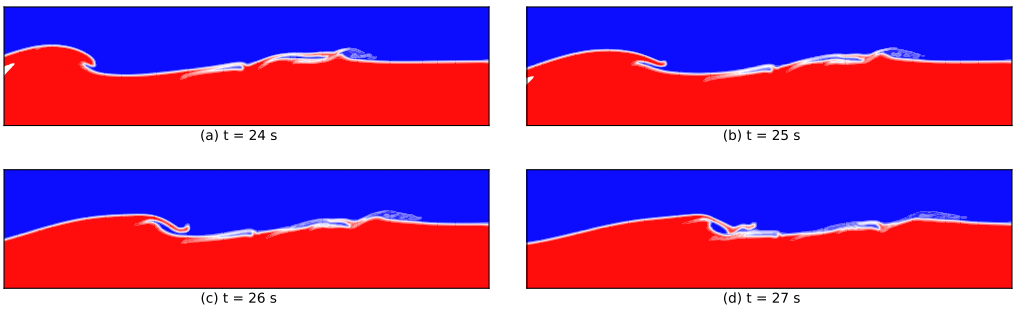


Figure 5: Wave episode 508: Schematic depiction of the wave breaking process in the empty numerical wave tank (without the presence of the monopile foundation).

In order to reproduce the short wave–structure interaction intervals, we use the *random phase model* (discussed in section 2, in equations (2) and (3)) to generate random waves from the examined sea state. The sea state is characterized by the significant wave height $H_s$, period $T_p$, and is described by the JONSWAP wave spectrum. For this procedure, we discretize the wave spectrum into $N = 90$ wave components with angular frequency equally spaced in the range $[\omega_{\min},\ \omega_{\max}] = [0.44,\ \ 2.26]$ rad/s.

## 4.1 Surrogate Model of the Monopile Foundation

We model the distortion introduced between the target (analytical) wave episode and the numerical wave episode as a nonlinear mapping. The numerical solver along with the boundary conditions and absorption, implicitly define the map, $f$, between the target and numerical wave episodes (this is the wave calibration from section 4). In Figure 4, we display the wave spectral density comparison of the analytical wave (in black) and OpenFOAM realization (in red). The numerical solver also defines the mapping $g$ from the numerical wave episode to the numerical force on the structure.

When we use a machine learning technique to construct a surrogate model from the target wave episode (parametrized by the coefficient vector $\boldsymbol{\alpha}$), we are learning the composed function $h = g \circ f$. This is *not* the same as $g$, which maps the numerical waves to numerical forces, unless the OpenFOAM realization perfectly matches the target wave (i.e., $f(\cdot)$ is very nearly the identity). In Figure 7, we show a schematic depiction the relationship between the target and OpenFOAM wave realization, the resulting force on the structure as estimated from OpenFOAM, and what the surrogate learns to map.

This shift is a problem for training a model that converges *pointwise*. That is to say, if we want the precise structural loads corresponding to the target wave episode (that we parametrized with $\vec{\alpha}$), the OpenFOAM solution for the force-on-structure does not correspond to the desired wave. However, if instead our goal is to train a model that merely calculates the correct steady state statistics, including nonlinearities from real physics, the developed surrogate model successfully achieves this goal.

How can we recover the correct statistics without a precise model? If, in Figure 7 the function $f(\cdot)$ does not change the *statistics* of the steady state, then the machine learning model may still learn the statistics of steady state forces on structure. One straightforward case where this could be

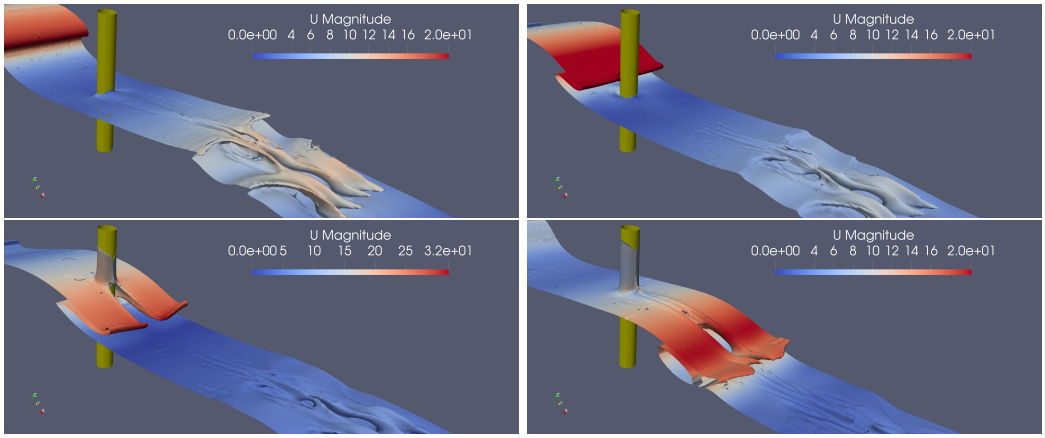


Figure 6: Isometric views of the wave - structure interaction (wave episode 508). The colormap expresses the velocity magnitude. **Upper Left:** Initiation of breaking wave at t = 23 s, **Upper Right:** Breaking wave hits the monopile foundation at t = 24.5 s, **Lower Left:** Wave run-up on the foundation at t = 25.5 s, **Lower Right:** Second-order effects around the monopile at t = 28 s.

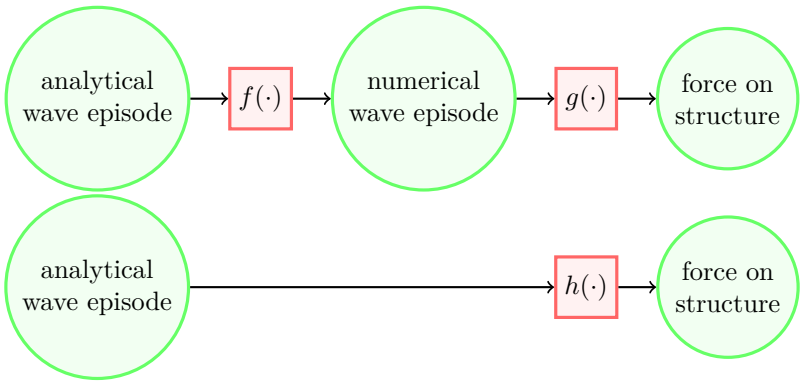


Figure 7: Flow chart showing the procedure to estimate the hydrodynamic force followed by (top) OpenFOAM simulation and (bottom) GPR surrogate model. In OpenFOAM simulations, the numerical wave may differ from the analytical wave due to the strong nonlinear hydrodynamic phenomena. The forces are estimated corresponding to the numerical wave. The surrogate model provides the map between the analytic wave episode and numerical force.

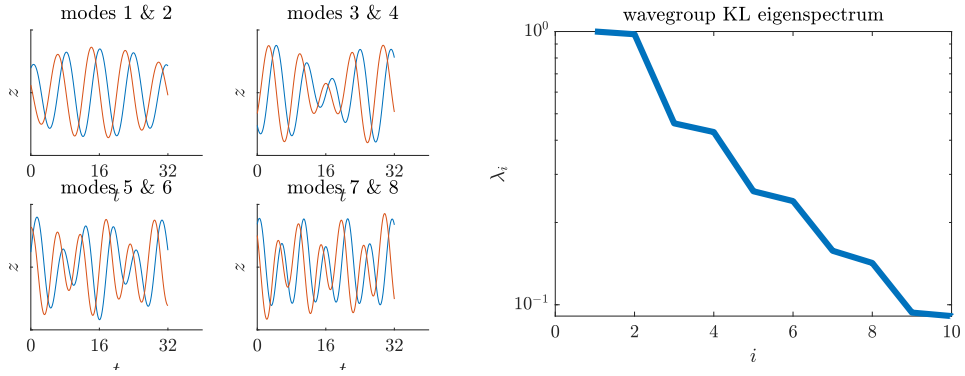


Figure 8: **Left**: KL mode shapes for the wave episodes, displayed in pairs due to approximate even/odd symmetries. Note that due to our normalization convention, the y-axis is arbitrary. **Right**: KL spectrum decay. X-axis counts the number of KL coefficient. The plots account for the quiescent sea state with wave height $H_s = 5$ m. For extreme sea state $H_s = 13$ m the plots are identical except for scale.

so is if $f(\cdot)$ were simply the forward time evolution operator. In this ideal case, our surrogate model for $h(\cdot) = g \circ f(\cdot)$ might still learn the correct statistics of forces, even if it incorrectly predicts the force associated with any particular wave episode.

Here the deviation of the OpenFOAM realization from the target wave episode is likely caused by nonlinear effects. Our target wave episode model assumes linear wave theory, which begins to break down in the extreme, steep waves (due to wave breaking, overtopping and other higher order hydrodynamic effects) that lead to extreme forces on structure (slamming effects - especially for the sea state 13 m).

This form of realization error is statistically important, because peak sharpening and trough broadening are likely to have some impact on the distribution of extreme loads. However, this effect is likely to *counteract* an earlier modeling error of (incorrectly) assuming that the extreme steep waves we examine are well modeled by linear wave theory. Taken as a whole, we cannot say that the divergence between target and numerical wave episode does not impact the expected results. However, we expect this effect to be less significant for statistical reconstruction relative to the recovery of specific force time series.

# 5 Numerical Results

## 5.1 Irregular wave episodes dimension reduction

In this study, we choose sea states described by the JONSWAP wave spectrum and parameters $H_s \in \{5m, 13m\}$ and $T_p = 8$ s. For the KL dimension reduction of the irregular wave episode, we choose $T = 32$ s and $n = 3$ distinct non-zero wave episode coefficients. Figure 8 (Left) displays the mode shapes for the KL basis, while Figure 8 (Right) displays the decay of eigenspectrum with the increasing number of KL modes, which practically confirms that higher KL modes carry negligible amount of energy compared the first KL modes. We note that in this study we choose $n = 3$, therefore, we only use the first three KL modes shown in the Figure.

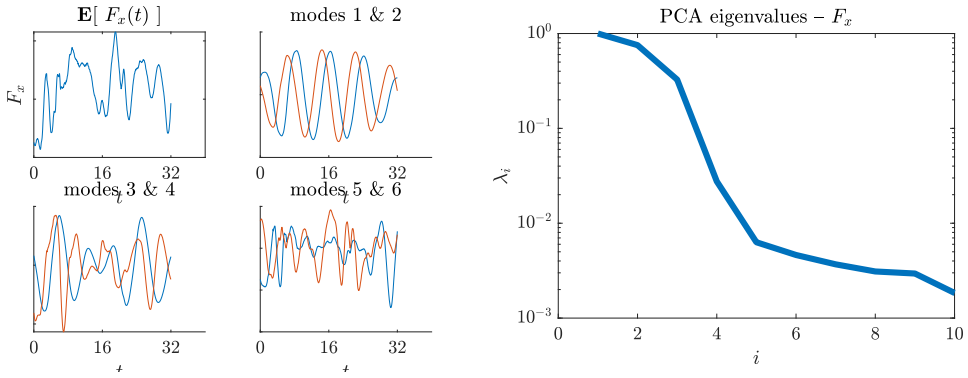


Figure 9: For quiescent sea state with $H_s = 5$ m. **Left**: PCA modes for the axial hydrodynamic force, $F_x(t)$. **Right**: PCA eigenspectrum for $F_x(t)$. X-axis denotes the number of PCA modes.

We construct the irregular wave episodes by drawing Latin hypercube samples from a hyperbox with size parameter $z^* = 3$ from which we choose the KL coefficients $\{\alpha_1, \alpha_2, \alpha_3\}$. The resulting irregular wave episode is decomposed via DFT into $N = 90$ distinct sinusoids which are the inputs to the OpenFOAM wave maker. The numerical wave is simulated for 52 s, i.e., the $T = 32$ s wave episode, along with a $T_{\text{pre}} = 20$ s ramp-up period to reduce the transients.

Finally, the GPR surrogate model is trained on irregular wave episodes of 32 s duration, which are obtained from a mild and an extreme sea state. In particular, from the wave spectrum of the sea state with $H_s = 5$ m, we initially generate 16 training irregular wave realizations, selected via Latin hypercube sampling, which are followed by 4 additional training irregular wave episodes using active sampling. Subsequently, we use the spectrum transfer technique described in Section 3.3 to adjust our existing training wave data to the spectrum of the sea spectrum with $H_s = 13$ m. From the extreme sea state, we collect an additional 9 irregular wave realizations using active sampling.

## 5.2 Computational Resources

In this study, the CFD simulations are performed on Tetralith HPC cluster utilizing 128 processors per simulation. Each irregular wave episode, used for training the surrogate model, is simulated to interact with the monopile foundation for 52 s. This simulation requires approximately 12 h - 18 h real time. The steady state (or full sea state) simulations for sea states with $H_s = 5$ m and $H_s = 13$ m model the wave–structure interaction for 30 min and 60 min, respectively. As mentioned in section 4, this time is split into 12 and 24 intervals, respectively, which has the advantage that the shorter interval simulations can run in parallel, and thus the modeling of the full sea state is accelerated. Each interval models 180 s of wave–structure interaction and requires 40 h - 45 h of real time on the HPC cluster.

## 5.3 Hydrodynamic force dimension reduction

Figure 9 (Left) shows the mean and the PCA modes associated with the axial hydrodynamic force time series, $F_x(t)$, in the quiscent sea state with $H_m = 5$ m. We note that, because input dimensionality is $n = 3$, based on Wiener–Khinchin theory we expect at least 3 significant PCA

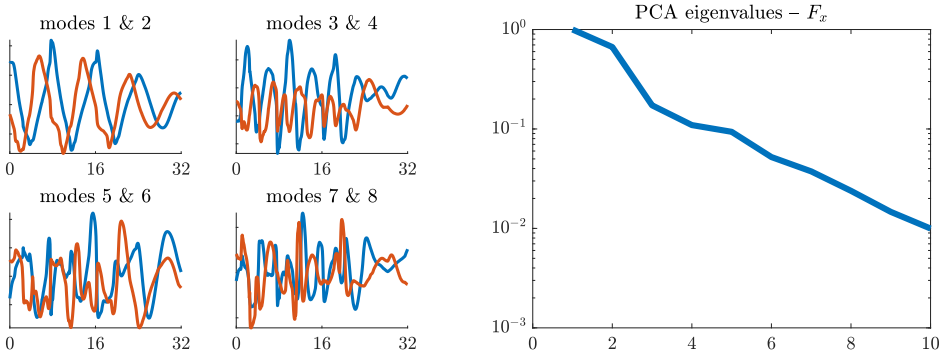


Figure 10: For extreme sea state with $H_s = 13$ m. **Left**: PCA modes for the axial hydrodynamic force, $F_x(t)$. **Right**: PCA eigenspectrum for $F_x(t)$. X-axis denotes the number of PCA modes.

modes to describe the output. Figure 9 (Right) also shows the spectral decay associated with those modes. We indeed see a major drop after the first 3 directly excited modes – the higher order output modes represent memory effects and nonlinear effects.

Figure 10 shows the equivalent information for the extreme sea state with $H_m = 13$ m. The primary difference we notice is that the eigenspectrum decay is slower, corresponding to more energy at higher PCA modes, because extreme waves contain more nonlinear behavior. Figure 11 shows contour slices for the KL-GPR surrogate for different combinations of PCA modes. The contours show the regions where the peak hydrodynamic force presents maxima and minima values. This is helpful for engineers in the design stage to extract information about how the waves (which are described by the KL coefficient vector $\boldsymbol{\alpha}$) impact the maximum hydrodynamic force. In other words, Figure 11 answers for which combination of $\alpha_i$ the force is maximized.

## 5.4 KL-GPR model predictions for the force time series and statistics

Figure 12 shows the KL-GPR-model predicted time series of the axial hydrodynamic force, $F_x(t)$, for wave episodes 401 - 404, which are used for the validation of the KL-GPR-based surrogate model (i.e., these wave episodes are not used for the construction of the surrogate model). The wave episodes 401 - 404 represent "large waves" associated with the $H_s = 5$ m sea state.

We note a few features here. First, the surrogate model presents higher uncertainty at the beginning of the interval. This is probably caused both by the encounter condition, which is stochastic, and by transients from initializing the numerical simulation. Second, we note that agreement is weakest near the extreme peaks, particularly in wave 404. This is likely because the nonlinear wave phenomena are most significant in this wave. Initially the KL-GPR model is trained on wave episodes from the quiescent sea state and their characteristics are listed in Table 2 Appendix A The training waves are characterized by smaller wave height, wave steepness and KC number compared to the corresponding values of the validation wave 402 and 404. The validation waves 401 and 403 are closer to the training waves. The characteristics of validation waves are listed in Table 3 Appendix A. This led to our decision to additionally train on extreme waves associated with the spectrum $H_s = 13$ m.

In Figure 13, we compare the steady state statistics for the quantity $F_x$ computed via steady state

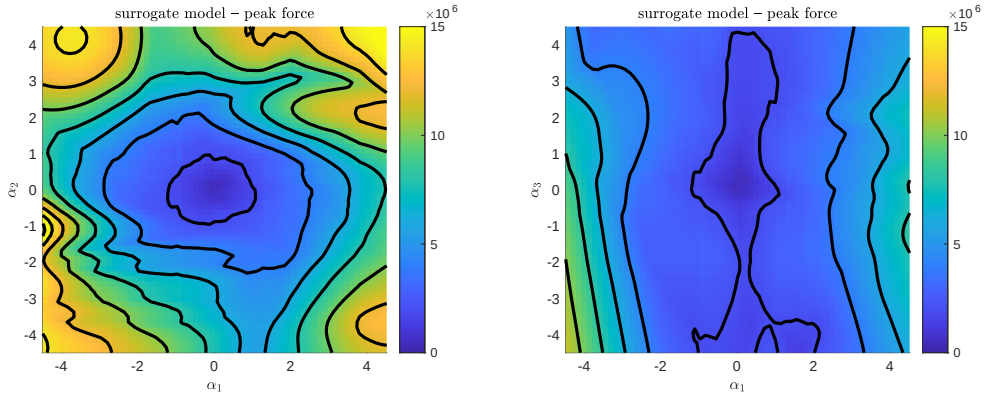


Figure 11: Maximum (unsigned) predicted force, $F_x$, of KL-GPR surrogate in $[N]$, for the extreme sea states with $H_s = 13$ m. The colormap shows the change in force for varying input waves, which are described by vector $\boldsymbol{\alpha}$. Left: $(\alpha_1, \alpha_2)$, $\alpha_3 = 0$. Right: $(\alpha_1, \alpha_3)$, $\alpha_2 = 0$.

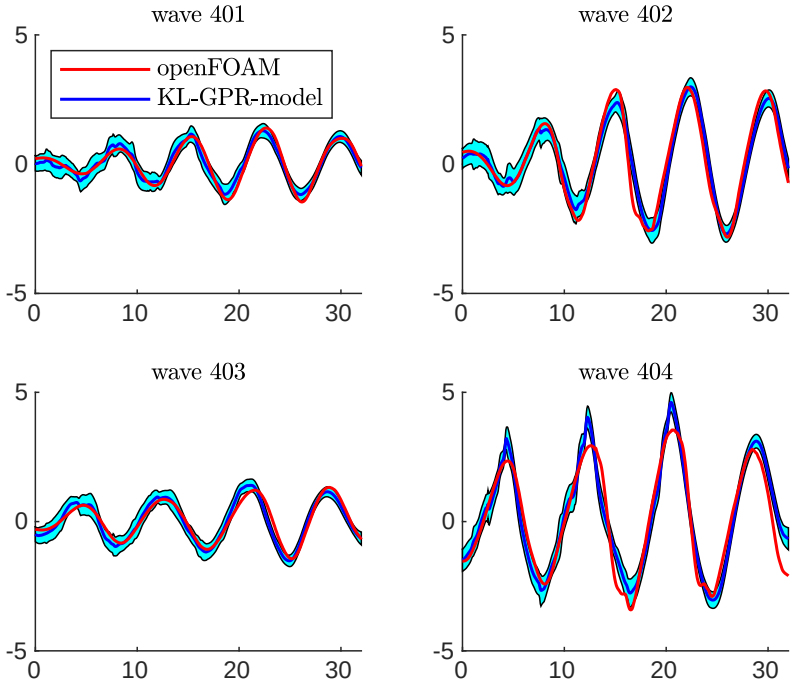


Figure 12: Comparisons of $F_x(t)$ from (red) OpenFoam and (blue) KL-GPR-model with posterior uncertainty, for quiescent sea state with $H_s = 5$ m. Y-axis counts $1 \times 10^5$ [N]. These waves are only used for validation purposes – not used during the model training.

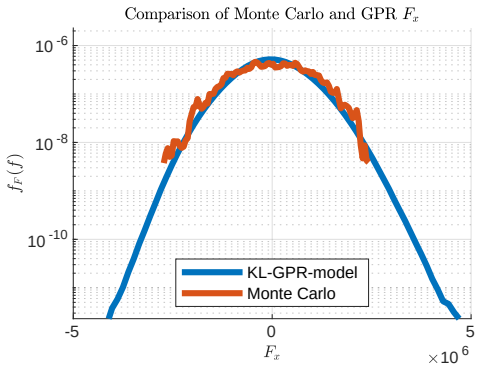


Figure 13: Comparison of the hydrodynamic force $F_x$ pdf predicted from KL-GPR-model, for $H_s = 5$ m.

sampling (red) and via surrogate reconstruction (blue). The steady state statistics are the simple ensemble statistics from the steady state simulations, while the surrogate statistics are generated by sampling a large number ($1 \times 10^5$) of wave episodes, and calculating the corresponding forces from the surrogate model. We estimate the pdf by concatenating each of these generated force time series, and then computing the time statistics via histogram. Alternately, we might have recovered the pdf using a kernel density estimator in place of simple histograms, but plots generated by histogram emphasize the statistical limit to our precision. We note reasonable agreement, up to the limit of statistical resolution. However, because the statistics for $F_x$ are very nearly Gaussian in the $H_s = 5$ m sea state, we chose to examine the statistical recovery of the hydrodynamic force for steeper waves.

In Figure 14 we show the recovered statistics for the extreme sea state with $H_s = 13$ m. We decompose the axial hydrodynamic force, $F_x$, into the pressure component, $F_x^P$, and the viscous component, $F_x^\nu$, which differ by nearly two orders of magnitude. Finally, we show the viscous contribution to the vertical hydrodynamic force, $F_z(t)$. For each dynamic quantity (force component), we train a separate surrogate model. We note that our recovery of the viscous forces, especially the $z$ component, are less accurate. This is in part due to our decision to prioritize the axial hydrodynamic force, $F_x$, in our active search acquisition function for samples selection, which in turn is dominated by the contribution from the pressure force, $F_x^P$.

In Figure 13 and Figure 14, the (signed) force statistics account for the hydrodynamic force direction which explains the asymmetry. We present signed statistics because, during the design phase, the structure may be constructed to resist positve and negative loads differently, for instance, by pre-loading. Through these plots, we conclude that the axial pressure force, $F_x^P$, has greater magnitude than the viscous effects due to the incident wave. Conversely, the viscous force is much more significant in the vertical direction , $z$. The pressure component of the vertical force is identically zero for this monopile design.

## 5.5 Effect of energy truncation

In this study we construct the wave episodes choosing a truncation order at $n = 3$. The energy decay associated with the number of KL coefficients are shown in Figure 8, and it is clear that a fraction

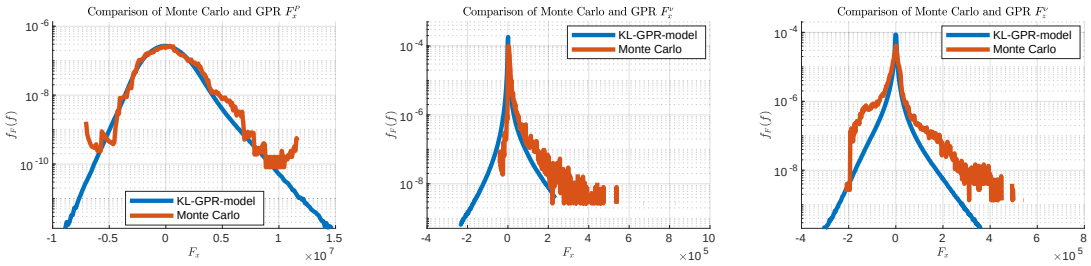


Figure 14: Comparison of KL-GPR-modeled pdfs of force components. **Left**: Pressure component $F_x^P$. **Middle**: Viscous component $F_x^\nu$. **Right**: Viscous component $F_z^\nu$.

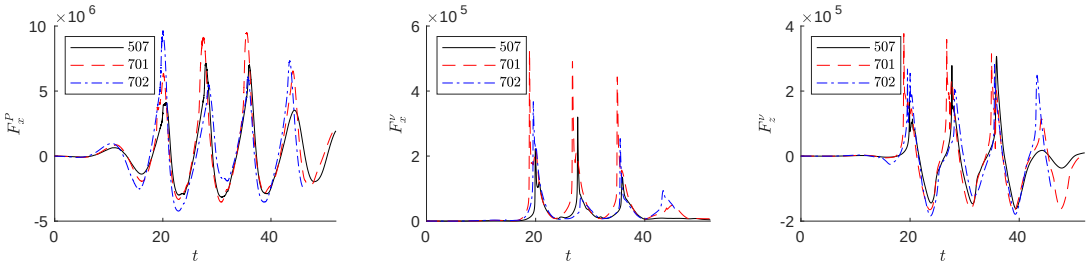


Figure 15: Comparison of truncated wave episode (507) with energy-corrected wave episodes (701 and 702). Wave episode 507 (black) is taken from our training set. Wave 701 (red) used the scaling coefficients correction, and wave 702 (blue) use the un-truncated coefficients correction. **Left**: Pressure component $F_x^P$. **Middle**: Viscous component $F_x^\nu$. **Right**: Viscous component $F_z^\nu$.

of the energy is truncated by choosing only the first three coefficients. At this point we examine whether this energy truncation effect the recovered force statistics.

We study the effects of energy truncation in the hydrodynamic force predicted from the KL-GPR surrogate model. We chose wave episode 507 to estimate the CFD axial and vertical force components, shown in Figure 15, for the basis of our comparison. This wave episode is constructed as previously described, i.e., with $n = 3$ nonzero components of the KL vector $\boldsymbol{\alpha}$, and a stochastic prelude. Importantly, our stochastic prelude technique means that any adjustment to wave episode 507 (other than global rescaling) will unavoidably add some aleatoric variance to a second OpenFOAM simulation, especially at the very beginning of the region of interest.

We consider two methods to correct for the effect of energy truncation. One option is to *scale the coefficient vector* $\boldsymbol{\alpha}$ by an energy factor. This would have the effect of shifting energy from high wave numbers to lower wave numbers. At the same time, by retaining the low dimensionality of $\boldsymbol{\alpha}$, this method maintains both the shape of the wave episode, and its easy applicability to machine learning. We constructed wave episode 701 by scaling the first three components of $\boldsymbol{\alpha}$ from wave episode 507. The corresponding CFD force to wave 701 is also shown in Figure 15. It is clear that each peak force on structure is significantly increased relative to wave episode 507. This increase is too large to be consistent with the steady state, and demonstrates that the distribution of wave energy across different scales is important on the force applied to the marine structure.

The second option is to simply avoid truncation of the higher order modes from the wave episode. That is to say, instead of a low order truncation $n = 3$, choose *a very high order truncation* (here, $n = 25$) to ensure that most of the wave energy is captured. This method has the advantage of modeling the waves with the highest fidelity. However, it has the disadvantage of increasing the dimensionality of the wave episode space, which makes the application of machine learning challenging. In Figure 15, we construct wave episode 702 by copying the first three components from $\boldsymbol{\alpha}$ from wave episode 507, and then drawing the remaining components ($k \in 4, ...25$) from the KL distribution. We can see that the first and fourth peak of the force on structure are significantly different between wave episodes 507 and 702, but the second and third peak are very similar. We explain this by noting that the first peak, at $t = 20$ s, is at the border of the stochastic prelude region and the wave-episode region. The divergence of this peak is well explained by the random draw of stochastic prelude. The significant difference for the fourth peak cannot be explained away so easily. By comparing with wave 702, and examining the viscous force, it is likely that that wave 507 has no significant wave crashing here. However, a small addition of energy, either in lower wave numbers or higher, was enough to cause a wave crash at the fourth peak.

In summary, correcting for the energy deficit by moving energy from high wave numbers to low wave numbers significantly and nonphysically overestimates the peak forces on the structure. However, correcting for the energy deficit by including the truncated KL modes increases the peak forces by a much less dramatic margin. This suggests that the KL truncation is likely responsible for some of the tail underestimate in Figure 14. We focus our optimal experimental design on the force component $F_x$, and in the quiescent sea states (with $H_s = 5$ m) the GPR model is able to accurately recover the load statistics with only a handful of wave episodes in the training set. In the extreme sea state (with $H_s = 13$ m), we need more data points to recover the heavier distribution right tail. When we examine other force components with more strongly non-Gaussian character, $F_x^\nu$ and $F_z^\nu$, (which are not considered during the active sampling phase) the KL-GPR model is less accurate. Finally, by studying how the energy truncation of the wave episodes effects our results, we determine that the truncated energy itself likely is not the cause of any remaining statistical error, nonlinear effects from the truncated shape complexity may play a subtle role in the distribution of

extreme loads.

# 6 Conclusions

We have presented a efficient computational framework for the statistical quantification of the force on a wind turbine monopile foundation caused by nonlinear interactions with irregular waves. The surrogate model is built using GPR, a data-driven machine learning technique combined with active learning and order reduction methods. To ensure that the developed model provides accurate predictions, we train with data from CFD simulations. CFD models are able to capture the nonlinear and complex phenomena in the wave–structure interaction, yet they command expensive computational costs and are thus unsuitable when massive simulations are required (e.g. uncertainty quantification, sensitivity analysis, statistical calculations for risk, etc).

As a adjunct to the active learned GPR model, our approach relies on two dimensionality reduction steps: to describe irregular wave episodes and to describe the force-on-structure time series. Specifically, we use the KL theorem to construct low dimensional wave episodes that nonetheless adequately represent the sea state associated with a specific power spectrum (i.e., JONSWAP). At the same time, we use KL to reduce the dimensionality of the force-on-structure time history. Together, these dimensionality reduction steps allow the construction of the efficient surrogate modeling with the GPR method, which otherwise has difficulty with the high dimensionality associated with directly modeling time series.

The wave episode samples are the inputs for the high-fidelity CFD simulations–the OpenFOAM code–which are able to capture nonlinear wave phenomena, such as breaking wave and slamming loads. While important for accurately resolving the interactions between extreme waves and offshore structures, these phenomena also raise the difficulty of choosing experimental designs due to a matching problem: the numerical irregular wave may not match the target analytic wave profile. We explicitly address this matching problem with a wave calibration procedure, but this mismatch is also implicitly accounted for via machine learning.

Once we obtain a first set of simulation data, we train a surrogate model using the GPR machine learning technique to learn the relationship from the wave episode to the structural load. Particularly, each surrogate model maps the KL coefficients ($\boldsymbol{\alpha}$) that describe the wave episode to a single output KL mode ($q_i$). Thus, we construct as many GPR models as the number of output KL modes. Finally, we model the force time series by combining the output KL coefficients from each trained GPR model.

To evaluate the generated force statistics as estimated from the surrogate model, we compare with longer time OpenFOAM simulations that give an accurate description of the steady state statistics . The prediction and the ground truth of the axial hydrodynamic force show a good agreement both in quiescent and extreme sea states. However, the viscous part of the force is not very well predicted-especially in the right tail. We attribute this to the fact that the active sampling method chooses wave episodes that improve the GPR model accuracy in predicting the pressure part of the total hydrodynamic force.

This framework has strong promise for application to design problems related to offshore wind turbine foundation or other offshore applications. The machine learned model presented in this study can facilitate the analysis of these structures combating the restraints of the classical modeling methods which suffer either from reduced accuracy (especially when it comes to extreme waves) or expensive computational cost. The presented surrogate modeling method allows for the investigation of more scenarios in the design stage, including extreme events, as well as efficient estimation of pdf

including their non-trivial tails. Therefore, it offers a reliable way for significant cost reduction and acceleration of the design process.

**Acknowledgments**

SG and TPS have been supported by the ONR grants N00014-20-1-2366 and N00014-21-1-2357. EK has been supported by the Swedish Centre of Natural Hazards and Disaster Science (CNDS), the Onassis Foundation (Scholarship ID: FZP 021-1/2019-2020, the Anna-Maria Lundins Scholarship (AMh2021-0023) and Liljewalch Scholarship. The CFD simulations were performed on computational resources provided by the Swedish National Infrastructure for Computing (SNIC) at the HPC cluster Tetralith at the National Supercomputer Centre, at Linköping University.

# References

[1] OpenFOAMv1906. https://www.openfoam.com/news/main-news/openfoam-v1906. [Online; accessed 13-January-2023].

[2] UNESCO. https://whc.unesco.org/en/renewable-energy/.

[3] A. Aggarwal, H. Bihs, D. Myrhaug, and M. A. Chella. Characteristics of breaking irregular wave forces on a monopile. *Applied Ocean Research*, 90:101846, 2019.

[4] P. A. Anastopoulos and K. J. Spyrou. Ship dynamic stability assessment based on realistic wave group excitations. *Ocean Engineering*, 120:256–263, 2016.

[5] P. A. Anastopoulos and K. J. Spyrou. Evaluation of the critical wave groups method in calculating the probability of ship capsize in beam seas. *Ocean Engineering*, 187:106213, 2019.

[6] P. A. Anastopoulos, K. J. Spyrou, C. C. Bassler, and V. Belenky. Towards an improved critical wave groups method for the probabilistic assessment of large ship motions in irregular seas. *Probabilistic Engineering Mechanics*, 44:18–27, 2016. Special Issue Based on Papers Presented at the 7th International Conference on Computational Stochastic Mechanics (CSM7).

[7] E. W. E. Association. Deep water: The next step for offshore wind energy. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2013.

[8] C. Baykal, B. M. Sumer, D. R. Fuhrman, N. G. Jacobsen, and J. Fredsøe. Numerical investigation of flow and scour around a vertical circular cylinder. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 373(2033):20140104, 2015.

[9] A. Blanchard and T. Sapsis. Output-weighted optimal sampling for Bayesian experimental design and uncertainty quantification. 2020.

[10] A. Blanchard and T. Sapsis. Bayesian optimization with output-weighted optimal sampling. *Journal of Computational Physics*, 425:109901, 2021.

[11] P. Boccotti. On mechanics of irregular gravity waves. *Atti della Accademia Nazionale dei Lincei*, 19:110–170, 1989.

[12] A. Bøhn. Fatigue loads on large diameter offshore wind monopile foundations in non-operational situations. Master's thesis, NTNU, 2016.

[13] K. Chaloner and I. Verdinelli. Bayesian experimental design: A review. *Statistical Science*, 10(3):273–304, 1995.

[14] F. Chasparis, Y. Modarres-Sadeghi, F. S. Hover, M. S. Triantafyllou, M. Tognarelli, and P. Beynet. Lock-in, transient and chaotic response in riser viv. volume Volume 5: Polar and Arctic Sciences and Technology; CFD and VIV of *International Conference on Offshore Mechanics and Arctic Engineering*, pages 479–485, 05 2009.

[15] C. Chung, J. Lee, and J.-S. Yang. National offshore wind strategy for late-mover countries. *Renewable Energy*, 192:472–484, 2022.

[16] W. Cousins and T. P. Sapsis. Reduced order precursors of rare events in unidirectional nonlinear water waves. *Journal of Fluid Mechanics*, 790:368–388, 2016.

[17] M. Elangovan. Simulation of irregular waves by CFD. *International Journal of Mechanical and Mechatronics Engineering*, 5(7):1379–1383, 2011.

[18] European Commission. A EU strategy to harness the potential of offshore renewable energy for a climate neutral future. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=COM:2020:741:FIN, 2020.

[19] European Commission. The EU Blue Economy Report 2022. *Publications Office of the European Union*, Luxembourg, 2022.

[20] M. Farazmand and T. P. Sapsis. Reduced-order prediction of rogue waves in two-dimensional deep-water waves. *Journal of Computational Physics*, 340:418–434, 2017.

[21] A. Forrester, A. Sobester, and A. Keane. *Exploring and Exploiting a Surrogate*, chapter 3, pages 77–107. John Wiley & Sons, Ltd, 2008.

[22] J. J. Gerbrands. On the relationships between svd, klt and pca. *Pattern Recognition*, 14(1):375–381, 1981. 1980 Conference on Pattern Recognition.

[23] A. Ghadirian, H. Bredmose, and M. Dixen. Breaking phase focused wave group loads on offshore wind turbine monopiles. In *Journal of Physics: Conference Series*, volume 753, page 092004. IOP Publishing, 2016.

[24] Y. Goda. *Random seas and design of maritime structures*, volume 33. World Scientific Publishing Company, 2010.

[25] S. Guth and T. P. Sapsis. Probabilistic characterization of the effect of transient stochastic loads on the fatigue-crack nucleation time. *Probabilistic Engineering Mechanics*, 66:103162, 2021.

[26] S. Guth and T. P. Sapsis. Wave episode based Gaussian process regression for extreme event statistics in ship dynamics: Between the Scylla of Karhunen–Loève convergence and the Charybdis of transient features. *Ocean Engineering*, 266:112633, 2022.

[27] K. Hasselmann, T. Barnett, E. Bouws, H. Carlson, D. Cartwright, K. Enke, J. Ewing, H. Gienapp, D. Hasselmann, P. Kruseman, A. Meerburg, P. Muller, D. Olbers, K. Richter, W. Sell, and H. Walden. Measurements of wind-wave growth and swell decay during the Joint North Sea Wave Project (JONSWAP). *Deut. Hydrogr. Z.*, 8:1–95, 01 1973.

[28] P. Higuera, J. L. Lara, and I. J. Losada. Realistic wave generation and active wave absorption for navier–stokes models: Application to OpenFOAM®. *Coastal Engineering*, 71:102–118, 2013.

[29] T. W. House. FACT SHEET: Biden-Harris administration announces new actions to expand U.S. offshore wind energy. 9 2022.

[30] X. Huan and Y. M. Marzouk. Simulation-based optimal Bayesian experimental design for nonlinear systems. *Journal of Computational Physics*, 232:288–317, 2013.

[31] L. J and Z. F. Global wind report 2021: Technical report. *Global Wind Energy Council (GWEC), Brussels, Belgium*, 2021.

[32] D. Kallehave, B. W. Byrne, C. LeBlanc Thilsted, and K. K. Mikkelsen. Optimization of monopiles for offshore wind turbines. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 373(2035):20140100, 2015.

[33] A. Kamath, M. A. Chella, H. Bihs, and Ø. A. Arntsen. Breaking wave interaction with a vertical cylinder and the effect of breaker location. *Ocean Engineering*, 128:105–115, 2016.

[34] K. Karhunen. Über lineare Methoden in der Wahrscheinlichkeitsrechnung. *Ann. Acad. Sci. Fennicae Ser. A. I. Math.-Phys.*, 1947(37):79, 1947.

[35] E. Katsidoniotaki, E. Nilsson, A. Rutgersson, J. Engström, and M. Göteman. Response of point-absorbing wave energy conversion system in 50-years return period extreme focused waves. *Journal of Marine Science and Engineering*, 9(3):345, 2021.

[36] E. Katsidoniotaki, E. Ransley, S. Brown, J. Palm, J. Engström, and M. Göteman. Loads on a point-absorber wave energy converter in regular and focused extreme wave events. In *International Conference on Offshore Mechanics and Arctic Engineering*, volume 84416, page V009T09A022. American Society of Mechanical Engineers, 2020.

[37] E. Katsidoniotaki, Z. Shahroozi, C. Eskilsson, J. Palm, J. Engström, and M. Göteman. Validation of a CFD model for wave energy system dynamics in extreme waves. *Ocean Engineering*, 268:113320, 2023.

[38] R. A. Khan and S. Ahmad. Dynamic Response and Fatigue Reliability Analysis of Marine Riser Under Random Loads. volume Volume 2: Structures, Safety and Reliability; Petroleum Technology Symposium of *International Conference on Offshore Mechanics and Arctic Engineering*, pages 183–191, 06 2007.

[39] S. Liu, M. C. Ong, and C. Obhrai. Numerical simulations of breaking waves and steep waves past a vertical cylinder at different Keulegan–Carpenter numbers. *Journal of Offshore Mechanics and Arctic Engineering*, 141(4), 2019.

[40] M. Loève. *Processus Stochastiques et Mouvement Brownien*, chapter Fonctions aléatoires du second ordre. Gauthier-Villars, 1948.

[41] M. K. McWilliam, M. Friis-Møller, N. Pollini, K. Dykes, and M. Jensen. A surrogate model of offshore wind farm support structures for wind farm design and financial valuation. In *Journal of Physics: Conference Series*, volume 2265, page 042048. IOP Publishing, 2022.

[42] M. A. Mohamad and T. P. Sapsis. Sequential sampling strategy for extreme event statistics in nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 115(44):11138–11143, 2018.

[43] M. Mohseni and C. G. Soares. CFD investigation of submerged geometry effect on wave run-up around a fixed, vertical monopile in regular head waves. In *Developments in Maritime Technology and Engineering*, pages 101–110. CRC Press, 2021.

[44] A. Morató, S. Sriramula, N. Krishnan, and J. Nichols. Ultimate loads and response analysis of a monopile supported offshore wind turbine using fully coupled simulation. *Renewable Energy*, 101:126–143, 2017.

[45] R. P. D. Offshore. DNV-RP-C203 Fatigue design of offshore steel structures. September, 2021.

[46] S. D. Offshore. DNV-ST-0126 Support structures for wind turbines. December, 2021.

[47] A. Ortega Nadal. Time domain simulation parameters for fatigue assessment of an offshore gravity based wind turbine. Master's thesis, NTNU, 2018.

[48] A. Otter, J. Murphy, V. Pakrashi, A. Robertson, and C. Desmond. A review of modelling techniques for floating offshore wind turbines. *Wind Energy*, 25(5):831–857, 2022.

[49] B. T. Paulsen, B. de Sonneville, M. van der Meulen, and N. G. Jacobsen. Probability of wave slamming and the magnitude of slamming loads on offshore wind turbine foundations. *Coastal Engineering*, 143:76–95, 2019.

[50] O. M. Phillips, D. Gu, and M. Donelan. Expected structure of extreme waves in a gaussian sea. Part I: Theory and SWADE buoy measurements. *Journal of Physical Oceanography*, 23:992–1000, 1992.

[51] E. Pickering, S. Guth, G. E. Karniadakis, and T. Sapsis. Discovering and forecasting extreme events via active learning in neural operators. *Nature Computational Science*, 2:823–833, 2012.

[52] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, USA, Jan. 2006.

[53] S. Rudy, D. Fan, J. d. A. Ferrandis, T. Sapsis, and M. S. Triantafyllou. Learning Optimal Parametric Hydrodynamic Database for Vortex-Induced Crossflow Vibration Prediction of Both Freely-Mounted Rigid and Flexible Cylinders. volume All Days of *International Ocean and Polar Engineering Conference*, 06 2021. ISOPE-I-21-3214.

[54] T. P. Sapsis. Output-weighted optimal sampling for Bayesian regression and rare event statistics using few samples. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 476(2234):20190834, 2020.

[55] T. P. Sapsis. Statistics of extreme events in fluid flows and waves. *Annual Review of Fluid Mechanics*, 53(1):85–111, 2021.

[56] H. A. Schäffer and G. Klopman. Review of multidirectional active wave absorption methods. *Journal of waterway, port, coastal, and ocean engineering*, 126(2):88–97, 2000.

[57] P. D. Sclavounos. Karhunen–Loève representation of stochastic ocean waves. *Proceedings of the Royal Society A*, 468(2145):2574–2594, 2012.

[58] S. Serebrinsky and M. Ortiz. A hysteretic cohesive-law model of fatigue-crack nucleation. *Scripta Materialia*, 53(1):1193–1196, 2005.

[59] K. W. Stevens. Adaptive sequential sampling for extreme event statistics in ship design. Master's thesis, Massachusetts Institute of Technology, 2018.

[60] S. Suryasentana, H. Burd, B. Byrne, A. Aghakouchak, and T. Sørensen. Comparison of machine learning models in a data-driven approach for scalable and adaptive design of laterally-loaded monopile foundations. 2020.

[61] Z.-h. Tang, B. Melville, N. Singhal, A. Shamseldin, J.-h. Zheng, D.-w. Guan, and L. Cheng. Countermeasures for local scour at offshore wind turbine monopile foundations: A review. *Water Science and Engineering*, 2022.

[62] K. K. Vu, C. D'Ambrosio, Y. Hamadi, and L. Liberti. Surrogate-based methods for black-box optimization. *International Transactions in Operational Research*, 24(3):393–424, 2017.

[63] S. Wang, T. J. Larsen, and H. Bredmose. Ultimate load analysis of a 10 MW offshore monopile wind turbine incorporating fully nonlinear irregular wave kinematics. *Marine Structures*, 76:102922, 2021.

[64] Y. Watanabe, H. Saeki, and R. J. Hosking. Three-dimensional vortex structures under breaking waves. *Journal of Fluid Mechanics*, 545:291–328, 2005.

[65] H. G. Weller. Derivation, modelling and solution of the conditionally averaged two-phase flow equations. *Nabla Ltd, No Technical Report TR/HGW*, 2:9, 2002.

[66] C. Windt, J. Davidson, P. Schmitt, and J. V. Ringwood. Wave–structure interaction of wave energy converters: A sensitivity analysis. *Proceedings of the Institution of Civil Engineers-Engineering and Computational Mechanics*, 173(3):144–158, 2020.

[67] Y. Yang, A. Blanchard, T. Sapsis, and P. Perdikaris. Output-weighted sampling for multi-armed bandits with extreme payoffs. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 478(2260):20210781, 2022.

[68] X. Zeng, W. Shi, C. Michailides, S. Zhang, and X. Li. Numerical and experimental investigation of breaking wave forces on a monopile-type offshore wind turbine. *Renewable Energy*, 175:501–519, 2021.

[69] J. Zhang, X. Zhao, and X. Wei. Data-driven structural control of monopile wind turbine towers based on machine learning. *IFAC-PapersOnLine*, 53(2):7466–7471, 2020.

# A Appendix A: CFD simulations

## A.1 Numerical Methods

### A.1.1 Governing equations

The Navier-Stokes equations are used to determine the motion of a fluid and can be seen as Newton's second law for fluid motion. For an incompressible Newtonian fluid flow, the Navier-Stokes equations take the form:

$$\frac{\partial u_i}{\partial t} + \frac{\partial}{\partial x_j}(u_i u_j) = -\frac{1}{\rho}\frac{\partial p}{\partial x_i} + \nu \nabla^2 u_i + f_b \tag{24}$$

where $u$ is the fluid velocity, $p$ is the fluid pressure, $\rho$ is the fluid density, $\nu$ is the fluid dynamic viscosity, and $f_b$ includes the external forces. The $i$ and $j$ denote the indices in $x$ and $y$ direction respectively. The left-hand side of this equation corresponds to the inertial forces while in the right-hand side the first term corresponds to pressure forces, the second term to viscous forces and the third term to the external forces applies in the fluid. These equations are always solved together with the continuity equation:

$$\frac{\partial u_i}{\partial x_i} = 0 \tag{25}$$

The CFD code the numerical model solves the Reynolds-averaged Navier-Stokes (RANS) equations, which are a reduced form of the general Navier-Stokes equations. The Reynolds decomposition is applied, according to which the variables $\mathbf{u}$ and $p$ of Navier-Stokes equations are written as the sum of the time-averaged and fluctuating part:

$$\begin{aligned} \mathbf{u} &= \overline{\mathbf{u}} + \mathbf{u}' \\ \mathbf{p} &= \overline{\mathbf{p}} + \mathbf{p}' \end{aligned} \tag{26}$$

The RANS equations and the mass conservation take the following form:

$$\frac{\partial \overline{u_i}}{\partial t} + \frac{\partial}{\partial x_j}(\overline{u_i u_j}) = -\frac{1}{\rho}\frac{\partial \overline{p}}{\partial x_i} + \nu \nabla^2 \overline{u_i} + \frac{1}{\rho}\frac{\partial \tau_{ij}}{\partial x_j} + f_b \tag{27}$$

$$\frac{\partial \overline{u_i}}{\partial x_i} = 0 \tag{28}$$

where $\tau_{ij}$ is the Reynolds stress tensor, $\tau_{ij} = -\overline{u_i' u_j'}$. To achieve a close form the RANS equations, the $k-\omega$ SST turbulence model is adopted. More details about the computational domain and the boundary conditions are provided in the Appendix A (subsection A.2).

### A.1.2 Free surface modeling

The free water surface is captured using the volume of fluid method (VOF) [65]. The two-phase fluid problem is treated as a single fluid and the phase fraction, $\alpha$, is used to indicate the mixture between air ($\alpha = 0$) and water ($\alpha = 1$) at each cell. The conservation of the phase fraction, $\alpha$, is essential and the transport equation should be added to describe the motion of the phases:

$$\frac{\partial \alpha}{\partial t} + \nabla \cdot (\mathbf{u}\alpha) + c_a \nabla \cdot (\mathbf{u}_c \alpha(1-\alpha)) = 0 \tag{29}$$

where the last term is an artificial compression to keep the surface sharp, $\mathbf{u}_c$ is the fluid velocity normal to the interface, where $c_\alpha$ is the compression coefficient. At each cell, the fluid properties (density, $\rho$, and dynamic viscosity, $\mu$) are computed as:

$$\rho = \alpha\rho_{water} + (1-\alpha)\rho_{air} \tag{30}$$

$$\mu = \alpha\mu_{water} + (1-\alpha)\mu_{air} \tag{31}$$

### A.1.3 Solver and algorithm

In this study, OpenFOAM version 1906 [1], is utilized to model the wave–structure interaction. OpenFOAM is an an open source CFD toolbox able to solve complex fluid applications based on the cell-centered finite volume method to solve the RANS of the two phase fluid flow, i.e., air and water. In this approach, the equations are integrated over each of the control volumes (i.e., each cell of the computational mesh). The volume integrals are converted to surface integrals using Gauss's theorem. The surface integrals are calculated as the weighted sum of the cell faces. The pressure-velocity coupling in RANS equations is solved via the PIMPLE algorithm, while the *interFoam* solver is operated for capturing the two incompressible, isothermal fluids using volume of fluid method for interface capturing.

## A.2 Numerical Wave Tank

A 3D numerical wave tank (NWT) is setup in order to reproduce the wave - structure interaction modeling. The total length of the NWT was defined by the wavelength and it is equal to 300 m (approximately $3\lambda$), the width is 42 m corresponding to $6D$, where $D$ is the diameter of the monopile. The water depth is 33 m. The NWT utilizes symmetric side planes in order to reduce the computational domain and thereby the computational cost. A monopile with diameter of 7 m was fixed at 100 m from the inlet boundary (approximately a wavelength). Figure 16 shows the computational domain with the boundary labeling.

The grid resolution is discretized based on the mesh sensitivity study presented in the following section. The quality of the wave propagation heavily depends on the aspect ratio of the cells close to the free water surface. The grid cells keep aspect ratio close to 1, which means the computational mesh consists of cubic cells. In a distance of one wavelength downstream the monopile the mesh grading technique is applied resulting in cells with larger aspect ratio. This is a technique for reducing the computational cost of each simulation.

For turbulence modeling the $k-\omega$ SST model was implemented with wall functions for the resolution of the boundary layer around cylinder. The boundary conditions are presented in the Table 1. The boundary conditions for the seabed and the side surfaces are type zeroGradient and symmetryPlane, respectively.

## A.3 Numerical Wave Generation and Absorption

The CFD modeling of waves requires a special set of boundary conditions. The wave is generated at the inlet which provides the appropriate time-dependent velocity field and surface elevation the outlet boundary offers the absorption capability. In OpenFOAM, several custom boundary conditions have been built for wave generation and absorption including a wide range of wave theories. In this study,

the IHFOAM library is utilized for the wave generation and absorption [28]. The irregular wave is composed of the superposition of regular wave components, i.e.; the first-order irregular wave episode is generated by the summation of the linear regular waves components. For the wave generation, the static boundary method is implemented at the inlet boundary in which the expressions for the free surface elevation (Equation 32) and wave velocities (Equation 33) are applied as Dirichlet boundary conditions based on the linear wave theory. The prescribed velocities determine the pressure and the free surface elevation.

$$\eta = \sum_{i=1}^{N} \frac{H_i}{2} \cos(k_{xi}x + k_{yi}y - \omega_i t + \phi_i), \tag{32}$$

where $N$ is the number of wave components, $H$ is the wave height, $k_x$ and $k_y$ are the components of the wave number vector $k$, $\omega$ is the angular frequency and $\phi$ is the wave phase. The orbital velocity components in horizontal and vertical directions of the wave propagation are given by:

$$\begin{aligned} u &= \sum_{i=1}^{N} \frac{H_i}{2} \omega_i \frac{\cosh[k_i(h+z)]}{\sinh(k_i h)} \cos(\theta_i), \\ w &= \sum_{i=1}^{N} \frac{H_i}{2} \omega_i \frac{\sinh[k_i(h+z)]}{\sinh(k_i h)} \sin(\theta_i), \end{aligned} \tag{33}$$

where $h$ is the water depth, $z$ is the water elevation from the still water surface and $\theta$ is the total phase of each wave component.

The absorption of the generated waves is a very important feature in the numerical modeling. The restrictive dimensions of the wave tank might causes unwanted reflections of the wave components which disturb the study zone of the domain. Therefore, the wave absorption should be handled properly in order to do not distort the results. In this study, the active wave absorption method, available on the IHFOAM library, was implemented. Based on Shallow Water Equation, the outlet boundary generates a velocity profile equal and opposite to the incident wave velocity that cancels the incoming wave. The active wave absorption presents good performance at long spurious waves and it was found that reduces the computational cost compared to other methods (e.g. relaxation zones). The methodology is further described in [56].

In OpenFOAM, for the construction of the irregular wave episode, the user chooses the *irregularMultiDirectional* wave model and provides the wave height, wave period and phase of each wave component. Further description about the wave generation is provided in the following subsections.

## A.4 Grid Convergence Study

The mesh sensitivity study is conducted without the presence of the monopile foundation. We want to evaluate the wave generation without being affected by the reflection effects from the foundation. The wave elevation is measured at the location where the monopile will be placed later. In the mesh sensitivity study the spectra wave density is evaluated. The numerical tests are performed for a two-dimensional numerical wave tank. In this study we consider two sea states; the milder sea state with Hs = 5 m and the extreme sea state with Hs = 13 m. The mesh resolution is different for the waves episodes of each sea state, i.e.; it will be computational costly to use the same fine resolution for the milder sea state, since the wave episodes are less extremes. Therefore, separate

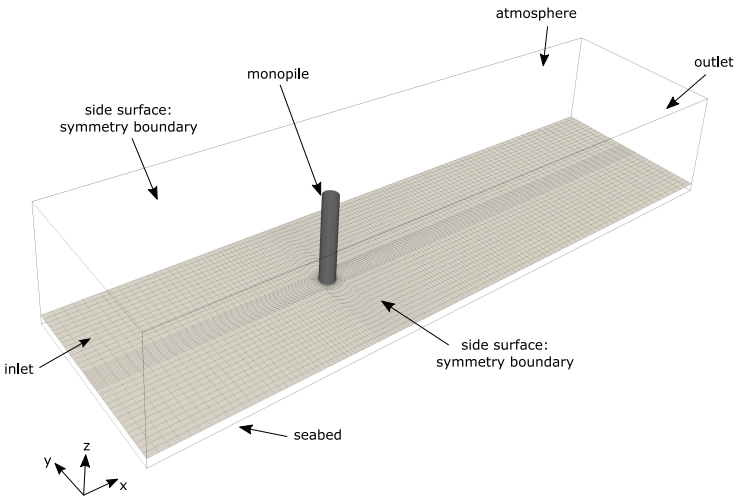


Figure 16: Boundaries of the computational domain on an example of mesh used for CFD simulations.

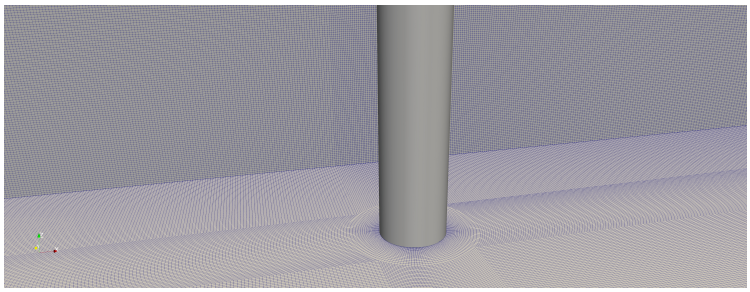

Figure 17: Spatial discretization of the NWT close the monopile.

| | Monopile | Atmposphere | Inlet | Outlet |
|---|---|---|---|---|
| $U$ | fixedValue | pressureInletOutletVelocity | waveVelocity | waveVelocity |
| $p$ | zeroGradient | totalPressure | zeroGradient | zeroGradient |
| $\alpha$ | zeroGradient | inletOutlet | waveAlpha | zeroGradient |
| $k$ | kqRWallFunction | zeroGradient | zeroGradient | zeroGradient |
| $\omega$ | omegaWallFunction | zeroGradient | zeroGradient | zeroGradient |

Table 1: The boundary conditions.

mesh study is carried out for each sea state. To evaluate the mesh resolution, steep irregular wave episodes are selected. The resolution is defined as cells per wave height (CPH), i.e; in our case the wave episode is irregular, therefore, the resolution is defined based on the mean wave height of the irregular realization.

In particular, for the milder sea state (Hs = 5 m), the resolution is evaluated based on the wave episodes 307 and 310. Four different resolutions are examined, i.e., 14 CPH, 17 CPH, 20 CPH, 28 CPH. For the extreme sea state (Hs = 13 m), six resolutions are examined since the wave episodes are strongly nonlinear, i.e.; 20 CPH, 25 CPH, 30 CPH, 35 CPH, 40 CPH, 45 CPH. Figure 18 presents the qualitative comparison of the mesh sensitivity study. The wave episodes 307 and 310 (milder sea state) present very good convergence for all the examined resolutions. The peak of the SWD is also captured accurately. For the waves 502 and 510 (extreme sea state), the SWD changes with the increasing mesh resolution, while the SWD better converges for 40 CPH and 45 CPH.

To quantitative evaluate the mesh convergence, we compute the $l_1$ norm on the SWD difference between each resolution with the finest mesh for each case on the interval $[f_l, f_u] = [0.09, 0.17]$. In particular the $l_1$ error metric is defined as:

$$l_1 = \int_{f_l}^{f_u} |SWD_i(f) - SWD_{finest}(f)|\, df \tag{34}$$

where the $SWD$ is the spectral wave density shown in Figure 18, the subscript $finest$ declares the mesh with the finest resolution for each examined wave, while the subscript $i$ declares the each of the other examined resolutions. $f_l$ and $f_u$ are the lower and upper boundaries of the examined frequency range. Figure 19 presents the l1 metric for each wave episode. For the milder sea state, the resolution 20 CPH satisfy $l_1 < 0.015$, therefore, the 20 CPH is used as resolution for all the wave episodes of the milder sea state. For the extreme sea state, $l_1 < 0.02$ is observed for resolution 40 CPH therefore this is used for all the wave episodes.

## A.5 Wave Characteristics

Some critical characteristics of the wave episodes are listed in the Table 2 and Table 3 (mild sea state), and Table 4 (extreme sea states)

| Training Wave Episodes | | | |
|---|---|---|---|
| **Wave episode** | **Max wave height [m]** | **Wave steepness** | **Keulegan-Carpenter** |
| **301** | 5.2 | 0.052 | 2.4 |
| **302** | 6.1 | 0.061 | 3.2 |
| **303** | 2.8 | 0.028 | 1.3 |
| **304** | 7.4 | 0.074 | 3.4 |
| **305** | 7.1 | 0.071 | 3.3 |
| **306** | 6.0 | 0.060 | 2.8 |
| **307** | 6.5 | 0.065 | 2.9 |
| **308** | 4.7 | 0.047 | 2.0 |
| **309** | 2.6 | 0.026 | 1.2 |
| **310** | 10.2 | 0.102 | 4.5 |
| **311** | 7.8 | 0.078 | 3.6 |
| **312** | 7.6 | 0.076 | 3.5 |
| **313** | 4.7 | 0.047 | 2.1 |
| **314** | 6.6 | 0.066 | 3.1 |
| **315** | 6.3 | 0.063 | 2.9 |
| **316** | 6.2 | 0.062 | 2.9 |
| **451** | 15.3 | 0.153 | 7.3 |
| **452** | 8.4 | 0.084 | 3.7 |
| **453** | 4.5 | 0.045 | 2.2 |
| **454** | 7.3 | 0.073 | 3.4 |

Table 2: Mild sea state ($H_s$ = 5 m, T = 8 s): Characteristics of the wave episodes used for training purposes. The waves 301 - 316 are randomly selected and the waves 451 - 454 are actively sampled.

| validation Wave Episodes | | | |
|---|---|---|---|
| **Wave episode** | **Max wave height [m]** | **Wave steepness** | **Keulegan-Carpenter** |
| **401** | 3.5 | 0.035 | 1.65 |
| **402** | 9.0 | 0.090 | 4.2 |
| **403** | 3.3 | 0.033 | 1.6 |
| **404** | 9.7 | 0.097 | 4.5 |
| **405** | 10.7 | 0.107 | 4.6 |
| **406** | 11.2 | 0.112 | 5.0 |
| **407** | 11.9 | 0.119 | 5.4 |
| **408** | 3.4 | 0.034 | 1.5 |
| **409** | 5.2 | 0.052 | 2.5 |
| **410** | 7.4 | 0.074 | 3.5 |
| **411** | 9.4 | 0.094 | 4.5 |

Table 3: Mild sea state ($H_s$ = 5 m, T = 8 s): Characteristics of the wave episodes used for testing purposes.

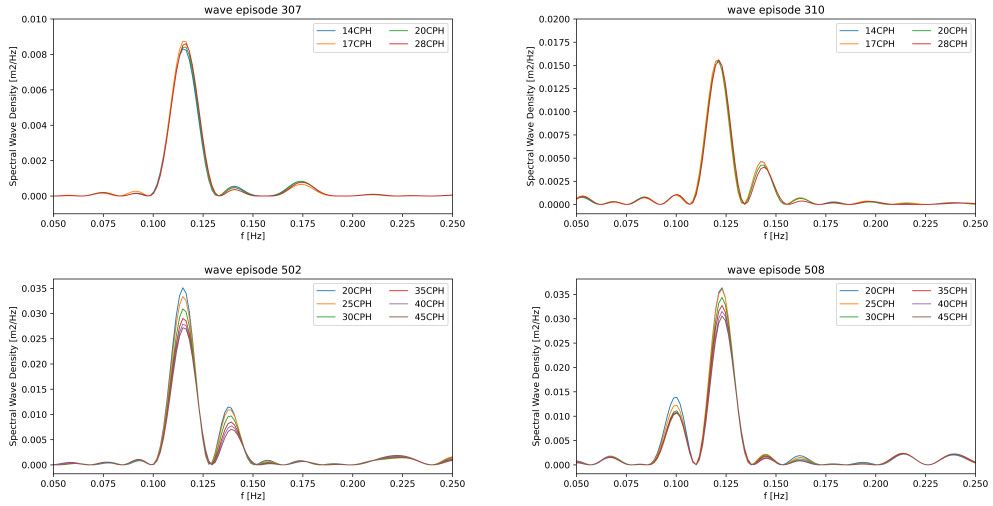


Figure 18: For the mesh sensitivity study the convergence of the spectra wave density is examined. For the sea state with Hs = 5 m the for wave episodes 307 and 310 are evaluated. For the sea state with hs = 13 m the wave episodes 502 and 508 are evaluated. The resolution is defined as cells per wave height (CPH). For the sea state with Hs = 13 m more resolutions examined due to the more complex nonlinear waves.

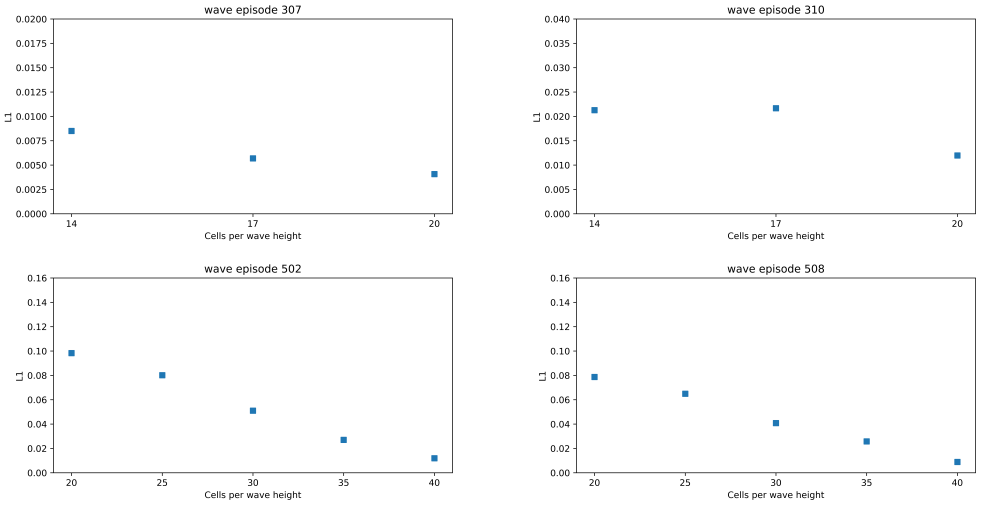


Figure 19: The l1 norm is considered as the error metric for the evaluation of the spectrum wave density convergence (Figure 18). Each resolution is compared with the finest resolution for each case.

| Wave episode | Max wave height [m] | Wave steepness | Keulegan-Carpenter |
|---|---|---|---|
| **501** | 18.8 | 0.188 | 8.6 |
| **502** | 25.3 | 0.253 | 11.8 |
| **503** | 18.5 | 0.185 | 8.6 |
| **504** | 16.7 | 0.167 | 7.5 |
| **505** | 23.3 | 0.233 | 9.5 |
| **506** | 22.7 | 0.227 | 10.4 |
| **507** | 14.0 | 0.140 | 6.4 |
| **508** | 27.3 | 0.273 | 12.3 |
| **509** | 25.5 | 0.255 | 11.5 |
| **510** | 26.0 | 0.260 | 11.5 |
| **511** | 23.2 | 0.232 | 10.9 |
| **512** | 26.3 | 0.263 | 12.0 |
| **513** | 21.7 | 0.217 | 9.7 |
| **514** | 26.0 | 0.260 | 11.8 |

Table 4: Extreme sea state ($H_s$ = 13 m, T = 8 s): Characteristics of the wave episodes used for training purposes. All the waves are actively selected.

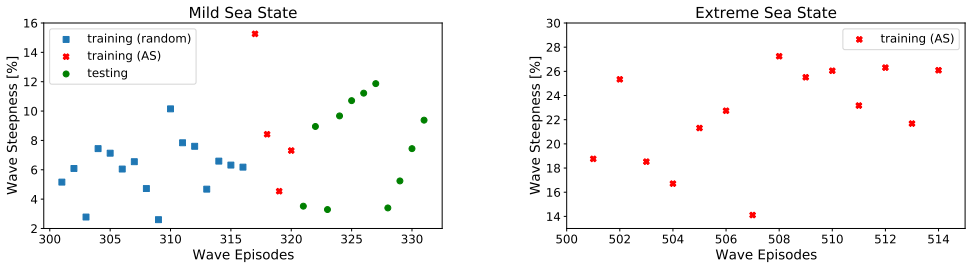


Figure 20: Visual depiction of the wave episodes steepness. Mild sea state (left): The initial training wave episodes are randomly selected (blue) and they are followed by wave episodes which are actively sampled (red). The testing waves are also depicted (green). Extreme sea state (right): The training wave episodes are randomly sampled.