

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

Stephen Guth<sup>1</sup>  | Eirini Katsidoniotaki<sup>1,2,3</sup>  | Themistoklis P. Sapsis<sup>1</sup> 

<sup>1</sup>Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

<sup>2</sup>Department of Electrical Engineering, Uppsala University, Uppsala, Sweden

<sup>3</sup>Centre of Natural Hazards and Disaster Science (CNDS), Uppsala, Sweden

## Correspondence

Stephen Guth, Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02138, USA.  
Email: [sguth@mit.edu](mailto:sguth@mit.edu)

## Funding information

Anna-Maria Lundins Scholarship, Grant/Award Number: AMh2021-0023; Liljewalch Scholarship; Office of Naval Research, Grant/Award Numbers: N00014-20-1-2366, N00014-21-1-2357; Swedish National Infrastructure for Computing; Swedish Centre of Natural Hazards and Disaster Science (CNDS); Onassis Foundation, Grant/Award Number: FZP 021-1/2019-2020

## Abstract

Accurately determining hydrodynamic force statistics is crucial for designing offshore engineering structures, including offshore wind turbine foundations, due to the significant impact of nonlinear wave–structure interactions. However, obtaining precise load statistics often involves computationally intensive simulations. Furthermore, the estimation of statistics using current practices is subject to ongoing discussion due to the inherent uncertainty involved. To address these challenges, we present a novel machine learning framework that leverages data-driven surrogate modeling to predict hydrodynamic loads on monopile foundations while reducing reliance on costly simulations and facilitate the load statistics reconstruction. The primary advantage of our approach is the significant reduction in evaluation time compared to traditional modeling methods. The novelty of our framework lies in its efficient construction of the surrogate model, utilizing the Gaussian process regression machine learning technique and a Bayesian active learning method to sequentially sample wave episodes that contribute to accurate predictions of extreme hydrodynamic forces. Additionally, a spectrum transfer technique combines computational fluid dynamics (CFD) results from both quiescent and extreme waves, further reducing data requirements. This study focuses on reducing the dimensionality of stochastic irregular wave episodes and their associated hydrodynamic force time series. Although the dimensionality reduction is linear, Gaussian process regression successfully captures high-order correlations. Furthermore, our framework incorporates built-in uncertainty quantification capabilities, facilitating efficient parameter sampling using traditional CFD tools. This paper provides comprehensive implementation details and demonstrates the effectiveness of our approach in delivering reliable statistics for hydrodynamic loads while overcoming the computational cost constraints associated with classical modeling methods.

## KEYWORDS

active sampling, heavy tails and extreme events, offshore structures, optimal experimental design, reduced order modeling, wave episodes

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Applied Organometallic Chemistry* published by John Wiley & Sons Ltd.

## 1 | INTRODUCTION

Offshore wind turbines (OWTs) are anticipated to play a key role in driving the global transition toward sustainable energy sources.<sup>1–3</sup> As wind turbine power levels continue to increase, there is a growing demand for larger OWT structures. However, a significant challenge faced by the offshore wind industry is to ensure the reliability and survivability of these structures while simultaneously reducing costs. Throughout their operational lifespan, OWTs are exposed to stochastic and extreme environmental loads (e.g., wind and waves) that can affect their lifetime.<sup>4–9</sup> Accurately capturing the aero-hydro-servo-elastic loads on OWT structures is essential. Specifically, precise analysis of the ultimate limit state (ULS) and fatigue limit state (FLS) enables less conservative designs and, consequently, cost reduction. To adhere to current best practices, such as designing based on the extreme loads corresponding to a 50-year return period,<sup>10–12</sup> engineers require prior knowledge of the short-term and long-term extreme load statistics on critical structural components. The extreme value theory is crucial for estimating the distribution of long-term extreme loads by leveraging information from short-term extreme loads.<sup>13</sup> While the IEC standard provides recommendations for estimating 50-year loads using statistical extrapolation methods,<sup>14</sup> it allows flexibility for designers to choose the specific details of the statistical procedure.<sup>12</sup> Commonly used peak extraction techniques, such as block maxima and peak over threshold, have been explored by researchers.<sup>15–18</sup> These techniques are accompanied by fitting functions, including normal, Weibull, three-parameter Weibull, Gumbel, and Lognormal distributions, to extrapolate the data. The choice of statistical extrapolation methods significantly impacts estimation results, although limited research has compared their effects. Addressing the challenge of handling large variations in environmental conditions, Monte Carlo-based sampling procedures offer a more realistic approach.<sup>19,20</sup> However, basic randomization may exhibit slow convergence in achieving desired 50-year return period estimates. Implementing these methods in engineering design poses challenges, with a primary focus on validating selected extrapolation methods through comparison with measurement or simulation data to ensure accuracy and reliability. Additionally, capturing dynamic load behavior and assessing extrapolation method performance necessitates a substantial number of transient simulations and/or physical experiments.

For industry and engineering purposes, semi-empirical analytical models are available and are commonly used for the determination of extreme loads in OWT structures.<sup>21,22</sup> Although lower fidelity numerical models are fast, they cannot capture high nonlinear phenomena while the peak loads are captured accurately. This is a critical issue for the reliable design of the offshore wind turbine foundations. In the literature, several studies are available that utilize CFD simulations to examine the monopile-type foundation in breaking waves,<sup>23–25</sup> analyze the slamming effects,<sup>26,27</sup> investigate the flow and scour around the monopile<sup>28,29</sup> and identify the second-order hydrodynamic loads.<sup>30,31</sup> The high-fidelity simulations provide an accurate solution yet demand expensive computational cost. Generalizing from simulation or experimental results to the recovery of short-term statistics (i.e., statistics of the steady state) requires the application of careful data-driven surrogate modeling.<sup>32–34</sup> Developing surrogate models to capture the behavior of offshore wind turbines is essential during the design stage. These models offer a cost-effective alternative to numerical modeling and experiments, enabling efficient and reliable design processes. They support various analyses, including sensitivity analysis, optimization, uncertainty quantification, and statistical reconstruction for defining ULS and FLS. Surrogate modeling, such as with Gaussian process regression (GPR),<sup>35,36</sup> additionally allows the deployment of techniques from the optimal experimental design literature,<sup>37–39</sup> where initial simulation or experimental results are used to improve subsequent designs.<sup>40–44</sup>

In addition, scientists have addressed the question of experimental design for these studies—what waves to simulate—with a number of methods, including stochastic wavegroups,<sup>45,46</sup> critical wavegroups,<sup>47–50</sup> equivalent waves,<sup>51</sup> reduced order wavegroups,<sup>40,52,53</sup> and Karhunen-Loève (KL) wave episodes.<sup>54</sup> While previous studies have explored surrogate models using methods like polynomial-chaos expansion and Kriging,<sup>55</sup> there is a growing interest in utilizing machine learning approaches. However, according to Elyasichamazkoti and Khajehpoor,<sup>56</sup> most existing studies develop surrogate models focusing on specific areas such as fault detection, wind speed and power forecasting, power optimization, and control. Limited research has been conducted on creating surrogate models that accurately capture the dynamic response of offshore wind turbines. For instance, Ahmad et al<sup>57</sup> utilized artificial neural networks to develop a surrogate model for the aero-hydro-servo elastic performance of hybrid floating offshore wind turbines. The model was trained on mid-fidelity data; however, there is a need to develop surrogates that can accurately capture extreme dynamic phenomena by utilizing high-fidelity datasets.

This work introduces a novel framework for modeling the short-term probability distribution of extreme loads on the monopile foundation of OWTs across several sea states. As the increasing power levels of wind turbines require larger monopiles with significant diameters, the nonlinear wave-monopile interaction, especially in breaking wave conditions, is crucial to consider. Therefore, we focus on the hydrodynamic loads. This study takes the advantage of the predictive capabilities and the computational simplicity of the GPR machine learning method jointly with the efficiency of the active sampling method to develop a surrogate model. The surrogate maps the relationship between irregular wave episodes and hydrodynamic force time series, enabling the estimation of extreme load statistics for various sea states through Monte Carlo sampling. The surrogate model is trained using high-fidelity load datasets from CFD simulations, capturing strong nonlinear hydrodynamic effects like wave breaking and snap loads. The study validates the predicted probability distributions against steady-state CFD simulations, confirming the method's validity and significant computational cost savings. A unique feature of the framework lies in the employed techniques. Specifically, the spectrum transfer approach utilizes simulation data from multiple sea states to describe both quiescent and extreme waves, reducing the need for additional numerical simulations. To handle time series prediction via the surrogate modeling, dimension reduction techniques such as KL and PCA are employed,

while the GPR model accounts for nonlinear dependencies, including those associated with extreme events. While similar techniques have been used in ship hull load calculations,<sup>54</sup> wave episode methods have not been applied to estimate hydrodynamic loads on monopile structures before. Overall, this work presents a pioneering framework to model short-term extreme hydrodynamic loads statistics on monopile foundations showcasing its potential in improving the understanding and prediction of extreme loads in offshore wind turbine design.

We structure this paper as follows. In Section 2, we briefly recapitulate the design of KL wave episodes first described in Guth and Sapsis.<sup>54</sup> 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 three sea states with  $H_s \in \{5\text{m}, 9\text{m}, 13\text{m}\}$  to represent quiescent, intermediate, and very extreme seas, respectively. While we have used  $H_s = 13\text{m}$  during data collection to emphasize extreme waves and nonlinear physics, we follow literature suggestions to evaluate on  $H_s = 9\text{m}$  as a better match for real applications.<sup>58–60</sup> Additionally, we include Appendix A with details about the CFD simulations setup and the characteristics of examined wave episodes.

## 2 | IRREGULAR WAVE EPISODES

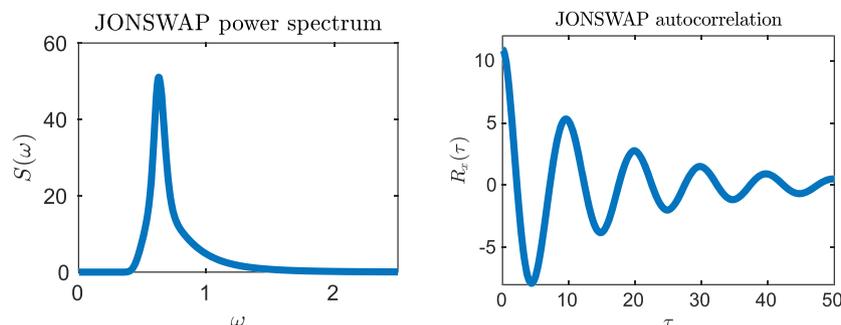
Our goal is to estimate the hydrodynamic load statistics for an offshore wind turbine monopile foundation subjected to a selected sea state. Traditionally, a sea state realization (i.e., steady state) is approximated as an irregular wave record with duration from 30 min to 3 h. To avoid the enormous 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<sup>54</sup> 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 (Figure 1).

### 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<sup>61</sup> to describe the time series  $x(\xi = 0, t)$  at a fixed spatial location:

$$\begin{aligned} S_j(\omega) &= \frac{\alpha g^2}{\omega^5} \exp\left[-\frac{5}{4}\left(\frac{\omega_p}{\omega}\right)^4\right] \gamma^r \\ r &= \exp\left[-\frac{(\omega - \omega_p)^2}{2\sigma^2 \omega_p^2}\right] \end{aligned} \quad (1)$$



**FIGURE 1** Left: One-sided power spectral density for a JONSWAP spectrum with  $H_s = 13$  meters. Right: Corresponding JONSWAP time autocorrelation function.

where constants are given by

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

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 multiple sea states, each of which with peak period  $T_p = \frac{2\pi}{\omega_p} = 8$  s and with a range of significant wave height crest to trough of  $5 \text{ m} \leq H_s \leq 13 \text{ m} = 8$  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 mild and extreme sea states 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  $[0, T]$ .<sup>47–49</sup> 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

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

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

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

However, the random phase model does not provide the low dimensional parametrization necessary to apply ML techniques. Instead, we follow Guth and Sapsis<sup>54</sup> in building a basis set corresponding to a particular sea state and interval. We present the KL theorem<sup>62,63</sup>:

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

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

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

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

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

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

are centered orthogonal random variables:

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

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

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

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  $\alpha_i, i = 1, 2, \dots$ . In the literature, this procedure is also known as principle component analysis (PCA), proper orthogonal decomposition (POD).<sup>64</sup> 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  $\alpha$  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.<sup>54</sup> 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  $\alpha$ . 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  $\alpha$ .

We follow the same technique described in Guth and Sapsis<sup>54</sup> to convert from the coefficient vector  $\alpha$  to the OpenFOAM input. As we describe in Section A.2.2 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. We emphasize, that while OpenFOAM uses these linear waves to define the boundary conditions, especially at the inlet, within the numerical wave tank the wave evolution is fully nonlinear.

Then, we extrapolate the signal onto the interval  $[0, 20]$  using the stochastic prelude technique described in Guth and Sapsis<sup>54</sup> 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 an initialization and ramp-up period and the rest is the real simulation time. We note that during 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 | Breaking wave criterion

Before performing the CFD simulations, we want to have a priori-knowledge whether the analytical irregular wave episodes appear breaking wave behavior. They are assessed in relation to the breaking wave criterion,<sup>65</sup> which is defined by

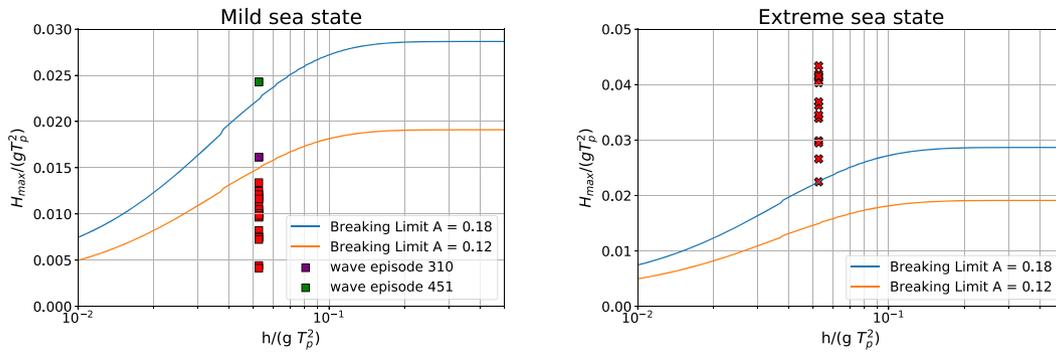
$$\frac{H_b}{gT_p^2} = A \left( 1 - \exp\left(-1.5\pi \frac{d}{L_p}\right) \right) \frac{\tanh(kd)}{2\pi}, \quad (9)$$

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,  $0.12 < A < 0.18$ . The wave number,  $k$  is estimated from the dispersion relation:

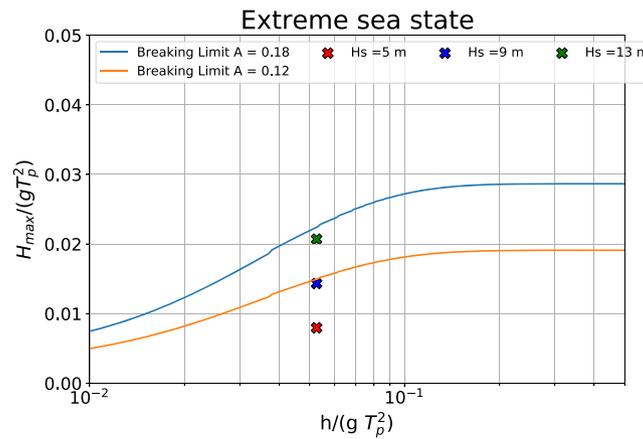
$$\omega_p^2 = k \cdot g \cdot \tanh(k \cdot d). \quad (10)$$

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 A2, Table A3, and Table A4. 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.

While in Figure 2, we presented the simulated wave episodes compared to the breaking limit; in Figure 3, we compare the recovered sea states in comparison to the breaking limit. We see that for the quiescent  $H_s = 5$  m case, the significant wave height is well below the wave breaking limit, while for the very extreme  $H_s = 13$  m, the significant wave height nearly exceeds the upper limit. In the intermediate  $H_s = 9$  m case, the



**FIGURE 2** The analytical wave episodes are located based on their wave breaking limit. **Left:** At the quiescent sea state ( $H_s = 5\text{m}$ ), the majority of the wave episodes are below the breaking region (orange), except the wave episode 451, which 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 ( $H_s = 13\text{m}$ ), all the wave episodes are above the maximum breaking limit (blue); therefore, breaking wave phenomena is expected to occur.



**FIGURE 3** Comparison of the three sea states considered ( $H_s \in \{5\text{m}, 9\text{m}, 13\text{m}\}$ ) with the breaking wave criterion from Equation 9.

significant wave height is approximately the lower limit, suggesting that nonlinear breaking effects will be important only for fraction of the steeper waves encountered in the steady state.

## 2.4 | 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:

$$F_x(t) = \sum_{i=1}^{n_{out}} q_i \hat{\mu}_{i,T}(t), \quad t \in [0, T]. \quad (11)$$

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,<sup>54</sup> a set of force-on-structure time series associated with the particular wave episodes we simulated. We additionally note that, while the reduced order model used to represent the force is linear, we are able to represent non-Gaussian loads and load statistics through the  $q_i$  model, described in the next section.

### 3 | SURROGATE MODELING WITH GAUSSIAN PROCESS REGRESSION

#### 3.1 | Gaussian process regression

In this section, we describe the construction of a data-drive surrogate model used to estimate the load coefficients  $q_i$  using the GPR technique. For an introduction to Gaussian process theory, see Rasmussen and Williams.<sup>35</sup> Briefly, the Gaussian process constructs a normal distribution for every choice of  $\alpha$  given by

$$p(q_i | \alpha) \sim \mathcal{N}(\mu(\alpha), \sigma(\alpha)), \quad (12)$$

where

$$\mu(\alpha) = K_*^\top (K + \sigma_n^2 I)^{-1} Q \quad (13)$$

$$\sigma^2(\alpha) = K_{**} - K_*^\top (K + \sigma_n^2 I)^{-1} K_*, \quad (14)$$

$\mu(\bar{\alpha})$  is the posterior mean,  $\sigma^2(\bar{\alpha})$  is the posterior variance,  $\sigma_n^2$  is the aleatoric uncertainty,  $K$ ,  $K_*$ , and  $K_{**}$  are kernel matrices, and  $Q$  is the observed output data vector. The mean of this posterior distribution is the model's best prediction of the loads associated with a particular wave episode, and the variance is a combination of epistemic uncertainty (from limited data) and aleatoric uncertainty (from irreducible randomness, e.g., associated with the stochastic prelude).

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

$$k_{SE}^{ARD}(\alpha_1, \alpha_2) = \sigma_k^2 \exp\left(-\frac{1}{2} \sum_i \frac{(\alpha_{1,i} - \alpha_{2,i})^2}{l_i^2}\right) = \sigma^2 \exp\left(-\frac{1}{2} (\alpha_1 - \alpha_2)^\top M (\alpha_1 - \alpha_2)\right), \quad (15)$$

where  $M = \text{diag}(l_i)$ . The kernel hyperparameter  $\sigma_k$  is a measure of variability due to choice of  $\alpha$ , 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 (11), 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  $\alpha$ . That is, for each  $i \in [1, n_{\text{out}}]$ , we build a distinct Gaussian process to estimate a posterior distribution

$$q_i(\alpha) \sim \mathcal{N}(\mu_i(\alpha), \sigma_i(\alpha)). \quad (16)$$

This collection of independent scalar models allows us to model a low dimensional projection of the full time series  $F_x(t|\alpha)$ . We can sample from the combined model by sampling each  $q_i$  individually from the Gaussian process posteriors, and then combining the coefficients using Equation (11). While the variance for each  $q_i$  is sampled independently, the means  $\mu_i(\alpha)$  are allowed to vary non-parametrically. This allows for the short-term distribution of each  $q_i$  to have non-Gaussian (joint) statistics, despite each conditional distribution in Equation (16) having Gaussian form. To restate, the Gaussian distribution inherited from GPR represents the uncertainty quantification of the model, not the predicted short-term statistics.

#### 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, because the GPR model is more reliable in interpolation than extrapolation. 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  $\alpha$ , 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 9 (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 primarily represent the effects of nonlinear physics.

Second, we make use of the initial uniform data points in order to optimally choose subsequent simulation designs for additional training data. Active learning, 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.<sup>43,66</sup>

For this problem, we use Likelihood Weighted Uncertainty Sampling (LW-US), an acquisition function developed in the literature<sup>42,43</sup> to preferentially sample inputs likely to contribute to the output statistical tails – rare events. The functional form of the acquisition function is given by

$$u_{LW-US}(\boldsymbol{\alpha}) = \frac{p_A(\boldsymbol{\alpha})}{p_q(q(\boldsymbol{\alpha}))} \sigma_q^2(\boldsymbol{\alpha}) \quad (17)$$

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 Guth and Sapsis,<sup>54</sup> 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  $(\sigma_q^2(\boldsymbol{\alpha}) - \sigma_n^2)$ , that is, 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

$$u_{LW-US_j}(\boldsymbol{\alpha}) = \frac{p_{\alpha}(\boldsymbol{\alpha})}{p_{q_j}(q_j(\boldsymbol{\alpha}))} (\sigma_{q_j}^2(\boldsymbol{\alpha}) - \sigma_{n_j}^2). \quad (18)$$

### 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.<sup>54</sup> For the present context, we have input-output pairs  $\boldsymbol{\alpha}_i, \mathbf{q}_i$  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.<sup>54</sup> First, we calculate the KL basis for each spectrum,  $S_1$  and  $S_2$ . Second, we calculate the eigenvalue ratios

$$r_j = \sqrt{\frac{\lambda_j^{(1)}}{\lambda_j^{(2)}}}. \quad (19)$$

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

$$\alpha_j^{(1)} = \alpha_j^{(2)} r_j \quad (20)$$

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, for example, 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.

## 4.1 | Wave induced force

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

$$F = \iint_A (p\hat{n} + \tau)dA, \quad (21)$$

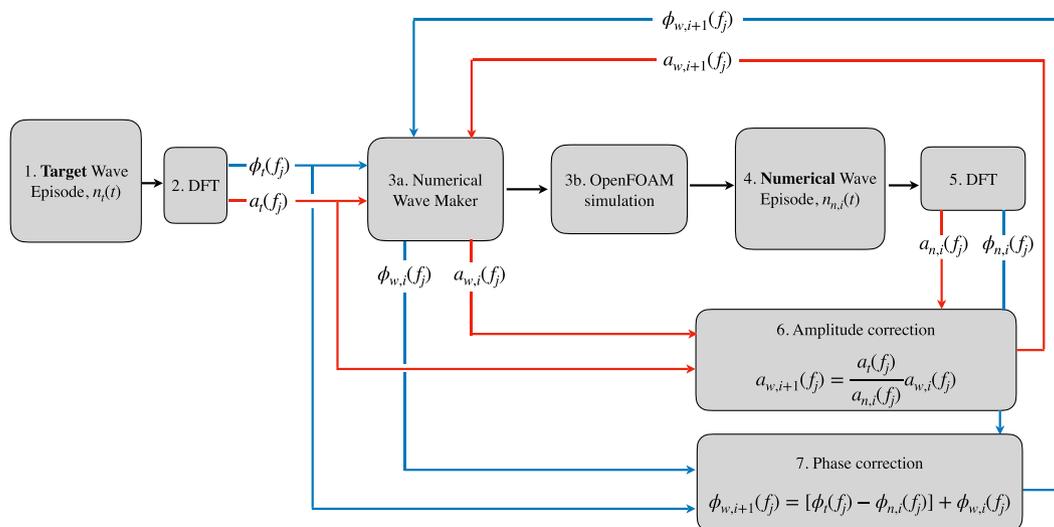
In our application, the horizontal component of the force, which is known as the *inline force*,  $F_x$ , has the major contribution; thus, it is chosen as the quantity of interest in this study.

## 4.2 | Numerical wave calibration

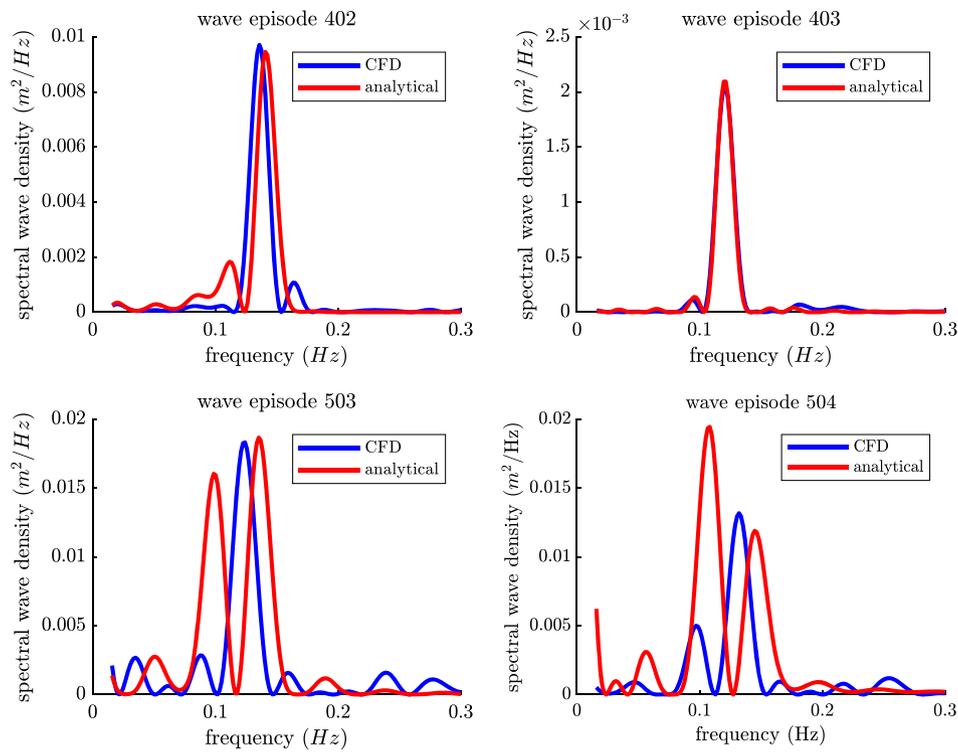
In OpenFOAM simulations, to produce the *target* irregular wave episode (e.g., analytical) at a specific location in the numerical wave tank (e.g., where the monopile foundation will be placed), we follow a *wave calibration* procedure.<sup>67</sup> Figure 4 shows the schematic depiction of the wave calibration, which is similar to the one presented in Windt et al.<sup>68</sup> and previously has been applied by previous studies.<sup>69,70</sup>

The target wave episode and the numerical wave obtained from the last iteration are compared through their *density wave spectrum*. Figure 5 compares four representative wave episodes. For the waves obtained from the quiescent sea state (i.e., 404 and 403), the analytical and numerical spectra 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 A.1), 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.

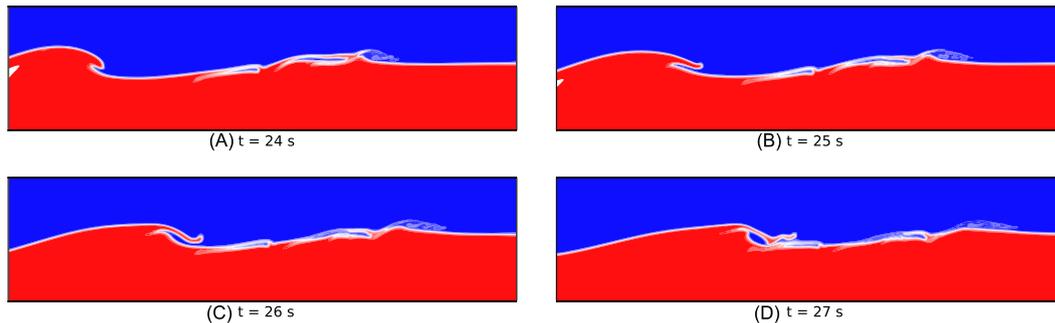
In Figure 2, we compare the analytical wave episodes to the two estimates of the breaking limit – defined from the breaking wave criterion (Section 2.3). We observe that, for the quiescent sea state ( $H_s = 5$  m), all but one of the waves are firmly below the breaking limit. However, for



**FIGURE 4** 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, that is,  $a_{w,i}(f_j) \rightarrow a_{w,i+1}(f_j)$  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 5** 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).



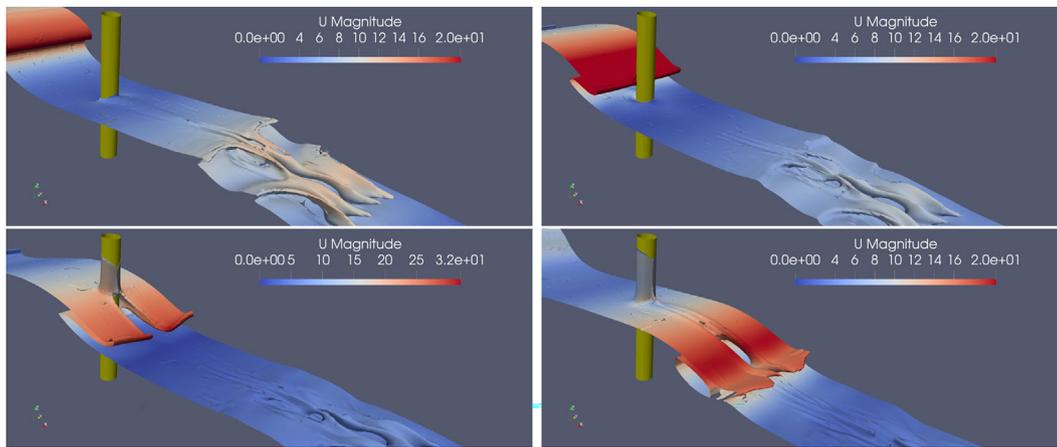
**FIGURE 6** Wave episode 508: Schematic depiction of the wave breaking process in the empty numerical wave tank (without the presence of the monopile foundation).

the extreme sea state ( $H_s = 13$  m), nearly all of the waves are above the breaking limit. In Figure 6 and Figure 7, we provide an OpenFOAM visualization of the breaking wave evolution and its interaction with the monopile foundation, respectively.

### 4.3 | Ground truth solution

To evaluate the accuracy of the GPR surrogate model, we compare the model prediction with a reference solution—that is, 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 three sea states ( $H_s \in \{5\text{m}, 9\text{m}, 13\text{m}\}$ ) by simulating the wave-structure interaction for a 30 min period ( $H_s = 5\text{m}$ ), a 75 min period ( $H_s = 9\text{m}$ ), or 60 min period ( $H_s = 13\text{m}$ ).

In CFD simulations, because of the expensive computational 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 realizations, 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 simulations into 12 and



**FIGURE 7** 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.

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.

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.

#### 4.4 | Surrogate model of the monopile foundation

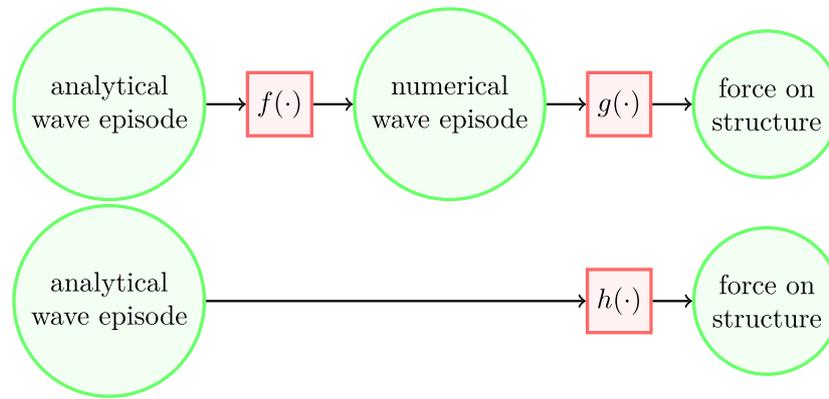
We model the distortion introduced between the target (analytical) wave episode and the numerical wave episode as a nonlinear mapping. The numerical solver along with the boundary conditions and absorption, implicitly define the map,  $f$ , between the target and numerical wave episodes (this is the wave calibration from Section 4.0.2). In Figure 5, 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  $\alpha$ ), 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 8, 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.

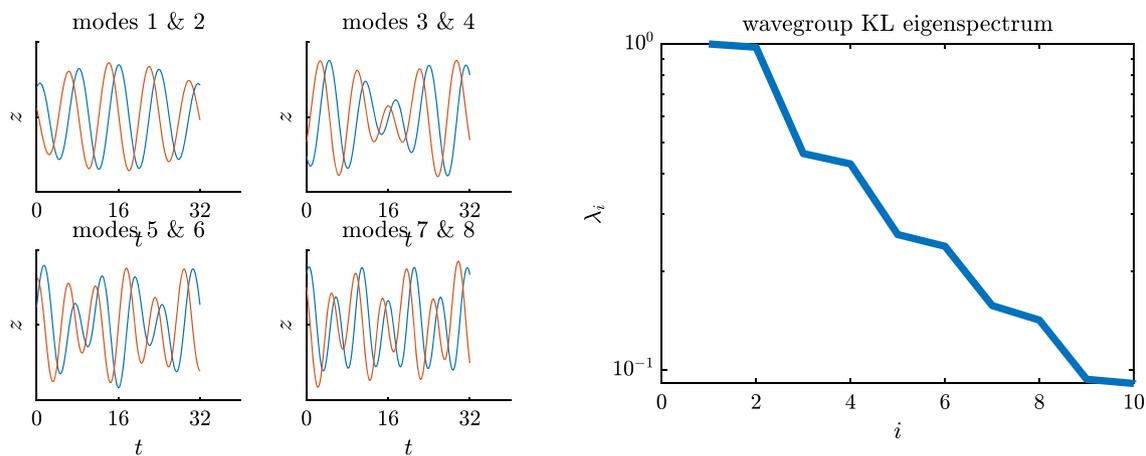
This shift is a problem for training a model that converges *pointwise*. That is to say, if we want the precise structural loads corresponding to the target wave episode (that we parametrized with  $\vec{\alpha}$ ), the OpenFOAM solution for the force-on-structure does not correspond to the desired wave. However, if instead our goal is to train a model that merely calculates the correct short-term 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 8 the function  $f(\cdot)$  does not change the short-term *statistics*, then the machine learning model may still learn the short-term statistics of the forces on structure. One straightforward case where this could be 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).



**FIGURE 8** 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 9** 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.

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  $H_s \in \{5\text{m}, 9\text{m}, 13\text{m}\}$  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 9 (Left) displays the mode shapes for the KL basis, while Figure 9 (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.

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, that is, 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 four 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 nine irregular wave realizations using active sampling. We additionally simulated assorted validation wave episodes.

## 5.2 | Computational resources

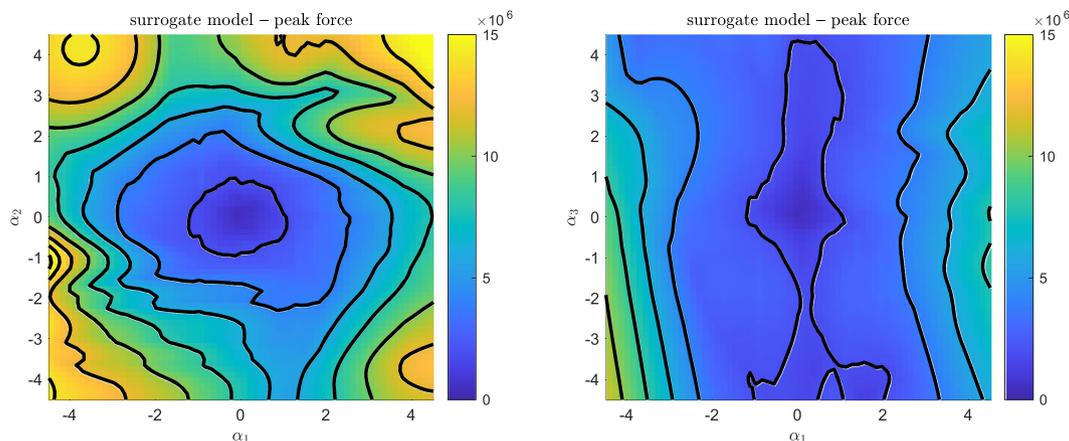
In this study, the CFD simulations are performed on the Tetralith HPC cluster utilizing 128 processors per simulation. Each irregular wave episode, used for training the surrogate model, is simulated to interact with the monopile foundation for 52 s. This simulation requires approximately 12–18 h real 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 4.0.3, this time is split into 12 and 24 intervals, 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 40–45 h of real time on the HPC cluster.

## 5.3 | Hydrodynamic force dimension reduction

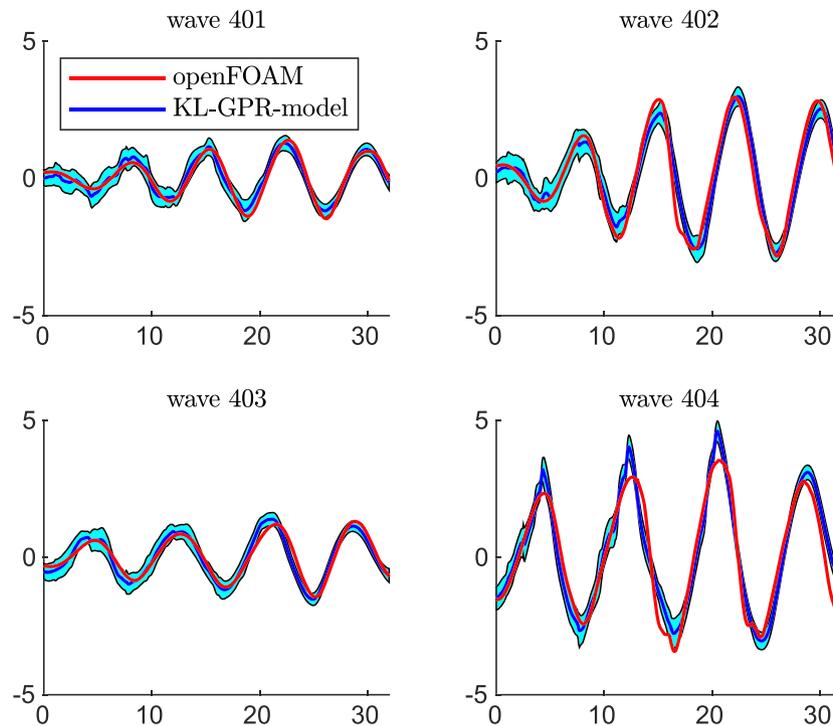
Figure 10 shows contour slices for the KL-GPR surrogate for different combinations of PCA modes. The contours show the regions where the peak hydrodynamic force presents maxima and minima values. This is helpful for engineers in the design stage to extract information about how the waves (which are described by the KL coefficient vector  $\alpha$ ) impact the maximum hydrodynamic force. In other words, Figure 10 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 11 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.



**FIGURE 10** Maximum (unsigned) predicted force,  $F_x$ , of KL-GPR surrogate in [N], for the extreme sea state with  $H_s = 13$  m. The colormap shows the change in force for varying input waves, which are described by vector  $\alpha$ . **Left:**  $(\alpha_1, \alpha_2), \alpha_3 = 0$ . **Right:**  $(\alpha_1, \alpha_3), \alpha_2 = 0$ .



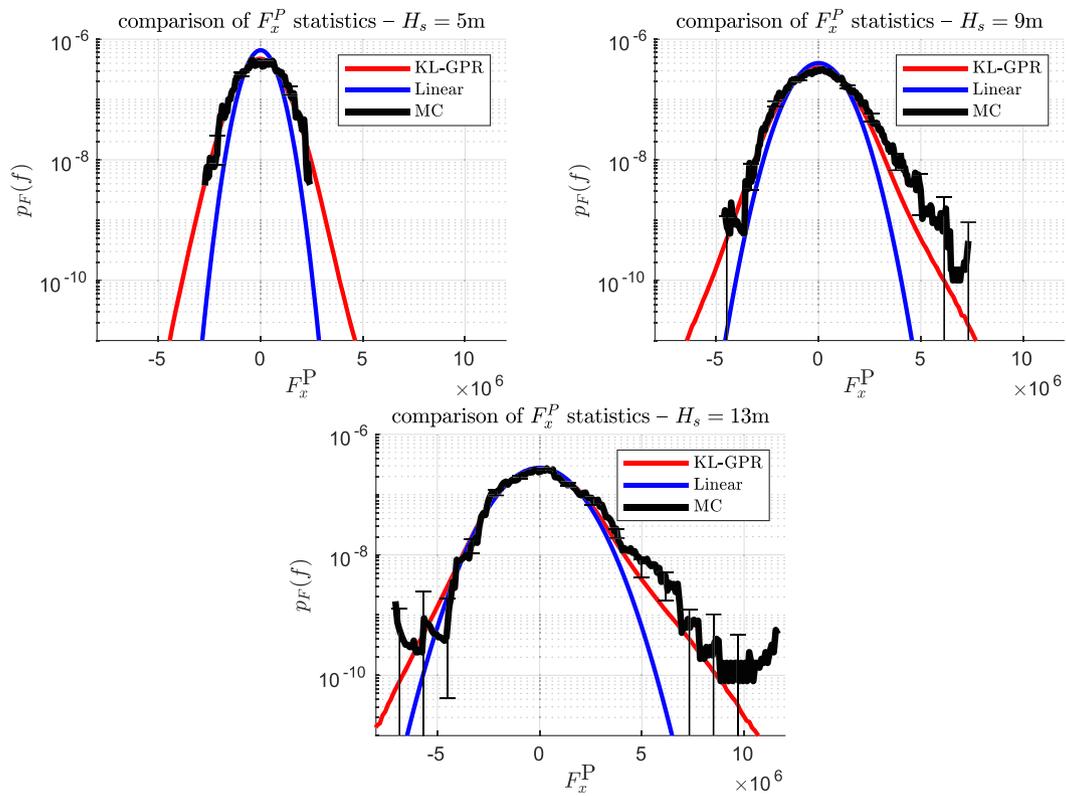
**FIGURE 11** Comparisons of  $F_x(t)$  from (red) OpenFoam and (blue) KL-GPR-model with posterior uncertainty, for quiescent sea state with  $H_s = 5$  m. Y-axis counts  $1 \times 10^5$  [N]. These waves are only used for validation purposes – not used during the model training.

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 A2 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 A3 Appendix A. This led to our decision to additionally train on extreme waves associated with the spectrum  $H_s = 13$  m.

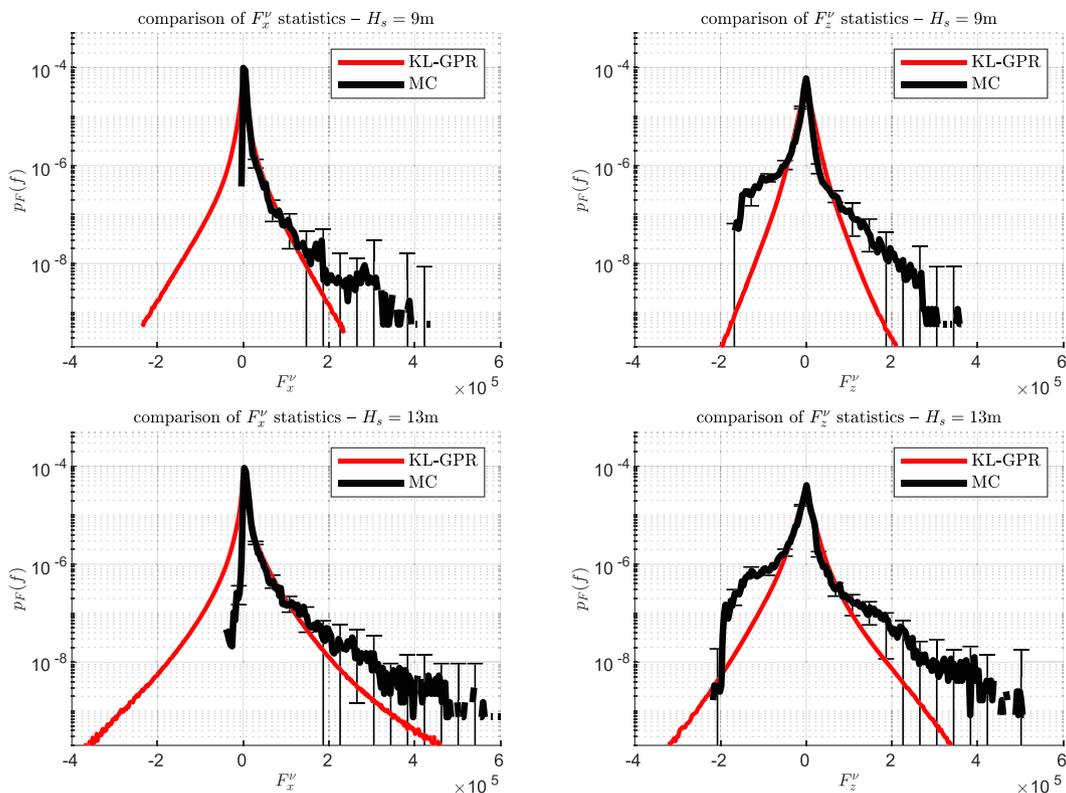
Ultimately, our goal is to recover the short-term statistics (statistics of the steady state) of the forces on the structure. The first force we examine is the signed hydrodynamic axial pressure force  $F_x^p$ , the dominating component of the total axial force. In Figure 12, we present the pdf of  $F_x^p$  for the three sea states we have considered:  $H_s \in \{5\text{m}, 9\text{m}, 13\text{m}\}$ . For each plot, we display the direct Monte Carlo statistics of the steady state (in black), a linear estimate derived from potential flow and the Wiener–Khinchine theorem (in blue),<sup>71–73</sup> and our reconstructed statistics from the KL-GPR model (in red). The KL-GPR statistics are generated by Monte Carlo sampling a large number ( $1 \times 10^5$ ) of wave episodes, and calculating the corresponding forces from the GPR 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 here that both the black line (sampling directly from CFD simulations of the steady state) and the red line (sampling from the KL-GPR model) employ the Monte Carlo sampling technique for accurate unbiased statistics. The major difference is that for the red line, our method, samples are evaluated from the surrogate model instead of a full new CFD simulation, which is many orders of magnitude faster.

It is clear that the reconstructed short-term statistical distribution does a better job of capturing both the heavy tail loads, and the asymmetry between positive and negative axial loads in heavier seas. At the same time, though the model outperforms linear estimate, its predictions slightly underestimate the heavy right tail, especially in the  $H_s = 9$  m case. This underestimate is likely from two sources, both ultimately stemming from limited data. First, the  $n = 3$  truncation order for the KL wave episodes truncates some of the high frequency wave energy, in the interest of making the statistical problem tractable. This truncation slightly reduces wave steepness and therefore may underpredict the the hydrodynamic pressure force. We consider this issue more fully below in Section 5.5. Second, our training simulations were actively chosen with  $H_s = 5$  m and  $H_s = 13$  m in mind specifically. While the simulations themselves are reusable due to the spectrum transfer technique, it may have created a ‘gap’ corresponding to waves that contribute most significantly to the extreme loads at  $H_s = 9$  m, but not in more extreme seas.

Finally, in Figure 13, we present the recovered viscous forces  $F_x^v$  and  $F_z^v$ , along with the steady state simulations, for each of  $H_s = 9$  m and  $H_s = 13$  m (the viscous forces are essentially zero in the  $H_s = 5$  m case). Because viscous forces are essentially nonlinear, we do not present linear



**FIGURE 12** Comparison of the hydrodynamic pressure force  $F_x^P$  pdf predicted from KL-GPR-model, for (Top Left)  $H_s = 5\text{m}$ , (Top Right)  $H_s = 9\text{m}$ , and (Bottom)  $H_s = 13\text{m}$ . Monte Carlo direct CFD statistics (black) with  $1\sigma$  bootstrap error bars as well as linear potential flow statistics (blue) included for comparison.



**FIGURE 13** Comparison of KL-GPR-modeled pdfs of viscous forces. Top:  $H_s = 9\text{m}$ . Bottom:  $H_s = 13\text{m}$ . Left: axial viscous force  $F_x^V$ . Right: vertical viscous force  $F_z^V$ . Monte Carlo direct CFD estimates with  $1\sigma$  bootstrap error bars included.

estimates. 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,  $F_x^p$ . Furthermore, the KL-GPR model fails to capture the essential asymmetry of the viscous forces. This is likely due to interaction between the typical wave episode, which has negligible viscous forces, and the energy preserving reduced order model, which pays insufficient attention to the rare, but extreme, viscous loads.

## 5.5 | Effect of energy truncation

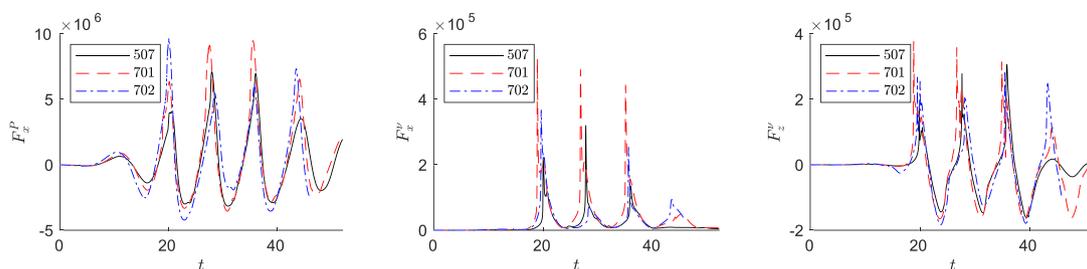
In this study, we construct the wave episodes choosing a truncation order at  $n=3$ . The energy decay associated with the number of KL coefficients are shown in Figure 9, and it is clear that a fraction of the energy is truncated by choosing only the first three coefficients. At this point we examine 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 14, for the basis of our comparison. This wave episode is constructed as previously described, that is, with  $n=3$  nonzero components of the KL vector  $\alpha$ , 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*  $\alpha$  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  $\alpha$ , 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  $\alpha$  from wave episode 507. The corresponding CFD force to wave 701 is also shown in Figure 14. 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 14, we construct wave episode 702 by copying the first three components from  $\alpha$  from wave episode 507, and then drawing the remaining components ( $k \in 4, \dots, 25$ ) 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 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 13. 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



**FIGURE 14** 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  $F_x^v$ . **Right:** Viscous component  $F_z^v$ .

non-Gaussian character,  $F_x^v$  and  $F_z^v$ , (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 affects 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 an 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 an 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 ( $\alpha$ ) 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 short-term 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.

In conclusion, this study's framework shows great promise in addressing design challenges for offshore wind turbine foundations or other offshore applications. By utilizing machine learning, it overcomes limitations of classical modeling, offering innovative ways to analyze complex structures. The introduced surrogate modeling method enables efficient exploration of more design scenarios, including extreme events, while estimating probability density functions effectively. This can contribute to the significant cost reduction and faster design processes, benefiting industry professionals.

This framework's adaptability extends to a range of offshore wind turbine foundations, including jackets, floating structures, and other offshore systems like wave energy converters. All that's needed is a reliable model or experimental setup to produce the required training data. In particular, while high-fidelity models, such as CFD simulations, are essential for capturing nonlinear wave–structure interactions under extreme conditions, the framework remains adaptable to data from experimental campaigns, even to lower-fidelity model when extremes are not a primary concern.

In consideration of higher-dimensional systems, it is acknowledged that the use of GPR may encounter challenges due to the significant increase in data requirements as the number of input features or dimensions grows. The computational expense associated with the inverse of the covariance matrix in GPR becomes prohibitive in such cases. To address this, we propose the utilization of deep neural networks or other machine learning techniques adept at handling high-dimensional data.

## ACKNOWLEDGEMENTS

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.

## PEER REVIEW

The peer review history for this article is available at <https://www.webofscience.com/api/gateway/wos/peer-review/10.1002/we.2880>.

## DATA AVAILABILITY STATEMENT

We have not made use of any proprietary databases in preparing our results. The details of the CFD and surrogate modeling code used is carefully detailed in the text of our submission.

## ORCID

Stephen Guth  <https://orcid.org/0000-0002-7434-5031>

Eirini Katsidoniotaki  <https://orcid.org/0000-0001-5096-3559>

Themistoklis P. Sapsis  <https://orcid.org/0000-0003-0302-0691>

## REFERENCES

1. European Commission. The EU Blue Economy Report 2022, Luxembourg. 2022. Publications Office of the European Union.
2. House TW. FACT SHEET: Biden-Harris administration announces new actions to expand U.S. offshore wind energy. 2022. <https://www.whitehouse.gov/briefing-room/statements-releases/2022/09/15/fact-sheet-biden-harris-administration-announces-new-actions-to-expand-u-s-offshore-wind-energy/>
3. Chung C, Lee J, Yang J-S. National offshore wind strategy for late-mover countries. *Renew Energy*. 2022;192:472-484.
4. Serebrinsky S, Ortiz M. A hysteretic cohesive-law model of fatigue-crack nucleation. *Scripta Materialia*. 2005;53(1):1193-1196.
5. Khan RA, Ahmad S. Dynamic Response and Fatigue Reliability Analysis of Marine Riser Under Random Loads. In: International Conference on Offshore Mechanics and Arctic Engineering, Vol. Volume 2: Structures, Safety and Reliability; Petroleum Technology Symposium; 2007:183-191. <https://doi.org/10.1115/OMAE2007-29235>
6. Chasparis F, Modarres-Sadeghi Y, Hover FS, Triantafyllou MS, Tognarelli M, Beynet P. Lock-in, transient and chaotic response in riser viv. In: International Conference on Offshore Mechanics and Arctic Engineering, Vol. Volume 5: Polar and Arctic Sciences and Technology; CFD and VIV; 2009:479-485. <https://doi.org/10.1115/OMAE2009-79444>
7. Bøhn A. Fatigue loads on large diameter offshore wind monopile foundations in non-operational situations. *Master's Thesis*; 2016.
8. Ortega Nadal A. Time domain simulation parameters for fatigue assessment of an offshore gravity based wind turbine. *Master's Thesis*: NTNU; 2018.
9. Guth S, Sapsis TP. Probabilistic characterization of the effect of transient stochastic loads on the fatigue-crack nucleation time. *Probab Eng Mech*. 2021;66:103162. <https://www.sciencedirect.com/science/article/pii/S0266892021000461>
10. S. D. Offshore. DNV-ST-0126 Support structures for wind turbines; 2021.
11. R. P. D. Offshore. DNV-RP-C203 Fatigue design of offshore steel structures; 2021.
12. IEC61400-1. Wind energy generation system-Part I: Design requirements. Geneva, Switzerland; 2019.
13. Kotz S, Nadarajah S. *Extreme value distributions: theory and applications*: World Scientific; 2000.
14. Fogle J, Agarwal P, Manuel L. Towards an improved understanding of statistical extrapolation for wind turbine extreme loads. *Wind Energy: An Int J Progr Appl Wind Power Convers Technol*. 2008;11(6):613-635.
15. Cheng PW, Van Bussel GJW, van Kuik GAM, Vugts JH. Reliability-based design methods to determine the extreme response distribution of offshore wind turbines. *Wind Energy: An Int J Progress Appl Wind Power Convers Technol*. 2003;6(1):1-22.
16. Wang Y, Xia Y, Liu X. Establishing robust short-term distributions of load extremes of offshore wind turbines. *Renew Energy*. 2013;57:606-619.
17. Dimitrov N. Comparative analysis of methods for modelling the short-term probability distribution of extreme wind turbine loads. *Wind Energy*. 2016;19(4):717-737.
18. Schinas PN, Manolas DI, Riziotis VA, Philippidis TP, Voutsinas SG. Statistical extrapolation methods for estimating extreme loads on wind turbine blades under turbulent wind conditions and stochastic material properties. *Wind Eng*. 2021;45(4):921-938.
19. Müller K, Cheng PW. Application of a monte carlo procedure for probabilistic fatigue design of floating offshore wind turbines. *Wind Energy Sci*. 2018;3(1):149-162.
20. Graf P, Dykes K, Damiani R, Jonkman J, Veers P. Adaptive stratified importance sampling: hybridization of extrapolation and importance sampling monte carlo methods for estimation of wind turbine extreme loads. *Wind Energy Sci*. 2018;3(2):475-487.
21. Morató A, Sriramula S, Krishnan N, Nichols J. Ultimate loads and response analysis of a monopile supported offshore wind turbine using fully coupled simulation. *Renew Energy*. 2017;101:126-143.
22. Wang S, Larsen TJ, Bredmose H. Ultimate load analysis of a 10 MW offshore monopile wind turbine incorporating fully nonlinear irregular wave kinematics. *Marine Struct*. 2021;76:102922.
23. Kamath A, Chella MA, Bihs H, Arntsen OA. Breaking wave interaction with a vertical cylinder and the effect of breaker location. *Ocean Eng*. 2016;128:105-115.
24. Aggarwal A, Bihs H, Myrhaug D, Chella MA. Characteristics of breaking irregular wave forces on a monopile. *Appl Ocean Res*. 2019;90:101846.
25. Zeng X, Shi W, Michailides C, Zhang S, Li X. Numerical and experimental investigation of breaking wave forces on a monopile-type offshore wind turbine. *Renew Energy*. 2021;175:501-519.
26. Ghadirian A, Bredmose H, Dixen M. Breaking phase focused wave group loads on offshore wind turbine monopiles. In: Journal of physics: Conference series, Vol. 753 IOP Publishing; 2016:92004.
27. Paulsen BT, de Sonnevile B, van der Meulen M, Jacobsen NG. Probability of wave slamming and the magnitude of slamming loads on offshore wind turbine foundations. *Coastal Eng*. 2019;143:76-95.
28. Baykal C, Sumer BM, Fuhrman DR, Jacobsen NG, Fredsøe J. Numerical investigation of flow and scour around a vertical circular cylinder. *Phil Trans R Soc A: Math, Phys Eng Sci*. 2015;373(2033):20140104.

29. Tang Z, Melville B, Singhal N, Shamseldin A, Zheng J, Guan D, Cheng L. Countermeasures for local scour at offshore wind turbine monopile foundations: A review. *Water Science and Engineering*; 2022.
30. Watanabe Y, Saeki H, Hosking RJ. Three-dimensional vortex structures under breaking waves. *J Fluid Mech.* 2005;545:291-328.
31. Mohseni M, Soares CG. CFD investigation of submerged geometry effect on wave run-up around a fixed, vertical monopile in regular head waves. *Developments in maritime technology and engineering*: CRC Press; 2021:101-110.
32. Forrester A, Sobester A, Keane A. Exploring and exploiting a surrogate. *Engineering design via surrogate modelling*: John Wiley & Sons, Ltd; 2008:77-107. <https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470770801.ch3>
33. Vu KK, D'Ambrosio C, Hamadi Y, Liberti L. Surrogate-based methods for black-box optimization. *Int Trans Oper Res.* 2017;24(3):393-424. <https://onlinelibrary.wiley.com/doi/abs/10.1111/itor.12292>
34. Rudy S, Fan D, Ferrandis JA, Sapsis T, Triantafyllou MS. Learning Optimal Parametric Hydrodynamic Database for Vortex-Induced Crossflow Vibration Prediction of Both Freely-Mounted Rigid and Flexible Cylinders. In: *International Ocean and Polar Engineering Conference, Vol. All Days*; 2021. ISOPE-I-21-3214.
35. Rasmussen CE, Williams CKI. *Gaussian processes for machine learning*, Adaptive Computation and Machine Learning. Cambridge, MA, USA: MIT Press; Max-Planck-Gesellschaft; 2006.
36. Stevens KW. Adaptive sequential sampling for extreme event statistics in ship design. *Master's Thesis*: Massachusetts Institute of Technology; 2018.
37. Chaloner K, Verdinelli I. Bayesian experimental design: A review. *Stat Sci.* 1995;10(3):273-304.
38. Huan X, Marzouk YM. Simulation-based optimal Bayesian experimental design for nonlinear systems. *J Comput Phys.* 2013;232:288-317.
39. Pickering E, Guth S, Karniadakis GE, Sapsis T. Discovering and forecasting extreme events via active learning in neural operators. *Nature Comput Sci.* 2012;2:823-833. <https://www.nature.com/articles/s43588-022-00376-0>
40. Mohamad MA, Sapsis TP. Sequential sampling strategy for extreme event statistics in nonlinear dynamical systems. *Proc Natl Acad Sci.* 2018;115(44):11138-11143. <https://www.pnas.org/content/115/44/11138>
41. Blanchard A, Sapsis T. Output-Weighted Optimal Sampling for Bayesian Experimental Design and Uncertainty Quantification. *SIAM ASA J of Uncertainty Quant.* 2021;9:564.
42. Sapsis TP. Output-weighted optimal sampling for Bayesian regression and rare event statistics using few samples. *Proc R Soc A: Math, Phys Eng Sci.* 2020;476(2234):20190834. <https://royalsocietypublishing.org/doi/abs/10.1098/rspa.2019.0834>
43. Blanchard A, Sapsis T. Bayesian optimization with output-weighted optimal sampling. *J Comput Phys.* 2021;425:109901.
44. Sapsis TP. Statistics of extreme events in fluid flows and waves. *Ann Rev Fluid Mech.* 2021;53(1):85-111. <https://doi.org/10.1146/annurev-fluid-030420-032810>
45. Boccotti P. On mechanics of irregular gravity waves. *Atti della Accademia Nazionale dei Lincei.* 1989;19:110-170.
46. Phillips OM, Gu D, Donelan M. Expected structure of extreme waves in a gaussian sea. Part I: Theory and SWADE buoy measurements. *J Phys Oceanography.* 1992;23:992-1000.
47. Sclavounos PD. Karhunen-Loève representation of stochastic ocean waves. *Proc R Soc A.* 2012;468(2145):2574-2594.
48. Anastopoulos PA, Spyrou KJ, Bassler CC, Belenky V. Towards an improved critical wave groups method for the probabilistic assessment of large ship motions in irregular seas. *Probab Eng Mech.* 2016;44:18-27. Special Issue Based on Papers Presented at the 7th International Conference on Computational Stochastic Mechanics (CSM7).
49. Anastopoulos PA, Spyrou KJ. Ship dynamic stability assessment based on realistic wave group excitations. *Ocean Eng.* 2016;120:256-263. <https://www.sciencedirect.com/science/article/pii/S0029801816300841>
50. Anastopoulos PA, Spyrou KJ. Evaluation of the critical wave groups method in calculating the probability of ship capsizing in beam seas. *Ocean Eng.* 2019;187:106213. <https://www.sciencedirect.com/science/article/pii/S002980181930397X>
51. Katsidoniotaki E, Ransley E, Brown S, Palm J, Engström J, Götteman M. Loads on a point-absorber wave energy converter in regular and focused extreme wave events. In: *International conference on offshore mechanics and arctic engineering, Vol. 84416 American Society of Mechanical Engineers*; 2020:V009T09A022.
52. Cousins W, Sapsis TP. Reduced order precursors of rare events in unidirectional nonlinear water waves. *J Fluid Mech.* 2016;790:368-388.
53. Farzmand M, Sapsis TP. Reduced-order prediction of rogue waves in two-dimensional deep-water waves. *J Comput Phys.* 2017;340:418-434. <https://www.sciencedirect.com/science/article/pii/S0021999117302553>
54. Guth S, Sapsis TP. 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 Eng.* 2022;266:112633.
55. Dimitrov N, Kelly MC, Vignaroli A, Berg J. From wind to loads: wind turbine site-specific load estimation with surrogate models trained on high-fidelity load databases. *Wind Energy Sci.* 2018;3(2):767-790.
56. Elyasichamazkoti F, Khajehpoor A. Application of machine learning for wind energy from design to energy-water nexus: a survey. *Energy Nexus.* 2021; 2:100011.
57. Ahmad I, Mzoughi F, Aboutaleb P, Garrido I, Garrido AJ. A regressive machine-learning approach to the non-linear complex fast model for hybrid floating offshore wind turbines with integrated oscillating water columns. *Sci Rep.* 2023;13(1):1499.
58. Neary VS, Ahn S, Seng BE, Allahdadi MN, Wang T, Yang Z, He R. Characterization of extreme wave conditions for wave energy converter design and project risk assessment. *J Marine Sci Eng.* 2020;8(4). <https://www.mdpi.com/2077-1312/8/4/289>
59. Wrang L, Katsidoniotaki E, Nilsson E, Rutgersson A, Rydn J, Gtteman M. Comparative analysis of environmental contour approaches to estimating extreme waves for offshore installations for the baltic sea and the north sea. *J Marine Sci Eng.* 2021;9(1):96. <https://www.mdpi.com/2077-1312/9/1/96>
60. de Hauteclocque G, Mackay E, Vanem E. Quantitative comparison of environmental contour approaches. *Ocean Eng.* 2022;245:110374. <https://www.sciencedirect.com/science/article/pii/S0029801821016693>
61. Hasselmann K, Barnett T, Bouws E, et al. Measurements of wind-wave growth and swell decay during the Joint North Sea Wave Project (JONSWAP). *Deut Hydrogr Z.* 1973;8:1-95.
62. Karhunen K. Über lineare Methoden in der Wahrscheinlichkeitsrechnung. *Ann Acad Sci Fennicae Ser A I Math-Phys.* 1947;1947(37):79.
63. Loève M. *Processus stochastiques et mouvement brownien* Levy P, ed.: Gauthier-Villars; 1948.
64. Gerbrands JJ. On the relationships between SVD, KLT and PCA. *Pattern Recog.* 1981;14(1):375-381. 1980 Conference on Pattern Recognition.

65. Goda Y. *Random seas and design of maritime structures*, Vol. 33: World Scientific Publishing Company; 2010.
66. Yang Y, Blanchard A, Sapsis T, Perdikaris P. Output-weighted sampling for multi-armed bandits with extreme payoffs. *Proc R Soc A: Math, Phys Eng Sci*. 2022;478(2260):20210781. <https://royalsocietypublishing.org/doi/abs/10.1098/rspa.2021.0781>
67. Elangovan M. Simulation of irregular waves by CFD. *Int J Mech Mechatron Eng*. 2011;5(7):1379-1383.
68. Windt C, Davidson J, Schmitt P, Ringwood JV. Wave-structure interaction of wave energy converters: A sensitivity analysis. *Proc Inst Civil Eng-Eng Comput Mech*. 2020;173(3):144-158.
69. Katsidoniotaki E, Nilsson E, Rutgersson A, Engström J, Götteman M. Response of point-absorbing wave energy conversion system in 50-years return period extreme focused waves. *J Marine Sci Eng*. 2021;9(3):345.
70. Katsidoniotaki E, Shahroozi Z, Eskilsson C, Palm J, Engström J, Götteman M. Validation of a CFD model for wave energy system dynamics in extreme waves. *Ocean Eng*. 2023;268:113320.
71. Wiener N. Generalized harmonic analysis. *Acta Math*. 1930;55(none):117-258. <https://doi.org/10.1007/BF02546511>
72. Khintchine A. Korrelationstheorie der stationären stochastischen Prozesse. *Mathematische Annalen*. 1934;109(1):604-615. <https://doi.org/10.1007/BF01449156>
73. Naess A, Moan T. *Stochastic dynamics of marine structures*: Cambridge University Press; 2013.
74. Liu S, Ong MC, Obhrai C. Numerical simulations of breaking waves and steep waves past a vertical cylinder at different Keulegan-Carpenter numbers. *J Offshore Mech Arctic Eng*. 2019;141(4):41806.
75. Weller HG. Derivation, modelling and solution of the conditionally averaged two-phase flow equations. Nabla Ltd, No Technical Report TR/HGW 29; 2002.
76. OpenFOAMv1906. <https://www.openfoam.com/news/main-news/openfoam-v1906>, [Online; accessed 13-January-2023].
77. Higuera P, Lara JL, Losada IJ. Realistic wave generation and active wave absorption for Navier-Stokes models: Application to OpenFOAM®. *Coastal Eng*. 2013;71:102-118.
78. Schäffer HA, Klopman G. Review of multidirectional active wave absorption methods. *J Waterway, Port, Coastal, Ocean Eng*. 2000;126(2):88-97.

**How to cite this article:** Guth S, Katsidoniotaki E, Sapsis TP. Statistical modeling of fully nonlinear hydrodynamic loads on offshore wind turbine monopile foundations using wave episodes and targeted CFD simulations through active sampling. *Wind Energy*. 2024;27(1):75-100. doi:10.1002/we.2880

## APPENDIX A: CFD SIMULATIONS

### A.1 | Basic wave parameters

The wave steepness,  $s_w$ , for the irregular waves is defined as<sup>24</sup>

$$s_w = \frac{2\pi H_s}{g T_p^2}, \quad (\text{A1})$$

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

$$KC = \frac{2\pi\eta}{D}, \quad (\text{A2})$$

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, that is, 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.<sup>74</sup> For each wave episode in this study, the wave steepness and the Keulegan-Carpenter number are listed in the Appendix A (Table A2, Table A3, and Table A4).

## A.2 | Numerical methods

### A.2.1 | 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

$$\frac{\partial u_i}{\partial t} + \frac{\partial}{\partial x_j} (u_i u_j) = -\frac{1}{\rho} \frac{\partial p}{\partial x_i} + \nu \nabla^2 u_i + f_b \quad (\text{A3})$$

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

$$\frac{\partial u_i}{\partial x_i} = 0 \quad (\text{A4})$$

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

$$\begin{aligned} \mathbf{u} &= \bar{\mathbf{u}} + \mathbf{u}' \\ p &= \bar{p} + p' \end{aligned} \quad (\text{A5})$$

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

$$\frac{\partial \bar{u}_i}{\partial t} + \frac{\partial}{\partial x_j} (\bar{u}_i \bar{u}_j) = -\frac{1}{\rho} \frac{\partial \bar{p}}{\partial x_i} + \nu \nabla^2 \bar{u}_i + \frac{1}{\rho} \frac{\partial \tau_{ij}}{\partial x_j} + f_b \quad (\text{A6})$$

$$\frac{\partial \bar{u}_i}{\partial x_i} = 0 \quad (\text{A7})$$

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 (Section A.2.1).

### A.2.2 | Free surface modeling

The free water surface is captured using the volume of fluid method (VOF).<sup>75</sup> 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:

$$\frac{\partial \alpha}{\partial t} + \nabla \cdot (\mathbf{u} \alpha) + c_\alpha \nabla \cdot (\mathbf{u}_c \alpha (1 - \alpha)) = 0 \quad (\text{A8})$$

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

$$\rho = \alpha \rho_{water} + (1 - \alpha) \rho_{air} \quad (\text{A9})$$

$$\mu = \alpha \mu_{water} + (1 - \alpha) \mu_{air} \quad (\text{A10})$$

### A.2.3 | Solver and algorithm

In this study, OpenFOAM version 1906,<sup>76</sup> is utilized to model the wave–structure interaction. OpenFOAM is 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, that is, 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.3 | Numerical wave tank

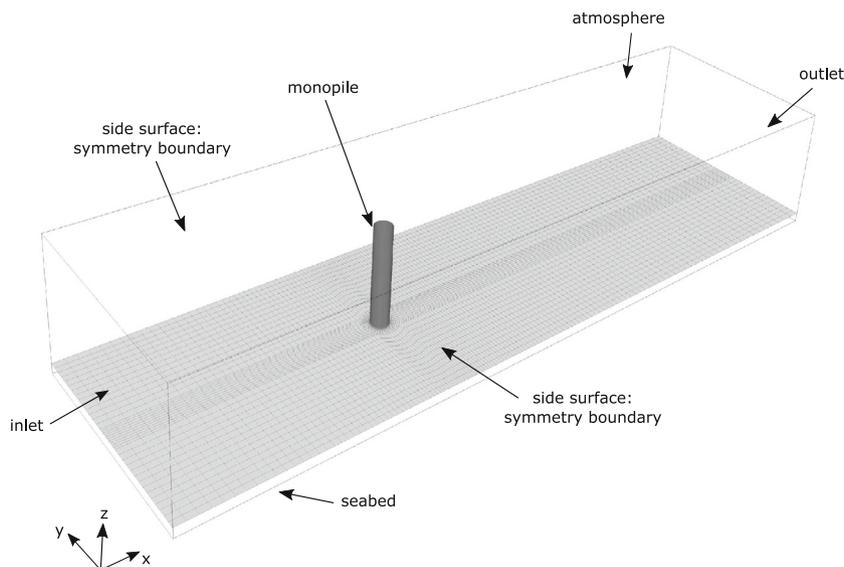
A 3D numerical wave tank (NWT) is setup in order to reproduce the wave–structure interaction modeling. The total length of the NWT was defined by the wavelength and it is equal to 300 m (approximately  $3\lambda$ ), the width is 42 m corresponding to  $6D$ , where  $D$  is the diameter of the monopile. 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 A1 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 of 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.

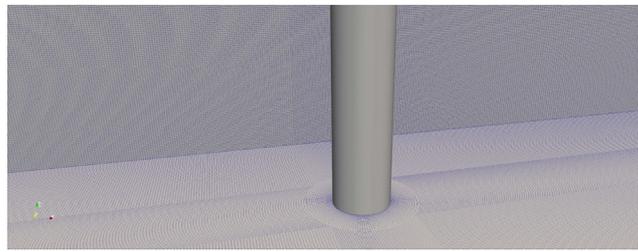
For turbulence modeling, the  $k-\omega$  SST model was implemented with wall functions for the resolution of the boundary layer around cylinder. The boundary conditions are presented in the Table A1. The boundary conditions for the seabed and the side surfaces are type zeroGradient and symmetryPlane, respectively.

### A.4 | 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.<sup>77</sup> The irregular wave is composed of the superposition of regular wave components; that is, the first-order irregular wave episode is generated by the summation of the linear regular waves components. For the wave generation, the static



**FIGURE A1** Boundaries of the computational domain on an example of mesh used for CFD simulations.



**FIGURE A2** Spatial (Figure A2) discretization of the NWT close the monopile.

**TABLE A1** The boundary conditions.

	Monopile	Atmosphere	Inlet	Outlet
<b>U</b>	fixedValue	pressureInletOutletVelocity	waveVelocity	waveVelocity
<b>p</b>	zeroGradient	totalPressure	zeroGradient	zeroGradient
<b><math>\alpha</math></b>	zeroGradient	inletOutlet	waveAlpha	zeroGradient
<b>k</b>	kqRWallFunction	zeroGradient	zeroGradient	zeroGradient
<b><math>\omega</math></b>	omegaWallFunction	zeroGradient	zeroGradient	zeroGradient

boundary method is implemented at the inlet boundary in which the expressions for the free surface elevation (Equation A11) and wave velocities (Equation A12) are applied as Dirichlet boundary conditions based on the linear wave theory. The prescribed velocities determine the pressure and the free surface elevation.

$$\eta = \sum_{i=1}^N \frac{H_i}{2} \cos(k_{xi}x + k_{yi}y - \omega_i t + \phi_i), \quad (\text{A11})$$

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

$$\begin{aligned} u &= \sum_{i=1}^N \frac{H_i}{2} \omega_i \frac{\cosh[k_i(h+z)]}{\sinh(k_i h)} \cos(\theta_i), \\ w &= \sum_{i=1}^N \frac{H_i}{2} \omega_i \frac{\sinh[k_i(h+z)]}{\sinh(k_i h)} \sin(\theta_i), \end{aligned} \quad (\text{A12})$$

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 Schäffer and Klopman.<sup>78</sup>

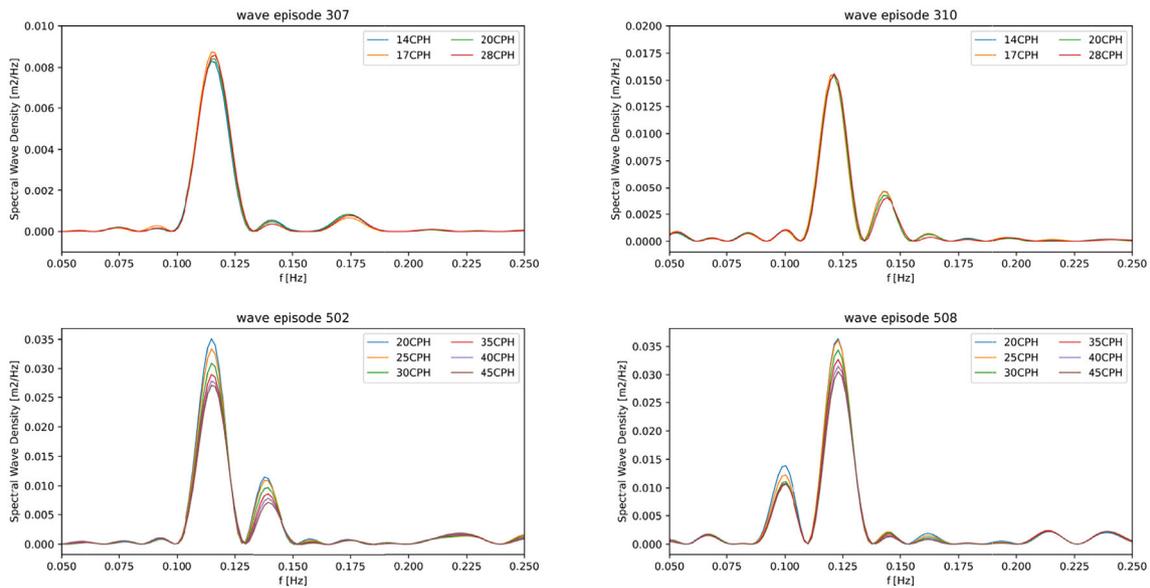
In OpenFOAM, for the construction of the irregular wave episode, the user chooses the *irregularMultiDirectional* wave model and provides the wave height, wave period and phase of each wave component. Further description about the wave generation is provided in the following subsections.

## A.5 | Grid convergence study

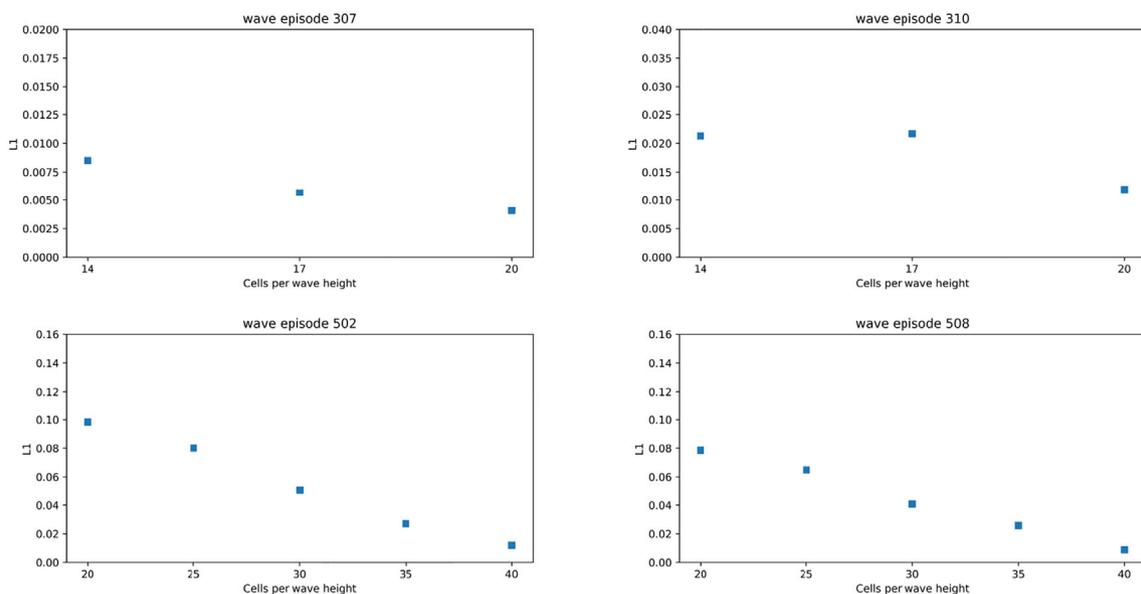
The mesh sensitivity study is conducted without the presence of the monopile foundation. We want to evaluate the wave generation without being affected by the reflection effects from the foundation. The wave elevation is measured at the location where the monopile will be placed

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  $H_s = 5$  m and the extreme sea state with  $H_s = 13$  m. The mesh resolution is different for the waves episodes of each sea state; that is, 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 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); that is, 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 ( $H_s = 5$  m), the resolution is evaluated based on the wave episodes 307 and 310. Four different resolutions are examined, that is, 14 CPH, 17 CPH, 20 CPH, and 28 CPH. For the extreme sea state ( $H_s = 13$  m), six resolutions are examined since the wave episodes are strongly nonlinear, that is, 20 CPH, 25 CPH, 30 CPH, 35 CPH, 40 CPH, and 45 CPH. Figure A3 presents the qualitative



**FIGURE A3** For the mesh sensitivity study, the convergence of the spectra wave density is examined. For the sea state with  $H_s = 5$  m, wave episodes 307 and 310 are evaluated. For the sea state with  $H_s = 13$  m, wave episodes 502 and 508 are evaluated. The resolution is defined as cells per wave height (CPH). For the sea state with  $H_s = 13$  m, more resolutions examined due to the more complex nonlinear waves.

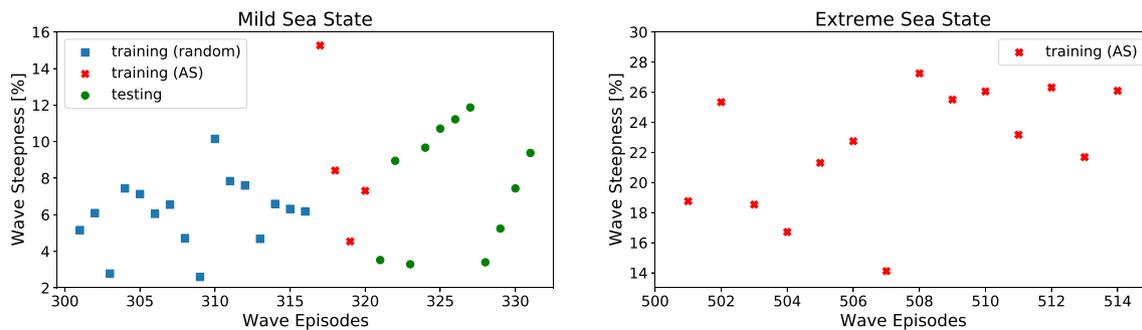


**FIGURE A4** The  $L_1$  norm is considered as the error metric for the evaluation of the spectrum wave density convergence (Figure A3). Each resolution is compared with the finest resolution for each case.

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 quantitatively 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:

$$l_1 = \int_{f_l}^{f_u} |SWD_i(f) - SWD_{finest}(f)| df \quad (\text{A13})$$



**FIGURE A5** 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 (Figure A5).

**TABLE A2** 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.

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

where the *SWD* is the spectral wave density shown in Figure A3, 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 A4 presents the  $I_1$  metric for each wave episode. For the milder sea state, the resolution 20 CPH satisfy  $I_1 < 0.015$ ; therefore, the 20 CPH is used as resolution for all the wave episodes of the milder sea state. For the extreme sea state,  $I_1 < 0.02$  is observed for resolution 40 CPH; therefore, this is used for all the wave episodes.

## A.6 | Wave characteristics

Some critical characteristics of the wave episodes are listed in the Table A2 and Table A3 (mild sea state), and Table A4 (extreme sea states)

**TABLE A3** Mild sea state ( $H_s = 5$  m,  $T = 8$  s): Characteristics of the wave episodes used for testing purposes.

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 A4** 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.

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