Designing for dangerous waves – a new 'Adaptive Screening' method to predict extreme values of non-linear marine and coastal structure responses to waves

van Essen, Sanne M.; Seyffert, Harleigh C.

DOI
10.1016/j.ress.2025.111404

Publication date
2025

Document Version
Final published version

Published in
Reliability Engineering and System Safety

Citation (APA)
van Essen, S. M., & Seyffert, H. C. (2025). Designing for dangerous waves – a new 'Adaptive Screening' method to predict extreme values of non-linear marine and coastal structure responses to waves. *Reliability Engineering and System Safety*, *264*, Article 111404. https://doi.org/10.1016/j.ress.2025.111404

Reliability Engineering and System Safety 264 (2025) 111404

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

Designing for dangerous waves – a new 'Adaptive Screening' method to predict extreme values of non-linear marine and coastal structure responses to waves

Sanne M. van Essen [a,b,*], Harleigh C. Seyffert [a]

[a] Department of Maritime & Transport Technology, Delft University of Technology (TU Delft), Delft, The Netherlands
[b] Ships Department, Maritime Research Institute Netherlands (MARIN), Wageningen, The Netherlands

ARTICLE INFO

Keywords:
Extreme value prediction
Non-linear responses
Ships and offshore structures
Coastal structures
Oceanography
Wave impacts
Design loads
Hydrodynamics
Probabilistic design
Reliability
Non-linear waves
Vertical bending moments
Green water loading

ABSTRACT

Predicting extreme values of strongly non-linear hydrodynamic responses (such as wave impact loads) is crucial for ensuring the safety and reliability of marine and coastal structures. However, this task is challenging due to the complexity and rarity of these responses. Existing methods are often limited to weakly non-linear responses or are very computationally expensive. This paper presents a new multi-fidelity method called 'Adaptive Screening', designed to efficiently predict extreme values of strongly non-linear wave-induced responses. These values are critical inputs for structural design and reliability analysis. Adaptive Screening combines elements of screening, multi-fidelity Gaussian Process Regression, and adaptive sampling. We validate its effectiveness through three applications: predicting the most probable maxima of second-order wave crests, vertical bending moments on a ferry, and green water impact loads on a containership. Our results demonstrate that Adaptive Screening outperforms conventional brute-force methods, achieving comparable accuracy in predicting extreme values while significantly reducing high-fidelity simulation times (especially for the most non-linear cases). Like many alternative methods, Adaptive Screening relies on a response-dependent low-fidelity indicator variable. We also show that the method performs well with realistic indicators for a range of applications. The test cases indicate that Adaptive Screening is very promising for the strongly non-linear responses it was designed for.

1. Introduction and objectives

1.1. Wave impacts

Impulsive 'wave impacts' may occur when a structure is hit by large and steep waves, when a floating structure experiences large wave-induced motions, or a combination of both. The resulting loads can cause significant damage, endanger crew or inhabitants, or decrease performance of the structure. Severe wave impact accidents are documented on e.g., production ships [1], a bulk carrier [2], semi-submersibles [3,4], a drilling rig [5], cruise ships [6,7] and several other ships [8]. Wave impacts also affect the structural reliability of various coastal structures such as breakwaters (e.g., [9,10]), lighthouses (e.g., [11]), bridges (e.g., [12]), complete ports (e.g., [13]) and offshore wind turbines (e.g., [14]). These accidents and studies illustrate that it is still relevant to improve the prediction of wave impact loads and their probability. In this context, we consider stochastic phenomena such as green water and slamming on ships, wave-in-deck impacts on fixed and floating offshore structures and wave impacts on wind turbines, breakwaters, dams, jetty's, bridges and other coastal structures. Wave impacts are strongly non-linear and Fig. 1 illustrates how violent such events can be. Predicting the extreme values of such strongly non-linear responses is therefore essential in designing safe and reliable structures.

1.2. Requirements for an extreme value prediction method for wave impacts

Unfortunately, obtaining extreme values for wave impact loads is a challenge due to their rare and complex nature. We need long simulations to obtain converged statistics of rare events, while at the same time we need high-fidelity (HF) simulations to resolve the complex physics. HF models here are Computational Fluid Dynamics (CFD) or physical experiments. CFD became very good at reproducing wave impact loads in a given wave event (see e.g., [16]), but full Monte Carlo Simulation (MCS) with such HF tools for long durations is presently not feasible in practical design contexts. We therefore need a dedicated extreme value prediction method (EVPM), that can

* Corresponding author at: Department of Maritime & Transport Technology, Delft University of Technology (TU Delft), Delft, The Netherlands.
E-mail addresses: s.m.vanessen@tudelft.nl, s.v.essen@marin.nl (S.M. van Essen), h.c.seyffert@tudelft.nl (H.C. Seyffert).

https://doi.org/10.1016/j.ress.2025.111404
Received 13 November 2024; Received in revised form 28 April 2025; Accepted 26 June 2025
Available online 11 July 2025

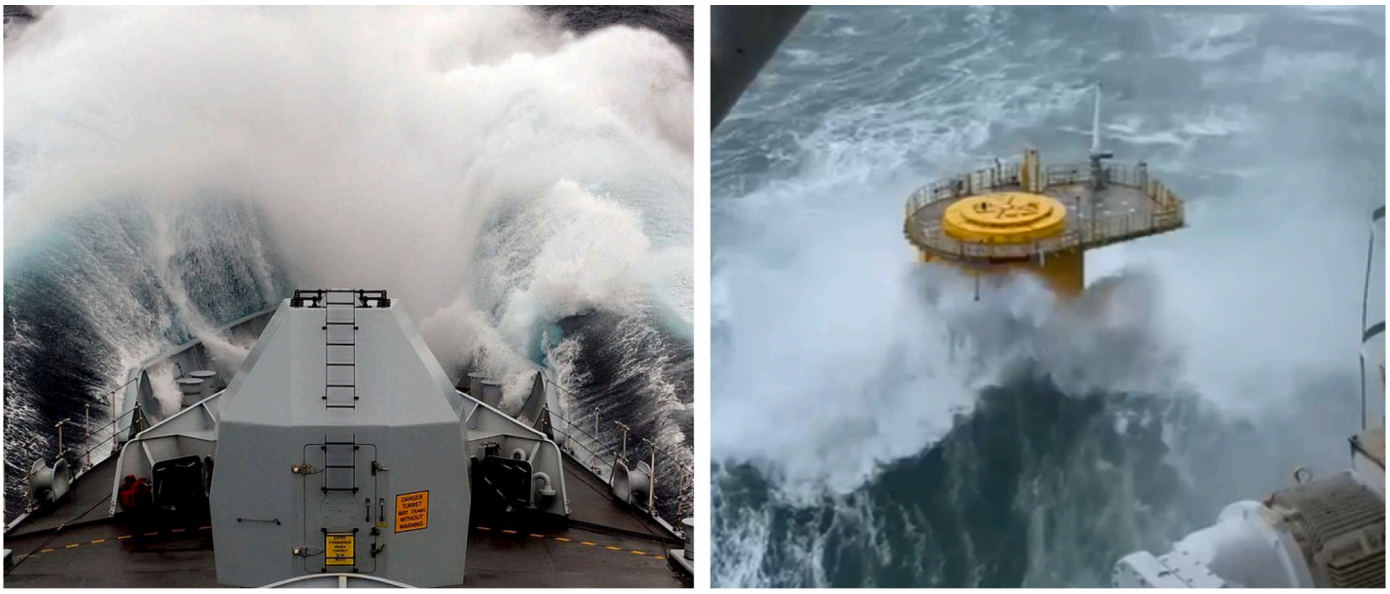


Fig. 1. Two examples of wave impacts on marine structures: a wave impacting *HMS Edinburgh* of the UK Royal Navy in 8 m waves on the South Atlantic (left; courtesy D. Rosenbaum, Royal Navy Media Archive [15], with permission) and a wave impacting a wind turbine foundation with a transition piece ~20 m above the calm water level close to Fécamp during storm Ciarán on 2 Nov. 2023 (right; courtesy K. King, with permission).

reduce this computational burden. Due to the stochastic nature of ocean waves and wave-induced responses, their maximum values increase with exposure duration; the longer the duration at sea, the larger the expected maximum ship response. Consequently, wave response extremes are defined by a probability of exceedance (PoE) or a return period. Keeping this in mind, an EVPM for strongly non-linear responses like wave impacts must meet some criteria:

- The EVPM should be multi-fidelity, integrating elements of HF models for accurate response modelling, and low-fidelity (LF) models to handle the long simulation times required for rare event statistics.
- For strongly non-linear responses, only a few wave events of 10–20 s can realistically be evaluated with a HF model in the design of a structure. The EVPM has to be able to cope with that.
- The wave impact complexity is mainly due to wave complexity (see Fig. 1; steepness, non-linearity, breaking, directionality etc.). Linear wave models disregard much of this complexity. To be able to identify critical wave events, the EVPM's LF model therefore must account for some wave non-linearity.
- Peak responses are not always most relevant; rise times or impulses can be more critical for structural dynamic behaviour. Therefore, the EVPM must consider consistent LF and HF time profiles, not just extreme values.
- The EVPM must be efficient and practical for use in design.

1.3. Long- and short-term statistics

Design for 'seakeeping' (the response of a ship to waves) often distinguishes long-term (over a ship's lifetime in various sea states) and short-term (within single wave conditions) responses (see e.g., [17]). Most available methods either predict short-term extreme values in a given sea state (wave condition), or predict long-term extreme values over all sea states without properly considering short-term variability. Here, we focus on short-term extreme value prediction. The selection of sea states and their long-term statistics can be handled by the environmental contour method [18], or by its improved versions (e.g., [19,20]), as is widely accepted in the offshore industry (e.g., [21]). As the method fails to characterise the joint *pdf* of the environmental variables in some cases [22], alternatively we could select critical sea states using methods such as proposed by Gramstad et al. [23].

1.4. Existing methods and their shortcomings

Various existing EVPMs were reviewed in [24]. Here, we briefly summarise this review for the present audience, add some recent studies and explain why current methods are not directly applicable to strongly non-linear wave-induced responses. Extreme value prediction problems are similar to reliability problems; the latter predict structural failure probabilities, whereas the former predict extreme load values independently of structural response. These quantities are related by Eq. (1), where S are loads, R is the structural 'resistance' of the structure, P_f is its probability of failure and g_l is some limit state function. The review below therefore covers both extreme value and reliability studies.

$$P_f = P(g_l \leq 0) = P((R - S) \leq 0) \tag{1}$$

Design wave or response-conditioning methods (RCMs) generate critical wave event profiles using response transfer functions, wave spectra, and wave phase assumptions for extreme events. They can produce single profiles (e.g., [25]) or multiple profiles that account for random wave backgrounds (e.g., [26–28]). Most of these RCMs are based on the classical reliability methods FORM or SORM. RCMs meet several EVPM criteria from Section 1.2: they are efficient, multi-fidelity, event-based and consider time profiles. However, most classical RCMs rely on linear Gaussian waves and struggle to incorporate significant non-linearity, requiring complex modifications like inverse solutions for higher-order wave terms or the use of higher-order wave models. Some recent studies include wave non-linearity in an RCM by using a higher-order wave model [29,30]. However, they are still applied only to weakly non-linear responses, and they do not include efficient adaptive sampling techniques.

Screening methods use LF indicators (or surrogates) to select critical wave events for HF analysis. Ideally, the indicator's order statistics closely match those of the HF response, allowing wave events to be ranked by their LF response (e.g., [31–35]). This approach allows efficient screening of sea states using MCS of the LF indicator, where identified critical events are then analysed using HF methods like CFD or experiments. van Essen and Seyffert [24] reviewed studies validating various LF indicators for wave impact loads. Screening methods meet several EVPM criteria from Section 1.2: they are multi-fidelity, event-based, can consider time profiles and accommodate weakly non-linear

$\overline{x}$	Mean of variable x
σ_x^2	Variance of variable x
$\hat{x}$	MPM of variable x
CFD	Computational Fluid Dynamics
CRS	Cooperative Research Ships
EVPM	Extreme Value Prediction Method
GPR	Gaussian Process Regression
HF	High-Fidelity
LF	Low-Fidelity
MCS	Monte-Carlo Simulation
MF-GPR	Multi-Fidelity Gaussian Process Regression
MPM	Most Probable Maximum
PoE	Probability of Exceedance
RCM	Response-Conditioning Method
RWE	Relative Wave Elevation
USMV	Acquisition func. (uncertainty sampling + mean value)
C' or C''	LF or HF wave crest height
C_{20}	Coefficient of variation over last 20 iterations
$\mathbf{d}^*$	New PoE prediction range for HF, increasing order
$\mathbf{d}_H^{\text{sel}}$	Est. PoE of selected HF peaks, increasing order
$\mathbf{d}_L^{\text{mcs}}$	PoE of all LF peaks in the MCS, increasing order
$\mathbf{d}_L^{\text{sel}}$	PoE of selected LF peaks, increasing order
$\overline{E}_{20}$	Max. abs. iteration difference, averaged over last 20
F_x' or F_x''	LF or HF green water load
H_s	Significant wave height
$\mathbf{h}^*$	Predicted HF peak values over range $\mathbf{d}^*$
$\mathbf{h}^{\text{sel}}$	HF peak values in the selected events
$\mathbf{l}^{\text{mcs}}$	All LF peak values in the MCS, decreasing order
$\mathbf{l}^{\text{sel}}$	Selected LF peak values, decreasing order
m	Number of initially selected HF samples
n	Number of LF indicator peaks in the MCS
n_w	Number of wave encounters in the MCS
N	Number of wave seeds in MCS
P	Probability
P_{exp}	PoE level corresponding to Texp and wave period
R	Monotonicity acceptance criterion
R'_{deck}	LF relative wave elevation on deck
T_{exp}	Target/exposure duration (typically 20 min–3 h)
T_p	Peak wave period
$T_{p,e}$	Peak encounter wave period
V_s	Ship forward speed
V' or V''	LF or HF hogging vertical bending moment
γ	Peak enhancement factor JONSWAP spectrum
ϵ_1	Stopping limit for $\overline{E}_{20}$
ϵ_2	Stopping limit for C_{20}
μ	Wave heading w.r.t. structure
ω	Wave frequency

Box I. Most important nomenclature.

LF models. They also seem suited for strongly non-linear responses. However, they are yet not very efficient and existing research has focused more on indicator selection than on handling extreme value statistics.

Sampling techniques can be used to reduce the number of required simulations compared to MCS. Such methods include importance sampling (e.g., [36–38]), subset sampling (e.g., [39]), adaptive sampling or combinations (e.g., [40]). Adaptive sampling, also known as sequential sampling or Bayesian design of experiments, combines surrogate modelling with data-driven sampling strategies that 'learn' where to sample next. Surrogates can be constructed in many different ways: using simple polynomial models, polynomial chaos expansion (e.g., [41,42]), support vector regression (e.g., [43,44]) or neural networks (e.g., [45,46]). In particular, Gaussian Process Regression (GPR or 'kriging') is well-suited for problems with sparse data, as it provides both predictions and associated uncertainty estimates. These uncertainty bands can be used to effectively guide sampling [47]. GPR can be applied in single- or multi-fidelity forms [48,49]. GPR-based adaptive sampling methods were applied to maritime extreme value problems by Gramstad et al. [23], Guth and Sapsis [50], Guth et al. [51], Tang et al. [52] and Abaei et al. [53] (using single-fidelity GPR) and by Guth et al. [54] (using multi-fidelity GPR). Acquisition functions to guide the sampling may target extreme values in distribution tails (e.g., [55]), and sampling is typically stopped when predicted values stabilise (e.g., [56,57]). Kim et al. [58] studied multi-fidelity sampling algorithms in the context of extreme ship motions. Response Surface Methods (RSM) for reliability problems are similar: they adaptively update a 'response surface function' (a surrogate for limit state function g_l in Eq. (1)). Again, this can be a polynomial (e.g., [59–61]), but this can lead to errors for strongly non-linear systems. GPR (e.g., [62,63]) or neural networks (e.g., [46,64,65]) can offer better accuracy. Marrel and Iooss [66,67] discuss ways to optimise GPR surrogates in the context of reliability problems. Adaptive sampling methods with GPR meet many EVPM criteria from Section 1.2: they are efficient, can be multi-fidelity, can handle weakly non-linear models and some examples consider time profiles. However, their elements and implementation are very case-specific, and none of the examples satisfies all requirements. They have yet to be applied to strongly non-linear wave-induced responses.

Due to the mentioned EVPM limitations for strongly non-linear responses, maritime designers typically follow established classification society guidelines. Organisations like maritime classification societies and the International Towing Tank Conference (ITTC) traditionally recommend physical experiments using several 0.5 to 3-h wave seeds for direct assessment of short-term extreme response values. We call this the 'conventional' industry approach, with guidelines summarised in Appendix D.

1.5. Paper objectives & novelty

As discussed above, we need a new EVPM for wave impacts (and other strongly non-linear wave-induced responses), because available methods are not suitable, efficient, and accurate enough, and accidents related to wave impacts still happen. The objectives of this paper are therefore to:

1. Introduce a new EVPM for strongly non-linear responses that complies with the requirements in Section 1.2.
2. Show that the method can accurately and efficiently predict extreme values for a range of realistic applications.

To this end, we introduce a novel method called Adaptive Screening. The method combines elements from screening, multi-fidelity GPR, and adaptive sampling, considering the requirements in Section 1.2. A pilot study for the new approach was presented in [68]; several important improvements have since been implemented.

In theory, Adaptive Screening can predict both long- and short-term extreme values, but our focus here is on the short term. Specifically, we predict the short-term most probable maximum (MPM) value, which represents the likeliest extreme within a given exposure duration. We

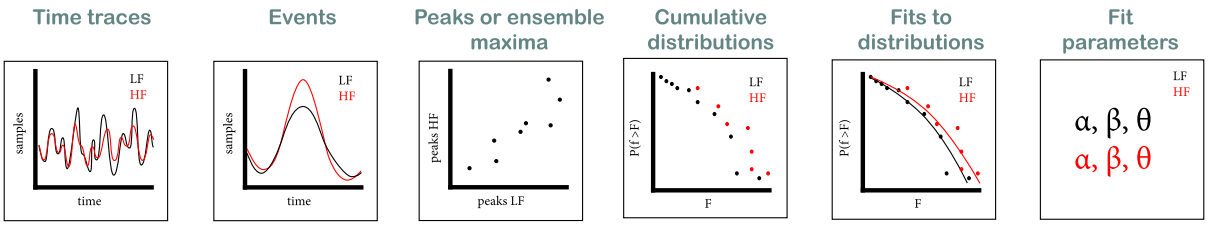


Fig. 2. Some of the possible 'statistical levels' where multi-fidelity methods can derive or learn the relation between a LF indicator (black) and HF non-linear response (red). Typical examples, the overview is not exhaustive.

assume that critical sea states were already selected beforehand with one of the methods discussed in Section 1.3 or Adaptive Screening with lower fidelity levels. Using MPM for design is common in maritime engineering (e.g., [69–71]). We also (qualitatively) evaluate the predicted distributions over a larger range of values than only the MPM to make sure that the MPM prediction is solid and converged.

To demonstrate the effectiveness of Adaptive Screening, we present three applications: (1) a weakly non-linear case predicting second-order wave crest extremes, (2) an intermediate case predicting extreme vertical bending moments on a ferry, and (3) a strongly non-linear case predicting green water impact loads on a containership. The complexity of the cases builds up. Case 1 is weakly non-linear: we consider only waves and the LF and HF data are the same variable. Case 2 is more non-linear and complex: we consider wave-induced responses, but the LF and HF data are still the same variable. Finally, case 3 is strongly non-linear and complex: we consider wave-induced responses and the LF and HF data are different variables. Case 3 therefore covers the complexity of the original problem the method was designed for (wave impact loads). The first two applications allow for method tuning and validation, while the third demonstrate the method's capability with strongly non-linear responses. Especially cases 2 and 3 illustrate real challenges designers face when developing safe and reliable structures exposed to waves. Both are maritime examples, but case 2 is representative for assessment of bending moments on any floating structure, and case 3 is representative for wave impact load assessment on any fixed or floating structure (breakwaters, bridges, wind turbines, etc.). Such assessments form an important element of the reliability analysis of these structures; they provide input for S in Eq. (1).

The novelty of Adaptive Screening lies in its specific design for strongly non-linear rare responses. It combines adaptive sampling with GPR applied directly to the distribution shape (to target the tail of the distribution) and with screening elements that allow for (weakly) non-linear low-fidelity tools. Another novelty of the present study is that we validate the EVPM a.o. with high-fidelity experimental wave impact data - a relevant case study for design and reliability of structures exposed to severe waves.

The method is introduced in Section 2, with applications in Sections 3 to 5, and final discussion and conclusions in Section 6.

2. New method: Adaptive screening

Multi-fidelity methods learn relationships between LF and HF responses at different 'statistical levels', as shown in Fig. 2. Methods that generate critical wave events, like RCM or some adaptive sampling implementations (e.g., [50,51,54]) are positioned on the left, where relationships retain detailed wave event data. Screening methods or approaches that select events from databases (e.g., [23] at a sea state- instead of event-level) are on the right, simplifying the problem to single-variate regression and reducing HF data requirements. E.g., [51] shows that efficient application of GPR to full wave event time traces requires a reduction of the number of input wave components, which can lead to underestimated response distribution tails. We therefore developed Adaptive Screening at the 'level' of cumulative distributions. We use (multi-fidelity) GPR to construct a surrogate for the HF exceedance probability distribution, combined with adaptive selection of new HF samples guided by the screening analysis. GPR was chosen over e.g., polynomial regression and neural networks because it (a) is efficient with limited data and (b) provides uncertainty bands that guide adaptive sampling. Additionally, GPR is very fast with limited numbers of samples. We first present the steps and detailed formulations of Adaptive Screening, after which we discuss the utilised assumptions and implementation.

2.1. Steps

A schematic overview of Adaptive Screening is provided in Fig. 3, and the method involves the steps described below. The numbers in the figure roughly correspond to these steps. Box I includes most notations used below.

Step 1 Define a LF indicator with a strong statistical relation to the target HF response, as would be done in a screening method. The indicator signal is not necessarily the same signal at a different fidelity level as the target HF response; it can also be another signal with similar order statistics as the target HF response. An ideal LF indicator has identical order statistics as the target HF non-linear response, where the highest LF indicator value appears in the same wave event as the highest HF non-linear response value, and so on. See [24] for a review of suitable wave impact indicators.

Step 2 Perform LF Monte-Carlo Simulations (MCS) for a large number N of wave seeds, each with the same exposure duration T_{exp}. This exposure duration is the duration for which you want to obtain the extreme value of the response in a single sea state. This is generally the duration for which the hydrodynamic response is expected to remain ergodic (roughly 15 min to 3 h, depending on wave ergodicity, speed and course changes, etc.). The total MCS should be significantly longer than T_{exp} to obtain converged extreme values. See [72,73] for an example of the required N for wave crests, green water impact forces and wave-in-deck impact forces. The total MCS duration T_{tot} follows from $T_{\text{tot}} = NT_{\text{exp}}$.

Step 3 Identify all n LF indicator peaks in the full MCS duration T_{tot}. This can be done in different ways; here we use peak-over-threshold crossings. Also find the number of zero up-crossing encountered wave crests n_w within T_{tot}, which can be estimated from Eq. (2) if there is no explicit wave record available. Here, $T_{p,e}$ is the peak wave encounter period, which is different from the peak wave period T_p if the (ship) response has forward speed. They are related using the absolute wave frequency $\omega = 2\pi/T$ and Eq. (3), where ω_e is the wave encounter frequency, V_s is the forward speed of the ship and μ is the wave heading with respect to the ship (π is head waves in the sign convention). n_w is the basis for all distributions presented from here on, which enables us to use combinations of LF indicators and HF responses with a different number of peaks.

$$n_w \approxeq NT_{\text{exp}}/T_{p,e} \tag{2}$$

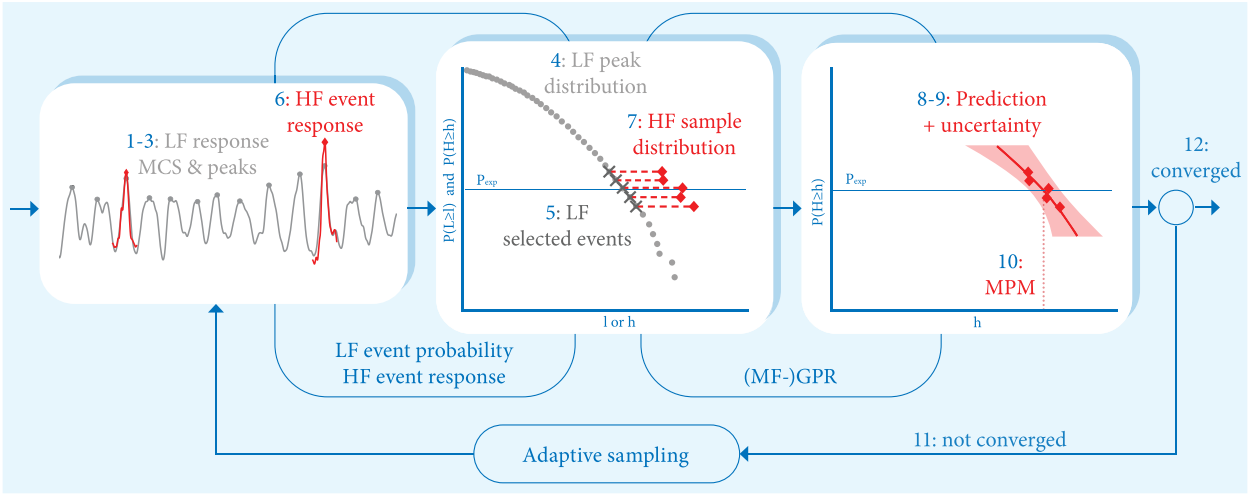


Fig. 3. Schematic illustration of the new Adaptive Screening method. The numbers in this figure roughly correspond to the method steps in Section 2.1. The left plot only shows a small part of the MCS time traces, and only a few HF samples are included in the middle and right distributions to illustrate the principle.

$$\omega_e = \omega - \omega^2 V_s cos(\mu)/g \tag{3}$$

Step 4 Calculate the LF probability of exceedance (PoE) for all indicator peaks $\mathbf{d}_L^{\text{mcs}} = \{d_{L,i}^{\text{mcs}} | i = 1, 2, \dots, n\}$, related to the number of wave encounters, by applying Eq. (4). The largest LF indicator peak value has a PoE of $1/n_w$ and the smallest n/n_w. We now have a LF indicator peak PoE distribution dataset $\left[\mathbf{d}_L^{\text{mcs}}, \mathbf{l}^{\text{mcs}}\right]$, where $\mathbf{l}^{\text{mcs}}$ contains all indicator peak values, in descending order corresponding to the ascending PoE values in $\mathbf{d}_L^{\text{mcs}}$.

$$\mathbf{d}_L^{\text{mcs}} = P(\mathbf{l}^{\text{mcs}} \geq l) \cdot \frac{n}{n_w} \tag{4}$$

Step 5 Select initial samples from the LF MCS dataset of the previous step. The selected set is called $\left[\mathbf{d}_L^{\text{sel}}, \mathbf{l}^{\text{sel}}\right]$, where $\mathbf{d}_L^{\text{sel}} = \{d_{L,k}^{\text{mcs}} | k = 1, 2, \dots, m\}$ and $\mathbf{l}^{\text{sel}} = \{l_k^{\text{mcs}} | k = 1, 2, \dots, m\}$. Different sampling strategies can be used to determine these indices k. Here we select these events around the PoE corresponding to T_{exp} based on the LF MCS dataset: the 'probability of interest'. This P_{exp} is given in Eq. (5). We pick m PoE values that span a range around P_{exp} and call them $[p_1, p_2, \dots, p_m]$. Now we use Eq. (6) to add the elements in $\mathbf{d}_L^{\text{mcs}}$ closest to these values to the selected set, and the corresponding elements from $\mathbf{l}^{\text{mcs}}$. The resulting selected set $\left[\mathbf{d}_L^{\text{sel}}, \mathbf{l}^{\text{sel}}\right]$ is a subset of the full available LF MCS set $\left[\mathbf{d}_L^{\text{mcs}}, \mathbf{l}^{\text{mcs}}\right]$. Note that selecting samples around P_{exp} is also used in alternative RCM's such as those by Torhaug et al. [27] and Dietz [74].

$$P_{\text{exp}} = N/n_w \approxeq T_{p,e}/T_{\text{exp}} \tag{5}$$

$$d_{L,k}^{\text{mcs}} = \left[\arg \min_{d \in \mathbf{d}_L^{\text{mcs}}} |p_k - d|\right] \text{ for } k = 1, 2, \dots, m \tag{6}$$

Step 6 Find the corresponding HF response for the wave events corresponding to $\left[\mathbf{d}_L^{\text{sel}}, \mathbf{l}^{\text{sel}}\right]$, by running CFD calculations or experiments for these selected events. This new dataset of HF samples is called $\mathbf{h}^{\text{sel}} = \{h_k | k = 1, 2, \dots, m\}$, here h_k is the HF non-linear response value maximum for event k.

Step 7 Estimate the sample HF distribution. This is done by assuming that the order statistics of $\mathbf{l}^{\text{sel}}$ and $\mathbf{h}^{\text{sel}}$ are identical. In other words, we assume that HF value h_k in event k from Step 6 is equally likely as the selected LF value l_k^{mcs} in the same event from Step 5. This is a critical screening assumption, stating that the HF distribution $[\mathbf{d}_H^{\text{sel}}, \mathbf{h}^{\text{sel}}]$ can be estimated using $\mathbf{d}_H^{\text{sel}} \approx \mathbf{d}_L^{\text{sel}}$. This only works if a suitable indicator signal is chosen in Step 1.

Step 8 Define a range of PoE between 1 and 0 where you want to estimate the HF distribution $\mathbf{h}^*$. We select this prediction range $\mathbf{d}^*$ between approximately a factor ten below and above P_{exp}, defined in high to low PoE order.

Step 9 Construct the surrogate HF distribution over ln ($\mathbf{d}^*$). We use the logarithm to construct a surrogate that focuses on the tail of the HF distribution. The estimated HF distribution from Step 7 only contains a few samples. We apply 1D single- or multi-fidelity GPR to the available samples to construct the surrogate HF distribution $\mathbf{h}^*$ (including uncertainty) over this range.

- With single-fidelity GPR, we use the HF sample dataset $[\ln(\mathbf{d}_H^{\text{sel}}), \mathbf{h}^{\text{sel}}]$ from Step 7 as input. The utilised GPR formulations are in Appendix B.1.
- With multi-fidelity MF-GPR, we use the same HF sample dataset $[\ln(\mathbf{d}_H^{\text{sel}}), \mathbf{h}^{\text{sel}}]$ from Step 7 as input, *and* LF dataset $\left[\ln(\mathbf{d}_L^{\text{mcs}}), \mathbf{l}^{\text{mcs}}\right]$ from Step 3. We used the linear autoregressive (AR1) multi-fidelity model with two levels [49]. The utilised MF-GPR formulations are in Appendix B.2.

For GPR and both parts of MF-GPR we used the Matern32 kernel; details can also be found in Appendix B. As it will not be possible to define a perfect indicator for most non-linear response problems, the HF sample data set $[\ln(\mathbf{d}_H^{\text{sel}}), \mathbf{h}^{\text{sel}}]$ will be 'noisy'. To avoid overfitting, we constrained the noise variance in both GPR and MF-GPR with a lower limit. For the same reason, and to help ensure a monotonic distribution, we also constrained the length-scale hyperparameter of all kernels with a lower limit. Both constraints are explained in detail in Appendix B.3. The result of the (MF-)GPR procedure is the HF prediction $[\ln(\mathbf{d}^*), \mathbf{h}^*]$. To avoid confusion, we emphasise that the regression model is linear and Gaussian, not the underlying HF process itself. We still predict extreme values of strongly non-linear responses.

Step 10 Estimate the target short-term extreme value from the (MF-) GPR prediction. As explained in Section 1.5, our target is the HF MPM value $\hat{H}$, found using Eq. (7). The uncertainty of the MPM can be derived from the predicted uncertainty band by GPR in the same way. This is illustrated by the horizontal line in the right inset of Fig. 3.

$$\mathbf{d}^*(\hat{H}) = P_{\text{exp}}, \quad \text{therefore:} \quad \hat{H} = \mathbf{d}^{*-1}(P_{\text{exp}}) \tag{7}$$

Step 11 Start the adaptive sampling procedure, iterating over Step 5 to Step 11. In each iteration, an acquisition function is applied to define a new sample, unless convergence is reached. As we already guided the prediction in steps Step 8 (by choosing a prediction range) and Step 9 (by focusing on the tail of the distribution using the logarithm of PoE), we selected an adaptive strategy focusing on exploration rather than exploitation (see e.g., [47]). The new samples are selected from $\left[\mathbf{d}_L^{\text{mcs}}, \mathbf{l}^{\text{mcs}}\right]$ defined in Step 4, without replacement. The required formulations are discussed in Section 2.2 (acquisition function) and Section 2.3 (stopping criterion). When a new HF sample is defined as discussed there, it is added to the LF selected set $\left[\mathbf{d}_L^{\text{sel}}, \mathbf{l}^{\text{sel}}\right]$ in Step 5, after which Step 6 to Step 11 are repeated.

Step 12 When convergence is reached according to the criterion in Section 2.3, the result is the converged prediction for the HF distribution $\mathbf{h}^*$ over prediction range $\mathbf{d}^*$, the associated MPM value $\hat{H}$, and the MPM uncertainty.

The steps of Adaptive Screening are presented in general terms, making the method adaptable to any strongly non-linear response with a suitable indicator. Step 1 to Step 6 focus on screening, Step 9 on GPR and Step 11 on adaptive sampling. A key advantage of the new method is that integrating (MF-)GPR with screening enables efficient generation of new HF samples at targeted low PoE values, avoiding the need for extensive HF simulations. This would not be possible with GPR alone. This adaptation makes Adaptive Screening especially effective for strongly non-linear responses. Using MF-GPR instead of GPR is expected to boost efficiency since MF-GPR leverages LF statistics both in event selection (Step 5) and prediction (Step 9).

To apply this general method, several choices must be made: the indicator variable and its calculation tool in Step 1, the MCS length in Step 2, the event definition method in Step 3, the selection of initial critical events in Step 5, the HF tool and initialisation of this tool based on the LF wave event in Step 6, the type of GPR and its settings in Step 9, the target extreme value in Step 10, the acquisition function, number of samples per iteration, and stopping criterion in Step 11. The following sections outline broadly applicable options for the acquisition function and stopping criterion, along with guidance on initialising critical events in the HF tool. A detailed explanation of problem-dependent choices for three case studies will be provided in Sections 3 to 5.

The MPM in Step 10 corresponds to the $q = 0.368$ quantile of the short-term distribution for linear Gaussian signals [75]. The offshore industry often prefers higher quantiles for design. We can estimate these by replacing P_{exp} in Eq. (7) with $P_q = 1 - q^{m/n}$ [76]. However, this necessitates an even longer LF MCS to achieve converged results. Further research is needed to evaluate the applicability of the method for higher quantiles.

A challenging aspect of screening-based methods is initialising HF events in Step 6. We use LF linear or weakly non-linear wave events as input for HF calculations, raising the question of how to define equivalent HF wave input conditions. Potential solutions include the event matching procedures by Johannessen and Lande [77] and Gramstad et al. [78] or using coarse mesh CFD as a LF tool, as demonstrated by van Essen et al. [35]. These approaches require further investigation, which is beyond the scope of this paper.

2.2. Adaptive sampling

Since HF event calculations are costly for strongly non-linear responses, the acquisition function in Step 11 should minimise the number of required HF samples. We define an acquisition function that selects one new HF sample per iteration, even when used with MF-GPR. Combining the notation from Section 2.1 and Appendix B, we denote the mean HF prediction for the current iteration as $\overline{\mathbf{h}^*}$ and its predicted variance as $\boldsymbol{\sigma}^2_{h^*}$. Below, $p^{\text{usmv}}_{\text{new}}$ represents the optimal PoE for adding a new sample to $\mathbf{d}_L^{\text{sel}}$ in Step 5. Insights from the pilot study with fixed samples [68] were used to design our acquisition function, named 'USMV', as defined in Eq. (8). This function balances selecting the best new sample with the highest predicted variance from the previous iteration (Uncertainty Sampling) while guiding the sample toward the PoE level with the largest mean value from the same iteration. The function tends to favour HF values in the distribution's tail. As said, $p^{\text{usmv}}_{\text{new}}$ would be the best new sample value to add to $\mathbf{d}_L^{\text{sel}}$ in Step 5. However, in Adaptive Screening, we select the next sample from the available LF MCS sample pool of Step 4. We therefore always select the sample from $\mathbf{d}_L^{\text{mcs}}$ with the closest value, without replacement (see Eq. (9)). The new sample $d^{\text{mcs}}_{L,\text{new}}$ is added to the existing LF pool $\left[\mathbf{d}_L^{\text{sel}}, \mathbf{l}^{\text{sel}}\right]$ in Step 5.

$$p^{\text{usmv}}_{\text{new}} = \text{argmax}\left[\boldsymbol{\sigma}^2_{h^*} \cdot \overline{\mathbf{h}^*}\right] \tag{8}$$

$$d^{\text{mcs}}_{L,\text{new}} = \left[\arg\min_{d \in \mathbf{d}_L^{\text{mcs}}} |p^{\text{usmv}}_{\text{new}} - d|\right] \tag{9}$$

2.3. Stopping criterion

We further limit the number of HF samples via the stopping criterion. Such a criterion can be based on practical considerations (available time or computational resources), comparison with a target outcome, the relative improvement of results across iterations (e.g., [47]) or on all predicted outcomes for the next iteration (e.g., [56]). Since a fully converged Adaptive Screening prediction cannot achieve 100% accuracy due to the imperfect indicator, we chose to stop when the predicted PoE distributions no longer show significant changes. This criterion consists of three parts. The first part checks whether the predicted distribution is a proper PoE distribution. The second part sets a limit for the maximum absolute difference between each set of subsequently predicted distributions. Because this difference may be far from P_{exp}, a third convergence criterion was based on the coefficient of variation (COV) of the MPM value over the last iterations. The complete formulations for this stopping criterion are detailed in Appendix C.1.

2.4. Assumptions

The key assumption in our screening method is the similarity between the order statistics of the LF indicator and the HF response in Step 7. The validity of this assumption hinges on the chosen indicator; a poor indicator can significantly reduce result accuracy. Most alternative methods, such as RCM and other screening techniques discussed in Section 1.4, rely on similar assumptions. van Essen and Seyffert [24] review suitable indicators for wave impact loads. Another critical assumption in Step 6 is that the HF tool accurately calculates the true HF event response. Previous studies have shown that CFD can effectively predict wave impact loads if wave kinematics and ship motions are modelled well (see e.g., [16,79,80]). Additionally, van Essen et al. [35] demonstrated that screening results could serve as effective inputs for such calculations. In Step 9, GPR also assumes a degree of surrogate smoothness, which is reasonable for a PoE distribution. MF-GPR further assumes that the LF and HF distribution shapes are similar, reinforcing the screening assumption. The utilised GPR formulations assume homoscedastic noise, meaning noise is uniform across $\mathbf{d}^*$ from Step 8. The HF 'noise' in Adaptive Screening comes from two sources: the stochastic (aleatoric) variability of the HF response, and the LF indicator quality. A bad indicator leads to increased scatter in HF samples. While both noise sources are likely heteroscedastic in reality, with increasing noise levels for higher response values (see [72,73]), restricting $\mathbf{d}^*$ to a small range around P_{exp} allows for a reasonable assumption of uniform noise distribution. This also enhances method efficiency by potentially reducing the number of HF points needed for convergence. Finally, we assume that the monotonicity criterion in Eq. (31) combined with the definition of $\mathbf{d}^*$ between 1 and 0 ensures the prediction of a proper exceedance probability distribution.

2.5. Implementation

We used Python toolboxes GPy ([81], v1.10.0) and Emukit ([82], v0.4.9) to define and optimise the (MF-)GPR formulations. The standard Emukit structure was used to combine adaptive sampling with GPR. This was not straightforward for MF-GPR. Adaptive sampling with MF-GPR was therefore achieved by creating a new surrogate 'mock' GPR prediction in each iteration, trained on the MF-GPR predictions and their variances.[1] The acquisition function was then applied to this mock GPR prediction to choose a new point. Afterward, MF-GPR was applied again to the updated training set in each iteration. The acquisition function and stopping criteria were newly implemented. All scripts are available in the 4TU repository: van Essen and Seyffert [83].

3. Application 1: second-order wave crest heights

In application 1 we predict extreme values of second-order wave crest heights. This can be seen as a weakly non-linear HF 'response'. The goal was to predict the HF one-hour MPM value, so $T_{\text{exp}} = 3600$ s. The test case used a long-crested JONSWAP wave spectrum with $H_s = 10$ m, $T_p = 11$ s, $\gamma = 3.3$. With $s = 2\pi H_s/(gT_p^2) = 0.053$, the waves were relatively steep, and the water depth was set to 30 m to enhance second-order effects. The HF response was a second-order wave, and the LF indicator a linear Gaussian wave with or without noise. To keep the root-mean-square error of the one-hour MPM wave crest height within 3% of H_s, a minimum of 8–22 one-hour wave seeds (random realisations) is required in a MCS according to van Essen et al. [72] and Scharnke et al. [73]. To be conservative, we used 50 one-hour seeds (so $N = 50$ and total duration $T_{\text{tot}} = 50$ hours). We define C' as the LF wave crest height, C'' as the HF wave crest height, and $\widehat{C''}$ as its one-hour MPM value. For this test case, we selected an analytically traceable HF response, allowing easy generation of long HF ground truth time traces and avoiding the need for CFD simulations for Step 6 of Adaptive Screening. This approach enabled us to validate the Adaptive Screening framework without introducing CFD calculation accuracy concerns. Sections 3.1 and 3.2 describe the HF and LF data sources, while Section 3.3 explains adaptive sampling for this application. For comparison, the conventional industry method for this application is described in Section 3.4. Results and computational time are discussed in Sections 3.5 and 3.6.

The probability of interest $P_{\text{exp}} = 3.06 \times 10^{-3}$ for this application follows from Eqs. (2) and (5), where $n_w \cong 1.64 \times 10^4$ is the approximate number of wave crests in the 50-h MCS at zero forward speed (so $T_{p,e} = T_p$).

3.1. High-fidelity data

We generated second-order wave elevation time traces analytically using Python toolbox PySeaWave from the Cooperative Research Ships (CRS), based on Sharma and Dean [84], the random phase method and a frequency bandwidth of 0–5 rad/s for second-order interactions. The zero up-crossing wave crests in these traces were used as HF data. To confirm convergence of the one-hour MPM, we generated 10 random 50-h realisations. The true HF one-hour MPM for the chosen 50-h realisation was 10.56 m, with a mean of 10.55 m across the realisations and U95% uncertainty of 0.10 m.

[1] Similar as all other GPRs in the present study, the mock GPR used a Matern32 kernel, and restricted Gaussian noise between 0.5–1.5 times the variance of the input data after subtracting a linear trend.

3.2. Low-fidelity indicator data

The basic LF indicator, 'GoodInd', was derived from linear wave time traces that deterministically correspond to the second-order waves described above. The true LF and HF zero up-crossing peak PoE distributions for the 50-h MCS are shown in Fig. 15(a) in Appendix E. This LF GoodInd distribution followed from Step 3 of Adaptive Screening, and the HF distribution was generated similarly. The HF results across ten 50-h realisations show that the one-hour MPM (at P_{exp}) is well-converged. The LF distribution matches the linear Rayleigh distribution, and the HF distribution the second-order distribution of Forristall [85],[2] confirming correct MCS setup. Linear wave crests are a good indicator to predict the occurrence of second-order wave crests. In a more non-linear response problem, the order statistics of the LF indicator and HF response are expected to be less similar (see e.g., [32,33,35]). To account for this, we defined a second, noisier LF indicator called 'WorseInd'. The extra noise was created by adding a secondary JONSWAP wave system with random, independent phases and $H_s = 2.0$ m, $T_p = 14.0$ s and $\gamma = 1.0$ to the GoodInd signal. This additional wave system introduces extra differences in the order statistics of LF and HF peaks. In summary: C'_{good} is an LF indicator consisting of linear wave crests and C'_{worse} is an LF indicator consisting of wave crests in the linear waves plus the additional noise waves.

To assess indicator quality, LF and HF crests were time-matched by identifying LF zero up-crossing crests and locating the corresponding HF maximum within each LF interval. This approach allows LF and HF crests to be matched even when they are slightly shifted in time. Fig. 16(a) (left) in Appendix E shows a scatter plot for GoodInd, where LF and HF crest statistics closely align, though not perfectly. In the more non-linear case WorseInd (right in the same plot), there is more scatter, reflecting differences in LF and HF order statistics. These originate not from measurement errors, but from model differences. The noise indicates a less effective indicator. We aimed to define representative indicators for real non-linear response problems in waves and the available indicators. To quantify this, we used the average 'screening quality index' (SQI_{mean}) from [35], where a value of one is ideal, and higher values indicate more LF samples are needed to predict true HF values. A short explanation and interpretation of this variable can be found in Appendix A. For GoodInd, $SQI_{mean} = 1.6$; for WorseInd $SQI_{mean} = 3.1$, showing the expected decrease in indicator quality with added noise. Still, the order statistics of these indicator/response pairs are relatively similar compared to actual cases; van Essen et al. [35] found SQI values around 10 with the best indicators for green water loads.

3.3. Adaptive sampling

We chose two PoE levels in the initial sampling for Step 5 of Adaptive Screening, 0.001 and 0.01 ($m = 2$), positioned around P_{exp} defined in Section 3.2. The prediction range $\mathbf{d}^*$ in Step 8 was selected between 5×10^{-4} and 2×10^{-2} in 200 uniform steps, containing 450 HF events. To generate a 'new' HF sample in each iteration, we matched LF and HF peaks as described in Section 3.2. When an event was selected based on its LF indicator value (in Step 5 or Step 11), the corresponding HF value was drawn from the matched LF-HF peaks. This is only possible because we have an analytically traceable HF variable; in reality, each iteration would require a new HF event calculation.

To monitor convergence, we set stopping criteria limits for $S(j)$ in Eq. (34): $\epsilon_1 = 0.05$ m for the maximum absolute wave crest distribution difference, and $\epsilon_2 = 0.003$ for the coefficient of variation of the one-hour MPM value (standard deviation 0.3% of mean). In

[2] We used mean wave period $T_1 = T_p/1.198$ for JONSWAP to obtain the reference Forristall distribution.

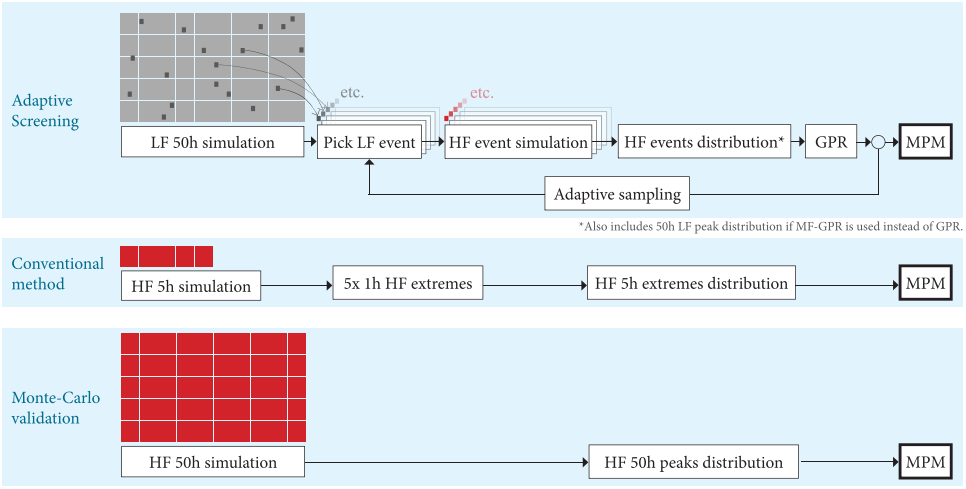


Fig. 4. Flow chart of application 1: Adaptive Screening, reference conventional method and HF Monte-Carlo simulation for validation. The area of the red blocks approximately indicates the simulation length.

real applications, less stringent criteria may suffice if reducing HF calculations is the main priority.

As explained in Section 2.2, each iteration picks a new event from the existing LF MCS pool based on the acquisition function applied over $\mathbf{d}^*$, without replacement. The theoretical new best point at PoE $p_{\text{new}}^{\text{usmv}}$ is therefore always within this range. However, the closest available PoE level in the (remaining) pool may be slightly outside this range. This is acceptable, as it still improves prediction accuracy within the range.

For MF-GPR, we also need the LF distribution as input. As explained in Appendix B.2, we used a subset of the full LF MCS samples in MF-GPR. This was done by applying a 'translated' three-parameter Weibull fit (see e.g., [17,69,86–89]) to the top 50% of LF samples, using Eq. (10). We then derived a subset of 50 equally spaced LF samples for MF-GPR from this fit.

$$P(X \geq x) = \exp\left(-\left(\frac{x-\theta}{\alpha}\right)^{\beta}\right) \tag{10}$$

3.4. Reference conventional industry method

We also applied a version of the 'conventional' industry method, described in Section 1.4 to application 1. Implementation details and results are discussed in Appendix D, for three versions of the conventional method (3, 5 or 7 seeds per sea state). In the following section, the 5-seed conventional results are compared to the full HF MCS and Adaptive Screening outcomes. Fig. 4 provides an overview of the three methods used to determine the one-hour MPM for application 1, with block sizes illustrating the relative duration of simulations, highlighting Adaptive Screening's efficiency in minimising HF simulation time (red blocks).

3.5. Results

We first checked the suitability of the selected stopping criteria (Sections 2.3 and 3.3) for application 1. Appendix C.2 shows that the criteria indeed seem appropriate. Fig. 5 shows the fully converged predicted MPM and its U95% for both indicators, two GPR versions and acquisition function USMV. It includes annotations with the accuracy of the mean predictions at convergence. For reference, the prediction accuracy after only 10 HF samples (8 iterations plus 2 initial samples) is also indicated.

The predicted final distributions are plotted in Fig. 6. They look very similar to the HF ground truth, for most of the prediction range and especially around P_{exp}. This is confirmed by Fig. 5; the HF one-hour MPM values are predicted with a deviation between −0.8% and +0.2% from the true MCS result at convergence. This accuracy is in the same range as the reference conventional method, which has a −0.7% deviation from the ground truth (see Appendix D). The uncertainties of these predictions are the same range as well (U95% of 0.6 m from the conventional method, compared to the shaded area in Fig. 5), although slightly higher with Adaptive Screening and the WorseInd indicator.

It was expected that Adaptive Screening with MF-GPR may converge quicker than with GPR, as this also leverages the LF distribution shape. The results show that the MF-GPR version indeed converges quicker for both cases (with fewer HF samples), and Fig. 6 shows that the predicted distributions from MF-GPR are also closer to the true distributions than from GPR. There seems to be a benefit in using MF-GPR over GPR especially for lower numbers of HF samples (as also observed in e.g., [54]).

Comparison of the GoodInd and WorseInd indicators shows that neither the accuracy of the distribution shape nor that of the MPM value is significantly affected by the quality of the indicator. However, the number of required samples to reach this result does increase with decreasing indicator quality. Adaptive Screening was run until convergence, meaning that differences will be visible in the number of required HF samples and in the uncertainty instead of the mean accuracy. The higher SQI_{mean}, the more HF samples are required for convergence and the larger the uncertainty of the predicted MPM. This is logical, as the reduced indicator quality introduces noise in the HF results, which is directly translated to uncertainty in the GPR output.

3.6. Computational time

Even more than accuracy, efficiency is the foreseen advantage of Adaptive Screening over the conventional method. To quantify this, we need to translate the required number of HF events to simulation time. As noted in Section 2.4, prior studies have demonstrated the ability of HF CFD to replicate wave impact loads accurately when wave and ship motion data from experiments are reproduced accurately. Simulation times in these studies vary, with durations of $4T_p$ [78], 50 s [90], 35 s [79], ~20 s [16], or even as short as 3 s [80]. When the initialisation for CFD events is obtained from a LF tool, it was shown that this can be done with 52 s [51] or 12 s [35]. A short duration can suffice if fully non-linear wave event kinematics close to the structure can be obtained from the LF tool, whereas a longer duration may be required if the waves need to be initialised from linear wave elevation and/or further from the structure.

Assuming 10–50 s of HF simulation per event, we can estimate the total simulation time for convergence. The required 10–62 HF events

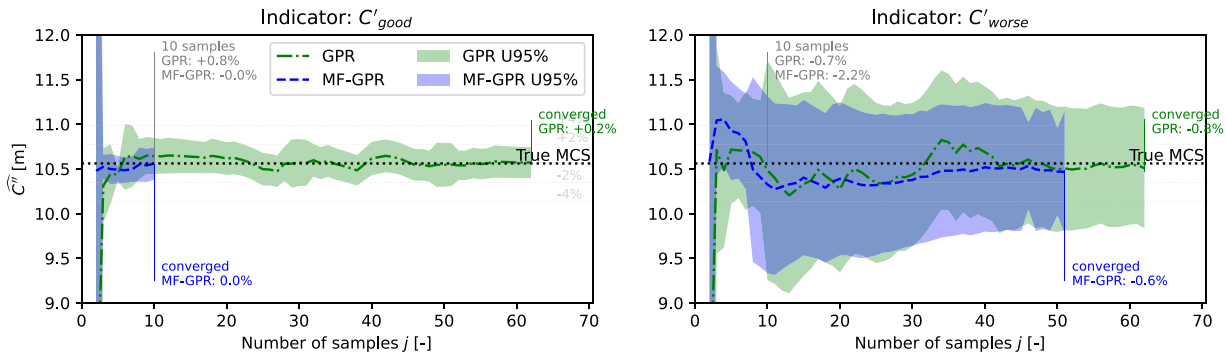
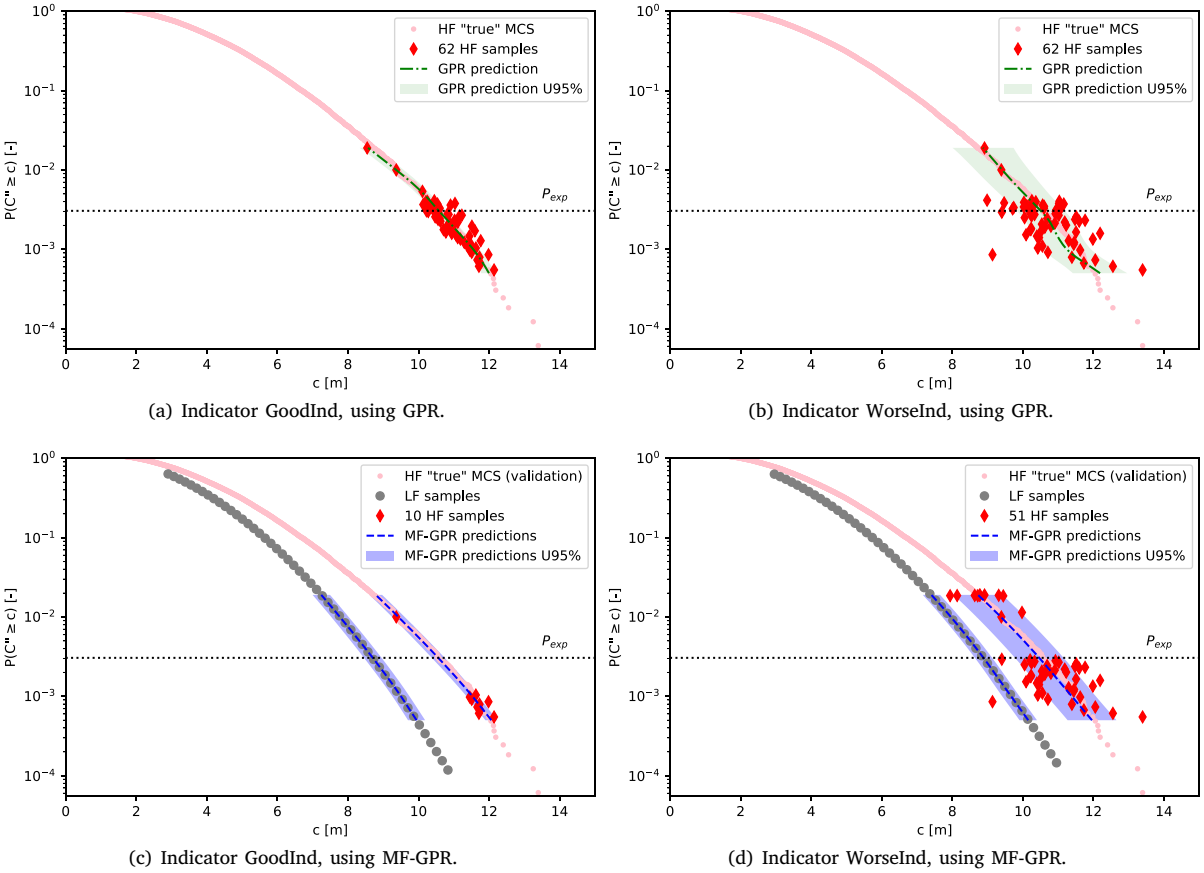


Fig. 5. Convergence of one-hour HF MPM and its uncertainty from Adaptive Screening with GPR or MF-GPR, for both indicators in application 1. Annotations indicate the mean prediction accuracy compared to the true MCS result.



(a) Indicator GoodInd, using GPR.

(b) Indicator WorseInd, using GPR.

(c) Indicator GoodInd, using MF-GPR.

(d) Indicator WorseInd, using MF-GPR.

Fig. 6. Converged distributions for application 1 for both indicators, from Adaptive Screening with GPR or MF-GPR, acquisition function USMV and the defined stopping criterion.

with Adaptive Screening thus translates to 0.03–0.9 h of HF simulation per sea state. In comparison, the conventional method requires 5 h per sea state (~10 times more), and a full HF Monte Carlo simulation (MCS) takes around 50 h (~100 times more). Thus, Adaptive Screening achieves accurate results far more efficiently, even with strict convergence criteria - and could converge even faster with relaxed criteria. Fig. 5 indicates that with only 10 HF samples, the deviation can be <2.2%, significantly reducing HF simulation time. Adaptive Screening also requires a 50 h LF MCS, which can be quick for an analytical indicator (as used here) or with linear potential flow. Additionally, each iteration requires a (MF-)GPR prediction. These are simulation times, not CPU times. CPU time depends heavily on the tools used for LF and HF simulations, where the LF simulations should be cheaper. For this application, the total computational cost (LF simulations, HF event calculations and iterative Adaptive Screening procedure) was ~ 1 CPU hour (CPUh), primarily driven by the iterative procedure. As the problem is analytically traceable, this is not necessarily shorter than with the reference method or brute force MCS. However, for more non-linear problems, the HF simulations are more computationally intensive, to the advantage of Adaptive Screening (see also Table 1).

4. Application 2: vertical bending moment on a ferry

Application 1 examined waves only, whereas Adaptive Screening was designed for hydrodynamic responses to waves. Thus, application 2 explores the moderately non-linear problem of predicting vertical

bending moments (VBM) on the 190 m MARIN ferry 2 [91]. We predicted the HF 30 min MPM value (so $T_{\text{exp}} = 1800$ s) of hogging VBM midships, using a sign convention where hogging peaks are identical to VBM troughs. The loading condition matched that in [72]. We selected an extreme irregular head wave condition, with a JONSWAP spectrum, $H_s = 13.2$ m, $T_p = 10.0$ s, $\gamma = 3.0$, at a forward speed of 10 knots (5.14 m/s). This extreme sea state (in reality with heavy wave breaking) was selected to make the test case similar to that described in [50,54]. Although our ship shape, and LF / HF tools differ, the wave and response conditions align, allowing us to obtain an idea of Adaptive Screening's performance compared to the alternative method in [50]. The HF response is non-linear VBM (V''), and the LF indicator is linear VBM (V'). Unlike in application 1, this HF response is not analytically traceable, but long-duration non-linear simulations were feasible. We selected a total MCS duration of 30 h. By performing LF and HF simulations for the same wave realisations and selecting HF samples from the existing HF simulations, we could again validate the probabilistic framework. Here we use Adaptive Screening only in combination with MF-GPR, as this worked best in application 1. This chapter follows the structure of application 1, but without the conventional reference analysis.

The probability of interest $P_{\text{exp}} = 4.18 \times 10^{-3}$ for this application follows from Eqs. (2), (3) and (5), where $T_{p,e} = 7.5$ s for $T_p = 10.0$ s at 5.14 m/s speed in head waves, and $n_w \approxeq 1.44 \times 10^4$ is the approximate number of wave crests in the 30-h MCS.

4.1. High-fidelity data

The HF VBM time traces were supplied by non-linear time-domain code PRETTI_R v19.0.1 of CRS [92,93] which includes Froude–Krylov non-linearity. The ground truth HF hogging peaks were defined by the zero up-crossing troughs in these VBM traces. The utilised panel distribution is visualised in Fig. 7. The true HF 30-minute MPM for the chosen 30-hour realisation is 1.05×10^9 Nm.

4.2. Low-fidelity indicator data

The LF VBM time traces were supplied by linear frequency-domain potential flow diffraction code SEACAL v7.2.0 of CRS, in the zero speed Green's function implementation. The true LF and HF zero up-crossing peak PoE distributions for the 30-h MCS are shown in Fig. 15(b) in Appendix E. We matched the LF and HF VBM hogging peaks at midships in the same way as described for the wave crests in Section 3.2. This results in the scatter plot in Fig. 16(b) in Appendix E. This shows that the hogging LF indicator and HF response peaks have less similar order statistics than those of the wave crests in application 1. This is also reflected in the SQI_{mean} value, which is 22.8.

4.3. Adaptive sampling

As in application 1, we chose two PoE levels (0.001 and 0.01) for initial sampling in Step 5, along with a prediction range of 1×10^{-3} to 5×10^{-2} with 200 uniform steps in Step 8, both centred around P_{exp}. To generate a 'new' HF sample in each iteration, we matched LF and HF peaks in time following the method described in Section 4.2. For each selected event, the corresponding HF value (determined by the LF indicator value in Step 5 or Step 11) was drawn from the HF simulations at the closest matching time. Unlike in application 1, where an analytically traceable HF variable was available, here we used the database of non-linear HF simulation results. To monitor convergence, we set the second limit of the stopping criteria $S(j)$ in Eq. (34) equal to the limit we set in application 1: $\epsilon_2 = 0.003$ for the coefficient of variation of the MPM value (standard deviation 0.3% of mean). The first limit, for the maximum absolute distribution difference, needs to be adapted as it has the same unit as the predicted variable. For application 1, we selected 0.05 m crest height, with maximum crest heights in the order of 15 m. To use a similar criterion, we use $\epsilon_1 = 5 \times 10^6$ N m, with maximum hogging peak values in the order of 1.5×10^9 N m. Again, less stringent criteria may suffice if reducing HF calculations is the main priority. The rest of the adaptive procedure for application 2 was identical to that of application 1.

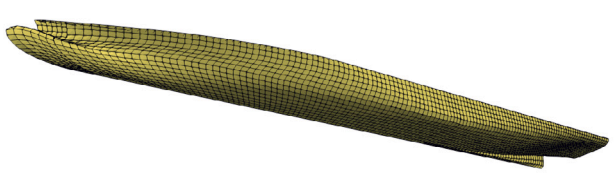


Fig. 7. Mesh of the MARIN ferry 2 for the SEACAL and PRETTI_R calculations in application 2, consisting of 3327 panels.

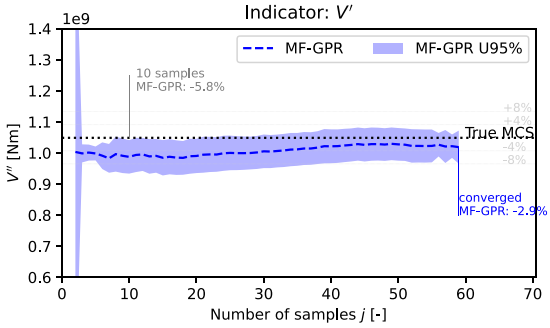


Fig. 8. Convergence of 30-min HF MPM and its uncertainty from Adaptive Screening with MF-GPR, for hogging VBM in application 2. Annotations indicate the mean prediction accuracy compared to the true MCS result.

4.4. Results

Fig. 8 shows the predicted 30 min HF MPM and its U95% for hogging, from Adaptive Screening with MF-GPR and acquisition function USMV. The converged distribution shapes are plotted in Fig. 9. These plots are similar to Figs. 6(c) and 6(d) for application 1, but here we plotted the LF and HF distributions separately to distinguish them better. These results show that, initially, the results start converging very quickly. However, it takes a while to reach proper convergence according to the defined stopping criterion. With 59 HF samples, the 30 min VBM hogging peak MPM is predicted with −2.9% accuracy. This accuracy is slightly less than that of the one-hour wave crest MPM results in application 1. The is explainable by the indicator quality (measured as SQI_{mean}), which is lower for application 2 than application 1. The U95% uncertainty of the 30 min MPM in the present application is also relatively large compared to the mean value, but convergence was reached with a similar number of HF samples as for the WorseInd indicator in application 1. For reference, the LF equivalent MPM deviates −8.2% from the true HF value, so Adaptive Screening does significantly improve the prediction compared to pure LF simulations. These results demonstrate a comparable level of accuracy to those obtained using an alternative multi-fidelity method for a similar VBM case in [50].

4.5. Computational time

Similar to application 1, we assume that each HF event simulation must be run for 10–50 s. The 59 HF events required for convergence in application 2 then correspond to ~10-50 min HF simulation time. We picked the HF results from the 30-h validation HF MCS with PRETTI_R. This total 30-h simulation took 26 CPUh. We therefore estimate that the required 10-50 min HF simulation time takes ~0.1–0.7 CPUh. We also

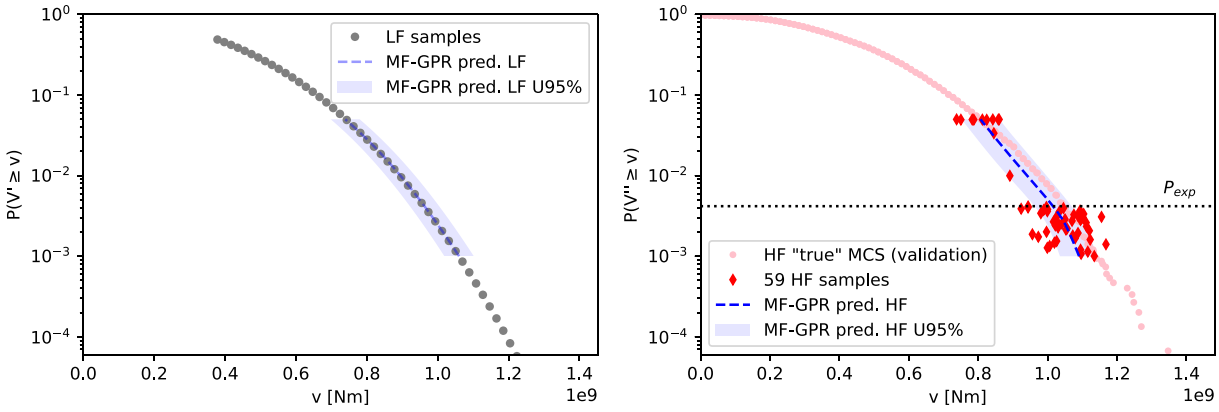


Fig. 9. Converged distribution for application 2, from Adaptive Screening with MF-GPR, acquisition function USMV and the defined stopping criterion.

needed 30 h of LF simulation time with frequency-domain code SEACAL (one speed, one heading) combined with linear analytical waves, which took ~0.25 CPUh. Additionally, Adaptive Screening roughly took 1 CPUh for 200 iterations. The total computational cost for the Adaptive Screening procedure (excluding HF validation data) was therefore 0.1–0.7 CPUh (HF simulations) + 0.25 CPUh (LF simulations) + 1 CPUh (iterations) ≈ 1.5-2 CPUh. The alternative would be 26 CPUh for the 30-h HF MCS, showing that the new method is considerably more efficient (see also Table 1). All calculations for application 2 were done on a single laptop core.

5. Application 3: green water impact forces on a containership

Applications 1 and 2 are weakly and moderately non-linear test cases, while Adaptive Screening is designed for strongly non-linear responses. These earlier applications allowed us to compare the method's converged outcome to reference HF MCS results (with a good outcome) and to assess the impact of various settings. The method still performed well with lower-quality indicators, which showed the method's potential for more non-linear problems. However, to properly show the method's suitability, a strongly non-linear test case is needed. This is presented here.

In application 3, we predicted extreme values of global green water impact force peaks on the breakwater of the 230 m KCS containership [94]. This HF force is noted as F_x'' in the remainder of this application. It was validated against experiments (see Fig. 10). The considered test campaign is extensively described in [35], including a discussion of the capability of many different numerical indicators to predict the occurrence of these green water forces. We selected an experiment with a full-scale duration of 30 min in irregular head waves with a JONSWAP spectrum and $H_s = 6.8$ m, $T_p = 9.7$ s, $\gamma = 3.3$ at 4.6 knots forward speed (2.37 m/s). The conditions were selected such that the impact frequency was high; there were 95 experimental impacts in this 30 min duration. Bandringa et al. [16] has shown that calculations using Cartesian-grid finite volume CFD method ComFLOW [95] on a fine mesh are able to reproduce the green water forces in 12-second events accurately for the same experimental data.

We did not run Adaptive Screening until proper convergence according to the stopping criterion of Section 2.3, because that would require a longer duration of the validation material. Instead, we demonstrate that we can get a good estimate of the 5 min MPM using only 10 HF events ($T_{\text{exp}} = 300$ s). This is a demonstration of the potential of the method; not a full validation case. Again, we only apply Adaptive Screening in combination with MF-GPR.

Here we do not have a HF response that is analytically traceable (as in application 1) nor do we calculable with a weakly non-linear tool (as in application 2), but we do have HF experimental data. By deterministically reproducing these experiments in the numerical indicator calculation, we could also avoid the CFD simulations in Step 6 of Adaptive Screening for this application. Similar as in application 2 from the HF simulations, new HF samples were picked here from the HF experiments. This chapter follows the same structure as that of applications 1 and 2.

The probability of interest $P_{\text{exp}} = 2.80 \times 10^{-2}$ for this application follows from Eqs. (2), (3) and (5), where $T_{p,e} = 8.4$ s for $T_p = 9.7$ s at 2.37 m/s speed in head waves, $N = T_{\text{tot}}/T_{\text{exp}} = 1800/300 = 6$ and $n_w \cong 215$ is the approximate number of wave crests in the available 30 min experiment. The number of wave encounters remains the leading reference for all PoE values. This enables the use of different variables for the LF and HF signals in the Adaptive Screening procedure, as demonstrated in this application.

5.1. High-fidelity data

The 'ground truth' HF distribution and 5 min MPM in application 3 were taken from the experimentally measured green water impact peak forces on the breakwater of the ship (the vertical structure perpendicular to the ship's length in Fig. 10). These forces were measured using 40 force panels on the breakwater surface. We focused on the peak total forces across this structure, obtained by summing the forces over time from all panels. A zero up-crossing distribution of these force peaks was made. As the number of HF green water impacts was limited to 96 (and very sparse in the tail), we fitted a Generalised Pareto distribution to the highest 10% available samples before deriving the ground truth MPM ($\widehat{F_x''}$). This was done by applying Eq. (7) to this fitted distribution. The thus found true HF 5-minute MPM from the experiments is 2.95×10^3 kN.

5.2. Low-fidelity indicator data

We selected two well-performing indicators from [35]: green water force on the breakwater and relative wave elevation (RWE) at a probe on the foredeck. Both indicators were calculated on a coarse ComFLOW mesh, known as 'CF3' in [35], with wave phases matched to experiments for deterministic reproduction. Details on the mesh, settings, and numerical setup can be found in the cited work. These indicators are non-linear, but lower in order than the fine-mesh CFD calculations or experiments that are needed for accurate green water force assessment. Here, the coarse-mesh LF green water force and RWE indicators are denoted F_x' and R'_{deck}, respectively. These indicators have SQI_{mean} values of 2.08 (F_x') and 2.06 (R'_{deck}). The true LF and HF peak PoE distributions of F_x' are shown in Fig. 15(c), and matched LF-HF peak scatter plots for both indicators are included in Fig. 16(c) (both in Appendix E). Although the relationship between LF and HF peaks is less linear than in application 1, this does not necessarily mean that the indicator is worse, as supported by similar SQI values. The scatter

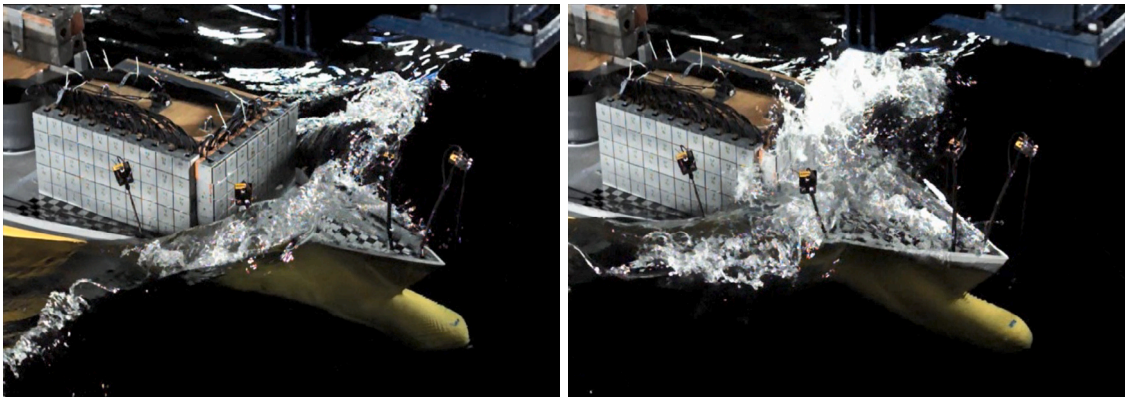


Fig. 10. Green water impact on the KCS during the CRS Green Water experiments; snapshots just before and during the same event.

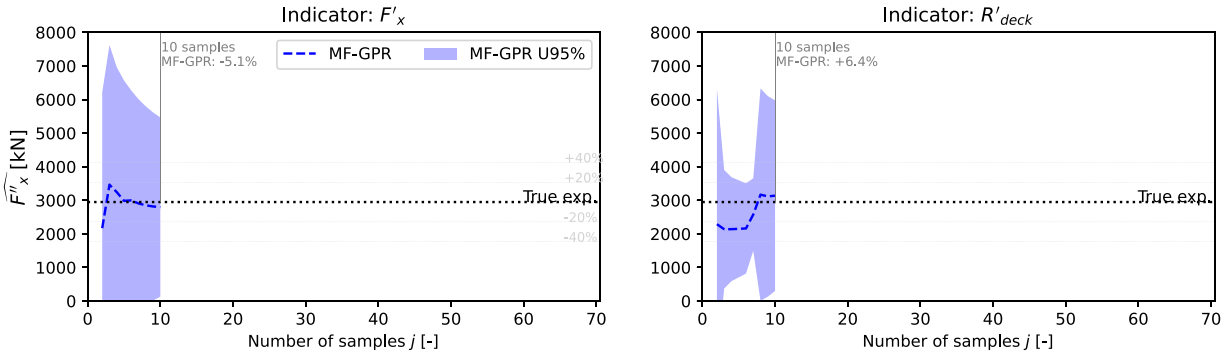


Fig. 11. Convergence of 5 min HF MPM and its uncertainty from Adaptive Screening with MF-GPR, for both indicators in application 3. Annotations indicate the mean prediction accuracy compared to the true MCS result.

plots illustrate why finer CFD meshes are essential for accurate HF impact forces: coarse mesh CFD predictions underestimate forces due to wave dissipation, sampling errors, and underestimated crest heights. However, here we use coarse mesh CFD only to quickly indicate critical event occurrence, not to predict the magnitude of the resulting forces. The latter is done using fine mesh calculations for a number of events in Step 6. This application with indicator R'_{deck} will demonstrate that it is indeed possible to use an LF indicator that is a different variable than the HF response.

5.3. Adaptive sampling

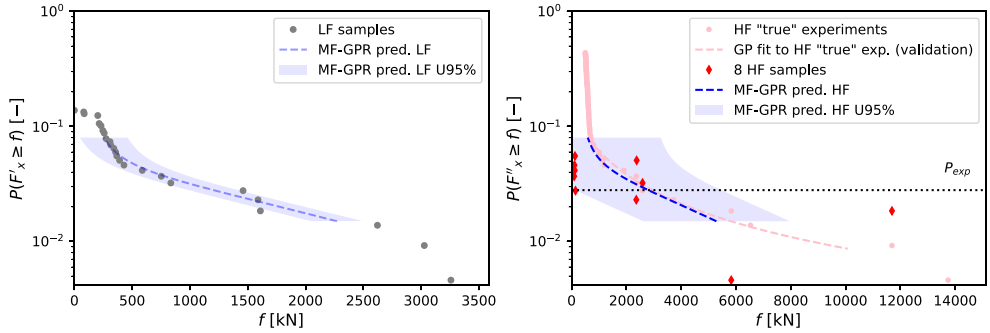
Similar as in application 1 and 2, we chose two PoE levels for the initial sampling points (0.005 and 0.05) for Step 5 of Adaptive Screening and a prediction range (1.5×10^{-2} to 8×10^{-2}, with 200 uniform steps) for Step 8. Both were selected around P_{exp}. We used the matching in time of LF and HF peaks described in Section 3.2 in order to obtain a 'new' HF sample in every iteration. Using this procedure, the HF value corresponding to a selected event (based on its LF indicator value in Step 5 or Step 11) was taken from the experiments around the same time. In application 1 this was possible because we had an analytically traceable HF variable; here we use the experiments as HF database. We also need LF distribution samples as direct input for MF-GPR. In application 1 and 2 we did not use all available LF wave crests to prevent unnecessarily large MF-GPR matrices. As the number of LF indicator peaks for both indicators in the 30 min duration are limited, this is not necessary in application 3. We used all available LF distribution samples in MF-GPR.

5.4. Results

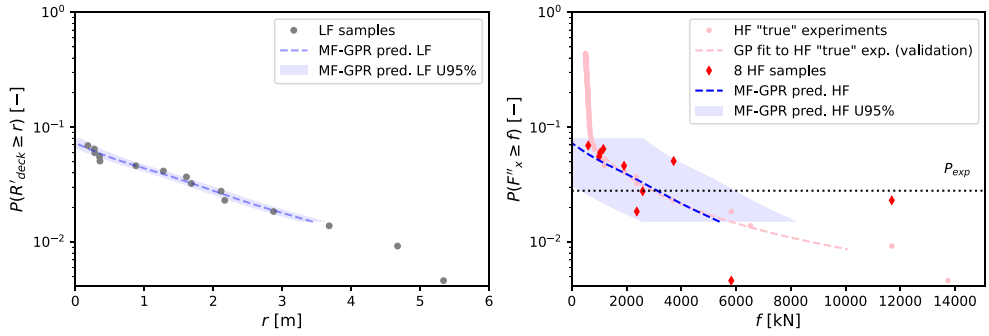
Fig. 11 shows the predicted 5 min HF MPM and its U95% for both indicators, from Adaptive Screening with MF-GPR and acquisition function USMV. The x-axis scale is the same as in the plots of the other two applications, for reference. As said, we only ran the procedure up to 10 samples (2 initial samples and 8 iterations), at which point the results are not converged yet. The predicted distribution shapes with 10 HF samples are plotted in Fig. 12.

Comparing these results to those in application 1 and 2 firstly shows that the uncertainty band and the deviations from the ground truth with 10 HF samples are an order larger for this strongly non-linear application than for weakly and moderately non-linear applications 1 and 2. Knowing that it is very hard to predict accurate green water impact loads, this is not surprising. However, using only 10 HF events, the shape of the green water impact force distribution is quite well approximated (Fig. 12). Using the coarse mesh force indicator, the 5 min MPM green water impact forces can be predicted with an error of only −5.1% (see Figs. 11 and 12(a)). Using indicator F'_x leads to slightly more accurate (and seemingly more converged) results than using R'_{deck}, but this difference may not be significant with the low number of HF samples. The results with R'_{deck} show that it is indeed possible to use different variables as LF and HF signals in Adaptive Screening, with satisfactory results. This can be beneficial in cases where it is less computationally expensive to calculate RWE than forces on a coarse mesh.

These results indicate that we can obtain quite accurate 5 min extreme green water forces using only 10 HF events. In the design stage of a new ship, an accuracy of 5%–10% is definitely acceptable for wave impact load extreme values (considering that the utilised safety factors for impacts are usually much higher). Using Adaptive Screening



(a) Indicator coarse mesh ComFLOW green water force peaks (F'_x).



(b) Indicator coarse mesh ComFLOW RWE peaks at fore deck (R'_{deck}).

Fig. 12. Distributions for application 3, from Adaptive Screening with MF-GPR, acquisition function USMV and 10 HF samples for two indicators. The left side show the LF indicator samples and the LF prediction from MF-GPR, the right side shows the HF samples, the HF prediction from MF-GPR and the reference 'true' experimental HF distribution (fitted with a Generalised Pareto distribution).

could therefore be considered an improvement compared to the need to do experiments or very lengthy fine mesh CFD calculations. Obviously we have not run Adaptive Screening until convergence (which would require more HF event calculations) and only considered 30 min screening time and a 5 min MPM. Real design loads would be required for a longer exposure duration (e.g., 30 min or one hour). However, this application shows that an acceptable accuracy could already be achieved with very few HF events, that the results can converge close to the true distribution with 'real' available indicators for a strongly non-linear response case, and that the computational time in order to do this reduces considerably compared to MCS.

5.5. Computational time

Similar to applications 1 and 2, we assume that each HF fine mesh CFD event calculation has to be run for 10–50 s. The 10 HF events analysed in application 3 therefore correspond to 100–500 s HF simulation time. We also need to run the coarse mesh CFD screening, which corresponds to 30 min LF simulation time (where 'LF' is in this case coarse mesh CFD). This took 56 CPUh on 8 cores (so 7 wall clock hours). We did not perform the HF fine mesh simulations, as we sampled the HF values from the experiments. However, we can estimate the computational cost for a real problem. One fine mesh CFD green water event calculation with ComFLOW in [16] for the same experiments took ~2.000 CPUh. We therefore estimate that the full procedure takes 56 CPUh (30 min LF screening) + 20.000 CPUh (10 HF events with a fine mesh, which can be run in parallel) + ~1 CPUh (iterative Adaptive Screening). This is long, but still significantly shorter than running fine mesh simulations for the full 30 min duration. For reference, linearly extrapolating the 2.000 CPUh for the 20 s events in [16] to 30 min would be ~180.000 CPUh (see also Table 1).

6. Discussion & conclusions

We introduced a new multi-fidelity extreme value prediction method, Adaptive Screening, for strongly non-linear (ship) responses to waves. We applied it to three case studies with increasing levels of non-linearity and complexity: second-order wave crests, VBM on a ferry and green water impact loads on a containership. The results of the applications are summarised in Table 1 at the end of this paper. Adaptive Screening accurately predicts MPM extreme values for both weakly and strongly non-linear responses, with significantly reduced computational cost compared to traditional methods. This makes the approach promising for strongly non-linear responses, where alternatives are often limited. However, ship design requires evaluating many wave conditions over the ship's lifetime, which is impractical with coarse mesh CFD as needed for the most realistic and non-linear application 3. We therefore envisage a recursive approach, where low-/ medium-fidelity Adaptive Screening can identify critical long-term sea states (similar to Gramstad et al. [23]), and medium-/ high-fidelity Adaptive Screening can then estimate short-term load distributions in these states (as presented here).

It can be concluded that Adaptive Screening provides quite accurate extreme values of non-linear responses, if an adequate indicator is selected. This indicator selection is crucial, as in other methods like RCM and alternative screening approaches. Ideally, the indicator's order statistics should closely resemble those of the HF response. Since perfect similarity cannot usually be achieved, some 'random noise' is introduced in selecting events around probability level P_{exp} (Step 5), causing the HF prediction to be a noisy surrogate for the true HF distribution (Step 7). This noise results in a consistent underestimation of extreme values by screening approaches, as also observed in research

projects JIP Green Water [96] and CRS SCREAM [97]. For instance, classical screening typically selects the top 10–20 indicator events from a given simulation duration. However, even with perfectly accurate CFD calculations, extreme HF values may be underestimated since events are drawn from the distribution tail. With an imperfect indicator, some selected events will inevitably have lower HF responses than the true maxima (as the latter are the largest in the available pool), leading to an underestimation of the HF extreme values. A weaker indicator further increases this effect. In this study, we mitigate this bias by selecting events at P_{exp} from a long LF MCS in Step 2 (rather than the largest events from a shorter MCS), ensuring sufficient HF events above the target extreme value. A realistic indicator may still select some 'wrong' events, but these can include both larger and smaller HF responses, reducing underestimation. While this approach does not fully eliminate bias (as tail events remain sparse), our results demonstrate that it can effectively reduce the underestimation to acceptable levels across various cases and levels of non-linearity with realistic indicators. Further work could consider modelling this bias to define a safety factor.

With the realistic indicators in the present applications, the one-hour MPM second-order wave crest height could be calculated with an accuracy of 0.2–0.8%; the 30 min MPM VBM with an accuracy of 2.9% and the 5 min MPM green water impact load with an accuracy of 5.1%. The results in application 1 had a similar accuracy and uncertainty as a reference conventional method, with a significant reduction in HF simulation time. These levels of accuracy are probably acceptable in the design process of a maritime structure, considering that safety factors for strongly non-linear responses are generally much higher. Using Adaptive Screening considerably reduces the required computational cost for all three applications, compared to the conventional method or MCS.

Adaptive Screening places minimal constraints on the predicted distribution, apart from the (MF-)GPR properties and a monotonicity requirement. Thus, it can produce distributions outside the Generalised Extreme Value (GEV) family. For the present applications, the results closely approximates ground truth, suggesting that this is not problematic, though it may be an issue for more badly behaving problems. An alternative approach could involve learning the parameters of a GEV fit to the HF distribution rather than its overall shape. The cost of the new method lies mostly in the HF simulations. The largest efficiency gain is therefore to be found in reducing the number of required HF samples before convergence, rather than in speeding up the iterative procedure. Further acceleration of the procedure may therefore be possible by adopting a more lenient convergence criterion, optimising the utilised acquisition function or selecting multiple new events per iteration (allowing HF calculations to run in parallel).

The presented application cases validate only the statistical model in Adaptive Screening, assuming that the HF calculations or experiments in Step 6 can be initialised from the LF simulations and are perfectly accurate. A follow-up study should focus on validation of the full procedure (including HF calculations) for a strongly non-linear application, and possibly on making Adaptive Screening even more efficient and robust. Since only limited-duration ground-truth data were available for application 3, this strongly non-linear case served more as a demonstration than full validation. Future work could focus on an additional validation study using a strongly non-linear test case to further assess the method. However, the present results are very promising for future application of the method. It has potential to predict extreme values of wave impact loads, and likely also of other non-linear problems in maritime/civil engineering or oceanography (e.g., wave height, wind speed, ship parametric roll, propeller ventilation, dike overtopping, local water levels, etc.), and in short- and long-term scenarios.

CRediT authorship contribution statement

Sanne M. van Essen: Writing – original draft, Validation, Methodology, Formal analysis, Conceptualization, Visualization, Software, Investigation, Data curation. **Harleigh C. Seyffert:** Supervision, Project administration, Resources, Funding acquisition, Writing – review & editing.

Declaration of competing interest

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

Acknowledgements

This publication is part of the project "Multi-fidelity Probabilistic Design Framework for Complex Marine Structures" (project TWM.BL.019.007) of the research programme "Topsector Water & Maritime: the Blue route" which is (partly) financed by the Dutch Research Council (NWO), The Netherlands. The authors also thank the CRS for the use of the PySeaWave, SEACAL and PRETTI_R tools and for the green water ComFLOW and experimental data in application 3.

Appendix A. Screening quality index

The screening quality index *SQI* is defined in [35]. This metric only applies to validation datasets, as it uses an LF-HF peak matching procedure such as described in Section 3.2. We define false negatives (indicator misses a critical event), false positives (indicator predicts a critical event erroneously) and correct positives (indicator correctly predicts a critical event). *SQI* does not allow for any false negatives above a given HF threshold, as formulated in Eq. (11). Here $\mathbf{l}^{\text{mcs}}$ are all LF indicator peaks, $\mathbf{h}^{\text{mcs}}_{\text{matched}}$ are the matched HF peaks (see Fig. 16), t_h is the HF threshold, and # indicates a count. The numerator finds the minimum LF value with all matched HF values above a threshold and counts their number, representing the total number of positives in the screening. The denominator counts real events above the HF threshold, representing correct positives. An ideal *SQI* is one (all positives are correct), with higher values indicating poorer performance. For example, an SQI value of 4 implies 4 events must be evaluated to find one true positive. Say there are 15 real HF events above the threshold, this would mean 60 HF CFD event calculations. *SQI* is defined as a function of HF response threshold. For practical comparison, we use the mean SQI_{mean} over the full range of thresholds, similar to van Essen et al. [35].

$$SQI(t_h) = \frac{\#\left[\mathbf{l}^{\text{mcs}} \geq \min\left(\mathbf{l}^{\text{mcs}} \middle| \mathbf{h}^{\text{mcs}}_{\text{matched}} \geq t_h\right)\right]}{\#\left[\mathbf{h}^{\text{mcs}}_{\text{matched}} \geq t_h\right]} \tag{11}$$

Appendix B. (Multi-fidelity) Gaussian Process Regression

B.1. GPR

Single-fidelity GPR (e.g., [98]) can be used to predict new values $\mathbf{y}^*$ and their uncertainty over range $\mathbf{x}^*$, using a known sampled data set $[\mathbf{x}, \mathbf{y}]$. In Step 9 of Adaptive Screening, $[\mathbf{x}, \mathbf{y}]$ is the existing HF data set $\left[\ln\left(\mathbf{d}_H^{\text{sel}}\right), \mathbf{h}^{\text{sel}}\right]$ from Step 7; $\mathbf{x}^*$ is the prediction range $\ln(\mathbf{d}^*)$ from Step 8 and $\mathbf{y}^*$ is the predicted HF solution $\mathbf{h}^*$. The underlying process is assumed to have an infinite-dimensional multivariate Gaussian distribution with a mean function μ, and covariance function or kernel k (see Eq. (12)). As customary, we assume that the mean is zero, the observations are related only by the kernel, the data are noisy and discrete. This reduces it to Eq. (13), where $\mathbf{I}$ is the identity

matrix and σ_n^2 is the noise variance. We assume homoscedastic noise (see Section 2.4).

$$y(x) \sim \mathcal{N}\left(\mu(x), k(x,x')\right) \tag{12}$$

$$\mathbf{y}(\mathbf{x}) \sim \mathcal{N}\left(\mathbf{0}, \mathbf{K}(\mathbf{x},\mathbf{x}') + \sigma_n^2\mathbf{I}\right) \tag{13}$$

The kernel is a covariance function that describes the relation between the input points. We used the Matern32 kernel in Eq. (14), which includes length hyperparameter l and maximum allowable covariance σ. If two points are far apart, their covariance is zero; l determines how quickly this interaction decreases. The hyperparameters are assembled in $\theta = \{\sigma, l, \sigma_n\}$. The noise parameter is learned together with the others. GPR needs the value of the kernel for interaction of each point with all the other points, resulting in the interaction matrix in Eq. (15).

$$k(x,x'|\theta) = \sigma^2\left(1 + \frac{\sqrt{3}|x-x'|}{l}\right)\exp\left[-\frac{\sqrt{3}|x-x'|}{l}\right] \tag{14}$$

$$\mathbf{K}(\mathbf{x},\mathbf{x}'|\theta) = \begin{bmatrix} k(x_1,x_1|\theta) & k(x_1,x_2|\theta) & \dots & k(x_1,x_m|\theta) \\ k(x_2,x_1|\theta) & k(x_2,x_2|\theta) & \dots & k(x_2,x_m|\theta) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_m,x_1|\theta) & k(x_m,x_2|\theta) & \dots & k(x_m,x_m|\theta) \end{bmatrix} \tag{15}$$

We started by assuming arbitrary *a priori* hyperparameters, and then optimised them by maximising the conditional probability $P(\theta|\mathbf{x},\mathbf{y})$. According to Bayes' theorem, this is equivalent to maximising the log-likelihood given in Eq. (16), where m is the number of existing observations (see e.g., [82,99]). We solved this for θ using a multivariate optimisation algorithm. Assuming a uniform prior distribution $\mathcal{U}(\theta_{min}, \theta_{max})$ gives the best parameters θ.

$$\ln\left(P(\mathbf{y}|\mathbf{x},\theta)\right) = -\frac{1}{2}\mathbf{y}^T\mathbf{K}^{-1}\mathbf{y} - \frac{1}{2}\ln|\mathbf{K}| - \frac{m}{2}\ln(2\pi) \tag{16}$$

In order to predict $\mathbf{y}^*$ for new input range $\mathbf{x}^*$ using the thus trained GP, we modified Eq. (13) further to obtain Eq. (17). Interaction matrix $\mathbf{C}$ is provided in Eq. (18).

$$\begin{bmatrix}\mathbf{y}\\ \mathbf{y}^*\end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{C} + \sigma_n^2\mathbf{I}\right) \tag{17}$$

$$\mathbf{C} = \begin{bmatrix}\mathbf{K}(\mathbf{x},\mathbf{x}'|\theta) & \mathbf{K}(\mathbf{x}^*,\mathbf{x}'|\theta) \\ \mathbf{K}(\mathbf{x},\mathbf{x}^{*'}|\theta) & \mathbf{K}(\mathbf{x}^*,\mathbf{x}^{*'}|\theta)\end{bmatrix} = \begin{bmatrix}\mathbf{K}_e & \mathbf{K}^{*T} \\ \mathbf{K}^* & \mathbf{K}^{**}\end{bmatrix} \tag{18}$$

Now Eq. (19) provides the conditional probability of new values based on existing observations, which yields the best estimate for mean new values $\overline{\mathbf{y}^*}$ and their variance $\sigma^2_{y^*}$.

$$\begin{aligned}&\mathbf{y}^*|\mathbf{y} \sim \mathcal{N}\left(\mathbf{K}^*\mathbf{K}_e^{-1}\mathbf{y}, \mathbf{K}^{**} - \mathbf{K}^*\mathbf{K}_e^{-1}\mathbf{K}^{*T}\right) \\ &\overline{\mathbf{y}^*} = \mathbf{K}^*\mathbf{K}_e^{-1}\mathbf{y} \quad \text{and} \quad \sigma^2_{y^*} = \text{var}(\mathbf{y}^*) = \mathbf{K}^{**} - \mathbf{K}^*\mathbf{K}_e^{-1}\mathbf{K}^{*T}\end{aligned} \tag{19}$$

B.2. MF-GPR

In multi-fidelity (MF-)GPR, we use both LF and HF data ($[\mathbf{x}_l, \mathbf{y}_l]$ and $[\mathbf{x}_h, \mathbf{y}_h]$, organised in Eq. (20)) to predict the HF values $\mathbf{y}_h^*$ over $\mathbf{x}^*$. In our applications, $[\mathbf{x}_h, \mathbf{y}_h]$ is given by $[\ln(\mathbf{d}_H^{\text{sel}}), \mathbf{h}^{\text{sel}}]$ from Step 7 (same as in the GPR procedure). We also have the LF MCS set $[\ln(\mathbf{d}_L^{\text{mcs}}), \mathbf{l}^{\text{mcs}}]$ from Step 2. However, GPR scales with the cube of the sample size, making it a poor choice for applications with many samples (see e.g., [54]), so we reduced the number of utilised LF samples in MF-GPR. We therefore reduce this to $[\mathbf{d}_L, \mathbf{l}]$, either by sub-sampling or by fitting a theoretical distribution and deriving new samples from this fit. This data set is used as $[\mathbf{x}_l, \mathbf{y}_l]$ in MF-GPR.

$$\mathbf{x}_v = \begin{bmatrix}\mathbf{x}_l & \mathbf{x}_h\end{bmatrix}^T \quad \text{and} \quad \mathbf{y}_v = \begin{bmatrix}\mathbf{y}_l & \mathbf{y}_h\end{bmatrix}^T \tag{20}$$

We used two fidelity levels (as in [48]) and the autoregressive (AR1) multi-fidelity model of Kennedy and O'Hagan [49] in Eq. (21). This model uses the Markov assumption in Eq. (22), which speeds up the computation by decomposing the multi-fidelity problem into independent single-fidelity problems: one for the LF data and one for the difference function $\boldsymbol{\delta}$ between the LF and HF data. It can be interpreted as follows: given the nearest LF point $y_l(x)$, we can learn no more about HF point $y_h(x)$ from any other point $y_l(x')$ for $x' \neq x$. The parameter ρ is used to establish the correlation between the two datasets.

$$\mathbf{y}_h = \rho\mathbf{y}_l + \boldsymbol{\delta} \tag{21}$$

$$k\left(y_h(x), y_l(x')|y_l(x)\right) = 0 \tag{22}$$

The LF hyperparameters $\theta_l = \{\sigma_1, l_1\}$ are optimised first, using the LF data $[\mathbf{x}_l, \mathbf{y}_l]$. Next, the difference function hyperparameters $\theta_\delta = \{\sigma_2, l_2, \delta\}$ are optimised, using the HF data $[\mathbf{x}_h, \mathbf{y}_h]$ and predicted LF values $[\mathbf{x}^*, \mathbf{y}_l^*]$. Here δ is the difference parameter, and l and σ are the kernel hyperparameters (defined separately for the LF and difference parts). Finally, the LF and HF predictions are made using the optimised hyperparameter sets θ_l and θ_δ. The AR1 covariance matrix of the existing data $\mathbf{B_e}$ is formulated in Eq. (23), using Eq. (15). Similarly, the AR1 covariance matrix of existing with new points $\mathbf{B}^*$ is given in Eq. (24) and the matrix of new points $\mathbf{B}^{**}$ in Eq. (25).

$$\mathbf{B_e} = \begin{bmatrix}\mathbf{B_{e,11}} & \mathbf{B_{e,21}}^T \\ \mathbf{B_{e,21}} & \mathbf{B_{e,22}}\end{bmatrix} \quad \text{where:} \quad \begin{cases}\mathbf{B_{e,11}} &= \mathbf{K}(\mathbf{x}_l,\mathbf{x}_l'|\theta_l) \\ \mathbf{B_{e,21}} &= \rho\mathbf{K}(\mathbf{x}_h,\mathbf{x}_l'|\theta_l) \\ \mathbf{B_{e,22}} &= \rho^2\mathbf{K}(\mathbf{x}_h,\mathbf{x}_h'|\theta_l) + \mathbf{K}(\mathbf{x}_h,\mathbf{x}_h'|\theta_\delta)\end{cases} \tag{23}$$

$$\mathbf{B}^* = \begin{bmatrix}\mathbf{B_{11}^*} & \mathbf{B_{12}^*}\end{bmatrix} \quad \text{where:} \quad \begin{cases}\mathbf{B_{11}^*} &= \rho\mathbf{K}(\mathbf{x}^*,\mathbf{x}_l'|\theta_l) \\ \mathbf{B_{12}^*} &= \rho^2\mathbf{K}(\mathbf{x}^*,\mathbf{x}_h'|\theta_l) + \mathbf{K}(\mathbf{x}^*,\mathbf{x}_h'|\theta_\delta)\end{cases} \tag{24}$$

$$\mathbf{B}^{**} = \rho^2\mathbf{K}(\mathbf{x}^*,\mathbf{x}^{*'}|\theta_l) + \mathbf{K}(\mathbf{x}^*,\mathbf{x}^{*'}|\theta_\delta) \tag{25}$$

The MF form of the full covariance matrix is given by Eq. (26) and its log-likelihood function by Eq. (27), where m_v is the number of elements in $\mathbf{x}_v$. We use the Matern32 kernel again in both parts of MF-GPR.

$$\mathbf{B} = \begin{bmatrix}\mathbf{B_e} & \mathbf{B^{*T}} \\ \mathbf{B^*} & \mathbf{B^{**}}\end{bmatrix} \tag{26}$$

$$\ln(P(\mathbf{y}_v|\mathbf{x}_v,\theta_l,\theta_\delta)) = -\frac{1}{2}\mathbf{y}_v^T\mathbf{B}^{-1}\mathbf{y}_v - \frac{1}{2}\ln|\mathbf{B}| - \frac{m_v}{2}\ln(2\pi) \tag{27}$$

Finally, the new mean predicted values $\overline{\mathbf{y}_v^*}$ and their variance $\boldsymbol{\sigma}^2_{\mathbf{y}_v^*}$ can be calculated using Eq. (28). These predicted values include both the LF and HF prediction over range $\mathbf{x}^*$.

$$\overline{\mathbf{y}_v^*} = \mathbf{B^*}\mathbf{B_e}^{-1}\mathbf{y}_v \quad \text{and} \quad \boldsymbol{\sigma}^2_{\mathbf{y}_v^*} = \text{var}(\mathbf{y}_v^*) = \mathbf{B^{**}} - \mathbf{B^*}\mathbf{B_e}^{-1}\mathbf{B^{*T}} \tag{28}$$

B.3. Constraints

It is not possible to define a perfect indicator for most non-linear response problems, so the HF samples resulting from Step 7 will be 'noisy'. To avoid over-fitting, the minimum GPR noise variance was constrained to 0.5 × the variance of the available samples, with respect to a least-squares linear fit to these samples (see Eq. (29), where $[\mathbf{x}^*(j), \overline{\mathbf{y}}^*(j)]$ is the set of mean predicted HF samples and $\sigma_n(j)$ is the noise variance for iteration j). In MF-GPR, this is done separately for the LF and difference parts. We also constrained the length hyperparameter $l(j)$ of all kernels to the lower limit in Eq. (30), where $\text{len}(\overline{\mathbf{y}}^*(j))$ is the number of HF samples in iteration j. This limit takes value ten for a few samples, and higher values for many samples. The relatively high number for the first iterations helps to ensure monotonically decreasing distributions and quick convergence, and the increasing limit with the

total number of training samples helps to avoid over-fitting of the noisy data for many samples.

$$\sigma_n(j) \geq 0.5 \cdot \text{var}\left(\overline{\mathbf{y}}^*(j) - \text{linear\_fit}\left[\mathbf{x}^*(j), \overline{\mathbf{y}}^*(j)\right]\right) \tag{29}$$

$$l(j) \geq \max\left(10, 0.1 \cdot \text{len}(\overline{\mathbf{y}}^*(j))\right) \tag{30}$$

Appendix C. Details of stopping criterion

C.1. Formulations

We set an acceptance criterion requiring the PoE distribution to decrease monotonically, with higher threshold values at low PoE than at high PoE. Enforcing monotonicity in GPR predictions is challenging, as it necessitates correlations across the entire domain, while kernels capture correlations only locally (see e.g., [100]). Although methods exist to address this (e.g., [101,102]), they may overly constrain regression. We therefore chose not to modify GPR for monotonicity, accepting that some early predictions with limited HF data may lack monotonicity. In our implementation (Step 8), we defined the prediction range $\mathbf{d}^*$ in ascending PoE order, so the first element of the HF prediction $\mathbf{h}^*$ is highest and the last the lowest. An iteration was rejected if it did not meet the acceptance criterion R in Eq. (31) (where $w = \text{len}(\overline{\mathbf{h}}^*(j))$): each element of $\overline{\mathbf{h}}^*(j)_i$ must be smaller than the previous one, and the difference between the first and last elements must exceed 1% of the first element's value.

$$R(j) = \begin{cases} \text{accept} & \text{if } \left[\forall i \in [2, \dots, w], \overline{\mathbf{h}}^*(j)_{i-1} > \overline{\mathbf{h}}^*(j)_i\right] \\ & \quad \cap \left[\overline{\mathbf{h}}^*(j)_1 - \overline{\mathbf{h}}^*(j)_w > 0.01\overline{\mathbf{h}}^*(j)_1\right] \\ \text{reject} & \text{otherwise} \end{cases} \tag{31}$$

Secondly, we calculated the maximum absolute difference $E(j)$ in HF value between each set of subsequently predicted distributions. As the distributions in the first few iterations can be quite erratic, we used the average of $E(j)$ over the last K_1 iterations: $\overline{E}_{K_1}(j)$. When $j < K_1$, all available iterations were used. Here, K_1 is a user-defined value.

$$\overline{E}_{K_1}(j) = \frac{1}{\psi}\sum_{i=\chi}^{j} E(i)$$

$$\text{where: } \begin{cases} E(j) = \max|\mathbf{h}^*(j) - \mathbf{h}^*(j-1)| & \\ \chi = 1 \text{ and } \psi = j & \text{for } j = 1, 2, \dots, K_1 - 1 \\ \chi = j - K_1 + 1 \text{ and } \psi = K_1 & \text{for } j \geq K_1 \end{cases} \tag{32}$$

The largest difference between the predicted distributions of the $E(j)$ criterion may be far from P_{exp}. For this reason, a third convergence criterion was based on the coefficient of variation (COV) of the MPM value over the last K_2 iterations: $C_{K_2}(j)$. This is expressed in Eq. (33), where $\widehat{H}(j)$ is the MPM value predicted in iteration j. Again, we took the COV over the available iterations when $j < K_2$, where K_2 is another user-defined value.

$$C_{K_2}(j) = \frac{\sigma_{K_2}(j)}{\mu_{K_2}(j)} \quad \text{where:}$$

$$\times \begin{cases} \mu_{K_2}(j) = \frac{1}{\psi}\sum_{i=\chi}^{j} \widehat{H}(i) & \\ \sigma_{K_2}(j) = \sqrt{\frac{1}{\psi}\sum_{i=\chi}^{j}\left(\widehat{H}(i) - \mu_{K_2}(i)\right)^2} & \\ \chi = 1 \text{ and } \psi = j & \text{for } j = 1, 2, \dots, K_2 - 1 \\ \chi = j - K_2 + 1 \text{ and } \psi = K_2 & \text{for } j \geq K_2 \end{cases} \tag{33}$$

The total stopping criterion $S(j)$ is provided in Eq. (34), where limits ϵ_1 and ϵ_2 are case-dependent. The defined values $K_1 = K_2 = 20$ balance limiting both the influence of outliers and the minimum number of iterations for which convergence can be detected.

$$S(j) = \begin{cases} \text{stop} & \text{if } (R(j) = \text{accept}) \cap \left(\overline{E}_{20}(j) < \epsilon_1\right) \cap \left(C_{20}(j) < \epsilon_2\right) \\ \text{continue} & \text{otherwise} \end{cases} \tag{34}$$

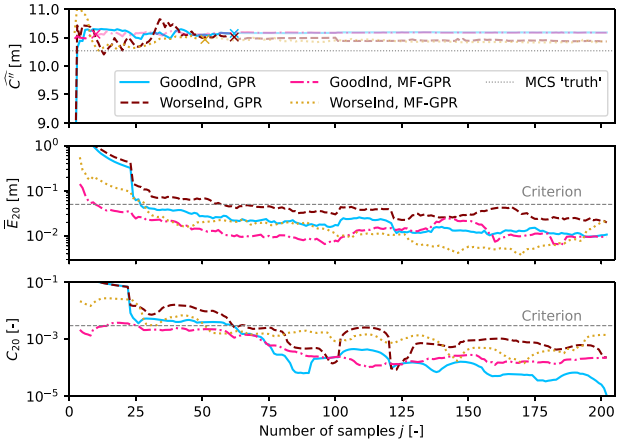


Fig. 13. MPM and stopping criteria, app. 1. $\overline{E}_{20}$ (Eq. (32); middle) and C_{20} (from Eq. (33); bottom) only shown for the accepted iterations by Eq. (31). MPM (top) is transparent from 'x', where convergence was reached based on the full stopping criterion (Eq. (34) and Section 3.3).

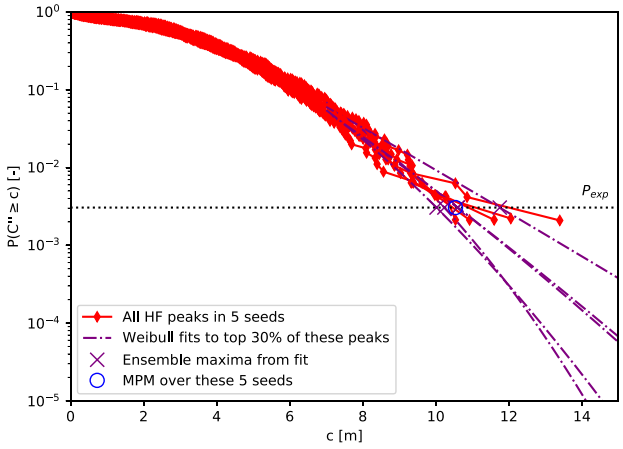
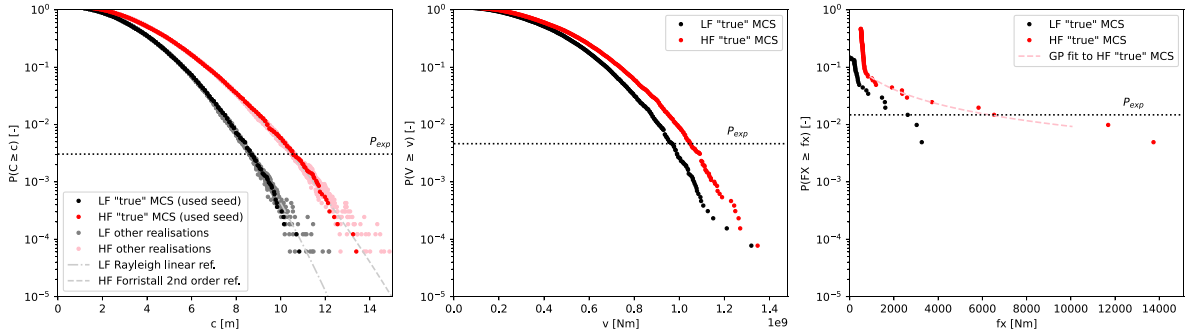


Fig. 14. Peak distributions for 5 randomly picked seeds, their Weibull fits and the derived MPM in one of the 500 seed picking realisations of the conventional procedure.

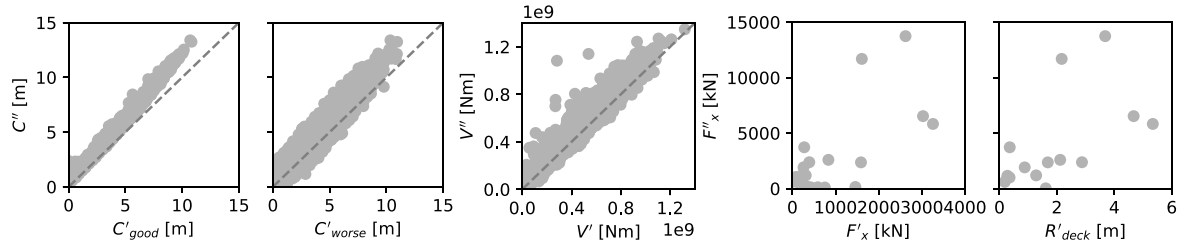
C.2. Evaluation of stopping criterion in application 1

The predicted one-hour MPM HF wave crest height $\widehat{C''}$ with Adaptive Screening and parts of the stopping criterion of Section 2.3 as a function of the number of HF samples are shown in Fig. 13 for application 1. The first part of the criterion (monotonocity check) is not explicitly shown, but the other parts are only plotted for the accepted iterations. The moment convergence is reached according to the limits set in Section 3.3 is indicated in the top MPM plot with a marker 'x'. These plots show that the presently selected limits seem appropriate, but may be quite strict; more lenient limits would lead to earlier convergence with only a small MPM accuracy loss.



(a) Application 1, indicator C'_{good} (b) Application 2, indicator V'. (c) Application 3, indicator F'_x.

Fig. 15. True peak distributions for the LF indicator and HF response (validation) in each application. Fig. 15(a) shows the utilised 50-h realisation, and nine other 50-h realisations to illustrate the variability of the distributions.



(a) Application 1. (b) Application 2. (c) Application 3.

Fig. 16. Scatter plots of matched LF and HF peaks for all three applications.

Appendix D. 'Conventional' industry method for application 1

As discussed in Section 1.4, class societies and ITTC recommend experiments with several 0.5 to 3-h wave seeds to determine short-term extreme response values. For strongly non-linear responses, guidelines suggest using 10–60 seeds per wave condition (e.g., [69,71,76,103–105]). Extreme values are then derived from the responses, before or after fitting. van Essen et al. [72] and Scharnke et al. [73] showed that a large number of seeds is needed to effectively reduce uncertainty. However, in practice, the limited available experimental time often restricts testing to 1 seed per wave condition for ships (as also advised by ITTC [106]) and up to 5–10 occasionally for offshore structures.

We applied this conventional method to application 1 for reference. No experiments were done, but it was assumed that they would supply HF data for several one-hour seeds. We used HF MCS results from Section 3.1 to simulate this, randomly picking 5 seeds ($N_{\text{conv}} = 5$) from the 50-h HF MCS. This was repeated 500 times in a bootstrap analysis to account for variability. To reduce variability in the one-hour maxima, a Weibull fit (Eq. (10)) was applied to the top 30% HF wave crests in each seed. We applied Eqs. (4) and (7) to the HF data of each fitted seed to assemble 5 seed maxima in C''_E, and then made an ensemble maxima distribution using Eq. (35). The MPM $\widehat{C''_E}$ was derived from Eq. (36), where 0.632 is the exceedance probability of the MPM in an ensemble maxima distribution for a Gaussian signal [75]. Fig. 14 shows the significant variation in the maxima derived from one of the 5-seed realisations. To evaluate the sensitivity of the results for the number of seeds, we have done the same for 3 and 7 seeds.

This variability yields a mean one-hour MPM of 10.62 m (3 seeds), 10.48 m (5 seeds) and 10.44 m (7 seeds), with a U95% uncertainty of 0.85 m (3 seeds), 0.61 m (5 seeds) and 0.49 m (7 seeds) over the 500 realisations. This mean value deviates less than 1% from the true 50-h MCS results in Section 3.1 for all three numbers of seeds. Obtaining these results requires 3 / 5 / 7 h of HF simulation or experiment time. We have used the 5-seed results for reference in the remainder of this publication.

$$\mathbf{d}_E(e) = P(C''_E \geq e) \tag{35}$$

$$\mathbf{d}_E\left(\widehat{C''_E}\right) = 0.632 \quad \text{, therefore:} \quad \widehat{C''_E} = \mathbf{d}_E^{-1}(0.632) \tag{36}$$

Appendix E. Summary of input and results of all three applications

This appendix assembles some input plots and results of all three cases, for a concise overview. Fig. 15 provides the PoE distributions of the LF and HF input; Fig. 16 provides the scatter plots of matched indicator and HF response peaks; and Table 1 provides an overview of the input and results in all three cases.

Data availability

Most data and scripts underlying this publication are available in (or can be re-generated using) the 4TU repository: [83]. This is valid for all scripts, and the data of application 1 and 2. The dataset of application 3 is proprietary, so this is not included in the repository.

Table 1
Summary of input, accuracy and computational time for all three applications. The very short computational times for the alternative methods in application 1 result from the analytical tractability of the problem.

	Application 1		Application 2	Application 3	
Ship	–		MARIN ferry 2	KCS containership	
Exposure duration MPM	1 h		30 min	5 min	
HF non-linear variable	Second-order waves		Hogging VBM midship	Green water load breakwater	
HF non-linear source	Analytical		Non-linear diffraction	Experiments	
LF indicator variable	Linear waves	Linear waves + noise	Hogging VBM midship	Green water load breakwater	Relative wave elevation deck
LF indicator source	Analytical	Analytical	Linear diffraction	Coarse mesh CFD	Coarse mesh CFD
Accuracy of predicted MPM compared to brute force MCS					
Adaptive Screening (USMV, GPR)	+0.2%	−0.8%	–	–	–
Adaptive Screening (USMV, MF-GPR)	+0.0%	−0.6%	−2.9%	−5.1%	+6.4%
Conventional reference method	−0.7%	−0.7%	–	–	–
Required LF simulation time					
Adaptive Screening	50 h	50 h	30 h	0.5 h	0.5 h
Required HF simulation time					
Adaptive Screening (estimate*)	0.03–0.9 h	0.03–0.9 h	0.2–0.8 h	0.03–0.1 h	0.03–0.1 h
Conventional reference method	5 h	5 h	–	–	–
Brute force HF MCS	50 h	50 h	30 h	0.5 h	0.5 h
Required computational time for the total procedure					
Adaptive Screening (estimate*)	1 CPUh		1.5-2 CPUh	20 000 CPUh	
Conventional reference method	<1 CPUh		–	–	
Brute force HF MCS	<1 CPUh		26 CPUh	180 000 CPUh	

* *Assuming HF events with a duration of 10-50 s.*

References

[1] Ersdal G, Kvitrud A. Green water on Norwegian production ships. In: 10th ISOPE Conf. Seattle, USA: Int Soc Offshore Polar Eng; 2000.

[2] Dallinga RP, Gaillarde G. Hatch cover loads experienced by M.V. Derbyshire during typhoon 'Orchid'. In: Int Works on Safety of Bulk Carriers. Glasgow, Scotland, UK; 2001.

[3] Halsne M, N. O, Ersdal G, Langøy M, Andersen T, Bjørheim LG. Semisubmersible in service experiences on the Norwegian Continental Shelf. In: 41th OMAE Conf. Hamburg, Germany: ASME; 2022, http://dx.doi.org/10.1115/OMAE2022-81289.

[4] Zhang N, Xiao L, Zou Q, Cummins C. Large-scale wave basin experimental study on the spatio-temporal distribution of wave impact loads on a semi-submersible platform. Ocean Eng 2025;327:120991. http://dx.doi.org/10.1016/j.oceaneng.2025.120991.

[5] Viste-Ollestad I, Andersen TL, Oma N, Zachariassen S. Investigation report Petroleumstilsynet - Investigation of an accident with fatal consequences on COSLInnovator, 30 December 2015. Technical Report, 2016, https://www.ptil.no/contentassets/34ba6b722c0c44a3a137240bae06f623/investigation-report—cosl-drilling—cosl-innovator.pdf.

[6] Pereira I. Coast guard probing deaths, injuries of americans on vessels in antarctic waters. ABC News 2023. https://abcnews.go.com/International/coast-guard-probing-deaths-injuries-americans-vessels-antarctic/story?id=96864073.

[7] Reuters. Norwegian cruise ship MS maud loses power in north sea during storm. Guardian 2023. https://www.theguardian.com/world/2023/dec/22/norwegian-cruise-ship-ms-maud-loses-power-navigation-system.

[8] Kapsenberg GK. On the slamming of ships (Ph.D. thesis), Delft, The Netherlands: Delft University of Technology; 2018, http://dx.doi.org/10.4233/uuid:14eac2bb-63ee-47e4-8218-1ba3830a97b4.

[9] Cuomo G, Lupoi G, Shimosako K, Takahashi S. Dynamic response and sliding distance of composite breakwaters under breaking and non-breaking wave attack. Coast Eng 2011;58:953–69. http://dx.doi.org/10.1016/j.coastaleng.2011.03.008.

[10] Zhou T, Yin Y, Ma Z, Chen J, Zhai G. Numerical investigation of breaking waves impact on vertical breakwater with impermeable and porous foundation. Ocean Eng 2023;280:114477. http://dx.doi.org/10.1016/j.oceaneng.2023.114477.

[11] Antonini A, Brownjohn JMW, Dassanayake D, Raby A, Bassit J, Pappas A, D'Ayala D. A Bayesian inverse dynamic approach for impulsive wave loading reconstruction. Coast Eng 2021;168:103920. http://dx.doi.org/10.1016/j.coastaleng.2021.103920.

[12] Cuomo G, ichiro Shimosako K, Takahashi S. Wave-in-deck loads on coastal bridges and the role of air. Coast Eng 2009;56:793–809. http://dx.doi.org/10.1016/j.coastaleng.2009.01.005.

[13] Lucio D, Lara J, Tomas A, Losada I. Probabilistic assessment of climate-related impacts and risks in ports. Reliab Eng Syst Saf 2024;251:110333. http://dx.doi.org/10.1016/j.ress.2024.110333.

[14] Han F, Wang W, Zheng X-W, Han X, Shi W, Li X. Investigation of essential parameters for the design of offshore wind turbine based on structural reliability. Reliab Eng Syst Saf 2025;254:110601. http://dx.doi.org/10.1016/j.ress.2024.110601.

[15] LeFevre K. Nasa earthdata; heavy weather, high seas – can dangerous sea states be predicted?. 2015, https://earthdata.nasa.gov/learn/sensing-our-planet/heavy-weather-high-seas.

[16] Bandringa H, Helder JA, van Essen SM. On the validity of CFD for simulating extreme green water loads on ocean-going vessels. In: 39th OMAE Conf. Virtual, Online: ASME; 2020, http://dx.doi.org/10.1115/OMAE2020-18290.

[17] Vanem E, Zhu T, Babanin A. Statistical modelling of the ocean environment – A review of recent developments in theory and applications. Mar Struct 2022;86:103297. http://dx.doi.org/10.1016/j.marstruc.2022.103297.

[18] Winterstein SR, Ude TC, Cornell CA, Bjerager P, Haver S. Environmental parameters for extreme response: inverse FORM with omission factors. In: Int Conf Struct Saf Reliab. Innsbruck, Austria; 1993.

[19] Huseby AB, Vanem E, Natvig B. A new approach to environmental contours for ocean engineering applications based on direct Monte Carlo simulations. Ocean Eng 2013;60:124–35. http://dx.doi.org/10.1016/j.oceaneng.2012.12.034.

[20] Mackay EBL, Murphy-Barltrop C, Jonathan P. The SPAR model: a new paradigm for multivariate extremes: application to joint distributions of metocean variables (OMAE-24-1018). J Offshore Mech Arct Eng 2025;147:011205:1–10. http://dx.doi.org/10.1115/1.4065968.

[21] DNV. Recommended Practice RP-C205: Environmental conditions and environmental loads. Oslo, Norway: Det Norske Veritas; 2019, https://www.dnv.com/energy/standards-guidelines/dnv-rp-c205-environmental-conditions-and-environmental-loads/.

[22] Speers M, Randell D, Tawn J, Jonathan P. Estimating metocean environments associated with extreme structural response to demonstrate the dangers of environmental contour methods. Ocean Eng 2024;311:118754. http://dx.doi.org/10.1016/j.oceaneng.2024.118754.

[23] Gramstad O, Agrell C, Bitner-Gregersen E, Guo B, Ruth E, Vanem E. Sequential sampling method using Gaussian process regression for estimating extreme structural response. Mar Struct 2020;72. http://dx.doi.org/10.1016/j.marstruc.2020.102780.

[24] van Essen SM, Seyffert HC. Finding dangerous waves – Review of methods to obtain wave impact design loads for marine structures (OMAE-22-1110). J Offshore Mech Arct Eng 2023;145(6):060801. http://dx.doi.org/10.1115/1.4056888.

[25] Adegeest L, Braathen A, Løseth R. Use of nonlinear sea loads simulations in design of ships. In: 12th PRADS Conf. The Hague, The Netherlands; 1998, p. 53–8.

[26] Taylor PH, Jonathan P, Harland LA. Time domain simulation of jack-up dynamics with the extremes of a Gaussian process. J Vib Acoust 1997;119:624–8. http://dx.doi.org/10.1115/1.2889772.

[27] Torhaug R, Winterstein SR, Braathen A. Nonlinear ship loads: stochastic models for extreme response. J Ship Res 1998;42, 1:46–55. http://dx.doi.org/10.5957/jsr.1998.42.1.46.

[28] Jensen JJ. Extreme value predictions and critical wave episodes for marine structures by FORM. Ships Offshore Struct 2009;3, 4:325–33. http://dx.doi.org/10.1080/17445300802370461.

[29] Takami T, Fujimoto W, Houtani H, Matsui S. Combination of HOSM and FORM for extreme wave-induced response prediction of a ship in nonlinear waves. Ocean Eng 2023;286:115643. http://dx.doi.org/10.1016/j.oceaneng.2023.115643.
[30] Dermatis A, Lasbleis M, Kim S, de Hauteclocque G, Bouscasse B, Ducrozet G. A multi-fidelity approach for the evaluation of extreme wave loads using nonlinear response-conditioned waves. Ocean Eng 2025;316:119919. http://dx.doi.org/10.1016/j.oceaneng.2024.119919.
[31] Stansberg CT. A wave impact parameter. In: 27th OMAE Conf. Estoril, Portugal: ASME; 2008, http://dx.doi.org/10.1115/OMAE2008-57801.
[32] Bunnik THJ, Stansberg CT, Pákozdi C, Fouques S, Somers L. Useful indicators for screening of sea states for wave impacts on fixed and floating platforms. In: 37th OMAE Conf. Madrid, Spain: ASME; 2018, http://dx.doi.org/10.1115/OMAE2018-78544.
[33] Bunnik THJ, Scharnke J, de Ridder E-J. Efficient indicators for screening of random waves for wave impacts on a jacket platform and a fixed offshore wind turbine. In: 38th OMAE Conf. Glasgow, UK: ASME; 2019, http://dx.doi.org/10.1115/OMAE2019-95481.
[34] Stansberg CT. Wave front steepness and influence on horizontal deck impact loads. J Mar Sci Eng 2020;8, 314. http://dx.doi.org/10.3390/jmse8050314.
[35] van Essen SM, Monroy C, Shen Z, Helder JA, Kim D-H, Seng S, Ge Z. Screening wave conditions for the occurrence of green water events on sailing ships. Ocean Eng 2021;234:109218. http://dx.doi.org/10.1016/j.oceaneng.2021.109218.
[36] Tabandeh A, Jia G, Gardoni P. A review and assessment of importance sampling methods for reliability analysis. Struct Saf 2022;97:102216. http://dx.doi.org/10.1016/j.strusafe.2022.102216.
[37] Chiron M, Genest C, Morio J, Dubreuil S. Failure probability estimation through high-dimensional elliptical distribution modeling with multiple importance sampling. Reliab Eng Syst Saf 2023;235:109238. http://dx.doi.org/10.1016/j.ress.2023.109238.
[38] Fernandez Castellon D, Fenerci A, Petersen ØW, Øiseth O. Full long-term buffeting analysis of suspension bridges using Gaussian process surrogate modelling and importance sampling Monte Carlo simulations. Reliab Eng Syst Saf 2023;235:109211. http://dx.doi.org/10.1016/j.ress.2023.109211.
[39] Chan J, Papaioannou I, Straub D. An adaptive subset simulation algorithm for system reliability analysis with discontinuous limit states. Reliab Eng Syst Saf 2022;225:108607. http://dx.doi.org/10.1016/j.ress.2022.108607.
[40] Cheng K, Papaioannou I, Lu Z, Zhang X, Wang Y. Rare event estimation with sequential directional importance sampling. Struct Saf 2023;100:102291. http://dx.doi.org/10.1016/j.strusafe.2022.102291.
[41] Nguyen PTT, Manuel L. Uncertainty quantification in low-probability response estimation using sliced inverse regression and polynomial chaos expansion. Reliab Eng Syst Saf 2024;242:109750. http://dx.doi.org/10.1016/j.ress.2023.109750.
[42] Nguyen PTT, Manuel L. A bi-fidelity surrogate model for extreme loads on offshore structures. Ocean Eng 2024;307:118175. http://dx.doi.org/10.1016/j.oceaneng.2024.118175.
[43] Roy A, Manna R, Chakraborty S. Support vector regression based metamodeling for structural reliability analysis. Probab. Eng Mech 2018;55:78–89. http://dx.doi.org/10.1016/j.probengmech.2018.11.001.
[44] Fang C, Xu Y-L, Li Y, Li J. Serviceability analysis of sea-crossing bridges under correlated wind and wave loads. Reliab Eng Syst Saf 2024;246:110077. http://dx.doi.org/10.1016/j.ress.2024.110077.
[45] Chojaczyk A, Teixeira A, Neves L, Cardoso J, Soares CG. Review and application of Artificial Neural Networks models in reliability analysis of steel structures. Struct Saf 2015;52:78–89. http://dx.doi.org/10.1016/j.strusafe.2014.09.002.
[46] Bao Y, Sun H, Guan X, Tian. Y. An active learning method using deep adversarial autoencoder-based sufficient dimension reduction neural network for high-dimensional reliability analysis. Reliab Eng Syst Saf 2024;247:110140. http://dx.doi.org/10.1016/j.ress.2024.110140.
[47] Fuhg JN, Fau A, Nackenhorst U. State–of–the–art and comparative review of adaptive sampling methods for kriging. Arch Comput Methods Eng 2021;28:2689–747. http://dx.doi.org/10.1007/s11831-020-09474-6.
[48] Forrester IJ, Sobester A, Keane AJ. Multi-fidelity optimization via surrogate modelling. Proc R Soc A 2007;463:3251–69. http://dx.doi.org/10.1098/rspa.2007.1900.
[49] Kennedy MC, O'Hagan A. Predicting the output from a complex computer code when fast approximations are available. Biometrika 2000;87, 1:1–13.
[50] Guth S, Sapsis TP. Wave episode based Gaussian process regression for extreme event statistics in ship dynamics: Between the Scylla of Karhunen–Loeve convergence and the Charybdis of transient features. Ocean Eng 2022;266:112633. http://dx.doi.org/10.1016/j.oceaneng.2022.112633.
[51] 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 2023;28:1–26. http://dx.doi.org/10.1002/we.2880.
[52] Tang T, Ding H, Dai S, Chen X, Taylor PH, Zang J, Adcock TAA. Data informed model test design with machine learning - an example in nonlinear wave load on a vertical cylinder. In: 42nd OMAE Conf. Melbourne, Australia: ASME; 2023, http://dx.doi.org/10.1115/OMAE2023-102682.
[53] Abaei MM, Leira BJ, vik SS, BahooToroody A. Integrating physics-based simulations with gaussian processes for enhanced safety assessment of offshore installations. Reliab Eng Syst Saf 2024;249:110235. http://dx.doi.org/10.1016/j.ress.2024.110235.
[54] Guth S, Champenois B, Sapsis TP. Application of Gaussian process multi-fidelity optimal sampling to ship structural modeling. In: 34th symp. naval hydrodynamics. Washington DC, USA; 2022.
[55] Mohamad MA, Sapsis TP. Sequential sampling strategy for extreme event statistics in nonlinear dynamical systems. Proc Natl Acad Sci 2018;115, 44. http://dx.doi.org/10.1073/pnas.1813263115.
[56] Huang D, Allen TT, Notz WI, Miller RA. Sequential kriging optimization using multiple-fidelity evaluations.. Struct Multidiscip Optim 2006;32:369–82. http://dx.doi.org/10.1007/s00158-005-0587-0.
[57] Zhang Y, Dong Y, Frangopol DM. An error-based stopping criterion for spherical decomposition-based adaptive Kriging model and rare event estimation. Reliab Eng Syst Saf 2024;241:109610. http://dx.doi.org/10.1016/j.ress.2023.109610.
[58] Kim M, O'Conner K, Pipiras V, Sapsis T. Sampling low-fidelity outputs for estimation of high-fidelity density and its tails. SIAM/ ASA J Uncertain Quantif 2025;13, 1:30–62. http://dx.doi.org/10.1137/24M1639142.
[59] Rigo P, Caprace J-D, Sekulski Z, Bayatfar A, Echeverry S. Structural design optimization – tools and methodologies. In: A holistic approach to ship design. Springer International Publishing; 2019, p. 295–327. http://dx.doi.org/10.1007/978-3-030-02810-7_9.
[60] Kim DH, Lee SG. Reliability analysis of offshore wind turbine support structures under extreme ocean environmental loads. Renew Energy 2015;79:161–6. http://dx.doi.org/10.1016/j.renene.2014.11.052.
[61] Zhao W, Fan F, Wang W. Non-linear partial least squares response surface method for structural reliability analysis. Reliab Eng Syst Saf 2017;161:69–77. http://dx.doi.org/10.1016/j.ress.2017.01.004.
[62] Kim H, Jeong S, Yang C, Noblesse F. Hull form design exploration based on response surface method. In: 21st ISOPE Conf. Int Soc Offshore Polar Eng; 2011, p. 816–25.
[63] Wilkie D, Galasso C. Gaussian process regression for fatigue reliability analysis of offshore wind turbines. Struct Saf 2021;88:102020. http://dx.doi.org/10.1016/j.strusafe.2020.102020.
[64] Kingston GB, Rajabali Nejad M, Gouldby BP, van Gelder PHAJM. Computational intelligence methods for the efficient reliability analysis of complex flood defence structures. Struct Saf 2011;33:1. http://dx.doi.org/10.1016/j.strusafe.2010.08.002.
[65] Saraygord Afshari S, Enayatollahi F, Xu X, Liang X. Machine learning-based methods in structural reliability analysis: A review. Reliab Eng Syst Saf 2022;108223. http://dx.doi.org/10.1016/j.ress.2021.108223.
[66] Marrel A, Iooss B. Probabilistic surrogate modeling by Gaussian process: A review on recent insights in estimation and validation. Reliab Eng Syst Saf 2024;247:110094. http://dx.doi.org/10.1016/j.ress.2024.110094.
[67] Marrel A, Iooss B. Probabilistic surrogate modeling by Gaussian process: A new estimation algorithm for more robust prediction. Reliab Eng Syst Saf 2024;247:110120. http://dx.doi.org/10.1016/j.ress.2024.110120.
[68] van Essen SM, Scholcz TP, Seyffert HC. Prediction of short-term non-linear response using screening combined with multi-fidelity Gaussian Proces Regression. In: 42nd OMAE Conf. Melbourne, Australia: ASME; 2023, http://dx.doi.org/10.1115/OMAE2023-100954.
[69] DNV. Class Guideline DNV-CG-0130: Wave loads. Oslo, Norway; 2018, https://rules.dnv.com/docs/pdf/DNV/CG/2018-01/DNV-CG-0130.pdf.
[70] BV. Rule Note NI638: Guidance for long-term hydro-structure calculations. Paris, France: Bureau Veritas; 2019, https://erules.veristar.com/dy/data/bv/pdf/638-NI_2019-02.pdf.
[71] ABS. Guidance notes on air gap and wave impact analysis for semisubmersibles. Spring, USA: American Bureau of Shipping; 2020, https://ww2.eagle.org/content/dam/eagle/rules-and-guides/current/offshore/249-gn-airgapanalysis-semisubmersibles-2018/air-gap-analysis-gn-may20.pdf.
[72] van Essen SM, Scharnke J, Seyffert HC. Required test durations for converged short-term wave and impact extreme value statistics - Part 1: ferry dataset. Mar Struct 2023;90:103410. http://dx.doi.org/10.1016/j.marstruc.2023.103410.
[73] Scharnke J, van Essen SM, Seyffert HC. Required test durations for converged short-term wave and impact extreme value statistics - Part 2: deck box dataset. Mar Struct 2023;90:103411. http://dx.doi.org/10.1016/j.marstruc.2023.103411.
[74] Dietz JS. Application of conditional waves as critical wave episodes for extreme loads on marine structures (Ph.D. thesis), Lyngby, Denmark: Technical University of Denmark; 2004.
[75] Ochi MK. Applied probability and stochastic processes in engineering and physical sciences. Singapore: John Wiley & Sons; 1990.
[76] DNV. Offshore Technical Guidance DNV-OTG-14: Horizontal wave impact loads for column stabilised units. Oslo, Norway: Det Norske Veritas; 2019, https://www.dnv.com/maritime/Offshore/technical-guidance-otg.html.
[77] Johannessen TB, Lande Ø. Long term analysis of steep and breaking wave properties by event matching. In: 37th OMAE Conf. Madrid, Spain: ASME; 2018, http://dx.doi.org/10.1115/OMAE2018-78283.

[78] Gramstad O, Johannessen TB, Lian G. Long-term analysis of wave-induced loads using High Order Spectral Method and direct sampling of extreme wave events. Mar Struct 2023;103473. http://dx.doi.org/10.1016/j.marstruc.2023.103473.
[79] Bandringa H, Helder JA. On the validity and sensitivity of CFD simulations for a deterministic breaking wave impact on a semi subsmersible. In: 37th OMAE Conf. Madrid, Spain: ASME; 2018, http://dx.doi.org/10.1115/OMAE2018-78089.
[80] Pákozdi C, Califano A, Akselsen A, Croonenborghs E, Kim J, Peric M, Loubeyre S, Bouscasse B, Ducrozet G, Xu-Haihua. Joint-industry effort to develop and verify CFD modeling practice for predicting wave impact. In: 41st OMAE Conf. Hamburg, Germany: ASME; 2022, http://dx.doi.org/10.1115/OMAE2022-79152.
[81] SheffieldML. GPy: A Gaussian process framework in python. 2012, http://github.com/SheffieldML/GPy.
[82] Paleyes A, Pullin M, Mahsereci M, Lawrence N, González J. Emulation of physical processes with emukit. In: 2nd workshop on machine learning and the physical sciences, neurIPS. Vancouver, Canada; 2019, http://dx.doi.org/10.48550/arXiv.2110.13293.
[83] van Essen SM, Seyffert HC. Scripts and data underlying the publication that defines and applies the new Adaptive Screening method, for extreme value prediction of non-linear wave-induced responses. 4TU Repos 2025. http://dx.doi.org/10.4121/f1348609-c912-4d06-82b8-197c01f3437b.
[84] Sharma JN, Dean RG. Development and evaluation of a procedure for simulating a random directional second order sea surface and associated wave forces. Ocean Eng Rep 1979;20.
[85] Forristall GZ. Wave crest distributions: observations and second-order theory. Phys Ocean 2000;30, 8:1931–43. http://dx.doi.org/10.1175/1520-0485(2000)030<1931:WCDOAS>2.0.CO;2.
[86] ITTC. Recommended Practice 7.5-02-07-02.6: Global loads seakeeping procedure. International Towing Tank Conf; 2017, https://www.ittc.info/media/8109/75-02-07-026.pdf.
[87] Haver S. Metocean modelling and prediction of extremes. Stavanger, Norway: Haver & havet, University in Stavanger, NTNU; 2017, p. 1–256.
[88] Mauro F, Braidotti L, la Monaca U, Nabergoj R. Extreme loads determination on complex slender structures. Int Shipbuild Prog 2019;66:57–76. http://dx.doi.org/10.3233/ISP-180256.
[89] de Valk C, van den Brink H. An appraisal of the value of simulated weather data for quantifying coastal flood hazard in the Netherlands . Natural Hazards and Earth System Sciences 2025;25:1769–88. http://dx.doi.org/10.5194/nhess-25-1769-2025, 5.
[90] Bunnik THJ, Helder JA, de Ridder E-J. Deterministic simulation of breaking wave impact and flexible response of a fixed offshore wind turbine. In: 34th OMAE Conf. St. John's, Newfoundland, Canada: ASME; 2015, http://dx.doi.org/10.1115/OMAE2015-41989.
[91] Ferrari V, Gornicz T, Kisjes A, Quadvlieg FHHA. Influence of skeg on ship manoeuvrability at high and low speeds. In: 14th PRADS Conf (2019). Lecture Notes Civil Eng, vol. 63, 2021, p. 384–403. http://dx.doi.org/10.1007/978-981-15-4624-2_23.
[92] Tuitman J. Hydro-elastic response of ship structures to slamming induced whipping (Ph.D. thesis), Delft, The Netherlands: Delft University of Technology; 2010.
[93] Lin F. User manual of PRETTI_R (Lloyd's Register report for Cooperative Research Ships). Technical Report, 2018.
[94] MOERI. KRISO Container Ship (KCS) geometry. Maritime and Ocean Engineering Research Institute (MOERI); 1997, http://www.simman2008.dk/KCS/container.html.
[95] Luppes R, van der Heiden HJL, van der Plas P, Düz B. Simulations of wave impact and two-phase flow with ComFLOW: Past and recent developments. In: 19th ISOPE Conf. Anchorage, USA: Int Soc Offshore Polar Eng; 2013.
[96] Rezende F, Voogt A, Qing Y. JIP FPSO Green Water, project plan. Technical Report, 2017, https://ferforum.com/wp-content/uploads/2017/06/JIP-Green-Water-Project-Plan_RevF-1.pdf.
[97] van Essen SM, Kapsenberg GK, Helmers JB, Monroy C, Thornhill E, Barcarolo D. Proposal CRS SCREAM: SCReening for extremes and maxima (in waves). Technical Report, 2019, https://www.crships.org/.
[98] Rasmussen CE, Williams CKI. Gaussian processes for machine learning. Massachusetts Institute of Technology: MIT Press; 2006, https://gaussianprocess.org/gpml/.
[99] Ebden M. Gaussian processes for regression: a quick introduction. 2008, https://www.robots.ox.ac.uk/~mebden/reports/GPtutorial_old.pdf.
[100] Benker M, Bliznyuk A, Zaeh MF. A Gaussian process based method for data-efficient remaining useful life estimation. IEEE Access 2021;9:137470–82. http://dx.doi.org/10.1109/ACCESS.2021.3116813.
[101] Riihimäki J, Vehtari A. Gaussian processes with monotonicity information. J Mach Learn Res 2010;9:645–52.
[102] Ustyuzhaninov I, Kazlauskaite I, Ek CH, Campbell N. Monotonic Gaussian process flows. In: 23rd Int Conf Artif Intel and Stats. Proc. of machine learning research, vol. 108, PMLR; 2020, p. 3057–67, http://proceedings.mlr.press/v108/ustyuzhaninov20a/ustyuzhaninov20a.pdf.
[103] BV. Rule Note NR583: Whipping and springing assessment. Paris, France: Bureau Veritas; 2015, https://erules.veristar.com/dy/data/bv/pdf/583-NR_2015-07.pdf.
[104] Naess A, Gaidai O, Teigen PS. Extreme response prediction for nonlinear floating offshore structures by Monte Carlo simulation. Appl Ocean Res 2007;29, 4:221–30. http://dx.doi.org/10.1016/j.apor.2007.12.001.
[105] Lian G. Slamming loads on large volume structures from breaking waves (Ph.D. thesis), Stavanger, Norway: University of Stavanger; 2018.
[106] ITTC. Recommended Practice 7.5-02-07-02.3: Experiments on rarely occurring events. International Towing Tank Conf; 2017, https://www.ittc.info/media/8105/75-02-07-023.pdf.