

Sequential Design for the Efficient Estimation of Offshore Structure Failure Probability

Matthew Speers, Jonathan Tawn, Philip Jonathan

School of Mathematical Sciences, Lancaster University LA1 4YF, United Kingdom.

Abstract

Estimation of the failure probability of offshore structures exposed to extreme ocean environments is critical to their safe design and operation. The conditional density of the environment (CDE) quantifies regions of the space of long term environment responsible for extreme structural response. Moreover, the probability of structural failure is obtained by simply integrating the CDE over the environment space. In this work, two methodologies for estimation of the CDE and failure probability are considered. The first (IS-PT) combines parallel tempering MCMC (Markov chain - Monte Carlo, for CDE estimation) with important sampling (for eventual estimation of failure probability). The second (AGE) combines adaptive Gaussian emulation with Bayesian quadrature. Both approaches provide large reductions in the number of function evaluations of the complex structural response to fluid loading, relative to naive brute-force calculations. We evaluate IS-PT and two variants of the AGE procedure in application to a simple synthetic structure with multimodal CDE, and a monopile structure exhibiting non-linear resonant response. IS-PT provides reliable results for both applications. The AGE procedures require balancing exploration and exploitation of the environment space, using a typically-unknown weight parameter, λ . When λ is known, perhaps from prior engineering knowledge, AGE provides an order of magnitude reduction in computational cost relative to IS-PT. However, with λ unknown, IS-PT is more reliable.

Keywords: Structural design, extreme, full probabilistic analysis, contour, importance sampling, bridge sampling, Gaussian process, active learning, significant wave height, wave steepness, monopiles.

PACS: 0000, 1111

2000 MSC: 0000, 1111

1. Introduction

1.1. Background

An offshore structure (such as an oil platform or wind turbine) is subject to environmental loading, e.g., from winds, waves and currents. The ocean engineer seeks to evaluate the risk posed to structural integrity by the environment, enabling the structure to be designed and

maintained to the required level of reliability. Often, this involves computationally demanding fluid loading and structural response calculations. Therefore, the design of computationally efficient approaches for assessment of structural risk is a topic of considerable importance.

Take an environmental variable $\mathbf{X}$ (such as significant wave height H_S) characterising the long term metocean environment on space $\mathcal{E}_{\mathbf{X}}$. The short term environment variable $\mathbf{Y}$ (such as individual wave height H) defined on space $\mathcal{E}_{\mathbf{Y}}$, depends on $\mathbf{X}$. This $\mathbf{Y}$ is stochastic given $\mathbf{X}$, in the sense that many values of $\mathbf{Y}$ may be summarised by a single value $\mathbf{x}$ of $\mathbf{X}$, in terms of the distribution for $\mathbf{Y}|\{\mathbf{X} = \mathbf{x}\}$. Given complete knowledge of the short term conditions $\mathbf{Y}$, along with a physical model for the response $\mathbf{R} \in \mathcal{E}_{\mathbf{R}}$ induced on the structure by $\mathbf{Y}$, it is possible to characterise the *multivariate* structural response induced on the structure fully. Typically, $\mathcal{E}_{\mathbf{R}} = \mathbb{R}^d$ for some dimension $d > 0$.

In our setting, we assume the existence of a deterministic function $g_{\mathbf{R}}(\mathbf{y}) : \mathcal{E}_{\mathbf{Y}} \mapsto \mathcal{E}_{\mathbf{R}}$ for the structural response $\mathbf{R} = \mathbf{r}$ induced by the environment $\mathbf{Y} = \mathbf{y}$. Practitioners do not typically have knowledge of the full short term environment $\mathbf{Y}$, but instead have information on the long term summary variable $\mathbf{X}$. Since $\mathbf{Y}$ is not a deterministic function of $\mathbf{X}$, practitioners estimate the density function $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) : \mathcal{E}_{\mathbf{Y}} \times \mathcal{E}_{\mathbf{X}} \mapsto \mathbb{R}^+$ of the short term environment $\mathbf{Y}|\{\mathbf{X} = \mathbf{x}\}$. In practice, evaluation of $g_{\mathbf{R}}$ and $f_{\mathbf{Y}|\mathbf{X}}$ can be computationally expensive, involving complex load calculations and the simulation of 3-dimensional wave and wind fields.

Given the functions $g_{\mathbf{R}}$ and $f_{\mathbf{Y}|\mathbf{X}}$, we can evaluate the density $f_{\mathbf{R}|\mathbf{X}}(\mathbf{r}|\mathbf{x}) : \mathcal{E}_{\mathbf{R}} \times \mathcal{E}_{\mathbf{X}} \mapsto \mathbb{R}^+$ as

$$f_{\mathbf{R}|\mathbf{X}}(\mathbf{r}|\mathbf{x}) = \int_{\mathcal{E}_{\mathbf{Y}}} g_{\mathbf{R}}(\mathbf{y}) f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) d\mathbf{y}, \quad (1)$$

for $\mathbf{R}|\{\mathbf{X} = \mathbf{x}\}$, the multivariate response conditioned on the long term environment. Again, evaluating $f_{\mathbf{R}|\mathbf{X}}$ can be prohibitively expensive, due the potential complexities of $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{x})$ and $g_{\mathbf{R}}(\mathbf{y})$. A natural approach to quantify the risk to a structure is then to estimate the probability of failure p due to response $\mathbf{R}$ and environment $\mathbf{X}$. For $\mathbf{R} = (R_1, \dots, R_d) \in \mathcal{E}_{\mathbf{R}}$, this can be written

$$p = \mathbb{P} \left(\bigcup_{i=1}^d (R_i > r_{\text{Cr}}^{(i)}) \right) = 1 - \mathbb{P} \left(\bigcap_{i=1}^d (R_i < r_{\text{Cr}}^{(i)}) \right), \quad (2)$$

the probability that at least one response component R_i , $i = 1, \dots, d$, exceeds its critical level $r_{\text{Cr}}^{(i)} \in \mathbb{R}$. This can be written using (1) as

$$p = \int_{\mathcal{E}_{\mathbf{X}}} \left\{ \int_{\mathcal{E}_{\mathbf{R}}} \left[1 - \left(\prod_{i=1}^d I(R_i < r_{\text{Cr}}^{(i)} | \{\mathbf{X} = \mathbf{x}\}) \right) \right] f_{\mathbf{R}|\mathbf{X}}(\mathbf{r}|\mathbf{x}) d\mathbf{r} \right\} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}, \quad (3)$$

where I is the indicator function which takes the value unity when its argument is true, and value zero otherwise, and the integral evaluated numerically by sampling repeatedly from models for $\mathbf{R}|\{\mathbf{X} = \mathbf{x}\}$ and $\mathbf{X}$. Throughout, we assume the density $f_{\mathbf{X}}$ of the long term environment $\mathbf{X}$ is either known or estimable, possibly using extreme value techniques (e.g., as in Section 4). Evaluation of (3) is thus solely made difficult by the computational expense

required to obtain draws of $\mathbf{R}|\{\mathbf{X} = \mathbf{x}\}$.

We aim to minimise the uncertainty in estimating (3) given a budget of a set number of realisations of $\mathbf{R}|\{\mathbf{X} = \mathbf{x}\}$. We use the available budget efficiently by making informed choices about the values of $\mathbf{X}$ at which to sample from $\mathbf{R}|\{\mathbf{X} = \mathbf{x}\}$. Typically, methods for the efficient evaluation of (3) target values of $\mathbf{X}$ contributing most to the integral. In the simplest terms, this is achieved by targetting regions where the value of the integrand

$$\tilde{f}_{\mathbf{X}}(\mathbf{x}; \mathbf{r}_{\text{Cr}}) = \left\{ \int_{\mathcal{E}_{\mathbf{R}}} \left[1 - \left(\prod_{i=1}^d I(R_i < r_{\text{Cr}}^{(i)} | \{\mathbf{X} = \mathbf{x}\}) \right) \right] f_{\mathbf{R}|\mathbf{X}}(\mathbf{r}|\mathbf{x}) d\mathbf{r} \right\} \times f_{\mathbf{X}}(\mathbf{x}) \quad (4)$$

$$= \mathbb{P}(\text{'failure'} | \{\mathbf{X} = \mathbf{x}\}) \times f_{\mathbf{X}}(\mathbf{x}), \quad (5)$$

in (3) is large, where $\mathbf{r}_{\text{Cr}} = (r_{\text{Cr}}^{(1)}, \dots, r_{\text{Cr}}^{(d)})$ is the vector of critical values of responses. That is, it is beneficial to target values of the long-term environmental variables that are both likely to occur (large $f_{\mathbf{X}}(\mathbf{x})$) and to induce structural failure (large $\mathbb{P}(\text{'failure'} | \{\mathbf{X} = \mathbf{x}\})$). We subsequently refer to $\tilde{f}_{\mathbf{X}}(\mathbf{x}; \mathbf{r}_{\text{Cr}})$ defined in (4) as the conditional density of the environment (CDE), as it is the unnormalised long-term environment density conditional on the occurrence of structural failure; we use $\tilde{f}$ (rather than f) to indicate an unnormalised density.

Peherstorfer et al. (2016), Yang et al. (2018) and Wang et al. (2021) show that minimising the uncertainty in (3) can be achieved for an arbitrary multi-dimensional response. We restrict ourselves to $d = 1$, with $\mathbf{R} = R$ and $\mathbf{r}_{\text{Cr}} = r_{\text{Cr}}$, for brevity and ease of presentation. In this case, equation (4) reduces to

$$\tilde{f}_{\mathbf{X}}(\mathbf{x}; r_{\text{Cr}}) = \mathbb{P}(R > r_{\text{Cr}} | \{\mathbf{X} = \mathbf{x}\}) \times f_{\mathbf{X}}(\mathbf{x}). \quad (6)$$

Existing methodologies reducing uncertainty in (3) by targetting (6) include: *sampling* methods such as importance sampling (see e.g., Castellon et al. 2022) and bridge sampling (Meng and Wong, 1996); *adaptive Gaussian emulation* (e.g., Gramstad et al. 2020 and Lystad et al. 2023); and approaches combining sampling and adaptive emulation (e.g., Castellon et al. 2023 and Xiao et al. 2020). Good sampling methods reduce the variance of a target integral for a given sampling budget, whilst emulation provides a cheaper approximate route to otherwise expensive complex function evaluation. Relevant recent reviews are given by Moustapha et al. (2022), Wang et al. (2022), Tabandeh et al. (2022) and Marrel and Iooss (2024).

In simple cases, we might expect that the CDE $\tilde{f}_{\mathbf{X}}$ is approximately elliptically-contoured (e.g., Speers et al. 2024), and therefore well-approximated by a unimodal Gaussian-like density in $\mathcal{E}_{\mathbf{X}}$. However, in reality there are good reasons to expect this not to be the case in general, due to e.g., the presence of multiple failure modes or resonant responses. In the current work, we are particularly interested in investigating methodologies to estimate such complex CDE structures well.

We choose to investigate the efficient estimation of (6) in the context of designing monopile structures. We choose this structural type for two reasons: firstly, because it provides a useful template structure for generic studies of fluid loading; and secondly, it is of itself a relevant

structural type for e.g., offshore wind applications. This thinking motivates the synthetic study of Section 3, and the wind turbine application of Section 4.

1.2. Objectives and outline

The objective of the current work is to explore methodologies based on efficient sampling or adaptive Gaussian emulation, to estimate the conditional density of the environment (CDE) and thereby failure probabilities for synthetic and real-world monopile structures. In Section 2, we first describe an approach, termed IS-PT, coupling importance sampling with parallel tempering Markov chain-Monte Carlo (MCMC) (Earl and Deem, 2005) for estimation of multi-modal CDEs, a scenario which has received little attention in the offshore reliability literature. Secondly, building on Gramstad et al. (2020) and Cohn (1993), we consider two variants of an alternative approach, termed AGE, based on adaptive Gaussian emulation, adopting an acquisition function promoting sampling which balances exploration and exploitation of regions of $\mathcal{E}_{\mathbf{X}}$ contributing to the CDE. In Section 3, the approaches from Section 2 are applied for a synthetic monopile structure exhibiting different resonant responses, to evaluate their respective performance. We find that all approaches provide good estimation of failure probability, but that AGE approaches require fewer expensive function evaluations provided that the required balance between exploitation and exploration of $\mathcal{E}_{\mathbf{X}}$ is assumed known. If this balance is unknown, IS-PT provides a more reliable procedure. In Section 4, we demonstrate good performance of all approaches in a more realistic setting, estimating the structural failure probability for oscillating monopiles, with harmonic response modelled using the T-FNV (Transformed - Faltinsen, Newman and Vinje, Faltinsen et al. 1995) model of Taylor et al. (2024). Our findings here regarding the relative computational complexities of IS-PT and AGE approaches are similar to those for the synthetic case. Discussion and conclusions are provided in Section 5. Online supplementary material (SM) provides supporting description of methodology and results.

2. Methodology

2.1. Overview of methodologies

We begin by discussing two methods for the efficient evaluation of integral (3). In Section 2.2, we introduce an importance sampling scheme coupled with an adaptive parallel tempering MCMC algorithm, designed for scenarios where the CDE $\tilde{f}_{\mathbf{X}}$ is multimodal. In Section 2.3, we describe an emulator replacing expensive draws of the structural response $R|\{\mathbf{X} = \mathbf{x}\}$ with predictions from a Gaussian process, and provide methods for the adaptive design of the emulator training set.

We emphasise that these methods are introduced as *alternative* options for the efficient estimation of (3), both seeking to minimise the target error within some set budget of expensive function evaluations. These approaches will then be compared in (3) to see which performs better.

2.2. MCMC-informed importance sampling

In offshore reliability, importance sampling methods select values $\mathbf{x}_1^*, \dots, \mathbf{x}_{n_{\text{IS}}}^*$, $n_{\text{IS}} > 0$, of $\mathbf{X}$ at which to evaluate $R|\{\mathbf{X} = \mathbf{x}\}$, to make efficient use of limited computational resources (see e.g., [Castellon et al. 2022](#)). These include traditional importance sampling techniques, or extensions such as bridge sampling (e.g., [Meng and Wong 1996](#)). In this article, we focus our attention on the former, since our initial investigations of bridge sampling showed no improvement in performance, despite increased computational cost.

Evaluation points are drawn from a proposal distribution g_{Pr} , chosen so that values $\mathbf{x}$ with higher density $g_{\text{Pr}}(\mathbf{x})$ are more informative to the target quantity. Our target quantity is the marginal structural failure probability (3), which may be written

$$p = \int_{\mathcal{E}_{\mathbf{X}}} \mathbb{P}(R > r_{\text{Cr}}|\{\mathbf{X} = \mathbf{x}\}) \frac{f_{\mathbf{X}}(\mathbf{x})}{g_{\text{Pr}}(\mathbf{x})} g_{\text{Pr}}(\mathbf{x}) d\mathbf{x}, \quad (7)$$

approximated by the importance sampling estimate

$$\hat{p}_{\text{IS}} = \frac{1}{n_{\text{IS}}} \sum_{i=1}^{n_{\text{IS}}} \mathbb{P}(R > r_{\text{Cr}}|\{\mathbf{X} = \mathbf{x}_i^*\}) \frac{f_{\mathbf{X}}(\mathbf{x}_i^*)}{g_{\text{Pr}}(\mathbf{x}_i^*)}, \quad (8)$$

for $\mathbf{x}_1^*, \dots, \mathbf{x}_{n_{\text{IS}}}^* \sim g_{\text{Pr}}$. The variance of $\hat{p}_{\text{IS}}$ is dependent on the proposal density g_{Pr} , with the optimal choice of proposal minimising the variance in the estimate for the fixed budget n_{IS} . Here, the choice of g_{Pr} minimising this variance is given by the CDE (6, [Rubinstein and Kroese 2016](#)), so methods typically attempt to find proposal densities approximately equal to the CDE, either by using MCMC (e.g., [Xiao et al. 2020](#)) or surrogate modelling of the response function (e.g., [Lystad et al. 2023](#)).

We choose to develop methodology to estimate the CDE for use as proposal density g_{Pr} utilising an MCMC scheme with the CDE $\tilde{f}_{\mathbf{X}}(\mathbf{x}; r_{\text{Cr}})$ as the posterior target distribution from Bayes' rule $\tilde{\pi}(\mathbf{x}|\theta) = \pi(\theta|\mathbf{x}) \times \pi(\mathbf{x})$, where $\pi(\theta|\mathbf{x})$ is an empirical estimate of the probability $\mathbb{P}(R > r_{\text{Cr}}|\{\mathbf{X} = \mathbf{x}\})$ obtained by repeated sampling of $R|\{\mathbf{X} = \mathbf{x}\}$, and $\pi(\mathbf{x}) = f_{\mathbf{X}}(\mathbf{x})$. Using this approach, we obtain a sample from $\tilde{f}_{\mathbf{X}}(\mathbf{x}; r_{\text{Cr}})$, and adopt a Gaussian smoothed version of $\tilde{\pi}(\mathbf{x}|\theta)$ as the proposal density g_{Pr} , see supplementary material SM3.1.

In simple scenarios, with lower dimensional environment space $\mathcal{E}_{\mathbf{X}}$ and unimodal, approximately elliptically-contoured $\tilde{f}_{\mathbf{X}}(\mathbf{x}; r_{\text{Cr}})$, MCMC samples can be obtained using traditional algorithms such as Metropolis-Hastings (see e.g., [Chib and Greenberg 1995](#)). In practice, however, the posterior $\tilde{f}_{\mathbf{X}}(\mathbf{x}; r_{\text{Cr}})$ may be more complex, e.g., exhibiting multi-modality or obvious departures from an elliptically-contoured density. Here we adopt parallel tempering MCMC as a more robust approach to estimate CDEs of arbitrary complexities.

Parallel tempering MCMC allows jumps between disjoint positive-density regions of the target distribution $\tilde{\pi}$ by combining some $n_{\text{Tm}} > 1$ MCMC chains, each targetting scaled forms of $\tilde{\pi}$. These chains are evaluated at different ‘temperatures’ $T_1, \dots, T_{n_{\text{Tm}}} > 0$, with the j th chain, $j = 1, \dots, n_{\text{Tm}}$, sampling from $\tilde{\pi}^{1/T_j}$; chains with a higher temperature target a ‘flatter’ form of the target posterior density $\tilde{\pi}$, allowing movement between otherwise disjoint regions

of positive density. Individual chains are sampled using a Metropolis-Hastings scheme with proposal density $\mathbf{x}'|\mathbf{x} \sim N(\mathbf{x}, \sigma_{\text{MH}}^2)$, $\sigma_{\text{MH}} > 0$, and acceptance probability α_{MH} . Swaps between chains i and j , ($i, j = 1, \dots, n_{\text{Tm}}, i \neq j$), are periodically proposed with acceptance probability $\alpha_{\text{Sw}}(i, j)$, allowing chains of lower temperature to move between disjoint high-density regions in $\mathcal{E}_{\mathbf{X}}$. [Sambridge \(2013\)](#) shows that, for a parallel tempering algorithm to satisfy detailed balance, individual-chain moves from $\mathbf{x}$ to $\mathbf{x}'$ should be accepted with probability $\alpha_{\text{MH}} = \min\{1, \tilde{\pi}(\mathbf{x}'|\theta)/\tilde{\pi}(\mathbf{x}|\theta)\}$, and swaps between the i th chain at state $\mathbf{x}_i$ and the j th chain at state $\mathbf{x}_j$ should be accepted with probability

$$\alpha_{\text{Sw}}(i, j) = \min \left\{ 1, \left(\frac{\tilde{\pi}(\mathbf{x}_j|\theta)}{\tilde{\pi}(\mathbf{x}_i|\theta)} \right)^{1/T_i} \left(\frac{\tilde{\pi}(\mathbf{x}_i|\theta)}{\tilde{\pi}(\mathbf{x}_j|\theta)} \right)^{1/T_j} \right\}. \quad (9)$$

Typically swaps are proposed only between adjacent chains, at predetermined set intervals. We use the approach of [Vousden et al. \(2015\)](#), adaptively selecting the temperature ladder $T_1, \dots, T_{n_{\text{Tm}}}$, as well as the step size standard deviation σ_{MH} for each individual chain. This method is implemented in the Python `pyPESTO` module ([Schälte et al., 2023](#)), employed for all MCMC sampling in this work.

A schematic of the resulting sequential design algorithm, henceforth referred to as importance sampling-parallel tempering (IS-PT), is given in [Figure 1](#).

2.3. Adaptive Gaussian emulation

2.3.1. Gaussian emulation

Importance sampling reduces the number of evaluations of the expensive response function needed for the calculation of failure probability (3). It does, however, still require some number of expensive evaluations, with this number being dependent on the convergence rate of the MCMC required for proposal distribution estimation. An alternative approach further reducing the need for computationally expensive evaluations is to replace draws from the true response function with estimates provided by a surrogate model, such as a Gaussian process emulator.

Various approaches to Gaussian process (GP) emulation have been reported in the offshore literature. [Gramstad et al. \(2020\)](#), [Castellon et al. \(2023\)](#) and [Lystad et al. \(2023\)](#) assume a parametric form for the distribution of the structural response $R|\{\mathbf{X} = \mathbf{x}\}$, and so model realisations as draws from this parametric distribution, with unknown parameters modelled as a GP. In our case, we choose to target the (logarithm of the) CDE (6). To do so, we make repeated draws from $R|\{\mathbf{X} = \mathbf{x}\}$ at n_{Tr1} training points $\mathbf{x}_1, \dots, \mathbf{x}_{n_{\text{Tr1}}}$, obtaining estimates of the conditional failure probability $\mathbb{P}(R > r_{\text{Cr}}|\{\mathbf{X} = \mathbf{x}\})$ at each of these values for $\mathbf{X}$. These values form the training set $\mathcal{D} \subset \mathcal{E}_{\mathbf{X}}$, the selection of which is discussed in [Section 2.3.2](#). After training the GP emulator on $\mathcal{D}$, we can then emulate the CDE at un-observed values $\mathbf{x} \notin \mathcal{D}$.

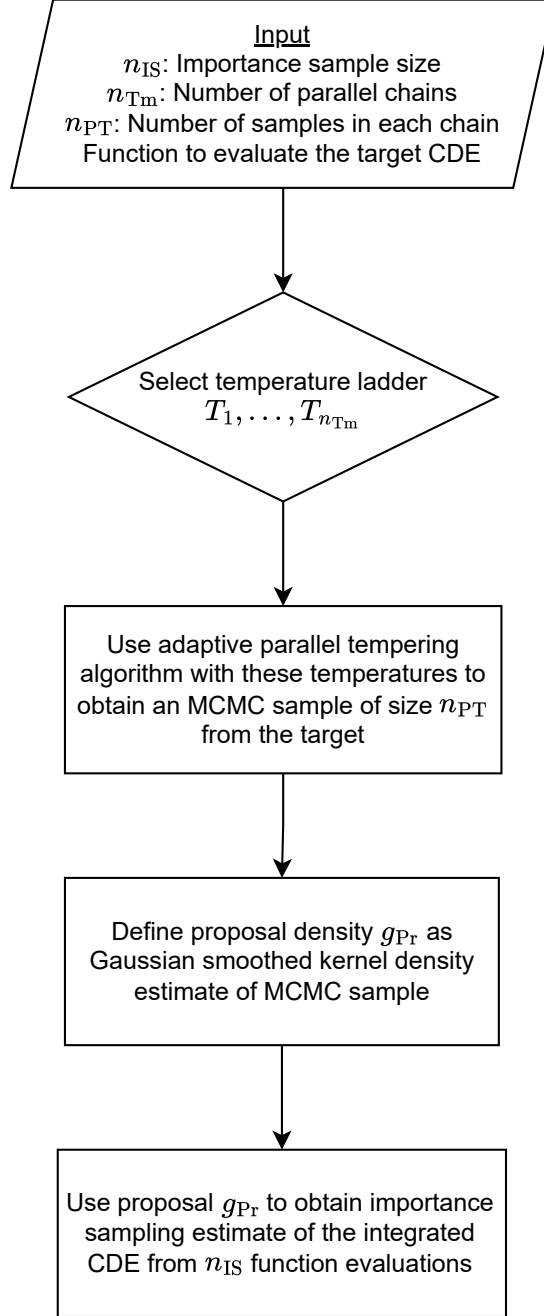


Figure 1: Schematic for the importance sampling-parallel tempering (IS-PT) sequential design algorithm of Section 2.2.

We define the GP emulator for the log-CDE as

$$w(\mathbf{x}) = \log\{\mathbb{P}(R > r_{\text{Cr}}|\{\mathbf{X} = \mathbf{x}\}) f_{\mathbf{X}}(\mathbf{x})\} \sim \mathcal{GP}(\mu_{\text{GP}}(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')), \quad w : \mathcal{E}_{\mathbf{X}} \mapsto \mathbb{R}, \quad (10)$$

for mean and covariance functions

$$\mu_{\text{GP}}(\mathbf{x}) = \mathbb{E}[w(\mathbf{x})], \quad \mu_{\text{GP}} : \mathcal{E}_{\mathbf{X}} \rightarrow \mathbb{R}, \quad (11)$$

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(w(\mathbf{x}) - \mu(\mathbf{x}))(w(\mathbf{x}') - \mu(\mathbf{x}'))], \quad k(\mathbf{x}, \mathbf{x}') : \mathcal{E}_{\mathbf{X}} \times \mathcal{E}_{\mathbf{X}} \rightarrow \mathbb{R}, \quad (12)$$

where the log transform is used to ensure positivity of the predicted CDE. Compared to direct emulation of the failure probability (see SM1.2) this approach is advantageous when the density $f_{\mathbf{X}}$ is itself not modelled continuously; see for instance the gridded density estimated in Section 4.4. Under this model, estimation of (3) is an application of Bayesian quadrature using a Gaussian process prior (e.g., Hennig et al. 2022).

For the kernel function k , we use the Matérn kernel (e.g., Genton 2001)

$$k_{\text{Mt}}(\mathbf{x}, \mathbf{x}') = \sigma_{\text{Mt}}^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} \frac{\|\mathbf{x} - \mathbf{x}'\|}{\ell} \right)^\nu K_\nu \left(\sqrt{2\nu} \frac{\|\mathbf{x} - \mathbf{x}'\|}{\ell} \right), \quad (13)$$

with variance and length scale parameters $\sigma_{\text{Mt}}^2, \ell > 0$, and smoothness parameter $\nu > 0$, where $\|\cdot\|$ is the Euclidean norm, $\Gamma : \mathbb{R} \mapsto \mathbb{R}$ is the gamma function, and $K_\nu : \mathbb{R}^+ \mapsto \mathbb{R}^+$ is the modified Bessel function of the second kind (e.g., Abramowitz and Stegun 1965); this kernel is chosen for its improved ability to capture sudden changes in the target function relative the squared exponential kernel. We combine the Matérn kernel (weighted by a multiplicative constant $C_{\text{Kr}} > 0$) with additive white noise kernel $k_{\text{WN}}(\mathbf{x}, \mathbf{x}') = \kappa$ when $\mathbf{x} = \mathbf{x}'$, and $= 0$ otherwise, for constant white noise variance $\kappa > 0$, yielding the full kernel function

$$k(\mathbf{x}, \mathbf{x}') = C_{\text{Kr}} k_{\text{Mt}}(\mathbf{x}, \mathbf{x}') + k_{\text{WN}}(\mathbf{x}, \mathbf{x}'). \quad (14)$$

The addition of white noise allows the model to perform well when evaluations of the true target are made with some uncertainty, as is typically the case in application, see Section 4.4. Kernel parameters $\sigma_{\text{Mt}}^2, \ell, C_{\text{Kr}}$ and κ are jointly estimated via maximum likelihood at each posterior update using the L-BFGS-B algorithm (see Pedregosa et al. 2011). The remaining parameter ν must be fixed; a brief sensitivity analysis suggests $\nu = 2.5$ as a sensible choice.

We assume a flat prior $\mu_{\text{GP}}(\mathbf{x}) = 0$ for all $\mathbf{x} \in \mathcal{E}_{\mathbf{X}}$. Given covariance function k as defined above, and training data $\mathbf{w} = (w(\mathbf{x}_1), \dots, w(\mathbf{x}_{n_{\text{Tr1}}}))$, the posterior predictive mean μ_{GP}^* and covariance function k^* for regression (10) can be found as,

$$\mu_{\text{GP}}^*(\mathbf{x}) = \mu_{\text{GP}}(\mathbf{x}) + k(\mathcal{D}, \mathbf{x})^T (k(\mathcal{D}, \mathcal{D}) + \alpha_{\text{Ng}} I_{n_{\text{Tr1}}})^{-1} (\mathbf{w} - \mu_{\text{GP}}(\mathcal{D})), \quad (15)$$

$$k^*(\mathbf{x}, \mathbf{x}') = k(\mathbf{x}, \mathbf{x}') - k(\mathcal{D}, \mathbf{x})^T (k(\mathcal{D}, \mathcal{D}) + \alpha_{\text{Ng}} I_{n_{\text{Tr1}}})^{-1} k(\mathcal{D}, \mathbf{x}'), \quad (16)$$

where α_{Ng} is an assumed observational nugget variance. In practice, we take $\alpha_{\text{Ng}} = 10^{-5}$. Given this trained GP emulator, the target marginal failure probability estimate $\hat{p}_{\text{GP}}$ can be

calculated using

$$\hat{p}_{\text{GP}} = \mathbb{E}_{W, \mathbf{X}}(\{\exp(w(\mathbf{x}))\}) = \int_{\mathcal{E}_{\mathbf{X}}} \left\{ \int_{\mathbb{R}} \exp(w) \phi(w; \mu_{\text{GP}}^*(\mathbf{x}), k^*(\mathbf{x}, \mathbf{x})) dw \right\} d\mathbf{x} \quad (17)$$

$$= \int_{\mathcal{E}_{\mathbf{X}}} \exp\left(\mu_{\text{GP}}^*(\mathbf{x}) + \frac{k^*(\mathbf{x}, \mathbf{x}')}{2}\right) d\mathbf{x}, \quad (18)$$

using the expression for the expectation of log-normal random variable $\exp(w)$, for $W|\{\mathbf{X} = \mathbf{x}\} \sim N(\mu_{\text{GP}}^*(\mathbf{x}), k^*(\mathbf{x}, \mathbf{x}))$ with parameters obtained from (10). For display purposes in the figures of Section 3, we evaluate the performance of our GP regression (10) in terms of the absolute difference Δ_{GP} between the true failure probability (3) and this estimate

$$\Delta_{\text{GP}} = |p - \hat{p}_{\text{GP}}|. \quad (19)$$

2.3.2. Active learning

The surrogate model (10) must be trained to provide reliable estimates of the CDE. Often, this training is carried out iteratively, with iteration n training the surrogate against true evaluations of $w(\mathbf{x})$ for all $\mathbf{x}$ in a training set $\mathcal{D}_n$, chosen inductively: at iteration $n + 1$, we update training set $\mathcal{D}_n$ to $\mathcal{D}_{n+1} = \{\mathcal{D}_n, \mathbf{x}_{n+1}\}$, where $\mathbf{x}_{n+1} = \arg\max_{\mathbf{x} \in \mathcal{E}_{\mathbf{X}}} U_n(\mathbf{x})$, for acquisition function U_n taking the form

$$U_n(\mathbf{x}; \lambda) = \lambda \Sigma_n(\mathbf{x}) + (1 - \lambda) M_n(\mathbf{x}), \quad (20)$$

for $\lambda \in [0, 1]$. Specification of the initial training set $\mathcal{D}_0$ is discussed in Section 3. Here $\Sigma_n : \mathcal{E}_{\mathbf{X}} \mapsto \mathbb{R}$ is an exploration term encouraging sampling at points far from existing members of $\mathcal{D}_n$, and $M_n : \mathcal{E}_{\mathbf{X}} \mapsto \mathbb{R}$ is an exploitation term encouraging sampling close to high values of the target function; see Pollatsek and Tversky (1970) for an early discussion of this utility form. In our setting, $M_n : \mathcal{E}_{\mathbf{X}} \mapsto \mathbb{R}$ is large at values $\mathbf{x}$ with high contributions to the integral (17), motivating our first acquisition function

$$U_n^{(1)}(\mathbf{x}; \lambda) = \lambda \log k_n^*(\mathbf{x}, \mathbf{x}) + (1 - \lambda) \log \hat{f}_{\mathbf{X}}^{(n)}(\mathbf{x}; r_{\text{Cr}}), \quad (21)$$

where k_n^* is the posterior GP kernel function obtained via (15) with training set $\mathcal{D}_n$, and $\hat{f}_{\mathbf{X}}^{(n)}(\mathbf{x}; r_{\text{Cr}})$ is the CDE estimate at iteration n . Gramstad et al. (2020), Lystad et al. (2023) and Wang et al. (2024) provide examples of iterative schemes using Gaussian process emulation with acquisition functions similar to (21), for their respective forms of GP emulator (10). Following (3) and (4), we estimate the CDE $\hat{f}_{\mathbf{X}}^{(n)}(\mathbf{x}; r_{\text{Cr}})$ as the integrand in (17), namely

$$\hat{f}_{\mathbf{X}}^{(n)}(\mathbf{x}; r_{\text{Cr}}) = \exp\left(\mu_n^*(\mathbf{x}) + \frac{k_n^*(\mathbf{x}, \mathbf{x}')}{2}\right), \quad (22)$$

where μ_n^* is the posterior GP mean function obtained via posterior update (15) with training set $\mathcal{D}_n$.

A similar approach is the active learning Cohn (ALC) scheme of Cohn (1993), aiming to

reduce the overall variance of the GP surrogate on $\mathcal{E}_{\mathbf{X}}$. They find the deduced reduction in variance across the *entire space* $\mathcal{E}_{\mathbf{X}}$, given the addition of a new query point $\mathbf{x}$ to the training set $\mathcal{D}_n$ at training iteration n . This is approximated over a reference set $\mathcal{P} = \{\mathbf{x}_j\}_{j=1}^{n_{\text{Rf}}}$ on $\mathcal{E}_{\mathbf{X}}$ as

$$\text{ALC}(\mathbf{x}) = \frac{1}{n_{\text{Rf}}} \sum_{j=1}^{n_{\text{Rf}}} k_n^*(\mathbf{x}_j, \mathbf{x}_j) - \tilde{k}_{n+1}^*(\mathbf{x}_j, \mathbf{x}_j; \mathbf{x}), \quad \mathbf{x} \in \mathcal{P}, \quad (23)$$

for positive semi-definite $\text{ALC}(\mathbf{x})$, where $\tilde{k}_{n+1}^*(\mathbf{x}_i, \mathbf{x}_i; \mathbf{x})$ is the variance of the GP (10) at iteration $n+1$, given that query point $\mathbf{x}$ is chosen as the next training point, thereby making $\mathcal{D}_{n+1} = \{\mathcal{D}_n, \mathbf{x}\}$. The summand in (23) can be written

$$k_n^*(\mathbf{x}_j, \mathbf{x}_j) - \tilde{k}_{n+1}^*(\mathbf{x}_j, \mathbf{x}_j; \mathbf{x}) = \frac{(\mathbf{k}_{n,j}^* \mathbf{C}_n^{*-1} \mathbf{m}_n^* - k_n^*(\mathbf{x}, \mathbf{x}_j))^2}{(k_n^*(\mathbf{x}, \mathbf{x}) - \mathbf{m}_n^{*T} \mathbf{C}_n^{*-1} \mathbf{m}_n^*)}, \quad (24)$$

see [Seo et al. 2000](#), where $\mathbf{C}_n^* = k_n^*(\mathcal{D}_n, \mathcal{D}_n)$ is the covariance matrix over current design points, $\mathbf{k}_{n,j}^* = k_n^*(\mathcal{D}_n, \mathbf{x}_j)$ is the vector of covariances between the training data and reference point $\mathbf{x}_j$ and $\mathbf{m}_n^* = k_n^*(\mathcal{D}_n, \mathbf{x})$ is the covariance vector between the training data and the query point $\mathbf{x}$.

[Seo et al. \(2000\)](#) recommend selecting the best next query point $\mathbf{x}$ by maximising a *weighted* sum of (24) over the reference grid $\mathcal{P}$. Instead, we employ an acquisition function of the form (20) utilising the ALC criterion. We find that the acquisition function

$$U_n^{(2)}(\mathbf{x}; \lambda) = \lambda \log \text{ALC}(\mathbf{x}) + (1 - \lambda) \log \hat{f}_{\mathbf{X}}^{(n)}(\mathbf{x}; r_{\text{Gr}}), \quad (25)$$

performs well for careful choice of λ . This is similar to the acquisition function (21), except that in (25) the exploration term $\Sigma_n(\mathbf{x}) = \log \text{ALC}(\mathbf{x})$ considers the effect of including a query point $\mathbf{x}$ in reducing error on the whole candidate space $\mathcal{E}_{\mathbf{X}}$, rather than just at the query point itself.

A schematic of the adaptive Gaussian emulation (AGE) procedure for sequential design is given in [Figure 2](#).

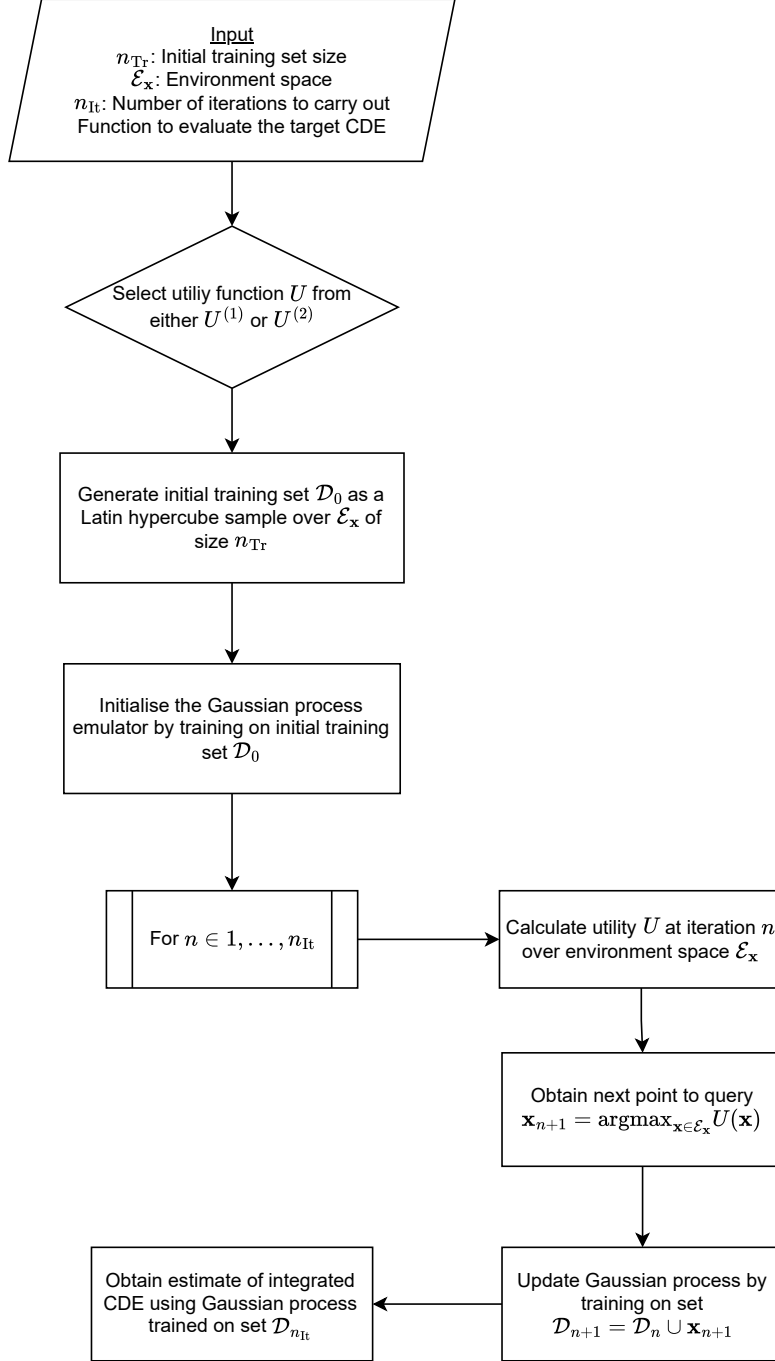


Figure 2: Schematic for the adaptive Gaussian emulation (AGE) sequential design algorithm of Section 2.3.

3. Synthetic simulation study

3.1. Synthetic scenario design

We now compare the methods introduced in Section 2 under a synthetic test scenario, designed to be simple enough to yield a valuable comparison, whilst being sufficiently complex to be representative of a real-world structure. We construct a synthetic response with an artificially bimodal CDE, intended to represent the extreme case of non-convex failure regions discussed in Section 1. We follow the approaches of the likes of [Gramstad et al. \(2020\)](#) and [Castellon et al. \(2023\)](#), who model structural responses (approximately) described by some parametric distribution function. The structural response R given long term environment $\mathbf{X}$ is modelled as a Weibull random variable, with distribution function

$$F_{R|\mathbf{X}}(r|\mathbf{x}) = 1 - \exp \left\{ - \left(\frac{r}{\eta(\mathbf{x})} \right)^k \right\}, \quad r > 0, \quad (26)$$

for fixed shape parameter $k = 2$ and scale parameter $\eta : \mathcal{E}_{\mathbf{X}} \mapsto \mathbb{R}$ dependent on the long term environment. Adoption of this conditional Weibull form allows straightforward sampling from $R|\{\mathbf{X} = \mathbf{x}\}$, as well as exact evaluation of the conditional failure probability $\mathbb{P}(R > r_{Cr}|\{\mathbf{X} = \mathbf{x}\})$. The environment $\mathbf{X}$ is assumed bivariate $\mathbf{X} = (X_1, X_2)$, with density function $f_{\mathbf{X}}(\mathbf{x}) = f_{(X_1, X_2)}(x_1, x_2) = f_{X_2|X_1}(x_2|x_1)f_{X_1}(x_1)$, where

$$f_{X_1}(x_1) = \frac{x_1}{\sigma_R^2} \exp\left(-\frac{x_1^2}{2\sigma_R^2}\right), \quad x_1 \geq 0 \quad (27)$$

is the Rayleigh density with scale parameter $\sigma_R > 0$ set to $\sigma_R = 12$ in this case, and

$$f_{X_2|X_1}(x_2|x_1) = \frac{1}{x_2\sigma_{LN}(x_1)\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left(\frac{\log(x_2) - \mu_{LN}(x_1)}{\sigma_{LN}(x_1)} \right)^2 \right\} \quad (28)$$

is the log-normal density with

$$\mu_{LN}(x_1) = 0.933 + 0.578x_1^{0.395}, \quad (29)$$

$$\sigma_{LN}(x_1) = 0.055 + 0.336 + \exp(-0.585x_1), \quad (30)$$

as in [Mathisen and Bitner-Gregersen \(1990\)](#) for wave period given significant wave height. The reader might therefore choose to consider this environmental specification as an example of an extreme environment of significant wave height (Rayleigh) and conditional significant wave period (log-normal), although this interpretation is not necessary.

The function η is constructed to provide a structural response with the desired multimodal behaviour. To achieve this, we define a scenario with scale $\eta(\mathbf{x})$ increasing around values $\mathbf{x}_{Pk1}$ and $\mathbf{x}_{Pk2}$, modelling the scale parameter using the multimodal function

$$\eta(\mathbf{x}) = C \{ A \max(\|\mathbf{x} - \mathbf{x}_{Pk1}\|, \nu) + B \max(\|\mathbf{x} - \mathbf{x}_{Pk2}\|, \nu) \}, \quad (31)$$

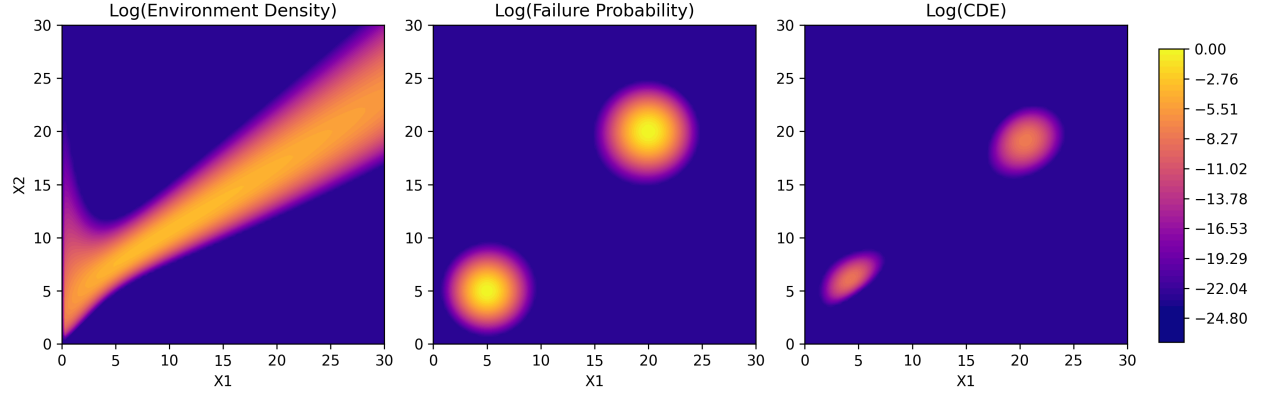


Figure 3: Panels summarising the synthetic response case study used in this section. From left to right the panels show: bivariate environment log-density (27); structural log-failure probability as Weibull exceedance probability of $r_{Cr} = 175$ (32); and log-CDE (6) obtained by multiplying failure probability by environment density.

for scaling parameters $A, B, C > 0$, and peak radius $\nu > 0$ surrounding ‘resonant’ $\mathbf{X}$ values $\mathbf{x}_{Pk1} = (5, 5)$ and $\mathbf{x}_{Pk2} = (20, 20)$. We set constants, $A = 1.3$, $B = 1.5$, $C = 100$, $\nu = 0.5$, and critical response $r_{Cr} = 175$, chosen in order to yield a true ‘synthetic’ failure probability

$$p_{Sn} = \mathbb{P}(R > r_{Cr}) = \int_{\mathcal{E}_{\mathbf{X}}} \exp \left\{ - \left(\frac{r_{Cr}}{\eta(\mathbf{x})} \right)^2 \right\} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = 1.3 \times 10^{-3}. \quad (32)$$

This yields a failure probability in order of magnitude comparable to the failure probabilities discussed in Section 4.4. Figure 3 shows the environment density, failure probability and CDE for this synthetic scenario, over the bivariate environment space $\mathcal{E}_{\mathbf{X}} = [0, 30]^2$.

In practice, estimates of conditional failure probability $\mathbb{P}(R > r_{Cr} | \{\mathbf{X} = \mathbf{x}\})$ are found empirically using realisations of $R | \{\mathbf{X} = \mathbf{x}\}$. We introduce further uncertainty in this synthetic case in the form of the conditional distribution for $R | \{\mathbf{X} = \mathbf{x}\}$ by making η stochastic, with

$$\eta_{\delta}(\mathbf{x}) = \eta(\mathbf{x}) \cdot (1 + \epsilon_{\delta}), \quad \text{for } \epsilon_{\delta} \sim N(0, \delta^2), \quad (33)$$

where we set $\delta = 0.05$. This applies an additive white noise to the scale of our observations with variance proportional to the value of the scale function, meaning that larger values of the scale function will correspond to ‘more uncertain’ observations. In the absence of uncertainty in η_{δ} (i.e., with $\delta = 0$), the expected distribution of $R | \{\mathbf{X} = \mathbf{x}\}$ is relatively easily identified from a smaller number of realisations of fluid loading simulation. However, for uncertain η_{δ} , the number of realisations required to be confident about the expected distribution of $R | \{\mathbf{X} = \mathbf{x}\}$ increases. That is, particularly with $\delta > 0$, we expect to need to sample from the same regions of $\mathcal{E}_{\mathbf{X}}$ multiple times to build confidence in our estimate of CDE.

Note that the IS-PT and AGE procedures are expected to give good performance for any reasonable combination of values for parameters σ_R , μ_{LN} , σ_{LN} , A , B , C , ν , x_{Pk1} and x_{Pk2} in this simulation study, and therefore that the actual values of parameters used here are of little direct relevance. The critical feature of the current simulation study is that the general

characteristics of the distribution of environmental variables and those of the structural response to the environment, reflect the general characteristics of actual environments and structural responses. It is for this reason that we refer to the environmental variables in this section as X_1 and X_2 rather than e.g., H_s and T_p .

3.2. Results of synthetic study

3.2.1. Overview

Here we apply the methods introduced in Section 2 to the synthetic scenario discussed above. We first present the results of the importance sampling-parallel tempering (IS-PT) approach of Section 2.2 in Section 3.2.2, followed by those of the adaptive Gaussian emulation (AGE) procedure of Section 2.3 in Section 3.2.3. We adjust the number of expensive function evaluations n_{Ev} used for each of IS-PT and AGE methods so they yield the same order of magnitude of root mean squared error

$$\text{RMSE}(\hat{p}) = \sqrt{\sum_{r=1}^{n_{\text{Rp}}} \frac{(\hat{p}_r - p_{\text{Sn}})^2}{n_{\text{Rp}}}}, \quad (34)$$

over some number n_{Rp} of replicate analyses, where $\hat{p}_r$ is the estimate provided by either IS-PT or AGE at replicate r . We also evaluate the bias

$$\text{Bias}(\hat{p}) = \sum_{r=1}^{n_{\text{Rp}}} \frac{(\hat{p}_r - p_{\text{Sn}})}{n_{\text{Rp}}}, \quad (35)$$

for each of the methods.

3.2.2. Importance sampling coupled with parallel tempering MCMC (IS-PT)

We apply the IS-PT framework of Section 2.2 under the synthetic scenario in three stages: first (a) parallel tempering MCMC sampling with the CDE (6) as target posterior density; followed by (b) kernel smoothing of the resulting sample to obtain proposal density p_{Pr} ; and finally (c) evaluation of importance sampling estimate (8) using $n_{\text{IS}} = 100$ draws from this proposal. Steps (a)-(c) are repeated $n_{\text{Rp}} = 100$ times, to estimate $\hat{p}_{\text{IS}}$.

Step (a) is achieved using the adaptive parallel tempering algorithm of Vousden et al. (2015) implemented in the `pyPESTO` module. We run $n_{\text{Tm}} = 5$ parallel chains, supplying an initial temperature ladder $T_1, \dots, T_5$ geometrically spaced between $T_1 = 1$ and $T_5 = 20$, with initial proposal variance $\sigma_{\text{MH}}^2 = 1$. The MCMC algorithm then adaptively tunes the temperature spacing and proposal variance, targetting equal acceptance probability of swaps between adjacent chains. Each of the five chains is run for $n_{\text{PT}} = 400$ time steps, with periodic swaps between chains proposed according to Vousden et al. (2015), requiring $n_{\text{Tm}} \times n_{\text{PT}} = 2000$ expensive function evaluations in total. An example trace plot from the T_1 chain is given in SM3.1. The chain at temperature $T_1 = 1$ is retained, and burn-in length n_{Br} automatically chosen using Geweke's diagnostic (Geweke, 1991). When $n_{\text{Br}} < n_{\text{PT}}$, this burn-in period

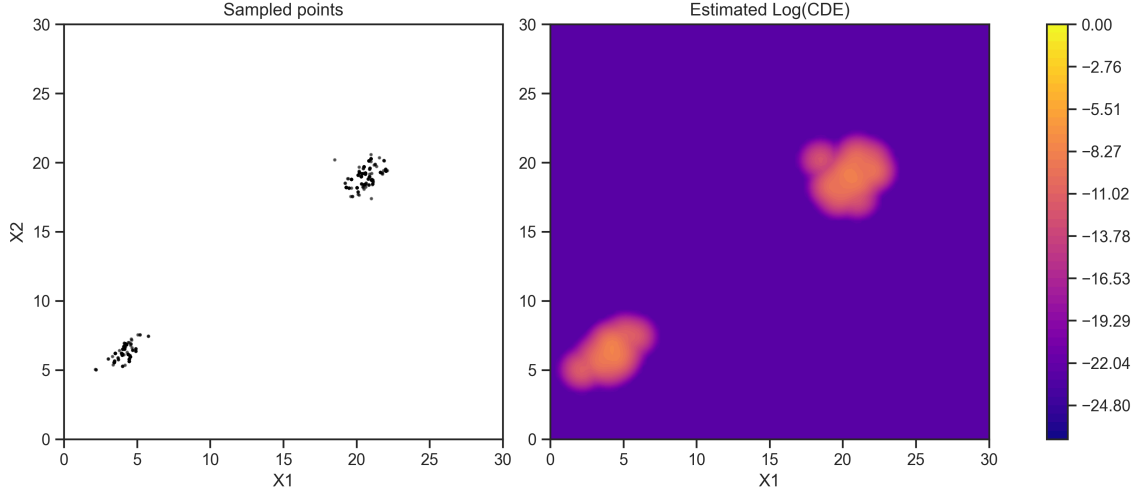


Figure 4: Example sample from CDE (6) under the synthetic structural scenario obtained using the adaptive parallel tempering MCMC algorithm of Vousden et al. (2015) (left), and corresponding smoothed log-CDE estimated using Gaussian kernel bandwidth selected according to Scott (2015) (right).

is discarded, leaving a sample of length $n_{PT} - n_{Br}$. For step (b), the sample is then used to provide a Gaussian kernel smoothed estimate of the CDE, with kernel bandwidth chosen according to Scott’s rule of thumb (Scott, 2015), see SM3.1 for details. Step (c) consists of evaluating importance sampling probability estimate $\hat{p}_{IS}$ given by (8), using $n_{IS} = 100$ draws from proposal density g_{Pr} found in step (b), requiring a further 100 expensive function evaluations. Figure 4 shows a typical sample obtained using this approach together with resulting CDE estimate g_{Pr} . The (root mean square error) RMSE (34) is estimated to be $\text{RMSE}(\hat{p}_{IS}) = 2.20 \times 10^{-4}$, using $n_{Rp} = 100$ replicates of the IS-PT analysis, with *each* of the n_{Rp} IS-PT estimates requiring $n_{Ev} = n_{Tm} \times n_{PT} + n_{IS} = 2100$ expensive function evaluations. The bias in the $\hat{p}_{IS}$ estimate over the 100 replicates is small, equal to $\text{Bias}(\hat{p}_{IS}) = 5.32 \times 10^{-5}$.

3.2.3. Adaptive Gaussian emulation (AGE)

The GP emulator (10) is used to model the log-CDE under this synthetic scenario, following the AGE procedure of Section 2.3.2. It is iteratively trained as in (15) on training sets $\mathcal{D}_1, \dots, \mathcal{D}_{n_{It1}}$ for n_{It1} iterations, with training set $\mathcal{D}_{n+1} = \{\mathcal{D}_n, \mathbf{x}^*\}$, $n > 1$, constructed with $\mathbf{x}^*$ chosen according to either $U^{(1)}$ (21, Variance case) or $U^{(2)}$ (25, ALC case). In each case, the initial training set $\mathcal{D}_0$ is a simple space-filling Latin hypercube design of $n_{Tr1} = 144$ points, chosen as a low, but adequate, number of starting points found to provide stable kernel parameter convergence at iteration zero. At each subsequent iteration, we begin the kernel parameter optimisation at the previous iteration’s estimates.

Figure 5 shows an example of how utility $U^{(1)}$ is constructed using the emulator (10) trained on initial set $\mathcal{D}_0$, for two example values of λ . The upper panels shows the exploration Σ_0 and exploitation M_0 terms as defined in (21), with lower panels showing utility functions obtained by prioritising exploration ($\lambda = 0.7$) or exploitation ($\lambda = 0.3$). Green points in the lower panels indicate the maximum $\mathbf{x}^* = \arg\max_{\mathbf{x} \in \mathcal{E}_X} U^{(1)}(\mathbf{x}; \lambda)$, illustrating that the choice

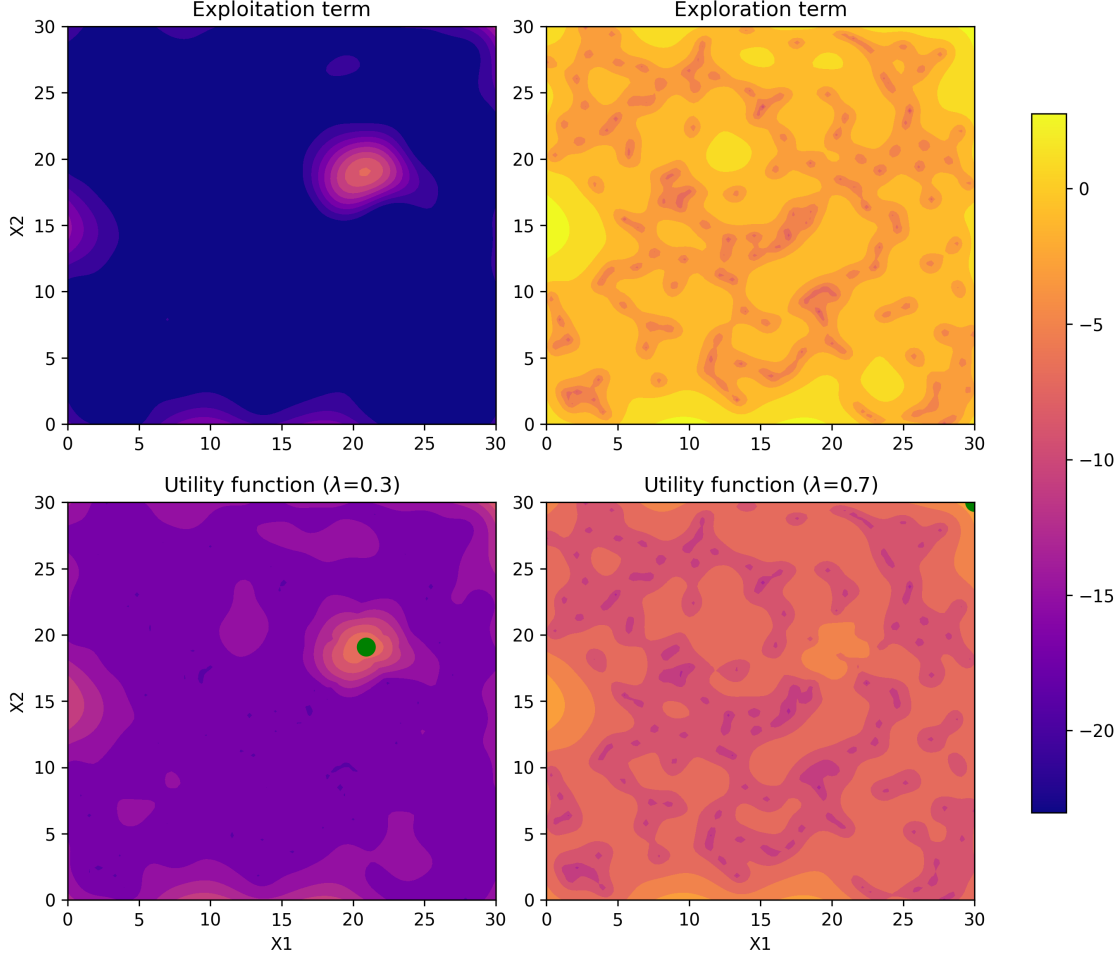


Figure 5: Behaviour of utility function $U^{(1)}(\mathbf{x}; \lambda)$ over the environment space $\mathcal{E}_{\mathbf{x}}$, for synthetic scenario. Upper panels show exploitation and exploration terms obtained from GP emulator (10) trained on initial Latin hypercube set $\mathcal{D}_0$ of size $n_{\text{Tr}1} = 144$. Lower panels show resulting utility functions for weights $\lambda = 0.3$ and $\lambda = 0.7$. In each lower panel, the optimal sampling point $\mathbf{x}^* = \operatorname{argmax}_{\mathbf{x} \in \mathcal{E}_{\mathbf{x}}} U^{(1)}(\mathbf{x}; \lambda)$ is indicated in green. In the lower right hand panel, $\mathbf{x}^*$ is located in the upper right corner of $\mathcal{E}_{\mathbf{x}}$.

of this tuning parameter can alter the design of the training set $\mathcal{D}_1 = \{\mathcal{D}_0, \mathbf{x}^*\}$ (and thus subsequent training sets $\mathcal{D}_2, \mathcal{D}_3, \dots$). In the lower right panel, the maximum is located on the edge of the environment space, due to the Latin hypercube sampling used to construct $\mathcal{D}_0$ placing no points on the boundary. (This can be prevented by adding initial training points along the boundary, however, this isn't necessary as subsequent iterations move away from the edge of the space once it has been explored.) See SM3.2 for an equivalent example for $U^{(2)}$.

The GP emulator is trained using both utility functions $U^{(1)}$ and $U^{(2)}$, for a range of $n_{\lambda} = 50$ values of weight parameter λ , equally spaced on the interval $[0.01, 0.99]$. For each value of λ , we perform $n_{\text{It}1} = 100$ iterations of the GP update (15) for each utility. This yields posterior estimates μ_n^* and k_n^* for $n = 1, \dots, n_{\text{It}1}$. This analysis is replicated $n_{\text{Rp}} = 100$ times, with randomised initial set $\mathcal{D}_0$ and conditional response scale η (33) at each replicate.

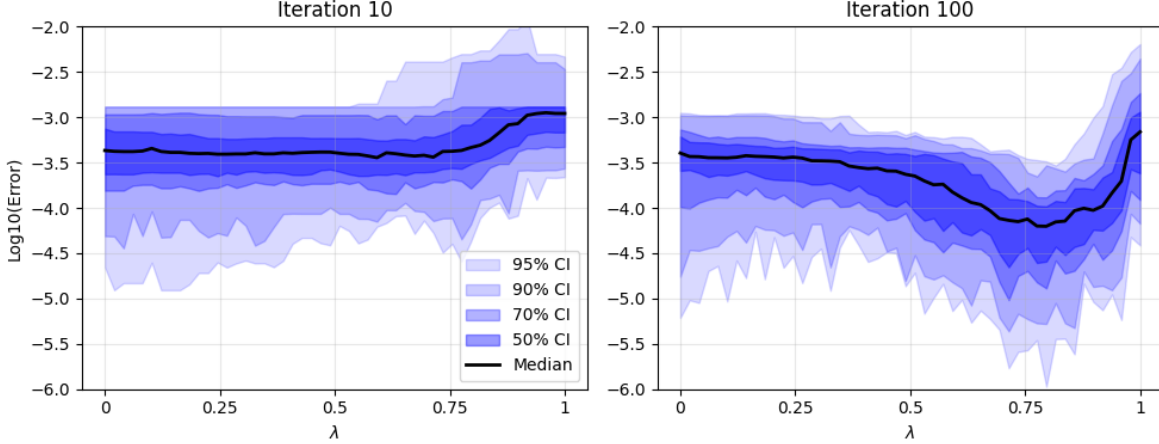


Figure 6: Log-scale absolute error Δ_{GP} of the GP probability estimate $\hat{p}_{\text{GP}}$ at specified iterations, for emulator (10) trained using $U^{(1)}$ over the range of weight $\lambda \in [0.01, 0.99]$ for the synthetic scenario. At iteration 100, the weight that minimises median error is $\lambda^* = 0.80$.

Each of the n_{Rp} replicate analyses produces $n_{\lambda} \times n_{\text{It1}}$ values of the failure probability estimate $\hat{p}_{\text{GP}}$ and error Δ_{GP} for each utility. Figure 6 shows the distribution of the resulting Δ_{GP} values under variance utility $U^{(1)}$ (21) with respect to λ , at iterations 10 and 100. Errors are plotted on the log scale, with 50%, 70%, 90% and 95% confidence bands indicated in different shades of blue. The median log-error trend with respect to λ is given as a black line. The GP emulator converges to the truth for weights in the interval I^* , which in this case corresponds approximately to $[0.5, 1]$. The location of I^* on the unit interval is determined by the bimodal nature of the synthetic response. For some initial training sets $\mathcal{D}_0$, at iteration zero, the emulator detects one peak in response but fails to detect the other; this can be seen in the top left panel of Figure 5, where the mode at $\mathbf{x}_{\text{Pk2}} = (20, 20)$ is found, but that at $\mathbf{x}_{\text{Pk1}} = (5, 5)$ is not. For low values of λ , the utility function $U^{(1)}$ sometimes does not place enough weight on the exploration term for the algorithm to detect the second peak in subsequent iterations (e.g., the lower left panel of Figure 5 shows a low value for utility at $\mathbf{x}_{\text{Pk1}}$, whereas the lower right panel has a higher utility there). That is, for low values of λ , the iterative algorithm tends not to allow the GP to ‘discover’ the second mode. The value of λ minimising the median error at the final iteration is $\lambda^* = 0.80$. Figure 7 shows the distribution of Δ_{GP} across all iterations when $\lambda = \lambda^*$. In general, there is a decrease in error with iteration, with ‘spike’ at around iteration 10 for some replicates; these spikes shows where the algorithm tends to detect the second mode, causing a temporary increase in bias due to the uncertainty in (33). Figures corresponding to Figures 6 and 7 for ALC utility $U^{(2)}$ can be found in SM3.2. For $U^{(2)}$, a minimum of Δ_{GP} is found in $I^* = [0.2, 0.5]$, and comparison of errors Δ_{GP} at the final iteration indicates that Δ_{GP} for $U^{(2)}$ is somewhat larger than for $U^{(1)}$.

Figure 8 shows an example GP emulator at the final iteration, trained on set $\mathcal{D}_{100}$ selected using $U^{(1)}$ with $\lambda^* = 0.80$. The left panel shows the posterior GP mean $\mu_{100}^*(\mathbf{x})$ and the right the posterior GP standard deviation $k_{100}^*(\mathbf{x}, \mathbf{x})^{1/2}$, both over $\mathbf{x} \in \mathcal{E}_{\mathbf{x}}$. The initial Latin hypercube training set $\mathcal{D}_0$ is shown as dark green crosses, and the iteratively selected new

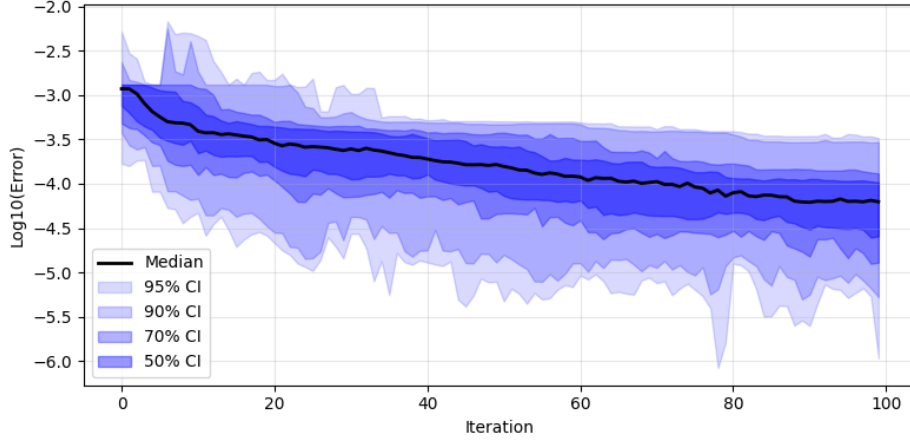


Figure 7: Distribution of log-scale absolute error Δ_{GP} in the GP probability estimate with respect to iteration, trained using $U^{(1)}$ with $\lambda = \lambda^*$. The trend in median error is indicated in black, with various confidence intervals shown in blue.

training points $\mathcal{D}_{100} \setminus \mathcal{D}_0$ are shown as light green crosses. The light green crosses, iteratively selected using the utility function, mostly cluster around the high-density regions of the synthetic CDE, whilst allowing some exploration into low-density regions of $\mathcal{E}_{\mathbf{x}}$.

We evaluate the RMSE (34) of the AGE approach using $\hat{p}_{\text{GP}}$ obtained from emulators trained under $U^{(1)}$ with $\lambda = \lambda^*$ over $n_{\text{Rp}} = 100$ replicate analysis. The resulting estimates yield $\text{RMSE}(\hat{p}_{\text{GP}}) = 1.16 \times 10^{-4}$, a similar value to $\text{RMSE}(\hat{p}_{\text{IS}})$ reported in Section 3.2.2. For the AGE approach, each replicate analysis involves a total of $n_{\text{Ev}} = |\mathcal{D}_0| + n_{\text{Itr1}} = 244$ expensive function evaluations, assuming λ^* is known. The bias in the $\hat{p}_{\text{GP}}$ estimate over the 100 replicates is $\text{Bias}(\hat{p}_{\text{GP}}) = 3.18 \times 10^{-5}$, comparable in size to that of $\hat{p}_{\text{IS}}$. Corresponding results using $U^{(2)}$ are similar, and summarised in the next section.

3.2.4. Comparison of IS-PT and AGE results

Figure 9 shows the distribution of the IS-PT estimate $\hat{p}_{\text{IS}}$ and the AGE estimates $\hat{p}_{\text{GP}}$ (from $U^{(1)}$ and $U^{(2)}$ at iteration 100, $\lambda = \lambda^*$) around the target failure probability p_{Sn} . A summary of the RMSEs and biases for these estimates can be seen in Table 1, along with the number of expensive function evaluations n_{Ev} required for each replicate analysis. Both variants of the $\hat{p}_{\text{GP}}$ estimate show an equivalent performance to $\hat{p}_{\text{IS}}$ for around 12% of required expensive function evaluations, provided we have knowledge of the optimal weight parameter λ^* . The AGE approach with utility $U^{(2)}$ is computationally somewhat more demanding than that using $U^{(1)}$, due to the required calculation of ALC (23) at each iteration.

However, if λ^* is unknown, and cannot be reliability estimated, we see that IS-PT provides a useful if computationally more demanding alternative. The current analysis shows that approximately 2000 expensive function iterations using IS-PT are sufficient to estimate a bimodal CDE well in two dimensions, avoiding the need to specify problematic hyperparameters such as λ . We explore the relative merits of IS-PT and AGE methodologies further for the monopile structure scenario of Section 4.

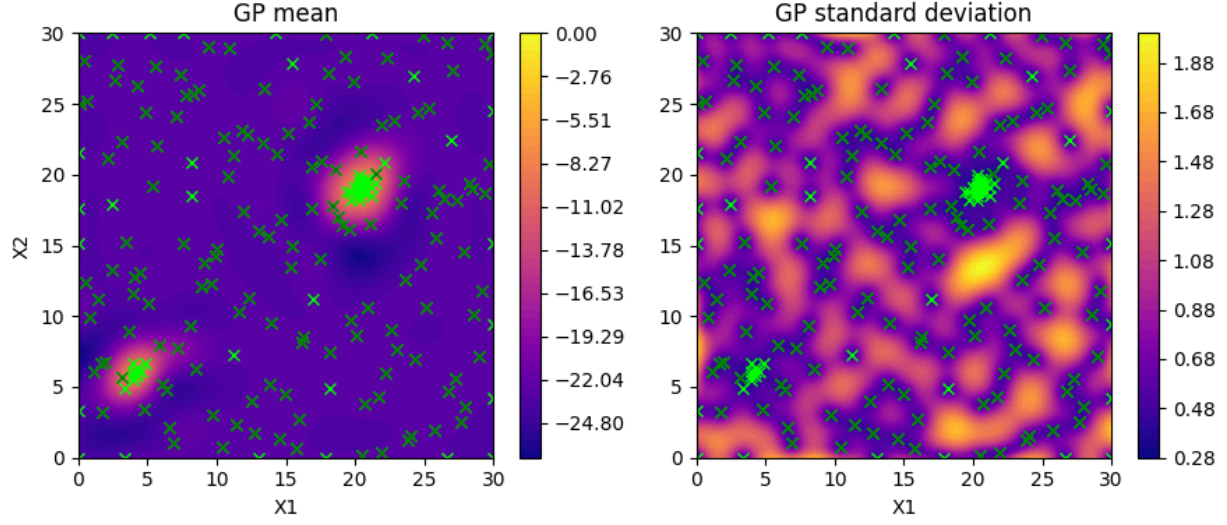


Figure 8: GP emulator at iteration 100 for variance utility $U^{(1)}$ (21), trained using the optimal value $\lambda^* = 0.80$ minimising median of error Δ_{GP} . The panels from left to right show: the posterior GP mean $\mu_{100}^*(\mathbf{x})$ over $\mathbf{x} \in \mathcal{E}_{\mathbf{X}}$; the posterior GP standard deviation $k_{100}^*(\mathbf{x}, \mathbf{x})^{1/2}$. The initial random Latin hypercube training set $\mathcal{D}_0$ is shown as dark green crosses, and the iteratively selected new training points $\mathcal{D}_{100} \setminus \mathcal{D}_0$ are shown as light green crosses.

	IS-PT	$U^{(1)}$ AGE	$U^{(2)}$ AGE
RMSE	2.20×10^{-4}	1.16×10^{-4}	2.40×10^{-4}
Bias	5.32×10^{-5}	3.18×10^{-5}	1.50×10^{-4}
Number of function evaluations, n_{Ev}	$n_{\text{Tm}} \times n_{\text{PT}} + n_{\text{IS}} = 2100$	$ \mathcal{D}_0 + n_{\text{Itr1}} = 244$	$ \mathcal{D}_0 + n_{\text{Itr1}} = 244$

Table 1: RMSEs and biases of $\hat{p}_{\text{IS}}$ and $\hat{p}_{\text{GP}}$ when targetting failure probability p_{Sn} , calculated for $n_{\text{Rp}} = 100$ replicate analyses. The true value of probability of failure is $p_{\text{Sn}} = 1.3 \times 10^{-3}$. Also shown is the number of expensive response function evaluations required for a single replicate analysis for each of IS-PT and AGE.

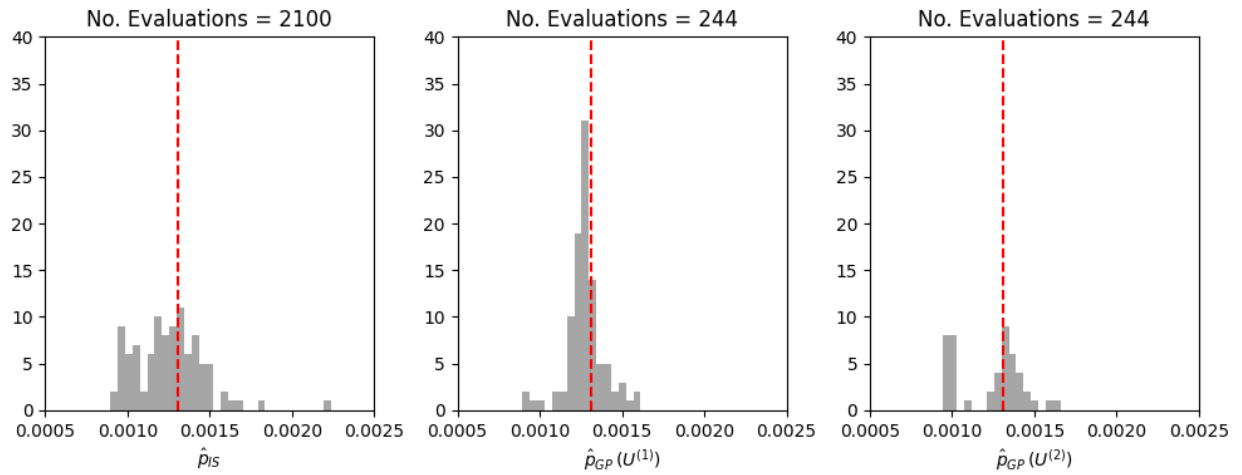


Figure 9: Distribution of $n_{\text{Rp}} = 100$ estimates $\hat{p}_{\text{IS}}$ (left), $\hat{p}_{\text{GP}}$ for $U^{(1)}$ iteration 100 with $\lambda = \lambda^*$ (centre) and $\hat{p}_{\text{GP}}$ for $U^{(2)}$ iteration 100 with $\lambda = \lambda^*$ (right), for true failure probability p_{Sn} (red). The number of function evaluations required for a single replicate analysis is indicated in the panel titles.

4. Application to monopile response models

4.1. Overview of case study

We now apply the IS-PT and AGE methodologies of Section 2 to a real-world case study, using hindcast data from a location around 1km offshore of Albany, Western Australia, produced by the Centre for Australian Weather and Climate Research, see Section 4.2 to estimate a model for the extreme ocean environment. We consider a model monopile structure situated in this environment, subject to wave-induced loading which in turn induces some resonant effect. To construct the test scenario, we first use the extreme value methods of Davison and Smith (1990) and Heffernan and Tawn (2004) to model the joint behaviour of a bivariate ocean $\mathbf{X}$ at this location, see Section 4.3. This is followed in Section 4.4 with numeric simulation from the T-FNV model of Taylor et al. (2024) to approximate the inertial load placed on an offshore wind turbine at this location. Finally, this load is propagated through the linear response function for a damped harmonic oscillator, yielding realisations of the harmonic response on our model structure. The results of these simulations are given in Section 4.5 to provide a ‘baseline’ estimate of the CDE. This baseline is then used to assess the performance of IS-PT and AGE methodologies in Sections 4.6 and 4.7. These methods are then compared in Section 4.8.

The reader is referred to articles including Riise et al. (2018), He and Zhu (2019), Wan et al. (2023), Liu et al. (2025) and Yang et al. (2025), as well as Orszaghova et al. (2025), for relevant discussion and illustration of wind turbine response characteristics.

4.2. Albany hindcast data

The data includes hourly hindcast observations over the period 1980-2017, consisting of sea state variables significant wave height H_s , peak wave period T_p , energy wave period T_e , and mean wave period T_m . There are a total of 333120 observations. We preprocess the data by isolating storm peak values of the sea state H_s . Given storm events that are sufficiently well spaced in time, this removes any temporal correlation in the storm peak data, simplifying the modelling process whilst retaining the observations most likely to induce structural failure.

To isolate the storms peaks, we follow the procedure of Ewans and Jonathan (2008). Firstly, a wave height h_{st} (in metres) is chosen as the storm threshold, such that an upcrossing above this height is considered the beginning of a storm event. The subsequent downcrossing of this height is considered the end of the storm event. We also merge any two storm events that occur within 48 hours of one another, retaining only the largest storm peak value. The value of h_{st} is determined by assessing the number of storm peaks recovered from the dataset using a given threshold against the practicality of observed storm lengths; this creates a trade-off between retaining enough points for statistical modelling, whilst avoiding identifying storms of unrealistically long duration. We choose to limit occurrences of storms lasting longer than three days, selecting $h_{st} = 4$ yielding a total of 976 storm peak observations, with around 6% of identified storm durations exceeding three days.

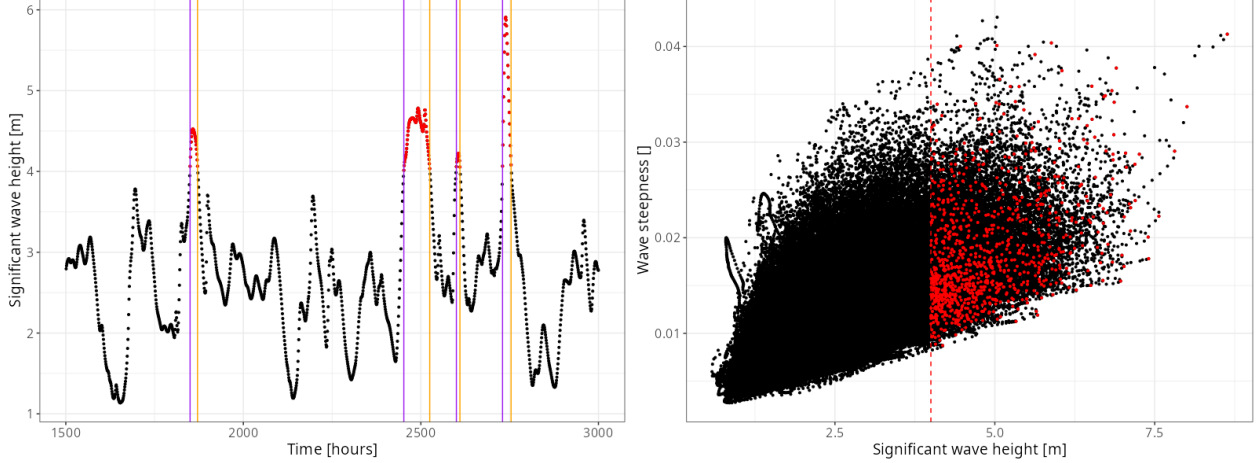


Figure 10: Illustration of storm peak isolation for storm threshold $h_{St} = 4$. The left panel shows H_s value against hourly index, with the beginning of each storm (defined as the first upcrossing of 4m) indicated in purple. The end of each storm (defined as the first downcrossing of 4m) is shown in orange. Sequences of within-storm H_s values are highlighted in red. In the right panel, the entire hindcast data of S_e against H_s are shown in black, with chosen storm peak values indicated in red. The storm threshold $h_{St} = 4$ is indicated as a dashed red vertical line.

Figure 10 illustrates this process, with the left panel showing identified storms. Given selection of storm peak H_s values, these can be matched with the corresponding T_p , T_e or T_m values to obtain a joint storm peak environment. We focus our attention on storm peak significant wave height H_s and (significant) wave steepness

$$S_e = \frac{2\pi H_s}{gT_e^2}, \quad (36)$$

for gravitational acceleration $g = 9.81\text{ms}^{-2}$, modelling the 2-dimensional environment (H_s, S_e) . We choose to model S_e over T_e (or other wave period variables) because the most extreme sea states tend to be the steepest. Using S_e , our interest therefore lies in characterising the pair of positively valued variables (H_s, S_e) , when at least one of the pair is very large, an appropriate setting for application of the conditional extremes method of [Heffernan and Tawn \(2004\)](#). Going forward, we let $\mathbf{X} = (H_s, S_e)$, referring to the joint storm peak values seen in the right panel of Figure 10, rather than the original hourly data.

4.3. Joint storm peak variable modelling

4.3.1. Outline of long term environment model

We describe the joint behaviour a long term environment, using the conditional extreme value model of [Heffernan and Tawn \(2004\)](#). This approach facilitates modelling of the joint extremes of $\mathbf{X}$, facilitating the extrapolation of joint behaviour beyond the range of the sample data. This asymptotically justified framework has been widely applied in capturing the tail dependence of environmental data due to its flexibility in capturing different extremal dependence types and its ease of use (e.g., [Jonathan et al. 2014](#); [Towe et al. 2019](#); [Shooter](#)

et al. 2021; Tendijck et al. 2023). It is also simple to extend this model to account for seasonality or long term trends in an environment through the addition of covariates (e.g., Ewans and Jonathan 2008), however, we omit the inclusion of covariate effects as this is not the focus of this work. Furthermore, whilst our synthetic environment and example data are of dimension $d = 2$, we present the conditional extremes method in the general case of $d > 1$.

The conditional extremes method consists of a two stage modelling process: first, the transformation of environment variable $\mathbf{X}$ to a standard marginal scale (typically Laplace); and second, modelling of the joint structure of the standardised variables. The first step is achieved using univariate extreme value techniques (e.g., Davison and Smith 1990), and the second via a series of d pairwise non-linear regressions. Details of these steps are discussed below.

4.3.2. Marginal modelling and transformation

We use the peaks over threshold method of Davison and Smith (1990) when modelling the marginal distributions of $X_1, \dots, X_d$. That is, for X_j , $j = 1, \dots, d$, we fit a generalised Pareto distribution (GPD) to sample exceedances of X_j above some high threshold u_j , modelling non-exceedances empirically. The full model for the marginal distribution F_{X_j} of X_j is

$$F_{X_j}(x) = \begin{cases} \tilde{F}_{X_j}(x) & x \leq u_j \\ \tilde{F}_{X_j}(u_j) + \{1 - \tilde{F}_{X_j}(u_j)\}F_{\text{GPD},j}(x; u_j, \sigma_j, \xi_j) & x > u_j, \end{cases} \quad (37)$$

for empirical distribution $\tilde{F}_{X_j}$ of X_j , and GPD distribution function

$$F_{\text{GPD},j}(x; u_j, \sigma_j, \xi_j) = 1 - \left(1 + \frac{\xi_j(x - u_j)}{\sigma_j}\right)_+^{-1/\xi_j}, \quad x > u_j, \quad (38)$$

for scale and shape parameters $\sigma_j > 0$ and $\xi_j \in \mathbb{R}$, with $y_+ = \max(y, 0)$ for $y \in \mathbb{R}$. The conditioning thresholds u_j , $j = 1, \dots, d$, are chosen so the asymptotic behaviour justifying the use of the GPD tail distribution holds approximately. Appropriate values of these thresholds are typically selected by either manually examining the stability of σ_j and ξ_j when fitting to exceedances above candidate values for u_j , or using automated methods such as those of Varty et al. (2021) and Murphy et al. (2025). We find parameter stability tests to be satisfactory for our data, see SM4.1. Given a choice of u_j , parameters σ_j and ξ_j are found using maximum likelihood techniques.

The marginal model F_{X_j} is used to map X_j onto X'_j with Laplace margins, via the probability integral transform

$$X'_j = \begin{cases} \log \left\{ 2F_{X'_j}(X'_j) \right\} & X'_j \leq F_{X'_j}^{-1}(0.5) \\ -\log \left\{ 2 \left[1 - F_{X'_j}(X'_j) \right] \right\} & X'_j > F_{X'_j}^{-1}(0.5), \end{cases} \quad (39)$$

for $j = 1, \dots, d$, obtaining the multivariate Laplace-scale environment variable $\mathbf{X}' = (X'_1, \dots, X'_d)$.

4.3.3. Joint dependence modelling

We now apply the conditional extremes framework to the Laplace-scale environment variable $\mathbf{X}' \in \mathbb{R}^d$. This requires specifying a conditioning environmental variable $X'_j \in \mathbb{R}$, followed by modelling the remaining variables $\mathbf{X}'_{-j} \in \mathbb{R}^{d-1}$, conditional on the event $X'_j > v_j$ for $v_j > 0$, $j = 1, \dots, d$. Fitting this model allows simulation of new multivariate events with extremal dependence structure representative of the original process $\mathbf{X}'$, facilitating estimation of joint extreme event set probabilities.

Broadly following [Keef et al. \(2013\)](#), [Heffernan and Tawn \(2004\)](#) assume that, for $j = 1, \dots, d$, there exist unique values $\alpha_{|j} \in [-1, 1]^{d-1}$, $\beta_{|j} \in (-\infty, 1]^{d-1}$ and $\mathbf{z}_{|j} \in \mathbb{R}^{d-1}$, such that

$$\lim_{v_j \rightarrow \infty} \mathbb{P} \left(\frac{\mathbf{X}'_{-j} - \alpha_{|j} X'_j}{X_j^{\beta_{|j}}} < \mathbf{z}_{|j}, X'_j - v_j > x | X'_j > v_j \right) = e^{-x} G_{|j}(\mathbf{z}_{|j}), \quad (40)$$

for $x > 0$ and distribution function $G_{|j} : \mathbb{R}^{d-1} \mapsto \mathbb{R}$ with non-degenerate marginals, where componentwise operations are assumed. In practice, the limit (40) is assumed to hold for some suitably large finite threshold v_j , yielding the regression

$$\mathbf{X}'_{-j} | \{X'_j = x\} = \alpha_{|j} x + x^{\beta_{|j}} \mathbf{Z}_{|j}, \quad (41)$$

for $x > v_j$, and residual random variable $\mathbf{Z}_{|j}$ independent of X'_j given $X'_j > v_j$, where elementwise operations are assumed. Regression (41) is then used to model all data in the region $\{\mathbf{X}'_j \in \mathbb{R}^d : X'_j > v_j\}$, and parameters $\alpha_{|j}$ and $\beta_{|j}$ are estimated using standard maximum likelihood estimation (MLE) techniques. For this estimation we utilise the additional parameter constraints of [Keef et al. \(2013\)](#) ensuring consistency of conditional return level values between extremal dependence types. For model fitting only, it is assumed that $G_{|j}$ corresponds to independent Gaussian distributions with unknown means and variances. Once parameter estimates have been obtained, we follow [Winter and Tawn \(2017\)](#) and model $G_{|j}$ using the Gaussian kernel smoothed density estimate of the observed values of residual

$$\mathbf{Z}_{|j} = \frac{\mathbf{X}'_{-j} - \alpha_{|j} X'_j}{X_j^{\beta_{|j}}}, \quad (42)$$

for $X'_j > v_j$, smoothed using kernel bandwidth $\delta_{\text{HT}} > 0$. The conditioning threshold v_j is chosen by studying parameter stability above candidate values, see SM4.1. The selection of δ_{HT} is considered in SM4.2. This model is fitted for all choices of the conditioning variable $\mathbf{X}'_j$, allowing simulation of $\mathbf{X}'$ in each of the corresponding regions $\{\mathbf{X}' \in \mathbb{R}^d : X'_j > v_j\}$ as described by [Heffernan and Tawn \(2004\)](#).

The environment density $f_{\mathbf{X}}$ is then estimated as in [Speers et al. \(2024\)](#), using prediction from fitted models in each of the the upper tail regions $\mathcal{E}_{\mathbf{X}}^{(j)} = \{\mathbf{X}' \in \mathbb{R}^d : X'_j > v_j\}$, $j = 1, \dots, d$, followed by empirical estimation in the remaining lower region $\mathcal{E}_{\mathbf{X}}^{\text{Lw}} = \{\mathbf{X}' \in \mathbb{R}^d : X_j \leq v_j \forall j\}$. In a given upper region $\mathcal{E}_{\mathbf{X}}^{(j)}$, we make Laplace-scale simulations from the joint dependence model (40) (using the parameter estimates found via MLE), followed by marginal transformation back to the physical scale using the inverse of transformation

(39). During simulation in $\mathcal{E}_{\mathbf{X}}^{(j)}$, we reject realisations for which $\max_{j':j' \neq j} X'_{j'} > X'_j$. This simulation yields a set of n_{Sm} realisations of $\mathbf{X}$ within $\mathcal{E}_{\mathbf{X}}^{(j)}$, from which we may empirically estimate the probability density $f_{\mathbf{X}}$ over a gridded set of subregions of $\mathcal{E}_{\mathbf{X}}^{(j)}$. Specifically, for a set D of feasible values of $\mathbf{X}$ such that $\mathbb{P}(\mathbf{X} \in \mathcal{E}_{\mathbf{X}} \setminus D) \approx 0$, we partition D using grid $(D_1, \dots, D_{n_{\text{Gr}}})$. We then assume that each $|D_i|$, $i = 1, \dots, n_{\text{Gr}}$, is small enough for the approximation

$$\mathbb{P}(\mathbf{X} \in D_i) = \int_{\mathbf{s} \in D_i} f_{\mathbf{X}}(\mathbf{s}) d\mathbf{s} \approx |D_i| f_{\mathbf{X}}(\mathbf{x}), \quad (43)$$

to be suitable, assuming that $f_{\mathbf{X}}$ is reasonably constant for all $\mathbf{x} \in D_i$. For any D_i within an upper tail region $\mathcal{E}_{\mathbf{X}}^{(j)}$, $j = 1, \dots, d$, we estimate the joint density with

$$\hat{f}_{\mathbf{X}}^{(j)}(\mathbf{x}) = \frac{n_{\text{Sm}}^{(i)}}{n_{\text{Sm}} |D_i|}, \quad \mathbf{x} \in D_i \subset \mathcal{E}_{\mathbf{X}}^{(j)}, \quad (44)$$

where $n_{\text{Sm}}^{(i)}$ is the number of simulated values of $\mathbf{x}$ in D_i . Combining these estimates with the empirical density $\tilde{f}_{\mathbf{X}}$ used in the lower region $\mathcal{E}_{\mathbf{X}}^{\text{Lw}}$, the full density estimate for $\mathbf{x} \in \mathcal{E}_{\mathbf{X}}$ is

$$\hat{f}_{\mathbf{X}}(\mathbf{x}) = \begin{cases} \tilde{f}_{\mathbf{X}}(\mathbf{x}) & \mathbf{x} \in \mathcal{E}_{\mathbf{X}}^{\text{Lw}} \cap D, \\ \hat{f}_{\mathbf{X}}^{(j)}(\mathbf{x}) & \mathbf{x} \in \mathcal{E}_{\mathbf{X}}^{(j)} \cap D, \quad j = 1, \dots, d, \\ 0 & \mathbf{x} \notin D. \end{cases} \quad (45)$$

4.3.4. Estimate of the environment density

Figure 11 shows the resulting estimate (45) of the environment density $f_{\mathbf{X}}$, found using the conditional extremes model. We take marginal thresholds $u_1 = \tilde{F}_{H_s}^{-1}(0.7)$ and $u_2 = \tilde{F}_{S_e}^{-1}(0.7)$, where $\tilde{F}_{H_s}$ and $\tilde{F}_{S_e}$ are the empirical distribution functions of H_s and S_e . The conditioning threshold for H_s is chosen as $v = \tilde{F}_{H_s}^{-1}(0.6)$. When simulating joint values of $\mathbf{X}$ conditional on $H_s > v$, we take $\delta_{\text{HT}} = 0.4$, $n_{\text{Sm}} = 10^5$, $D = [3, 12] \times [0.01, 0.05]$ and $n_{\text{Gr}} = 90 \times 45$; these values yield $|D_i|$ small enough for approximation (43) to be reasonable, for negligible computational cost when estimating density $f_{\mathbf{X}}$. We choose not to fit the conditional extremes model to the region where S_e is large since in typical offshore applications, large values of H_s rather than S_e dominate structural failure. We model the density empirically for values of H_s below the conditioning threshold v , and apply a Gaussian kernel smoother with bandwidth chosen according to Scott (2015).

4.4. Non-linear harmonic structural response simulation

4.4.1. Overview of response simulation

We now obtain empirical distributions of the structural response $R|\{\mathbf{X} = \mathbf{x}\}$, $\mathbf{X} = (H_s, S_e)$, in our model monopile scenario, given a fixed environment $\mathbf{x} \in D$. This is achieved using realisations $R_j|\{\mathbf{X} = \mathbf{x}\}$, $j = 1, \dots, n_{\text{RI}}$, of the structural response, obtained via repeated direct simulation using physical models of environmental loading on the monopile. For each of n_{RI} realisations of fluid loading, this requires: (a) simulation of hour-long time series

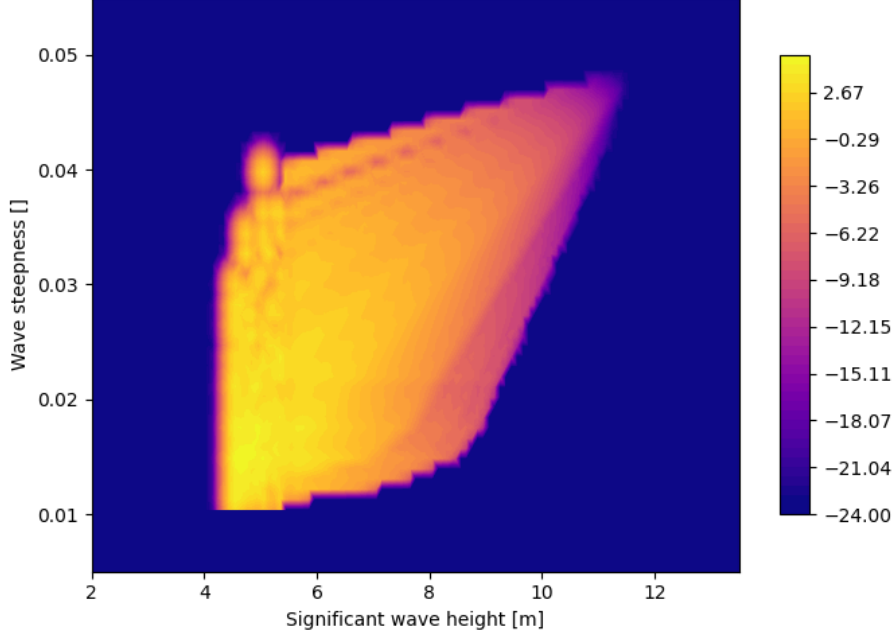


Figure 11: Estimate of the joint density $f_{\mathbf{X}}$ of environment variable $\mathbf{X} = (H_S, S_e)$, using the conditional extremes model of [Heffernan and Tawn \(2004\)](#) fitted to storm peak data from a location 1km offshore of Albany.

realisations of the stochastic linear wave elevation $\{E_j(t; \mathbf{x}) : t \in [0, 60^2]\}$ for an underlying environment $\mathbf{x}$; (b) conversion of the surface elevation $E_j(t; \mathbf{x})$ to load $L_j(t; \mathbf{x})$ induced on a monopile; (c) transformation of this load via a linear response function to resonant response $R_j(t; \mathbf{x})$ time series observed on the model structure; and (d) isolation of maximum response $R_j|\{\mathbf{X} = \mathbf{x}\}$ from the time series $R_j(t; \mathbf{x})$. Note that all physical quantities are given in SI units throughout this section. Further, $R_j(t; \mathbf{x})$ refers to a time series of response whereas $R_j|\{\mathbf{X} = \mathbf{x}\}$ is the maximum response observed over the time series,

$$R_j|\{\mathbf{X} = \mathbf{x}\} = \max_{t \in [0, 60^2]} R_j(t; \mathbf{x}). \quad (46)$$

Step (a) is achieved by modelling the surface elevation according to linear wave theory (see e.g., [Holthuijsen 2010](#)). We obtain the load in (b) using linear surface elevation as input to the methods of [Taylor et al. \(2024\)](#) and [Orszaghova et al. \(2025\)](#), outputting an approximation to the non-linear inertial load $L_j(t; \mathbf{x})$ that $E_j(t; \mathbf{x})$ induces on a monopile. For (c) we pass this load through the linear response function for a damped harmonic oscillator, obtaining a realisation of harmonic response time series $R_j(t; \mathbf{x})$. Steps (a)-(d) are repeated for all n_{RI} realisations, yielding the a numerical estimate of the environment conditioned response $R|\{\mathbf{X} = \mathbf{x}\}$.

4.4.2. Simulation details

Under linear wave theory, the surface elevation $E_j(t; \mathbf{x})$ at time $t > 0$, in a sea state with parameter $\mathbf{X} = \mathbf{x}$, is modelled as the finite sum of Fourier components at n_{Fr} evenly spaced

frequencies $f_1, \dots, f_{n_{\text{Fr}}} > 0$, $f_2 - f_1 = \Delta_{\text{Fr}}$, with contributions determined by underlying wave spectrum $S(f; \mathbf{x})$. We take $n_{\text{Fr}} = 60^2$, $f_1 = 10^{-3}$ and $f_{n_{\text{Fr}}} = 1$, and the JONSWAP (Hasselmann et al., 1973) parametric form for $S(f; \mathbf{x})$ (see SM2) and then model the surface elevation at the location of the structure as

$$E_j(t; \mathbf{x}) = \sum_{i=1}^{n_{\text{Fr}}} \left\{ A_i | \{ \mathbf{X} = \mathbf{x} \} \cdot \cos(2\pi f_i t) + B_i | \{ \mathbf{X} = \mathbf{x} \} \cdot \sin(2\pi f_i t) \right\}, \quad t > 0, \quad (47)$$

where $A_i | \{ \mathbf{X} = \mathbf{x} \}$, $B_i | \{ \mathbf{X} = \mathbf{x} \} \sim N(0, \Delta_{\text{Fr}} S(f_i; \mathbf{x}))$, $i = 1, \dots, n_{\text{Fr}}$, are random Gaussian coefficients with variance equal to the wave energy in frequency band $(f_i - \Delta_{\text{Fr}}/2, f_i + \Delta_{\text{Fr}}/2)$ of the discretised wave spectrum. Model (47) assumes the monopile is placed at the spatial origin and is concentrated at this point with no spatial dimensions. The wave surface elevation (47) is stochastic due to the random Gaussian coefficients, requiring multiple realisations of hourly time series $\{E_j(t; \mathbf{x}) : t \in [0, 60^2]\}$ to capture the full behaviour of the wave surface when $\mathbf{X} = \mathbf{x}$.

The method of Orszaghova et al. (2025) takes these linear surface elevation time series $E_j(t; \mathbf{x})$, recovers the non-linear higher-order harmonics of the wave signal (see SM4.3) and outputs a time series of non-linear horizontal monopile loading $L_j(t; \mathbf{x})$ using the T-FNV model of Taylor et al. (2024). We omit the full details of this methodology here as it is beyond the scope of our case study; in short, the method allows evaluation of structural load without the need to calculate full wave-kinematic profiles (see e.g., Speers et al. 2024), greatly increasing computational efficiency. For this reason, the T-FNV approach is well-suited for our example scenario, as it provides physically-accurate model output (thus testing our methodology in a realistic setting) at a low computational cost (allowing us to generate the full true CDE (6) as the target). This method requires specifying a water depth d , for which we take $d = 30$.

To approximate the effect of wave-induced oscillation on the model monopile, we pass the load $L_j(t; \mathbf{x})$ through the linear response, or transfer, function of a damped harmonic oscillator. For input signal $L_j(t; \mathbf{x})$, the output signal $R_j(t; \mathbf{x})$ is then defined as

$$\chi_{R_j}(f) = \chi_T(f; \gamma) \chi_{L_j}(f), \quad f > 0, \quad (48)$$

where transfer function $\chi_T(f; \gamma)$, the ratio of Fourier transform of the output to the input, takes the form

$$\chi_T(f; \gamma) = \frac{1}{f_0 - f^2 + i\gamma f}, \quad (49)$$

Alternatively,

$$R_j(t; \mathbf{x}) = \mathcal{F}^{-1} \{ \chi_T(f) \cdot \mathcal{F}(L_j(t; \mathbf{x})) \}, \quad (50)$$

where $\mathcal{F} : \mathbb{R} \mapsto \mathbb{R}^+$ is the Fourier transform mapping functions in the time domain to the frequency domain. See SM4.3 for further discussion of the transfer function (49). From time series (50), we obtain a realisation of maximum response (46). Steps (a)-(d) are carried out over a grid of environment values $\mathbf{x}_1, \dots, \mathbf{x}_{n_{\text{Gr}}}$, chosen as the centre points of the cells D_i , $i = 1, \dots, n_{\text{Gr}}$, of $\mathcal{E}_{\mathbf{X}}$ used to estimate the environment density via (45). At each grid point $\mathbf{x}_i$, we obtain $n_{\text{Rl}} = 1000$ realisations of the response $R | \{ \mathbf{X} = \mathbf{x}_i \}$, for centre value $\mathbf{x}_i$ of

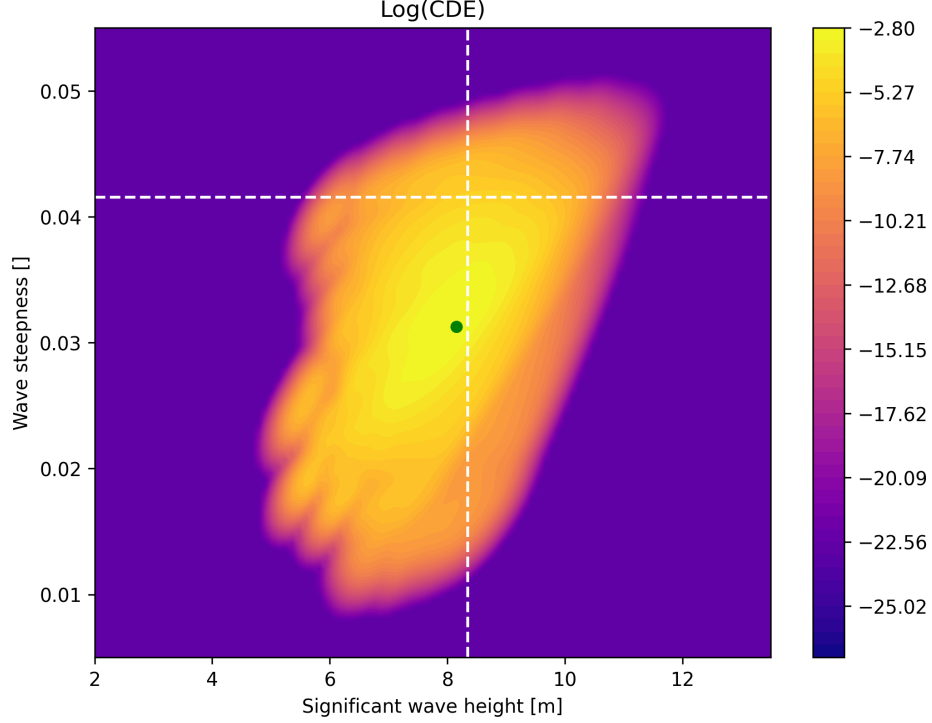


Figure 12: Conditional density of the environment (CDE) for the oscillating monopile scenario, conditioned on exceedance of the 50-year response event. White lines show the marginal 50-year events. The mode of the CDE is indicated in green.

D_i . This procedure provides an empirical estimate $\tilde{F}_{R|\mathbf{X}}(\cdot|\mathbf{x})$ of the distribution of response $R|\{\mathbf{X} = \mathbf{x}\}$, for $\mathbf{x} = \mathbf{x}_1, \dots, \mathbf{x}_{n_{\text{Gr}}}$, from $n_{\text{Rl}} = 1000$ realisations $R_j|\{\mathbf{X} = \mathbf{x}\}$, $j = 1, \dots, n_{\text{Rl}}$, for each $\mathbf{x}$.

4.5. Benchmarking: obtaining a good estimate of CDE and probability of failure

We choose to exploit knowledge of the distribution of $R|\{\mathbf{X} = \mathbf{x}\}$ over the full grid of $\mathbf{x}_1, \dots, \mathbf{x}_{n_{\text{Gr}}} \in \mathcal{E}_{\mathbf{X}}$ in order to obtain a good estimate of CDE. In practical application, we would not attempt this estimation, since it requires a prohibitively expensive total of $n_{\text{Rl}} \times n_{\text{Gr}}$ function evaluations of $R|\{\mathbf{X} = \mathbf{x}\}$. However, with this good estimate of CDE, we are able to evaluate the performance of the IS-PT and AGE procedures, the main objective of this section, reported in Section 4.7. The CDE is estimated using the simulation of Section 4.4 as

$$\tilde{f}_{\mathbf{X}}(\mathbf{x}; r_{\text{Cr}}) = \left\{ 1 - \tilde{F}_{R|\mathbf{X}}(r_{\text{Cr}}|\mathbf{x}) \right\} \times \hat{f}_{\mathbf{X}}(\mathbf{x}), \quad (51)$$

for empirical distribution $\tilde{F}_{R|\mathbf{X}}(\cdot|\mathbf{x})$. Further, $\hat{f}_{\mathbf{X}}$ is the estimated environment density (45) and r_{Cr} is the critical response. The resulting CDE estimate, smoothed using a Gaussian kernel smoother with Scott (2015) bandwidth, is shown in Figure 12. The white dashed lines show the marginal 50-year events for both H_s and S_e , found using the marginal extreme value models (37). The modal point of the estimated CDE (51) is indicated in green.

In practice, the critical response r_{Cr} is specified by domain experts from detailed knowledge

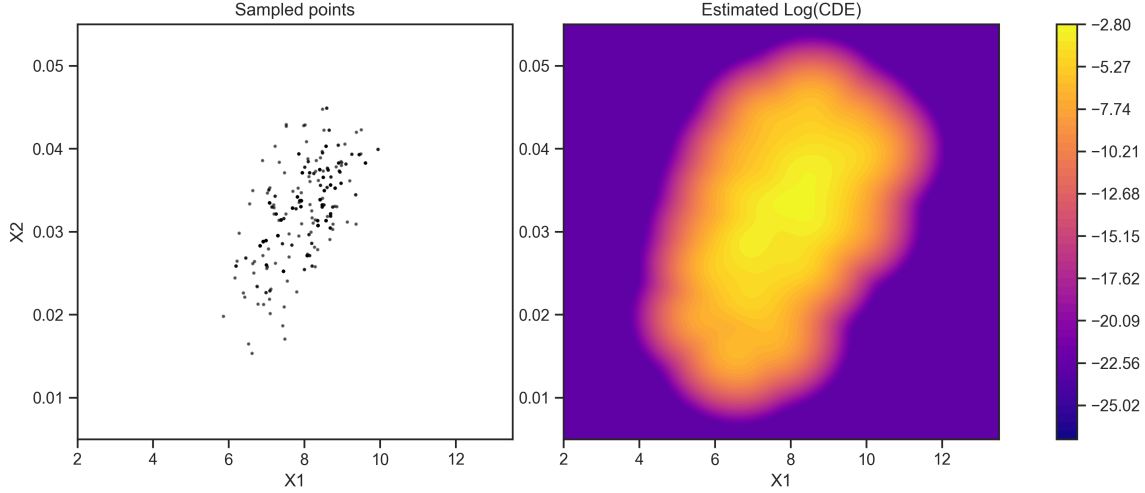


Figure 13: Example sample from CDE (6) under the synthetic structural scenario obtained using the adaptive parallel tempering MCMC algorithm of Vousden et al. (2015) (left), and corresponding smoothed log-CDE estimated using Gaussian kernel bandwidth selected according to Scott (2015) (right).

of the structure, and the corresponding failure probability then estimated as discussed in Section 1. Here, we set the value of r_{Cr} in order to yield a known failure probability for testing purposes. The critical response r_{Cr} is set to $r_{Cr} = \tilde{F}_{R_A}^{-1}(1 - 1/50)$, the 50-year response event, where

$$\tilde{F}_{R_A} = \sum_{m=0}^{\infty} [\tilde{F}_R(r)]^m \frac{\rho_{St} e^{-\rho_{St}}}{m!} = \exp[-\rho_{St}(1 - \tilde{F}_R(r))], \quad (52)$$

is the empirical distribution of the annual maximum response R_A . Further, $\rho_{St} = 26$ is the expected number of storms per annum estimated empirically from the data, and

$$\tilde{F}_R(r) = \int_{\mathcal{E}_X} \tilde{F}_{R|\mathbf{X}}(r|\mathbf{x}) \hat{f}_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}, \quad (53)$$

the empirical distribution of the marginal response R for a single storm event. See Section 3.1.2 of Speers et al. 2024 for further discussion of annual response distribution estimation. The resulting ‘single storm’ failure probability in this case becomes $p_{TFNV} = 1.1 \times 10^{-3}$.

4.6. IS-PT results

We use the IS-PT approach of Section 2 to emulate the CDE (51), taking sampling parameter values $n_{Tm} = 5$, $n_{PT} = 400$, $n_{IS} = 100$, initial proposal variance $\sigma_{MH}^2 = 1$, and bounding temperatures $T_1 = 1$, $T_5 = 20$ seen to perform well in Section 3.2.2. For $n_{Rp} = 100$ replicates, the adaptive algorithm of Vousden et al. (2015) is used to obtain a sample of size $n_{PT} = 400$ (minus burn-in length n_{Br}) from the CDE. A Gaussian kernel smoothed estimate with Scott’s bandwidth is then used as proposal density p_{Pr} in importance sampling estimate (8). Figure 13 shows an example MCMC sample and resulting proposal estimate of the CDE at a single replicate. Over all $n_{IS} = 100$ replicates, we ob-

tain $\text{RMSE}(\hat{p}_{\text{IS}}) = \left(\sum_{r=1}^{n_{\text{RP}}} (\hat{p}_{\text{IS}}^{(r)} - p_{\text{TFNV}})^2 / n_{\text{RI}} \right)^{1/2} = 5.10 \times 10^{-5}$ where $\hat{p}_{\text{IS}}^{(r)}$ is the probability estimate (8) obtained at replicate r , for true $p_{\text{TFNV}} = 1.1 \times 10^{-3}$. The corresponding bias $\text{Bias}(\hat{p}_{\text{IS}}) = 1.83 \times 10^{-6}$ is also small.

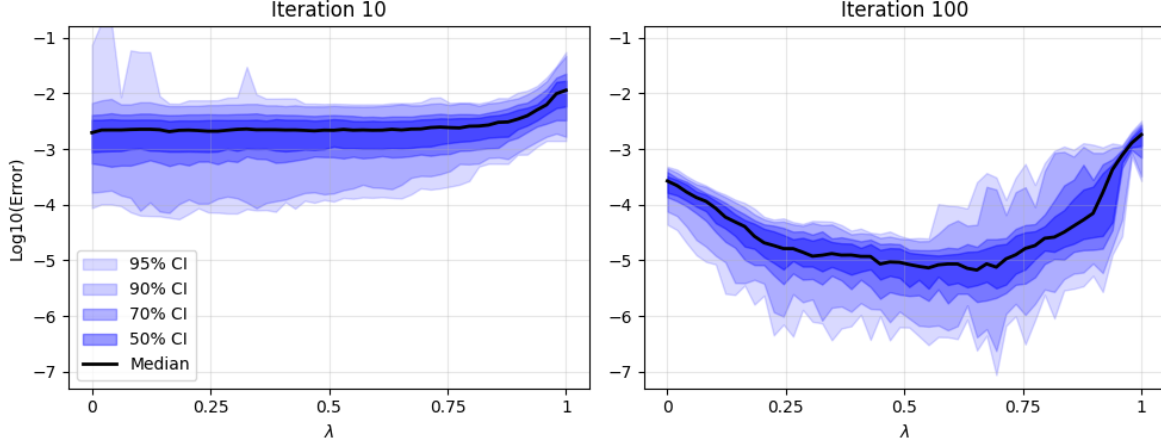


Figure 14: Log-scale absolute error Δ_{GP} of the GP probability estimate $\hat{p}_{\text{GP}}$ at specified iterations, for emulator (10) trained using $U^{(1)}$ over the range of weight $\lambda \in [0.01, 0.99]$ for the TFNV scenario. At iteration 100, the weight that minimises median error is $\lambda^* = 0.67$.

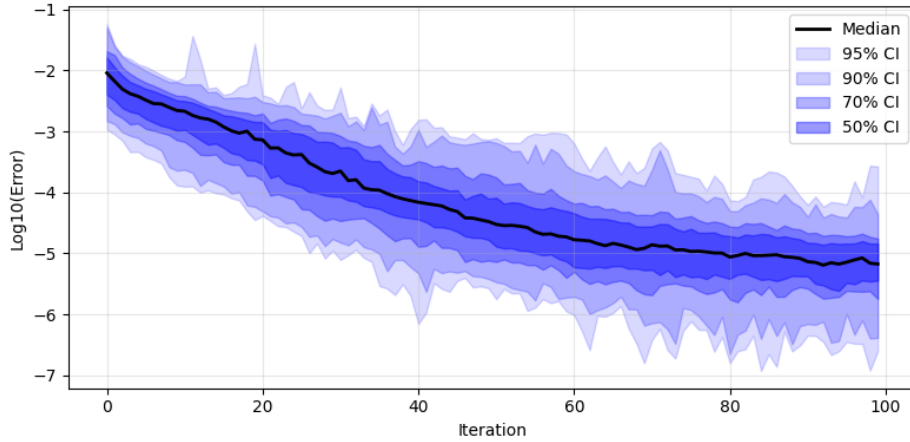


Figure 15: Distribution of log-scale absolute error Δ_{GP} in the GP probability estimate with respect to iteration, trained using $U^{(1)}$ with $\lambda = \lambda^*$. The trend in median error is indicated in black, with various confidence intervals shown in blue.

4.7. AGE results

We now use the AGE methods of Section 2 to emulate the log-CDE, seeking a reasonable estimate of probability of failure with considerably fewer than the $n_{\text{Gr}} \times n_{\text{RI}}$ function evaluations of $R|\{\mathbf{X} = \mathbf{x}\}$ used for the IS-PT estimate in Section 4.5. The emulator for CDE (12) is defined as in (10). For $n_{\text{RP}} = 100$ replicates, it is initialised using Latin hypercube sample $\mathcal{D}_0$,

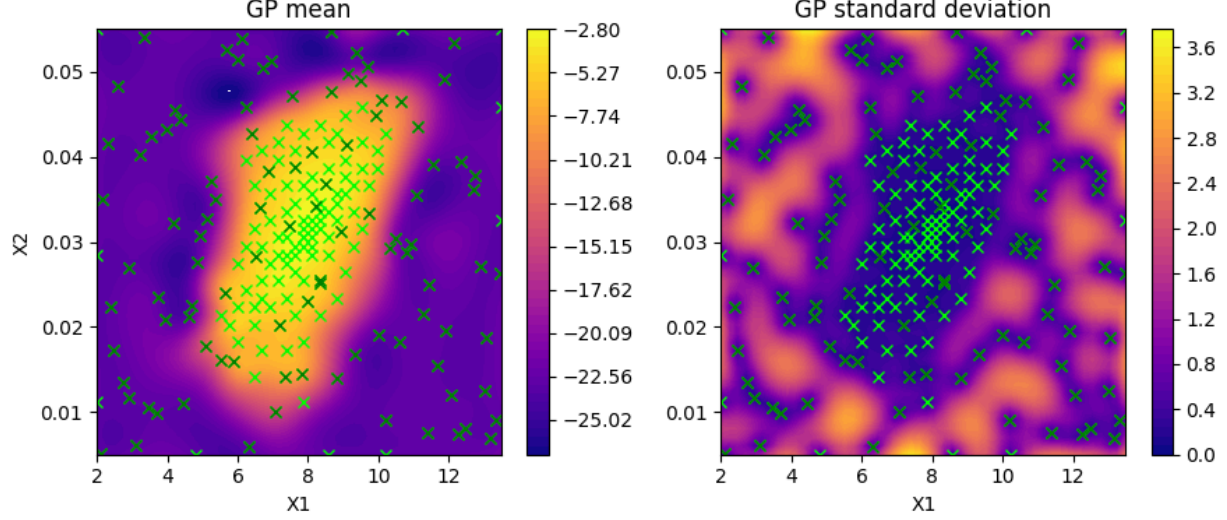


Figure 16: GP emulator at iteration 100 for variance utility $U^{(1)}$ (21), trained using the optimal value $\lambda^* = 0.67$ minimising median of error Δ_{GP} . The panels from left to right show the posterior GP mean $\mu_{100}^*(\mathbf{x})$ over $\mathbf{x} \in \mathcal{E}_{\mathbf{x}}$, and the posterior GP standard deviation $k_{100}^*(\mathbf{x}, \mathbf{x})^{1/2}$. The initial random Latin hypercube training set $\mathcal{D}_0$ is shown as dark green crosses, and the iteratively selected new training points $\mathcal{D}_{100} \setminus \mathcal{D}_0$ are shown as light green crosses.

$|\mathcal{D}_0| = 100$, then trained inductively over n_{RI} realisations of $n_{\text{It2}} = 100$ iterations, using $U^{(1)}$ and $U^{(2)}$ for a range of weights $\lambda \in [0.01, 0.99]$. For utility $U^{(1)}$, Figure 14 shows the relationship between error Δ_{GP} and the value of λ , by comparison with the estimate of p_{TFNV} from Section 4.5. The weight $\lambda^* = 0.67$ is found to minimise the median value of error Δ_{GP} at iteration 100. Figure 15 shows Δ_{GP} with iteration for λ^* , and Figure 16 shows the emulator trained using λ^* at iteration 100. For $|\mathcal{D}_{100}| = 200$ total function evaluations at $\mathbf{x} \in \mathcal{D}_{100} \setminus \mathcal{D}_0$ chosen by $U^{(1)}$ with $\lambda = \lambda^*$, we obtain $\text{RMSE}(\hat{p}_{\text{GP}}) = \left(\sum_{r=1}^{n_{\text{RP}}} (\hat{p}_{\text{GP}}^{(r)} - p_{\text{TFNV}})^2 / n_{\text{RI}} \right)^{1/2} = 6.99 \times 10^{-5}$ where $\hat{p}_{\text{GP}}^{(r)}$ is the probability estimate (17) obtained at iteration 100 and replicate r , for true $p_{\text{TFNV}} = 1.1 \times 10^{-3}$. The corresponding bias $\text{Bias}(\hat{p}_{\text{GP}}) = 1.57 \times 10^{-5}$ is also small. Results using utility $U^{(2)}$ are reported in SM5, and summarised in the next section.

4.8. Comparison of IS-PT and AGE performance

Figure 17 shows the distribution of the IS-PT estimate $\hat{p}_{\text{IS}}$ and the AGE estimates $\hat{p}_{\text{GP}}$ (based on variance and ALC utilities $U^{(1)}$ and $U^{(2)}$ at iteration 100, $\lambda = \lambda^*$) around the target failure probability p_{TFNV} . These results are summarised in Table 1. For the given budgets of expensive function evaluation set, as in Section 3.2.4, methods demonstrate essentially equivalent performance. Again, the key issue is specification of λ for AGE procedures. With λ known, AGE procedures are computationally more efficient. However, specification of λ is in general problematic, suggesting that IS-PT is a more reliably applicable approach.

	IS-PT	AGE $U^{(1)}$	AGE $U^{(2)}$
RMSE	5.10×10^{-5}	6.99×10^{-5}	4.99×10^{-5}
Bias	1.83×10^{-6}	1.57×10^{-5}	9.40×10^{-6}
Number of function evaluations, n_{Ev}	$n_{\text{Tm}} \times n_{\text{PT}} + n_{\text{IS}} = 2100$	$ \mathcal{D}_0 + n_{\text{Itr1}} = 244$	$ \mathcal{D}_0 + n_{\text{Itr1}} = 244$

Table 2: RMSEs and biases of $\hat{p}_{\text{IS}}$ and $\hat{p}_{\text{GP}}$ when targetting failure probability p_{TFNV} , calculated for $n_{\text{Rp}} = 100$ replicate analyses. The true value of probability of failure is $p_{\text{TFNV}} = 1.3 \times 10^{-3}$. Also shown is the number of expensive response function evaluations required for a single replicate analysis for each of IS-PT and AGE.

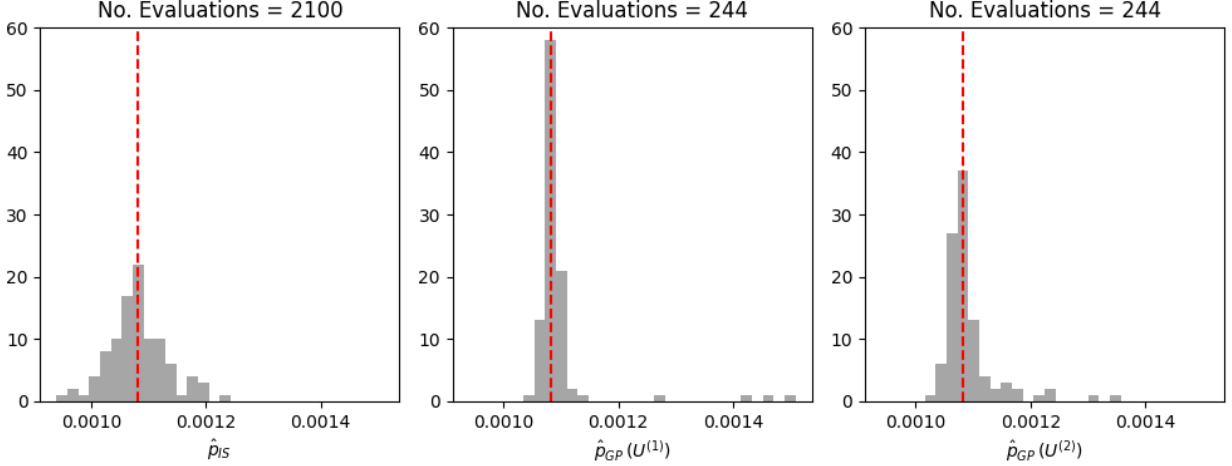


Figure 17: Distribution of $n_{\text{Rp}} = 100$ estimates $\hat{p}_{\text{IS}}$ (left), $\hat{p}_{\text{GP}}$ for $U^{(1)}$ iteration 100 with $\lambda = \lambda^*$ (centre), and $\hat{p}_{\text{GP}}$ for $U^{(2)}$ iteration 100 with $\lambda = \lambda^*$ (right) for true failure probability p_{TFNV} (red). The number of function evaluations required for a single replicate analysis is indicated in the panel titles.

5. Discussion

Estimation of failure probability for marine structures can be a computationally demanding task. In earlier work (Speers et al., 2024) we showed that the conditional density of the environment (CDE) for a structure is a useful design diagnostic, preferable to design contours. Moreover, CDE also provides a natural starting point for estimation of failure probability: the integral of the CDE over the environment space is the probability of structural failure.

The mode of the CDE represents the combination of long term environmental conditions most likely to induce structural failure at the 50-year level. In practice, the location of the mode depends both on the extremal dependence characteristics of the environment variables, and the nature of the fluid-structure interaction. Interestingly, for the oscillating monopile application discussed in Section 4, the location of the mode calculated by brute force in Section 4.5, corresponds approximately to the combination of the 50-year storm peak significant wave height and the 1-year storm peak significant wave steepness.

In this work, we develop, demonstrate and compare two methods to estimate the CDE and hence failure probability for simple monopile structures. The first methodology (IS-PT) incorporates parallel tempering MCMC to estimate the CDE, together with importance sampling to estimate failure probability. The second methodology uses adaptive Gaussian

emulation (AGE) to estimate the CDE and thence Bayesian quadrature to estimate failure probability.

Whereas use of either methodology requires the specification of hyperparameters, the AGE approach is particularly problematic, necessitating the specification of a key weight (λ in e.g., (20)) to control the extent to which adaptive emulation is encouraged to explore the environmental space as opposed to exploiting already-identified informative structure in that space. Specification of λ is in general case dependent.

The computational complexity of each methodology is typically dictated by the number of expensive evaluations of the structural response R given long term environmental conditions $\mathbf{X}$. If the value of explore-exploit λ is known, then procedures adopting AGE provide a good estimate of failure probability requiring an order of magnitude fewer expensive function evaluations than IS-PT. We demonstrate the good performance of IS-PT and two AGE procedures on a simple synthetic structure with complex fluid loading behaviour (and bimodal CDE), and on a more realistic monopile structure (with more straightforward unimodal CDE). Good performance for AGE procedures requires knowledge of the optimal choice of explore-exploit λ , which was evaluated by us in this work using assumed knowledge of the true structural response; obviously, this information will not generally be available to the structural designer. Nevertheless, if it is anticipated that the CDE is likely to be unimodal, we speculate that the choice of λ is likely to be less critical than for more complex CDEs. This can be seen, e.g., by comparison of the intervals I^* of minimum median error in the right hand panels of Figure 6 (bimodal CDE, AGE with variance utility $U^{(1)}$, $I^* \approx [0.6, 0.9]$), Figure 14 (unimodal, $U^{(1)}$, $I^* \approx [0.25, 0.75]$), Figure SM4 (bimodal, AGE with ALC utility $U^{(2)}$, $I^* \approx [0.2, 0.35]$) and Figure SM11 (unimodal CDE, $U^{(2)}$, $I^* \approx [0.1, 0.5]$). Intervals I^* of acceptable values for λ are wider for unimodal CDEs, and moreover the intervals corresponding to utilities $U^{(1)}$ and $U^{(2)}$ overlap. However, for bimodal CDEs, the intervals I^* corresponding to $U^{(1)}$ and $U^{(2)}$ are disjoint. Given the importance of selecting λ well, investigation into the performance acquisition functions not reliant on the specification of weight parameter λ (e.g., Osborne et al. 2012, Gunter et al. 2014) is warranted. These methods, however, incur additional theoretical assumptions and computational complexity, which may limit their usefulness in general offshore applications.

More generally, sampling from high-dimensional, multimodal distributions is receiving increasing attention in the academic literature, since it is a fundamental issue in modern statistical learning. Tempering methods are often used when multimodality is suspected, yet these can in principle be difficult to tune as noted by e.g., Miasojedow et al. (2013), especially in high dimensional settings. Solutions have been proposed, including Graham and Storkey (2017), Luo et al. (2018) and Park (2025) in the context of Hamiltonian Monte Carlo sampling. Other authors (e.g., Qiu and Wang 2024) consider sampling using tempered distribution flows. In fact, the parallel tempering package exploited in the current work has already been applied successfully in five dimensions by Vousden et al. (2015); this may be sufficient for less complex metocean design studies. Nevertheless, exploration of the performance of IS-PT in higher dimensions for more complex structural types is an interesting areas for future work; studies along these lines are currently being conducted by the authors.

If the optimal value of λ for AGE is unknown (as will generally be the case), IS-PT provides a reliable general-purpose approach to estimation of CDE and failure probability useful even for challenging multimodal CDEs. In the current work, we consider univariate responses in a two-dimensional environment. We anticipate that, for higher-dimensional responses and environments, the structure of the CDE will be more complex (and multimodal) in general. Given this, it appears reasonable to assume that IS-PT will prove a more reliable route to estimation of CDE and probability of structural failure. In contrast, we view AGE as a method requiring further development before it can be considered a mature engineering solution.

Acknowledgments

The work was completed while Matthew Speers was part of the EPSRC funded STOR-i centre for doctoral training (grant no. EP/S022252/1), with part-funding from the ARC TIDE Industrial Transformational Research Hub at the University of Western Australia. The authors wish to acknowledge the support of Jana Orszaghova and Paul Taylor at the University of Western Australia, and David Randell at Shell.

References

- Abramowitz, M., Stegun, I.A., 1965. Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables. volume 55. Courier Corporation.
- Bishop, C.M., Nasrabadi, N.M., 2006. Pattern Recognition and Machine Learning. volume 4. Springer.
- Castellon, D.F., Fenerci, A., Øiseth, O., Petersen, Ø.W., 2022. Investigations of the long-term extreme buffeting response of long-span bridges using importance sampling Monte Carlo simulations. *Engineering Structures* 273, 114986.
- Castellon, D.F., Fenerci, A., Petersen, Ø.W., Øiseth, O., 2023. Full long-term buffeting analysis of suspension bridges using Gaussian process surrogate modelling and importance sampling Monte Carlo simulations. *Reliability Engineering & System Safety* 235, 109211.
- Chib, S., Greenberg, E., 1995. Understanding the Metropolis-Hastings algorithm. *The American Statistician* 49, 327–335.
- Cohn, D., 1993. Neural network exploration using optimal experiment design, in: Cowan, J., Tesauro, G., Alspector, J. (Eds.), *Advances in Neural Information Processing Systems*, Morgan-Kaufmann.
- Davison, A.C., Smith, R.L., 1990. Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society Series B* 52, 393–425.
- Earl, D.J., Deem, M.W., 2005. Parallel tempering: theory, applications, and new perspectives. *Physical Chemistry Chemical Physics* 7, 3910–3916.

- Ewans, K., Jonathan, P., 2008. The effect of directionality on northern North Sea extreme wave design criteria. *Journal of Offshore Mechanics and Arctic Engineering* 130, 041604.
- Faltinsen, O.M., Newman, J.N., Vinje, T., 1995. Nonlinear wave loads on a slender vertical cylinder. *Journal of Fluid Mechanics* 289, 179–198.
- Genton, M.G., 2001. Classes of kernels for machine learning: a statistics perspective. *Journal of machine learning research* 2, 299–312.
- Geweke, J., 1991. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. Technical Report. Federal Reserve Bank of Minneapolis.
- Graham, M.M., Storkey, A.J., 2017. Continuously tempered Hamiltonian Monte Carlo. [arXiv:1704.03338](https://arxiv.org/abs/1704.03338).
- Gramstad, O., Agrell, C., Bitner-Gregersen, E., Guo, B., Ruth, E., Vanem, E., 2020. Sequential sampling method using Gaussian process regression for estimating extreme structural response. *Marine Structures* 72, 102780.
- Gunter, T., Osborne, M.A., Garnett, R., Hennig, P., Roberts, S.J., 2014. Sampling for inference in probabilistic models with fast Bayesian quadrature, in: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K. (Eds.), *Advances in Neural Information Processing Systems*, Curran Associates, Inc.
- Hasselmann, K., Barnett, T.P., Bouws, E., Carlson, H., Cartwright, D.E., Enke, K., Ewing, J., Gienapp, A., Hasselmann, D., Kruseman, P., et al., 1973. Measurements of wind-wave growth and swell decay during the Joint North Sea Wave Project (JONSWAP). *Ergaenzungsheft zur Deutschen Hydrographischen Zeitschrift, Reihe A* .
- He, R., Zhu, T., 2019. Model tests on the frequency responses of offshore monopiles. *Journal of Marine Science and Engineering* 7, 430.
- Heffernan, J.E., Tawn, J.A., 2004. A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 66, 497–546.
- Hennig, P., Osborne, M.A., Kersting, H.P., 2022. *Probabilistic Numerics: Computation as Machine Learning*. Cambridge University Press.
- Holthuijsen, L.H., 2010. *Waves in Oceanic and Coastal Waters*. Cambridge university press.
- Jonathan, P., Ewans, K., Randell, D., 2014. Non-stationary conditional extremes of northern North Sea storm characteristics. *Environmetrics* 25, 172–188.
- Keef, C., Papastathopoulos, I., Tawn, J.A., 2013. Estimation of the conditional distribution of a multivariate variable given that one of its components is large: Additional constraints for the Heffernan and Tawn model. *Journal of Multivariate Analysis* 115, 396–404.

- Liu, J., Xu, S., Mao, H., Han, C., Han, J., 2025. High-frequency resonance of bottom-fixed monopile-supported offshore wind turbines under nonlinear waves. *Ocean Engineering* 340, 122355.
- Luo, R., Wang, J., Yang, Y., Wang, J., Zhu, Z., 2018. Thermostat-assisted continuously-tempered Hamiltonian Monte Carlo for Bayesian learning, in: Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R. (Eds.), *Advances in Neural Information Processing Systems*, Curran Associates, Inc.
- Lystad, T.M., Fenerci, A., Øiseth, O., 2023. Full long-term extreme buffeting response calculations using sequential Gaussian process surrogate modeling. *Engineering Structures* 292, 116495.
- Marrel, A., Iooss, B., 2024. Probabilistic surrogate modeling by Gaussian process: A review on recent insights in estimation and validation. *Reliability Engineering & System Safety* 247, 110094.
- Mathisen, J., Bitner-Gregersen, E., 1990. Joint distributions for significant wave height and wave zero-up-crossing period. *Applied Ocean Research* 12, 93–103.
- Meng, X.L., Wong, W.H., 1996. Simulating ratios of normalizing constants via a simple identity: A theoretical exploration. *Statistica Sinica* 6, 831–860.
- Miasojedow, B., Moulines, E., Vihola, M., 2013. An adaptive parallel tempering algorithm. *Journal of Computational and Graphical Statistics* 22, 649–664.
- Moustapha, M., Marelli, S., Sudret, B., 2022. Active learning for structural reliability: survey, general framework and benchmark. *Structural Safety* 96, 102174.
- Murphy, C., Tawn, J.A., Varty, Z., 2025. Automated threshold selection and associated inference uncertainty for univariate extremes. *Technometrics* 67, 215–224.
- Orszaghova, J., Taylor, P.H., Wolgamot, H., McCauley, G., Kurniawan, A., Wu, Q., Tan, B., George, A.E., 2025. Wave loads on monopile foundations revisited – new high-quality experiments for validation of a novel engineering model, in: *Proc. ASME OMAE 2025*, ASME, Vancouver, British Columbia, Canada. OMAE2025-156732.
- Osborne, M., Garnett, R., Ghahramani, Z., Duvenaud, D.K., Roberts, S.J., Rasmussen, C., 2012. Active learning of model evidence using Bayesian quadrature, in: Pereira, F., Burges, C., Bottou, L., Weinberger, K. (Eds.), *Advances in Neural Information Processing Systems*, Curran Associates, Inc.
- Park, J., 2025. Sampling from high-dimensional, multimodal distributions using automatically tuned, tempered Hamiltonian Monte Carlo. [arXiv:2111.06871](https://arxiv.org/abs/2111.06871).
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Édouard Duchesnay, 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12, 2825–2830.

- Peherstorfer, B., Cui, T., Marzouk, Y., Willcox, K., 2016. Multifidelity importance sampling. *Computer Methods in Applied Mechanics and Engineering* 300, 490–509.
- Pollatsek, A., Tversky, A., 1970. A theory of risk. *Journal of Mathematical Psychology* 7, 540–553.
- Qiu, Y., Wang, X., 2024. Efficient multimodal sampling via tempered distribution flow. *Journal of the American Statistical Association* 119, 1446–1460.
- Riise, B.H., Grue, J., Jensen, A., Johannessen, T.B., 2018. High frequency resonant response of a monopile in irregular deep water waves. *Journal of Fluid Mechanics* 853, 564–586.
- Rubinstein, R.Y., Kroese, D.P., 2016. *Simulation and the Monte Carlo method*. John Wiley & Sons.
- Sambridge, M., 2013. A parallel tempering algorithm for probabilistic sampling and multimodal optimization. *Geophysical Journal International* 196, 357–374.
- Schälte, Y., Fröhlich, F., Jost, P.J., Vanhoefer, J., Pathirana, D., Stapor, P., Lakrisenko, P., Wang, D., Raimúndez, E., Merkt, S., Schmiester, L., Städter, P., Grein, S., Dudkin, E., Doresic, D., Weindl, D., Hasenauer, J., 2023. pyPESTO: a modular and scalable tool for parameter estimation for dynamic models. *Bioinformatics* 39, 711.
- Scott, D.W., 2015. *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley & Sons.
- Seo, S., Wallat, M., Graepel, T., Obermayer, K., 2000. Gaussian process regression: active data selection and test point rejection, in: *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium*, pp. 241–246 vol.3.
- Shooter, R., Tawn, J.A., Ross, E., Jonathan, P., 2021. Basin-wide spatial conditional extremes for severe ocean storms. *Extremes* 24, 241–265.
- Speers, M., Randell, D., Tawn, J.A., Jonathan, P., 2024. Estimating metocean environments associated with extreme structural response to demonstrate the dangers of environmental contour methods. *Ocean Engineering* 311, 118754.
- Tabandeh, A., Jia, G., Gardoni, P., 2022. A review and assessment of importance sampling methods for reliability analysis. *Structural Safety* 97, 102216.
- Taylor, P.H., Tang, T., Adcock, T.A., Zang, J., 2024. Transformed-FNV: wave forces on a vertical cylinder—a free-surface formulation. *Coastal Engineering* 189, 104454.
- Tendijck, S., Tawn, J., Jonathan, P., 2023. Extremal characteristics of conditional models. *Extremes* 26, 139–156.
- Towe, R.P., Tawn, J.A., Lamb, R., Sherlock, C.G., 2019. Model-based inference of conditional extreme value distributions with hydrological applications. *Environmetrics* 30, e2575.

- Varty, Z., Tawn, J.A., Atkinson, P.M., Bierman, S., 2021. Inference for extreme earthquake magnitudes accounting for a time-varying measurement process. [arXiv:2102.00884](https://arxiv.org/abs/2102.00884).
- Vousden, W.D., Farr, W.M., Mandel, I., 2015. Dynamic temperature selection for parallel tempering in markov chain monte carlo simulations. *Monthly Notices of the Royal Astronomical Society* 455, 1919–1937.
- Wan, J.H., Bai, R., Li, X.Y., Liu, S.W., 2023. Natural frequency analysis of monopile supported offshore wind turbines using unified beam-column element model. *Journal of Marine Science and Engineering* 11, 628.
- Wang, C., Qiang, X., Xu, M., Wu, T., 2022. Recent advances in surrogate modeling methods for uncertainty quantification and propagation. *Symmetry* 14, 1219.
- Wang, H., Gramstad, O., Schär, S., Marelli, S., Vanem, E., 2024. Comparison of probabilistic structural reliability methods for ultimate limit state assessment of wind turbines. *Structural Safety* 111, 102502.
- Wang, J., Sun, Z., Cao, R., 2021. An efficient and robust kriging-based method for system reliability analysis. *Reliability Engineering & System Safety* 216, 107953.
- Winter, H.C., Tawn, J.A., 2017. kth-order Markov extremal models for assessing heatwave risks. *Extremes* 20, 393–415.
- Xiao, N.C., Zhan, H., Yuan, K., 2020. A new reliability method for small failure probability problems by combining the adaptive importance sampling and surrogate models. *Computer Methods in Applied Mechanics and Engineering* 372, 113336.
- Yang, H., Zhang, Y., Lin, J., Mao, H., 2025. Higher-harmonic resonance response of a monopile-supported offshore wind turbine in waves and wind. *Ocean Engineering* 342, 123120.
- Yang, X., Liu, Y., Mi, C., Tang, C., 2018. System reliability analysis through active learning kriging model with truncated candidate region. *Reliability Engineering & System Safety* 169, 235–241.

Supplementary Material to ‘Sequential Design for the Efficient Estimation of Offshore Structure Failure Probability’

Matthew Speers, Philip Jonathan, Jonathan Tawn

School of Mathematical Sciences, Lancaster University LA1 4YF, United Kingdom

SM1. Alternative methodology to that presented in Section 2 of the main text

SM1.1. Gaussian process-informed importance sampling

Referring to Section 2.2 of the main text, authors Xiao et al. (2020) and Lystad et al. (2023), use Gaussian emulation to inform their choice of importance sampling density. We briefly propose a similar approach using the GP emulators of the main article. This avoids the need to run the MCMC sampler described in the main article, allowing either (a) more budget allocation to the evaluation of importance sampling estimate $\hat{p}_{\text{IS}}$, or (b) a reduction in total computational cost.

Our approach mirrors that of Lystad et al. (2023), who create a uniform proposal density with support informed by a GP estimate of the CDE. Given an estimate $\hat{f}_{\mathbf{X}|R>r_{\text{Cr}}}^{(n)}$ of the CDE, found via (22) using the n th-iterate GP emulator defined in the main article, we define a proposal density

$$g_{\text{Pr}}^{(n)}(\mathbf{x}) = \frac{1}{A_n} \quad \text{for} \quad A_n = \int_{\mathcal{E}_{\mathbf{X}}} \mathbb{I} \left\{ \hat{f}_{\mathbf{X}|R>r_{\text{Cr}}}^{(n)}(\mathbf{x}) > \delta \right\} d\mathbf{x}, \quad (\text{SM.1})$$

for some $\delta \in [0, 1]$. That is, we draw a proposal sample $\mathbf{x}_1^*, \dots, \mathbf{x}_N^*$ uniformly on the region where the estimate of the CDE at the current iteration n is greater than some δ .

SM1.2. Gaussian process emulation of failure probability

We consider an alternate emulator construction to that shown in Section 2.3 of the main article. Here, we emulate only the conditional failure probability $\Pr(R > r_{\text{Cr}} | \{\mathbf{X} = \mathbf{x}\})$, rather than the entire integrand $\Pr(R > r_{\text{Cr}} | \{\mathbf{X} = \mathbf{x}\}) f_{\mathbf{X}}(\mathbf{x})$. Since probabilities must always be observed on the unit interval, we map the Gaussian emulator output $w(\mathbf{x}) \in \mathbb{R}$ onto the range $[0, 1]$ via the logistic function $g_{\text{Lg}} : \mathbb{R} \mapsto [0, 1]$, $g_{\text{Lg}}(w) = e^w / (1 + e^w)$, modelling

$$w(\mathbf{x}) = g_{\text{Lg}}^{-1}(\Pr(G_R(\mathbf{x}) > r_{\text{Cr}})) \sim \mathcal{GP}(\mu_{\text{GP}}(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')), \quad w : \mathcal{E}_{\mathbf{X}} \mapsto \mathbb{R}, \quad (\text{SM.2})$$

for mean and covariance functions

$$\mu_{\text{GP}}(\mathbf{x}) = \mathbb{E}[w(\mathbf{x})], \quad \mu_{\text{GP}} : \mathcal{E}_{\mathbf{X}} \rightarrow \mathbb{R}, \quad (\text{SM.3})$$

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(w(\mathbf{x}) - \mu(\mathbf{x}))(w(\mathbf{x}') - \mu(\mathbf{x}'))], \quad k(\mathbf{x}, \mathbf{x}') : \mathcal{E}_{\mathbf{X}} \times \mathcal{E}_{\mathbf{X}} \rightarrow \mathbb{R}. \quad (\text{SM.4})$$

This emulator may then be trained according to the posterior update steps defined in the main article. The target marginal failure probability estimate $\hat{p}_{\text{GP}}$ can then be summarised using

$$\hat{p}_{\text{GP}} = \mathbb{E}_{W, \mathbf{X}}(\{g_{\text{Lg}}(w(\mathbf{x}))\}) \quad (\text{SM.5})$$

$$= \int_{\mathcal{E}_{\mathbf{X}}} \left\{ \int_{\mathbb{R}} g_{\text{Lg}}(w) \phi(w; \mu_{\text{GP}}^*(\mathbf{x}), k^*(\mathbf{x}, \mathbf{x})) dw \right\} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}, \quad (\text{SM.6})$$

for $W|\{\mathbf{X} = \mathbf{x}\} \sim N(\mu_{\text{GP}}^*(\mathbf{x}), k^*(\mathbf{x}, \mathbf{x}))$ with parameters obtained from (SM.2). The estimate (SM.6) can be written

$$\hat{p}_{\text{GP}} \approx \int_{\mathcal{E}_{\mathbf{X}}} g_{\text{Lg}} \left(\frac{\mu_{\text{GP}}^*(\mathbf{x})}{\sqrt{1 + \pi k(\mathbf{x}, \mathbf{x}')/8}} \right) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}, \quad (\text{SM.7})$$

by the approximation for the convolution of a logistic sigmoid function with a Gaussian density given in Section 4.5.2 of Bishop and Nasrabadi (2006). Values for (SM.7) may then be obtained via numerical integration. In this setting, the CDE estimate of the GP emulator at iteration n becomes

$$\hat{f}_{\mathbf{X}|R>r_{\text{Cr}}}^{(n)}(\mathbf{x}) = \frac{g_{\text{Lg}} \left(\mu_n^*(\mathbf{x}) (1 + \pi k_n^*(\mathbf{x}, \mathbf{x}')/8)^{-\frac{1}{2}} \right) f_{\mathbf{X}}(\mathbf{x})}{\int_{\mathcal{E}_{\mathbf{X}}} g_{\text{Lg}} \left(\mu_n^*(\mathbf{x}) (1 + \pi k_n^*(\mathbf{x}, \mathbf{x}')/8)^{-\frac{1}{2}} \right) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}}. \quad (\text{SM.8})$$

SM2. JONSWAP wave spectrum discussed in Section 4.4 of the main text

The JONSWAP spectral density (Hasselmann et al., 1973) is used to simulate linear random waves in the monopile application of the main article. It is defined, in terms of angular frequency $\omega = 2\pi f$, as

$$S(\omega; \mathbf{x}) = \alpha \omega^{-r} \exp \left\{ -\frac{r}{4} \left(\frac{|\omega|}{\omega_p(\mathbf{x})} \right)^{-4} \right\} \gamma^{\delta(\omega; \mathbf{x})}, \quad (\text{SM.9})$$

for $\omega > 0$, where $\mathbf{X} = (H_s, S_e)$ and $\omega_p(\mathbf{x}) = 2\pi/t(\mathbf{x})$, where $t(\mathbf{x})$ is the observed value of the second spectral moment wave period $T_2 = \sqrt{(2\pi H_S)/(g S_e)}$ in sea state $\mathbf{X} = \mathbf{x}$, with

$$\delta(\omega; \mathbf{x}) = \exp \left\{ -\frac{1}{2(0.07 + 0.02 \cdot I\{\omega_p(\mathbf{x}) > |\omega|\})^2} \left(\frac{|\omega|}{\omega_p(\mathbf{x})} - 1 \right)^2 \right\}, \quad (\text{SM.10})$$

and constants $r, \alpha, \gamma > 0$. The Phillips constant α is chosen so that

$$4 \cdot \left\{ \int_{-\infty}^{\infty} S(\omega; \mathbf{x}) d\omega \right\}^{\frac{1}{2}} = h(\mathbf{x}), \quad (\text{SM.11})$$

where $h(\mathbf{x})$ is the observed value of significant wave height H_S in sea state $\mathbf{X} = \mathbf{x}$.

SM3. Supplementary results to case studies of Section 3.2 of the main text

SM3.1. Importance sampling coupled with parallel tempering MCMC (IS-PT)

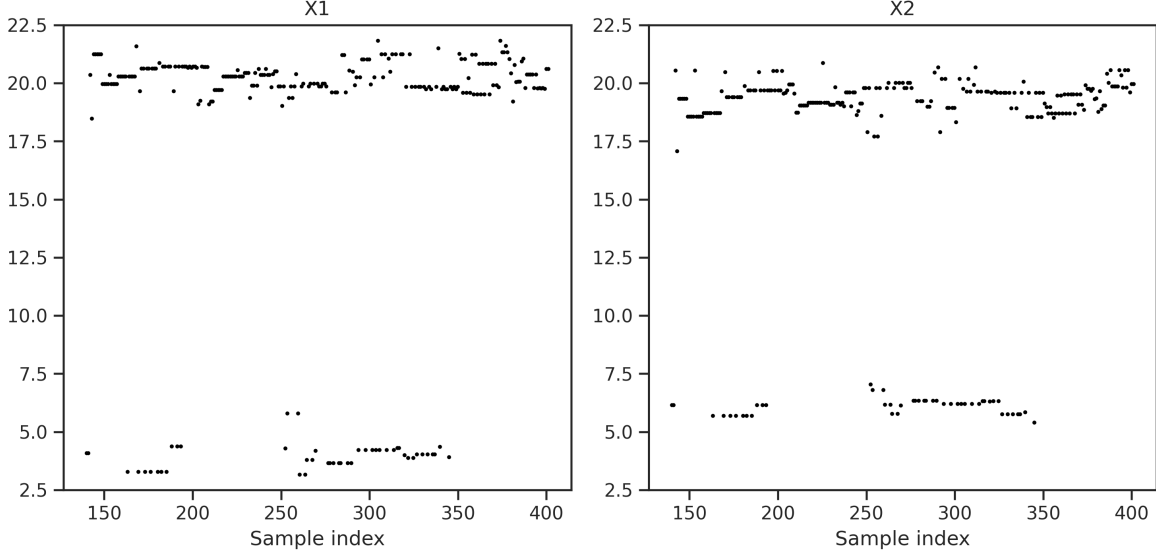


Figure SM1: Example trace plot of sample from synthetic CDE discussed in Section 3, obtained from adaptive parallel tempering algorithm of [Vousden et al. \(2015\)](#) at temperature $T_1 = 1$, used in IS-PT approach for estimation of proposal density. Sample is initially of length $n_{PT} = 400$, with $n_{Br} = 144$ discarded (not plotted).

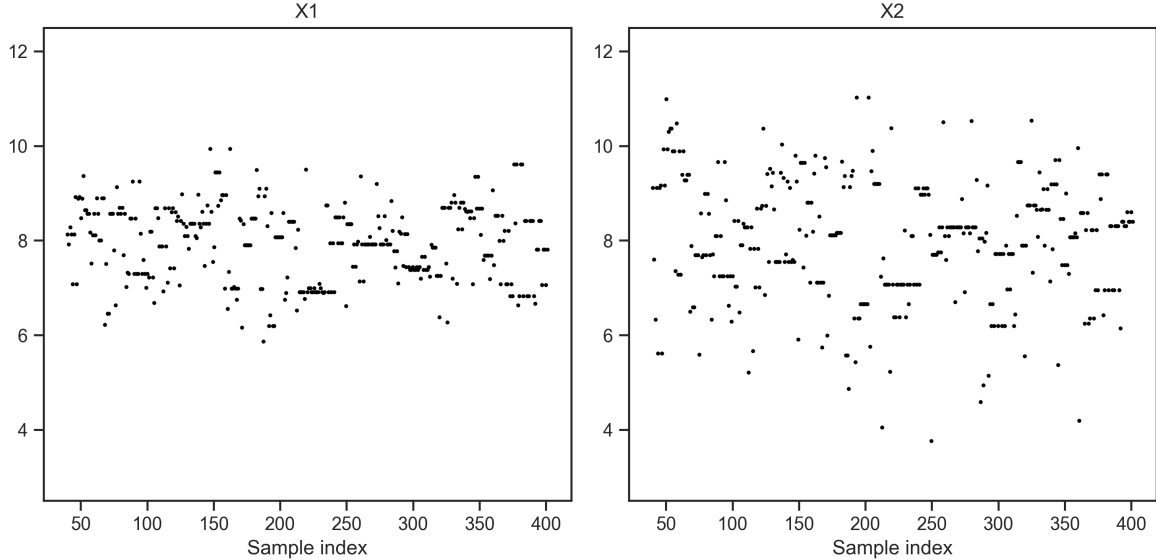


Figure SM2: Example trace plot of sample from monopile CDE discussed in Section 4, obtained from the adaptive parallel tempering algorithm of [Vousden et al. \(2015\)](#) at temperature $T_1 = 1$, used in IS-PT approach for estimation of proposal density. Sample is initially of length $n_{PT} = 400$, with $n_{Br} = 40$ discarded (not plotted).

SM3.2. Adaptive Gaussian emulation (AGE)

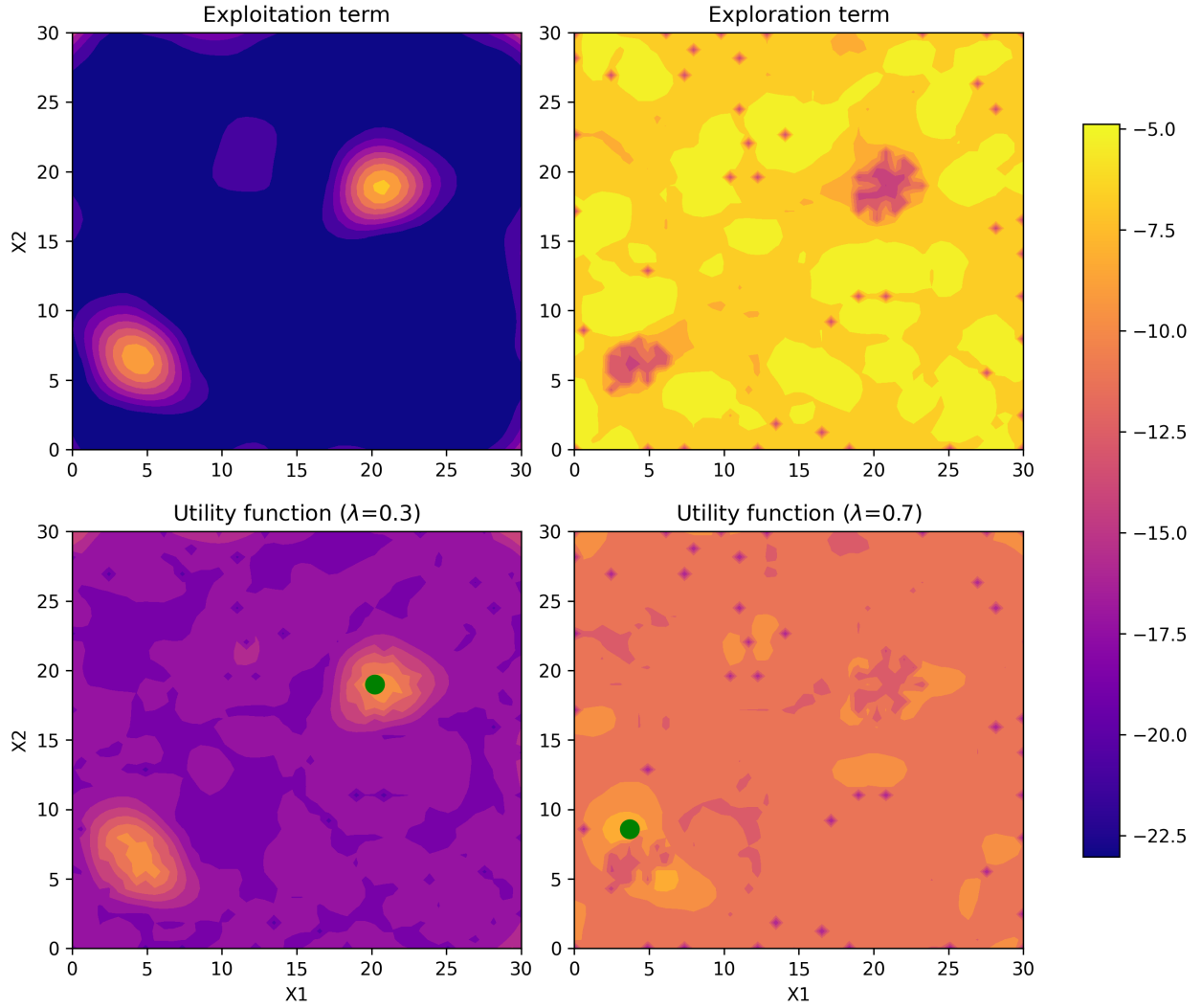


Figure SM3: Behaviour of utility function $U^{(2)}(\mathbf{x}; \lambda)$ over the environment space $\mathcal{E}_{\mathbf{X}}$, for synthetic scenario. Upper panels show exploitation and exploration terms obtained from GP emulator (10) trained on initial Latin hypercube set $\mathcal{D}_0$ of size $n_{\text{Tr1}} = 144$. Lower panels show resulting utility functions for weights $\lambda = 0.3$ and $\lambda = 0.7$. In each lower panel, the optimal sampling point $\mathbf{x}^* = \arg\max_{\mathbf{x} \in \mathcal{E}_{\mathbf{X}}} U^{(2)}(\mathbf{x}; \lambda)$ is indicated in green. To be compared with Figure 5 of the main text.

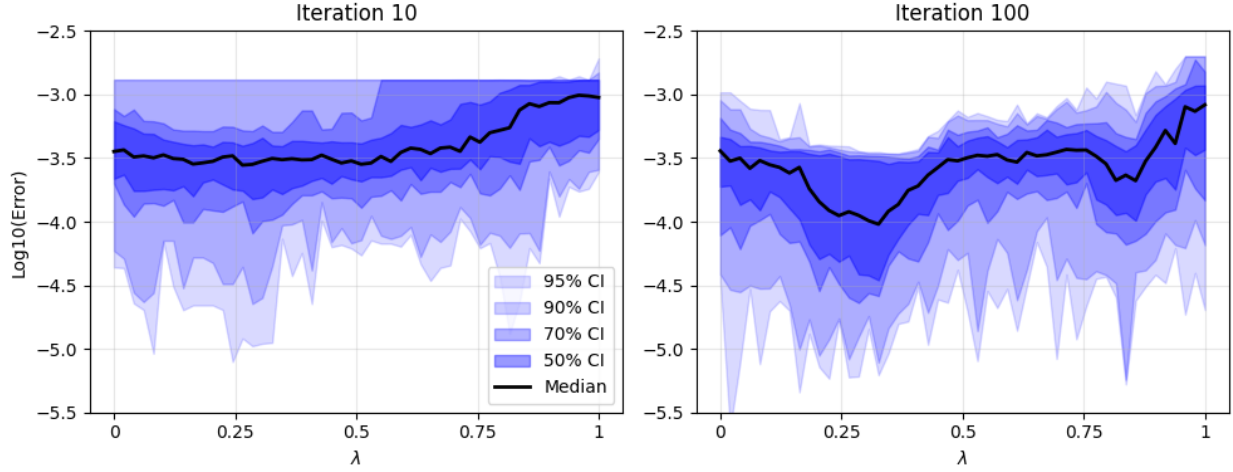


Figure SM4: Log-scale absolute error Δ_{GP} of the GP probability estimate $\hat{p}_{\text{GP}}$ at specified iterations, for emulator (10) trained using $U^{(2)}$ over the range of weight $\lambda \in [0.01, 0.99]$ for the synthetic scenario. At iteration 100, the weight that minimises median error is $\lambda^* = 0.33$. To be compared with Figure 6 of the main text.

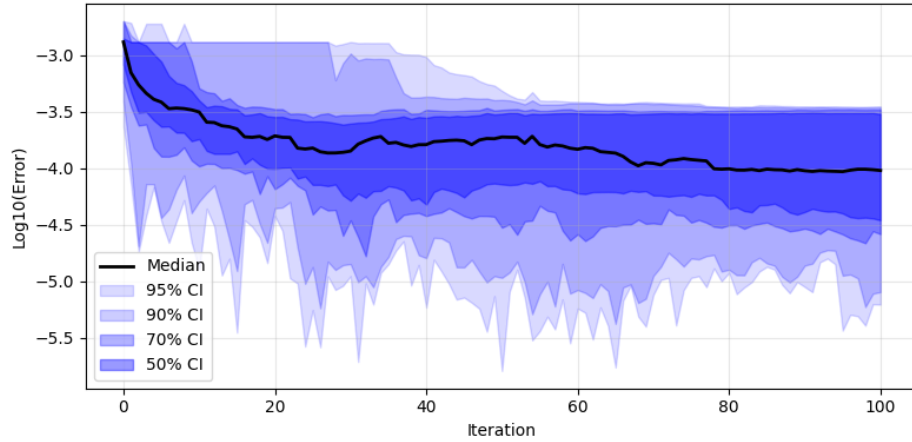


Figure SM5: Distribution of log-scale absolute error Δ_{GP} in the GP probability estimate with respect to iteration, trained using $U^{(2)}$ with $\lambda = \lambda^*$. The trend in median error is indicated in black, with various confidence intervals shown in blue. To be compared with Figure 7 of the main text.

SM4. Supplementary results to monopile case study of Section 4 of the main text

SM4.1. Extreme value model threshold selection supporting the discussion of Section 4.3.4 of the main text

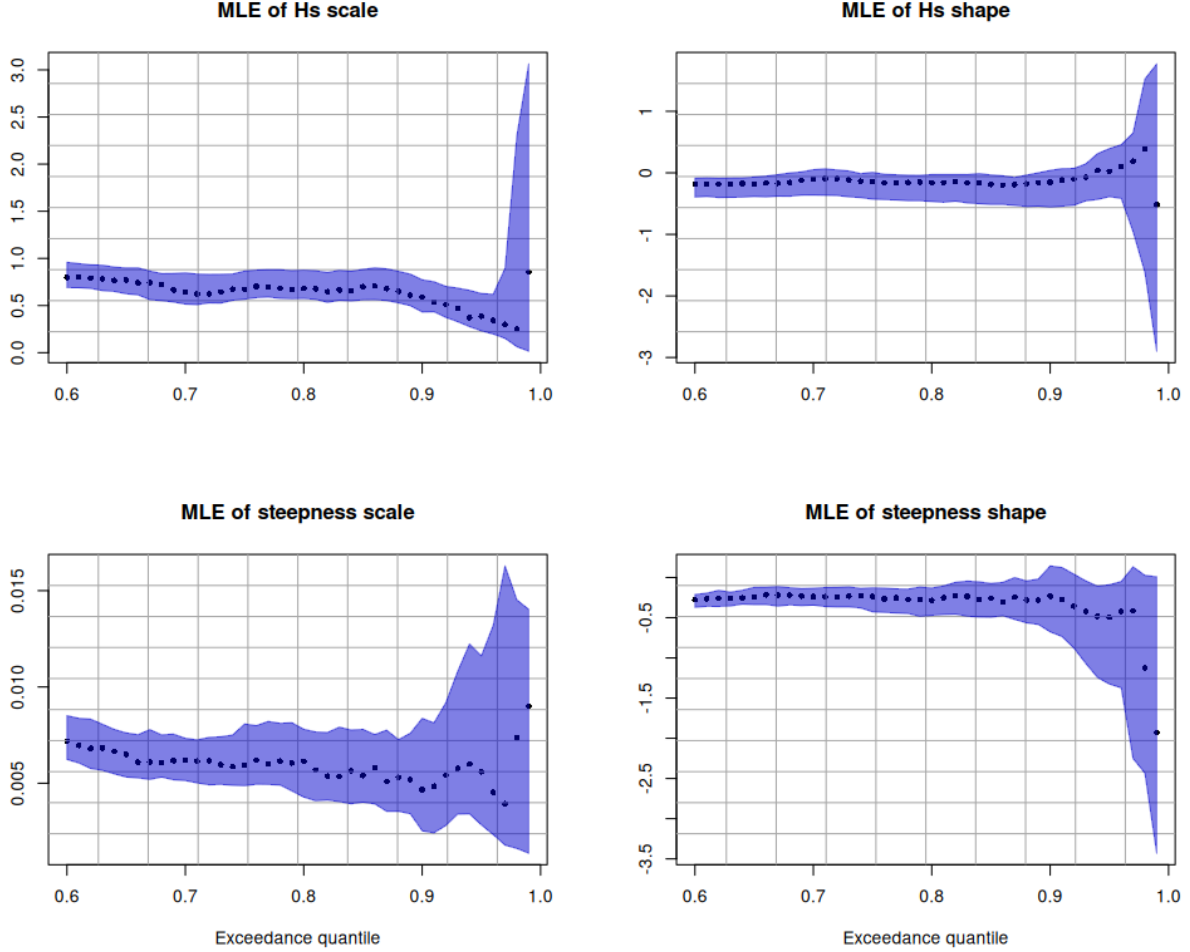


Figure SM6: Threshold stability plots for estimates of the generalised Pareto scale and shape parameters σ and ξ when fitted to H_s and S_e above a range of values of the conditioning threshold $u > 0$. The estimates of σ and ξ are given on the y-axes, with the respective threshold quantile $q_u = \tilde{F}_{H_s}^{-1}(u)$ and $q_u = \tilde{F}_{S_e}^{-1}(u)$ indicated on the x-axes, for empirical distributions $\tilde{F}_{H_s}$ of H_s and $\tilde{F}_{S_e}$ of S_e . Point estimates from original Albany data are given in black, and bootstrapped 95% confidence intervals are shown as a blue region. Stability of estimates for ξ and linearity of estimates of σ above a threshold quantile q_u indicates that u is a suitable choice for GPD model threshold. Following visual analysis of the four panels we select threshold $u_1 = \tilde{F}_{H_s}^{-1}(0.7)$ and $u_2 = \tilde{F}_{S_e}^{-1}(0.7)$ for marginal modelling of H_s and S_e respectively.

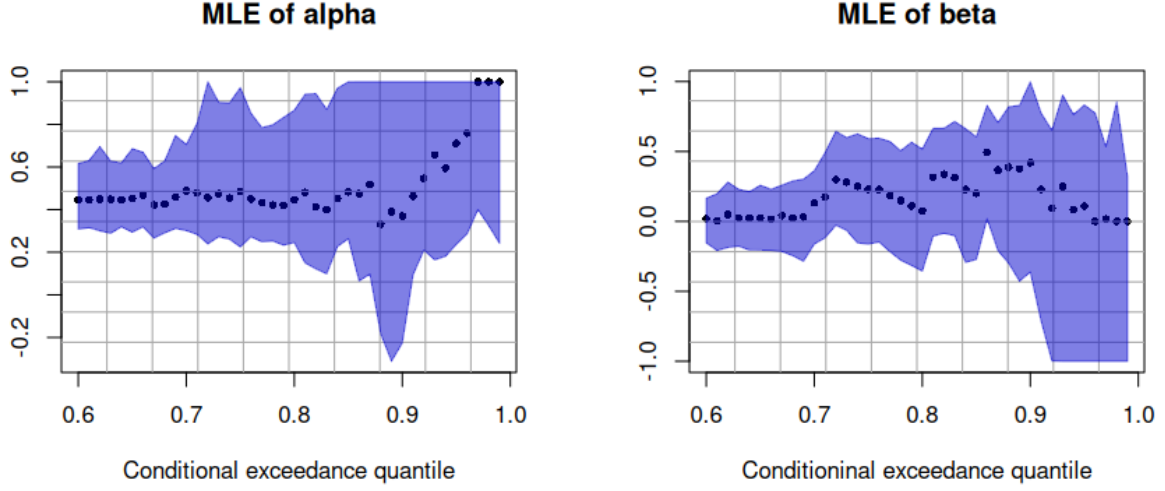


Figure SM7: Threshold stability plots for estimates of the conditional extremes parameters α and β when fitted to $(H_s, S_e) | \{H_s > v\}$, for a range of values of the conditioning threshold $v > 0$. The estimates of α and β are given on the y-axes and the respective quantile $q_v = \tilde{F}_{H_s}^{-1}(u)$ on the x-axes. Point estimates from original Albany data are given in black, and bootstrapped 95% confidence intervals are shown as a blue region. Stability of parameter estimates above a threshold quantile q_v indicates that v is a suitable choice for conditional model threshold. Following visual analysis of the two panels, we select threshold $v = \tilde{F}_{H_s}^{-1}(0.6)$ for joint modelling of H_s and S_e .

SM4.2. Extreme value density estimation supporting the discussion of Section 4.3.4 of the main text.

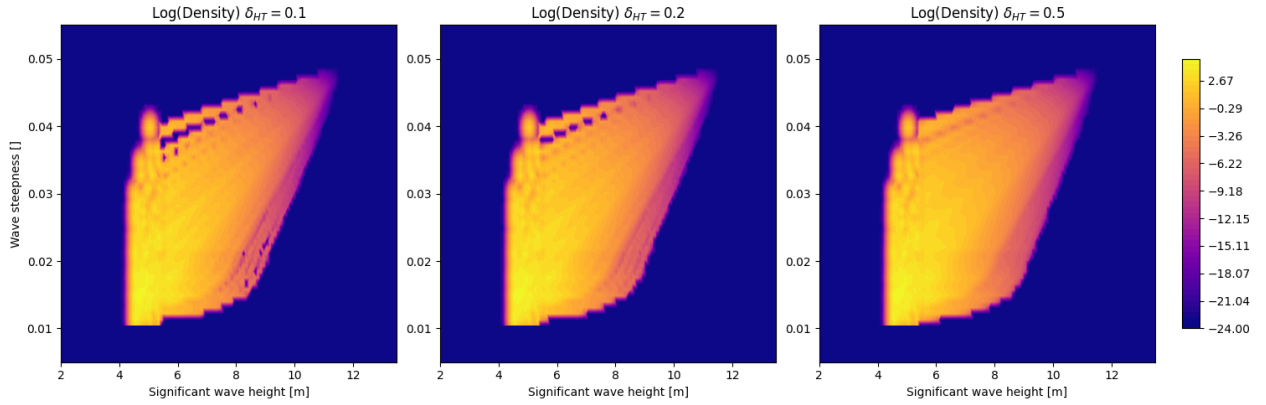


Figure SM8: Sensitivity analysis of estimated joint density $\hat{f}_{\mathbf{X}}$ of $\mathbf{X} = (H_s, S_e)$ with conditional extremes smoothing parameter δ_{HT} . We aim to obtain the smallest value of δ_{HT} which eliminates ‘gaps’ in the extrapolated region. Following visual inspection of the three panels and Figure 11 of the main article, we take $\delta_{HT} = 0.4$.

SM4.3. Non-linear harmonic response simulation

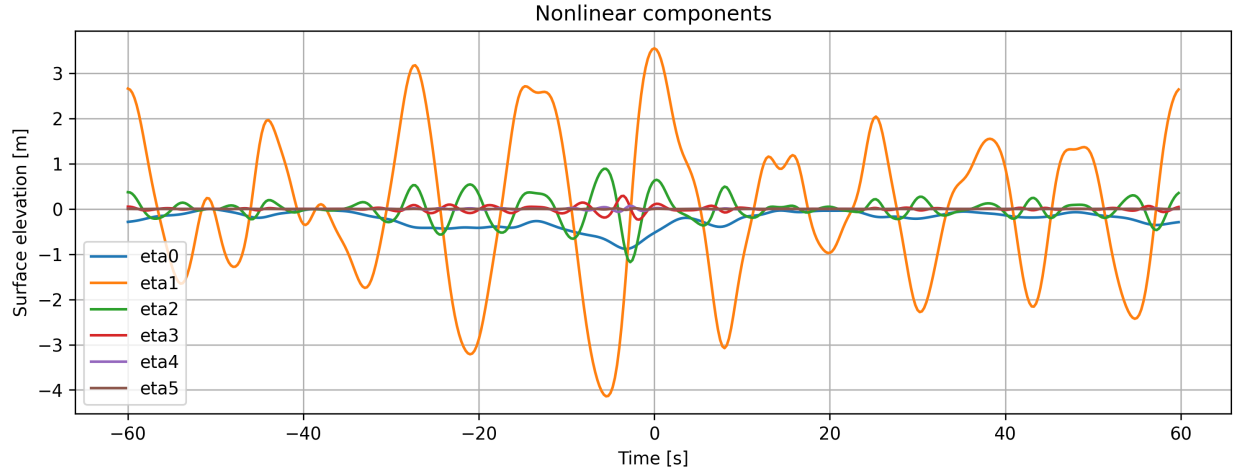


Figure SM9: Harmonic signals constructed using the method of Orszaghova et al. (2025), from a linear surface elevation input. The 0-5th order harmonics are shown. To support the discussion in Section 4.4.2.

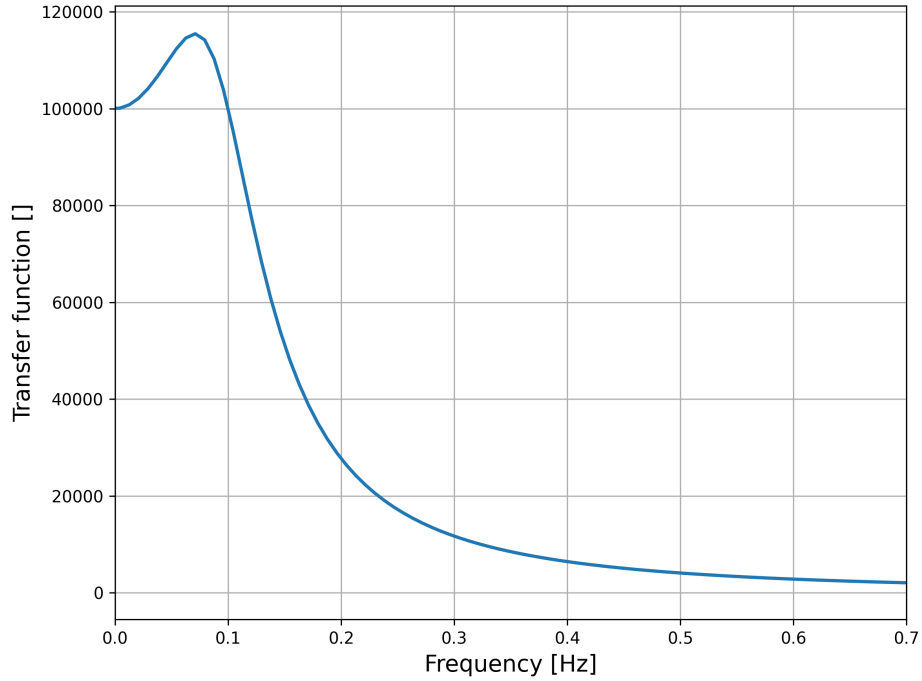


Figure SM10: Transfer function of a damped harmonic oscillator (49), used in the case study of Section 4.4 of the main text. The transfer function is plotted against input frequency [Hz], with resonant frequency $f_0 = 1/10$. The output of the transfer function is assumed unit-less for our case study.

SM5. Supplementary results to the AGE results of Section 4.7 of the main text

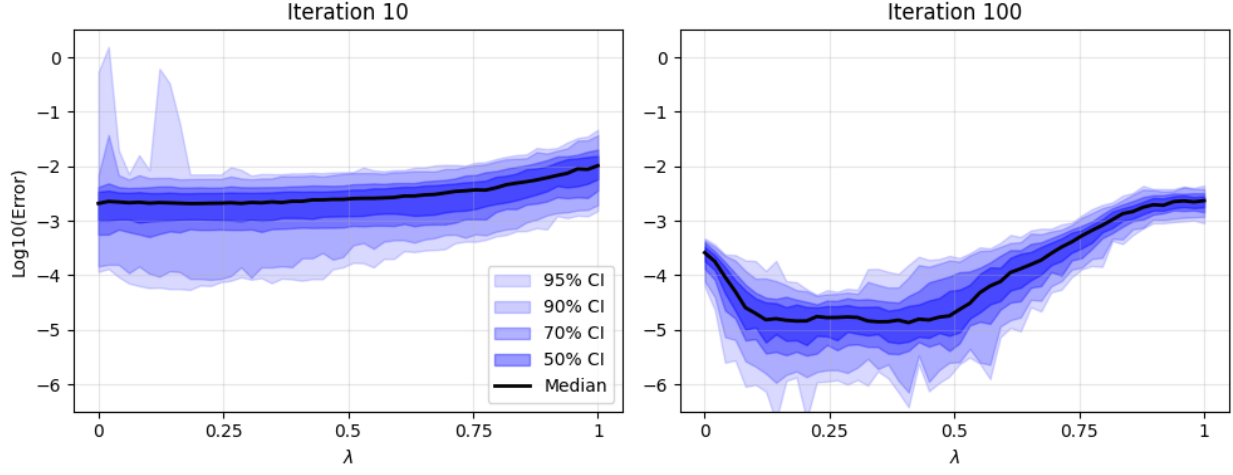


Figure SM11: Log-scale absolute error Δ_{GP} of the GP probability estimate $\hat{p}_{GP}$ at specified iterations, for emulator (10) trained using $U^{(2)}$ over the range of weight $\lambda \in [0.01, 0.99]$ for the TFNV scenario. At iteration 100, the weight that minimises median error is $\lambda^* = 0.41$. To be compared with Figure 14 of the main text.

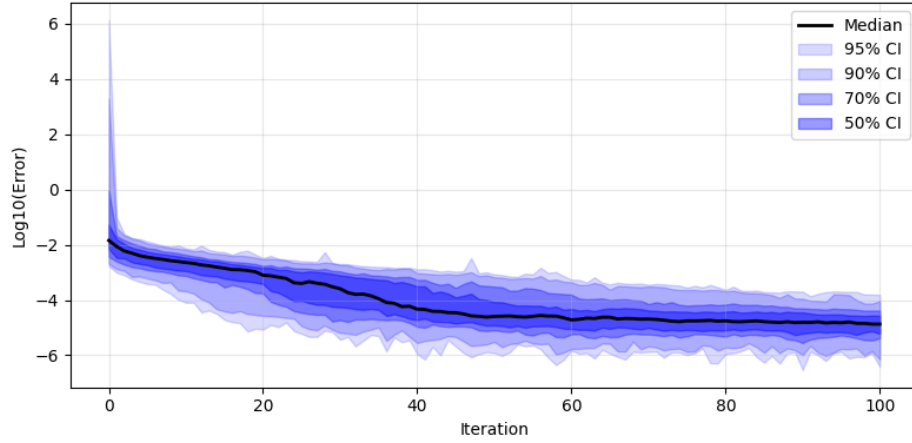


Figure SM12: Distribution of log-scale absolute error Δ_{GP} in the GP probability estimate with respect to iteration, trained using $U^{(2)}$ with $\lambda = \lambda^*$. The trend in median error is indicated in black, with various confidence intervals shown in blue. To be compared with Figure 15 of the main text.