

A sequential sampling strategy for extreme event statistics in nonlinear dynamical systems

Mustafa A. Mohamad and Themistoklis P. Sapsis

Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge MA 02139

April 20, 2018

Abstract

We develop a method for the evaluation of extreme event statistics associated with nonlinear dynamical systems, using a small number of samples. From an initial dataset of design points, we formulate a sequential strategy that provides the ‘next-best’ data point (set of parameters) that when evaluated results in improved estimates of the probability density function (pdf) for a scalar quantity of interest. The approach utilizes Gaussian process regression to perform Bayesian inference on the parameter-to-observation map describing the quantity of interest. We then approximate the desired pdf along with uncertainty bounds utilizing the posterior distribution of the inferred map. The ‘next-best’ design point is sequentially determined through an optimization procedure that selects the point in parameter space that maximally reduces uncertainty between the estimated bounds of the pdf prediction. Since the optimization process utilizes only information from the inferred map it has minimal computational cost. Moreover, the special form of the metric emphasizes the tails of the pdf. The method is practical for systems where the dimensionality of the parameter space is of moderate size, i.e. order $\mathcal{O}(10)$. We apply the method to estimate the extreme event statistics for a very high-dimensional system with millions of degrees of freedom: an offshore platform subjected to three-dimensional irregular waves. It is demonstrated that the developed approach can accurately determine the extreme event statistics using limited number of samples.

Keywords Extreme events | Adaptive sampling | Sequential experimental design

Understanding the statistics of extreme events in dynamical systems of high complexity is of vital importance for reliability assessment and design. We formulate a new method to pick samples optimally so that we have rapid convergence of the full statistics (complete probability distribution) of a quantity of interest, including the tails that describe extreme events. This is important for large scale problems in science and engineering where we desire to predict the statistics of relevant quantities but can only afford a limited number of simulations or experiments due to their cost. We demonstrate our approach in a hydromechanical system with millions of degrees of freedom where only 10-20 carefully selected samples can lead to accurate approximation of the extreme event statistics.

1 Introduction

For many natural and engineering systems, extreme events, corresponding to large excursions, have significant consequences and are important to predict. Examples include extreme economic events, such as credit shocks [8], rogue waves in the ocean [12], and extreme climate events [11]. Extreme events ‘live’ in the tails of a probability distribution function (pdf), thus it is critical to quantify the pdf many standard deviations away from the mean. For most real-world problems, the underlying processes are far too complex to enable estimation of the tails through direct simulations or repeated experiments. This is a result of the low probabilities of extreme events, which necessitates a large number of experiments or ensembles to resolve their statistics. For random dynamical systems with inherently nonlinear dynamics (expressed through intermittent events, nonlinear energy transfers, broad energy spectrum, and large intrinsic dimensionality) we are usually limited to a few ensemble realizations.

The setup in this article involves a stochastic dynamical system that depends on a set of random parameters with known probability distribution. We assume that the dimensionality of the random parameters is or can be reduced to a moderate size $\mathcal{O}(10)$. Because of the inherent stochastic and transient character of extreme

responses, it is not sufficient to consider the dynamical properties of the system independently from the statistical characteristics of solutions. A statistical approach to this problem has important limitations, such as requiring various extrapolation schemes due to insufficient sample numbers (see extreme value theorems [18]). Another strategy is large deviations theory [24, 7], a method for the probabilistic quantification of large fluctuations in systems, which involves identifying a large deviations principle that explains the least unlikely rare event. While applied to many problems, for complex systems estimating the rate function can be very costly and the principle does not characterize the full probability distribution. The resulting distributions via such approaches cannot always capture the non-trivial shape of the tail, dictated by physical laws in addition to statistical characteristics. On the other hand, in a dynamical systems approach there are no sufficiently generic efficient methods to infer statistical information from dynamics. For example, the Fokker-Planck equation [22] is challenging to solve even in moderate to low dimensions [14]. To this end, it is essential to consider blended strategies. The utilization of combined dynamic-stochastic models for the prediction of extreme events have also been advocated and employed in climate science and meteorology by others [9, 13, 2]. In [15, 16] a probabilistic decomposition of extreme events was utilized to efficiently characterize the probability distribution of complex systems, which considered both the statistical characteristics of trajectories and the mechanism triggering the instabilities (extreme events). While effective, the proposed decomposition of intermittent regimes requires explicit knowledge of the dynamics triggering the extremes, which may not be available or easily determined for arbitrary dynamical systems.

We formulate a sequential method for capturing the statistics of an observable that is, for example, a functional of the state of a dynamical system or a physical experiment. The response of the observable is modeled using a machine learning method that infers the functional form of the quantity of interest by utilizing only a few strategically sampled numerical simulations or experiments. Combining the predictions from the machine learning model, using the Gaussian process regression framework, with available statistical information on the random input parameters, we formulate an optimization problem that provides the next-best or most informative experiment that should be performed to maximally reduce uncertainty in the pdf prediction of extreme events (tails of the distribution) according to a proposed ‘distance’ metric. To account for tail features the metric utilize a logarithmic transformation of the pdfs, which is similar to the style of working on the rate function in the large deviation principle. The optimization process relies exclusively on the inferred properties on the parameter-to-observation map and no additional simulations are required in the optimization. For the optimization problem to be practically solvable we require the parameter space to be of moderate size, on the order $\mathcal{O}(10)$. The proposed method allows us to sequentially sample the parameter space in order to rapidly capture the pdf and, in particular, the tails of the distribution of the observable of interest.

2 Problem setup and method overview

Consider a dynamical system with state variable $u \in \mathbb{R}^n$,

$$\frac{du}{dt} = g(t, u; \theta(\omega)), \quad \omega \in \Omega, \quad (1)$$

where Ω is the sample space in an appropriate probability space (we denote the density of the random variable X by f_X and its cumulative density by F_X). The random variable $\theta : \Omega \rightarrow U \subset \mathbb{R}^m$ parameterizes sources of uncertainty, such as stochastic forcing terms or system parameters with *a priori* known distribution f_θ . For fixed $\omega \in \Omega$, the response u is a deterministic function in time. We are interested in estimating the pdf \hat{f} of a scalar quantity of interest or observable $q \in \mathbb{R}$ given by

$$q = \hat{T}(\theta) \triangleq \mathcal{F}(u) + \varepsilon \quad (2)$$

where $\hat{T} : U \subset \mathbb{R}^m \rightarrow \mathbb{R}^1$ is a continuous *parameter-to-observation* map, \mathcal{F} is an arbitrary functional of u , and ε is some observational or numerical noise term, which we take as zero, without loss of generality. In our setup the unknown parameter-to-observation-map \hat{T} is *expensive* to evaluate, representing, for example, a large scale numerical simulation or a costly physical experiment, and so we desire to minimize the number of evaluations of this mapping. Note that the true statistics of the random variable q induced by \hat{T} is,

$$\hat{f}(s) = \frac{d}{ds} \hat{F}(s) = \frac{d}{ds} \mathbb{P}(\hat{T}(\theta) \leq s) = \frac{d}{ds} \int_{A(s)} f_\theta(\theta) d\theta, \quad (3)$$

where $A(s) = \{\theta \in U : \hat{T}(\theta) \leq s\}$ and \hat{F} is the cumulative density function of q .

Our objective is to estimate the statistics, especially non-Gaussian features, of the observable q , i.e. the pdf of the random variable induced by the mapping $\hat{T}(\theta)$ which we denote by \hat{f} :

Consider the quantity of interest q with pdf \hat{f} induced by the unknown mapping $\hat{T}(\theta)$, where θ is a random valued parameter with known pdf f_θ . Given a dataset $\mathcal{D}_{n-1} = \{(\theta_i, \hat{T}(\theta_i))\}_{i=1}^{n-1}$ of size $n-1$, so that the estimated distribution of q using a learning algorithm from this dataset is $f_{\mathcal{D}_{n-1}}$, determine the next input parameter θ such that when the map is evaluated at this new sample the error between $f_{\mathcal{D}_n}$ and \hat{f} is minimized, placing special emphasis on the tails of the distribution where $|\hat{T}(\theta)|$ is large.

If we consider the augmented θ -parameterized dataset $\mathcal{D}_n(\theta) = \mathcal{D}_{n-1} \cup \{(\theta, \hat{T}(\theta))\}$ for all $\theta \in U$, and denote the learned density by pdf $f_{\mathcal{D}_n(\theta)}$, the next parameter θ_n is obtained by minimizing a distance metric between two probability distributions, $\theta_n = \arg \min_\theta Q(f_{\mathcal{D}_n(\theta)}, \hat{f})$

Determining the next-best experiment θ_n should not involve the expensive-to-evaluate map \hat{T} nor \hat{f} . The main computational savings of our proposed method involves (1) using a inexpensive-to-compute surrogate model to replace $\hat{T}(\theta)$ appearing in the θ -parametrized dataset $\mathcal{D}_n(\theta)$ and using a version of the distance metric without explicit dependence on \hat{f} . Here we utilize Gaussian process regression (GPR), as the learning algorithm to construct the surrogate for \hat{T} . Using the posterior distribution of the inferred map through the GPR scheme, we estimate the pdf for the quantity of interest as well as the pdfs that correspond to the confidence intervals on the map. The distance metric Q , is then based on minimization of the logarithmic transform of the pdfs that correspond to the map upper and lower confidence intervals from the posterior variance, which does not involve \hat{f} . This optimization problem provides the ‘next-best’ point in a sequential fashion.

The overall aim is to accurately capture the tail statistics of \hat{f} through a minimum number of observations of q . Note that the resulting method should not need to densely sample all regions in U , since not all regions have significant probability (f_θ may be negligible) or importance ($|\hat{T}|$ may be small). The formulated sampling strategy should accurately predict the tail region of the pdf taking into account both the magnitude of the map $|\hat{T}|$ and the value of the probability of the sample θ , see Fig. 1.

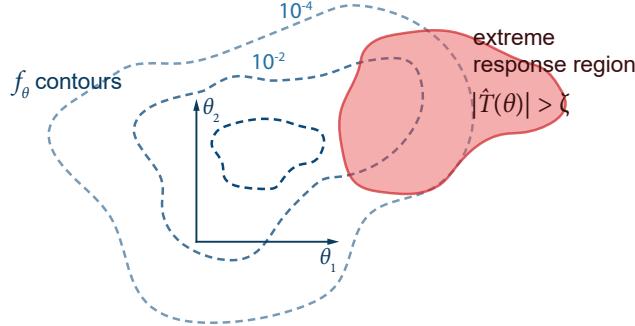


Figure 1 Areas with large probability in θ are not necessarily associated with regions where $|\hat{T}|$ is large. The proposed criterion focuses on sampling regions where both the probability and the magnitude of $|\hat{T}|$ are significant.

3 Method description

An important component of our proposed method is the construction of an inexpensive surrogate for the map \hat{T} . We utilize Gaussian process regression (GPR) as our learning method. GPR considers the function as a Gaussian process, in terms of a distribution in function space (see appendix A and numerous references, such as [21]). An important property of GPR is that the posterior distribution is a Gaussian process with an explicit mean and kernel function. The variance of the posterior can be used as a proxy for the error or uncertainty of the prediction, which we utilize along with f_θ to guide selection of the next sample point.

We learn the parameter-to-observation map $\hat{T}: U \subset \mathbb{R}^m \rightarrow \mathbb{R}$ using GPR from the dataset $\mathcal{D}_{n-1} = \{(\hat{\theta}_i, \hat{T}(\hat{\theta}_i))\}_{i=1}^{n-1}$. The GPR method then provides analytic expressions for the mean $T_{n-1}(\theta)$ and kernel $k_{n-1}(\theta, \theta')$ for the posterior distribution of the Gaussian random function $\mathbf{T}_{n-1}(\theta) \sim \text{GP}(T_{n-1}(\theta), k_{n-1}(\theta, \theta'))$ (see appendix A)

for the expressions). We can then construct the following estimate for the pdf \hat{f} using the posterior mean T_{n-1} of the learned surrogate model \mathbf{T}_{n-1} :

$$f_{n-1}(s) = \frac{dF_{n-1}}{ds} = \frac{d}{ds} \int_{A_{n-1}(s)} f_\theta(\theta) d\theta, \quad (4)$$

where $A_{n-1}(s) = \{\theta \in U : T_{n-1}(\theta) \leq s\}$ and F_{n-1} is the cumulative distribution function.

We now formulate the optimization problem for the next sample point θ^* . Consider the augmented θ -parameterized dataset $\tilde{\mathcal{D}}_n(\theta) = \mathcal{D}_{n-1} \cup \{(\theta, T_{n-1}(\theta))\}$, which approximates $D_n(\theta) = \mathcal{D}_{n-1} \cup \{(\theta, \hat{T}(\theta))\}$ by using the GPR mean T_{n-1} instead of \hat{T} . Then let $\tilde{\mathbf{T}}_n(\theta; \theta^*) \sim \text{GP}(\tilde{T}_n(\theta; \theta^*), \tilde{k}_n(\theta, \theta'; \theta^*))$ denote the random function trained on the augmented dataset $\tilde{\mathcal{D}}_n(\theta^*)$. The pdf of the random variable $\hat{T}(\theta)$, where $\theta \sim f_\theta$, is now replaced by a θ^* -parameterized random probability measure induced by the Gaussian random function $\tilde{\mathbf{T}}_n(\theta; \theta^*)$, where $\theta \sim f_\theta$ and θ^* is a sample point. Note, that the mean $\tilde{T}_n(\theta; \theta^*)$ of $\tilde{\mathbf{T}}_n(\theta; \theta^*)$ in the θ -parameterized dataset is identical to $T_{n-1}(\theta)$ for all θ^* , since the prediction of the value of the map at the sample point is given by the posterior mean $T_{n-1}(\theta)$ at iteration $n - 1$.

The proposed criterion Q is then based on minimization of a distance metric between the pdfs of the confidence bounds of \mathbf{T}_n . Specifically, let $\tilde{f}_n^\pm(\cdot; \theta^*)$ denote the pdfs of the two random variables $\tilde{T}_n(\theta; \theta^*) \pm \alpha \tilde{\sigma}_n(\theta; \theta^*)$, where $\tilde{\sigma}_n(\theta) = \tilde{k}_n(\theta, \theta)$, which are the upper and lower bounds of the confidence interval based on the α -scaled standard deviation of the posterior distribution. The pdfs corresponding to the confidence bounds are explicitly given by

$$\tilde{f}_n^\pm(s; \theta^*) = \frac{d}{ds} \tilde{F}_n^\pm(s; \theta^*) = \frac{d}{ds} \int_{A_n^\pm(s; \theta^*)} f_\theta(\theta) d\theta, \quad (5)$$

where $A_n^\pm(s) = \{\theta \in U : \tilde{T}_n(\theta; \theta^*) \pm \alpha \tilde{\sigma}_n(\theta; \theta^*) \leq s\}$. We employ the 95% interval bounds, so that the standard deviation is scaled by a factor $\alpha = 1.96$. The cdf \tilde{F}^+ corresponding to the upper confidence bound $\tilde{T}_n(\theta; \theta^*) + \alpha \tilde{\sigma}_n(\theta; \theta^*)$, is a lower bound for \tilde{F}_n , and we have the relation $\tilde{F}^+(s) \leq \tilde{F}(s) \leq \tilde{F}^-(s)$, for all s . Note, although the map mean of the Gaussian random function $\tilde{\mathbf{T}}_n(\theta; \theta^*)$ based on the θ -parameterized dataset is identical to the posterior mean $T_{n-1}(\theta)$ at iteration $n - 1$, the value of the variance $\tilde{\sigma}_n(\theta; \theta^*)$ now vanishes at the sample point: $\tilde{\sigma}_n(\theta^*; \theta^*) = 0$.

The distance metric we propose for the selection of the next sample point is given by

$$Q(\tilde{f}_n^+(\cdot; \theta), \tilde{f}_n^-(\cdot; \theta)) \triangleq \frac{1}{2} \int |\log(\tilde{f}_n^-(s; \theta)) - \log(\tilde{f}_n^+(s; \theta))| ds, \quad (6)$$

where the integral is computed over the intersection of the two domains that the pdfs are defined over. The next sample point is then determined by solving the optimization problem:

$$\theta_n = \arg \min_{\theta} Q(\tilde{f}_n^+(\cdot; \theta), \tilde{f}_n^-(\cdot; \theta)). \quad (7)$$

This is a L_1 based metric of the logarithmic transform of the pdfs. The logarithmic transform of the densities in the criterion effectively emphasizes extreme and rare events, as we explicitly show in Theorem 1. The computational savings comes from the construction of a criterion Q that avoids \hat{f} and instead uses \tilde{f}_n^\pm , which involves evaluating the GPR emulator T_{n-1} (inexpensive) and an additional GPR prediction.

We make a few comments on the optimization problem and the sequential algorithm. The sequential strategy is summarized in pseudocode in appendix C. For the optimization problem, we utilize a derivative free method, specifically a particle swarm optimizer. The integrations for the pdfs are computed explicitly from the definition of the Lebesgue integral, by partitioning the co-domain of map T . This is much more efficient compared with a Monte-Carlo approach that would result in a very expensive computational task, as long as the dimensionality of the parameter space is low. For high-dimensional parameter spaces, the computational cost of the integration can become prohibitive and in such cases an order reduction in the parameter space should be first attempted, or alternatively an importance sampling algorithm could be used to compute the integral. In the numerical problems, the optimization problem is practically solvable for low dimensional parameter spaces $\mathcal{O}(5)$. The starting design plan size n_s , if not already provided, should be small; a Latin hypercube based sampling plan should be utilized for this purpose. We also recommend as a pre-training period to process a small number of iterations using a

metric without the logarithmic transform of the densities, such as the following L_2 based metric

$$Q'(\tilde{f}_n^+(\cdot; \theta), \tilde{f}_n^-(\cdot; \theta)) \triangleq \frac{1}{2} \int |\tilde{f}_n^-(s; \theta) - \tilde{f}_n^+(s; \theta)|^2 ds, \quad (8)$$

in order to capture the main probability mass (low order moments) before utilizing the proposed metric that emphasizes extreme and rare events. In addition, it is not necessary to retrain the GPR hyperparameters after every iteration, which can remain fixed after being calibrated from a few iterations. Updating the Gaussian process emulator after the addition of new data points can be done in $\mathcal{O}(n^2)$ if the hyperparameters are fixed, otherwise the GPR emulator must be performed anew in $\mathcal{O}(n^3)$ (see appendix B)¹.

3.1 Asymptotic behavior

The first theoretical result relates to the convergence of the proposed method to the true pdf as the number of samples goes to infinity (see appendix D for the proof). The second result shows that the asymptotic form of the criterion is given by (see appendix D for the proof):

Theorem 1. *Let $T_n(\theta)$ and $\sigma_n(\theta)$ from the GPR scheme be sufficiently smooth functions of θ . The asymptotic behavior of Q for large n (ensuring small σ) and small $\|\nabla\sigma\|/\|\nabla T\|$ is given by*

$$\tilde{Q}_n \triangleq \frac{1}{2} \int |\log f_n^+(s) - \log f_n^-(s)| ds \quad (9)$$

$$\approx \int \left| \frac{\frac{d}{ds} \mathbb{E}(\sigma_n(\theta) \cdot \mathbf{1}_{T_n(\theta)=s})}{f_n(s)} \right| ds, \quad (10)$$

where \mathbb{E} denotes the expectation over the probability measure \mathbb{P}_θ .

Note that the pdf in the denominator under the integral in 30 is a direct implication of our choice to consider the difference between the logarithms of the pdfs in the optimization criterion. The pdf of the parameter-to-observation map f_n in the denominator of the integrand guarantees that even values of the parameter θ , where the probability is low (rare events) are sampled. This point is clearly demonstrated by the following corollary (proof in appendix D).

Corollary 1. *Let $\theta: \Omega \rightarrow [u_1, u_2] \subset \mathbb{R}$ be a one-dimensional random variable and in addition to the assumptions of Theorem 1 we assume that $T'_n(\theta)$ is monotonically increasing function. Then, the asymptotic value of \tilde{Q}_n for large n has the following property:*

$$\tilde{Q}_n \gtrsim \left| \int_U \sigma_n(\theta) d(\log f_\theta(\theta)) + \int_U \sigma_n(\theta) d(\log T'_n(\theta)) + \sigma_n(u_2) - \sigma_n(u_1) \right|. \quad (11)$$

Therefore for large value of n , \tilde{Q}_n bounds (within higher order error) a quantity that consists of boundary terms and two integral terms involving sampling the function $\sigma(\theta)$ over the contours of $\log f_\theta(\theta)$ and $\log T'_n(\theta)$. Consider the first term on the right, which can be discretized as:

$$\int_U \sigma(\theta) d(\log f_\theta(\theta)) = \lim_{N \rightarrow \infty} \sum_{i=1}^N \Delta z \sum_{\{\theta: \log f_\theta(\theta)=z_i\}} \sigma(\theta), \quad (12)$$

where $\{z_i\}_{i=1}^N$ is an equipartition of the range of $\log f_\theta$. This summation is summarized in Fig. 2 (left). The way the integral is computed guarantees that the integrand $\sigma_n(\theta)$ will be sampled even in locations where the pdf f_θ has small value (rare events) or else the criterion would not converge.

On the other hand, if we had instead chosen a criterion that focused directly on the convergence of low order statistics for $\sigma_n(\theta)$, such as the same criterion, but *without the logarithmic transformation of the densities*, we would have

$$\tilde{Q}'_n \triangleq \frac{1}{2} \int |f_n^+(s) - f_n^-(s)| ds \geq \left| \int_U \sigma_n(\theta) d(F_\theta(\theta)) \right|,$$

where F_θ is the cumulative distribution function of θ . The corresponding integration is shown in Fig. 2 (right). In such case, sampling the function $\sigma_n(\theta)$ in regions of high probability would be sufficient for the criterion to converge. However, such a strategy would most likely lead to large errors in regions associated with rare events since the sampling will be sparse.

¹For low-dimensional θ , since we presume the dataset size is small, the cost difference may be negligible.

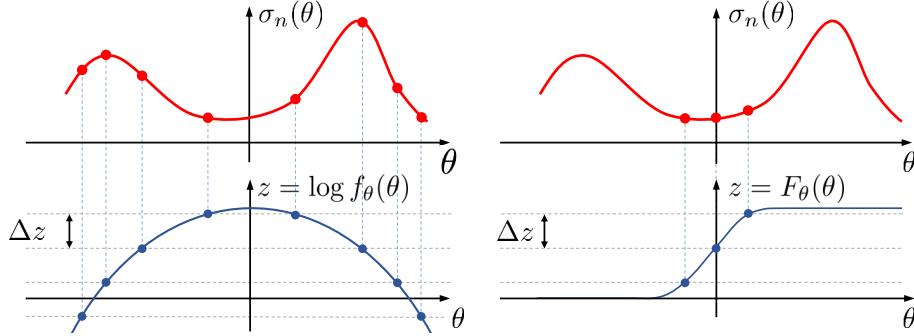


Figure 2 (left) Integration of $\sigma_n(\theta)$ over contours of $\log f_\theta(\theta)$ implies sampling of $\sigma_n(\theta)$ in low probability regions of θ . (right) On the other hand, low-order moments of $\sigma_n(\theta)$ rely only on values of $\sigma_n(\theta)$ close to high probability regions of θ , thus rare events are not sampled sufficiently.

4 Applications

We illustrate the proposed algorithm to two problems. The first example consists of a nonlinear oscillator stochastically forced by a colored noise process; this application, serves to illustrate the main ideas of the proposed method. The second application, involving three-dimensional hydrodynamic wave impacts on an offshore platform (a system with millions of degrees of freedom) showcases the applicability of the proposed method to real world setups where computation of extreme event statistics using traditional approaches are prohibitively expensive, since the simulation times of experimental runs are on the order of several hours.

4.1 Nonlinear oscillator driven by correlated stochastic noise

Consider the nonlinear oscillator,

$$\ddot{x} + \delta\dot{x} + F(x) = \zeta(t), \quad (13)$$

forced by a stationary, colored noise with correlation function $C(\tau) = \sigma_\zeta^2 e^{-\tau^2/2\ell_\zeta^2}$ and the nonlinear restoring term given by

$$F(x) = \begin{cases} \alpha x, & 0 \leq |x| \leq x_1 \\ \alpha x_1, & x_1 < |x| \leq x_2 \\ \alpha x_1 + \beta(x - x_2)^3, & x_2 \leq |x|. \end{cases} \quad (14)$$

Since the system is stochastically forced it is necessary to use an expansion to obtain a parameterization in terms of a finite number of random variables. We use a Karhunen-Loève expansion (see appendix E) to obtain

$$\ddot{x}(t) + \delta\dot{x}(t) + F(x(t)) = \sum_{i=1}^m \theta_i(\omega) e_i(t), \quad t \in [0, T], \quad (15)$$

which is truncated to a suitable number m . For illustration, we take our quantity of interest as the average value of the response, so that the parameter-to-observation map is defined by $T(\theta) \triangleq \overline{x(t; \theta)} = \frac{1}{T} \int_0^T x(t; \theta) dt$.

We consider a three term truncation $m = 3$. The system parameters are given by $\delta = 1.5$, $\alpha = 1.0$, $\beta = 0.1$, $x_1 = 0.5$, $x_2 = 1.5$ and the forcing parameters are $\sigma_\zeta = 4$ and $\ell_\zeta = 0.1$, with $t \in [0, 25]$. For comparisons the exact pdf is obtained by sampling the true map from 64000 points on a $40 \times 40 \times 40$ grid. In Fig. 3 we illustrate the sampling as determined by the proposed algorithm in addition to the L_1 log error between the exact pdf and the GPR mean prediction. In these simulations we start from a dataset of 6 points selected according to a Latin-Hypercube (LH) design. In order to capture the main mass of the pdf, before focusing on the tails of the distribution, we perform 12 iterations using the d_{L_2} error metric before moving on to the criterion using the d_{L_1} error of the logarithms of the pdfs. Observe the unique shape that the sampling algorithm has identified in θ space, which spans regions in θ associated with finite probability and large values of q . Fig. 4 demonstrates the progression of the estimated pdf as a function of the iteration count. Even after only 100 samples we have already captured the qualitative features of the exact pdf and have very good quantitative agreement.

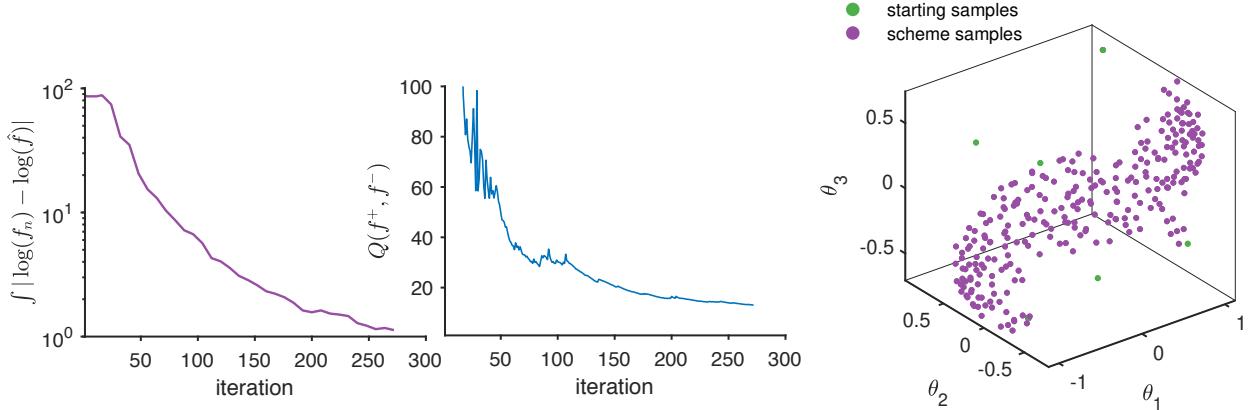


Figure 3 (Right) A scatter plot of the algorithm sampling of the parameter space (green points denote the initial random LH samples). (Left) The corresponding L_1 error of the logarithmic transform of the pdf between the GPR mean and truth and (middle) value of the criterion in 5 as a function of the iteration number.

We have explored the convergence properties of the algorithm and in Fig. 5 we compare the proposed sampling method to space-filling Latin Hypercube sampling. The LH strategy is not iterative and thus must be started anew, which puts the LH sampling at a large disadvantage. Nonetheless, this serves as a benchmark to a widely used reference method for the design of experiments due to its simplicity. In the figure, the purple curve represents the mean LH design error and the shaded region represents the standard deviation about the mean, which are computed by evaluating 250 number of random LH designs per fixed dataset size. Even considering the variance of the LH curve, the proposed algorithm under various parameters (initial dataset size or number of ‘core’ iterations where the Q criterion uses the L_2 metric) is observed to outperform the LH strategy by nearly an order of magnitude in the L_1 error of the logarithm of the pdfs. This demonstrates the favorable properties of the proposed sampling strategy for accurately estimating the tail of target distribution.

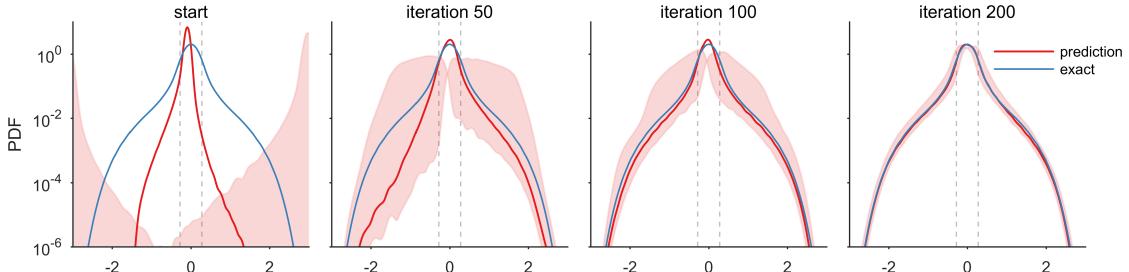


Figure 4 Progression of the pdf estimation as we iteratively sample more points. We shade the region between the pdfs f^\pm in red purely as a visualization of the convergence of the pdfs f^\pm as more points are sampled. Dashed vertical lines denote one standard deviation.

4.2 Hydrodynamic forces and moments on an offshore platform

Here we apply the sampling algorithm to compute the probability distributions describing the loads on an offshore platform in irregular seas. The response of the platform is quantified through direct, three-dimensional numerical simulations of Navier-Stokes utilizing the smoothed particle hydrodynamics (SPH) method [5] (Fig. 6). Our numerical setup parallels that of a physical wave tank experiment and consists of a wave maker on one end and a sloping ‘beach’ on the other end of the tank to quickly dissipate the energy of incident waves and avoid wave reflections. Further details regarding the simulations are provided in appendix F.

Wind generated ocean waves are empirically described by their energy spectrum. Here, we consider irregular seas with JONSWAP spectral density (see appendix F for details and parameters). While realizations of the random waves have the form of time series, an alternative description can be obtained by considering a sequence

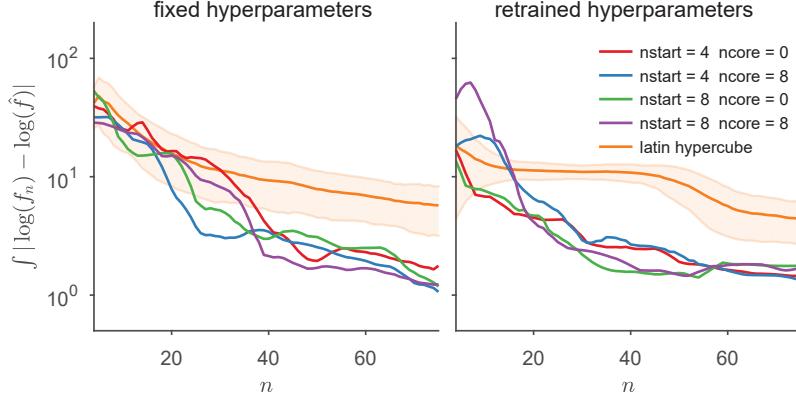


Figure 5 Comparison of the errors between LH sampling and proposed algorithm under different parameters measured against the exact pdf for the case where $m = 2$ in 15. The parameter `ncore` is the number of ‘core’ iterations performed according to a L_2 metric and `nstart` is initial dataset size (where the points are sampled from an LH design).

of primary wave groups, each characterized by a random group length scale L and height A (see e.g. [3]). This formulation allows us to describe the input space through just two random variables (much fewer than what we would need with a Karhunen-Loeve expansion). Following [3] we describe these primary wavegroups by the representation $u(x) = A \operatorname{sech}(x/L)$, which is an explicit parameterization in terms of L and A . Thus, L and A correspond to θ_1 and θ_2 in the notation of Eq. section 2. The statistical characteristics of the wave groups associated with a random wave field (such as the one given by the JONSWAP spectrum) can be obtained by applying the scale-selection algorithm described in [4]. Specifically, by generating many realizations consistent with the employed spectrum we use a group detection algorithm to identify coherent group structures in the field along with their lengthscale and amplitude (L and A). This procedure provides us with the empirical probability distribution f_θ of the wave field and thus a nonlinear parametrization of the randomness in the input process. The quantities of interest in this problem are the forces and moments acting on the platform. The incident wave

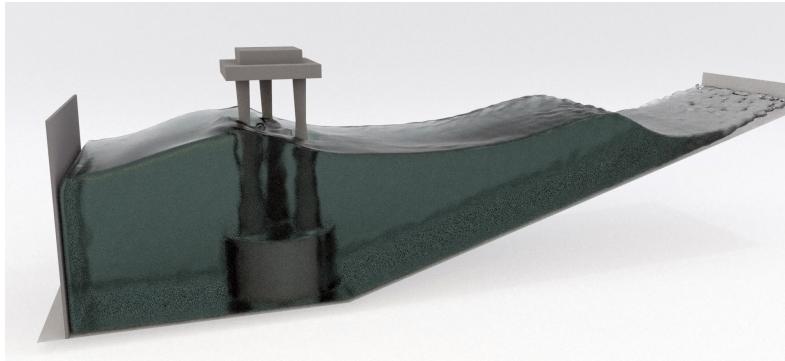


Figure 6 SPH simulation at $t = 103.5$ s with $\theta_1 = 4.63$ and $\theta_2 = 0.662$.

propagates in the x direction and as such we consider the pdf of the force in the x direction F_x and the moment M_y about the bottom-center of the platform:

$$q_f = \max_{t \in [0, T]} |F_x(t)| \quad \text{and} \quad q_m = \max_{t \in [0, T]} |M_y(t)|. \quad (16)$$

In Fig. 6 we show the results of the progression of density prediction for the force variable. In these experiments we begin by arbitrary selecting 4 initial sample points from a Latin Hypercube sampling strategy. Next, we perform 4 iterations using the L_1 distance metric to quickly capture the main mass of the distribution before focusing on the distribution away from the mean that utilizes the L_1 metric of the logarithmic of the pdf. The lightly shaded red region in the pdf plots is a visualization of the uncertainty in the pdf, obtained by sampling the GPR prediction and computing the pdf for 200 realizations and then computing the upper (lower) locus of the

maximum (minimum) value of the pdf at each value. The figures demonstrate that with 15 (i.e 14 total sample points) iterations (together with the 4 samples in the initial configuration) we are able to approximate the pdf to good agreement with the ‘exact’ pdf, which was computed from a densely sampled grid. In appendix F we also present the sampled map where it can be seen that the algorithm selects points associated with large forces and non-negligible probability of occurrence. In the same figures results for the momentum and an additional spectrum are included. Note, for this problem the GP regression operates on the logarithm of the observable because the underlying function is always positive.

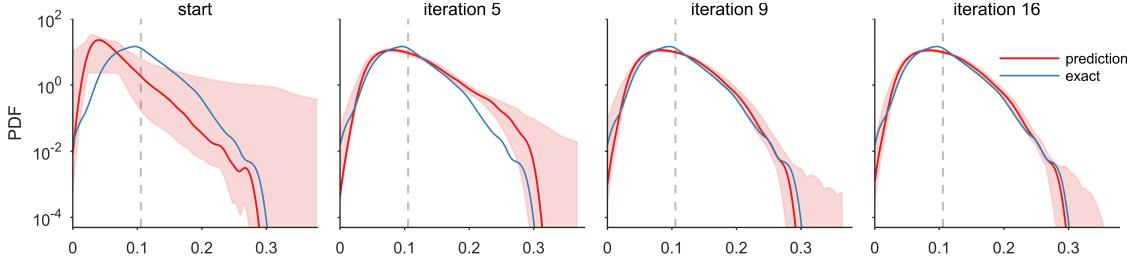


Figure 7 Progression for pdf density prediction for the force variable.

5 Conclusions

We developed and analyzed a computational algorithm for the evaluation of extreme event statistics associated with nonlinear dynamical systems that depend on a set of random parameters. The algorithm is practical even for very high dimensional systems but with parameter spaces of moderate dimensionality and it provides a sequence of points that lead to improved estimates of the probability distribution for a scalar quantity of interest. The criterion for the selection of the next design point emphasizes the tail statistics. We have proven asymptotic convergence of the algorithm and provided analysis for its asymptotic behavior. We have also demonstrated its applicability through two problems, one of them involving a demanding system with millions degrees of freedom.

Acknowledgments

This work has been supported through the ONR grant N00014-15-1-2381, the AFOSR grant FA9550-16- 1-0231, and the ARO grant W911NF-17-1-0306. We are grateful to the referees for providing numerous suggestions that led to important improvements and corrections. We also thank A. Crespo for helpful comments regarding the SPH simulations.

References

- [1] Corrado Altomare, Alejandro J. C. Crespo, Jose M. Domnguez, Moncho Gómez-Gesteira, Tomohiro Suzuki, and Toon Verwaest. Applicability of smoothed particle hydrodynamics for estimation of sea wave impact on coastal structures. *Coastal Engineering*, 96:1–12, 02 2015.
- [2] Nan Chen and Andrew J. Majda. Simple stochastic dynamical models capturing the statistical diversity of el niño southern oscillation. *Proceedings of the National Academy of Sciences*, 114(7):1468–1473, 2017.
- [3] Will Cousins and Themistoklis P. Sapsis. Unsteady evolution of localized unidirectional deep-water wave groups. *Physical Review E*, 91:063204, 2015 06.
- [4] Will Cousins and Themistoklis P. Sapsis. Reduced-order precursors of rare events in unidirectional nonlinear water waves. *Journal of Fluid Mechanics*, 790:368–388, 02 2016.
- [5] A. J. C. Crespo, J. M. Domnguez, B. D. Rogers, M. Gómez-Gesteira, S. Longshaw, R. Canelas, R. Vacondio, A. Barreiro, and O. Garca-Feal. DualSPHysics: Open-source parallel CFD solver based on smoothed particle hydrodynamics (sph). *Computer Physics Communications*, 187:204–216, 02 2015.

- [6] Timothy A. Davis and William W. Hager. Row modifications of a sparse Cholesky factorization. *SIAM Journal on Matrix Analysis and Applications*, 26(3):621–639, 2005.
- [7] Giovanni Dematteis, Tobias Grafke, and Eric Vanden-Eijnden. Rogue waves and large deviations in deep sea. *Proceedings of the National Academy of Sciences*, 2018.
- [8] Jean-Pierre Fouque, George Papanicolaou, Ronnie Sircar, and Knut Sølna. *Multiscale Stochastic Volatility for Equity, Interest-Rate and Credit Derivatives*. Cambridge University Press, 09 2011.
- [9] Christian L. E. Franzke. Extremes in dynamic-stochastic systems. *Chaos*, 27(1):012101, 2017.
- [10] Peter Frigaard, Michael Høgedal, and Morten Christensen. Wave generation theory, 06 1993.
- [11] Andreas Hense and Petra Friederichs. *Wind and precipitation extremes in the Earth's atmosphere*, pages 169–187. Springer Berlin Heidelberg, 2006.
- [12] Paul C. Liu. A chronology of freaque wave encounters. *Geofizika*, 24(1):57–70, 2007.
- [13] Andrew J. Majda. *Introduction to Turbulent Dynamical Systems in Complex Systems*. Springer International Publishing, 2016.
- [14] A. Masud and L. A. Bergman. Solution of the four dimensional Fokker-Planck equation: still a challenge. In *ICOSSAR*, volume 2005, pages 1911–1916, 2005.
- [15] Mustafa A. Mohamad, Will Cousins, and Themistoklis P. Sapsis. A probabilistic decomposition-synthesis method for the quantification of rare events due to internal instabilities. *Journal of Computational Physics*, 322:288–308, 10 2016.
- [16] Mustafa A. Mohamad and Themistoklis P. Sapsis. Probabilistic description of extreme events in intermittently unstable dynamical systems excited by correlated stochastic processes. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):709–736, 01 2015.
- [17] Arvid Naess and Torgeir Moan. *Stochastic Dynamics of Marine Structures*. Cambridge University Press, 2013.
- [18] Mario Nicodemi. *Extreme value statistics*, pages 1066–1072. Springer New York, 2012.
- [19] Athanasios Papoulis and S. Unnikrishna Pillai. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill Education, 4 edition, 2002.
- [20] Grigorios A. Pavliotis. *Stochastic Processes and Applications*, volume 60 of *Texts in Applied Mathematics*. Springer-Verlag New York, 2014.
- [21] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. MIT Press, 2005.
- [22] Kazimierz Sobczyk. *Stochastic Differential Equations*. Mathematics and Its Applications (East European Series). Springer Netherlands, 1991.
- [23] Andrew M. Stuart and Aretha L. Teckentrup. Posterior consistency for gaussian process approximations of bayesian posterior distributions. *ArXiv e-prints*, 2016.
- [24] S. R. Srinivasa Varadhan. *Large Deviations and Applications*. CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics, 1984.
- [25] Holger Wendland. *Scattered Data Approximation*, volume 17 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, 2004.

A Overview of Gaussian process regression

An important component of our algorithm is the construction of a surrogate for the map \hat{T} . We utilize the Gaussian process regression (GPR) method for this purpose. A feature of critical importance is that GPR specifies the posterior distribution as a Gaussian random function, with explicit formulas for the posterior mean and kernel function. The variance of the posterior can be used as an error or uncertainty estimate of the current prediction, which can in turn be used to guide optimization and explore parameter space. We briefly provide an overview of GPR since it's a crucial component of our proposed method, but refer to the book [21] and numerous other references in the literature for further details.

We estimate the parameter-to-observation map, $\hat{T}(\theta): U \rightarrow \mathbb{R}^1$, with $U \in \mathbb{R}^m$, via a GPR scheme from an observed dataset $\mathcal{D}_n = \{(\theta_i, \hat{T}(\theta_i))\}_{i=1}^n$, using n design points. These are the points that we have already sampled. Specifically, to estimate \hat{T} we place a Gaussian process prior over $\hat{T}(\theta)$ and consider the function values as a realization of the GP. In particular, with $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$, we have the following posterior mean $T_n(\theta)$ and covariance $k_n(\theta, \theta')$:

$$T_n(\theta) = \bar{T}(\theta) + k_0(\theta, \Theta)^T k_0(\Theta, \Theta)^{-1} (\hat{T}(\Theta) - \bar{T}(\Theta)) \quad (17)$$

$$k_n(\theta, \theta') = k_0(\theta, \theta') - k_0(\theta, \Theta)^T k_0(\Theta, \Theta)^{-1} k_0(\Theta, \theta') \quad (18)$$

where,

- $\bar{T}(\theta)$ is an arbitrary regression mean function, often chosen to be a constant or zero,
- $k_0(\theta, \theta') = \sigma^2 \exp\left(-\frac{\|\theta - \theta'\|^2}{2\lambda}\right)$ is the regression covariance function, here the exponential squared kernel, with σ and λ being positive hyperparameters,
- $k_0(\Theta, \Theta) \in \mathbb{R}^{n \times n}$ is the covariance matrix, with the ij^{th} entry given by $k_0(\theta_i, \theta_j)$, and $k_0(\theta, \Theta)$, $\bar{T}(\Theta)$, $\hat{T}(\Theta)$ are n -dimensional vectors with the i^{th} entries given by $k_0(\theta, \theta_i)$, $\bar{T}(\theta_i)$, $\hat{T}(\theta_i)$, respectively, and
- $\sigma_n^2(\theta) = k_n(\theta, \theta)$ is the variance at θ .

There are several important properties to emphasize in the GPR scheme related to the posterior [23]. Firstly, for any choice of the regression function the GPR mean estimate is an interpolant of the exact map at the design points, that is $T_n(\Theta) = \hat{T}(\Theta)$. Another property to note in the sequential framework, is that since $k_0(\Theta, \Theta)$ is positive definite:

$$\sigma_n(\theta) \leq \sigma_{n-1}(\theta) \leq \dots \leq \sigma_0(\theta), \quad \theta \in U, \quad (19)$$

in other words, additional data points lead to non-increasing local variance. Moreover, at each of the design points Θ the estimated variance vanishes, i.e. $\sigma_n(\theta_i) = 0$, for $i = 1, \dots, n$.

We recall an important result concerning the convergence of the posterior mean estimate of $T_n(\theta)$ [23, 25]. This result is used in appendix D to prove asymptotic convergence of the proposed sampling method.

Proposition 1. *Let $U \subset \mathbb{R}^M$ be a bounded Lipschitz domain that satisfies an interior cone condition. Also, let $T_n(\theta)$ be the mean given by the GPR method. Then there exists constants $p > M/2$ and C , independent of $\hat{T}(\theta)$, n , or Θ such that*

$$\|T_n - \hat{T}\|_{L^2} \leq Ch_\Theta^p \|\hat{T}\|_{L^2}, \quad (20)$$

where h_Θ is the fill distance defined as $h_\Theta = \sup_{\theta \in U} \inf_{\theta_n \in \Theta} \|\theta - \theta_n\|$.

As shown in [23], convergence of the mean estimate implies convergence of the estimated variance to zero, i.e.

$$\sigma_n(\theta) \rightarrow 0 \quad \text{as } n \rightarrow \infty \quad \text{and} \quad h_\Theta \rightarrow 0. \quad (21)$$

For uniform tensor grids Θ the fill distance is of the order $n^{-1/M}$ and thus the convergence of the map to the truth, in this case, is at least of order $\mathcal{O}(n^{-1/2})$. The precise rate of convergence may differ, depending on the exact criterion used.

B Iterative updates to a Gaussian process emulator with fixed hyperparameters

At every iteration a new data point is added to the design matrix Θ . This necessitates refactorization of the covariance matrix k_n . Since the Cholesky decomposition in general costs $\mathcal{O}(n^3)$, we would like to update the factorization without recomputing the Cholesky decomposition from scratch at each iteration. We can update the Cholesky factorization for a new data point at cost $\mathcal{O}(n^2)$ if the covariance matrix has fixed parameters (efficient formula's for the removal of a data point and similarly for block updates can be developed following the same approach, see e.g. [6]).

Denote the current covariance matrix with $n - 1$ data points by C_{11} with Cholesky factorization $C_{11} = A_1 A_1^\top$ and the covariance matrix after addition of a new design point θ_n by $\bar{C} = \bar{A} \bar{A}^\top$. We can write,

$$A = \begin{bmatrix} A_1 \\ 0 \end{bmatrix}, \quad \bar{A} = \begin{bmatrix} A_1 \\ \bar{a}_2^\top \end{bmatrix}, \quad (22)$$

so that

$$C = \begin{bmatrix} A_1 A_1^\top & \mathbf{0} \\ \mathbf{0}^\top & 0 \end{bmatrix} = \begin{bmatrix} C_{11} & \mathbf{0} \\ \mathbf{0}^\top & 0 \end{bmatrix}, \quad (23)$$

$$\bar{C} = \begin{bmatrix} A_1 A_1^\top & A_1 \bar{a}_2 \\ \bar{a}_2^\top A_1 & \bar{a}_2^\top \bar{a}_2 \end{bmatrix} = \begin{bmatrix} C_{11} & c_{21} \\ c_{21}^\top & c_{22} \end{bmatrix}, \quad (24)$$

where $c_{21} = k_0(\Theta, \theta_n)$ is the covariance of the new data point with the dataset and $c_{22} = k_0(\theta_n, \theta_n)$ is variance of the data point. The matrix \bar{C} can be decomposed as

$$\bar{C} = \begin{bmatrix} L_{11} & \\ \ell_{21}^\top & \ell_{22} \end{bmatrix} \begin{bmatrix} L_{11}^\top & \ell_{21} \\ & \ell_{22} \end{bmatrix}, \quad (25)$$

and thus

$$L_{11} L_{11}^\top = C_{11}, \quad (26)$$

$$L_{11} \ell_{21} = c_{21} \implies \ell_{21} = L_{11} \setminus c_{21}, \quad (27)$$

$$\ell_{21}^\top \ell_{21} + \ell_{22}^2 = c_{22} \implies \ell_{22} = \sqrt{c_{22} - \ell_{21}^\top \ell_{21}}, \quad (28)$$

which we solve for ℓ_{21} and ℓ_{22} at costs $\mathcal{O}(n^2)$ due to the linear system solve in the second equation above (L_{11} is already known from the previous iteration).

C Algorithm pseudocode

Below we summarize the main loop of the sequential algorithm in pseudocode:

```

input initial dataset  $\mathcal{D}_{n_s} = \{(\theta_i, \hat{T}(\theta_i))\}_{i=1}^{n_s}$ 
repeat
     $T_n, \sigma_n \leftarrow$  predict Gaussian process mean and variance
     $f_n, f_n^+, f_n^- \leftarrow$  integration using  $T_n, \sigma_n$ , and  $f_\theta$ 
     $\theta^* \leftarrow \arg \min_\theta Q(\theta, T_n, \mathcal{D}_n)$ 
    Append  $(\theta^*, \hat{T}(\theta^*))$  to dataset  $\mathcal{D}_n$ 
until desired error level  $d(f_n^+, f_n^-) < \epsilon$ 
return  $f_n$ 
```

This is an iterative procedure that leads to a series of pdfs for the quantity of interest q which, as we show in appendix D, converges to the true pdf under appropriate conditions. The function Q that computes the distance criterion at a test point θ is summarized below:

```

function  $Q(\theta, T_n, \mathcal{D}_n)$ 
    Append  $(\theta, T_n(\theta))$  to dataset  $\mathcal{D}_n$ 
     $\tilde{\sigma}_{n+1} \leftarrow$  predict Gaussian process variance
     $\tilde{f}_{n+1}^+, \tilde{f}_{n+1}^- \leftarrow$  integration using  $T_n, \tilde{\sigma}_{n+1}$ , and  $f_\theta$ 
    return  $d_{L_1}(\log(\tilde{f}_{n+1}^+), \log(\tilde{f}_{n+1}^-))$ 
```

Above d is a user defined distance metric in the stopping condition and $d_{L_1}(f, g) = \int |f - g|$.

D Convergence and asymptotic behavior of the optimization criterion

Here we provide details on the proposed sampling algorithm's convergence and the asymptotic behavior of the proposed criterion, which elucidate its favorable extreme event sampling properties. We first prove the convergence of the estimated probability distribution to the true distribution in the large sample size n limit. This is under the assumption that the sampling criterion does not lead to resampling, i.e. we have $h_\Theta \rightarrow 0$ as $n \rightarrow \infty$. Under this condition the result in appendix A guarantees convergence of the map. We show that L^2 convergence of the map implies convergence of the sequence of the transformed pdfs f_n .

Theorem 2. *Let the sequence of random variables be denoted by q_n with cumulative distribution function F_n . Moreover, assume that the sampling criterion does not lead to resampling of existing points. Then the L_2 convergence of the sequence of maps T_n to \hat{T} by proposition 1 implies q_n converges in probability to q , and this then implies F_n converges in distribution to \hat{F} .*

Proof. The assumption that the criterion does not lead to resampling of existing samples guarantees $h_\Theta \rightarrow 0$ as $n \rightarrow \infty$ and thus the sequence T_n converges in L_2 to \hat{T} by proposition 1.

We show that L^2 convergence implies convergence in measure. For every $\epsilon > 0$, by Chebyshev inequality, we have

$$\|T_n - \hat{T}\|_{L_2}^2 \geq \mathbb{E}(|T_n - \hat{T}|^2) \geq \epsilon^2 \mathbb{P}(|T_n - \hat{T}| \geq \epsilon). \quad (29)$$

Therefore since the left hand side tends to zero as $n \rightarrow \infty$, for every $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}(|T_n - \hat{T}| \geq \epsilon) = 0,$$

thus T_n converges in probability to \hat{T} . This also implies T_n converges in distribution to \hat{T} : $\lim_{n \rightarrow \infty} F_n(s) = \hat{F}(s)$, for all s that are points of continuity of $\hat{F}(s)$ (see [19]). \square

The next result is on the asymptotic behavior of the criterion. Specifically, we show that for large n , so that the distance between the upper and lower map bounds decrease and become small, the proposed criterion Q measures and attempts to minimize the variance of the surrogate map. We show that the metric appropriately weights regions with large $|T_n|$ and small probability f_θ with more importance.

Theorem 3. *Let $T_n(\theta)$ and $\sigma_n(\theta)$ from the GPR scheme be sufficiently smooth functions of θ . The asymptotic behavior of Q for large n (ensuring small σ) and small $\|\nabla\sigma\|/\|\nabla T\|$ is given by*

$$\tilde{Q}_n \triangleq \frac{1}{2} \int |\log f_n^+(s) - \log f_n^-(s)| ds \quad (30)$$

$$\approx \int \left| \frac{\frac{d}{ds} \mathbb{E}(\sigma_n(\theta) \cdot \mathbf{1}_{T_n(\theta)=s})}{f_n(s)} \right| ds, \quad (31)$$

where \mathbb{E} denotes the expectation over the probability measure \mathbb{P}_θ .

Proof. To simplify notation we drop the index n in what follows. We want to determine the asymptotic behavior of the integral for $n \rightarrow \infty$. In this case, we can assume that $\sigma(\theta)$ is uniformly bounded. We focus on the integrand and determine the behavior of the difference $f^+ - f^-$ appearing in the integrand. First focus on the cumulative distribution function difference

$$F^+(s) - F^-(s) = - \int_{A^-(s) \setminus A^+(s)} f_\theta(\theta) d\theta, \quad (32)$$

where $A^-(s) \setminus A^+(s)$ denotes the difference between the sets $A^\pm(s) = \{\theta \in U : T(\theta) \pm \sigma(\theta) \leq s\}$. Now since $A^-(s) \setminus A^+(s)$ is equivalent to $T(\theta) \in (s - \sigma(\theta), s + \sigma(\theta))$,

$$F^+(s) - F^-(s) = - \int_{T(\theta) \in (s - \sigma(\theta), s + \sigma(\theta))} f_\theta(\theta) d\theta, \quad (33)$$

$$= - \int_{T(\bar{\theta})=s} (h^-(\bar{\theta}) - h^+(\bar{\theta})) f_\theta(\bar{\theta}) dS_{\bar{\theta}}, \quad (34)$$

where $\int dS_{\bar{\theta}}$ denotes surface integration and $\bar{\theta}$ is on the surface $T(\bar{\theta}) = s$. The functions h^\pm satisfy

$$T(\bar{\theta} + h^\pm(\bar{\theta})\mathbf{n}) \pm \sigma(\bar{\theta} + h^\pm(\bar{\theta})\mathbf{n}) = s, \quad (35)$$

where $\mathbf{n} = \nabla T / \|\nabla T\|$ is the unit normal to the surface $T(\bar{\theta}) = s$. Note by Taylor theorem we have

$$T(\bar{\theta} + \mathbf{n}t) = T(\bar{\theta}) + (\mathbf{n} \cdot \nabla T(\bar{\theta}))t + \frac{(\mathbf{n} \cdot \nabla)^2 T(\bar{\theta} + \mathbf{n}t_1)}{2} t^2, \quad (36)$$

$$\sigma(\bar{\theta} + \mathbf{n}t) = \sigma(\bar{\theta}) + (\mathbf{n} \cdot \nabla \sigma(\bar{\theta}))t + \frac{(\mathbf{n} \cdot \nabla)^2 \sigma(\bar{\theta} + \mathbf{n}t_1)}{2} t^2, \quad (37)$$

for $0 < t_1 < t$, and therefore from (35) we find,

$$h^\pm(\bar{\theta}) = \frac{\mp \sigma(\bar{\theta})}{(\nabla T(\bar{\theta}) + \nabla \sigma(\bar{\theta})) \cdot \mathbf{n}} + \mathcal{O}(\sigma^2). \quad (38)$$

We then have

$$\begin{aligned} h^- - h^+ &= \left(\frac{\sigma}{(\nabla T - \nabla \sigma) \cdot \mathbf{n}} + \frac{\sigma}{(\nabla T + \nabla \sigma) \cdot \mathbf{n}} \right) + \mathcal{O}(\sigma^2) \\ &= \frac{2\sigma}{\|\nabla T\| \left(1 - \frac{(\nabla \sigma \cdot \nabla T)^2}{\|\nabla T\|^3} \right)} + \mathcal{O}(\sigma^2) \end{aligned} \quad (39)$$

$$= \frac{2\sigma}{\|\nabla T\|} + \mathcal{O}\left(\frac{\|\nabla \sigma\|}{\|\nabla T\|}, \sigma^2\right), \quad (40)$$

where for the second equality we use the Cauchy–Schwarz inequality $\nabla \sigma \cdot \nabla T \leq \|\nabla \sigma\| \|\nabla T\|$ to bound the error in the last line. From this result and (34) we obtain,

$$f^+ - f^- = -\frac{d}{ds} \int_{T(\bar{\theta})=s} (h^-(\bar{\theta}) - h^+(\bar{\theta})) f_\theta(\bar{\theta}) dS_{\bar{\theta}} \quad (41)$$

$$= -\frac{d}{ds} \int_{T(\bar{\theta})=s} \frac{2\sigma(\bar{\theta})}{\|\nabla T(\bar{\theta})\|} f_\theta(\bar{\theta}) dS_{\bar{\theta}} + \mathcal{O}\left(\frac{\|\nabla \sigma\|}{\|\nabla T\|}, \sigma^2\right) \quad (42)$$

$$= -\frac{d}{ds} \int_{T(\bar{\theta})=s} 2\sigma(\bar{\theta}) f_\theta(\bar{\theta}) d\bar{\theta} + \mathcal{O}\left(\frac{\|\nabla \sigma\|}{\|\nabla T\|}, \sigma^2\right) \quad (43)$$

$$\approx -2 \frac{d}{ds} \mathbb{E}(\sigma(\theta) \cdot \mathbf{1}_{T(\theta)=s}) \quad (44)$$

Next, note the following approximation

$$\log f^+ - \log f^- = \log(1 + \Delta f/f^-) \approx \Delta f/f^- \approx \Delta f/f, \quad (45)$$

where $\Delta f = f^+ - f^-$, under the small σ assumption. We therefore have the following result on the behavior of the difference between f^+ and f^-

$$\log f^+(s) - \log f^-(s) \approx -2 \frac{\frac{d}{ds} \mathbb{E}(\sigma(\theta) \cdot \mathbf{1}_{T(\theta)=s})}{f(s)}, \quad (46)$$

which implies (30). \square

Note that the pdf in the denominator in the integrand (10) is a direct implication of our choice to consider the difference between the logarithmic transform of the pdfs. This guarantees that values of the parameter θ where the probability is small (rare events) are sampled sufficiently to capture tail behavior of the map's pdf. This is seen clearly in the following corollary, where we consider the special case of a one-dimensional parameter space and a family of maps which are monotonically increasing.

Corollary 2. Let $\theta: \Omega \rightarrow [u_1, u_2] \subset \mathbb{R}$ be a one-dimensional random variable and in addition to the assumptions of Theorem 1 we assume that $T'_n(\theta)$ is monotonically increasing function. Then, the asymptotic value of \tilde{Q}_n for large n has the following property:

$$\tilde{Q}_n \gtrsim \left| \int_U \sigma_n(\theta) d(\log f_\theta(\theta)) + \int_U \sigma_n(\theta) d(\log T'(\theta)) + \sigma_n(u_2) - \sigma_n(u_1) \right|. \quad (47)$$

Proof. For simplicity, we drop the n index in the following. From the proof of Theorem 1 and using standard inequalities we obtain

$$\tilde{Q} \approx \int \left| \frac{\frac{d}{ds} \int_{T(\theta)=s} \sigma(\theta) f_\theta(\theta) d\theta}{f(s)} \right| ds \geq \left| \int \frac{\frac{d}{ds} \int_{T(\theta)=s} \sigma(\theta) f_\theta(\theta) d\theta}{f(s)} ds \right| = \left| \int \frac{\frac{d^2}{ds^2} \int_{T(\theta) \leq s} \sigma(\theta) f_\theta(\theta) d\theta}{\frac{d}{ds} \int_{T(\theta) \leq s} f_\theta(\theta) d\theta} ds \right| \quad (48)$$

Since we assume $T'_n(\theta) > 0$ by applying the transformation $\theta = T^{-1}(s)$ to the numerator of the right hand side we obtain

$$\frac{d^2}{ds^2} \int_{T(\theta) \leq s} f_\theta(\theta) \sigma(\theta) d\theta = \frac{d^2}{ds^2} \int_s^s \frac{f_\theta(T^{-1}(s)) \sigma(T^{-1}(s))}{T'(T^{-1}(s))} ds \quad (49)$$

$$= \frac{d}{ds} \left(\frac{f_\theta(T^{-1}(s)) \sigma(T^{-1}(s))}{T'(T^{-1}(s))} \right), \quad (50)$$

which is equivalent to

$$\frac{f_\theta(T^{-1}(s)) \sigma(T^{-1}(s))}{(T'(T^{-1}(s)))^2} + \frac{f_\theta(T^{-1}(s))}{T'(T^{-1}(s))} \left(\frac{\sigma(T^{-1}(s))}{T'(T^{-1}(s))} \right)' . \quad (51)$$

Similarly, we have the following for the denominator term:

$$\frac{d}{ds} \int_{T(\theta) \leq s} f_\theta(\theta) d\theta = \frac{d}{ds} \int_s^s \frac{f_\theta(T^{-1}(s))}{T'(T^{-1}(s))} ds = \frac{f_\theta(T^{-1}(s))}{T'(T^{-1}(s))}. \quad (52)$$

Therefore we obtain,

$$\tilde{Q} \gtrsim \left| \int_{T(u_1)}^{T(u_2)} \left[\frac{f'_\theta(\theta)}{f_\theta(\theta)} \frac{\sigma(\theta)}{T'(\theta)} + \left(\frac{\sigma(\theta)}{T'(\theta)} \right)' \right]_{\theta=T^{-1}(s)} ds \right| = \left| \int_{u_1}^{u_2} \left[\frac{f'_\theta(\theta)}{f_\theta(\theta)} \sigma(\theta) - \frac{T''(\theta)}{T'(\theta)} \sigma(\theta) + \sigma'(\theta) \right] d\theta \right|. \quad (53)$$

For each of the first two terms we apply the transformations $z = \log f_\theta(\theta)$ and $z = \log T'(\theta)$ respectively to obtain the final result:

$$\tilde{Q} \gtrsim \left| \int_{\log f_\theta(\theta)=z} \sigma(\theta) dz + \int_{\log T'(\theta)=z} \sigma(\theta) dz + \sigma(u)|_{u_1}^{u_2} \right|. \quad (54)$$

□

E Parameterization of a random process using the Karhunen-Loève expansion

To apply the developed algorithm it is important to have a parameterization of the uncertainty in the model; in other words a description of the stochastic forcing, random parameters, or initial conditions in terms of a finite set of random variables. There are a plethora of methods that can be used to achieve this parameterization with a minimal number of random parameters, which include linear methods such as the Karhunen-Loève expansion, also known as principal component analysis, and nonlinear methods, e.g. kernel principal component analysis.

For completeness we describe the Karhunen-Loève (KL) expansion (see e.g. [20, 22]) as one strategy, which is commonly applied in practice due to its simplicity and general applicability. The KL expansion provides an optimal linear parameterization of a random process in the mean square sense. An important caveat is that the KL expansion is not the only strategy, and might be ill-suited for certain systems depending on the structure of the stochastic process, and thus a nonlinear parameterization may be better suited to capture the desired physics with a lower dimensional parameterization.

Consider an L^2 process $\zeta(t)$, $0 \leq t \leq \tau$, with mean $\mathbb{E}(\zeta(t))$ and continuous covariance $k(t_1, t_2)$, we can expand this in terms of a double orthogonal process

$$\zeta(t; \omega) = \mathbb{E}(\zeta(t)) + \sum_{i=1}^{\infty} Z_i(\omega) e_i(t), \quad 0 \leq t \leq \tau, \quad (55)$$

where the eigenfunctions $e_i(t)$ of the covariance are orthonormal and the random variables $Z_i(\omega)$ are orthogonal with mean zero and variance λ_i , that is $\mathbb{E}(Z_i Z_j) = \lambda_i \delta_{ij}$. Through this procedure we can obtain a desired parameterization of the noise in terms of the orthogonal random coefficients of the expansion,

$$\theta(\omega) = (Z_1(\omega), \dots, Z_n(\omega)), \quad (56)$$

where the random variable $\theta \in \mathbb{R}^n$ has pdf f_θ .

Suppose now that $\zeta(t)$ is a Gaussian process, we can then show that the random variables Z_i are uncorrelated and Gaussian, and are therefore independent. Thus, for a mean zero Gaussian process with covariance $k(t_1, t_2) = \text{cov}(\zeta(t_1), \zeta(t_2))$ its KL expansion is given by

$$\zeta(t) = \sum_{i=1}^{\infty} \sqrt{\lambda_i} \xi_i e_i(t), \quad (57)$$

where $\{\lambda_i, e_i(t)\}$ are the eigenvalues and eigenfunctions of the covariance function and $\xi_i \sim \mathcal{N}(0, 1)$ are draws from the normal distribution. The variances $\lambda_i > 0$ are ordered $\lambda_1 > \lambda_2 > \dots > 0$ so that λ_1 corresponds to the eigenspace explaining the largest variance. We thus obtain an n -dimensional parameterization of the process by truncating the series to a desired accuracy. For such a case, we have $\theta(\omega) = (\sqrt{\lambda_1} \xi_1(\omega), \dots, \sqrt{\lambda_n} \xi_n(\omega))$ and $f_\theta = \mathcal{N}(0, \text{diag}(\lambda_1, \dots, \lambda_n))$.

F Hydrodynamic forces and moments on an offshore platform

F.1 Numerical experiments

The numerical simulations are performed using the open-source code DualSPHysics [5], which utilizes the smoothed particle hydrodynamics (SPH) framework, a meshless Lagrangian method. DualSPHysics has been validated on numerous test cases in offshore engineering applications, including forces on structures and also wave propagation, see e.g. [1].

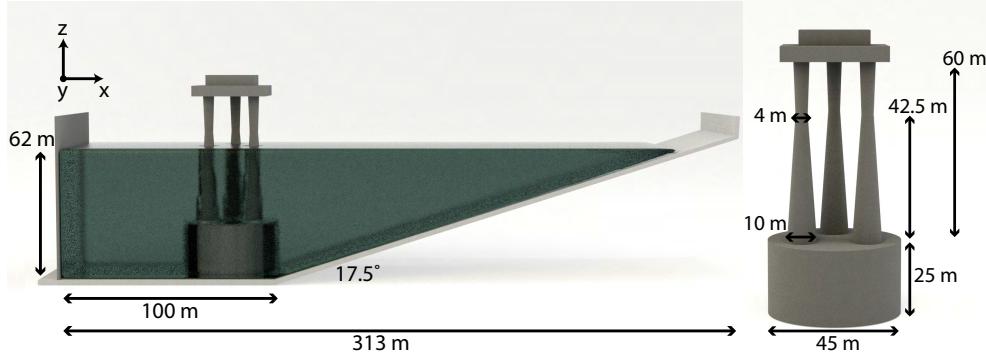


Figure 8 The numerical domain (left) and a close up view of the offshore platform (right).

A sketch of the numerical domain is provided in Fig. 8. We are interested in analyzing the forces and moments on offshore structures in the deep ocean, where the dispersion relation is given by $\omega_0^2 = gk_0$. We consider waves with peak period 10 s so that the characteristic wavelength is $\lambda_0 = \frac{2\pi}{k_0} = 156$ m. In addition, the depth of the wave tank is selected so that $\tanh(k_0 h) = 0.99$, thus the water depth is $h = 62$ m. The beach is setup at a 17.5° angle and the length of the horizontal tank dimension, that is excluding the sloping beach, is 100 m. The structure we consider is an offshore gravity platform (Fig. 8, right) and the dimensions of the model are based on

prototypical values. In particular, the base width of the platform is 45 m with height 25 m, three columns with base diameter 10 m extend from the bottom platform and narrow to a 4 m width at height 42.5 m. To generate waves, we implemented a hinged-type wave maker utilizing the corresponding Bi  sel transfer function [10] to relate the wave height to the stroke of the paddle. For the flap-type wave maker we use, the Bi  sel transfer function is given by

$$\frac{H}{S_0} = \frac{2 \sinh(kh)(1 - \cosh(kh) + kh \sinh(kh))}{kh(\sinh(kh) \cosh(kh) + kh)}, \quad (58)$$

where S_0 is the stroke at the free surface, H is the wave height in the far-field, k is the wavenumber, and h is the water depth.

Wind generated ocean waves are empirically described by their energy spectrum. Here, we consider irregular seas with JONSWAP spectral density:

$$S(f) = \frac{\alpha g^2}{(2\pi)^4 f^5} \exp\left[-\frac{5}{4}\left(\frac{f_p}{f}\right)^4\right] \cdot \gamma \exp\left[\frac{-(f-f_p)^2}{2\delta^2 f_p^2}\right], \quad (59)$$

where $\delta = 0.07$ for $k \leq k_0$ and $\delta = 0.09$ for $k > k_0$ and f_p is the peak frequency. In the original formulation α is related to the fetch and the mean wind speed, however for offshore applications, especially in the North Sea, the following modified version is often adopted [17] $\alpha = 5.058 H_s^2 f_p^4 (1 - 0.287 \log \gamma)$.

F.2 Additional cases

Here we provide results for the momentum variable for two different spectra with parameters:

- JONSWAP 1: $\alpha = 0.060$, $\gamma = 3.0$, $H_s = 13.2$ m.
- JONSWAP 2: $\alpha = 0.100$, $\gamma = 2.0$, $H_s = 15.7$ m.

These parameters are representative of North Sea conditions. The chosen peak wave period is $T_p = 10$ s. For these spectra, in Fig. 9 we plot the distribution of the wavefield parameterization f_θ .

The quantities of interest in this problem are the forces and moments acting on the platform:

$$q_f = \max_{t \in [0, T]} |F_x(t)| \quad \text{and} \quad q_m = \max_{t \in [0, T]} |M_y(t)|. \quad (60)$$

The forces and moments are normalized by $k_0^3/\rho g$ and $k_0^2/\rho g$, respectively.

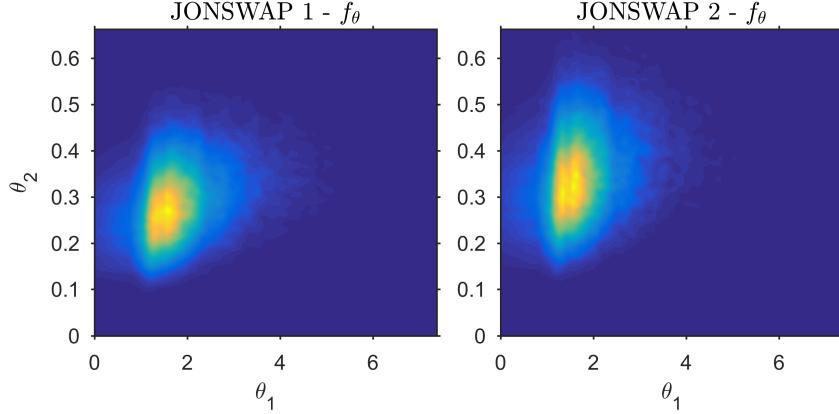


Figure 9 The parameter to observation map $T(\theta)$ for the force F_x (left) and moment M_y (right). Recall that θ_1 is the lengthscale L of a wave group and θ_2 is the amplitude or height A of the group.

In Fig. 10 we show the exact parameter to observation maps for the variables under consideration, which is computed by a cubic interpolation of 48 samples (i.e. numerical simulations of the SPH simulations for various θ values). It is important to note that the parameter to observation map *remain fixed regardless of the spectra of the*

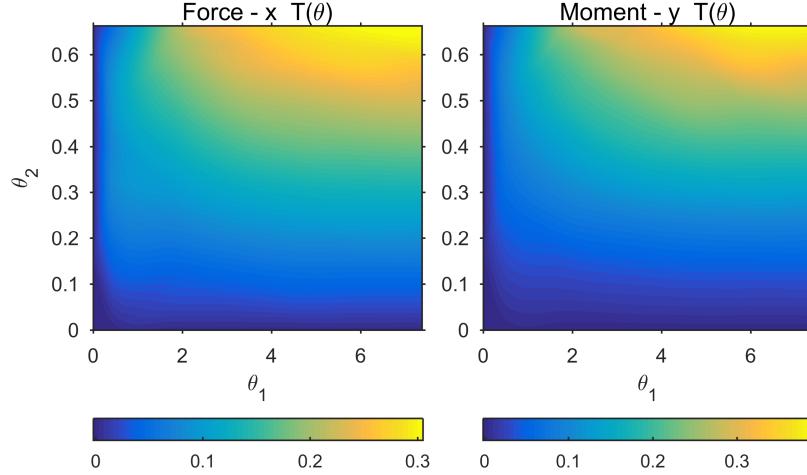


Figure 10 The parameter to observation map $T(\theta)$ for the force F_x (left) and moment M_y (right). Recall that θ_1 is the lengthscale L of a wave group and θ_2 is the amplitude or height A of the group.

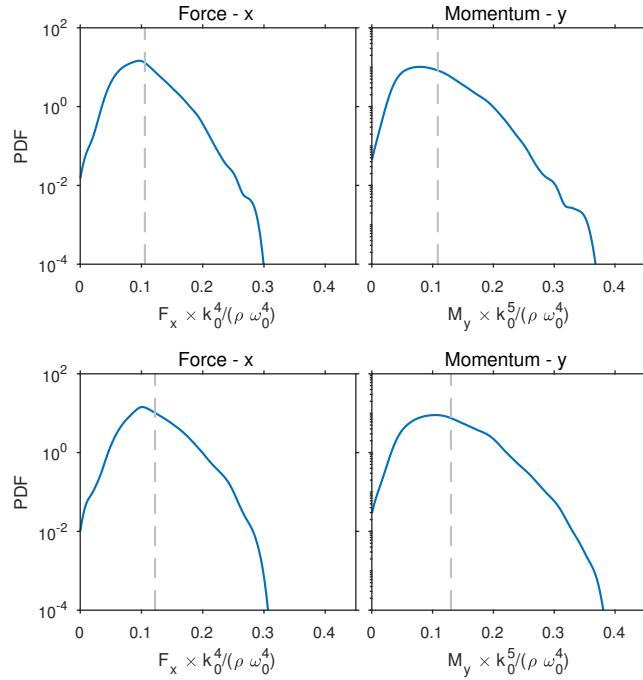


Figure 11 The exact pdf of the force and momentum variables (JONSWAP 1 left column and JONWAP 2 right column).

wave field. Fig 11 display the target, i.e the true, pdf from the densely sampled map shown in Fig. 10 for both JONSWAP spectra.

In Fig. 13 we show the result for the momentum variable for JONSWAP spectrum 1 (force variable results are provided in the main text). The results for JONSWAP spectrum 2 are summarized in Figs. 14 and 15.

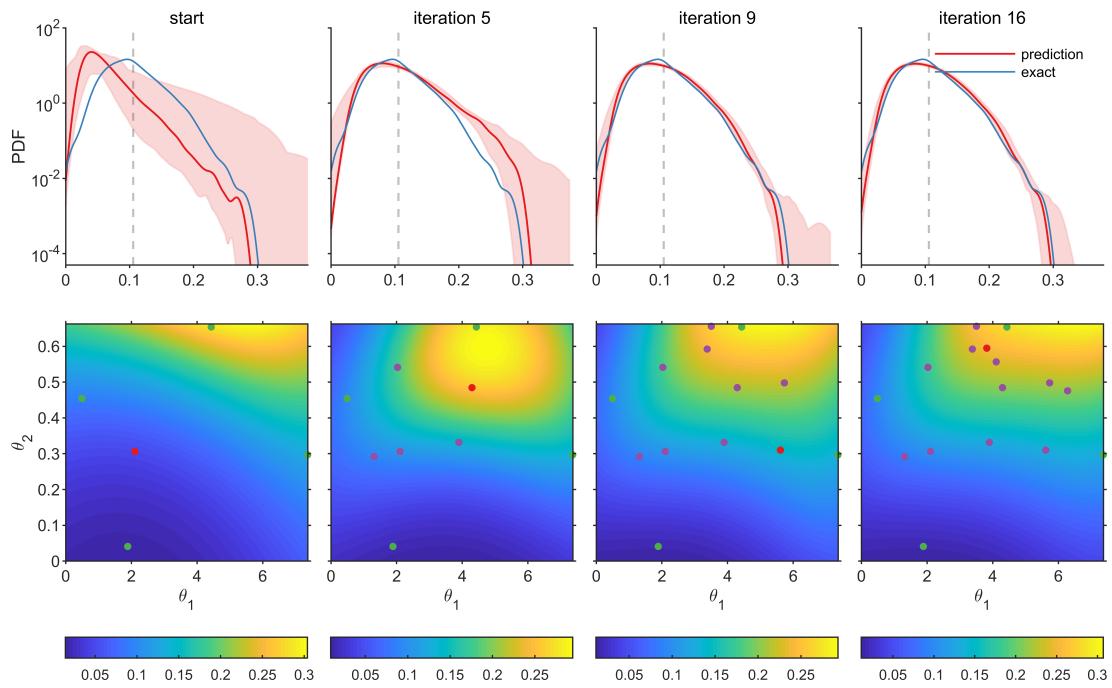


Figure 12 [JONSWAP 1] Algorithm progression for the force variable.

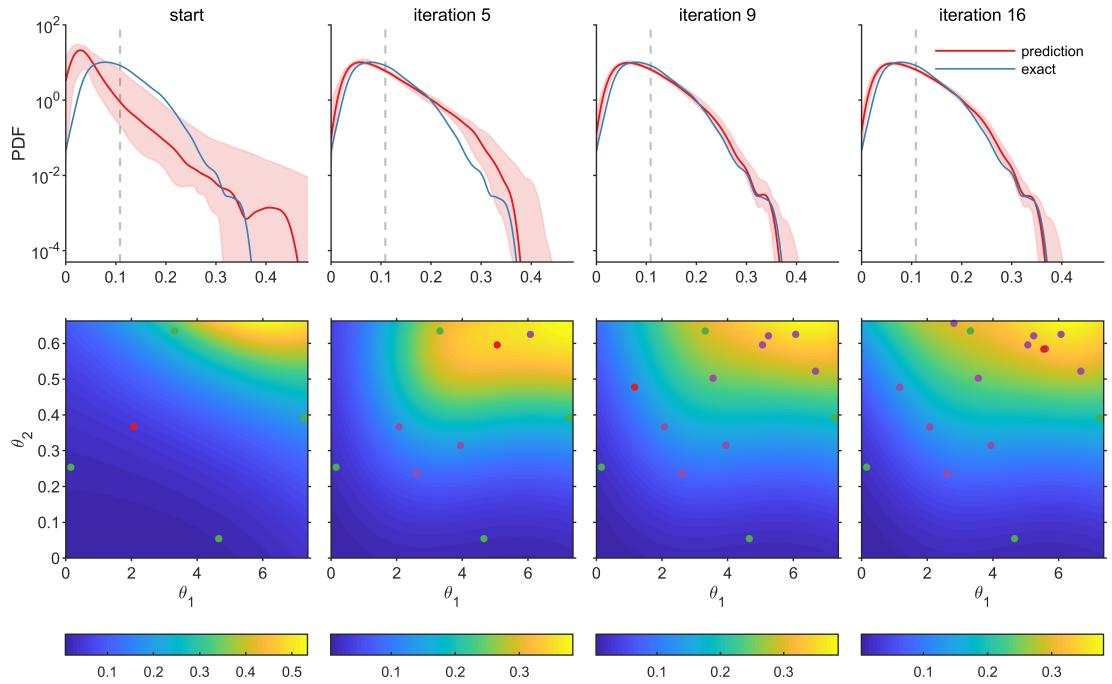


Figure 13 [JONSWAP 1] Algorithm progression for the momentum variable.

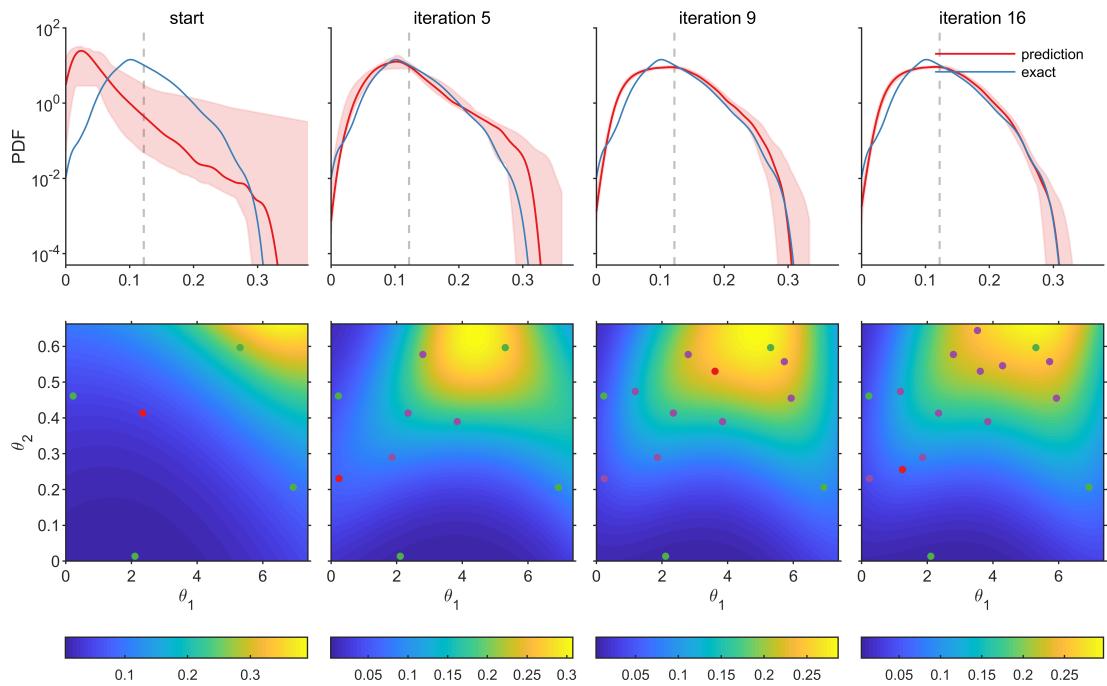


Figure 14 [JONSWAP 2] Algorithm progression for the force variable.

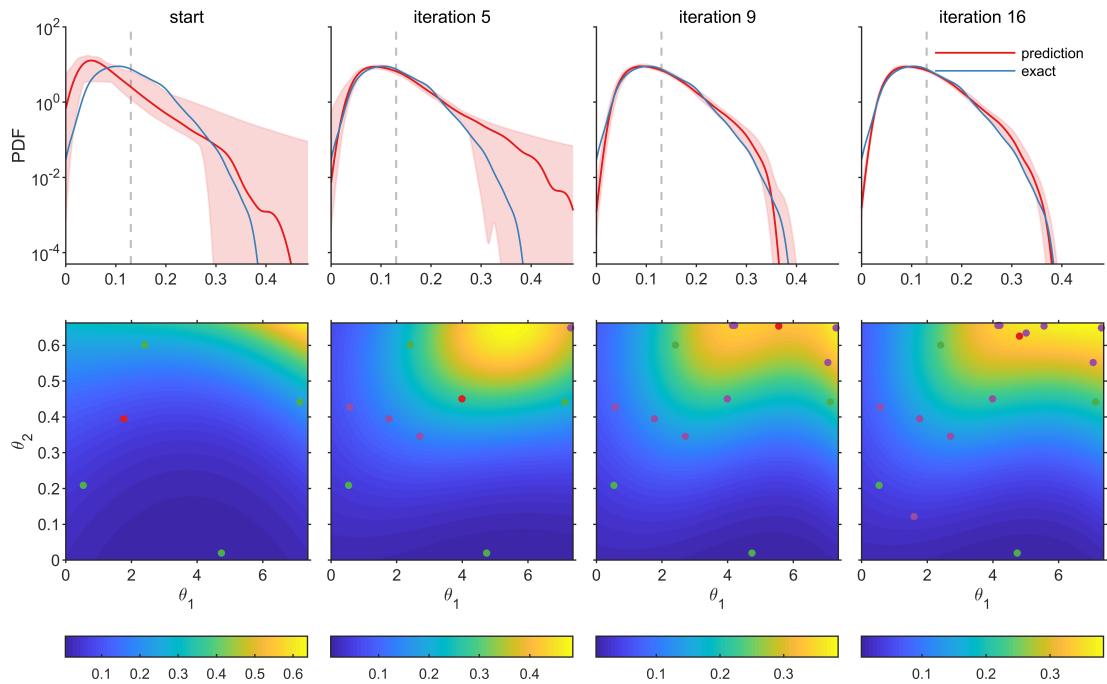


Figure 15 [JONSWAP 2] Algorithm progression for the momentum variable.