Sequential sampling strategy for extreme event statistics in nonlinear dynamical systems

Mustafa A. Mohamad[a] and Themistoklis P. Sapsis[a,1]

[a]Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge MA 02139

We develop a method for the evaluation of extreme event statistics associated with nonlinear dynamical systems from 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 uses 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 using 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 uses 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 and for problems where each sample is very expensive to obtain. 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 3D irregular waves. It is demonstrated that the developed approach can accurately determine the extreme event statistics using a limited number of samples.

extreme events | adaptive sampling | sequential experimental design

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 (1), rogue waves in the ocean (2), and extreme climate events (3). 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 necessitate 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. 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) (4). Another strategy is large deviations theory (5, 6), 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 nontrivial 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 (7) is challenging to solve even in moderate to low dimensions (8). 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 used in climate science and meteorology by others (9–11). In refs. 12 and 13, a probabilistic decomposition of extreme events was used 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

Significance

Understanding the statistics of extreme events in dynamical systems of high complexity is of vital importance for reliability assessment and design. We formulate a method to pick samples optimally so that we have rapid convergence of the full statistics 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 very expensive 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.

response of the observable is modeled using a machine learning method that infers the functional form of the quantity of interest by using only a few strategically sampled numerical simulations or experiments. Combining the predictions from the machine learning model, using the Gaussian process regression (GPR) 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 uses 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. The proposed method allows us to sequentially sample the parameter space to rapidly capture the pdf and, in particular, the tails of the distribution of the observable of interest.

Problem Setup and Method Overview

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

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

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 $\theta : \Omega \to U \subset \mathbb{R}^m$ 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

$$q = \hat{T}(\theta) \triangleq \mathcal{F}(u), \quad \text{[2]}$$

where $\hat{T} : U \subset \mathbb{R}^m \to \mathbb{R}$ is a continuous parameter-to-observation map and $\mathcal{F}$ is an arbitrary functional of u. In addition, it is possible to consider observations q that are corrupted by some observational or numerical noise term, but for simplicity, we assume zero noise. 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

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

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—that is, 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 $f_{D_{n-1}}$, determine the next input parameter θ such that when the map is evaluated at this

new sample the error between f_{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 an inexpensive-to-compute surrogate model to replace $\hat{T}(\theta)$ appearing in the θ-parametrized dataset $D_n(\theta)$ and using a version of the distance metric without explicit dependence on $\hat{f}$. Here, we use 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).

Method Description

An important component of our proposed method is the construction of an inexpensive surrogate for the map $\hat{T}$. We use GPR as our learning method. GPR considers the function as a Gaussian process, in terms of a distribution in function space (see *SI Appendix*, 1 and numerous references, such as ref. 14). 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 use along with f_θ to guide selection of the next sample point.

We learn the parameter-to-observation map $\hat{T} : U \subset \mathbb{R}^m \to \mathbb{R}$ using GPR from the dataset $\mathcal{D}_{n-1} = \{(\hat{\theta}_i, \hat{T}(\theta_i))\}_{i=1}^{n-1}$. The GPR
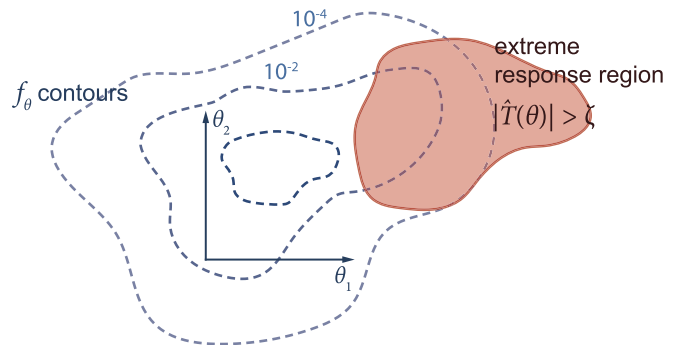

Fig. 1. Areas with large probability in θ are not necessarily associated with regions where $|\hat{T}|$ is large. The proposed criterion focuses on sampling regions where both the probability and the magnitude of $|\hat{T}|$ are significant.
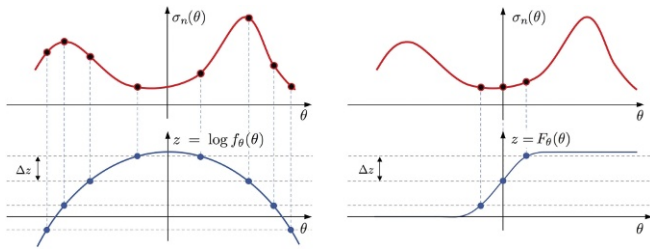
Fig. 2. (*Left*) Integration of $\sigma_n(\theta)$ over contours of $\log f_\theta(\theta)$ implies sampling of $\sigma_n(\theta)$ in low probability regions of θ. (*Right*) On the other hand, low-order moments of $\sigma_n(\theta)$ rely only on values of $\sigma_n(\theta)$ close to high probability regions of θ; thus, rare events are not sampled sufficiently.

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 $\boldsymbol{T}_{n-1}(\theta) \sim \mathrm{GP}(T_{n-1}(\theta), k_{n-1}(\theta, \theta'))$ (see *SI Appendix*, 1 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 $\boldsymbol{T}_{n-1}$:

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

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 $\widetilde{\mathcal{D}}_n(\theta) = \mathcal{D}_{n-1} \cup \{(\theta, T_{n-1}(\theta))\}$, 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 $\widetilde{\boldsymbol{T}}_n(\theta; \theta^*) \sim \mathrm{GP}(\widetilde{T}_n(\theta; \theta^*), \widetilde{k}_n(\theta, \theta'; \theta^*))$ denote the random function trained on the augmented dataset $\widetilde{\mathcal{D}}_n(\theta^*)$. 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 $\widetilde{\boldsymbol{T}}_n(\theta; \theta^*)$, where $\theta \sim f_\theta$ and θ^* is a sample point. Note that the mean $\widetilde{T}_n(\theta; \theta^*)$ of the random function $\widetilde{\boldsymbol{T}}_n(\theta; \theta^*)$ 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 $\boldsymbol{T}_n$. Specifically, let $\tilde{f}_n^{\pm}(\cdot; \theta^*)$ denote the pdfs of the two random variables $\widetilde{T}_n(\theta; \theta^*) \pm \alpha \widetilde{\sigma}_n(\theta; \theta^*)$, where $\widetilde{\sigma}_n^2(\theta) = \widetilde{k}_n(\theta, \theta)$, which are the upper and lower bounds of the confidence interval based on the α-scaled SD of the posterior distribution. The pdfs corresponding to the confidence bounds are explicitly given by

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

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

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

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

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:

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

This is an 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 come 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 *SI Appendix*, 3. For the optimization problem, we use 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 codomain 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
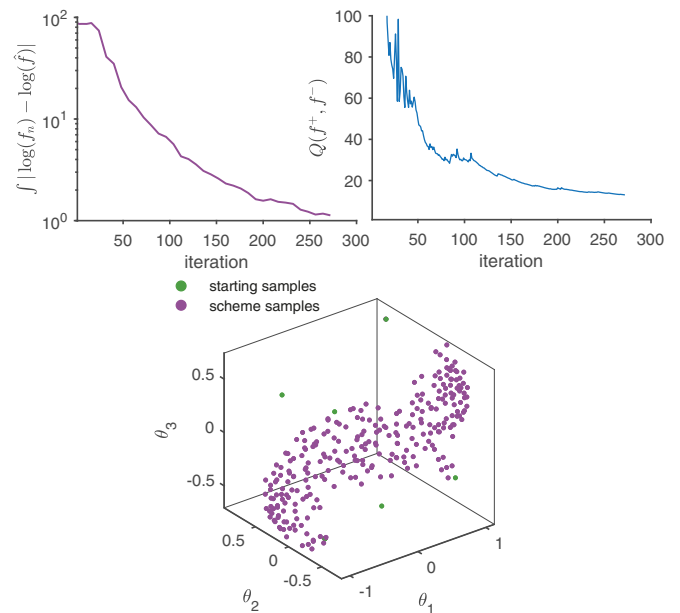

Fig. 3. (*Bottom*) A scatterplot of the algorithm sampling of the parameter space (green points denote the initial random LH samples). (*Top Left*) The corresponding L_1 error of the logarithmic transform of the pdf between the GPR mean and truth, and (*Top Right*) value of the criterion in 5 as a function of the iteration number.
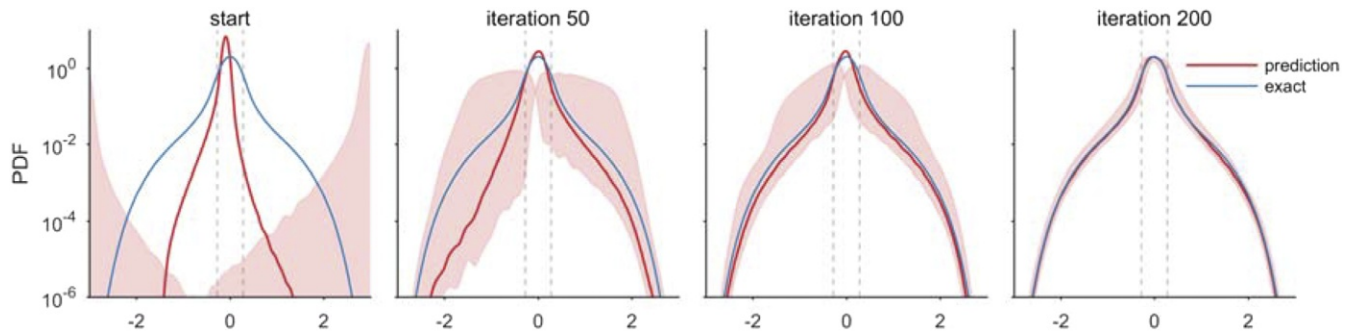
Mohamad and Sapsis

Fig. 4. Progression of the pdf estimation as we iteratively sample points. We shade the region between the pdfs $f^\pm$ in red, as a visualization of the convergence of the pdfs $f^\pm$. Dashed vertical lines denote 1 SD.

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 (LH)-based sampling plan should be used for this purpose. We also recommend as a pretraining 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,

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

to capture the main probability mass (low order moments) before using 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 *SI Appendix*, 2). (For low-dimensional θ, since we presume the dataset size is small, the cost difference may be negligible.)

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 *SI Appendix*, 4 for the proof). The second result shows that the asymptotic form of the criterion is given by (see *SI Appendix*, 4 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

$$\widetilde{Q}_n \triangleq \frac{1}{2}\int \left|\log f_n^+(s) - \log f_n^-(s)\right| ds \quad [9]$$

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

where $\mathbb{E}$ denotes the expectation over the probability measure $\mathbb{P}_\theta$.

Note that the pdf in the denominator under the integral in 9 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 demonstrated by the following corollary (proof in *SI Appendix*, 4).

Corollary 1: Let $\theta\colon \Omega \to [u_1, u_2] \subset \mathbb{R}$ be a one-dimensional random variable, and in addition to the assumptions of theorem 1,

we assume that $T_n(\theta)$ is a monotonically increasing function. Then, the asymptotic value of $\widetilde{Q}_n$ for large n has the following property:

$$\widetilde{Q}_n \geq \left|\int_U \sigma_n(\theta)\, d(\log f_\theta(\theta)) + \int_U \sigma_n(\theta)\, d(\log T'(\theta)) \right.$$
$$\left. + \sigma_n(u_2) - \sigma_n(u_1)\right|, \quad [11]$$

where $'$ denotes differentiation and where it is understood that the inequality above is based on the asymptotic result in Eq. 10.

Therefore, for a large value of n, $\widetilde{Q}_n$ 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

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

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 a 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
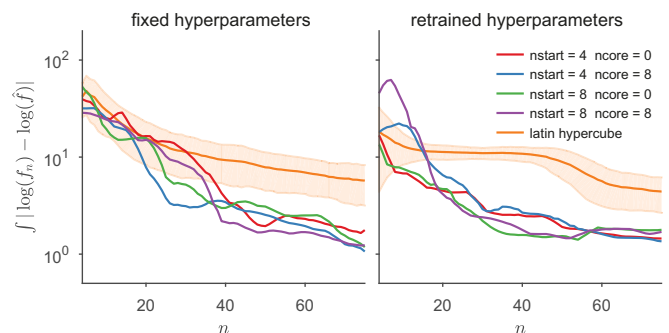

Fig. 5. Comparison of the errors between LH sampling and proposed algorithm under different parameters measured against the exact pdf for the case where $m = 2$ in Eq. 15. The parameter ncore is the number of core iterations performed according to an L_2 metric, and nstart is the initial dataset size (where the points are sampled from an LH design).
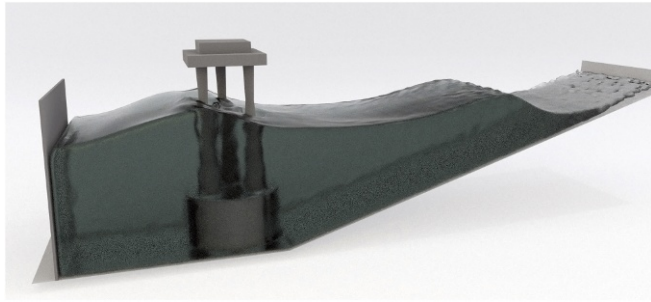
Fig. 6. SPH simulation at $t = 103.5$ s with $\theta_1 = 4.63$ and $\theta_2 = 0.662$.

$\sigma_n(\theta)$, such as the same criterion, but without the logarithmic transformation of the densities, we would have

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

where F_θ is the cumulative distribution function of θ. The corresponding integration is shown in Fig. 2, *Right*. In such a 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.

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 3D 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 is prohibitively expensive, since the simulation times of experimental runs are on the order of several hours.

Nonlinear Oscillator Driven by Correlated Stochastic Noise. Consider the nonlinear oscillator

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

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

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

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 *SI Appendix*, 5) to obtain

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

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 $\delta = 1.5$, $\alpha = 1.0$, $\beta = 0.1$, $x_1 = 0.5$, and $x_2 = 1.5$, 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 64,000 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 an LH design. To capture the main mass of the pdf, before focusing on the tails of the distribution, we perform 12 iterations using the Q' distance in Eq. 8 before moving on to the criterion using the distance metric Q in Eq. 6 involving 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.

We have explored the convergence properties of the algorithm, and in Fig. 5, we compare the proposed sampling method to space-filling LH 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 SD 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.
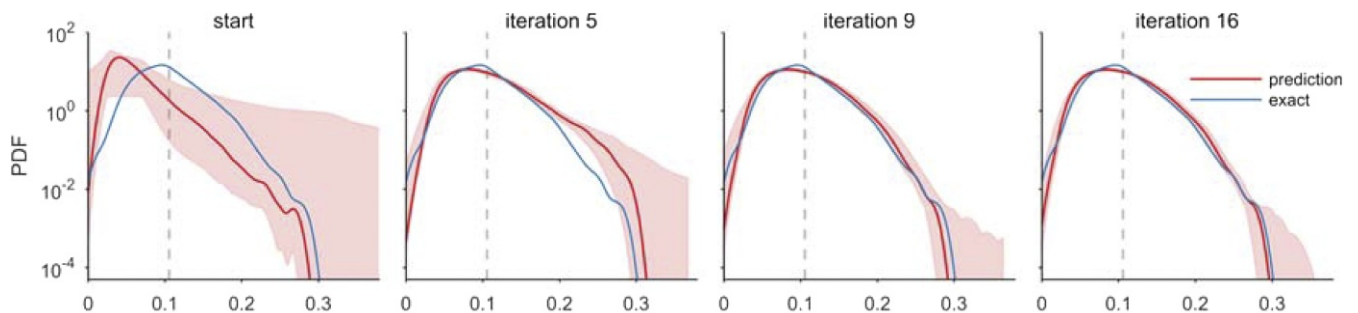

Fig. 7. Progression for pdf density prediction for the force variable.

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, 3D numerical simulations of Navier–Stokes using the smoothed particle hydrodynamics (SPH) method (15) (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 *SI Appendix*, 6.

Wind-generated ocean waves are empirically described by their energy spectrum. Here, we consider irregular seas with JONSWAP spectral density (see *SI Appendix*, 6 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 of primary wave groups, each characterized by a random group length scale L and height A (see, e.g., ref. 16). 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 ref. 16, we describe these primary wave groups by the representation $u(x) = A sech(x/L)$, 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. **1**. 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 ref. 17. Specifically, by generating many realizations consistent with the used spectrum, we use a group detection algorithm to identify coherent group structures in the field along with their length scale 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 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:

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

In Fig. 7, we show the results of the progression of density prediction for the force variable. In these experiments, we begin by arbitrarily selecting four initial sample points from an LH sampling strategy. Next, we perform four 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 uses 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 four 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 *SI Appendix*, Figs. S5–S7, we also present the sampled map where it can be seen that the algorithm selects points associated with large forces and nonnegligible probability of occurrence. In the same figures, results for the momentum and an additional spectrum are included. Note that for this problem, the GPR operates on the logarithm of the observable because the underlying function is always positive.

Conclusions

We developed and studied a computational algorithm for the evaluation of extreme event statistics associated with systems that depend on a set of random parameters. The algorithm is practical for high-dimensional systems but with parameter spaces of moderate dimensionality and where each sample is very expensive to obtain relative to the optimization of the next sample. The method 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 prove asymptotic convergence of the algorithm and provided analysis for its asymptotic behavior. We also demonstrated its applicability to two problems, one involving a demanding system with millions degrees of freedom.

1. Fouque J, Papanicolaou G, Sircar R, Sølna K (2011) *Multiscale Stochastic Volatility for Equity, Interest-Rate and Credit Derivatives* (Cambridge Univ Press, Cambridge, UK).
2. Liu P (2007) A chronology of freaque wave encounters. *Geofizika* 24:57–70.
3. Hense A, Friederichs P (2006) Wind and precipitation extremes in Earth's atmosphere. *Extreme Events in Nature and Society*, eds Albeverio S, Jentsch V, Kantz H (Springer, Berlin, Germany), pp 169–187.
4. Nicodemi M (2012) *Extreme value statistics*. Computational Complexity: Theory, Techniques, and Applications, ed Meyers RA (Springer, New York), pp 1066–1072.
5. Varadhan S (1984) Large deviations and applications, *CBMS-NSF Regional Conference Series in Applied Mathematics* (Society for Industrial and Applied Mathematics, Philadephia).
6. Dematteis G, Grafke T, Vanden-Eijnden E (2018) Rogue waves and large deviations in deep sea. *Proc Natl Acad Sci USA* 115:855–860.
7. Sobczyk K (1991) *Stochastic Differential Equations*, Mathematics and Its Applications (East European Series) (Springer, The Netherlands).
8. Masud A, Bergman L (2005) Solution of the four dimensional Fokker-Planck equation: Still a challenge. *ICOSSAR* 2005:1911–1916.
9. Franzke C (2017) Extremes in dynamic-stochastic systems. *Chaos* 27:012101.
10. Majda A (2016) *Introduction to Turbulent Dynamical Systems in Complex Systems* (Springer International Publishing, Basel).
11. Chen N, Majda A (2017) Simple stochastic dynamical models capturing the statistical diversity of El Niño southern oscillation. *Proc Natl Acad Sci USA* 114:1468–1473.
12. Mohamad M, Cousins W, Sapsis T (2016) A probabilistic decomposition-synthesis method for the quantification of rare events due to internal instabilities. *J Comput Phys* 322:288–308.
13. Mohamad M, Sapsis T (2015) Probabilistic description of extreme events in intermittently unstable dynamical systems excited by correlated stochastic processes. *SIAM/ASA J Uncertain Quantification* 3:709–736.
14. Rasmussen C, Williams C (2005) *Gaussian Processes for Machine Learning*, Adaptive Computation and Machine Learning (MIT Press, Cambridge, MA).
15. Crespo AJC (2015) DualSPHysics: Open-source parallel CFD solver based on smoothed particle hydrodynamics (SPH). *Comput Phys Commun* 187:204–216.
16. Cousins W, Sapsis T (2015) Unsteady evolution of localized unidirectional deep-water wave groups. *Phys Rev E* 91:063204.
17. Cousins W, Sapsis T (2016) Reduced-order precursors of rare events in unidirectional nonlinear water waves. *J Fluid Mech* 790:368–388.