

A stochastically precluded Karhunen-Loève representation for recovering extreme statistics in ship dynamics

Stephen Guth, Massachusetts Institute of Technology, sguth@mit.edu
Themistoklis P. Sapsis, Massachusetts Institute of Technology, sapsis@mit.edu

ABSTRACT

Several methods have been proposed in the literature for the representation and/or parametrization of random seas and the subsequent efficient reliability characterization of relevant engineering systems, e.g. ship motions and loads, offshore platforms, etc. The typical approach involves the expansion of the random wave field in terms of a Karhunen-Loève (KL) series, consisting of random coefficients (to represent the uncertainty) and deterministic wave-groups with prescribed time duration (that contain the temporal dependence). While these ideas successfully realize statistics consistent with the underlying spectrum, the choice of time duration is caught between the Scylla of slow KL series convergence for large T and the Charybdis of transient behavior for small T . To circumvent this issue, we propose the new notion of spectrum-consistent stochastically-precluded (SCSP) wave-groups. These are short-duration wave-groups obtained from a KL expansion but with a stochastically extrapolated preceding part (stochastic-prelude) that allows for the system to reach a proper statistical steady state, representative of the waves that have preceded before it enters the wave-group. We study the proposed method in the context of ship motions and loads (in particular pitch and vertical bending moments) and illustrate its favorable properties both in terms of convergence rate but also in terms of statistical representation skill for arbitrary quantities of interest.

Keywords: Vertical Bending Moment; Extreme Events; Karhunen Loève Expansion; Gaussian Process; Kriging

1 INTRODUCTION

Structural design and reliability analysis of marine systems require careful study of the effects of stochastic sea states on structural models, and in particular the effects of extreme waves [13]. While much work has gone into describing the higher order statistics of sea states (e.g. [19, 7]), describing the output extrema of nonlinear dynamical models, especially with memory effects is an open problem. There are currently two complementary approaches to extreme response assessment: one is to study the extreme value properties of the response itself with the use of exactly solvable prototype systems that are simple enough to solve analytically but also complex enough to

capture the essential physics (see e.g. [4]).

The other one is focused on the environment leading to extreme response. This is implemented through the study of wave groups as elementary units of a sea state [6, 17]. Work has been done to model the shapes of wave-groups, and equations that describe their behavior from a theoretical perspective. Several of these ideas have been used successfully for the assessment of ship stability [20, 3, 2]. Another approach involves the generation of short but extreme wave time series based on non-uniform phase distributions [1]. These result in large linear responses that can potentially indicate the behavior of a ship in different sea conditions. Two-parameter wave-groups have also been employed to quantify

extreme event statistics. However, these operate well only for ocean structures subjected to narrow-banded spectra that can be effectively parametrized by low-dimensional representations [12, 10].

In this work our aim is to expand on the second class of methods (i.e. model the environment leading to extreme responses), by building a finite-time Karhunen-Loève (KL) expansion of arbitrarily broad spectra that is capable to capture accurately the non-Gaussian structure of any quantity of interest related to ship motions or loads. While KL expansions have been employed previously in the context of ship stability [3] it has not been assessed for representing tail probabilities for the response and load statistics in the context of nonlinear ship motions.

In addition, particular emphasis in this work is given to the fact that beyond the effect of random excitation there is also the effect of initial conditions, i.e. the state of the system when it ‘meets’ the wavegroup. To model this effect we introduce the notion of a spectrum-consistent stochastic prelude (SCSP) that complements each deterministic wavegroup. The derived waves are combined with a high fidelity ship dynamics code: LAMP [18]. A low-dimensional (or even moderate-size, e.g. $O(10 - 100)$) parametrization that is capable to capture extreme event statistics paves the way for the utilization of statistical techniques to significantly reduce the number of samples/simulations/experiments required to capture the output statistics [16]. These have recently been made possible even for moderate-dimensionality problems through the development of output-weighted methods [5].

2 KARHUNEN LOÈVE CONSTRUCTION

As a first step, we will need a low dimensional representation of a realization of a random process on a finite interval. We will do this by first using the random process to define a set of basis vectors. Second, we will represent an

particular realization as a set of coefficients. To this end, we present the KL Theorem:

Theorem 1 (Karhunen Loève [9])

Consider the continuous time random process $\zeta(t)$ on $[a, b]$ which is zero mean and square integrable on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Define the covariance function

$$K_{\zeta}(s, t) = \mathbb{E}[X(s)X(t)] \quad (1)$$

with corresponding integral operator

$$T_{K_{\zeta}}\phi(t) = \int_a^b K_{\zeta}(t, s)\phi(s)ds \quad (2)$$

Then by Mercer’s theorem, the operator $T_{K_{\zeta}}$ has an orthonormal basis of eigenvectors $\{e_i(t)\}$ and corresponding eigenvalues $\{\lambda_i\}$. The variables

$$Z_i = \int_a^b \zeta(t)e_i(t)dt \quad (3)$$

(“the coefficients”) are centered orthogonal random variables with

$$\text{Var}(Z_i) = \mathbb{E}[Z_i^2] = \lambda_i \quad (4)$$

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

$$\zeta(t) = \sum_{i=1}^{\infty} Z_i e_i(t) \quad (5)$$

In summary, the eigenvectors of the covariance matrix of the sea surface form an orthonormal basis. The decomposition of $\zeta(t; \omega)$ onto this basis produces a set of (centered) orthogonal (random) coefficients.

In particular, this means that we can change back and forth between the function representation $\zeta(t; \omega)$ and the coefficient representation $Z_i(\omega)$. If we truncate after only a finite number of coefficients, we expect to retain ‘most’ of the variance in the reconstructed random process $\tilde{\zeta}(t; \omega)$, where ‘most’ is expressed in terms of the eigenvalues λ_i .

Consider the interval $[-a, a]$, sampled discretely at n_t equally spaced points. Then we

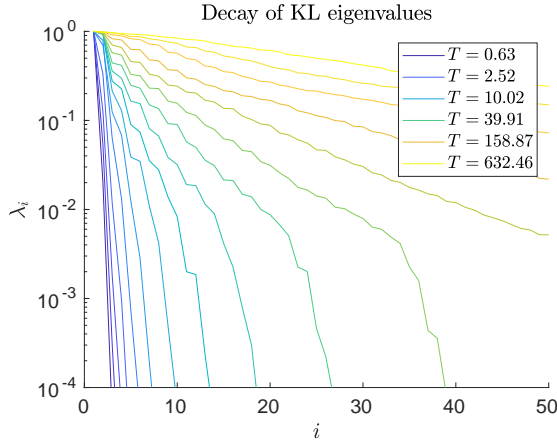


Figure 1: Exhibition of the decay of the discrete KL eigenvalues for different temporal windows T for a JONSWAP spectrum.

have $t_i = -a + \frac{i-1}{n_t-1}(2a)$. Sample the random process $\zeta(t; \omega)$ at each of these points, so that $\zeta_i(\omega) = \zeta(t_i; \omega)$ is an n_t -dimensional random vector.

Define the covariance matrix \mathbf{K} as

$$K_{jk} = \mathbb{E}[\zeta_j \zeta_k] \quad (6)$$

In the discrete case, we can approximate the expectation by sampling a large number of independent copies of the random process:

$$\mathbb{E}[\zeta_i \zeta_j] = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N \zeta_i(\omega_k) \zeta_j(\omega_k) \quad (7)$$

$$\hat{K}_{jk}^N = \frac{1}{N} \sum_{k=1}^N \zeta_i(\omega_k) \zeta_j(\omega_k) \quad (8)$$

We will elide explicit dependence on the number of samples N used to compute the covariance matrix.

Finally, the eigenvectors of the matrix $\hat{\mathbf{K}}$, \hat{e}_i^j are discrete approximations of the basis vectors $e^i(x)$, with corresponding discrete eigenvalues $\hat{\lambda}_i$. These eigenvalues correspond to the part of the total variance carried by the particular component.

The discrete random process ζ_i can be expressed in terms of the the basis vectors \hat{e}_i^j and a set of coefficients α_j by a discrete analogy

to equation 5:

$$\zeta_i = \sum_{j=1}^J \alpha_j \hat{e}_i^j \quad (9)$$

In turn, the coefficients α_j can be determined by the inner product

$$\alpha_j = \langle \zeta_i, \hat{e}_i^j \rangle = \sum_{i=1}^I \zeta_i \hat{e}_i^j \quad (10)$$

We will note, that in the special case of a Gaussian process, the coefficients α_j are not merely orthogonal with mean 0 and variance $\hat{\lambda}_j$, but are jointly Gaussian themselves. From here on, we will work entirely in the discrete setting, and will drop the circumflexes on $\hat{\mathbf{K}}$, \hat{e}_i^j , and $\hat{\lambda}_j$.

3 SPECTRUM CONSISTENT STOCHASTIC PRELUDING

Sea states may be described by a power spectrum, which gives the energy content of the surface gravity waves as a function of temporal frequency. Here we employ the JONSWAP spectrum (Fig. 2a):

$$S_J(\omega) = \frac{\alpha g^2}{\omega^5} \exp \left[-\frac{5}{4} \left(\frac{\omega_p}{\omega} \right)^4 \right] \gamma^r \quad (11)$$

$$r = \exp \left[-\frac{(\omega - \omega_p)^2}{2\sigma^2 \omega_p^2} \right]$$

where constants are given by

$$\alpha = 0.076 \left(\frac{U_{10}^2}{Fg} \right)^{0.22}, \quad \omega_p = 22 \left(\frac{g^2}{U_{10}F} \right)^{\frac{1}{3}}$$

$$\gamma = 3.3, \quad \sigma = \begin{cases} 0.07 & \omega \leq \omega_p \\ 0.09 & \omega > \omega_p \end{cases}$$

We emphasize that this choice of spectrum is not dictated by any particular need or constrain associated with the mathematical modeling.

For this work, we will assume long crested head seas, which correspond to zero angular spread and collapse to a one-dimensional

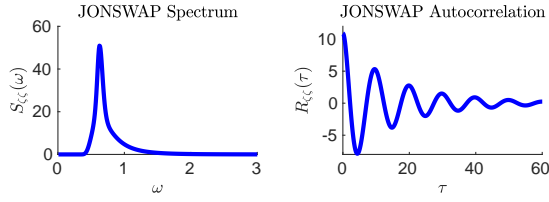


Figure 2: a) Power spectral density for the JONSWAP spectrum. b) Autocorrelation function.

description. We will also make use of the autocorrelation function (Fig. 2b), which is defined for a stationary random process as

$$R_{\zeta\zeta}(\tau) = \mathbb{E} [\zeta(t; \omega)\zeta(t + \tau; \omega)]. \quad (12)$$

Autocorrelation is analytically connected with the spectrum through the Fourier relation

$$R_{\zeta\zeta}(\tau) = \mathcal{F}^{-1}\{S_J(\omega)\}. \quad (13)$$

Deterministic part of the samples

We split the waveform construction into two steps: construction of the fixed region, and extrapolation to the variable prelude and postlude.

We will note that sea surface gravity waves are governed by a hyperbolic advection-dispersion equation. This means that the solution $\zeta(x, t)$ in the well behaved domain \mathbb{R} is determined by its behavior on a Cauchy surface. For simplicity, we choose the line $x = 0, t \in \mathbb{R}$, which we call the Eulerian Cauchy surface. If we fix the sea surface elevation on this line, we can determine the sea elevation at any $(x, t) \in \mathbb{R}^2$ using the dispersion relation

$$\omega^2 = gk \tanh(kH). \quad (14)$$

We generate samples of the random process $\zeta(0, t)$ from the power spectrum using the cosine construction, given by

$$\hat{\zeta}(t) = \sum_{i=1} \sqrt{2S_{\zeta\zeta}(\omega_i)\delta_\omega} \cos(\omega_i t + \phi_i) \quad (15)$$

where ω_i is a suitable discretization of angular frequency, δ_i is the frequency spacing, $S_{\zeta\zeta}(\omega)$

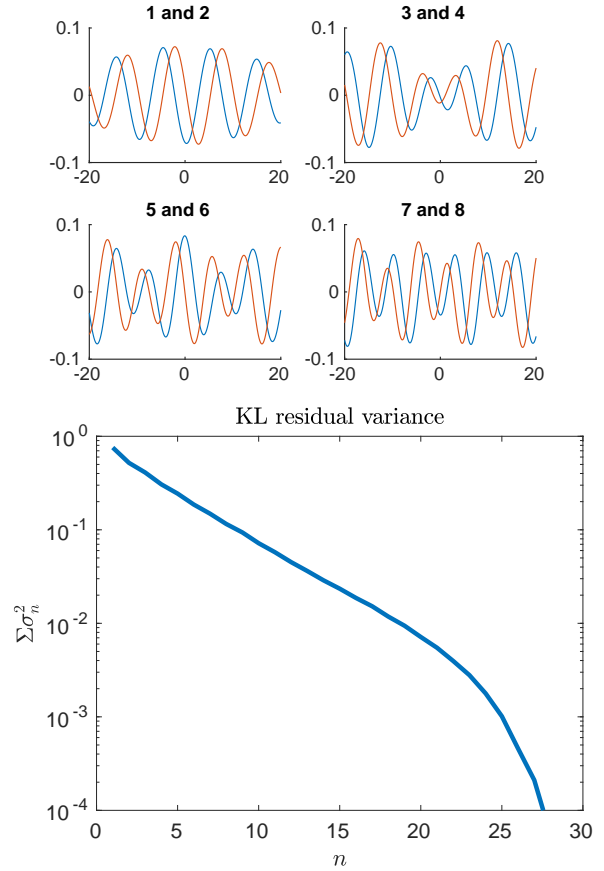


Figure 3: a) First 8 modes for the KL basis associated with JONSWAP spectrum and length $T = 40$ interval. b) Plot of the fractional residual variance after removing the first n KL modes.

is the power spectrum, and ϕ_i is a random phase uniformly distributed on $[0, 2\pi)$.

Realizations of this process exist for all time $t \in \mathbb{R}$ and are opaque parametrized by a (large) vector of phases $\{\phi_i\}$. To reduce this to a more efficient parametrization, we apply the finite-time KL reduction to this random process.

We truncate the spectrum of eigenvectors at J using a suitable criterion, such as retaining a certain fraction of the process variance. For our problem, we truncate after the first $J = 25$ modes, corresponding to a ‘knee’ in figure 3 approximately 99.5% retained variance.

By the KL theorem, we can sample this random process on the region $[0, 40]$ by gener-

ating J normally distributed random variables α_j , and then computing the sum

$$\zeta_i = \sum_{j=1}^J \alpha_j \sqrt{\lambda_j} e_i^j. \quad (16)$$

Gaussian Process Extrapolation

For each fixed set of random coefficients, $\alpha_j|_{j=1}^J$ we have a prescribed (or deterministic), finite-time sample. We now complement this deterministic signal with a spectrum-consistent stochastically-prelude (SCSP). This is generated moving ‘backwards’ from the beginning of the deterministic signal by performing a stochastic extrapolation. Specifically, for generation of the SCSP region, we follow the standard Gaussian Process sample generation procedure for stationary covariance kernel (see e.g. [15]).

Briefly, we begin with a fixed region of n_f equally spaced sample points. Next, we choose a memory length $n_m \leq n_f$. Then, we compute a Töplitz covariance matrix

$$\Sigma = \begin{bmatrix} R(0) & R(\Delta t) & \dots \\ R(\Delta t) & R(0) & \dots \\ R(2\Delta t) & R(\Delta t) & \dots \\ \dots & \dots & \dots \\ R((n_m - 1)\Delta t) & R((n_m - 2)\Delta t) & \dots \end{bmatrix} \quad (17)$$

where $R(\tau)$ is our previously computed auto-correlation function (section 3) with the subscript dropped. We also compute the marginal row

$$\Sigma^* = [R(n_m \Delta t) \quad R((n_m - 1)\Delta t) \quad \dots \quad R(\Delta t)]. \quad (18)$$

Note that while Σ and Σ^* depend on the memory length n_m , they are independent of the data (from the fixed region or otherwise). We extrapolate from the fixed region in an iterative manner:

1. Using the last n_m points, calculate the conditional mean and variance of the next point in the sequence.

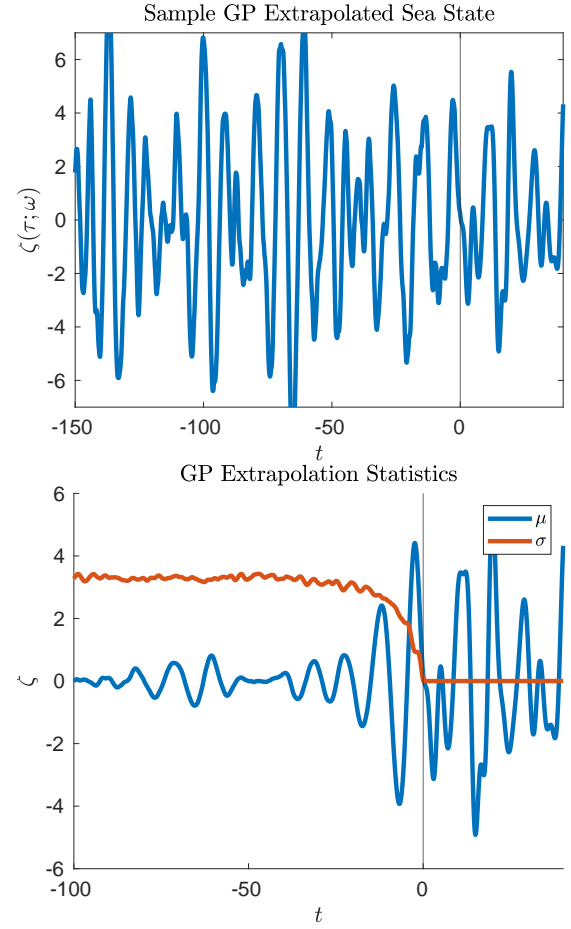


Figure 4: a) Sample extrapolation (for $t < 0$) by Gaussian Process sample generation (using the wave spectrum) conditioned on $t > 0$. b) Ensemble mean and standard deviation of all the extrapolated signals (corresponding to the same wavegroup for $t > 0$). Parameters: $n_m = 512$, $\Delta t = 0.078$, $n_{\text{samples}} = 1 \times 10^3$.

2. Sample the next point in the sequence from its conditional distribution.
3. Shift the ‘reading frame’ by Δt .

The mean and variance of the conditional posterior are given by

$$\mu = \Sigma^* \Sigma^{-1} \zeta \quad (19)$$

$$\sigma^2 = R(0) - \Sigma^* \Sigma^{-1} \Sigma^{*\top} \quad (20)$$

where ζ is the vector of the last n_m sampled points. Note that we may precompute σ^2 completely (posterior sampling variance is in-

dependent of the sample), and $\Sigma^* \Sigma^{-1}$ (approximately a convolutional stencil).

For best results, n_m ought not be larger than the size of the fixed region, but $n_m \Delta t$ ought be larger than the typical correlation length, that is, $|R(\tau)| \ll R(0) \quad \forall \tau > n_m \Delta t$.

Figure 4 shows the conditional mean and standard deviation for a set of Gaussian Process extrapolated signals. Note that in the fixed region $t \in [0, 40]$, the standard deviation vanishes and the mean is simply the KL constructed waveform. After the extrapolation begins, the mean decreases, but long range order (and statistical error) causes some residual spread in the mean. Finally, the signal variance increases quickly to a steady state over the course of a small number of wave periods.

Advection and Dispersion

So far, we have used Gaussian Process extrapolation to construct a set of spectrum-consistent stochastically-primed time series describing the sea surface elevation ζ at $x = 0$ for a time interval of interest $[t_i, t_f]$. In particular, the stochastic process has a fixed region at $x = 0, t \in [0, 40]$. However, LAMP simulates a vessel moving in the x -direction with some non-zero velocity.

Sea surface waves advect (propagate) and deform with time, according to the dispersion relation given in equation 14. For our spectrum, across the time period of interest ($T = 40$), deformation is insignificant. However, in order to ensure that the vessel encounters the fixed region, we should arrange the vessel trajectory so that it passes through the spatial point $x = 0$ during the interval $[0, 40]$.

For our simulation, we choose a fixed x velocity $u_x = 5.144$ m/s. We choose $x_i n$ so that $x_i n + u_x(T/2 - t_i) = 0$, where t_i is the initial (negative) simulation time, at the beginning of the ramp zone, and T is the temporal duration of the fixed region. In figure 5, we show the fixed region in the space-time domain, overlaid with the vessel trajectory.

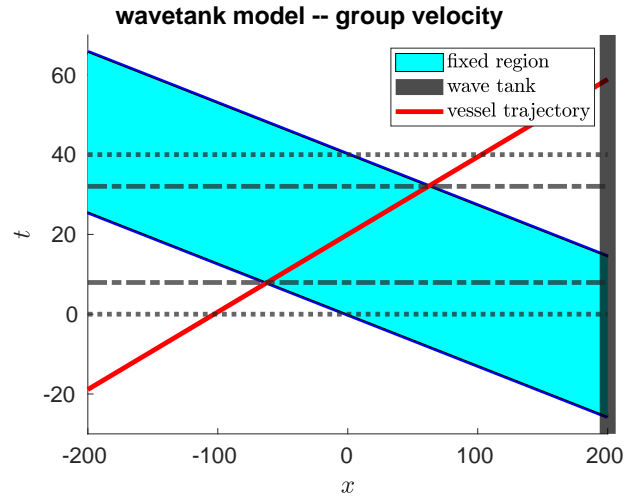


Figure 5: Representation of the fixed region in space and time, with overlaid trajectory of vessel.

Finally, while we consider the fixed region corresponding to $T \in [0, 40]$ in the Eulerian (input) frame of reference, we consider the corresponding output region to be $T \in [11.5, 18.5]$ in the Lagrangian frame of reference. This interval of length $T_L = 17$ in the output time series is long enough to observe approximately three successive wavelengths.

We emphasize: the sea state we fix is not the instantaneous sea elevation that the vessel ‘sees’ (Lagrangian frame), but the sea environment that the vessel passes through (Eulerian frame).

4 STATISTICS OF VESSEL RESPONSE

Effect of Uncertainty Before the Wavegroup

Our goal with stochastic prelude is to create a set of LAMP input files that have a common fixed region, but variable prelude regions. Simulations in LAMP were performed using the ONR Topsides Faredhull. Our procedure is the following:

- 1 Precompute the KL basis functions
- 2 For each fixed region --
- 3 Draw a set of KL coefficients
- 4 Reconstruct the fixed region

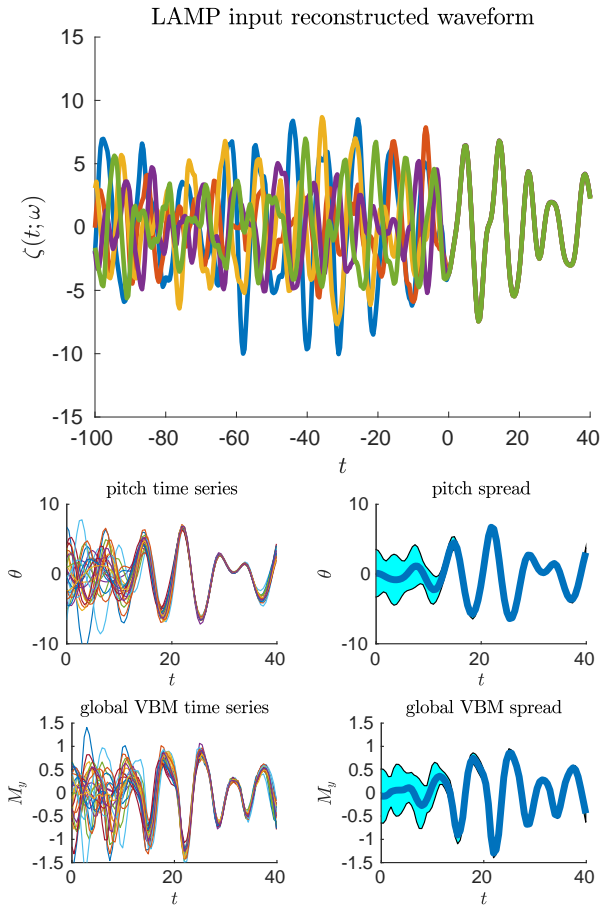


Figure 6: a) Fixed region ($t > 0$) and variable precluded region ($t < 0$) for **typical** SCSP sea surfaces. b) Representation of the statistics associated with exhibited fixed region. first column: time series for pitch and VBM; second column: mean and 1σ spread.

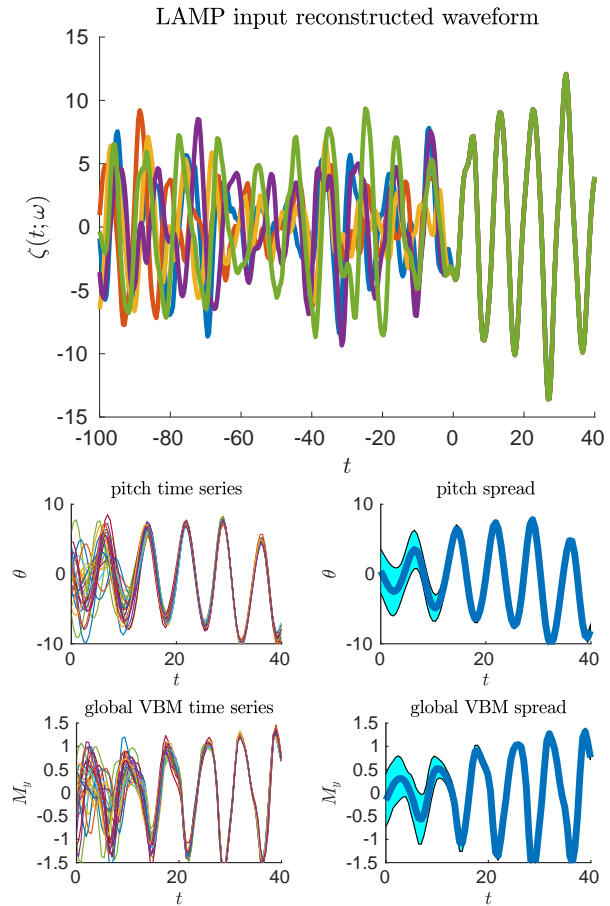


Figure 7: a) Fixed region ($t > 0$) and variable precluded region ($t < 0$) for **extreme** SCSP sea surfaces. b) Representation of the statistics associated with exhibited fixed region. first column: time series for pitch and VBM; second column: mean and 1σ spread.

- 5 For each prelude --
- 6 Extrapolation a Spectrum-
Consistent waveform
- 7 Discrete Fourier Transform
(DFT) for sum of sinusoids
- 8 Construct the LAMP input with
linear superposition of waves

We note that while the KL coefficients in step [3] are independent Gaussian random variables, we can instead use importance sampling, or even some structured sampling method to choose them, such as Latin Hypercube Sampling. Alternately, we may use an active search technique to use previous

simulation results to inform later simulation parameters [5].

From here out, we will examine the time series that LAMP saves as output. That is, while we perform the SCSP procedure to create realizations of a random process for the LAMP input, we will examine the statistics of two LAMP output time series. We will examine 1) θ , the vessel pitch angle and 2) M_y , the total Vertical Bending Moment (VBM) across the vessel.

In figure 6, we show the time series associated with a number of SCSP sea surfaces drawn from the same fixed region. Then, we show summary statistics for the corre-

sponding LAMP outputs. The region $T \in [11.5, 28.5]$ corresponds to the passage of the vessel through the fixed region.

The first column of figure 6 shows the output time series overlaid. It is clear that time series begin to converge in the fixed region, and afterward take some time to fully diverge. The second column shows this more carefully: the central dark blue line is the mean of the LAMP output time series, and the shaded region represents the 1σ neighborhood.

For comparison, 7 shows the same time series and output statistics for a different fixed region, chosen to be unusually large in magnitude.

We draw two conclusions from this exhibit that includes an exhaustive numerical investigation for a large number of KL samples - not presented here due to limited space. First, spectrum consistent stochastic precluding solves the consistent initial conditions problem for LAMP simulations. That is, we can fix the sea surface for only a small interval, but still provide LAMP with consistent inputs that provide a wide variety of internal "states" without risking a situation where the vessel is nonphysically above or below the water line.

Moreover, while spectrum consistent stochastic precluding leads to some variance in the recovered statistics in the fixed region (primarily due to memory effects), *this residual variance is both 1) 'small' and 2) much less significant, near local extrema, especially output extrema corresponding to significant waves. To this end, there is no need to employ multiple SCSP for each KL sample or model uncertainty due to initial conditions when the focus is on the statistics of local peaks.* For this reason we will employ, for each KL sample, a single randomly drawn SCSP.

In order to best reduce the set of output time series to a small number of summary statistics, we will extract the **global maximum and minimum** on the interval $T \in [11.5, 28.5]$. We chose this measure, instead of statistics representing one-point val-

ues or local extrema, because over a short interval directly estimating statistics other than the interval extrema is difficult without envelope approximations or parametric models. Figure 8 shows a comparison of the interval extrema with local extrema and one-point pdf for long time Monte Carlo simulations.

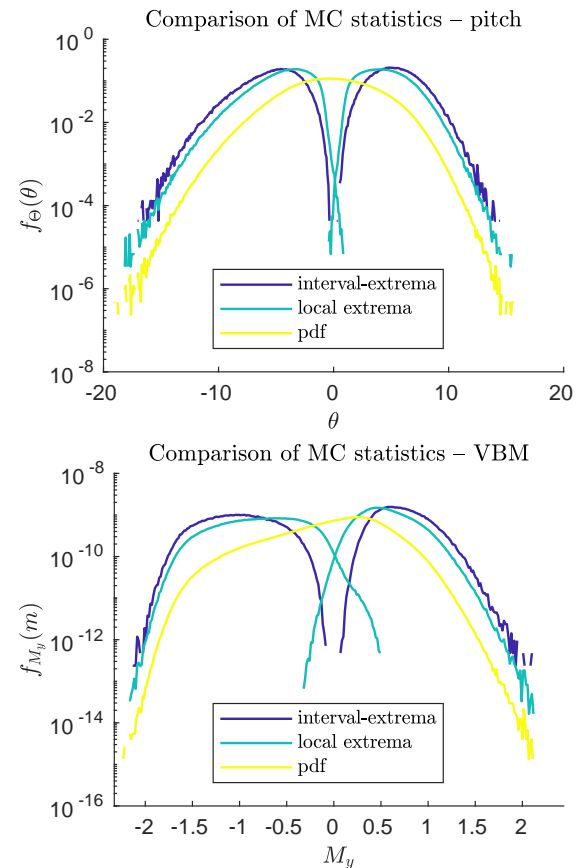


Figure 8: Comparison of our recovered statistics (output extrema over an interval) to two other common statistics: local extrema, and pdf.

Reconstruction of Tail Statistics

In the design of marine structures, we are typically interested in both the 'typical' response range, and the likelihood of tail responses—extreme events. Standard Monte Carlo experiment design—draw a number of sea states, simulate each in LAMP for a long period—does an adequate job of describing the peak of the output distributions through mean and

variance. However, Monte Carlo simulations run into an iron law in reproducing the distribution tails: to accurately resolve tails with probability p , Monte Carlo requires on the order of p^{-1} samples.

Importance sampling methods may provide an escape hatch. Instead of sampling input sea states according to the statistical steady state, we choose a new sampling distribution that emphasizes regions of the sample space expected to contribute to the tails to the output distribution. We then use this structured set of experiments in order to build a surrogate model (interpolation). Finally, we resample the surrogate model to estimate the pdf of the output statistics.

Importance Sampling The KL expansion is the starting point for our structured sampling. Truncated at $J = 25$, the joint distribution of Karhunen-Loève coefficients is independent and jointly Gaussian. In particular, the normalized α_j coefficients from equation 16 are independent identically distributed (iid) Gaussian random variables with $\mu = 0$ and $\sigma^2 = 1$.

The Monte Carlo sampling would draw J iid samples from $\mathcal{N}(0, 1)$, with extreme waves predictably sampled rarely. Instead, we sample from J iid random variables from the uniform distribution $\mathcal{U}(-z^*, z^*)$, where z^* is a judiciously chosen cut-off z-score. For the following work, we chose $z^* = 3$, as a balance between covering as much of the support of $\mathcal{N}(0, 1)$ as possible (99.5% of the probability mass for each coefficient) without unduly slowing convergence.

For each sample of coefficients, we construct a single input waveform using the SCSP procedure outlined above. We find that 1 SCSP realization-simulation is sufficient to recover extreme statistics, at the cost of increased error resolving the quiescent statistics. From the LAMP output, we extract the global minimum and maximum from the interval $t \in [11.5, 28.5]$. We concatenate the coeffi-

icients into a matrix of input data X , and we concatenate each output summary statistic as an output vector Y .

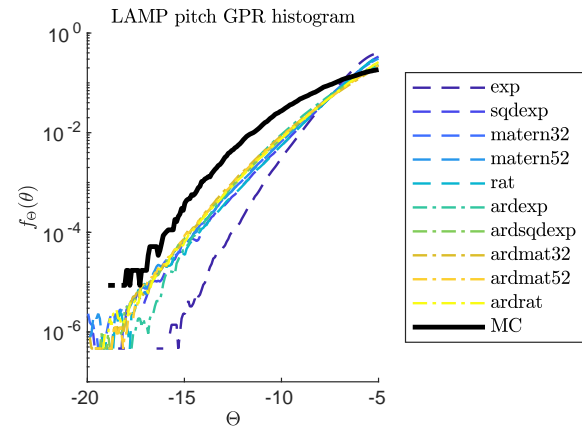


Figure 9: Comparison of recovered pdf tails for different choices of Gaussian Process kernel.

Surrogate Modeling Our goal with the surrogate model is to construct a function that interpolates (and in some cases extrapolates) between the structured sample points. To be specific, we wish to construct a function $\hat{f}(\alpha_1, \alpha_2, \dots, \alpha_n)$ that maps the KL coefficient representation of the fixed region to the extreme value of the output statistics on our interval of interest. That is to say, we want the surrogate $\hat{f}(\cdot)$ to “learn” the relationship between the shape of the sea waveform (on the fixed region) and the maximum and minimum values of the vessel structural responses (over the corresponding time interval).

We will encounter three difficulties: first, $J = 25$ is not low dimensional; second, our worry about distribution tails will require extrapolation; and third, we expect the dependencies among and between the J different coefficients to be non-trivial. A fourth minor difficulty is that the stochastic nature of the precluding operation will introduce a small noise term.

To best meet these goals, we choose a kriging approach making use of Gaussian Process regression (GPR) with Automatic Relevance

Determination (ARD). In GPR, we use a Machine Learning framework to select the optimal kernel function from a given class. The kernel, or covariance, is a function measuring ‘similarity,’ and can be thought of as a reciprocal distance function in an abstract space. Common kernel classes include exponential, squared exponential, and a fractional family developed by Matérn [8].

ARD is a technique to use machine learning to automatically determine which dimensions are more significant [11, 14]. The KL spectrum λ_j ranks the modes by energy, so it is reasonable to expect that component distances in the less energetic modes ought contribute less than component distances in the more energetic modes. Indeed, figure 9 shows that the ARD kernel classes generally slightly outperform their non-ARD equivalents.

The major advantage of a surrogate model is that evaluating a trained Gaussian Progress is many orders of magnitude faster than conducting a full LAMP simulation. While there is a time cost associated with training the GPR model, this cost is not prohibitive. For the number of samples we encountered during testing, training a GPR model with the ARD Matern 5/2 kernel took the same amount of time as performing 1 – 5 short KL LAMP simulations.

Resampling Finally, we construct the recovered output pdfs by resampling. First, we sample a large number n_s of samples from the true joint coefficient distribution. Next, we evaluate our surrogate model for each of these input samples. Finally, we histogram the outputs.

Because we draw the resampled inputs from the steady state distribution ($\mathcal{N}(0, 1)^J$), we do not require any importance weighting.

For a control and point of comparison, we compared the resampled pdfs with pdfs generated from traditional long time Monte Carlo sampling. To generate these pdfs, we first generated input sea states directly from the

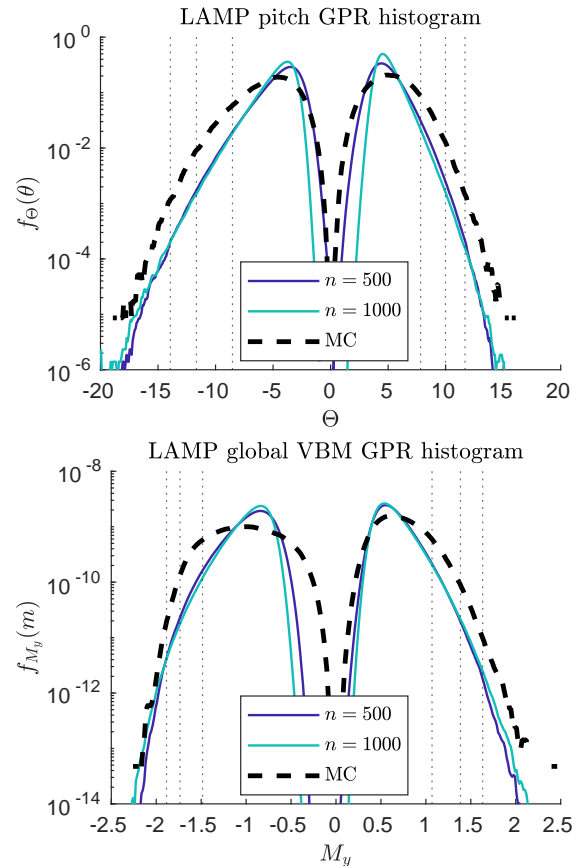


Figure 10: Comparison of recovered resampled pdf for different numbers of samples.

spectrum. Then, we ran LAMP simulations for long time periods. We discarded an initial transient period from each simulation.

Then, we subsampled a number of $T_L = 17$ intervals from the resulting output time series. We chose subsampling intervals to be well separated by the coherence length of the input process, so that each interval was approximately independent. Finally, from each subsampled interval we extracted the global maximum and global minimum for each output statistic. We expect that the histogram of all such interval-extrema quantities generated via the Monte Carlo method should converge to the same histogram generated from the surrogate resampling method.

Analysis of Tail Recovery

Figure 10 compares the recovered resampled pdf tails to the direct Monte Carlo pdfs for different numbers of full simulations. By $n_s = 500$, the recovered pdf tails have nearly converged for both the vessel pitch and VBM. On consumer grade desktop hardware, 500 LAMP simulations with the SCSP procedure can be performed in approximately 8 hours.

On the other hand, the converged pdf tails appear to have a small but consistent underestimate although the shape of the tail is captured adequately. The magnitude of this underestimate varies with the precise choice of Gaussian Process kernel and size of sampling region, and additionally appears to vary with number of retained KL modes. More work is needed to quantify the tail errors, and characterize the best practices for the reconstruction step.

5 CONCLUSION

In this paper, we developed a method to parametrically design realizations of a random process with stochastic extrapolation, consistent with the underlying spectrum. This prelude ensures that numerical simulation of short samples corresponding to a KL expansion, has random initial conditions that are consistent with the statistical steady state, and ameliorates the effects of transient behaviors on output time series.

We applied the SCSP procedure to the ship dynamics code LAMP, and estimated the extrema statistics for the pitch and vertical bending moment by an importance sampling method, maintaining 25 KL modes, a number that is sufficient to model even broad spectra. For the considered spectral parameters and ship geometries we did not identify a significant effect of the SCSP variability to the extreme events resulted by each generated wavegroup.

While previous work by our group [12, 10] has demonstrated that even very low-

dimensional parametrization of wavegroups can have some success in certain cases (narrow banded spectra) and for specific observables (e.g. ship pitch angle), it is not a general approach that can be used for statistical quantification of ship motions and loads [10]. The presented work is the first attempt to model the full pdf of pitch and VBM based on a truncated KL expansion. However, more work is needed to achieve better agreement with the direct MC simulations. Beyond that important issue of accuracy, an obvious future step is to use active (sequential) sampling techniques, which use an intermediate surrogate in order to choose later sample points to optimize information recovery about the pdf tails [5].

6 ACKNOWLEDGMENTS

The work described in this paper has been funded by the Office of Naval Research (ONR) under Dr. Woei-Min Lin. The authors are grateful to Kenneth Weems and Vadim Belenky (NSWC-Carderock) for providing support for the LAMP simulations, as well as for many stimulating discussions.

References

- [1] L. K. Alford and A. W. Troesch. Generating extreme ship responses using non-uniform phase distributions. *Ocean Engineering*, 2009.
- [2] P. A. Anastopoulos and K. J. Spyrou. Evaluation of the critical wave groups method in calculating the probability of ship capsizing in beam seas. *Ocean Engineering*, 187(January):106213, 2019.
- [3] P. A. Anastopoulos, K. J. Spyrou, C. C. Bassler, and V. Belenky. Towards an improved critical wave groups method for the probabilistic assessment of large ship motions in irregular seas. *Probabilistic Engineering Mechanics*, 44:18–27, 2016.

- [4] V. Belenky, D. Glotzer, V. Pipiras, and T. P. Sapsis. Distribution tail structure and extreme value analysis of constrained piecewise linear oscillators. *Probabilistic Engineering Mechanics*, 57:1–13, 2019.
- [5] A. Blanchard and T. Sapsis. Output-Weighted Optimal Sampling for Bayesian Experimental Design and Uncertainty Quantification. *SIAM ASA J. of Uncertainty Quantification*, 2021.
- [6] P. Boccotti. On mechanics of irregular gravity waves. *Atti della Accademia Nazionale dei Lincei*, 19:110–170, 1989.
- [7] G. Z. Forristall. Wave crest distributions: Observations and second-order theory. *American Meteorology Society*, 30:1931–1943, 2000.
- [8] M. G. Genton. Classes of kernels for machine learning: A statistics perspective. *J. Mach. Learn. Res.*, 2:299–312, Mar. 2002.
- [9] P. Holmes, J. L. Lumley, and G. Berkooz. *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*. Cambridge University Press, 1998.
- [10] S. Kevin. *Adaptive Sequential Sampling for Extreme Event Statistics in Ship Design*, MSc Thesis. MIT, Mechanical Engineering, 2018.
- [11] D. J. C. MacKay. Bayesian Interpolation. *Neural Computation*, 4(3):415–447, 05 1992.
- [12] M. A. Mohamad and T. P. Sapsis. Sequential sampling strategy for extreme event statistics in nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 115(44):11138–11143, 2018.
- [13] A. Naess and T. Moan. *Stochastic Dynamics of Marine Structures*. Cambridge University Press, 2013.
- [14] R. Neal. *Bayesian Learning for Neural Networks*. Lecture Notes in Statistics. Springer New York, 2012.
- [15] D. B. Percival. Simulating Gaussian Random Processes with Specified Spectra. *Computing Science and Statistics*, 24:534–538, 1992.
- [16] T. P. Sapsis. Statistics of extreme events in fluid flows and waves. *Annual Review of Fluid Mechanics*, 53(1):null, 2021.
- [17] P. D. Sclavounos. Karhunen-Loeve representation of stochastic ocean waves. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2012.
- [18] Y. Shin, V. Belenky, K. Weems, W. Lin, and A. Engle. Nonlinear time domain simulation technology for seakeeping and wave-load analysis for modern ship design. *Transactions, Society of Naval Architects and Marine Engineers*, 111, 2003.
- [19] M. A. Tayfun. Narrow band nonlinear sea waves. *Journal of Geophysical Research*, 85(C3):1548–1552, 3 1980.
- [20] N. Themelis and K. Spyrou. Probabilistic assessment of ship stability. *Transactions - Society of Naval Architects and Marine Engineers*, 117:181–206, 2007.