

ORIGINAL ARTICLE

In Search of Rogue Waves: A Novel Proposal Distribution for Parallelized Rejection Sampling of the Truncated KdV Gibbs Measure

Nicholas J. Moore | Brendan Foerster

Department of Mathematics, Colgate University, Hamilton, New York, USA

Correspondence: Nicholas J. Moore (nickmoore83@gmail.com)

Received: 26 November 2024 | **Revised:** 17 February 2025 | **Accepted:** 6 March 2025

Funding: This study was supported by the Office of Naval Research (award number N000142412617) and the Colgate University.

Keywords: extreme events | rogue waves | sampling algorithm

ABSTRACT

The Gibbs ensemble of the truncated KdV (TKdV) equation has been shown to accurately describe the anomalous wave statistics observed in laboratory experiments, in particular the emergence of extreme events. Here, we introduce a novel proposal distribution that facilitates efficient rejection sampling of the TKdV Gibbs measure. Within parameter regimes accessible to laboratory experiments and capable of producing extreme events, the proposal distribution generates 1–6 orders of magnitude more accepted samples than does a naive, uniform distribution. When equipped with the new proposal distribution, a simple rejection algorithm enjoys key advantages over a Markov chain Monte Carlo algorithm, include better parallelization properties and generation of uncorrelated samples.

1 | Introduction

Abnormally large water waves—variously known as anomalous waves, extreme waves, or rogue waves—have in recent decades captured the interest of the scientific community, ocean practitioners, and the public at large [1–8]. Although several mechanisms for their existence have been proposed, one line of research has demonstrated that abrupt changes in bottom topography can trigger anomalous wave activity [7, 9–24]. In particular, laboratory measurements of topography-induced anomalous wave statistics [7, 16] were successfully reconciled with a theoretical framework based on statistical and dynamical analysis of the variable-depth Korteweg–De Vries (KdV) equation [11, 12, 16], a partial differential equation (PDE) that describes nonlinear dispersion of shallow waves [25–27]. Specifically, the theory relies on a Galerkin truncation of KdV, known as the truncated KdV (TKdV) system [16].

The framework exploits the Hamiltonian structure of TKdV to define a Gibbs measure from which wave events can be sampled without the need to directly simulate the underlying PDEs [16]. In particular, the variable-depth TKdV equation gives rise to separate Gibbs ensembles upstream and downstream of the depth change. The inverse temperatures of these two systems are linked through a statistical matching condition enforced at the depth change [11, 16]. This framework has been shown to successfully predict outgoing wave statistics as a function of the incoming wave field characteristics and the size of the depth change.

While the framework has proven successful in predicting system behavior, the task of sampling the Gibbs ensembles presents some challenges. Previous studies have employed brute-force sampling by taking the spectrally uniform measure as a proposal distribution in a rejection or sampling-importance resampling (SIR) algorithm [16, 28], or they have employed more

sophisticated methods such as Markov chain Monte Carlo (MCMC) [11, 16]. However, each of these approaches has drawbacks. Taking the proposal distribution of a rejection (or SIR) algorithm to be uniform in spectral space results in an extremely low acceptance rate (or extremely low sample weights) in the parameter regime of greatest interest. Although the MCMC algorithm enjoys higher efficiency, disadvantages include: (1) a number of tunable parameters to calibrate; (2) long “burn-in” times and consequently poor parallelization properties; and (3) correlation between subsequent samples. An ideal sampling algorithm would combine the easy parallelization and uncorrelated nature of rejection sampling with the high efficiency of MCMC.

Of course, it is possible to improve rejection sampling considerably with a better choice of the proposal distribution. An ideal proposal distribution satisfies two criteria: (1) it approximates the target distribution well, and (2) it is easy to sample. Although the spectrally uniform measure satisfies the second requirement, it does not approximate the TKdV Gibbs measure particularly well. On the other hand, Sun and Moore [28], in their efforts to prove theoretical results regarding TKdV wave statistics, constructed a distribution that satisfies both criteria. The goal of the present paper is to employ the distribution of Sun and Moore as a proposal distribution in a highly efficient and parallelizable sampling algorithm. We employ this proposal distribution in a *rejection* sampling algorithm, although the same distribution could just as easily be employed in an SIR method. To demonstrate the desirable numerical features, it suffices to apply the algorithm to the downstream Gibbs ensemble, assuming that its inverse temperature has already been set by the statistical matching condition [28]. We will therefore consider the constant-depth TKdV system throughout this paper. The application of the algorithm to piecewise constant-depth topography is straightforward and will be done in future work.

The outline of the paper is as follows. Section 2 discusses the relevant background information, including the TKdV measure, its Hamiltonian structure, and the canonical–microcanonical Gibbs measure. In Section 3, we introduce the proposal distribution, including a recap of the relevant analysis of Sun and Moore [28]. Section 4 outlines the new sampling algorithm. In Section 5, we present results from numerical tests, including evaluation of the algorithm’s performance and physical interpretation of sampled wave fields. Section 6 offers some concluding remarks.

2 | Mathematical Preliminaries

This section provides the mathematical background for the new sampling algorithm. Below, we introduce the KdV and TKdV systems, as well as the associated Gibbs measure. The KdV system derives from the more primitive Euler equations in the long-wave limit and retains the physical effects of nonlinearity and dispersion [25–27]. Analysis of the KdV system complements studies that directly simulate the Euler equations [10, 29], many of which employ conformal mapping of multiply connected domains [30, 31] to represent obstacles or topographical changes.

2.1 | The KdV Equation

The KdV equation is given by

$$u_t + C_3 uu_\xi + C_2 u_{\xi\xi\xi} = 0 \quad \text{for } \xi \in [-\pi, \pi], \quad (1)$$

where $u(\xi, t)$ is the surface displacement measured in a reference frame traveling with the leading-order wave speed. Boundary conditions are periodic on the normalized domain $\xi \in [-\pi, \pi]$, where ξ is the horizontal location in the traveling frame. All variables u, ξ, t are assumed dimensionless already, and the dimensionless coefficients C_3 and C_2 characterize the strength of nonlinearity and dispersion, respectively [16]. Depth variations correspond to variation of the coefficients C_3 and C_2 with respect to time (since the reference frame is traveling) [11, 16]. In particular, an abrupt depth change results in distinct Gibbs ensembles upstream and downstream of the depth change, where the downstream inverse temperature is determined by the statistical matching condition [11, 16]. To demonstrate the features of the new sampling algorithm, it suffices to consider the downstream Gibbs ensemble only, supposing its inverse temperature has already been set by the statistical matching condition. We therefore consider constant values of C_3 and C_2 throughout this paper.

The KdV equation possesses a Hamiltonian structure given by

$$\mathcal{H}_3[u] = \frac{1}{6} \int_{-\pi}^{\pi} u^3 d\xi, \quad \mathcal{H}_2[u] = \frac{1}{2} \int_{-\pi}^{\pi} \left(\frac{\partial u}{\partial \xi} \right)^2 d\xi, \quad (2)$$

$$\mathcal{H}[u] = C_2 \mathcal{H}_2[u] - C_3 \mathcal{H}_3[u]. \quad (3)$$

Equation (1) can then be written as

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial \xi} \frac{\delta \mathcal{H}}{\delta u}, \quad (4)$$

where ∂_ξ is a symplectic operator. Hence, Equation (1) is a Hamiltonian system and, consequently, the Hamiltonian, Equation (3), is conserved during evolution.

In addition to the Hamiltonian, momentum and energy are conserved under KdV dynamics

$$\mathcal{M}[u] \equiv \int_{-\pi}^{\pi} u d\xi = 0, \quad \mathcal{E}[u] \equiv \frac{1}{2} \int_{-\pi}^{\pi} u^2 d\xi = \mathcal{E}_0. \quad (5)$$

The momentum vanishes, as indicated above, because u is measured as displacement from equilibrium. Meanwhile, the energy has been normalized to the value \mathcal{E}_0 . We remark that setting $\mathcal{E}_0 = \pi/2$ recovers the setup of Sun and Moore [28]. In the present paper, we will keep \mathcal{E}_0 arbitrary for slightly greater flexibility.

2.2 | The TKdV System

We next perform a finite Galerkin truncation of Equation (1) [11, 16, 28]. To this end, consider a spatial Fourier representation of the state variable

$$u(\xi, t) = \sum_{k=-\infty}^{\infty} \hat{u}_k(t) e^{ik\xi} = \sum_{k=1}^{\infty} a_k(t) \cos(k\xi) + b_k(t) \sin(k\xi), \quad (6)$$

$$\hat{u}_k = \frac{1}{2}(a_k - ib_k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} u(\xi, t) e^{-ik\xi} d\xi. \quad (7)$$

For convenience, we have recorded both the real and the complex Fourier representations, $\hat{u}_k \in \mathbb{C}$ and $a_k, b_k \in \mathbb{R}$. Note that $\hat{u}_{-k} = \hat{u}_k^*$ since $u(\xi, t)$ is real-valued and $\hat{u}_0 = 0$ due to momentum vanishing. Next, consider the Galerkin truncation at wavenumber K

$$u_K(\xi, t) = \mathcal{P}_K[u] = \sum_{k=-K}^K \hat{u}_k e^{ik\xi} = \sum_{k=1}^K a_k \cos(k\xi) + b_k \sin(k\xi). \quad (8)$$

Here, the truncation operator \mathcal{P}_K projects an infinite-dimensional function space onto a finite-dimensional one. Note that the formula for the Fourier coefficients Equation (7) still holds, even after truncation. Inserting the projected variable, u_K , into the KdV equation and applying the projection operator, \mathcal{P}_K , again where necessary produces the TKdV equation [11, 16, 32]

$$\frac{\partial u_K}{\partial t} + \frac{1}{2} C_3 \frac{\partial}{\partial \xi} \mathcal{P}_K[(u_K)^2] + C_2 \frac{\partial^3 u_K}{\partial \xi^3} = 0 \quad \text{for } \xi \in [-\pi, \pi]. \quad (9)$$

Equation (9) represents a *finite*-dimensional dynamical system. The quadratic nonlinearity, $\partial_{\xi} \mathcal{P}_K[(u_K)^2]$, mixes the modes during evolution. The presence of the additional projection operator in this term removes the aliased modes of wavenumber larger than K . Typical values of the cutoff wavenumber used in the previous studies are $K = 8$ – 32 [11, 16, 28, 33].

The TKdV equation enjoys nearly the same Hamiltonian structure as KdV, with the only modification being the inclusion of the projection operator

$$\mathcal{H}_K = C_2 \mathcal{H}_2[u_K] - C_3 \mathcal{H}_3[u_K], \quad (10)$$

$$\frac{\partial}{\partial t} u_K = \frac{\partial}{\partial \xi} \mathcal{P}_K \left[\frac{\delta \mathcal{H}_K}{\delta u_K} \right], \quad (11)$$

where now $\partial_{\xi} \mathcal{P}_K$ is the symplectic operator of interest.

The system's microstate can either be described in physical space $u_K(\xi, t)$, in complex spectral space, $(\hat{u}_1, \hat{u}_2, \dots, \hat{u}_K) \in \mathbb{C}^K$, or in real spectral space $(a_1, a_2, \dots, a_K, b_1, b_2, \dots, b_K) \in \mathbb{R}^{2K}$. All are equivalent through Equations (7) and (8). The momentum and energy defined in Equation (5) are also conserved in the truncated system and have the same normalized values $\mathcal{M}[u_K] = 0$ and $\mathcal{E}[u_K] = \mathcal{E}_0$. Parseval's identity implies

$$\mathcal{E}[u_K] = 2\pi \sum_{k=1}^K |\hat{u}_k|^2 = \frac{\pi}{2} \sum_{k=1}^K a_k^2 + b_k^2 = \mathcal{E}_0. \quad (12)$$

2.3 | The Gibbs Ensemble

Following previous studies [11, 16, 28, 32], we employ a *mixed canonical-microcanonical* Gibbs measure of TKdV; that is, a

measure that is canonical in the Hamiltonian and microcanonical in the energy, formally expressed as

$$d\mathcal{G} \propto \exp(-\beta \mathcal{H}_K) \delta(\mathcal{E} - \mathcal{E}_0). \quad (13)$$

Above, β is the inverse temperature. The exponential dependence with respect to the Hamiltonian is the well-known canonical distribution, which, under suitable conditions, maximizes entropy [34]. The Dirac-delta term $\delta(\mathcal{E} - \mathcal{E}_0)$ confines the distribution to the compact set $\mathcal{E} = \mathcal{E}_0$, thereby avoiding the far-field, sign-indefinite divergence of \mathcal{H}_3 and rendering the distribution normalizable [11, 16, 28, 32, 35].

This mixed Gibbs measure can be defined rigorously via integration against test functions on the unit hypersphere $\mathbb{S}^{2K-1} = \{\mathbf{x} \in \mathbb{R}^{2K} : |\mathbf{x}| = 1\}$. Throughout, we let $\phi : \mathbb{S}^{2K-1} \rightarrow \mathbb{R}$ represent an arbitrary, measurable test function. First, consider the standard normal distribution on \mathbb{R}^{2K} ,

$$d\gamma(\mathbf{x}) := \prod_{k=1}^{2K} \frac{1}{\sqrt{2\pi}} e^{-x_k^2/2} dx_k. \quad (14)$$

The *uniform* measure $d\mu_0$ on \mathbb{S}^{2K-1} can be defined by projecting $d\gamma(\mathbf{x})$ onto \mathbb{S}^{2K-1} ; that is, for any measurable $\phi : \mathbb{S}^{2K-1} \rightarrow \mathbb{R}$, let

$$\int_{\mathbb{S}^{2K-1}} \phi(\hat{\mathbf{x}}) d\mu_0(\hat{\mathbf{x}}) = \int_{\mathbb{R}^{2K}} \phi(\hat{\mathbf{x}}) d\gamma(\mathbf{x}), \quad (15)$$

where $\hat{\mathbf{x}} = \mathbf{x}/|\mathbf{x}| \in \mathbb{S}^{2K-1}$ represents a unit-vector projection. Above, we have chosen a free constant to render $d\mu_0$ a probability density function (p.d.f.). That is, $\int_{\mathbb{S}^{2K-1}} d\mu_0 = 1$, as can be seen by taking $\phi \equiv 1$ and recalling that $d\gamma$ is a p.d.f. We remark that $d\mu_0$ is a uniform measure in *spectral* space (not, e.g., a uniform distribution of surface displacements).

Next, for each unit vector, $\hat{\mathbf{x}} = (\hat{x}_1, \hat{x}_2, \dots, \hat{x}_{2K}) \in \mathbb{S}^{2K-1}$, we can identify a corresponding microstate

$$\hat{u}_k = \hat{U}_k(\hat{\mathbf{x}}) = \sqrt{\frac{\mathcal{E}_0}{2\pi}} (\hat{x}_k - i\hat{x}_{K+k}) \quad \text{for } k = 1, 2, \dots, K \quad (16)$$

with the conjugate relation, $\hat{u}_{-k} = \hat{u}_k^*$, defining the negative modes. By definition, Equation (16) satisfies the energy constraint Equation (12).

With this nomenclature established, we can precisely define the Gibbs measure $d\mathcal{G}$ from Equation (13). In particular, for any measurable $\phi : \mathbb{S}^{2K-1} \rightarrow \mathbb{R}$, let

$$\int_{\mathbb{S}^{2K-1}} \phi(\hat{\mathbf{x}}) d\mathcal{G}(\hat{\mathbf{x}}) = Z^{-1} \int_{\mathbb{S}^{2K-1}} \phi(\hat{\mathbf{x}}) \exp(-\beta \mathcal{H}_K[\hat{\mathbf{U}}(\hat{\mathbf{x}})]) d\mu_0(\hat{\mathbf{x}}) \quad (17)$$

$$= Z^{-1} \int_{\mathbb{R}^{2K}} \phi(\hat{\mathbf{x}}) \exp(-\beta \mathcal{H}_K[\hat{\mathbf{U}}(\hat{\mathbf{x}})]) d\gamma(\mathbf{x}), \quad (18)$$

where Z is the partition function (a normalization constant chosen to satisfy the law of total probability) and $\hat{\mathbf{U}} = (\hat{U}_1, \hat{U}_2, \dots, \hat{U}_K) : \mathbb{S}^{2K-1} \rightarrow \mathbb{C}^K$ as given by Equation (16). The Gibbs measure is associated with a probability density supported on the hypersphere, $f : \mathbb{S}^{2K-1} \rightarrow \mathbb{R}$,

$$f(\hat{\mathbf{x}}) = Z^{-1} \exp(-\beta \mathcal{H}_K[\hat{\mathbf{U}}(\hat{\mathbf{x}})]) \quad \text{for } \hat{\mathbf{x}} \in \mathbb{S}^{2K-1}. \quad (19)$$

We remark that if the inverse temperature vanishes, $\beta = 0$, then the Gibbs measure, $d\mathcal{G}$ from Equation (17), reduces to the uniform measure, $d\mu_0$, on \mathbb{S}^{2K-1} . Earlier studies have shown that if $\beta \neq 0$, $d\mathcal{G}$ can differ substantially from the uniform measure and produce anomalous surface-displacement statistics [16, 28]. More specifically, positive inverse temperature, $\beta > 0$, corresponds to the anomalous wave behavior observed in laboratory experiments [7, 16]. Our main goal is to devise an algorithm to sample the Gibbs measure (Equation 17) in this physically relevant and nontrivial regime of $\beta > 0$.

3 | The Proposal Distribution

In this section, we introduce the proposal distribution that will enable efficient rejection sampling of the Gibbs measure (Equation 17). This distribution was first constructed by Sun and Moore [28] who used it to establish rigorous results on the conditions that lead to Gaussian versus non-Gaussian surface-displacement statistics. The construction is based on analysis of the *linear* TKdV system, that is, $C_3 = 0$, even though the resulting distribution has relevance to the fully nonlinear system. Here, we briefly recap the derivation of the distribution, with emphasis placed on the intuition behind its construction since the rigor has already been established by Sun and Moore [28].

3.1 | Linear TKdV and Change of Measure

Setting $C_3 = 0$ eliminates the only nonlinearity in TKdV, Equation (9), rendering it a *linear* dynamical system. In this case, the Hamiltonian, Equation (10), reduces to its quadratic component $H_K = C_2 H_2$, which, by Parseval's identity, can be written as

$$H_2 = 2\pi \sum_{k=1}^K k^2 |\hat{u}_k|^2 = \mathcal{E}_0 \sum_{k=1}^K k^2 (\hat{x}_k^2 + \hat{x}_{K+k}^2). \quad (20)$$

To simplify notation, consider a normalized inverse temperature

$$\beta' = \mathcal{E}_0 C_2 K^2 \beta. \quad (21)$$

Using Equations (16), (18), (20), and (21), we can rewrite the main part of the linear TKdV Gibbs measure in terms of the integration variable $\mathbf{x} \in \mathbb{R}^{2K}$,

$$\exp(-\beta C_2 H_2) d\gamma(\mathbf{x}) = \exp\left(-\frac{\beta'}{K^2 |\mathbf{x}|^2} \sum_{k=1}^K k^2 (x_k^2 + x_{K+k}^2)\right) d\gamma(\mathbf{x}). \quad (22)$$

The analysis of Sun and Moore [28] builds upon a few observations of this formula. First, for a random variable $\mathbf{X} \in \mathbb{R}^{2K}$ selected from the standard Gaussian distribution $d\gamma$, the law of large numbers (LLN) implies $|\mathbf{X}|^2 / (2K) \rightarrow 1$ as $K \rightarrow \infty$, where the convergence is in probability. This convergence might suggest the crude substitution $|\mathbf{x}|^2 \approx 2K$ in Equation (22), which can be made even more general by introducing an extra degree of freedom $|\mathbf{x}|^2 \approx 2K/\alpha$, where $\alpha > 0$. As it turns out, by selecting α judiciously, this second substitution leads to an approximation of the Gibbs measure that can be justified rigorously [28].

A brief overview of the derivation is as follows. First, notice that the formula

$$\frac{1}{|\mathbf{x}|^2} = \frac{\alpha}{2K} + \frac{1}{2K} \left(\frac{2K}{|\mathbf{x}|^2} - \alpha \right) \quad (23)$$

holds for any α . To offer some intuition, our desired substitution $|\mathbf{x}|^2 \approx 2K/\alpha$ suggests the main part of $1/|\mathbf{x}|^2$ will be captured by the first term, $\alpha/(2K)$, and that the second term above will be small.

Next, consider an anisotropic Gaussian measure γ_Σ with covariance matrix

$$\Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_K^2, \sigma_1^2, \sigma_2^2, \dots, \sigma_K^2), \quad (24)$$

$$\sigma_k^2 = \frac{1}{1 + \alpha \beta' k^2 / K^3} \quad \text{for } k = 1, 2, \dots, K. \quad (25)$$

More explicitly,

$$d\gamma_\Sigma(\mathbf{x}) = \prod_{k=1}^K \frac{1}{2\pi\sigma_k^2} \exp\left(-\frac{x_k^2 + x_{K+k}^2}{2\sigma_k^2}\right) dx_k dx_{K+k} \quad (26)$$

$$= C \exp\left(-\frac{\alpha\beta'}{2K^3} \sum_{k=1}^K k^2 (x_k^2 + x_{K+k}^2)\right) d\gamma(\mathbf{x}), \quad (27)$$

where C is a constant. In particular, this anisotropic measure $d\gamma_\Sigma$ is chosen to equal the right-hand side of Equation (22) when the crude substitution $|\mathbf{x}|^2 \approx 2K/\alpha$ is made. If instead the exact Equation (23) is used, then Equation (22) can be rewritten *with no approximation made as*

$$\begin{aligned} & \exp(-\beta C_2 H_2) d\gamma(\mathbf{x}) \\ &= C^{-1} \exp\left(-\frac{\beta'}{2K^3} \left(\frac{2K}{|\mathbf{x}|^2} - \alpha \right) \sum_{k=1}^K k^2 (x_k^2 + x_{K+k}^2) \right) d\gamma_\Sigma(\mathbf{x}). \end{aligned} \quad (28)$$

If $(2K/|\mathbf{x}|^2 - \alpha) \rightarrow 0$, as expected from Equation (23), then the right-hand side of Equation (28) converges to a constant multiple of $d\gamma_\Sigma$. This is the main intuition for how the distribution $d\gamma_\Sigma$ approximates the Gibbs measures in the case of linear TKdV.

Finally, Sun and Moore [28] showed that if $\alpha = \alpha^*$ is selected as the root of the function

$$F(\alpha) := 1 - \frac{\alpha}{K} \sum_{k=1}^K \frac{1}{1 + \alpha \beta' k^2 / K^3} \quad (29)$$

then for any continuous $\phi : \mathbb{S}^{2K-1} \rightarrow \mathbb{R}$, the following convergence holds for linear TKdV:

$$\int_{\mathbb{S}^{2K-1}} \phi(\hat{\mathbf{x}}) d\mathcal{G}(\hat{\mathbf{x}}) - \int_{\mathbb{R}^{2K}} \phi(\hat{\mathbf{x}}) d\gamma_\Sigma(\mathbf{x}) \rightarrow 0 \quad \text{as } K \rightarrow \infty \quad (30)$$

for any fixed $\beta' \geq 0$. In this precise sense, the anisotropic Gaussian $d\gamma_\Sigma$, with α^* chosen as the root of Equation (29), approximates the linear-TKdV Gibbs measure as the cutoff wavenumber grows large.

We remark that α^* is not necessarily equal to one, as might have been guessed from the earlier intuitive argument involving

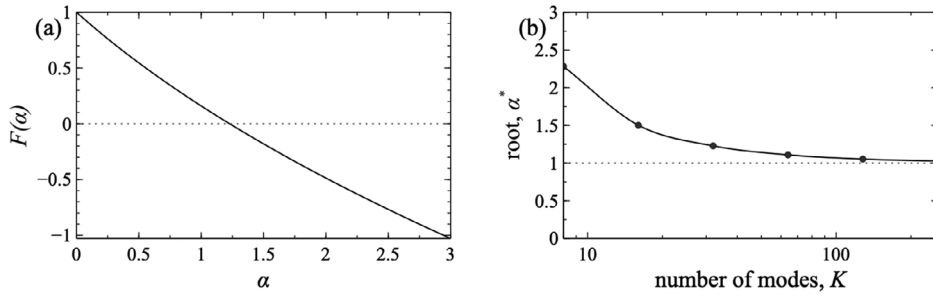


FIGURE 1 | The value α^* used in the approximating measure Equation (27) is selected as the root of Equation (29). (a) $F(\alpha)$ from Equation (29) with $K = 32$ and $\beta' = 20$. (b) Root α^* for increasing values of K with $\beta' = 20$.

the LLN (i.e., if $\mathbf{X} \in \mathbb{R}^{2K}$ is selected from the standard Gaussian distribution $d\gamma$, then $|\mathbf{X}|^2/(2K) \rightarrow 1$ as $K \rightarrow \infty$ in probability). The reason is that, with change of measure Equation (28), the random variable $\mathbf{X} \in \mathbb{R}^{2K}$ should now be regarded as coming from the anisotropic Gaussian $d\gamma_\Sigma$. The value $\alpha = \alpha^*$ is selected self-consistently, so that when the random variable \mathbf{X} is sampled from $d\gamma_\Sigma$ (which depends on α), the convergence $2K/|\mathbf{X}|^2 \rightarrow \alpha^*$ holds as $K \rightarrow \infty$.

Figure 1 provides a visualization of how α^* is selected and how its value depends on the cutoff wavenumber K . Figure 1a shows the function $F(\alpha)$ defined in Equation (29) for values $K = 32$ and $\beta' = 20$. The root occurs at approximately $\alpha^* = 1.2$. Figure 1b shows the value of α^* , as computed with a numerical root finder, for the values $K = 8, 16, 32, 64$, and 128 . Notice that $\alpha^* \rightarrow 1$ as $K \rightarrow \infty$, which can be confirmed by directly taking the limit of Equation (29).

3.2 | Evaluating the Proposal Distribution

Recall that the Gibbs measure (Equation 17) has an associated probability density, $f : \mathbb{S}^{2K-1} \rightarrow \mathbb{R}$, supported on the hypersphere. This f is the target distribution we seek to sample, and Equation (19) shows how to compute $f(\hat{\mathbf{x}})$ for any $\hat{\mathbf{x}} \in \mathbb{S}^{2K-1}$. Likewise, the approximating measure Equation (27) has an associated probability density $g : \mathbb{S}^{2K-1} \rightarrow \mathbb{R}$ supported on the hypersphere and defined by

$$\int_{\mathbb{S}^{2K-1}} \phi(\hat{\mathbf{x}}) g(\hat{\mathbf{x}}) d\mu_0(\hat{\mathbf{x}}) = \int_{\mathbb{R}^{2K}} \phi(\mathbf{x}) d\gamma_\Sigma(\mathbf{x}) \quad (31)$$

for any measurable $\phi : \mathbb{S}^{2K-1} \rightarrow \mathbb{R}$. The density g will be the proposal distribution of the rejection algorithm. Since the underlying measure $d\gamma_\Sigma$ from Equation (26) is Gaussian, it is straightforward to sample g . One first draws a sample \mathbf{X} from $d\gamma_\Sigma$ and then projects onto \mathbb{S}^{2K-1} via $\hat{\mathbf{X}} = \mathbf{X}/|\mathbf{X}|$. However, to use g as a proposal distribution, it is necessary to compute the ratio $f(\hat{\mathbf{x}})/g(\hat{\mathbf{x}})$ to determine the acceptance probability of a given $\hat{\mathbf{x}} \in \mathbb{S}^{2K-1}$. While Equation (19) shows how to compute $f(\hat{\mathbf{x}})$, we have not yet shown how to compute $g(\hat{\mathbf{x}})$.

In a rejection algorithm, it suffices to compute the ratio $f(\hat{\mathbf{x}})/g(\hat{\mathbf{x}})$ up to a multiplicative constant. Therefore, the value Z in Equation (19) and other constants that arise will not be needed. For this reason, we will use $Const$ to represent various constants that

arise in this section. Even though these constants are represented by the same symbol, they do not necessarily have the same value.

As seen in Equation (31), evaluating $g(\hat{\mathbf{x}})$ for a particular value of $\hat{\mathbf{x}} \in \mathbb{S}^{2K-1}$ requires integrating the measure $d\gamma_\Sigma(\mathbf{x})$ over all values of $\mathbf{x} \in \mathbb{R}^{2K}$ for which $\mathbf{x}/|\mathbf{x}| = \hat{\mathbf{x}}$. These values comprise a ray emanating from the origin and passing through the point $\hat{\mathbf{x}}$, given explicitly by $\mathbf{x} = r\hat{\mathbf{x}}$ for $r \in (0, \infty)$. Evaluating Equation (26) along this ray gives

$$d\gamma_\Sigma \Big|_{\mathbf{x}=r\hat{\mathbf{x}}} \propto \exp(-a(\hat{\mathbf{x}})r^2), \quad (32)$$

where the parameter $a(\hat{\mathbf{x}})$ depends on the direction vector $\hat{\mathbf{x}} \in \mathbb{S}^{2K-1}$ as given by

$$a(\hat{\mathbf{x}}) = \sum_{k=1}^K \frac{\hat{x}_k^2 + \hat{x}_{K+k}^2}{2\sigma_k^2}. \quad (33)$$

Inserting into Equation (31) gives

$$g(\hat{\mathbf{x}}) = Const \cdot \int_0^\infty \exp(-a(\hat{\mathbf{x}})r^2) r^{2K-1} dr. \quad (34)$$

To compute the integral above, we first perform a change of variables with $u = r^2$,

$$\int_0^\infty \exp(-ar^2) r^{2K-1} dr = \frac{1}{2} \int_0^\infty \exp(-au) u^{K-1} du. \quad (35)$$

We then use a known Laplace transform

$$\mathcal{L}[t^{K-1}] = \int_0^\infty \exp(-st) t^{K-1} du = (K-1)! s^{-K}. \quad (36)$$

Using Equations (34)–(36), the proposal distribution $g(\hat{\mathbf{x}})$ is then given by

$$g(\hat{\mathbf{x}}) = Const \cdot a(\hat{\mathbf{x}})^{-K}. \quad (37)$$

Inserting Equation (33) and the standard deviations given by Equation (25) then gives

$$g(\hat{\mathbf{x}}) = Const \cdot \left(1 + \frac{\alpha^* \beta'}{K^3} \sum_{k=1}^K k^2 (\hat{x}_k^2 + \hat{x}_{K+k}^2) \right)^{-K}. \quad (38)$$

This is the formula we will use to evaluate the proposal distribution g on the unit hypersphere.

We remark that a nice check on the above formula is possible. Taking the limit as $K \rightarrow \infty$ of Equation (38) and using the limit definition of the exponential gives

$$g(\hat{\mathbf{x}}) \rightarrow \text{Const} \cdot \exp\left(-\frac{\alpha^* \beta'}{K^2} \sum_{k=1}^K k^2 (\hat{x}_k^2 + \hat{x}_{K+k}^2)\right) \quad \text{as } K \rightarrow \infty. \quad (39)$$

Inspection of Equation (29) shows that $\alpha^* \rightarrow 1$ in this limit, in which case Equation (39) matches Equation (22).

4 | The Rejection Algorithm

We now describe the rejection algorithm that will enable efficient sampling of the Gibbs measure (Equation 17). Associated with this Gibbs measure is the probability density $f : \mathbb{S}^{2K-1} \rightarrow \mathbb{R}$ supported on the hypersphere. The density f is straightforward to evaluate through Equation (19), however, it is not straightforward to sample f directly. We therefore use density $g : \mathbb{S}^{2K-1} \rightarrow \mathbb{R}$, defined by Equation (31), as a proposal distribution. Since the measure $d\gamma_{\mathbb{S}}$ defining g is Gaussian, it is straightforward to sample g ; one first draws a sample \mathbf{X} from $d\gamma_{\mathbb{S}}$ and then projects onto \mathbb{S}^{2K-1} via $\hat{\mathbf{X}} = \mathbf{X}/|\mathbf{X}|$. Furthermore, Equation (30) shows that g converges to f in a particular limit. It is therefore reasonable to expect that g lies fairly close to f in general, at least more so than the naive, uniform distribution. For these reasons, we expect g to satisfy the two criteria for a good proposal distribution: it approximates f and is easy to sample. The numerical tests in Section 5 confirm that g yields acceptance rates much higher than the uniform distribution, even well outside the strict validity of Equation (30).

We will employ a standard rejection algorithm; that is, we first draw a sample from the proposal distribution $g(\hat{\mathbf{x}})$ and then accept the sample with probability proportional to $f(\hat{\mathbf{x}})/g(\hat{\mathbf{x}})$. Repeating this process yields a collection of samples from the target distribution $f(\hat{\mathbf{x}})$ [36]. Alternatively, one could employ the SIR algorithm with g as the proposal distribution [37], although we do not implement it here.

4.1 | Computing the Acceptance Probability and Rejection Constant

In the proposed rejection algorithm, the acceptance probability is proportional to the ratio f/g . For simplicity, we take Const in Equation (38) to equal Z^{-1} so that constants cancel to give

$$\frac{f(\hat{\mathbf{x}})}{g(\hat{\mathbf{x}})} = \exp(-\beta H_K[\hat{\mathbf{U}}(\hat{\mathbf{x}})]) \left(1 + \frac{\alpha^* \beta'}{K^3} \sum_{k=1}^K k^2 (\hat{x}_k^2 + \hat{x}_{K+k}^2)\right)^K. \quad (40)$$

In order to compute H_K above it is necessary to compute H_2 and H_3 . H_2 can be efficiently computed with Equation (20). For H_3 , Abramov et al. [35] derived the triple-summation formula

$$H_3 = \frac{\pi}{3} \sum_{\substack{k_1+k_2+k_3=0 \\ |k_1|, |k_2|, |k_3| \leq K}} \hat{u}_{k_1} \hat{u}_{k_2} \hat{u}_{k_3}. \quad (41)$$

This formula can be rearranged as a double summation

$$H_3 = 2\pi \sum_{n=1}^K \text{Re} \left(\hat{u}_n^* \sum_{k=1}^{n-1} \hat{u}_k \hat{u}_{n-k} \right). \quad (42)$$

Therefore, H_3 can be computed with $O(K^2)$ operations. We remark that, alternatively, H_3 could be computed from the definition in Equation (2) using the FFT with $O(K \log K)$ operations. However, for moderate values of K in the range $8 \leq K \leq 64$, Equation (42) is typically more efficient and hence the form that we will use to compute H_3 .

In practice, we reduce the number of control parameters in Equation (40) by recasting in terms of β' and the ratio C_3/C_2 . That is, we use Equations (10) and (21) to rewrite

$$\beta H_K = \frac{\beta'}{\mathcal{E}_0 K^2} \left(H_2 - \frac{C_3}{C_2} H_3 \right). \quad (43)$$

Therefore, the only parameters that must be specified are K, \mathcal{E}_0, β' , and C_3/C_2 ; it is not necessary to specify β, C_2 , and C_3 individually. In Appendix A, we provide exact solutions for H_2 and H_3 in the case of $K = 2$. These solutions are used to validate our numerical computation of H_2 and H_3 .

We now have everything needed to compute the ratio f/g through Equation (40). However, it remains to compute the rejection constant. That is, we seek a constant M to ensure that

$$\frac{f(\hat{\mathbf{x}})}{Mg(\hat{\mathbf{x}})} \leq 1 \quad \text{for all } \hat{\mathbf{x}} \in \mathbb{S}^{2K-1}, \quad (44)$$

so that the acceptance probability is no greater than one. The optimal such constant is given by

$$M = \sup_{\hat{\mathbf{x}} \in \mathbb{S}^{2K-1}} \frac{f(\hat{\mathbf{x}})}{g(\hat{\mathbf{x}})}. \quad (45)$$

Since the unit hypersphere \mathbb{S}^{2K-1} is a compact set, the supremum is equal to the maximum value. We can therefore solve a numerical optimization problem to determine M . Since the dimension of \mathbb{S}^{2K-1} is not necessarily small (e.g., the typical range $8 \leq K \leq 32$ yields $15 \leq \dim(\mathbb{S}^{2K-1}) \leq 63$), it is important to have a good initial guess for the numerical optimization.

To obtain a good initial guess, we consider the derivation of the proposal distribution g and the physical meaning of the ratio f/g . In particular, g was chosen specifically to account for the presence of H_2 in the Hamiltonian, whereas f includes contributions from both H_2 and H_3 . In particular, since f is proportional to $\exp(-\beta C_2 H_2 + \beta C_3 H_3)$ and g is approximately proportional to $\exp(-\beta C_2 H_2)$, the H_2 contributions approximately cancel in the ratio, giving

$$\frac{f(\hat{\mathbf{x}})}{g(\hat{\mathbf{x}})} \approx \text{Const} \cdot \exp(\beta C_3 H_3). \quad (46)$$

Thus maximizing H_3 should provide a good starting guess for maximizing the ratio f/g .

Recall from Equation (2), that H_3 measures the skewness, $\int_{-\pi}^{\pi} u^3 d\xi$, of the sampled surface displacement. We seek to

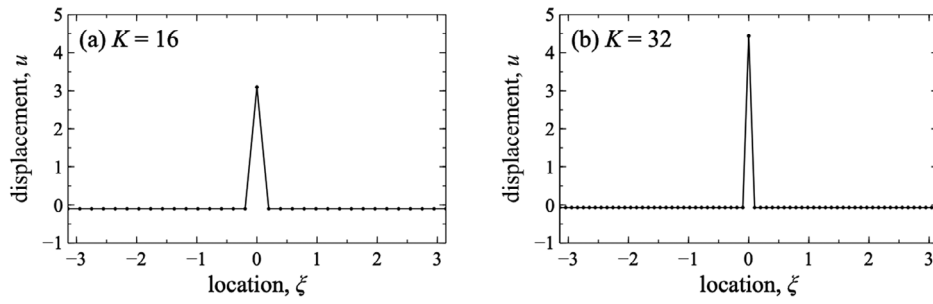


FIGURE 2 | Visualization of the Dirichlet kernel for $K = 16$ and 32 . This function is used as the initial guess to numerically optimize the ratio f/g on the unit hypersphere.

maximize this quantity within the space of functions having Fourier modes up to wavenumber K and satisfying the constraints of Equation (5) of zero momentum and fixed energy. In physical space, the maximum is achieved by a Dirac-delta approximation, that is, a spike-like function with a single, large positive value. This function must be shifted down slightly to satisfy the zero-mean condition, $\int_{-\pi}^{\pi} u d\xi = 0$. The best finite-mode Dirac-delta approximation is known as the Dirichlet kernel [38], characterized by a constant spectrum. In spectral space, the zero-mean Dirichlet kernel with peak centered at $\xi = 0$ is given by

$$\hat{u}_k = \sqrt{\frac{\mathcal{E}_0}{2\pi K}} \quad \text{for } k = \pm 1, 2, \dots, K. \quad (47)$$

Figure 2 illustrates this Dirichlet kernel in the cases $K = 16$ and 32 . Both show a large positive value at $\xi = 0$ and a small negative value for $\xi \neq 0$, where these two values are chosen to satisfy zero mean and unit energy.

Using Equation (47) as the initial guess, we numerically optimize the ratio f/g with the Nelder–Mead algorithm [39] in order to determine the rejection constant M . We find that the numerical optimization typically converges in only a few iterations, confirming that Equation (47) provides an effective starting guess.

4.2 | The Rejection Algorithm

All of the pieces are now in place to implement the rejection algorithm. To summarize, the algorithm is as follows:

1. Precompute the rejection constant M by numerically maximizing Equation (45) using Equation (47) as the initial guess.
2. Draw a sample $\mathbf{X} \in \mathbb{R}^{2K}$ from the anisotropic Gaussian distribution $d\gamma_2$ given by Equations (25) and (26) with $\alpha = \alpha^*$ selected as the root of Equation (29).
3. Project the sample \mathbf{X} onto \mathbb{S}^{2K-1} via $\hat{\mathbf{X}} = \mathbf{X}/|\mathbf{X}|$. By doing so, $\hat{\mathbf{X}}$ represents a random sample drawn from density g .
4. For the sample $\hat{\mathbf{X}}$, compute the acceptance probability $f(\hat{\mathbf{X}})/(Mg(\hat{\mathbf{X}}))$ using Equation (40), with Equations (20) and (42) used to compute H_2 and H_3 . Decide to accept the sample or not based on the computed probability. If accepted, $\hat{\mathbf{X}}$ represents a random sample drawn from density f . Through Equation (16), an accepted $\hat{\mathbf{X}}$ produces a Fourier-coefficient set, \hat{u}_k for $k = 1, \dots, K$, of an accepted wave field.

5. Go to step 2 and repeat until a desired number of samples have been accepted.

We note that it is trivial to parallelize the above algorithm. In particular, step 1 should be precomputed and then steps 2–5 can be performed in parallel with no communication required between computational nodes. In the following section, we report numerical results on both serial and parallelized versions of this algorithm.

5 | Results

We now present numerical results of the novel rejection algorithm. To provide physical context, we first visualize some wave fields and statistical data generated by the algorithm in different parameter regimes of TKdV. We then evaluate the performance of the algorithm, including: (i) a comparison against a naive version that uses the spectrally uniform measure as a proposal distribution, and (ii) results on the parallel speedup of the algorithm. We then examine characteristics of individual wave fields sampled by the algorithm, including extreme events.

5.1 | Physical Wave Fields Generated by the Algorithm

The main purpose of the new rejection algorithm is to efficiently generate independent, random samples of wave fields arising from the TKdV system without directly simulating the dynamics. To orient the reader to the physics of interest, we begin by showing a few example wave fields generated by the algorithm in different parameter regimes. In all of these examples, we fix the maximum wavenumber $K = 16$ and the energy $\mathcal{E}_0 = 1$, while varying the normalized inverse temperature β' and the nonlinearity-to-dispersion ratio C_3/C_2 . Recall that due to Equation (43), these are all of the control parameters to be specified (i.e., it is not necessary to specify β , C_2 , and C_3 individually). One feature we analyze below is the distribution of surface displacements that results from the Gibbs measure. We remind the reader that the Gibbs measure exists in spectral space, that is, it is a distribution of the wave field's Fourier coefficients, whereas the surface-displacement distribution exists in physical space. Gaussianity (or non-Gaussianity) of surface-displacement statistics does not imply Gaussianity (or non-Gaussianity) of the underlying Gibbs measure. For example, the spectrally uniform Gibbs measure $d\mu_0$,

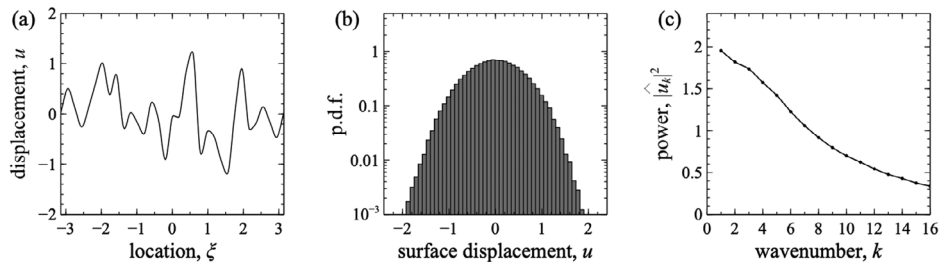


FIGURE 3 | Visualizing wave field statistics in the case of linear TKdV ($C_3/C_2 = 0$) with $\beta' = 40$. (a) A representative wave field sampled by the rejection algorithm. (b) Histogram of surface displacements over 5000 samples. The distribution is symmetric and nearly Gaussian. (c) Power spectrum averaged over all 5000 samples. The spectrum decays gradually with wavenumber.

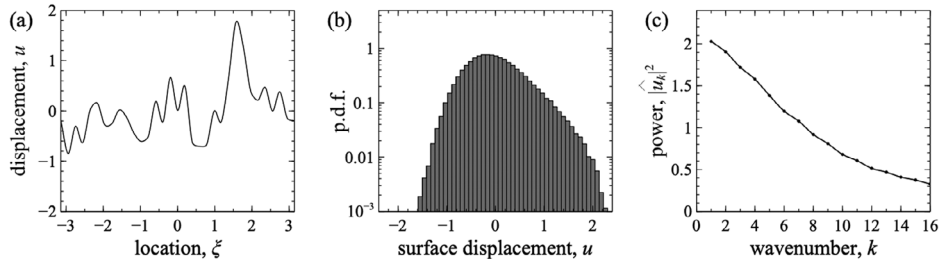


FIGURE 4 | The effects of strong nonlinearity, $C_3/C_2 = 120$, with $\beta' = 40$ as before. (a) The representative wave field features a prominent peak occurring near $\xi = 1.5$. (b) The histogram exhibits positive skewness, indicating that large, positive displacements are consistently favored in the ensemble of 5000 accepted samples. (c) The spectrum decays gradually, similar to the previous case.

that arises in the case $\beta' = 0$, produces Gaussian statistics of the surface displacement and not a uniform distribution of surface displacements [28].

To begin, Figure 3 displays wave information for the case of *linear* TKdV, that is, $C_3/C_2 = 0$, with a moderate inverse temperature $\beta' = 40$. We use the rejection algorithm to sample the Gibbs measure until 5000 samples are accepted. Figure 3a shows a representative wave field, obtained by performing a Fourier transform of a spectrum sampled from the Gibbs measure. The wave exhibits positive and negative surface displacements of roughly equal magnitudes. This plot represents an ordinary wave field with no extreme values. Figure 3b combines all 5000 samples into a histogram of surface displacements. Given the logarithmic vertical axis, the roughly parabolic shape implies a nearly normal distribution. That is, this set of parameters produces *Gaussian statistics of the surface displacement*, as is typically observed for linear TKdV [11, 16, 28]. Note that the Gibbs distribution itself, which exists in spectral space, is not Gaussian. Figure 3c shows the power spectrum averaged over all 5000 samples. The gradual decay of the spectrum is due to H_2 , which penalizes higher wavenumbers as seen through Equation (20).

Figure 4 shows the case of stronger nonlinearity, $C_3/C_2 = 120$, with the other parameters held fixed. Figure 4a shows the representative wave field, distinguished in this case by a prominent peak near $\xi = 1.5$. This wave field is not simply an outlier. Indeed, the histogram of 5000 accepted samples in Figure 4b exhibits a strong positive skewness (skewness = 0.59), indicating that large, positive values of u are more likely in the ensemble. The reason for this behavior is that the stronger nonlinearity (i.e., larger C_3/C_2) increases the relative contribution of $H_3 \propto \int_{-\pi}^{\pi} u^3 d\xi$, which favors positive skewness. Figure 4c shows that the spectrum decays

gradually, as before. To leading order, the spectrum is unaffected by nonlinearity [28].

These two cases illustrate some general principles underlying TKdV wave statistics. In particular, for $\beta' > 0$, the H_2 component penalizes high frequencies in the Gibbs ensemble, creating a gradually decaying power spectrum. If nonlinearity is absent, the surface-displacement distribution is symmetric [28], as seen in Figure 3b. If nonlinearity is present, the H_3 component favors positively skewed wave fields, leading to skewness in the ensemble histogram (Figure 4b) and increasing the likelihood of large, positive displacements. In particular, nonlinearity increases the likelihood of extreme events, as illustrated by the large peak seen in Figure 4a. In the more trivial case of $\beta' = 0$ (not shown), the Gibbs measure is independent of the Hamiltonian, which results in nearly Gaussian surface-displacement statistics [28].

5.2 | Performance of the Algorithm

We now evaluate the performance of the rejection algorithm equipped with the new proposal distribution g , defined in Equation (31). In particular, the average acceptance rate determines the speed at which random wave fields can be generated. Therefore, comparing the average acceptance rate achieved by g versus the uniform measure, Equation (15), indicates the degree of performance improvement. To simplify nomenclature, we refer to the use of g as the “improved rejection algorithm” and the use of the uniform measure as the “naive” version.

Table 1 summarizes results from a series of tests with cutoff wavenumber $K = 16$, energy $\mathcal{E}_0 = 1$, and three inverse temperatures, $\beta' = 20, 40$, and 60 . For each β' , we increment the

TABLE 1 | Performance of the improved rejection algorithm for $K = 16$. Columns 1–2: the parameter values β' and C_3/C_2 chosen for the test; $\mathcal{E}_0 = 1$ in all cases. Column 3: the skewness of the resulting ensemble, which serves as a proxy for anomalous wave behavior. Column 4: the average acceptance rate of the improved algorithm. The acceptance rate is extremely high for linear TKdV, and it decreases as nonlinearity, C_3/C_2 , increases. Column 5: the factor by which the novel proposal distribution improves the acceptance rate over the uniform distribution. The test shows an improvement factor of 1–6 orders of magnitude.

Performance test, $K = 16$				
β'	C_3/C_2	Skewness	Acceptance rate	Improvement
20	0	0	0.95	430
	60	0.11	2.4×10^{-2}	60
	120	0.23	6.9×10^{-4}	25
	180	0.37	2.1×10^{-5}	16
	240	0.53	7.7×10^{-7}	12
	300	0.73	3.2×10^{-8}	11
40	0	0	0.88	49000
	30	0.13	2.3×10^{-2}	3400
	60	0.25	6.7×10^{-4}	690
	90	0.40	2.2×10^{-5}	280
	120	0.59	8.0×10^{-7}	130
	150	0.81	3.6×10^{-8}	74
60	0	0	0.80	1600000
	20	0.14	2.3×10^{-2}	120000
	40	0.27	7.0×10^{-4}	23000
	60	0.42	2.4×10^{-5}	4300
	80	0.66	9.2×10^{-7}	1200
	100	0.85	4.6×10^{-8}	590

nonlinearity parameter, C_3/C_2 , to obtain the same increments of $\beta'C_3/C_2$, since this combination is the prefactor of \mathcal{H}_3 in the Hamiltonian. As seen in Equation (2), \mathcal{H}_3 controls the skewness of the wave field, and so equal increments of $\beta'C_3/C_2$ results in nearly equal increments of ensemble skewness. In this way, the numerical test probes a similar physical regime for each β' . The second column of Table 1 shows the ensemble skewness measured in the numerical tests and confirms that, for each β' , the skewness increments by roughly the same amount as nonlinearity increases.

Column 3 of Table 1 shows acceptance rates achieved by the improved rejection algorithm, while Column 4 shows how many times greater this acceptance rate is than that achieved by the naive version. In all cases, the improvement factor is much greater than one, and in most cases by several orders of magnitude. These results indicate that g vastly outperforms the uniform measure as a proposal distribution, increasing the number of accepted samples by 1–6 orders of magnitude.

More specifically, the improved rejection algorithm produces extremely high acceptance rates when nonlinearity is small. For $C_3/C_2 = 0$, the acceptance rates are all above 80%, indicating that

TABLE 2 | Performance of the improved rejection algorithm as compared to the naive version for $K = 32$. As before, the improved proposal distribution increases the number of accepted samples by several orders of magnitude. To measure the improvement ratio, we require at least 2000 samples from the improved algorithm and at least 100 from the naive version. The asterisk indicates cases in which the naive version was unable to produce 100 samples due to an extremely low acceptance rate.

Performance test, $K = 32$				
β'	C_3/C_2	Skewness	Acceptance rate	Improvement
20	0	0	0.99	360
	150	0.04	3.4×10^{-2}	13
	300	0.08	1.2×10^{-3}	41
	450	0.11	4.2×10^{-5}	27
	600	0.15	1.5×10^{-6}	20
	750	0.19	5.5×10^{-8}	14
40	900	0.25	2.1×10^{-9}	14
	0	0	0.96	66000
	75	0.05	3.3×10^{-2}	1600
	150	0.07	1.1×10^{-3}	79
	225	0.12	4.1×10^{-5}	2
	300	0.16	1.4×10^{-6}	700
60	375	0.22	5.5×10^{-8}	410
	450	0.26	2.1×10^{-9}	330*
	0	0	0.92	2500000
	50	0.04	3.1×10^{-2}	71000
	100	0.08	1.1×10^{-3}	1900
	150	0.13	3.9×10^{-5}	390
	200	0.17	1.4×10^{-6}	13
	250	0.23	5.5×10^{-8}	2.3
	300	0.26	2.0×10^{-9}	5700*

the proposal distribution g approximates the true Gibbs density, Equation (19), extremely well. In these cases, the performance gap between the improved and naive proposal distributions is the largest. In the most extreme case of $\beta' = 60$, the improved algorithm produces 1.6 million times more accepted samples than does the naive version.

As nonlinearity, C_3/C_2 , increases, the acceptance rate of the improved algorithm decreases, but it still outperforms the naive version by a wide margin. The most physically interesting cases are those of strong skewness (roughly skewness > 0.4) indicative of highly anomalous statistics and an increased likelihood of extreme events. In this regime, the improved algorithm achieves acceptance rates in the range 10^{-5} to 10^{-8} . Interestingly, the improvement factor over the naive version increases with β' . That is, in the regime of high skewness, the novel proposal distribution improves performance by an order of magnitude for $\beta' = 20$ but three orders of magnitude for $\beta' = 60$.

Table 2 summarizes results of a similar test with larger cutoff wavenumber $K = 32$. As before, we increment C_3/C_2 so as

to obtain the same increments in $\beta' C_3/C_2$ and thus similar increments in skewness. When nonlinearity is absent ($C_3/C_2 = 0$) the improved algorithm still enjoys extremely high acceptance rates ($>90\%$), indicating that the proposal distribution continues to approximate the Gibbs measure extremely well. The improvement over the uniform measure is extremely large, up to 2.5 million in the case $\beta' = 60$. As nonlinearity increases, the ensemble skewness increases and the acceptance rate decreases. The improved algorithm still enjoys much better performance than the naive version in this regime, increasing the number of accepted samples by 1–5 orders of magnitude.

A few surprises emerge in Table 2. First, in the case $\beta' = 40$, the improvement factor depends nonmonotonically on C_3/C_2 ; it initially decreases with increasing C_3/C_2 , reaches a minimum at $C_3/C_2 = 225$, and then increases. We have verified this non-monotone behavior is repeatable by rerunning the tests with different random seeds, but we do not have a simple explanation for it. Second, comparing cases of large C_3/C_2 across the different values of β' , it seems that the improved algorithm enjoys a sweet spot at $\beta' = 40$, where its advantage over the naive version is maximal. For the case $\beta' C_3/C_2 = 15000$ (second to largest value of C_3/C_2 in each group), the improvement factor is 410 for $\beta' = 40$, but only 14 and 2.3 for $\beta' = 20$ and 60, respectively.

We note that to measure the improvement ratio, we required at least 2000 samples to be drawn by the improved algorithm and at least 100 by the naive algorithm. In the two cases $(\beta', C_3/C_2) = (40, 450)$ and $(60, 300)$, the acceptance rate of the naive version was so low that the test could not be completed over the course of weeks. We therefore estimated the naive acceptance rate and the corresponding improvement ratio from the low number of samples that were drawn. These values are indicated with an asterisk in Table 2.

5.3 | Parallel Speedup of the Algorithm

A significant advantage of a rejection algorithm is the favorable parallelization properties. When equipped with a good proposal distribution and run in parallel, a simple rejection algorithm can outperform more sophisticated methods like MCMC. To demonstrate the favorable parallelization properties, we run the improved rejection algorithm in parallel with up to 16 processors. Figure 5 shows the speedup achieved versus the number of processors. As seen in the figure, the algorithm achieves nearly optimal speedup for up to eight processors. Beyond eight processors, the speedup diminishes as would be expected for any algorithm due to the required communication between processors and memory. The reduction in speedup, however, is small. For the largest case of 16 processors, the speedup is still 80% of the optimal value.

5.4 | Sampling Individual Wave Fields

A major advantage of a rejection algorithm over importance sampling is the ability to draw individual samples from a distribution, rather than merely evaluating statistical properties of the distribution (e.g., computing means or generating histograms). To demonstrate this capability, we now examine the characteristics

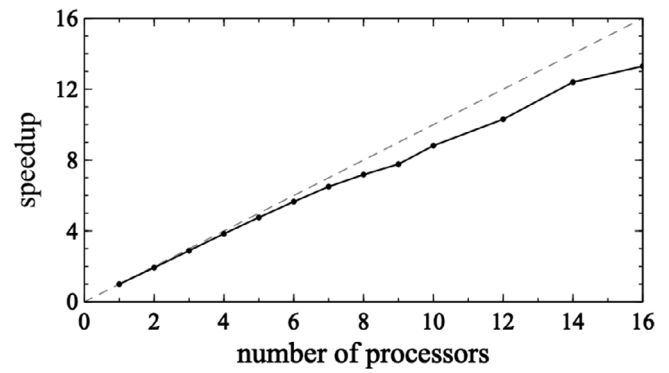


FIGURE 5 | The parallel speedup of the rejection algorithm versus the number of processors. Test conducted on an Apple M2 Ultra chip with 16 performance cores. The cutoff wavenumber is $K = 128$ and the number of nominal samples per thread is 10^4 .

of individual samples drawn from the Gibbs measure, with particular attention paid to the most extreme waves.

Figure 6 shows the first test with cutoff wavenumber $K = 16$ and inverse temperature $\beta' = 40$. We consider three values of the nonlinearity ratio, $C_3/C_2 = 0, 75, 150$. For each, we use the improved rejection algorithm to draw 500 individual wave fields from the Gibbs measure. We then extract the wave field with the largest positive displacement value to display in Figure 6; that is, each wave field shown in Figure 6 represents a 1-in-500 event. For reference, the faint dotted line shows a displacement equal to four standard deviations from the mean, $4\sigma = 2.26$ since $\sigma = \sqrt{E_0/\pi}$ and $E_0 = 1$. For a Gaussian distribution, roughly one in 16000 samples lies at least four standard deviations from the mean, and this is sometimes taken to define a rogue-wave event [7]. We note that each wave field contains $2K = 32$ surface-displacement values and so a crude estimate for the number of wave field samples required to achieve the 4σ threshold under Gaussian statistics is $16000/32 = 500$. Figure 6a shows that in the linear case, $C_3/C_2 = 0$, the largest surface displacement lies below the 4σ threshold. Figure 6b,c shows that as nonlinearity increases, the 1-in-500 wave exceeds the 4σ threshold, with the largest value, $C_3/C_2 = 150$, giving a surface displacement 11% above the threshold. To appreciate this example from a different light, we note that, unlike a Gaussian distribution, the maximum displacement of a finite-mode wave field with fixed energy is bounded above. In particular, the Dirichlet kernel achieves the maximum possible displacement (see Figure 2). For $K = 16$ and $E_0 = 1$, the maximum possible displacement is $u_{\max} = 3.09$, and the displacement seen in Figure 6c is 81% of this value.

Figure 7 shows a similar test with the same cutoff wavenumber $K = 16$, but a larger inverse temperature $\beta' = 60$. The values of the nonlinearity parameter, $C_3/C_2 = 0, 50, 100$, are chosen to yield the same increments in $\beta' C_3/C_2$. As seen in Figure 7, the first two values of C_3/C_2 yields a 1-in-500 event that does not meet the 4σ threshold. Figure 7c shows that the last case does exceed the threshold by the same 11% observed in Figure 6c. Once again, the fixed energy and cutoff wavenumber imposes a hard cap of $u_{\max} = 3.09$ on the displacement. Our tests suggest that it is difficult to surpass roughly 80% of this cap, even with

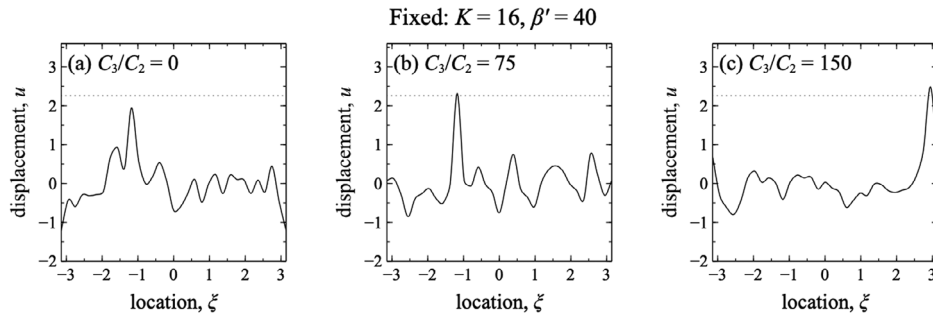


FIGURE 6 | Sampling individual wave fields to search for extreme events. For each C_3/C_2 , we show the individual wave field drawn from an ensemble of 500 that possesses the largest surface displacement, that is, a 1-in-500 event. Increasing nonlinearity (left to right) increases the value of the maximum. In the last two cases, the maximum displacement exceeds the 4σ threshold (faint dotted line). The parameters $K = 16$, $\mathcal{E}_0 = 1$, and $\beta' = 40$ are fixed.

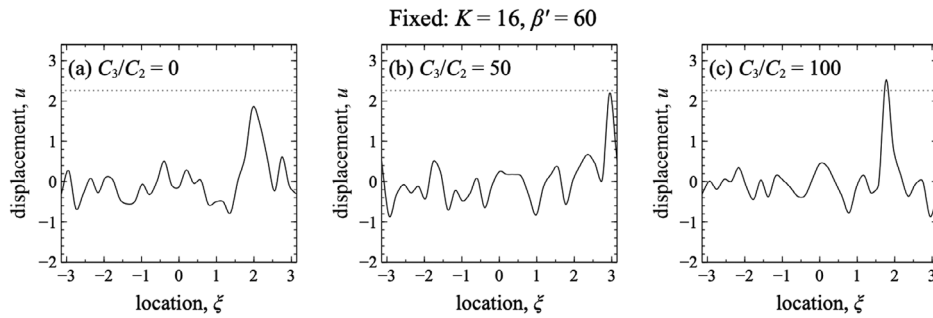


FIGURE 7 | Same test as in Figure 6 except with $\beta' = 60$ and different increments in C_3/C_2 . The first two cases of C_3/C_2 produce a 1-in-500 event that falls short of the 4σ threshold. In the last case, the 1-in-500 wave exceeds this threshold.

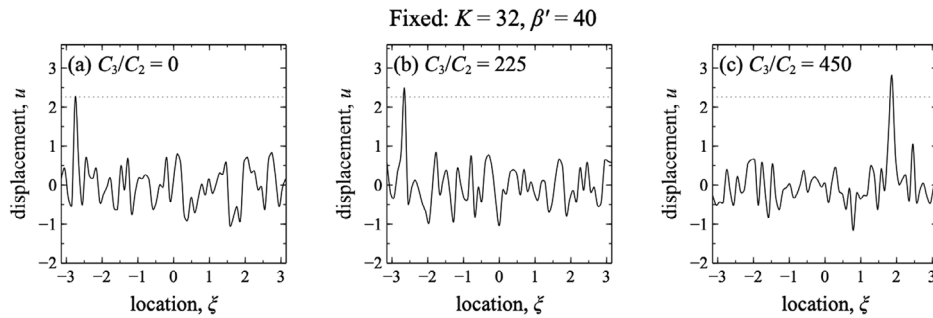


FIGURE 8 | Sampling of individual wave fields with $K = 32$. For each case, we show the wave field out of 500 that achieves the largest positive displacement. (b)–(c) The latter two cases with strong nonlinearity produce a displacement that exceeds the 4σ threshold. Fixed parameters are $\beta' = 40$ and $\mathcal{E}_0 = 1$.

significant skewness present in the ensemble (skewness = 0.85 for the ensemble of Figure 7c).

By increasing the cutoff wavenumber, we can increase the hard cap on the maximum surface displacement. In Figures 8, 9, we examine wave fields sampled with $K = 32$, for which the Dirichlet kernel achieves a maximum possible displacement of $u_{\max} = 4.44$. Figure 8 shows the 1-in-500 events for $\beta' = 40$ and three values of C_3/C_2 . As before, the linear case $C_3/C_2 = 0$ produces a maximal displacement slightly below the 4σ threshold, while the two nonlinear cases surpass the threshold. In the case of highest nonlinearity, the 1-in-500 maximal displacement exceeds

the 4σ threshold by 24%. This value is significantly greater than the 11% achieved in the tests with $K = 16$, despite significantly lower ensemble skewness (0.25 vs. 0.85). This test demonstrates the significance of increasing the hard cap, u_{\max} , by increasing the cutoff wavenumber. Unfortunately, achieving comparable values of ensemble skewness with higher K requires greater computational resources, as seen in Tables 1 and 2.

Figure 9 shows a similar test with $K = 32$ but with a larger inverse temperature of $\beta' = 60$ and the increments of C_3/C_2 chosen accordingly. Once again, the linear case produces a 1-in-500 event that falls short of the 4σ threshold, while both nonlinear cases

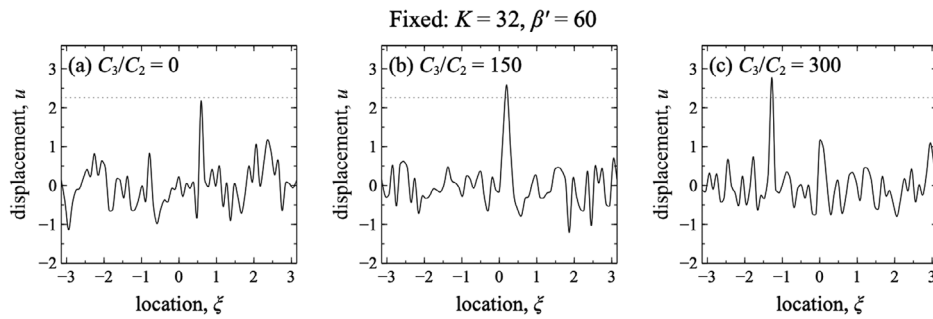


FIGURE 9 | Same test as in Figure 8 except with larger inverse temperature $\beta' = 60$. (c) The most extreme case $C_3/C_2 = 300$ produces a displacement that exceeds 4σ by 24%, which is the widest margin observed in all tests.

exceed the threshold. The most nonlinear case, Figure 9c, exceeds the threshold by roughly the same 24% seen in Figure 8c.

6 | Discussion

This paper introduces a novel proposal distribution for generating independent, random samples from the TKdV Gibbs measure. The proposal distribution stems from the analysis of Sun and Moore [28], who showed that it converges to the Gibbs measure in the infinite-mode limit of linear TKdV. Roughly speaking, this proposal distribution captures the spectral decay of sampled wave fields accurately so that there remains only one criterion for acceptance: sufficient skewness. Numerical tests indicate that the novel proposal distribution produces much higher acceptance rates than the spectrally uniform distribution, typically by 1–6 orders of magnitude, even for parameters well outside of the regime in which convergence can be established rigorously. Advantages of the algorithm over more sophisticated methods, like MCMC, include favorable parallelization properties and generation of uncorrelated samples.

Numerical tests demonstrate the algorithm's capability to generate extreme displacements over broad parameter ranges, with several sampled wave fields exceeding the 4σ criterion of a rogue wave. With a cutoff wavenumber of $K = 16$, tests indicate that 4σ events can only be generated when the ensemble skewness is sufficiently high (skewness >0.5). With $K = 32$, 4σ events can be generated with lower ensemble skewness (skewness >0.1 , although the required computational resources are similar).

The algorithmic foundations established here open many exciting avenues for future research. First, the sampling algorithm will be used to build a database of wave field ensembles corresponding to different regions of parameter space. The database will enable one to efficiently search across parameter space for extreme-wave events and characterize in detail the extreme events that occur in different parameter regimes. The database will also provide a repository of wave fields to be used as initial conditions in dynamic simulations. In particular, running extreme-wave initial conditions *backward* in time will reveal the wave characteristics that precede rogue events. When combined with neural-network feature detection [40–42], this process could identify warning signs of impending extreme wave activity. Lastly, it is an attractive prospect to couple nonlinear waves to larger-scale circulation patterns [43–46] in order to assess feedback between the two.

Acknowledgments

N. J. Moore recognizes support from the Office of Naval Research, award number N000142412617. B. Foerster recognizes support from the Colgate University student-initiated summer fellowship.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. S. Haver, "A Possible Freak Wave Event Measured at the Draupner Jacket January 1 1995," in *Rogue waves*, Vol. 2004 (2004), 1–8.
2. P. Müller, C. Garrett, and A. Osborne, "Rogue Waves: The Fourteenth 'Aha Huli' Hawaiian Winter Workshop," *Oceanography* 18, no. 3 (2005): 66–75.
3. C. Garrett and J. Gemmrich, "Rogue Waves," *Physics Today* 62, no. 6 (2009): 62–63.
4. A. Hadjihosseini, J. Peinke, and N. Hoffmann, "Stochastic Analysis of Ocean Wave States With and Without Rogue Waves," *New Journal of Physics* 16, no. 5 (2014): 053037.
5. G. Dematteis, T. Grafke, and E. Vanden-Eijnden, "Rogue Waves and Large Deviations in Deep Sea," *Proceedings of the National Academy of Sciences* 115, no. 5 (2018): 855–860.
6. G. Dematteis, T. Grafke, M. Onorato, and E. Vanden-Eijnden, "Experimental Evidence of Hydrodynamic Instantons: The Universal Route to Rogue Waves," *Physical Review X* 9, no. 4 (2019): 041057.
7. C. T. Bolles, K. Speer, and M. Moore, "Anomalous Wave Statistics Induced by Abrupt Depth Change," *Physical Review Fluids* 4, no. 1 (2019): 011801.
8. J. M. Dudley, G. Genty, A. Mussot, A. Chabchoub, and F. Dias, "Rogue Waves and Analogies in Optics and Oceanography," *Nature Reviews Physics* 1, no. 11 (2019): 675–689.
9. K. Trulsen, H. Zeng, and O. Gramstad, "Laboratory Evidence of Freak Waves Provoked by Non-Uniform Bathymetry," *Physics of Fluids* 24 (2012): 097101.
10. C. Viotti and F. Dias, "Extreme Waves Induced by Strong Depth Transitions: Fully Nonlinear Results," *Physics of Fluids* 26, no. 5 (2014): 051705.
11. A. J. Majda, M. Moore, and D. Qi, "Statistical Dynamical Model to Predict Extreme Events and Anomalous Features in Shallow Water Waves With Abrupt Depth Change," *Proceedings of the National Academy of Sciences* 116, no. 10 (2019): 3982–3987.
12. A. J. Majda and D. Qi, "Statistical Phase Transitions and Extreme Events in Shallow Water Waves With an Abrupt Depth Change," *Journal of Statistical Physics* 179, no. 5 (2020): 1718–1741.

13. D. Qi and A. J. Majda, "Using Machine Learning to Predict Extreme Events in Complex Systems," *Proceedings of the National Academy of Sciences* 117, no. 1 (2020): 52–59.
14. J. G. Herterich and F. Dias, "Extreme Long Waves Over a Varying Bathymetry," *Journal of Fluid Mechanics* 878 (2019): 481–501.
15. K. Trulsen, A. Raustøl, S. Jorde, and L. B. Rye, "Extreme Wave Statistics of Long-Crested Irregular Waves Over a Shoal," *Journal of Fluid Mechanics* 882 (2020).
16. N. J. Moore, C. T. Bolles, A. J. Majda, and D. Qi, "Anomalous Waves Triggered by Abrupt Depth Changes: Laboratory Experiments and Truncated KdV Statistical Mechanics," *Journal of Nonlinear Science* 30, no. 6 (2020): 3235–3263.
17. Y. Zheng, Z. Lin, Y. Li, T. Adcock, Y. Li, and T. S. van den Bremer, "Fully Nonlinear Simulations of Unidirectional Extreme Waves Provoked by Strong Depth Transitions: The Effect of Slope," *Physical Review Fluids* 5, no. 6 (2020): 064804.
18. D. Qi and E. Vanden-Eijnden, "Anomalous Statistics and Large Deviations of Turbulent Water Waves Past a Step," *AIP Advances* 12, no. 2 (2022): 025016.
19. Y. Li, Y. Zheng, Z. Lin, T. A. Adcock, and T. S. van den Bremer, "Surface Wavepackets Subject to an Abrupt Depth Change. Part 1. Second-Order Theory," *Journal of Fluid Mechanics* 915 (2021).
20. Y. Li, S. Draycott, T. A. Adcock, and T. S. van den Bremer, "Surface Wavepackets Subject to an Abrupt Depth Change. Part 2. Experimental Analysis," *Journal of Fluid Mechanics* 915 (2021).
21. Y. Li, S. Draycott, Y. Zheng, Z. Lin, T. A. Adcock, and T. S. van den Bremer, "Why Rogue Waves Occur Atop Abrupt Depth Transitions," *Journal of Fluid Mechanics* 919 (2021).
22. C. Lawrence, K. Trulsen, and O. Gramstad, "Statistical Properties of Wave Kinematics in Long-Crested Irregular Waves Propagating Over Non-Uniform Bathymetry," *Physics of Fluids* 33, no. 4 (2021): 046601.
23. S. Mendes, A. Scotti, M. Brunetti, and J. Kasparian, "Non-Homogeneous Analysis of Rogue Wave Probability Evolution Over a Shoal," *Journal of Fluid Mechanics* 939 (2022).
24. M. S. Afzal and L. Kumar, "Propagation of Waves Over a Rugged Topography," *Journal of Ocean Engineering and Science* 7, no. 1 (2022): 14–28.
25. P. D. Lax, "Periodic Solutions of the KdV Equation," *Communications on Pure and Applied Mathematics* 28, no. 1 (1975): 141–188.
26. R. S. Johnson, *A Modern Introduction to the Mathematical Theory of Water Waves*. Vol. 19 (Cambridge University Press, 1997).
27. G. B. Whitham, *Linear and Nonlinear Waves*, Vol. 42 (Wiley, 2011).
28. H. Sun and N. J. Moore, "On Normal and Non-Normal Wave Statistics Implied by a Canonical–Microcanonical Gibbs Ensemble of the Truncated KdV System," *Journal of Statistical Physics* 190, no. 4 (2023).
29. D. M. Ambrose, R. Camassa, J. L. Marzuola, R. M. McLaughlin, Q. Robinson, and J. Wilkening, "Numerical Algorithms for Water Waves With Background Flow Over Obstacles and Topography," *Advances in Computational Mathematics* 48, no. 4 (2022): 46.
30. D. Crowdy, *Solving Problems in Multiply Connected Domains* (SIAM, 2020).
31. P. J. Baddoo, N. J. Moore, A. U. Oza, and D. G. Crowdy, "Generalization of Waving-Plate Theory to Multiple Interacting Swimmers," *Communications on Pure and Applied Mathematics* 76, no. 12 (2023): 3811–3851.
32. J. Bajars, J. E. Frank, and B. J. Leimkuhler, "Weakly Coupled Heat Bath Models for Gibbs-Like Invariant States in Nonlinear Wave Equations," *Nonlinearity* 26, no. 7 (2013): 1945–1974.
33. H. Sun, N. J. Moore, and F. Bao, "Parameter Estimation for the Truncated KdV Model through a Direct Filter Method," *Journal of Machine Learning for Modeling and Computing* 4, no. 1 (2023): 109–132.
34. A. Majda and X. Wang, *Nonlinear Dynamics and Statistical Theories for Basic Geophysical Flows* (Cambridge University Press, 2006).
35. R. V. Abramov, G. Kovačič, and A. J. Majda, "Hamiltonian Structure and Statistically Relevant Conserved Quantities for the Truncated Burgers-Hopf Equation," *Communications on Pure and Applied Mathematics* 56, no. 1 (2003): 1–46.
36. S. Ross, *A First Course in Probability*, 8th ed. (Pearson Education, 2010).
37. Ø. Skare, E. Bølviken, and L. Holden, "Improved Sampling-Importance Resampling and Reduced Bias Importance Sampling," *Scandinavian Journal of Statistics* 30, no. 4 (2003): 719–737.
38. W. A. Strauss, *Partial Differential Equations: An Introduction* (Wiley, 2007).
39. J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence Properties of the Nelder–Mead Simplex Method in Low Dimensions," *SIAM Journal on Optimization* 9, no. 1 (1998): 112–147.
40. Z. Zhang, P. Jaiswal, and R. Rai, "FeatureNet: Machining Feature Recognition Based on 3D Convolution Neural Network," *Computer-Aided Design* 101 (2018): 12–22.
41. D. M. Pelt and J. A. Sethian, "A Mixed-Scale Dense Convolutional Neural Network for Image Analysis," *Proceedings of the National Academy of Sciences* 115, no. 2 (2018): 254–259.
42. D. Qi, "Unambiguous Models and Machine Learning Strategies for Anomalous Extreme Events in Turbulent Dynamical System," *Entropy* 26, no. 6 (2024): 522.
43. J. Pedlosky, *Geophysical Fluid Dynamics* (Springer Science & Business Media, 2013).
44. H. A. Dijkstra and M. Ghil, "Low-Frequency Variability of the Large-Scale Ocean Circulation: A Dynamical Systems Approach," *Reviews of Geophysics* 43, no. 3 (2005).
45. N. J. Moore and J. Mac Huang, "Large-Scale Circulation Reversals Explained by Pendulum Correspondence," *Journal of Fluid Mechanics* 993 (2024): A3.
46. J. A. Whitehead, *Energy Flow and Earth* (Springer, 2024).

Appendix A: Validation With Two-Mode Exact Solutions

In the case $K = 2$, it is possible to solve exactly for the Hamiltonian components \mathcal{H}_2 and \mathcal{H}_3 , offering a way to validate the numerical computation of these quantities. In particular, consider arbitrary values of two wave modes \hat{u}_1 and \hat{u}_2 written in polar form:

$$\hat{u}_1 = R_1 e^{i\theta_1}, \quad \hat{u}_2 = R_2 e^{i\theta_2}. \quad (\text{A1})$$

The energy constraint, Equation (12), gives

$$2\pi(|\hat{u}_1|^2 + |\hat{u}_2|^2) = 2\pi(R_1^2 + R_2^2) = \mathcal{E}_0. \quad (\text{A2})$$

Therefore, the moduli can then be written as

$$(R_1, R_2) = \sqrt{\frac{\mathcal{E}_0}{2\pi}} (\cos(\phi), \sin(\phi)) \quad (\text{A3})$$

for some $\phi \in [0, 2\pi)$. Thus, to select a set of arbitrary modes \hat{u}_1 and \hat{u}_2 satisfying the energy constraint, it suffices to specify the values $\theta_1, \theta_2, \phi \in [0, 2\pi)$. For example, these three angles could be selected randomly from a uniform distribution. Calculating \mathcal{H}_2 from Equation (20) and \mathcal{H}_3 from Equation (42) gives the exact solutions:

$$\mathcal{H}_2 = 2\pi(R_1^2 + 4R_2^2), \quad \mathcal{H}_3 = 2\pi R_2 R_1^2 \cos(2\theta_1 - \theta_2). \quad (\text{A4})$$

These exact solutions can be used to validate the numerical computation of \mathcal{H}_2 and \mathcal{H}_3 for any values of $\theta_1, \theta_2, \phi \in [0, 2\pi)$.