Statistical dynamical model to predict extreme events and anomalous features in shallow water waves with abrupt depth change

Andrew J. Majda[a,b,1], M. N. J. Moore[c,d], and Di Qi[a,b,1]

[a]Department of Mathematics, Courant Institute of Mathematical Sciences, New York University, New York, NY 10012; [b]Center for Atmosphere and Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, NY 10012; [c]Department of Mathematics, Florida State University, Tallahassee, FL 32304; and [d]Geophysical Fluid Dynamics Institute, Florida State University, Tallahassee, FL 32306

Contributed by Andrew J. Majda, December 18, 2018 (sent for review November 30, 2018; reviewed by Themistoklis P. Sapsis and Joseph Tribbia)

Understanding and predicting extreme events and their anomalous statistics in complex nonlinear systems are a grand challenge in climate, material, and neuroscience as well as for engineering design. Recent laboratory experiments in weakly turbulent shallow water reveal a remarkable transition from Gaussian to anomalous behavior as surface waves cross an abrupt depth change (ADC). Downstream of the ADC, probability density functions of surface displacement exhibit strong positive skewness accompanied by an elevated level of extreme events. Here, we develop a statistical dynamical model to explain and quantitatively predict the above anomalous statistical behavior as experimental control parameters are varied. The first step is to use incoming and outgoing truncated Korteweg–de Vries (TKdV) equations matched in time at the ADC. The TKdV equation is a Hamiltonian system, which induces incoming and outgoing statistical Gibbs invariant measures. The statistical matching of the known nearly Gaussian incoming Gibbs state at the ADC completely determines the predicted anomalous outgoing Gibbs state, which can be calculated by a simple sampling algorithm verified by direct numerical simulations, and successfully captures key features of the experiment. There is even an analytic formula for the anomalous outgoing skewness. The strategy here should be useful for predicting extreme anomalous statistical behavior in other dispersive media.

extreme anomalous event | statistical TKdV model | matching Gibbs measures | surface wave displacement | surface wave slope

Understanding and predicting extreme events and their anomalous statistics in complex nonlinear systems are a grand challenge in climate, material, and neuroscience as well as for engineering design. This is a very active contemporary topic in applied mathematics with qualitative and quantitative models (1–7) and novel numerical algorithms, which overcome the curse of dimensionality for extreme event prediction in large complex systems (2, 8–11). The occurrence of Rogue waves as extreme events within different physical settings of deep water (12–16) and shallow water (17–19) is an important practical topic.

Recent laboratory experiments in weakly turbulent shallow water reveal a remarkable transition from Gaussian to anomalous behavior as surface waves cross an abrupt depth change (ADC). A normally distributed incoming wave train, upstream of the ADC, transitions to a highly non-Gaussian outgoing wave train, downstream of the ADC, that exhibits large positive skewness of the surface height and more frequent extreme events (20). Here, we develop a statistical dynamical model to explain and quantitatively predict this anomalous behavior as experimental control parameters are varied. The first step is to use incoming and outgoing truncated Korteweg–de Vries (TKdV) equations matched in time at the ADC. The TKdV equation is a Hamiltonian system, which induces incoming and outgoing Gibbs invariant measures. The statistical matching of the known nearly Gaussian incoming Gibbs state at the ADC completely determines the predicted anomalous outgoing Gibbs state, which can be calculated by a simple Markov chain Monte Carlo algorithm verified by direct numerical simulations, and successfully captures key features of the experiment. There is even an analytic formula for the anomalous outgoing skewness. The strategy here should be useful for predicting extreme anomalous statistical behavior in other dispersive media in different settings (21, 22).

1. Experiments Showing Anomalous Wave Statistics Induced by an ADC

Controlled laboratory experiments were carried out in ref. 20 to examine the statistical behavior of surface waves crossing an ADC. In these experiments, nearly unidirectional waves are generated by a paddle wheel and propagate through a long, narrow wave tank. Midway through, the waves encounter a step in the bottom topography and thus, abruptly transition to a shallower depth. The paddle wheel is forced with a pseudorandom signal intended to mimic a Gaussian random sea upstream of the ADC. In particular, the paddle angle is specified as

$$\theta(t) = \theta_0 + \Delta\theta \sum_{n=1}^{N} a_n \cos(\omega_n t + \delta_n),\ E \sim (\Delta\theta)^2 \sum_n a_n^2 \omega_n^2,$$

Significance

Understanding and predicting extreme events and their anomalous statistics in complex nonlinear systems are a grand challenge in applied sciences as well as for engineering design. Recent controlled laboratory experiments in weakly turbulent shallow water with abrupt depth change exhibit a remarkable transition from nearly Gaussian statistics to extreme anomalous statistics with large positive skewness of the surface height. We develop a statistical dynamical model to explain and quantitatively predict the anomalous statistical behavior. Incoming and outgoing waves are modeled by the truncated Korteweg–de Vries equations statistically matched at the depth change. The statistical matching of the known nearly Gaussian incoming Gibbs state completely determines the predicted anomalous outgoing Gibbs state and successfully captures key features of the experiment.

Author contributions: A.J.M. designed research; A.J.M., M.N.J.M., and D.Q. performed research; and A.J.M., M.N.J.M., and D.Q. wrote the paper.

Reviewers: T.P.S., Massachusetts Institute of Technology; and J.T., National Center for Atmospheric Research.

The authors declare no conflict of interest.

[1]To whom correspondence may be addressed. Email: jonjon@cims.nyu.edu or qidi@cims.nyu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1820467116/-/DCSupplemental.

Published online February 13, 2019.

where the weights a_n are Gaussian in spectral space with peak frequency ω_p and the phases δ_n are uniformly distributed random variables. The peak frequency gives rise to a characteristic wavelength λ_c, which can be estimated from the dispersion relation. The energy E injected into the system is determined by the angle amplitude $\Delta\theta$, which is the main control parameter varied in ref. 20. Optical measurements of the free surface reveal a number of surprising statistical features.

- Distinct statistics are found between the incoming and outgoing wave disturbances: incoming waves display near-Gaussian statistics, while outgoing waves skew strongly toward positive displacement.
- The degree of non-Gaussianity present in the outgoing waves depends on the injected energy E: larger energies generate stronger skewness in the surface displacement probability density functions (PDFs) and more extreme events.
- Compared with the incoming wave train, the power spectrum of the outgoing wave field decays more slowly, which indicates that the anomalous behavior is associated with an elevated level of high frequencies.

2. Surface Wave Turbulence Modeled by TKdV Equation with Depth Dependence

The Korteweg–de Vries (KdV) equation is a one-dimensional, deterministic model capable of describing surface wave turbulence. More specifically, KdV is leading-order approximation for surface waves governed by a balance of nonlinear and dispersive effects valid in an appropriate far-field limit (23). Moreover, KdV has been adapted to describe waves propagating over variable depth (23). Here, we consider the variable-depth KdV equation truncated at wavenumber Λ (with $J=2\Lambda+1$ grid points) to generate weakly turbulent dynamics. The surface displacement is described by the state variable $u_\Lambda^\pm(x,t)$, with superscript $-$ for the incoming waves and superscript $+$ for the outgoing waves. The Galerkin truncated variable $u_\Lambda=\sum_{1\le|k|\le\Lambda}\hat{u}_k(t)e^{ikx}$ is normalized with zero mean $\hat{u}_0=0$ and unit energy $2\pi\sum_{k=1}^{\Lambda}|\hat{u}_k|^2=1$, which are conserved quantities. Here, $u_\Lambda\equiv\mathcal{P}_\Lambda u$ denotes the subspace projection. The evolution of $u_\Lambda^\pm$ is governed by the truncated KdV equation with depth change $D_\pm$:

$$\frac{\partial u_\Lambda^\pm}{\partial t}+\frac{D_\pm^{-3/2}}{2}E_0^{1/2}L_0^{-3/2}\frac{\partial}{\partial x}\mathcal{P}_\Lambda\left(u_\Lambda^\pm\right)^2+D_\pm^{1/2}L_0^{-3}\frac{\partial^3 u_\Lambda^\pm}{\partial x^3}=0, \qquad [1]$$

Eq. **1** is nondimensionalized on the periodic domain $x\in[-\pi,\pi]$. The depth is assumed to be unit $D_-=1$ before the ADC and $D_+<1$ after the ADC. The conserved Hamiltonian can be decomposed as

$$\mathcal{H}_\Lambda^\pm=D_\pm^{-3/2}E_0^{1/2}L_0^{-3/2}H_3\left(u_\Lambda^\pm\right)-D_\pm^{1/2}L_0^{-3}H_2\left(u_\Lambda^\pm\right)$$
$$H_3(u)=\frac{1}{6}\int_{-\pi}^{\pi}u^3dx,\ H_2(u)=\frac{1}{2}\int_{-\pi}^{\pi}\left(\frac{\partial u}{\partial x}\right)^2dx.$$

where we refer to H_3 as the cubic term and H_2 as the quadratic term. We introduce parameters (E_0,L_0,Λ) based on the following assumptions.

- The wavenumber truncation Λ is fixed at a moderate value for generating weakly turbulent dynamics.
- The state variable $u_\Lambda^\pm$ is normalized with zero mean and unit energy, $\mathcal{M}(u_\Lambda)=0,\mathcal{E}(u_\Lambda)=1$, which are conserved during evolution. Meanwhile, E_0 characterizes the total energy injected into the system based on the driving amplitude $\Delta\theta$.
- The length scale of the system L_0 is chosen so that the resolved scale $\Delta x=2\pi L_0/J$ is comparable with the characteristic wave length λ_c from the experiments.

Some intuition for how Eq. **1** produces different dynamics on either side of the ADC can be gained by considering the relative contributions of H_3 and H_2 in the Hamiltonian $\mathcal{H}_\Lambda^\pm$. The depth change, $D_+<1$, increases the weight of H_3 and decreases that of H_2, thereby promoting the effects of nonlinearity over dispersion and creating conditions favorable for extreme events. Since $\frac{\partial u}{\partial x}$ is the slope of the wave height, $H_2(u)$ measures the wave slope energy.

A deterministic matching condition is applied to the surface displacement $u_\Lambda^\pm$ to link the incoming and outgoing wave trains. Assuming that the ADC is met at $t=T_{\mathrm{ADC}}$, the matching condition is given by

$$u_\Lambda^-(x,t)|_{t=T_{\mathrm{ADC}}-}=u_\Lambda^+(x,t)|_{t=T_{\mathrm{ADC}}+}.$$

Eq. **1** is not designed to capture the short scale changes in rapid time. Rather, since we are interested in modeling statistics before and after the ADC, we will examine the long-time dynamics of large-scale structures.

Interpreting Experimental Parameters in the Dynamical Model. The model parameters (E_0,L_0,Λ) in [**1**] can be directly linked to the basic scales from the physical problem. The important parameters that characterize the experiments (20) include $\epsilon=\frac{a}{H_0}$, the wave amplitude a to (upstream) water depth H_0 ratio; $\delta=\frac{H_0}{\lambda_c}$, the water depth to wavelength λ_c ratio; and $D_0=\frac{d}{H_0}$, the depth ratio with upstream value $d=H_0$ and downstream value $d<H_0$. By comparing the characteristic physical scales, the normalized TKdV Eq. **1** can be linked to these experimental parameters via

$$L_0=6^{\frac{1}{3}}\left(M\epsilon^{\frac{1}{2}}\delta^{-1}\right),\ E_0=\frac{27}{2}\gamma^{-2}\left(M\epsilon^{\frac{1}{2}}\delta^{-1}\right), \qquad [2]$$

where M defines the computational domain size $M\lambda_c$ as M multiple of the characteristic wavelength λ_c, and $\gamma=\frac{U}{a}$, with U a scaling factor for the state variable u_Λ, normalizes the total energy of the system to 1.

Consider a spatial discretization with $J=2\Lambda+1$ grid points so that the smallest resolved scale is comparable with the characteristic wavelength

$$\Delta x=\frac{2\pi M\lambda_c}{J}\lesssim\lambda_c\Rightarrow M=\frac{J}{2\pi}\sim5,\ J=32.$$

In the practical numerical simulations, we select $M=5$ and let γ vary in the range $[0.5,1]$. Using reference values from the experiments (20), we find $\epsilon\in[0.0024,0.024]$, $\delta\sim0.22$, and D_0 changes from 1 to 0.24 at the ADC. These values yield the following ranges for the model parameters: $L_0\in[2,6]$ and $E_0\in[50,200]$. These are the values that we will test in the direct numerical simulations. Details about the derivation from scale analysis are in *SI Appendix, Section A*.

3. Equilibrium Statistical Mechanics for Generating the Stationary Invariant Measure

Since the TKdV equation satisfies the Liouville property, the equilibrium invariant measure can be described by a statistical formalism (24–26) based on a Gibbs measure with the conserved energy $\mathcal{E}_\Lambda$ and Hamiltonian $\mathcal{H}_\Lambda$. The equilibrium invariant measure is dictated by the conservation laws in the TKdV equation. In the case of fixed energy E_0, this is the mixed Gibbs measure with microcanonical energy and canonical Hamiltonian ensembles (24)

$$\mathcal{G}_\theta^\pm\left(u_\Lambda^\pm;E_0\right)=C_\theta^\pm\exp\left(-\theta^\pm\mathcal{H}\left(u_\Lambda^\pm\right)\right)\delta\left(\mathcal{E}\left(u_\Lambda^\pm\right)-E_0\right), \qquad [3]$$

where θ represents the "inverse temperature." The distinct statistics in the upstream and downstream waves can be

controlled with θ. As shown below, we find that negative temperature, $\theta^{\pm}<0$, is the appropriate regime to predict the experiments. In the incoming wave field, θ^{-} is chosen so that $\mathcal{G}_{\theta}^{-}$ has nearly Gaussian statistics. Using the above invariant measures [3], the expectation of any functional $F(u)$ can be computed as

$$\langle F\rangle_{\mathcal{G}_\theta} \equiv \int F(u)\,\mathcal{G}_\theta(u)\,du.$$

The value of θ in the invariant measure is specified from $\langle H_\Lambda\rangle_{\mathcal{G}_\theta}$ (24, 26).

In addition to producing equilibrium PDFs of u_Λ, the invariant measure can be used to predict the equilibrium energy spectrum without the direct simulation of TKdV. Direct simulation, however, is required to recover transient statistics of u_Λ and time autocorrelations.

Statistical Matching Condition of the Invariant Measures Before and After the ADC. The Gibbs measures $\mathcal{G}_{\theta}^{\pm}$ are defined based on the different inverse temperatures $\theta^{\pm}$ on the two sides of the solutions

$$\mu_t^{-}\left(u_\Lambda^{-};D_{-}\right),\quad u_\Lambda|_{t=T_{\mathrm{ADC}-}}=u_0,\ t<T_{\mathrm{ADC}};$$
$$\mu_t^{+}\left(u_\Lambda^{+};D_{+}\right),\quad u_\Lambda|_{t=T_{\mathrm{ADC}+}}=u_0,\ t>T_{\mathrm{ADC}},$$

where u_0 represents the deterministic matching condition between the incoming and outgoing waves. The two distributions, μ_t^{-},μ_t^{+}, should be matched at T_{ADC}, giving

$$\mu^{-}_{t=T_{\mathrm{ADC}}}(u_\Lambda)=\mu^{+}_{t=T_{\mathrm{ADC}}}(u_\Lambda).$$

In matching the flow statistics before and after the ADC, we first use the conservation of the deterministic Hamiltonian H_Λ^{+} after the depth change. Then, assuming ergodicity (24, 25), the statistical expectation for the Hamiltonian $\langle H_\Lambda^{+}\rangle$ is conserved in time after the depth change at $t=T_{\mathrm{ADC}}$ and should remain at this value as the system approaches equilibrium as $t\to\infty$. The final statistical matching condition to get the outgoing flow statistics with parameter θ^{+} can be found by

$$\langle H_\Lambda^{+}\rangle_{\mathcal{G}_\theta^{+}}=\langle H_\Lambda^{+}\rangle_{\mathcal{G}_\theta^{-}},\qquad [4]$$

with the outgoing flow Hamiltonian H_Λ^{+} and the Gibbs measures $\mathcal{G}_{\theta}^{\pm}$ before and after the ADC.

4. Nearly Gaussian Incoming Statistical State

For the parameters explored in ref. 20, the incoming wave field is always characterized by a near-Gaussian distribution of the surface displacement. It is found that a physically consistent Gibbs measure should take negative values in the inverse temperature parameter $\theta<0$, where a proper distribution function and a decaying energy spectrum are generated (ref. 26 and *SI Appendix*, *Section B.1* have explicit simulation results). The upstream Gibbs measure $\mathcal{G}_{\theta}^{-}$ with $D_{-}=1$ displays a wide parameter regime in (θ^{-},E_0) with near-Gaussian statistics. In Fig. 1*A*, *Left*, the inflow skewness κ_3^{-} varies only slightly with changing values of E_0 and θ^{-}. The incoming flow PDF then can be determined by picking the proper parameter value θ^{-} in the near-Gaussian regime with small skewness. In contrast, the downstream Gibbs measure $\mathcal{G}_{\theta}^{+}$ with $D_{+}=0.24$ shown in Fig. 1*A*, *Right* generates much larger skewness κ_3^{+} as the absolute value of θ^{+} and the total energy level E_0 increases. The solid lines in Fig. 1*C* offer a further confirmation of the transition from near-Gaussian statistics with small κ_3^{-} to a strongly skewed distribution κ_3^{+} after the depth change.

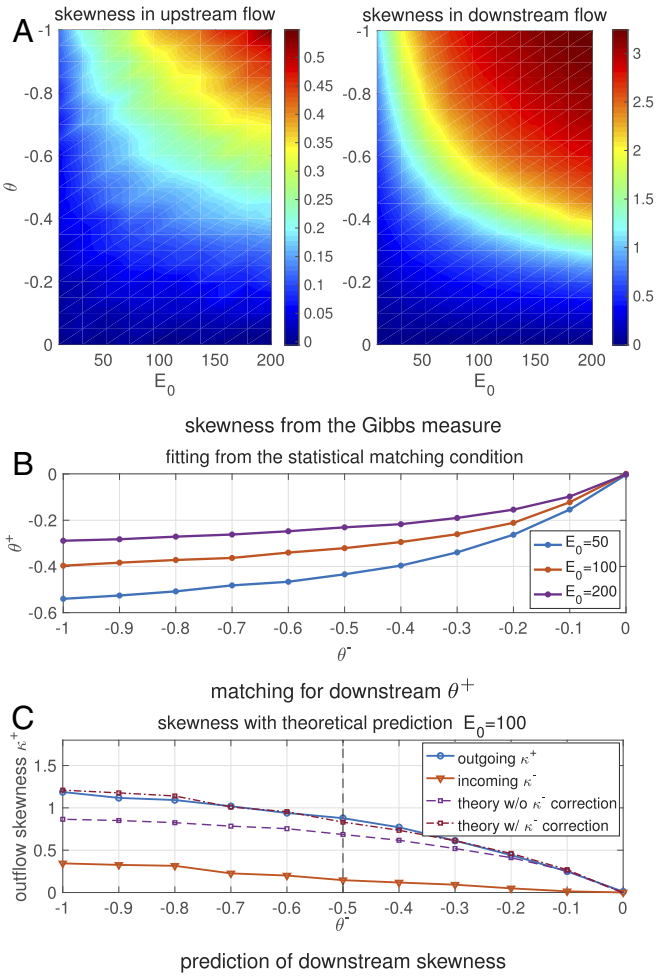

Fig. 1. (*A*) Skewness from the Gibbs measures in incoming and outgoing flow states with different values of total energy E_0 and inverse temperature θ (notice the different scales in the incoming and outgoing flows). (*B*) Outgoing flow parameter θ^{+} as a function of the incoming flow θ^{-} computed from the statistical matching condition with three energy levels E_0. (*C*) Skewness in the outgoing flow with the matched value of θ^{+} as a function of the inflow parameter θ^{-} (the theoretical predictions using [5] are compared).

In the next step, the value of the downstream θ^{+} is determined based on the matching condition [4]. The expectation $\langle H_\Lambda^{+}\rangle_{\mathcal{G}_\theta^{-}}$ about the incoming flow Gibbs measure can be calculated according to the predetermined parameter values of θ^{-} as well as E_0 from the previous step. For the direct numerical experiments shown later in Fig. 2, we pick proper choices of test parameter values as $L_0=6$, $E_0=100$, and $\theta^{-}=-0.1,-0.3,-0.5$. More test cases with different system energy E_0 can be found in *SI Appendix*, *Section B.2*, where similar transitions from near-Gaussian symmetric PDFs to skewed PDFs in the flow state $u_\Lambda^{\pm}$ can always be observed.

Direct Numerical Model Simulations. Other than the prediction of equilibrium statistical measures from the equilibrium statistical approach, another way to predict the downstream model statistics is through running the dynamical model [1] directly. The TKdV equation is found to be ergodic with proper mixing property as measured by the decay of autocorrelations as long as the system starts from a negative inverse temperature state as described before. For direct numerical simulations of the TKdV equations, a proper symplectic integrator is required to

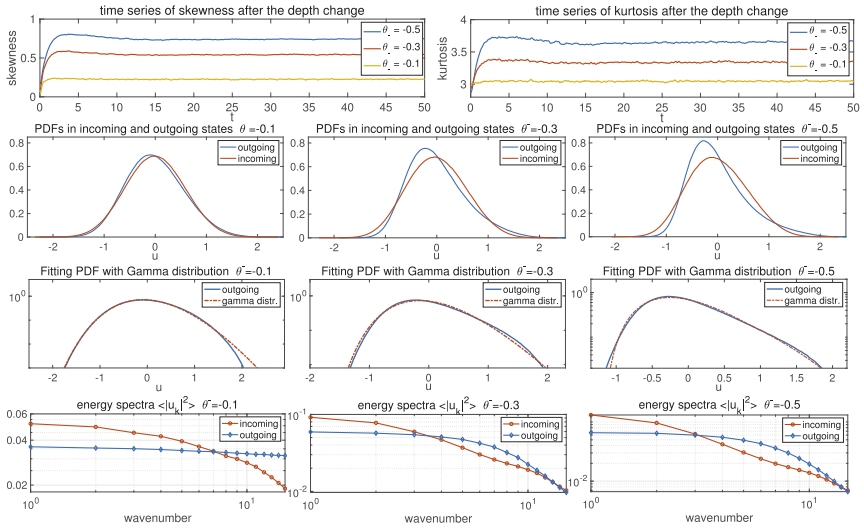

Fig. 2. Changes in the statistics of the flow state going through the ADC. The initial ensemble is set with the incoming flow Gibbs measure with different inverse temperature θ^-. (Row 1) Time evolution of the skewness and kurtosis. The ADC is taking place at $t=0$. (Row 2) Inflow and outflow PDFs of u_Λ. (Row 3) The downstream PDFs fitted with gamma distributions with consistent variance and skewness (in log coordinate in y). (Row 4) Energy spectra in the incoming and outgoing flows.

guarantee that the Hamiltonian and energy are conserved in time. It is crucial to use the symplectic scheme to guarantee the exact conservation of the energy and Hamiltonian, since they are playing the central role in generating the invariant measure and the statistical matching. The symplectic scheme used here for the time integration of the equation is the fourth-order midpoint method (27). Details about the mixing properties from different initial states and the numerical algorithm are described in *SI Appendix*, *Section C*.

5. Predicting Extreme Anomalous Behavior After the ADC by Statistical Matching

With the inflow statistics well described and the numerical scheme set up, we are able to predict the downstream anomalous statistics starting from the near-Gaussian incoming flow going through the ADC from $D_-=1$ to $D_+=0.24$. We first consider the statistical prediction in the downstream equilibrium measure directly from the matching condition. The downstream parameter value θ^+ is determined by solving the nonlinear Eq. **4** as a function of θ^+, $F\left(\theta^+\right)=\left\langle H_\Lambda^+\right\rangle_{\mathcal{G}_\theta^+}\left(\theta^+\right)-\left\langle H_\Lambda^+\right\rangle_{\mathcal{G}_\theta^-}=0$. In the numerical approach, we adopt a modified secant method, avoiding the stiffness in the parameter regime (*SI Appendix*, *Section B.2* shows the algorithm). The fitted solution is plotted in Fig. 1*B* as a function of the proposed inflow θ^-. A nonlinear $\theta^- - \theta^+$ relation is discovered from the matching condition. The downstream inverse temperature θ^+ will finally saturate at some level. The corresponding downstream skewness of the wave displacement u_Λ predicted from the statistical matching of Gibbs measures is plotted in Fig. 1*C*. In general, a large positive skewness for outgoing flow κ_3^+ is predicted from the theory, while the incoming flow skewness κ_3^- is kept in a small value in a wide range of θ^-. Note that, with $\theta^-\sim 0$ (that is, using the microcanonical ensemble only with energy conservation), the outflow statistics are also near Gaussian with weak skewness. The skewness in the outflow statistics grows as the inflow parameter value θ^- increases in amplitude.

For a second approach, we can use direct numerical simulations starting from the initial state sampled from the incoming flow Gibbs measure $\mathcal{G}_\theta^-$ and check the transient changes in the model statistics. Fig. 2 illustrates the change of statistics as the flow goes through the ADC. Fig. 2, row 1 plots the changes in the skewness and kurtosis for the state variable u_Λ after the depth change at $t=0$. The PDFs in the incoming and outgoing flow states are compared with three different initial inverse temperatures θ^-. After the depth changes to $D_0=0.24$ abruptly at $t=0$, both the skewness and kurtosis jump to a much larger value in a short time, implying the rapid transition to a highly skewed non-Gaussian statistical regime after the depth change. Furthermore, from Fig. 2, different initial skewness (but all relatively small) is set due to the various values of θ^-. With small $\theta^-=-0.1$, the change in the skewness is not very obvious (Fig. 2, row 2 shows the incoming and outgoing PDFs of u_Λ). In comparison, if the incoming flow starts from the initial parameter $\theta^-=-0.3$ and $\theta^-=-0.5$, a much larger increase in the skewness is induced from the ADC. Furthermore, in the detailed plots in Fig. 2, row 3 for the downstream PDFs under logarithmic scale, fat tails toward the positive direction can be observed, which represent the extreme events in the downstream flow (Fig. 3 shows the time series of u_Λ).

As a result, the downstream statistics in final equilibrium predicted from the direct numerical simulations here agree with the equilibrium statistical mechanics prediction illustrated in Fig. 1. The predictions from these two different approaches confirm each other.

6. Analytic Formula for the Upstream Skewness After the ADC

A statistical link between the upstream and downstream energy spectra can be found for an analytical prediction of the skewness in the flow state u after the ADC. The skewness of the state variable u_j at one spatial grid point is defined as the ratio between the third and second moments

$$\kappa_3=\left\langle u_j^3\right\rangle_\mu / \left\langle u_j^2\right\rangle_\mu^{\frac{3}{2}}.$$

Now, we introduce mild assumptions on the distribution functions.

- The upstream equilibrium measure μ_- has a relatively small skewness so that

$$\left\langle H_3\right\rangle_{\mu_-}=\frac{1}{6}\int_{-\pi}^{\pi}\left\langle u^3\right\rangle_{\mu_-}dx\equiv\epsilon;$$

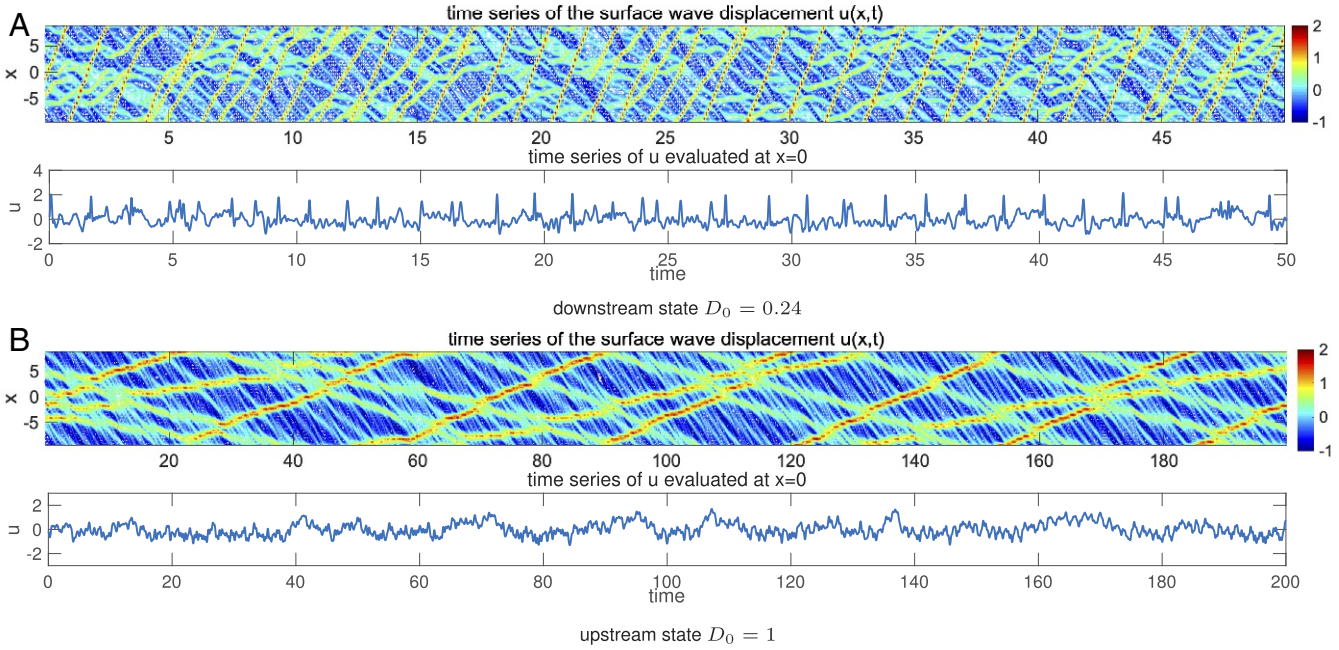

Fig. 3. Realization of the downstream (*A*) and upstream (*B*) flow solutions $u_\Lambda^\pm$. Note the larger vertical scale in the downstream time series plot.

- The downstream equilibrium measure μ_+ is homogeneous at each physical grid point so that the second and third moments are invariant at each grid point

$$\langle u_j^2 \rangle_{\mu_+} = \sigma^2 = \pi^{-1}, \quad \langle u_j^3 \rangle_{\mu_+} = \sigma^3 \kappa_3 = \pi^{-\frac{3}{2}} \kappa_3^+.$$

Then, the skewness of the downstream state variable u_Λ^+ after the ADC is given by the difference between the inflow and outflow wave slope energy of u_x:

$$\kappa_3^+ = \frac{3}{2} \pi^{\frac{1}{2}} L_0^{-\frac{3}{2}} E_0^{-\frac{1}{2}} D_+^2 \int_{-\pi}^{\pi} \left[\langle u_x^2 \rangle_{\mu_+} - \langle u_x^2 \rangle_{\mu_-} \right] dx + 3\pi^{\frac{1}{2}} \epsilon. \qquad [5]$$

The detailed derivation is shown in *SI Appendix*, *Section B.2*. In particular, the downstream skewness with near-Gaussian inflow statistics $\epsilon \ll 1$ is positive if and only if the difference of the incoming and outgoing wave slope energy is positive. This means that there is more small-scale wave slope energy in the outgoing state. As evidence, in Fig. 2, row 4, in all of the weak and strong skewness cases, the outflow energy spectrum always has a slower decay rate than the inflow energy spectrum, which possesses stronger energy in larger scales and weaker energy in the smaller scales.

In Fig. 1*C*, we compare the accuracy of the theoretical estimation [**5**] with numerical tests. In the regime with small incoming inverse temperature θ^-, the theoretical formula offers a quite accurate approximation of the third-order skewness using only information from the second-order moments of the wave-slope spectrum.

7. Key Features from Experiments Captured by the Statistical Dynamical Model

In this final section, we emphasize the crucial features generated by the statistical dynamical model [**1**] by making comparison with the experimental observations in ref. 20. As from the scale analysis displayed in *Section 2*, the theory is set in the same parameter regime as the experimental setup.

- The transition from near-Gaussian to skewed non-Gaussian distribution as well as the jump in both skewness and kurtosis observed in the experiment observations (figure 1 in ref. 20) can be characterized by the statistical model simulation results (Fig. 2, rows 1 and 2). Notice that the differences in the decay of third and fourth moments in the far end of the downstream regime from the experimental data are due to the dissipation effect in the flow from the wave absorbers that is not modeled in the statistical model here. The model simulation time series plotted in Fig. 3 can be compared with the observed time sequences from experiments (figure 1 in ref. 20). The downstream simulation generates waves with strong and frequent intermittency toward the positive displacement, while the upstream waves show symmetric displacements in two directions with, at most, small peaks in slow time. Even in the time series at a single location $x = 0$, the long-time variation displays similar structures.
- The downstream PDFs in experimental data are estimated with a gamma distribution in figure 2 in ref. 20. Here, in the same way, we can fit the highly skewed outgoing flow PDFs from the numerical results with the gamma distribution

$$\rho(u; k, \alpha) = \frac{e^{-k} \alpha^{-1}}{\Gamma(k)} \left(k + \alpha^{-1} u\right)^{k-1} e^{-\alpha^{-1} u}.$$

- The parameters (k, α) in the gamma distribution are fitted according to the measured statistics in skewness and variance: that is, $\sigma^2 = k\alpha^2, \kappa_3 = 2/\sqrt{k}$. Additionally, the excess kurtosis of the gamma distribution can be recovered as $\kappa_4 = 6/k$. As shown in Fig. 2, row 3, excellent agreement in the PDFs with the gamma distributions is reached in consistency with the experimental data observations. The accuracy with this approximation increases as the initial inverse temperature θ^- increases in value to generate more skewed distribution functions.
- Experimental measurements of the power spectra (figure 4 in ref. 20) reveal the downstream measurements to contain more energy at small scales (i.e., a relatively slower decay rate of the spectrum). This result is also observed in the direct numerical

simulations here (detailed results are shown in *SI Appendix*, *Section C.2*), as the outgoing state contains more energetic high frequencies.

8. Concluding Discussion

We have developed a statistical dynamical model to explain and predict extreme events and anomalous features of shallow water waves crossing an ADC. The theory is based on the dynamical modeling strategy consisting of the TKdV equation matched at the ADC with conservation of energy and Hamiltonian. Predictions can be made of the extreme events and anomalous features by matching incoming and outgoing statistical Gibbs measures before and after the abrupt depth transition. The statistical matching of the known nearly Gaussian incoming Gibbs state completely determines the predicted anomalous outgoing Gibbs state, which can be calculated by a simple sampling algorithm verified by direct numerical simulations, and successfully captures key features of the experiment. An analytic formula for the anomalous outgoing skewness is also derived. The strategy here should be useful for predicting extreme statistical events in other dispersive media in different settings.

ACKNOWLEDGMENTS. The research of A.J.M. is partially supported by Office of Naval Research Grant MURI N00014-16-1-2161, and D.Q. is supported as a postdoctoral fellow on the grant. M.N.J.M. acknowledges support from Simons Grant 524259.

1. Majda AJ, Branicki M (2012) Lessons in uncertainty quantification for turbulent dynamical systems. *Discrete Contin Dyn Syst A* 32:3133–3221.
2. Mohamad MA, Sapsis TP (2018) A sequential sampling strategy for extreme event statistics in nonlinear dynamical systems. *Proc Natl Acad Sci USA* 115:11138–11143.
3. Qi D, Majda AJ (2016) Predicting fat-tailed intermittent probability distributions in passive scalar turbulence with imperfect models through empirical information theory. *Commun Math Sci* 14:1687–1722.
4. Majda AJ, Tong XT (2015) Intermittency in turbulent diffusion models with a mean gradient. *Nonlinearity* 28:4171–4208.
5. Majda AJ, Chen N (2018) Model error, information barriers, state estimation and prediction in complex multiscale systems. *Entropy* 20:644.
6. Majda AJ, Tong XT (2018) Simple nonlinear models with rigorous extreme events and heavy tails. arXiv:1805.05615.
7. Thual S, Majda AJ, Chen N, Stechmann SN (2016) Simple stochastic model for el niño with westerly wind bursts. *Proc Natl Acad Sci USA* 113:10245–10250.
8. Chen N, Majda AJ (2018) Efficient statistically accurate algorithms for the Fokker–Planck equation in large dimensions. *J Comput Phys* 354:242–268.
9. Chen N, Majda AJ (2017) Beating the curse of dimension with accurate statistics for the Fokker–Planck equation in complex turbulent systems. *Proc Natl Acad Sci USA* 114:12864–12869.
10. Chen N, Majda AJ (2018) Conditional Gaussian systems for multiscale nonlinear stochastic systems: Prediction, state estimation and uncertainty quantification. *Entropy* 20:509.
11. Qi D, Majda AJ (2018) Predicting extreme events for passive scalar turbulence in two-layer baroclinic flows through reduced-order stochastic models. *Commun Math Sci* 16:17–51.
12. Adcock TA, Taylor PH (2014) The physics of anomalous ('rogue') ocean waves. *Rep Prog Phys* 77:105901.
13. Cousins W, Sapsis TP (2015) Unsteady evolution of localized unidirectional deep-water wave groups. *Phys Rev E* 91:063204.
14. Farazmand M, Sapsis TP (2017) Reduced-order prediction of rogue waves in two-dimensional deep-water waves. *J Comput Phys* 340:418–434.
15. Onorato M, Osborne AR, Serio M, Bertone S (2001) Freak waves in random oceanic sea states. *Phys Rev Lett* 86:5831–5834.
16. Dematteis G, Grafke T, Vanden-Eijnden E (2018) Rogue waves and large deviations in deep sea. *Proc Natl Acad Sci USA* 115:855–860.
17. Sergeeva A, Pelinovsky E, Talipova T (2011) Nonlinear random wave field in shallow water: Variable Korteweg-de Vries framework. *Nat Hazards Earth Syst Sci* 11:323–330.
18. Trulsen K, Zeng H, Gramstad O (2012) Laboratory evidence of freak waves provoked by non-uniform bathymetry. *Phys Fluids* 24:097101.
19. Viotti C, Dias F (2014) Extreme waves induced by strong depth transitions: Fully nonlinear results. *Phys Fluids* 26:051705.
20. Bolles CT, Speer K, Moore MNJ (2019) Anomalous wave statistics induced by abrupt depth change. *Phys Rev Fluids* 4:011801.
21. Solli DR, Ropers C, Koonath P, Jalali B (2007) Optical rogue waves. *Nature* 450:1054–1057.
22. Höhmann R, Kuhl U, Stöckmann H-J, Kaplan L, Heller EJ (2010) Freak waves in the linear regime: A microwave study. *Phys Rev Lett* 104:093901.
23. Johnson RS (1997) *A Modern Introduction to the Mathematical Theory of Water Waves* (Cambridge Univ Press, Cambridge, UK), Vol 19.
24. Rafail VA, Gregor K, Majda AJ (2003) Hamiltonian structure and statistically relevant conserved quantities for the truncated Burgers-Hopf equation. *Commun Pure Appl Math A J Courant Inst Math Sci* 56:1–46.
25. Majda A, Wang X (2006) *Nonlinear Dynamics and Statistical Theories for Basic Geophysical Flows* (Cambridge Univ Press, Cambridge, UK).
26. Bajars J, Frank JE, Leimkuhler BJ (2013) Weakly coupled heat bath models for Gibbs-like invariant states in nonlinear wave equations. *Nonlinearity* 26:1945–1973.
27. McLachlan R (1993) Symplectic integration of Hamiltonian wave equations. *Numerische Mathematik* 66:465–492.