

A Statistical Dynamical Model to Predict Extreme Events and Anomalous Features in Shallow Water Waves with Abrupt Depth Change

Andrew J. Majda^{a,1}, M. N. J. Moore^b, and Di Qi^{a,1}

^aDepartment of Mathematics and Center for Atmosphere and Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, NY 10012;

^bDepartment of Mathematics and Geophysical Fluid Dynamics Institute, Florida State University, Tallahassee, FL

This manuscript was compiled on November 26, 2018

Understanding and predicting extreme events and their anomalous statistics in complex nonlinear systems is a grand challenge in climate, material, and neuroscience, as well as for engineering design. Recent controlled laboratory experiments in weakly turbulent shallow water with abrupt depth change (ADC) exhibit a remarkable transition from nearly Gaussian statistics in incoming wave trains before the ADC to outgoing waves trains after the ADC with extreme anomalous statistics with large positive skewness of the surface height. 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 which are statistically matched at the ADC. 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 in different settings.

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

Understanding and predicting extreme events and their anomalous statistics in complex nonlinear systems is 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 statistical Gibbs invariant measures which are statistically matched at the ADC. 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 MCMC 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).

If you like, we could cite papers on ‘Optical rogue waves’ and/or rogue waves in microwave systems here. (references added)

1. Experiments showing anomalous wave statistics induced by an abrupt depth change

Controlled laboratory experiments were carried out in (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,

Significance Statement

Understanding and predicting extreme events and their anomalous statistics in complex nonlinear systems is 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. The 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.

A.J.M. designed the research. A.J.M., M.N.J.M., and D.Q. performed the research. A.J.M., M.N.J.M., and D.Q. wrote the paper.

The authors declare no conflict of interest.

¹To whom correspondence should be addressed. E-mail: qidi@cims.nyu.edu, jonjon@cims.nyu.edu

125 narrow wave tank. Midway through, the waves encounter a
 126 step in the bottom topography, and thus abruptly transition
 127 from one depth to another. The paddle wheel is forced with
 128 angle
 129

$$130 \quad \theta(t) = \theta_0 + \Delta\theta \sum_{n=1}^N a_n \cos(\omega_n t + \delta_n), \quad E \sim (\Delta\theta)^2 \sum_n a_n^2 \omega_n^2.$$

133 where the weights a_n are Gaussian in spectral space. The
 134 peak frequency ω_p gives rise to a characteristic wavelength λ_c
 135 of water waves, which can be estimated using the dispersion
 136 relation. Above, the total energy E in the system is determined
 137 by the angle amplitude $\Delta\theta$. Optical measurements of the free
 138 surface reveal a number of surprising statistical features:
 139

- 140 • Distinct statistics are found between the incoming and
 141 outgoing wave disturbances: incoming waves display near-
 142 Gaussian statistics, while outgoing waves skew strongly
 143 towards positive displacement.
- 144 • The degree of non-Gaussianity present in the outgoing
 145 wave field depends on the total energy of the system:
 146 larger driving amplitudes $\Delta\theta$ generate stronger skewness
 147 in the surface displacement PDFs.
- 148 • **The waves also show different slopes in the power spectra
 149 at upstream and downstream locations.**

152 I am not sure we need to include this last item. It is not
 153 really an experimental observation, simply a calculation
 154 that can be done using the dispersion relation and the
 155 values of the two depths. Could we simply take this last
 156 item out? Would you prefer to replace it with another
 157 observation? I could easily come up with one. Perhaps
 158 a comment on the decay rate of the power spectra. (I
 159 will confirm with Andy whether he would just delete
 160 this item or keep something else like the one above)

163 **2. Surface wave turbulence modeled by truncated KdV 164 equation with depth dependence**

165 The surface wave turbulence is modeled by a one-dimensional
 166 deterministic dynamical model.

167 This first sentence seems awkward to me.

168 The Korteweg-de Vries (KdV) equation (23) is a leading-
 169 order approximation of the surface waves that are determined
 170 by the balance of nonlinear and dispersive effects in an ap-
 171 propriate far-field limit. Here, the KdV equation is truncated
 172 in the first Λ modes (with $J = 2\Lambda + 1$ grid points) to gen-
 173 erate weakly turbulent dynamics. The surface displacement
 174 is described by the state variable $u_\Lambda^\pm(x, t)$ with superscript
 175 ‘–’ for the incoming waves and ‘+’ for the outgoing waves.
 176 The Galerkin truncated variable $u_\Lambda = \sum_{1 \leq |k| \leq \Lambda} \hat{u}_k(t) e^{ikx}$
 177 is normalized with zero mean $\hat{u}_0 = 0$ and unit energy
 178 $2\pi \sum_{k=1}^\Lambda |\hat{u}_k|^2 = 1$, which are conserved quantities, and
 179 $u_\Lambda \in \mathcal{P}_\Lambda u$ denotes the subspace projection. The evolution of
 180 u_Λ^\pm is governed by the truncated KdV equation with depth
 181 change D_\pm

$$182 \quad \frac{\partial u_\Lambda^\pm}{\partial t} + \frac{D_\pm^{-3/2}}{2} E_0^{1/2} L_0^{-3/2} \frac{\partial}{\partial x} \mathcal{P}_\Lambda (u_\Lambda^\pm)^2 + D_\pm^{1/2} L_0^{-3} \frac{\partial^3 u_\Lambda^\pm}{\partial x^3} = 0, \quad [1]$$

187 on the normalized periodic domain $x \in [-\pi, \pi]$. The conserved
 188 Hamiltonian can be decomposed into a cubic and a quadratic
 189 term

$$190 \quad \mathcal{H}_\Lambda^\pm = D_\pm^{-3/2} E_0^{1/2} L_0^{-3/2} H_3(u_\Lambda^\pm) - D_\pm^{1/2} L_0^{-3} H_2(u_\Lambda^\pm),$$

$$191 \quad H_3(u) = \frac{1}{6} \int_{-\pi}^{\pi} u^3 dx, \quad H_2(u) = \frac{1}{2} \int_{-\pi}^{\pi} \left(\frac{\partial u}{\partial x} \right)^2 dx.$$

192 Equation [1] is non-dimensionalized in the periodic domain.
 193 The depth is assumed to be unit $D_- = 1$ before the ADC and
 194 becomes $D_+ < 1$ after the ADC. We introduce parameters
 195 (E_0, L_0, Λ) based on the following assumptions:

- 196 • The wavenumber truncation Λ is fixed in a moderate
 197 value for generating weakly turbulent dynamics;
- 198 • The state variable u_Λ^\pm is normalized with zero mean and
 199 unit energy, $\mathcal{M}(u_\Lambda) = 0, \mathcal{E}(u_\Lambda) = 1$, which are conserved
 200 during evolution. Meanwhile, E_0 characterizes the total
 201 energy injected into the system based on the paddle
 202 amplitude $(\Delta\theta)^2$;
- 203 • The length scale of the system is denoted L_0 . The value
 204 is chosen so that the resolved scale $\Delta x = 2\pi L_0/J$ is
 205 comparable to the the characteristic wave length λ_c in
 206 the experiments.

207 Some intuition for how Eq. (1) produces different dynamics
 208 on either side of the ADC can be gained by considering the
 209 balance of the cubic and quadratic terms in the Hamiltonian
 210 \mathcal{H}_Λ^\pm . The depth change, $D_+ < 1$, increases the weight of
 211 H_3 and decreases that of H_2 , thus promoting the effects of
 212 nonlinearity and reducing dispersion. Since $\frac{\partial u}{\partial x}$ is the slope of
 213 the wave height, $H_2(u)$ measures the wave slope energy.

214 A *deterministic matching condition* is applied to the surface
 215 displacement u_Λ^\pm to link the incoming and outgoing wave trains.
 216 Assuming the abrupt depth change is met at $t = T_{\text{ADC}}$, the
 217 matching condition is given by

$$218 \quad u_\Lambda^-(x, t)|_{t=T_{\text{ADC}}} = u_\Lambda^+(x, t)|_{t=T_{\text{ADC}}} +,$$

219 Equation [1] is not designed to capture the short scale changes
 220 in rapid time. Rather, since we are interested in modeling
 221 statistics before and after the depth change, we will examine
 222 the long-time dynamics of large-scale structures.

223 I modified the last sentence above. Feel free to revert to
 224 your original sentence if you do not like mine. (the change
 225 looks good)

226 **Interpreting experimental parameters in the dynamical model.**
 227 The model parameters (E_0, L_0, Λ) in [1] can be directly linked
 228 to the basic scales from the physical problem. The important
 229 characterizing parameters measured from the experiments
 230 include: $\epsilon = \frac{a}{H_0}$ the wave amplitude a to water depth H_0
 231 ratio; $\delta = \frac{H_0}{\lambda_c}$ the water depth to wavelength scale λ_c ratio;
 232 and $D_0 = \frac{d}{H_0}$ the normalized wave depth ratio with incoming
 233 flow depth $d = H_0$ to the outgoing flow depth $d < H_0$. The
 234 interpretations and reference values of these model parameters
 235 are based on the experimental setup (20). By comparing the
 236 characteristic physical scales, the normalized TKdV equation
 237 [1] can be linked directly with the measured non-dimensional
 238 quantities by

$$239 \quad L_0 = 6^{\frac{1}{3}} \left(M \epsilon^{\frac{1}{2}} \delta^{-1} \right), \quad E_0 = \frac{27}{2} \gamma^{-2} \left(M \epsilon^{\frac{1}{2}} \delta^{-1} \right), \quad [2]$$

249 where M defines the computational domain size $M\lambda_c$ as M -
 250 multiple of the characteristic wavelength λ_c , and $\gamma = \frac{U}{a}$ rep-
 251 resents the factor to normalize the total energy in the state
 252 variable u_Λ to one.

253 Is there a difference between L_0 and L_d ? typo? (These two
 254 are actually different. L_d is the ‘true’ length scale defined in
 255 SI eqn. [10]. I guess here it’s better to delete this notation
 256 to avoid confusion.)

257 Consider the spatial discretization $J = 2\Lambda + 1$ so that the
 258 smallest resolved scale is comparable with the characteristic
 259 wavelength

$$261 \quad \Delta x = \frac{2\pi M \lambda_c}{J} \lesssim \lambda_c \Rightarrow M = \frac{J}{2\pi} \sim 5, J = 32.$$

262 Therefore in the practical numerical simulations, we pick
 263 $M = 5$ and γ varies in the range $[0.5, 1]$. Using the reference
 264 experimental measurements (20), $\epsilon \in [0.0024, 0.024]$, $\delta \sim 0.22$,
 265 and D_0 changes from 1 to 0.24 before and after the depth
 266 change. The reference values for the model scales can be esti-
 267 mated in the range $L_0 \in [2, 6]$ and $E_0 \in [50, 200]$. These are
 268 the values we will test in the direct numerical simulations. See
 269 details about the derivation from scale analysis in *SI Appendix*,
 270 A.

271 3. Equilibrium statistical mechanics for generating the 272 stationary invariant measure

273 In this section title, did you mean ‘statistical mechanism’ or
 274 ‘statistical mechanics’? (typos corrected)

275 Since the TKdV equation satisfies the Liouville property,
 276 the equilibrium invariant measure can be described by an equi-
 277 librium statistical formulism (24–26) using a Gibbs measure
 278 with the conserved energy \mathcal{E}_Λ and Hamiltonian \mathcal{H}_Λ . The equi-
 279 librium invariant measure is dictated by the conservation laws
 280 in the TKdV equation. In the case with fixed total energy E_0 ,
 281 this is the *mixed Gibbs measure* in the truncated model with
 282 microcanonical energy and canonical Hamiltonian ensembles
 283 (24)

$$284 \quad \mathcal{G}_\theta^\pm(u_\Lambda^\pm; E_0) = C_\theta^\pm \exp(-\theta^\pm \mathcal{H}(u_\Lambda^\pm)) \delta(\mathcal{E}(u_\Lambda^\pm) - E_0), \quad [3]$$

285 with θ representing the “inverse temperature”. The distinct
 286 statistics in the upstream and downstream waves can be con-
 287 trolled by the value of θ . Negative temperature, $\theta^\pm < 0$, is
 288 the appropriate regime to predict the experiments as shown
 289 below. In the incoming flow field, the inverse temperature θ^-
 290 is chosen so that \mathcal{G}_θ^- has Gaussian statistics. Using the above
 291 invariant measures [3], the expectation of any functional $F(u)$
 292 can be computed based on the Gibbs measure

$$293 \quad \langle F \rangle_{\mathcal{G}_\theta} \equiv \int F(u) \mathcal{G}_\theta(u) du.$$

294 The value of θ in the invariant measure is specified from $\langle H_\Lambda \rangle_{\mathcal{G}_\theta}$
 295 (24, 26). The invariant measure also predicts an equilibrium
 296 energy spectrum without running the TKdV equation directly.
 297 On the other hand, the time autocorrelation and transient
 298 statistics about the state variable u_Λ cannot be recovered from
 299 the statistical theory.

300 **Statistical matching condition in invariant measures before
 301 and after the abrupt depth change.** The Gibbs measures \mathcal{G}_θ^\pm
 302 are defined based on the different inverse temperatures θ^\pm on
 303 the two sides of the solutions

$$304 \quad \mu_t^-(u_\Lambda^-; D_-), \quad u_\Lambda |_{t=T_{\text{ADC}}^-} = u_0, \quad t < T_{\text{ADC}};$$

$$305 \quad \mu_t^+(u_\Lambda^+; D_+), \quad u_\Lambda |_{t=T_{\text{ADC}}^+} = u_0, \quad t > T_{\text{ADC}},$$

306 where u_0 represents the deterministic matching condition be-
 307 tween the incoming and outgoing waves. The two distributions,
 308 μ_t^-, μ_t^+ should also be matched at the depth change location
 309 T_{ADC} , so that,

$$310 \quad \mu_{t=T_{\text{ADC}}}^-(u_\Lambda) = \mu_{t=T_{\text{ADC}}}^+(u_\Lambda).$$

311 I do not believe we have defined μ_∞^- anywhere. (removed to
 312 avoid confusion)

313 In matching the flow statistics before and after the abrupt
 314 depth change, we first use the conservation of the determinis-
 315 tic Hamiltonian H_Λ^+ after the depth change. Then assuming
 316 ergodicity (24, 25), the statistical expectation for the Hamil-
 317 tonian $\langle H_\Lambda^+ \rangle$ is conserved in time after the depth change at
 318 $t = T_{\text{ADC}}$ and should stay in the same value as the system ap-
 319 proaches equilibrium as $t \rightarrow \infty$. The final statistical matching
 320 condition to get the outgoing flow statistics with parameter
 321 θ^+ can be found by

$$322 \quad \langle H_\Lambda^+ \rangle_{\mathcal{G}_\theta^+} = \langle H_\Lambda^+ \rangle_{\mathcal{G}_\theta^-}, \quad [4]$$

323 with the outgoing flow Hamiltonian H_Λ^+ and the Gibbs mea-
 324 sures \mathcal{G}_θ^\pm before and after the abrupt depth change.

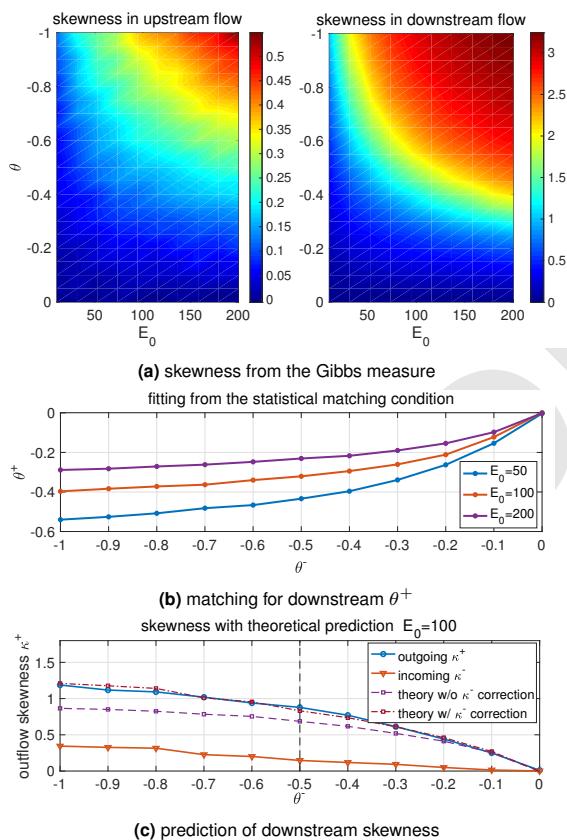
325 4. The nearly Gaussian incoming statistical state

326 The incoming flow is always characterized by a near-Gaussian
 327 distribution in the wave displacement. It is found that a
 328 physically consistent Gibbs measure should take negative val-
 329 ues in the inverse temperature parameter $\theta < 0$, where a
 330 proper distribution function and a decaying energy spectrum
 331 are generated (see (26) and *SI Appendix*, B.1 for the explicit
 332 simulation results). The upstream Gibbs measure \mathcal{G}_θ^- with
 333 $D_- = 1$ displays a wide parameter regime in (θ^-, E_0) with
 334 near-Gaussian statistics. In the left panel of Figure 1 (a), the
 335 inflow skewness κ_3^- varies only slightly with changing values of
 336 E_0 and θ^- . The incoming flow PDF then can be determined
 337 by picking the proper parameter value θ^- in the near Gaus-
 338 sian regime with small skewness. In contrast, the downstream
 339 Gibbs measure \mathcal{G}_θ^+ with $D_+ = 0.24$ shown in the right panel
 340 of Figure 1 (a) generates much larger skewness κ_3^+ as the
 341 absolute value of θ^+ and the total energy level E_0 increases.
 342 The solid lines in Figure 1 (c) offer a further confirmation of
 343 the transition from near-Gaussian statistics with small κ_3^- to
 344 a strongly skewed distribution κ_3^+ after the depth change.

345 In the next step, the value of the downstream θ^+ is deter-
 346 mined based on the matching condition [4]. The expectation
 347 $\langle H_\Lambda^+ \rangle_{\mathcal{G}_\theta^-}$ about the incoming flow Gibbs measure can be cal-
 348 culated according to the predetermined parameter values of
 349 θ^- as well as E_0 from the previous step. For the direct nu-
 350 merical experiments shown later in Figure 2, we pick proper
 351 choices of test parameter values as $L_0 = 6, E_0 = 100$ and
 352 $\theta^- = -0.1, -0.3, -0.5$. More test cases with different system

373 energy E_0 can be found in *SI Appendix, B.2* where similar
 374 transition from near Gaussian symmetric PDFs to skewed
 375 PDFs in the flow state u_Λ^\pm can always be observed.

376
 377
 378 **Direct numerical model simulations.** Besides the prediction of
 379 equilibrium statistical measures from the equilibrium statisti-
 380 cal approach, another way to predict the downstream model
 381 statistics is through running the dynamical model [1] directly.
 382 The TKdV equation is found to be ergodic with proper mixing
 383 property as measured by the decay of autocorrelations as long
 384 as the system starts from a negative inverse temperature state
 385 as described before. For direct numerical simulations of the
 386 TKdV equations, a proper symplectic integrator is required to
 387 guarantee the Hamiltonian and energy are conserved in time.
 388 It is crucial to use the symplectic scheme to guarantee the
 389 exact conservation of the energy and Hamiltonian since they
 390 are playing the central role in generating the invariant measure
 391 and the statistical matching. The symplectic schemes used
 392 here for the time integration of the equation is the 4th-order
 393 midpoint method (27). Details about the mixing properties
 394 from different initial states and the numerical algorithm are
 395 described in *SI Appendix, C*.



427 **Fig. 1.** First row: skewness from the Gibbs measures in incoming and outgoing flow
 428 states with different values of total energy E_0 and inverse temperature θ (notice
 429 the different scales in the incoming and outgoing flows); Second row: outgoing flow
 430 parameter θ^+ as a function of the incoming flow θ^- computed from the statistical
 431 matching condition with three energy level E_0 ; Last row: skewness in the outgoing
 432 flow with the matched value of θ^+ as a function of the inflow parameter θ^- (the
 433 theoretical predictions using [5] are compared).

5. Predicting extreme anomalous behavior after the ADC by statistical matching

435
 436
 437 With the inflow statistics well described and the numerical
 438 scheme set up, we are able to predict the downstream anom-
 439 alous statistics starting from the near-Gaussian incoming flow
 440 going through the abrupt depth change from $D_- = 1$ to
 441 $D_+ = 0.24$. First, we consider the statistical prediction in the
 442 downstream equilibrium measure directly from the matching
 443 condition. The downstream parameter value θ^+ is determined
 444 by solving the nonlinear equation [4] as a function of θ^+ ,
 445 $F(\theta^+) = \langle H_\Lambda^+ \rangle_{g_\theta^+}(\theta^+) - \langle H_\Lambda^+ \rangle_{g_\theta^-} = 0$. In the numerical
 446 approach, we adopt a modified secant method avoiding the
 447 stiffness in the parameter regime (see the *SI Appendix, B.2*
 448 for the algorithm). The fitted solution is plotted in Figure
 449 1 (b) as a function of the proposed inflow θ^- . A nonlinear
 450 $\theta^- - \theta^+$ relation is discovered from the matching condition. The
 451 downstream inverse temperature θ^+ will finally saturate at
 452 some level. The corresponding downstream skewness of the
 453 wave displacement u_Λ predicted from the statistical matching
 454 of Gibbs measures is plotted in Figure 1 (c). In general, a
 455 large positive skewness for outgoing flow κ_3^+ is predicted from
 456 the theory, while the incoming flow skewness κ_3^- is kept in
 457 a small value in a wide range of θ^- . Note that with $\theta^- \sim 0$
 458 (that is, using the microcanonical ensemble only with energy
 459 conservation), the outflow statistics are also near Gaussian
 460 with weak skewness. The skewness in the outflow statistics
 461 grows as the inflow parameter value θ^- increases in amplitude.

462
 463 For a second approach, we can use direct numerical sim-
 464 ulations starting from the initial state sampled from the incoming
 465 flow Gibbs measure \mathcal{G}_θ^- and check the transient changes in the
 466 model statistics. Figure 2 illustrates the change of statistics
 467 as the flow goes through the abrupt depth change. The first
 468 row plots the changes in the skewness and kurtosis for the
 469 state variable u_Λ after the depth change at $t = 0$. The PDFs
 470 in the incoming and outgoing flow states are compared with
 471 three different initial inverse temperatures θ^- . After the depth
 472 changes to $D_0 = 0.24$ abruptly at $t = 0$, both the skewness
 473 and kurtosis jump to a much larger value in a short time,
 474 implying the rapid transition to a highly skewed non-Gaussian
 475 statistical regime after the depth change. Further from Figure
 476 2, different initial skewness (but all relatively small) is set
 477 due to the various values of θ^- . With small $\theta^- = -0.1$, the
 478 change in the skewness is not very obvious (see the second row
 479 of Figure 2 for the incoming and outgoing PDFs of u_Λ). In
 480 comparison, if the incoming flow starts from the initial par-
 481 ameter $\theta^- = -0.3$ and $\theta^- = -0.5$, much larger increase in the
 482 skewness is induced from the abrupt depth change. Further-
 483 more, in the detailed plots in the third row of Figure 2 for the
 484 downstream PDFs under logarithmic scale, fat tails towards
 485 the positive direction can be observed, which represent the
 486 extreme events in the downstream flow (see also Figure 3 for
 487 the time-series of u_Λ).

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

493 Again, do we mean statistical ‘mechanism’ or ‘mechanics’?
 494 (should be ‘mechanics’)

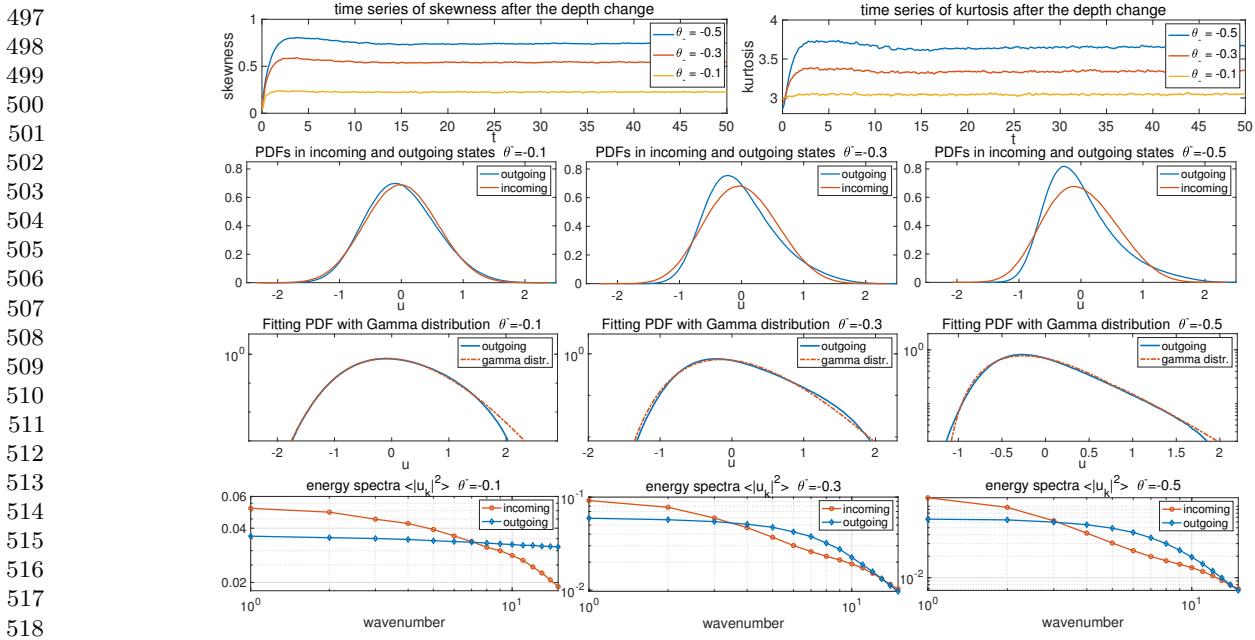


Fig. 2. Changes in the statistics of the flow state going through the abrupt depth change. The initial ensemble is set with the incoming flow Gibbs measure with different inverse temperature θ^- . First row: time evolution of the skewness and kurtosis. The abrupt depth change is taking place at $t = 0$; Second row: inflow and outflow PDFs of u_Λ ; Third row: the downstream PDFs fitted with Gamma distributions with consistent variance and skewness (in log coordinate in y); Last row: energy spectra in the incoming and outgoing flows.

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

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

Now we introduce mild assumptions on the distribution functions:

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

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

- 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

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

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

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

The detailed derivation is shown in [SI Appendix, 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 an evidence, in the last row of Figure 2 in all 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 Figure 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 (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 (Fig. 1 of (20)) can be characterized by the statistical model simulation results (see the first and second row of Figure 2). Notice that the difference in the decay of third and fourth moments in the far end of the downstream regime from the experimental data is 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 Figure 3 can be compared with

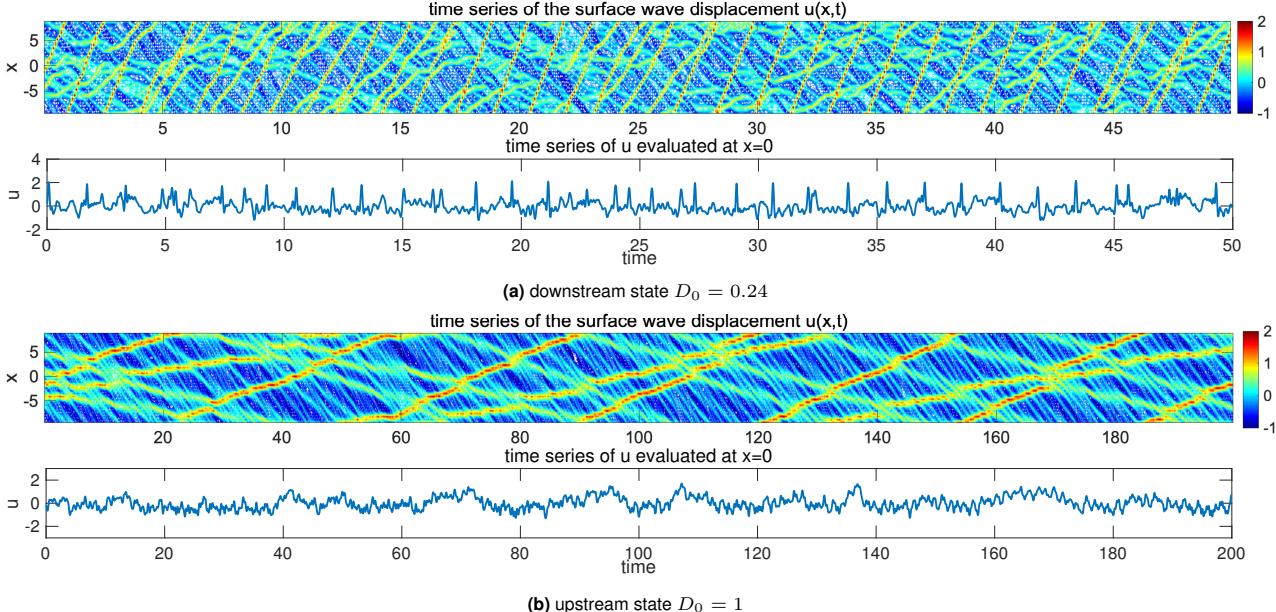


Fig. 3. Realization of the downstream and upstream flow solutions u_A^\pm . Note the larger vertical scale in the downstream time-series plot.

the observed time sequences from experiments (Fig. 1 of (20)). The downstream simulation generates waves with strong and frequent intermittency towards 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 Fig. 2 of (20). Here in the same way, we can fit the highly skewed outgoing flow PDFs from the numerical results with the Gamma distribution

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

The parameters (k, α) in the Gamma distribution are fitted according to the measured statistics in skewness and variance, that is, $\sigma^2 = k\alpha^2$, $\kappa_3 = 2/\sqrt{k}$. And the excess kurtosis of the Gamma distribution can be recovered as $\kappa_4 = 6/k$. As shown in the third row of Figure 2, 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.

- The experiments also have the up and down stream power spectra in time (Fig. 4 of (20)), which shows more energy at small time scales, i.e., a relatively slower decay rate in the downstream compared with the upstream case. This is also observed in the direct numerical simulations here (detailed results shown in *SI Appendix, C.2*). The

downstream state contains more energetic high frequencies. The peak frequency illustrates the characteristic time scale of the transporting wave trains along the water tank.

I do not understand the last sentence above. Why does the peak frequency illustrate the occurrence of transporting water waves?

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 abrupt depth change. The theory is based on the dynamical modeling strategy consisting of the TKdV equation matched at the abrupt depth change 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. This research of A. J. M. is partially supported by the Office of Naval Research through MURI N00014-16-1-2161. D. Q. is supported as a postdoctoral fellow on the second grant. M.N.J.M would like to acknowledge support from Simons grant 524259.

1. Majda AJ, Branicki M (2012) Lessons in uncertainty quantification for turbulent dynamical systems. *Discrete & Continuous Dynamical Systems-A* 32(9):3133–3221.
2. Mohamad MA, Sapsis TP (2018) A sequential sampling strategy for extreme event statistics in nonlinear dynamical systems. *Proceedings of the National Academy of Sciences* 115(44):11138–11143.
3. Qi D, Majda AJ (2016) Predicting fat-tailed intermittent probability distributions in passive scalar convection with imperfect models through empirical information theory. *Communications in Mathematical Sciences* 14(6):1687–1722.
4. Majda AJ, Tong XT (2015) Intermittency in turbulent diffusion models with a mean gradient. *Nonlinearity* 28(11):4171–4208.
5. Majda AJ, Chen N (2018) Model error, information barriers, state estimation and prediction in time series of the surface wave displacement $u(x,t)$
6. | www.pnas.org/cgi/doi/10.1073/pnas.XXXXXXXX
7. Farazmand M, Sapsis TP (2017) Reduced-order prediction of rogue waves in two-dimensional deep-water waves. *Journal of Computational Physics* 340:418–434.
8. Onorato M, Osborne AR, Serio M, Bertone S (2001) Freak waves in random oceanic sea states. *Physical Review Letters* 86(25):5831–5834.
9. Demattei G, Grafke T, Vanden-Eijnden E (2018) Rogue waves and large deviations in deep sea. *Proceedings of the National Academy of Sciences* 115(5):855–860.
10. Sergeeva A, Pelinovsky E, Talipova T (2011) Nonlinear random wave field in shallow water: variable Korteweg-de Vries framework. *Natural Hazards and Earth System Sciences* 11(2):323–330.
11. Trulsen K, Zeng H, Gramstad O (2012) Laboratory evidence of freak waves provoked by non-uniform bathymetry. *Physics of Fluids* 24(9):097101.