Rogue waves and large deviations in deep sea

Giovanni Dematteis[a,b], Tobias Grafke[a,c], and Eric Vanden-Eijnden[a,1]

[a]Courant Institute of Mathematical Sciences, New York University, New York, NY 10012; [b]Dipartimento di Scienze Matematiche, Politecnico di Torino, I-10129 Torino, Italy; and [c]Mathematics Institute, University of Warwick, Coventry CV4 7AL, United Kingdom

Edited by David A. Weitz, Harvard University, Cambridge, MA, and approved December 11, 2017 (received for review June 13, 2017)

The appearance of rogue waves in deep sea is investigated by using the modified nonlinear Schrödinger (MNLS) equation in one spatial dimension with random initial conditions that are assumed to be normally distributed, with a spectrum approximating realistic conditions of a unidirectional sea state. It is shown that one can use the incomplete information contained in this spectrum as prior and supplement this information with the MNLS dynamics to reliably estimate the probability distribution of the sea surface elevation far in the tail at later times. Our results indicate that rogue waves occur when the system hits unlikely pockets of wave configurations that trigger large disturbances of the surface height. The rogue wave precursors in these pockets are wave patterns of regular height, but with a very specific shape that is identified explicitly, thereby allowing for early detection. The method proposed here combines Monte Carlo sampling with tools from large deviations theory that reduce the calculation of the most likely rogue wave precursors to an optimization problem that can be solved efficiently. This approach is transferable to other problems in which the system's governing equations contain random initial conditions and/or parameters.

Laplace method | JONSWAP spectrum | peregrine soliton | intermittency | Monte Carlo

Rogue waves, long considered a figment of sailors' imaginations, are now recognized to be a real, and serious, threat for boats and naval structures (1, 2). Oceanographers define them as deep-water waves whose crest-to-trough height H exceeds twice the significant wave height H_s, which itself is four times the SD of the ocean surface elevation. Rogue waves appear suddenly and unpredictably and can lead to water walls with vertical size on the order of 20–30 m (3, 4), with enormous destructive power. Although rare, they tend to occur more frequently than predicted by linear Gaussian theory (5, 6). While the mechanisms underlying their appearance remain under debate (7–9), one plausible scenario has emerged over the years: It involves the phenomenon of modulational instability (10, 11), a nonlinear amplification mechanism by which many weakly interacting waves of regular size can create a much larger one. Such an instability arises in the context of the focusing nonlinear Schrödinger (NLS) equation (11–17) or its higher-order variants (18–22), which are known to be good models for the evolution of a unidirectional, narrow-banded surface wave field in a deep sea. Support for the description of rogue waves through such envelope equations recently came from experiments in water tanks (23–26), where Dysthe's modified NLS (MNLS) equation in one spatial dimension (18, 19) was shown to accurately describe the mechanism creating coherent structures which soak up energy from its surroundings. While these experiments and other theoretical works (27, 28) give grounds for the use of MNLS to describe rogue waves, they have not addressed the question of their likelihood of appearance. Some progress in this direction has been recently made in ref. 29, where a reduced model based on MNLS was used to estimate the probability of a given amplitude within a certain time, and thereby compute the tail of the surface height distribution. These calculations were done by using an ansatz for the solutions of MNLS, effectively making the problem 2D. The purpose of this work is to remove this approximation and study the problem in its full generality. Specifically, we consider the MNLS with random initial data drawn from a Gaussian distribution (30). The spectrum of this field is chosen to have a width comparable to that of the Joint North Sea Wave Project (JONSWAP) spectrum (31, 32) obtained from observations in the North Sea. We calculate the probability of occurrence of a large amplitude solution of MNLS out of these random initial data and thereby also estimate the tail of the surface height distribution. These calculations are performed within the framework of large deviations theory (LDT), which predicts the most likely way by which large disturbances arise and therefore also explains the mechanism of rogue wave creation. Our results are validated by comparison with brute-force Monte Carlo simulations, which indicate that rogue waves in MNLS are indeed within the realm of LDT. Our approach therefore gives an efficient way to assess the probability of large waves and their mechanism of creation.

1. Problem Setup

Our starting point will be the MNLS equation for the evolution of the complex envelope $u(t,x)$ of the sea surface in deep water (18), in terms of which the surface elevation reads $\eta(t,x) = \Re\big(u(t,x)e^{i(k_0x-\omega_0t)}\big)$ (here k_0 denotes the carrier wave number, $\omega_0=\sqrt{gk_0}$, and g is the gravitational acceleration). Measuring u and x in units of k_0^{-1} and t in ω_0^{-1}, we can write MNLS in nondimensional form as

$$\partial_t u + \frac{1}{2}\partial_x u + \frac{i}{8}\partial_x^2 u - \frac{1}{16}\partial_x^3 u + \frac{i}{2}|u|^2 u + \frac{3}{2}|u|^2\partial_x u + \frac{1}{4}u^2\partial_x\bar{u} - \frac{i}{2}|\partial_x|\,|u|^2 = 0, \quad x\in[0,L], \qquad [1]$$

where the bar denotes complex conjugation. We will consider Eq. **1** with random initial condition $u_0(x)\equiv u(0,x)$, constructed via their Fourier representation,

Significance

Quantifying the departure from Gaussianity of the wave-height distribution in the seas and thereby estimating the likelihood of appearance of rogue waves is a long-standing problem with important practical implications for boats and naval structures. Here, a procedure is introduced to identify ocean states that are precursors to rogue waves, which could permit their early detection. Our findings indicate that rogue waves obey a large deviation principle—i.e., they are dominated by single realizations—which our method calculates by solving an optimization problem. The method generalizes to estimate the probability of extreme events in other deterministic dynamical systems with random initial data and/or parameters, by using prior information about the nature of their statistics.

Author contributions: G.D., T.G., and E.V.-E. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. Email: eve2@cims.nyu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1710670115/-/DCSupplemental.

$$u_0(x) = \sum_{n\in\mathbb{Z}} e^{ik_n x}(2\hat{C}_n)^{1/2}\theta_n, \quad \hat{C}_n = Ae^{-k_n^2/(2\Delta^2)}, \quad \textbf{[2]}$$

where $k_n = 2\pi n/L$, θ_n are complex Gaussian variables with mean zero and covariance $\mathbb{E}\theta_n\bar{\theta}_m = \delta_{m,n}$, $\mathbb{E}\theta_n\theta_m = \mathbb{E}\bar{\theta}_n\bar{\theta}_m = 0$. This guarantees that $u_0(x)$ is a Gaussian field with mean zero and $\mathbb{E}(u_0(x)\bar{u}_0(x')) = 2\sum_{n\in\mathbb{Z}} e^{ik_n(x-x')}\hat{C}_n$. To make contact with the observational data, the amplitude A and the width Δ in Eq. **2** are picked so that $\hat{C}_n$ has the same height and area as the JONSWAP spectrum (31, 32); see *Supporting Information* for details.

Because the initial data for Eq. **1** are random, so is the solution at time $t > 0$, and our aim is to compute

$$P_T(z) \equiv \mathbb{P}\left(F(u(T)) \geq z\right), \quad \textbf{[3]}$$

where $\mathbb{P}$ denotes probability over the initial data and F is a scalar functional depending on u at time $T > 0$. Even though our method is applicable to more general observables, here we will focus on

$$F(u(T)) = \max_{x\in[0,L]} |u(T,x)|. \quad \textbf{[4]}$$

2. LDT Approach

A brute-force approach to calculate Eq. **3** is Monte Carlo sampling: Generate random initial conditions $u_0(x)$ by picking random θ_n's in Eq. **2**, evolve each of these $u_0(x)$ deterministically via Eq. **1** up to time $t = T$ to get $u(T,x)$, and count the proportion that fulfill $F(u(T)) \geq z$. While this method is simple, and will be used below as benchmark, it loses efficiency when z is large, which is precisely the regime of interest for the tails of the distribution of $F(u(T))$. In that regime, a more efficient approach is to rely on results from LDT which assert that Eq. **3** can be estimated by identifying the most likely initial condition that is consistent with $F(u(T)) \geq z$. To see how this result comes about, recall that the probability density of u_0 is formally proportional to $\exp(-\frac{1}{2}\|u_0\|_C^2)$, where $\|u_0\|_C^2$ is given by

$$\|u_0\|_C^2 = \sum_{n\in\mathbb{Z}} \frac{|\hat{a}_n|^2}{\hat{C}_n}, \quad \hat{a}_n = \frac{1}{L}\int_0^L e^{-ik_n x}u_0(x)dx. \quad \textbf{[5]}$$

To calculate Eq. **3** we should integrate this density over the set $\Omega(z) = \{u_0 : F(u(T,u_0)) \geq z\}$, which is hard to do in practice. Instead, we can estimate the integral by Laplace's method. As shown in *Materials and Methods*, this is justified for large z, when the probability of the set $\Omega(z)$ is dominated by a single $u_0(x)$ that contributes most to the integral and can be identified via the constrained minimization problem

$$\frac{1}{2}\min_{u_0\in\Omega(z)} \|u_0\|_C^2 \equiv I_T(z), \quad \textbf{[6]}$$

which then yields the following LDT estimate for Eq. **3**

$$P_T(z) \asymp \exp\left(-I_T(z)\right). \quad \textbf{[7]}$$

Here, $\asymp$ means that the ratio of the logarithms of both sides tends to 1 as $z \to \infty$. As discussed in *Materials and Methods*, a multiplication prefactor can be added to Eq. **7**, but it does not affect significantly the tail of $P_T(z)$ on a logarithmic scale.

In practice, the constraint $F(u(T,u_0)) \geq z$ can be imposed by adding a Lagrange multiplier term to Eq. **6**, and it is easier to use this multiplier as control parameter and simply see a posteriori what value of z it implies. That is to say, perform for various values of λ the minimization

$$\min_{u_0}\left(\frac{1}{2}\|u_0\|_C^2 - \lambda F(u(T,u_0))\right) \equiv S_T(\lambda), \quad \textbf{[8]}$$

over all u_0 of the form in Eq. **2** (no constraint), then observe that this implies the parametric representation

$$I_T(z(\lambda)) = \frac{1}{2}\|u_0^\star(\lambda)\|_C^2, \quad z(\lambda) = F(u(T,u_0^\star(\lambda))). \quad \textbf{[9]}$$

where $u_0^\star(\lambda)$ denotes the minimizer obtained in Eq. **8**. It is easy to see from Eqs. **6** and **8** that $S_T(\lambda)$ is the Legendre transform of $I_T(z)$ since:

$$S_T(\lambda) = \sup_{z\in\mathbb{R}}(\lambda z - I_T(z)) = \sup_{z\in\mathbb{R}}\left(\lambda z - \frac{1}{2}\inf_{u_0\in\Omega(z)}\|u_0\|_C^2\right), \quad \textbf{[10]}$$

3. Results

We considered two sets of parameters. In set 1, we took $A = 5.4\cdot 10^{-5}k_0^{-2}$ and $\Delta = 0.19k_0$. Converting back into dimensional units by using $k_0^{-1} = 36$ m consistent with the JONSWAP spectrum (31, 32), this implies a significant wave height $H_s = 4\sqrt{C(0)} = 3.3$ m classified as a "rough sea" (33). It also yields a Benjamin–Feir index (BFI) $= 2\sqrt{2C(0)}/\Delta = 0.34$, (32, 34), meaning that the modulational instability of a typical initial condition is of medium intensity. In set 2, we took $A = 3.4\cdot 10^{-4}k_0^{-2}$ and $\Delta = 0.19k_0$, for which $H_s = 8.2$ m is that of a "high sea" and the BFI is 0.85, meaning that the modulational instability of a typical initial condition is stronger.

Fig. 1, *Upper* shows the time evolution of $|u(t,x)|$ starting from an initial condition from set 1 optimized so that $\max_x|u(T,x)| = 8$ m at $T = 20$ min. For comparison, Fig. 1, *Lower* shows $|u(t,x)|$ for a typical initial condition drawn from its Gaussian distribution. To illustrate what is special about the initial conditions identified by our optimization procedure, in Fig. 2, we show snapshots of the surface elevation $\eta(t,x)$ at three different times, $t = 0, 10, 20$ min (black lines), using the constraint that $\max_x|u(T,x)| \geq 4.8$ m at $T = 20$ min. Additionally, we average all Monte Carlo samples achieving $\max_x|u(t,x)| \geq 4.8$ m, translated to the origin. Snapshots of this mean configuration are shown in Fig. 2 (blue lines). They agree well with those of the optimized solution (black lines). The one SD spread around the mean Monte Carlo realization (light blue) is reasonably small, especially around the rogue wave at final time. This indicates that the event $\max_x|u(T,x)| \geq 4.8$ m is indeed realized with probability close to 1 by starting from the most likely initial condition consistent with this event, as predicted by LDT. The

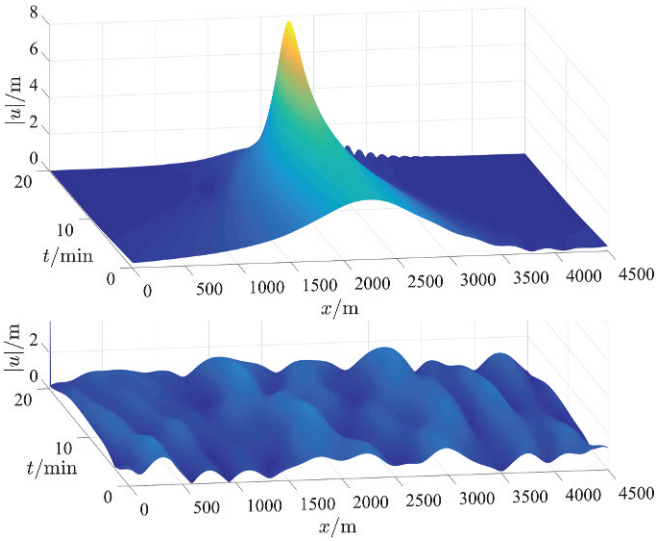

Fig. 1. (*Upper*) Time evolution of $|u(t,x)|$ from an initial condition optimized for $\max_x|u(T,x)| \geq 8$ m at $T = 20$ min. (*Lower*) Same for a typical Gaussian random initial condition.

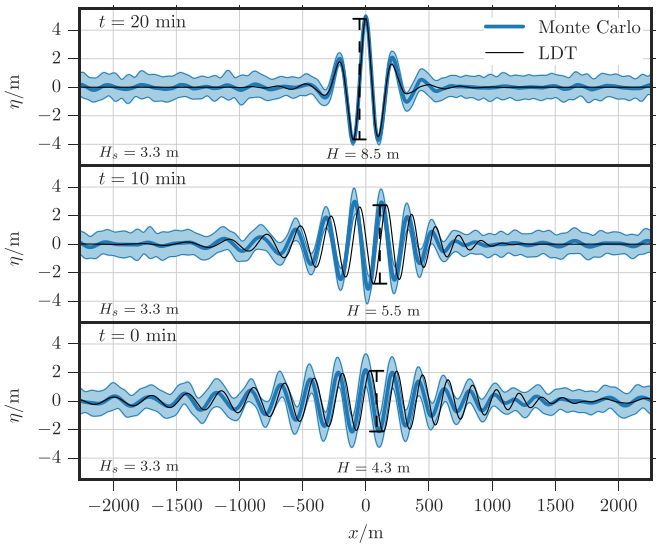

Fig. 2. Comparison between the average realization reaching $\max_x|u(T, x)| \geq 4.8$ m at $T = 20$ min (dark blue) and one SD around this mean (light blue), with the solution reaching the same amplitude starting from the maximum-likelihood initial condition (black) for $t = 0$, 10, 20 min.

usefulness of LDT is confirmed in Fig. 3, depicting the probabilities of $\max_x|u(T, x)|$ for both sets 1 and 2 calculated via LDT optimization (lines), compared with Monte Carlo sampling (dots). As can be seen, the agreement is remarkable, especially in the tail corresponding to the rogue wave regime. As expected, the Monte Carlo sampling becomes inaccurate in the tail, since there, the probabilities are dominated by unlikely events. The LDT calculation, in contrast, remains efficient and accurate far in the tail.

The probabilities plotted in Fig. 3 show several remarkable features. First, they indicate that, as T gets larger, their tails fatten significantly. For example, in set 1 $P_{T=20\,\mathrm{min}}(6\,\mathrm{m}) \approx 10^{-5}$, which is 5 orders of magnitude larger than initially, $P_{t=0\,\mathrm{min}}(6\,\mathrm{m}) \approx 10^{-10}$. Secondly, the probabilities converge to a limiting density for large T. This occurs after some decorrelation time $\tau_c \approx 10$ min in set 1 and $\tau_c \approx 3$ min in set 2. Similarly, the LDT results converge. In fact, this convergence can be observed at the level of the trajectories generated from the optimal $u_0^\star$. As Fig. 4 shows, reading these trajectories backward from $t = T$, their end portions coincide, regardless of whether $T = 20$ min, $T = 15$ min, or $T = 10$ min. The implications of these observations, in particular on the mechanism of creation of rogue waves and their probability of appearance within a time window, are discussed in *Interpretation*.

Before doing so, let us discuss the scalability of our results to larger domain sizes; see *Supporting Information* for more details. As shown above, the optimization procedure based on LDT predicts that the most likely way a rogue wave will occur in the domain is via the apparition of a single large peak in $|u(t, x)|$. In the setup considered before, this prediction is confirmed by the brute-force simulations using Monte Carlo sampling. It is clear, however, that for increased domain size—e.g., by taking a domain size of NL with $N \gg 1$—it will become increasingly likely to observe multiple peaks, for the simple reason that large waves can occur independently at multiple sufficiently separated locations. In these large domains, the large deviation predictions remain valid if we look at the maximum of $|u(t, x)|$ in observation windows that are not too large (that is, about the size of the domain L considered above). However, they deteriorate if we consider this maximum in the entire domain of size NL, in the sense that the value $\mathbb{P}\left(\max_{x\in[0,NL]}|u(t, x)| \geq z\right)$ predicted by LDT matches that from Monte Carlo sampling at values of z that are pushed further away in the tails. This is an entropic effect, which is easy to correct for: Events in different subwindows must be considered independent, and their probabilities superposed. That is, if we denote by

$$P_T^N(z) = \mathbb{P}\Big(\max_{x\in[0,NL]}|u(T, x)| \geq z\Big), \qquad [11]$$

it can be related to $P_T(z) = \mathbb{P}(\max_{x\in[0,L]}|u(T, x)| \geq z)$ via

$$P_T^N(z) = 1 - (1 - P_T(z))^N. \qquad [12]$$

This formula is derived in *Supporting Information* and shown to accurately explain the numerical results. For efficiency, L is chosen to be the smallest domain size for which boundary effects can be neglected, in the sense that the shape of the optimal trajectories does no longer change if L is increased further. In effect, this provides us with a method to scale up our results to arbitrary large observation windows.

4. Interpretation

The convergence of $P_T(z)$ toward a limiting function $P(z)$ has important consequences for the significance and interpretation

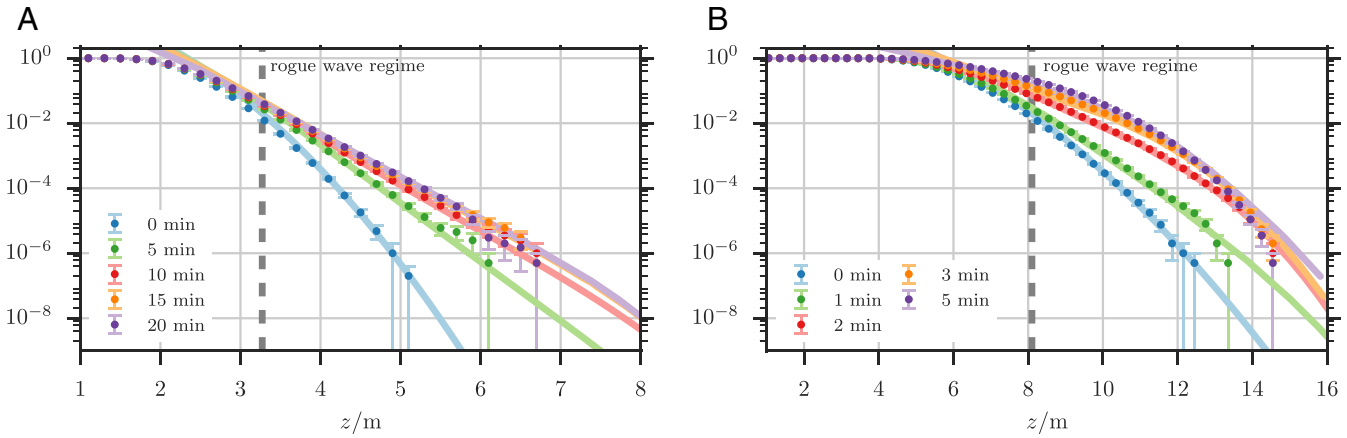

Fig. 3. Probabilities $P_T(z)$ of $\max_x|u(T, x)|$ for different times T for sets 1 (*A*) and 2 (*B*). The probabilities estimated by Monte Carlo sampling with 10^6 realizations (dots) are compared with those predicted by LDT (lines). These probabilities agree over ∼5 orders of magnitude in probability, although LDT allows for the calculation of the tail where Monte Carlo becomes prohibitively costly. The error bars in the Monte Carlo results represent the statistical error of 2 SD (95% confidence interval) for the Bernoulli distribution with parameter $P_T(z)$.

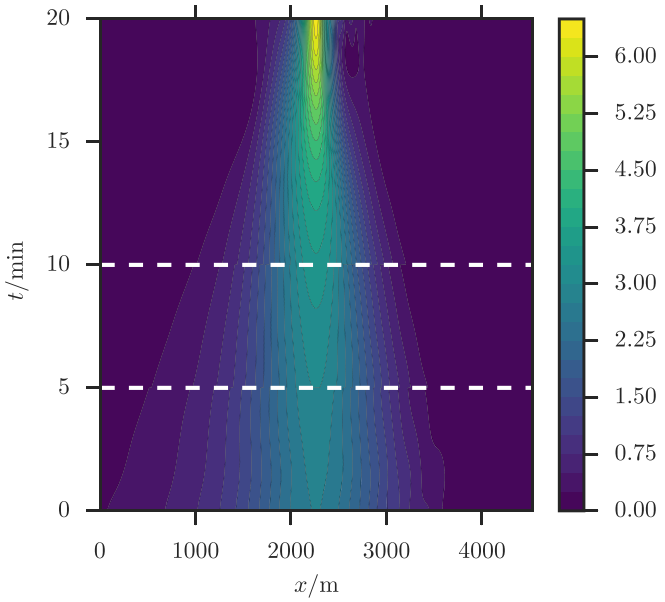

Fig. 4. Contour plot of the optimal trajectories from LDT for $T = 10$, 15, and 20 min in set 1. The trajectories, superposed to match at $t = T$, coincide, which is consistent with the convergence of the probabilities $P_T(z)$ for large T.

of our method and its results. Notice first that this convergence can be explained if we assume that the probability distribution of the solutions to Eq. **1** with Gaussian initial data converges to an invariant measure. In this case, for large T, the Monte Carlo simulations will sample the value of $\max_x |u|$ on this invariant measure, and the optimization procedure based on LDT will do the same. The timescale τ_c over which convergence occurs depends on how far this invariant measure is from the initial Gaussian measure of $u_0(x)$. Interestingly, the values we observe for τ_c are in rough agreement with the time scales predicted by the semiclassical limit of NLS that describes high-power pulse propagation (35, 36). As recalled in *Supporting Information*, this approach predicts that the timescale of apparition of a focusing solution starting from a large initial pulse of maximal amplitude U_i and length-scale L_i is $\tau_c = \sqrt{T_{\rm nl}\, T_{\rm lin}}$, where $T_{\rm nl} = \left(\frac{1}{2}\omega_0 k_0^2\, U_i^2\right)^{-1}$ is the nonlinear timescale for modulational instability and $T_{\rm lin} = 8\omega_0^{-1} k_0^2 L_i^2$ is the linear timescale associated with group dispersion. Setting $U_i = H_s$ (the size at the onset of rogue waves) and $L_i = \sqrt{2\pi}\Delta^{-1}$ (the correlation length of the initial field) gives $\tau_c \simeq 18$ min for set 1 and $\tau_c \simeq 8$ min for set 2, consistent with the convergence times of $P_T(z)$. This observation has implications in terms of the mechanism of apparition of rogue waves, in particular, their connection to the so-called Peregrine soliton, that has been invoked as prototype mechanism for rogue wave creation (5, 13, 37–40), in particular for water waves (24, 25, 41), plasmas (42), and fiber optics (36, 43, 44). This connection is discussed in *Supporting Information*.

Our findings also indicate that, even though the assumption that $u_0(x)$ is Gaussian is incorrect in the tail [that is, $P_{T=0}(z)$ is not equal to the limiting $P(z)$ in the tail], it contains the right seeds to estimate $P(z)$ via $P_T(z)$ if $T \gtrsim \tau_c$. (This convergence occurs on the timescale τ_c which is much smaller than the mixing time for the solutions of Eq. **1**, i.e., the time it would take from a given initial condition, rather than an ensemble thereof, to sample the invariant measure.) Altogether, this is consistent with the scenario put forward by Sapsis and coworkers in refs. 45 and 46 to explain how extreme events arise in intermittent dynamical systems and calculate their probability: They occur when the system hits small instability pockets which trigger a large transient excursion. In this scenario, as long as the initial probability distribution in these pockets is accurate, the dynamics will permit precise estimation of the distribution tail. In some sense, the distribution of the initial condition plays a role of the prior distribution in Bayesian inference, and the posterior can be effectively sampled by adding the additional information from the dynamics over short periods of time during which instabilities can occur. [Note in particular that the Gaussian field in Eq. **2** is the random field that maximizes entropy given the constraint on its covariance $C(x)$.] In ref. 45, this picture is made predictive by using a 2D ansatz for the initial condition $u_0(x)$ to avoid having to perform sampling in high dimension over the original $u_0(x)$. What our results show is that this approximation can be avoided all together by using LDT to perform the calculations directly with the full Gaussian initial condition in Eq. **2**.

Interestingly, we can use the results above to calculate the probability of occurrence of rogue waves in a given time window. More precisely, the probability $p(z, T_I)$ that a rogue wave of amplitude larger than z be observed in the domain $[0, L]$ during $[0, T_I]$ [i.e., that $\max_{t\in[0,T_I]} \max_{x\in[0,L]} |u(t,x)| \geq z$] can be estimated in terms of $P(z)$ and τ_c as

$$p \equiv \mathbb{P}\Big(\max_{t\in[0,T_I]} \max_{x\in[0,L]} |u(t,x)| \geq z\Big) \sim 1 - (1 - P(z))^{T_I/\tau_c}, \qquad [13]$$

where we use the fact that rogue waves can be considered independent on timescales larger than τ_c and assume $T_I \gg \tau_c$. The function p is plotted in Fig. 5 as a function of z and T_I. For example, for set 1, Eq. **13** indicates a 50% chance to observe a rogue wave of height $z = 4$ m (that is, ∼8 m from crest-to-trough) after 11 h [using $\tau_c = 10$ min and $P(z = 4\ \text{m}) = 1.1 \cdot 10^{-2}$]; if we wait 30 h, the chance goes up to 85%. Similarly, for set 2, the chance to observe a wave of 11 m height is ∼50% after 3 h and ∼85% after 8 h ($\tau_c = 3$ min and $P(z = 11\ \text{m}) = 1.2 \cdot 10^{-2}$).

5. Concluding Remarks

We have shown how an optimization problem building on LDT can be used to predict the pathway and likelihood of appearance of rogue waves in the solutions of MNLS fed by random initial data consistent with observations. This setup guarantees accuracy of the core of the initial distribution, which in turn permits the precise estimation of its tail via the dynamics. Our results give quantitative estimate for the probabilities of observing high-amplitude waves within a given time window. These results also show that rogue waves have very specific precursors, a feature that was already noted in ref. 47 in the context of a reduced model and could potentially be used for their early detection.

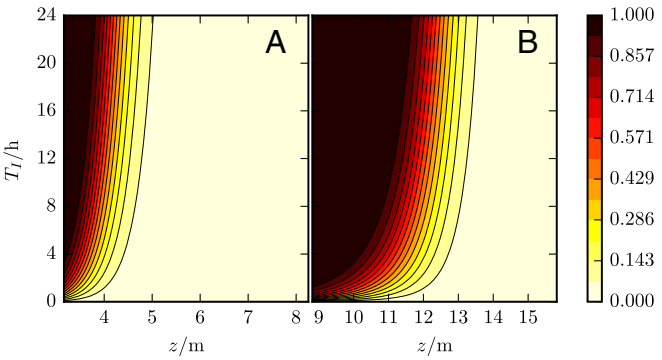

Fig. 5. Contour plots of the probability to observe a wave whose amplitude exceeds z in the time window $[0, T_I]$ for sets 1 (*A*) and 2 (*B*).

6. Materials and Methods

Laplace Method and Large Deviations. Here, we recall some standard large deviation results that rely on the evaluation of Gaussian integrals by Laplace's method and are at the core of the method we propose. It is convenient to rephrase the problem abstractly and consider the estimation of

$$P(z) = \mathbb{P}(\phi(\theta) > z), \tag{14}$$

where $\theta \in \mathbb{R}^D$ are Gaussian random variables with mean zero and covariance Id, and $\phi : \mathbb{R}^D \to \mathbb{R}$ is some real valued function—as long as we truncate the sum in Eq. **2** to a finite number of modes, $|n| \leq M$, the problem treated in this work can be cast in this way, with θ playing the role of $C^{-1/2}u_0$ and $\phi(\cdot)$ that of $F(u(T, C^{-1/2}\cdot))$. The probability $P(z)$ in Eq. **14** is given by

$$P(z) = (2\pi)^{-D/2} \int_{\Omega(z)} e^{-\frac{1}{2}|\theta|^2} d\theta, \tag{15}$$

where $\Omega(z) = \{\theta : \phi(\theta) > z\}$. The interesting case is when this set does not contain the origin, $0 \notin \Omega(z)$, which we will assume is true when $z > 0$. We also make two additional assumptions:

1. The point on the boundary $\partial\Omega(z)$ that is closest to the origin is isolated: Denoting this point as

$$\theta^\star(z) = \underset{\theta \in \partial\Omega(z)}{\operatorname{argmin}} |\theta|^2, \tag{16}$$

we assume that

$$\begin{aligned} &\frac{1}{2}|\theta^\star(z)|^2 \text{ is strictly increasing with } z \geq 0; \\ &\lim_{z\to\infty} \frac{1}{2}|\theta^\star(z)|^2 = \infty. \end{aligned} \tag{17}$$

2. The connected piece of $\partial\Omega(z)$ that contains $\theta^\star(z)$ is smooth with a curvature that is bounded by a constant independent of z.

The point $\theta^\star(z)$ satisfies the Euler–Lagrange equation for Eq. **16**, with the constraint incorporated via a Lagrange multiplier term:

$$\theta^\star(z) = \lambda \nabla\phi(\theta^\star(z)) \tag{18}$$

for some Lagrange multiplier λ. This implies that

$$\frac{\theta^\star(z)}{|\theta^\star(z)|} = \frac{\nabla\phi(\theta^\star(z))}{|\nabla\phi(\theta^\star(z))|} = \hat{n}(z). \tag{19}$$

where $\hat{n}(z)$ denotes the inward pointing unit vector normal to $\partial\Omega(z)$ at $\theta^\star(z)$. If we move inside the set $\Omega(z)$ from $\theta^\star(z)$ in the direction of $\hat{n}(z)$, the norm $|\theta|^2$ increases under the assumptions in Eq. **17**. Indeed, setting $\theta = \theta^\star(z) + \hat{n}(z)u$ with $u \geq 0$, we have

$$\begin{aligned} |\theta|^2 &= |\theta^\star(z)|^2 + 2\langle \hat{n}(z), \theta^\star(z)\rangle u + u^2 \\ &= |\theta^\star(z)|^2 + 2|\theta^\star(z)|z + z^2, \end{aligned} \tag{20}$$

where we use Eq. **19**. In fact, if we were to perform the integral in that direction, the natural variable of integration would be to rescale $u \to u/|\theta^\star(z)|$. In particular, if we were to replace $\Omega(z)$ by the half space $P(z) = \{\theta \,|\, \hat{n}(z) \cdot (\theta - \theta^*(z)) > 0\}$, it would be easy to estimate the integral in Eq. **15** by introducing a local coordinate system around $\theta^*(z)$, whose first coordinate is in the direction of $\hat{n}(z)$. Indeed, this would give:

$$\begin{aligned} &(2\pi)^{-D/2} \int_{P(z)} e^{-\frac{1}{2}|\theta|^2} d\theta \\ &= (2\pi)^{-D/2} \int_0^\infty e^{-\frac{1}{2}|\theta^\star(z)|^2 - |\theta^\star(z)|u - \frac{1}{2}u^2} du \int_{\mathbb{R}^{N-1}} e^{-\frac{1}{2}|\eta|^2} d\eta \\ &= (2\pi)^{-1/2} e^{-\frac{1}{2}|\theta^\star(z)|^2} \int_0^\infty e^{-|\theta^\star(z)|u - \frac{1}{2}u^2} du \\ &= (2\pi)^{-1/2} |\theta^\star(z)|^{-1} e^{-\frac{1}{2}|\theta^\star(z)|^2} \int_0^\infty e^{-v - \frac{1}{2}|\theta^\star(z)|^{-2}v^2} dv \\ &\sim (2\pi)^{-1/2} |\theta^\star(z)|^{-1} e^{-\frac{1}{2}|\theta^\star(z)|^2} \qquad \text{as } z \to \infty. \end{aligned} \tag{21}$$

The last approximation goes beyond a large deviations estimate (i.e., it includes the prefactor), and it implies

$$\lim_{z\to\infty} |\theta^\star(z)|^{-2} \log\left((2\pi)^{-D/2} \int_{P(z)} e^{-\frac{1}{2}|\theta|^2} d\theta \right) = -\frac{1}{2}. \tag{22}$$

This log-asymptotic estimate is often written as

$$\int_{P(z)} e^{-\frac{1}{2}|\theta|^2} d\theta \asymp e^{-\frac{1}{2}|\theta^\star(z)|^2} \qquad \text{as } z \to \infty. \tag{23}$$

Interestingly, while the asymptotic estimate in Eq. **21** does not necessarily apply to the original integral in Eq. **15** [that is, the prefactor may take different forms depending on the shape of $\partial\Omega(z)$ near $\theta^\star(z)$], the rougher log-asymptotic estimate in Eq. **23** does as long as the boundary $\partial\Omega(z)$ is smooth, with a curvature that is bounded by a constant independent of z. This is because the contribution (positive or negative) to the integral over the region between $\Omega(z)$ and $P(z)$ is subdominant in that case, in the sense that the log of its amplitude is dominated by $|\theta^\star(z)|$. This is the essence of the large deviations result that we apply in this work.

Numerical Aspects. To perform the calculations, we solve Eq. **1** with $L = 40\pi$ and periodic boundary conditions, and check that this domain is large enough to make the effect of these boundary conditions negligible (*Supporting Information*). The spatial domain is discretized by using 2^{12} equidistant grid points, which is enough to resolve the solution of Eq. **1**. To evolve the field $u(t, x)$ in time, we use a pseudospectral second-order exponential time-differencing (ETD2RK) method (48, 49).

When performing the Monte Carlo simulations, we use 10^6 realizations of the random initial data constructed by truncating the sum in Eq. **2** over the $M = 23$ modes with $-11 \leq n \leq 11$—i.e., $-3\Delta \leq k_n \leq 3\Delta$: These modes carry most of the variance, and we check that adding more modes to the initial condition did not affect the results in any significant way (*Supporting Information*).

Optimization Procedure. As explained above, the large deviation rate function $I_T(z)$ in Eq. **6** can be evaluated by solving the dual optimization problem in Eq. **8**, which we rewrite as $S_T(\lambda) = \inf_{u_0} E(u_0, \lambda)$, where we define the cost function

$$E(u_0, \lambda) \equiv \frac{1}{2}\|u_0\|_C^2 - \lambda F(u(T, u_0)). \tag{24}$$

We perform this minimization using steepest descent with adaptive step (line search) and preconditioning of the gradient (50). This involves evaluating the (functional) gradient of $E(u_0, \lambda)$ with respect to u_0. Using the chain rule, this gradient can be expressed as (using compact vectorial notation)

$$\frac{\delta E}{\delta u_0} = C^{-1}u_0 - \lambda J^T(T, u_0)\frac{\delta F}{\delta u} \tag{25}$$

where $J(t, u_0) = \delta u(t, u_0)/\delta u_0$ is the Jacobian of the transformation $u_0 \to u(t, u_0)$. Collecting all terms on the right-hand side of the MNLS Eq. **1** into $b(u)$, this equation can be written as

$$\partial_t u = b(u), \quad u(t = 0) = u_0, \tag{26}$$

and it is easy to see that in this notation $J(t, u_0)$ satisfies

$$\partial_t J = \frac{\delta b}{\delta u} J, \qquad J(t = 0) = \mathrm{Id}. \tag{27}$$

Consistent with what was done in the Monte Carlo sampling, to get the results presented above, we truncate the initial data u_0 over $M = 23$ modes using the representation

$$u_0(x) = \sum_{n=-11}^{11} e^{ik_n x}\hat{a}_n, \quad k_n = 2\pi n/L. \tag{28}$$

This means that minimization of Eq. **24** is performed in the $2M - 1 = 45$ dimensional space spanned by the modes $\hat{a}_n$, accounting for invariance by an overall phase shift—to check convergence, we also repeat this calculation using larger values of M and find no noticeable difference in the results (*Supporting Information*).

In practice, the evaluation of the gradient in Eq. **25** is performed by integrating both $u(t)$ and $J(t)$ up to time $t = T$. Eq. **27** is integrated by using the same pseudospectral method as for Eq. **1** on the same grid. To perform the steepest descent step, we then precondition the gradient through scalar multiplication by the step-independent, diagonal metric with the components of the spectrum $\hat{C}_n$ as diagonal elements.

ACKNOWLEDGMENTS. We thank W. Craig and M. Onorato for helpful discussions; O. Bühler, M. Mohamad, and T. Sapsis for interesting comments; and an anonymous reviewer for drawing our attention to the semiclassical theory for the nonlinear Schrödinger equation. G.D. is supported by the joint Math PhD program of Politecnico and Università di Torino. E.V.-E. is supported in part by National Science Foundation (NSF) Materials Research Science and Engineering Center Program Award DMR-1420073; and by NSF Award DMS-1522767.

1. Müller P, Garrett C, Osborne A (2005) Rogue waves. *Oceanography* 18:66–75.
2. White BS, Fornberg B (1998) On the chance of freak waves at sea. *J Fluid Mech* 355:113–138.
3. Haver S (2004) A possible freak wave event measured at the Draupner jacket January 1 1995. Proceedings of Rogue Waves 2004, October 20–22, Brest, France (Ifremer, Plouzané, France), pp 1–8.
4. Nikolkina I, Didenkulova I (2011) Rogue waves in 2006–2010. *Nat Hazards Earth Syst Sci* 11:2913–2924.
5. Onorato M, Residori S, Bortolozzo U, Montina A, Arecchi F (2013) Rogue waves and their generating mechanisms in different physical contexts. *Phys Rep* 528:47–89.
6. Nazarenko S, Lukaschuk S (2016) Wave turbulence on water surface. *Annu Rev Condens Matter Phys* 7:61–88.
7. Akhmediev N, Soto-Crespo JM, Ankiewicz A (2009) Extreme waves that appear from nowhere: On the nature of rogue waves. *Phys Lett A* 373:2137–2145.
8. Akhmediev N, Pelinovsky E (2010) Editorial—introductory remarks on "Discussion & debate: Rogue waves—towards a unifying concept?". *Eur Phys J Spec Top* 185:1–4.
9. Onorato M, Proment D, El G, Randoux S, Suret P (2016) On the origin of heavy-tail statistics in equations of the nonlinear Schrödinger type. *Phys Lett A* 380:3173–3177.
10. Benjamin TB, Feir JE (1967) The disintegration of wave trains on deep water: Part 1. Theory. *J Fluid Mech* 27:417–430.
11. Zakharov VE (1968) Stability of periodic waves of finite amplitude on the surface of a deep fluid. *J Appl Mech Tech Phys* 9:190–194.
12. Kuznetsov EA (1977) Solitons in a parametrically unstable plasma. *Akademiia Nauk SSSR Doklady* 236:575–577.
13. Peregrine DH (1983) Water waves, nonlinear Schrödinger equations and their solutions. *ANZIAM J* 25:16–43.
14. Akhmediev N, Eleonskii V, Kulagin N (1987) Exact first-order solutions of the nonlinear Schrödinger equation. *Theor Math Phys* 72:809–818.
15. Osborne AR, Onorato M, Serio M (2000) The nonlinear dynamics of rogue waves and holes in deep-water gravity wave trains. *Phys Lett A* 275:386–393.
16. Zakharov VE, Ostrovsky LA (2009) Modulation instability: The beginning. *Physica D* 238:540–548.
17. Onorato M, et al. (2009) Statistical properties of directional ocean waves: The role of the modulational instability in the formation of extreme events. *Phys Rev Lett* 102:114502.
18. Dysthe KB (1979) Note on a modification to the nonlinear Schrödinger equation for application to deep water waves. *Proc R Soc Lond A* 369:105–114.
19. Stiassnie M (1984) Note on the modified nonlinear Schrödinger equation for deep water waves. *Wave motion* 6:431–433.
20. Trulsen K, Dysthe KB (1996) A modified nonlinear Schrödinger equation for broader bandwidth gravity waves on deep water. *Wave motion* 24:281–289.
21. Craig W, Guyenne P, Sulem C (2010) A Hamiltonian approach to nonlinear modulation of surface water waves. *Wave Motion* 47:552–563.
22. Gramstad O, Trulsen K (2011) Hamiltonian form of the modified nonlinear Schrödinger equation for gravity waves on arbitrary depth. *J Fluid Mech* 670: 404–426.
23. Onorato M, et al. (2004) Observation of strongly non-Gaussian statistics for random sea surface gravity waves in wave flume experiments. *Phys Rev E* 70:067302.
24. Chabchoub A, Hoffmann N, Akhmediev N (2011) Rogue wave observation in a water wave tank. *Phys Rev Lett* 106:204502.
25. Chabchoub A, Hoffmann N, Onorato M, Akhmediev N (2012) Super rogue waves: Observation of a higher-order breather in water waves. *Phys Rev X* 2:011015.
26. Goullet A, Choi W (2011) A numerical and experimental study on the nonlinear evolution of long-crested irregular waves. *Phys Fluids* 23:016601.
27. Lo E, Mei CC (1985) A numerical study of water-wave modulation based on a higher-order nonlinear Schrödinger equation. *J Fluid Mech* 150:395–416.
28. Cousins W, Sapsis TP (2015) Unsteady evolution of localized unidirectional deep-water wave groups. *Phys Rev E* 91:063204.
29. Cousins W, Sapsis TP (2016) Reduced-order precursors of rare events in unidirectional nonlinear water waves. *J Fluid Mech* 790:368–388.
30. Nazarenko S (2011) *Wave Turbulence*, Lecture Notes in Physics (Springer, Berlin), Vol 825.
31. Hasselmann K, et al. (1973) *Measurements of wind-wave growth and swell decay during the Joint North Sea Wave Project (JONSWAP)*, Hydraulic Engineering Reports (Deutsches Hydrographisches Institut, Hamburg, Germany).
32. Onorato M, Osborne AR, Serio M, Bertone S (2001) Freak waves in random oceanic sea states. *Phys Rev Lett* 86:5831–5834.
33. World Meteorological Organization, ed (2016) Manual on Codes—International Codes, Volume I.1, Annex II to the WMO Technical Regulations: Part A—Alphanumeric Codes, WMO No. 306 (Secretariat of the World Meteorological Organization, Geneva).
34. Janssen PAEM (2003) Nonlinear four-wave interactions and freak waves. *J Phys Oceanogr* 33:863–884.
35. Bertola M, Tovbis A (2013) Universality for the focusing nonlinear Schrödinger equation at the gradient catastrophe point: Rational breathers and poles of the Tritronquée solution to Painlevé I. *Commun Pure Appl Math* 66:678–752.
36. Tikan A, et al. (2017) Universality of the peregrine soliton in the focusing dynamics of the cubic nonlinear Schrödinger equation. *Phys Rev Lett* 119:033901.
37. Akhmediev N, Ankiewicz A, Taki M (2009) Waves that appear from nowhere and disappear without a trace. *Phys Lett A* 373:675–678.
38. Shrira VI, Geogjaev VV (2010) What makes the peregrine soliton so special as a prototype of freak waves? *J Eng Math* 67:11–22.
39. Akhmediev N, Dudley JM, Solli D, Turitsyn S (2013) Recent progress in investigating optical rogue waves. *J Opt* 15:060201.
40. Toenger S, et al. (2015) Emergent rogue wave structures and statistics in spontaneous modulation instability. *Sci Rep* 5:10380.
41. Chabchoub A (2016) Tracking breather dynamics in irregular sea state conditions. *Phys Rev Lett* 117:144103.
42. Bailung H, Sharma S, Nakamura Y (2011) Observation of peregrine solitons in a multicomponent plasma with negative ions. *Phys Rev Lett* 107:255005.
43. Kibler B, et al. (2010) The peregrine soliton in nonlinear fibre optics. *Nat Phys* 6:790–795.
44. Suret P, et al. (2016) Single-shot observation of optical rogue waves in integrable turbulence using time microscopy. *Nat Commun* 7:13136.
45. Mohamad MA, Cousins W, Sapsis TP (2016) A probabilistic decomposition-synthesis method for the quantification of rare events due to internal instabilities. *J Comp Phys* 322:288–308.
46. Farazmand M, Sapsis TP (2017) A variational approach to probing extreme events in turbulent dynamical systems. arXiv:1704.04116.
47. Farazmand M, Sapsis TP (2017) Reduced-order prediction of rogue waves in two-dimensional deep-water waves. *J Comput Phys* 340:418–434.
48. Cox SM, Matthews PC (2002) Exponential time differencing for stiff systems. *J Comp Phys* 176:430–455.
49. Kassam AK, Trefethen LN (2005) Fourth-order time-stepping for stiff PDEs. *SIAM J Sci Comput* 26:1214–1233.
50. Borzì A, Schulz V (2011) *Computational Optimization of Systems Governed by Partial Differential Equations* (Society for Industrial and Applied Mathematics, Philadelphia).