Statistics of Extreme Events in Fluid Flows and Waves

Citation: Sapsis, Themistoklis P. 2021. "Statistics of Extreme Events in Fluid Flows and Waves." Annual Review of Fluid Mechanics, 53 (1).

As Published: 10.1146/ANNUREV-FLUID-030420-032810

Publisher: Annual Reviews

Persistent URL: https://hdl.handle.net/1721.1/141386

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Statistics of Extreme Events in Fluid Flows and Waves

Themistoklis P. Sapsis[1]

[1]Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, 01239, USA; sapsis@mit.edu

Keywords

Rare extreme events, intrinsic instabilities, heavy tails, risk modeling.

Abstract

Extreme events in fluid flows, waves, or structures interacting with them is critical for a wide range of areas, including reliability and design in engineering, as well as risk modeling due to natural disasters. Such events are characterized by the co-existence of high intrinsic dimensionality, complex nonlinear dynamics, and stochasticity. These properties severely restrict the application of standard mathematical approaches, which have been successful in other areas. This review focuses on methods specifically formulated to deal with these properties and it is structured around two cases: i) problems where an accurate but expensive model exists, and ii) problems where a small amount of data and possibly an imperfect reduced-order model that encodes some physics about the extremes can be employed.

1. INTRODUCTION

Extreme events in fluid flows, water waves, as well as engineering systems operating and interacting with those media is a topic of vital importance for risk evaluation, optimization and design, as well as for predictive control and autonomy. There is a plethora of relevant phenomena in fluids and waves, spanning a wide range of temporal and spatial scales and associated with a variety of dynamical mechanisms (Figure 1). Some examples include extreme dissipation and enstrophy events in hydrodynamic turbulence (Ishihara et al. 2007; Yeung et al. 2015), cavitation inception due to extremely low values of the pressure occurring in turbulent shear layers (Iyer & Ceccio 2002; Tsuji & Ishihara 2003; Tsuji et al. 2007, 2012), high energy acoustic bursts in turbulent jets (Aubert & McKinley 2011; Brès et al. 2017; Schmidt & Schmid 2019), transitions between chaotic and regular regimes in vortex-induced vibrations and their impact on structural fatigue life (Chasparis et al. 2009; Modarres-Sadeghi et al. 2010, 2011), extreme ship motions and loads in irregular water waves (St Denis & Pierson 1953; Ochi & Wang 1976; Belenky et al. 2019), and atmospheric blocking events leading to extreme weather phenomena (Tibaldi & Molteni 1990; Davini et al. 2012).

Extreme events are manifested by selected observables of the system, $q(u)$, which is described by the state variables, $u \in \mathbb{R}^N$. These observables undergo sporadic bursts with values spanning several standard deviations. The observables are often specified by the nature of the application and therefore their form has to be assumed nonlinear. In water waves, for example, a natural quantity of interest is the local maxima of the wave elevation, while in fluid-structure interaction problems it is the loads acting on the structure.

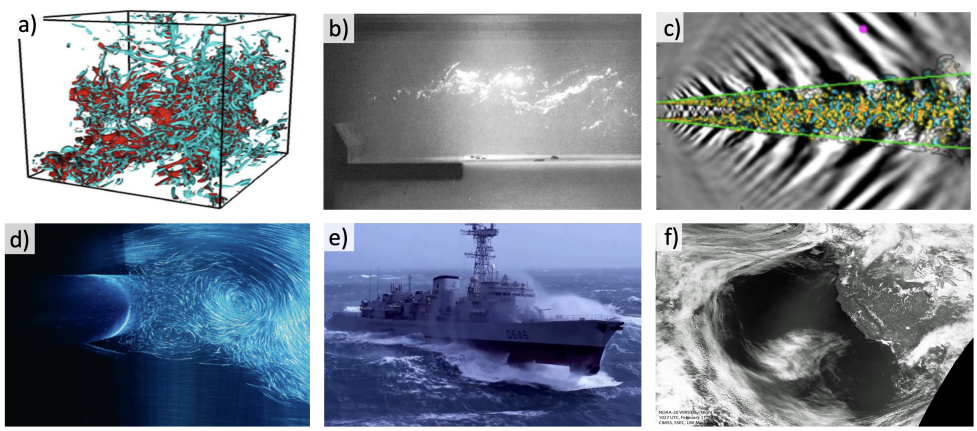

Figure 1

a) Extreme events of dissipation (red) and enstrophy (cyan) in 3D homogenous and isotropic turbulence (Yeung et al. 2015). b) Cavitation due to extremely low values of pressure in a turbulent shear layer (Iyer & Ceccio 2002). c) A high-energy burst in the acoustic far field of a turbulent jet: streamwise fluctuation velocity in the centre plane (coloured contours) and pressure component (greyscale contours) (Schmidt & Schmid 2019). d) Transitions between chaotic and quasi-periodic regimes in vortex-induced-vibrations (image courtesy: M. Bernitsas). e) Extreme ship motions and loads. f) A giant high-pressure ridge (atmospheric blocking event) over the Pacific Ocean has caused the most dry winter in California over the last 150 years (Secon 2020) (image taken by a National Oceanic and Atmospheric Administration (NOAA) satellite on February 11, 2020).

A usual statistical quantification measure of extreme events is the probability distribution function (pdf) of the observable. Due to the rare character of extreme events, the value of the pdf in regions associated with extremes is very small, making its accurate computation a challenging task. For a variety of problems it is also important to know the shape of extreme events, which is often captured through conditional statistical measures - the latter require also a large amount of data describing extreme events. In this review we discuss the challenges connected with extreme events in fluids and waves and present some recent computational methods which are motivated by these properties.

2. THE STOCHASTIC NONLINEAR NATURE OF EXTREME EVENTS

What is special about extreme events in fluid flows and waves? The associated models for these problems are typically continuous dynamical systems that 'live' in infinite dimensional spaces. Moreover, for most geophysical problems, as well as high-Re engineering flows the intrinsic dimensionality is very large, most often infinite. In other words, it is not straightforward to apply order reduction methods and describe the system with a low-dimensional model without losing critical dynamics and information, especially for extreme events (Sapsis & Majda 2013a,b; Majda et al. 2014). On the other hand, although high-performance computing allows for direct simulation of high-dimensional systems, analyzing probability of events that occur once every 20 or 100 years or optimizing the corresponding designs is still out of the question. From a mathematical viewpoint, extreme events in these high-dimensional systems are characterized by two attributes that limit the application of several mathematical methods that have been successfully used in other areas, namely,

- **Stochasticity**: The existence of uncertainty either due to intrinsic chaotic dynamics and/or due to exogenous causes such as stochastic excitation or uncertain parameters.
- **Nonlinearity**: The existence of essential nonlinearity inherent either in the dynamics or the observables.

The stochastic character leads to uncertainty in the time of occurrence of an extreme event and, to this end, there is no specific frequency associated with rare events. For this reason, methods that focus on spectral analysis of the underlying dynamics such as dynamic mode decomposition (Schmid 2010) or Koopman modes (Mezić 2013), albeit very successful with the analysis of statistical steady states, cannot capture effectively these intermittently occurring episodes (see section IV.F in Farazmand & Sapsis (2016)).

On the other hand, depending on the form of the nonlinearity in the dynamics, the system may be characterized by extreme events due to i) weak energy transfers between wavenumbers without the occurrence of positive Lyapunov exponents, or ii) violent energy transfers between degrees of freedom associated with intermittently positive Lyapunov exponents in the underlying dynamics. In both cases the synergistic action of nonlinearity and stochasticity leads to uncertainty in the magnitude of the transient episodes.

An additional challenge associated with extreme events is the fact that their magnitude, although distinguishably larger, cannot exceed a certain bound. In practice, this is due to global quantities being conserved, changing at slower time scales, or bounded, e.g. energy, momentum, enstrophy, etc. These intrinsic barriers are manifested through additional nonlinear mechanisms that prevent the indefinite growth of transient episodes, independently of the stochastic parameters. This property introduces important technical obstacles on the application of statistical extrapolation methods, such as extreme value theorems (Lucarini

et al. 2012, 2014), which assume a uniform decay law of the pdf tail; this is violated due to these intrinsic physical barriers.

An alternative approach for the analysis of extreme events in fluids and waves is given by large deviations theory (LDT) (Varadhan 1984, 2008). This has been formulated for rogue waves described by nonlinear Schrödinger equation in Dematteis et al. (2018, 2019). LDT utilizes the dynamical equations to approximate the asymptotic behavior of the tail for any given up-crossing level by identifying the most likely random event that corresponds to a trajectory that crosses this level. This method involves an optimization problem in the high-dimensional phase space of the dynamical system, which can be very expensive to solve. Moreover, the method provides information only for the tail of the pdf and not its full form. Concomitantly, numerical methods have been developed for efficient randomized sampling of rare events. These ideas have recently been applied on the quantification of rare-event statistics of the drag acting on an obstacle in a two-dimensional turbulent flow (Lestang et al. 2019), as well as quantification of rare climate events (Ragone & Bouchet 2019).

2.1. Low-dimensional prototype models for extreme events

Depending on the nature of the system, the rare-event properties and statistics may be dominated by exogenous stochastic causes, intrinsic instabilities, or both. In the first case the underlying dynamical system is *stable* and all the large events come as a result of the stochasticity introduced by the forcing or the parameters, combined with weak nonlinearities of the dynamics and possibly of the observable. Typical examples in this class of systems include fluid-structure interaction problems (Mohamad & Sapsis 2018), weakly nonlinear water waves (Fedele & Tayfun 2009), and ship motions and loads (Belenky et al. 2019). To make ideas more concrete we consider as an example a prototype system where extreme events occur due to exogenous causes: a pendulum modeling the roll motion of ships under random waves (Belenky & Sevastianov 2007),

$$\ddot{x} + c\dot{x} + a\sin x = f(t;\omega), \quad x \in [-\pi, \pi), \tag{1.}$$

where $a > 0$, $c > 0$, (stable system) and $f(t;\omega)$ is a stochastic process with known statistics, and ω is the random argument. In this case large-amplitude excitation leads to large response. The non-Gaussian, heavy-tailed pdf for the displacement of the system is shown in Fig. 2a together with a sample time series. While the nonlinearity defines the form of the pdf tails, there are no instabilities present in the dynamics. All rare events in this case are the direct result of large excitation amplitudes, which are amplified due to the nonlinearity of the system. In this sense, it is the recent history of the excitation that is the dominant factor for the formation of a large event.

On the other hand, we have systems that are stable most of the time but not for all times. *Transient instabilities* may be triggered by small noise and then result in an extreme event. One of the most popular examples are modes in turbulent fluid flows and nonlinear water waves subjected to nonlinear energy exchanges that occur in a random fashion and result in intermittent responses (Pedlosky 1998; Salmon 1998; DelSole 2004; Majda et al. 2005, 1997; Dysthe et al. 2008; Xiao et al. 2013). Low-dimensional prototype systems that mimic these properties were introduced in Majda & Branicki (2012); Majda & Harlim (2012). The simplest example from this class of systems is a single-mode model with multiplicative excitation which mimics the interaction with other degrees of freedom (Mohamad & Sapsis 2015). This can be brought in the form of the well-known Mathieu

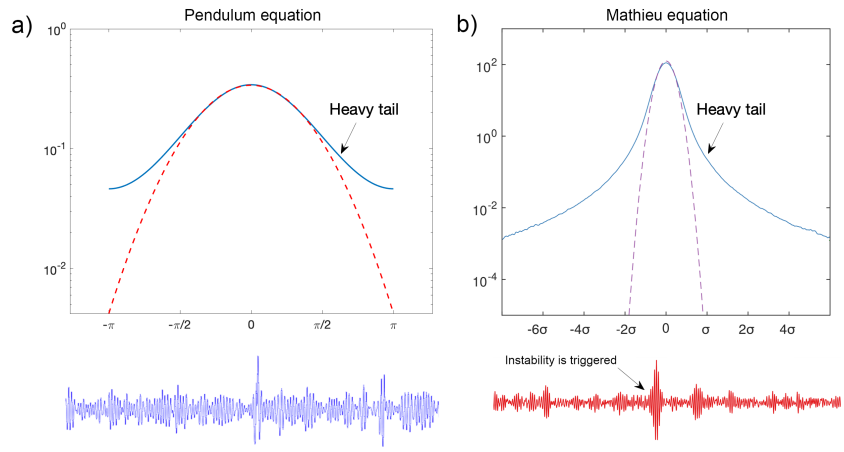

Figure 2

a) Pdf for a stable nonlinear system (pendulum) driven by stochastic noise (red dashed curve shows the optimal Gaussian approximation); Lower plot (blue curve) shows a segment of the time series for the pendulum displacement. b) Pdf for the Mathieu equation driven by small intensity noise; a small segment from the response time series is shown with red.

equation from mechanics:

$$\ddot{x} + c\dot{x} + (a - bg(t;\omega))x = f(t;\omega), \quad x \in \mathbb{R}, \qquad 2.$$

where a, b, c are positive numbers, $f(t;\omega)$ is small-intensity noise and $g(t,\omega)$ is a narrow-band processes with spectrum centered around the natural frequency of the (unforced) oscillator and possibly small intensity. The same model has also been used to describe transient parametrical resonances in ship roll motions (Mohamad & Sapsis 2016). Analytical approximations for the tail structure of the above equation have been presented in Mohamad & Sapsis (2015). In Fig. 2b we present the pdf of the response as well as a sample time series showing one extreme event. For this case, it is the intrinsic instability mechanism (when b is sufficiently large) that defines the properties and statistics of extreme events rather than the magnitude or the properties of the signals $f(t;\omega)$ and $g(t;\omega)$. Clearly, this simple model does not contain saturation mechanisms that bound the magnitude of extreme events and this is demonstrated by the homogeneous decay rate of the tail. In the next two subsections we review several high-dimensional models for fluid flows and waves exhibiting extreme events due to either exogenous or intrinsic instabilities.

2.2. High-dimensional models with transient instabilities

The phase space of systems exhibiting extreme events is typically characterized by a high-dimensional background set where the system state resides with large probability and smaller regions where instabilities occur (Mohamad et al. 2016; Farazmand & Sapsis 2017a; Sapsis 2018; Farazmand & Sapsis 2019b). In Figure 3 a schematic for the phase space of such system is shown, exhibiting extreme events due to internal instabilities. The first component is a stochastic attractor or more generally a set where the system state lies most of the time, represented as the brown shaded region. This can be formed due to persistent

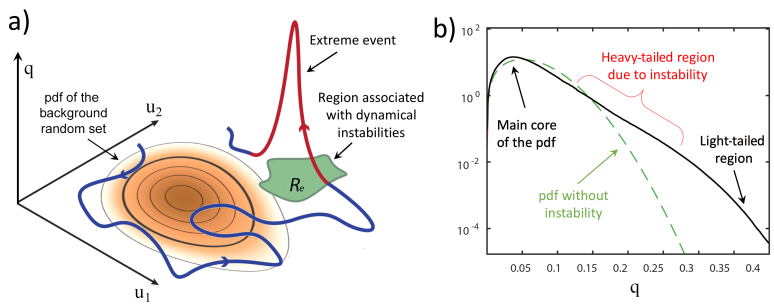

Figure 3

a) Extreme events are associated with large excursions due to the random triggering of dynamical instabilities. The shaded region indicates the pdf associated with the attractor of the system or more generally the set where the system lies most of the time. The instability region is in green. b) A typical heavy-tailed distribution for extreme events. The heavy-tailed region has finite extent because instabilities cannot lead to arbitrarily large magnitude. (Reproduced from Mohamad et al. (2016)).

instabilities (chaotic dynamics), stochastic parameters, or stochastic excitation of the system. The second component is an instability region, represented with green color. When the system enters this neighborhood certain observables grow rapidly indicating the formation of extreme events, due to nonlinear effects. These large excursions are manifested in the pdf of the observable as heavy-tailed regimes (Figure 3b). Because of limiting physical mechanisms, the heavy-tailed regime in the pdf must have finite support, regardless of the size and topology of the instability regions.

2.2.1. Extreme events in nonlinear dispersive waves. Models for nonlinear dispersive waves include nonlinear Schrödinger equations (NLS) for the description of modulation instabilities that lead to rogue water waves (Zakharov 1968; Trulsen & Dysthe 1996; Trulsen et al. 2000; Xiao et al. 2013), extreme events in dispersive wave models for weak turbulence (Majda et al. 1997), and transition to extreme anomalous statistics in weakly turbulent shallow water waves going through an abrupt depth change, i.e. instabilities induced by geometry (Majda et al. 2019; Majda & Qi 2019). For many of these problems the mechanism behind transient instabilities is nonlinear focusing in spatially localized wavegroups (Osborne 2001; Adcock et al. 2012; Chabchoub et al. 2011). These intense localization events can randomly occur due to the random superposition of harmonic waves that form the random background i.e. the 'core' of the pdf. This is demonstrated in Figure 4 for the Majda-McLaughlin-Tabak (MMT) model, a one-dimensional nonlinear and dispersive wave model (Cousins & Sapsis 2014). We note that for these problems the random background is characterized by almost linear dynamics and non-linearity occurs only when there is sufficient energy concentration in spatial region. When the amplitude of the wave field is small, energy remains constant over a range of different wavenumbers and there is only phase mixing due to dispersion, since different wavenumbers propagate at different speeds.

To visualize the transient dynamics during an extreme event we have projected the wave field from the physical space (Figure 4a) on the localized basis given by

$$\hat{u}_i(x) = c_i \exp\left(-2\frac{(x-x_c)^2}{L^2} + \frac{2\pi i n x}{L}\right), \quad n = 0, 1, 2, ... \qquad 3.$$

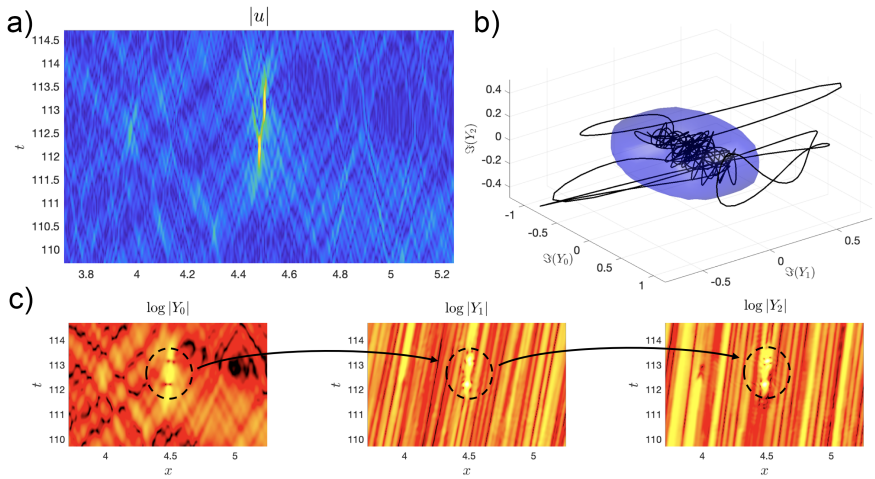

Figure 4

a) Extreme event in the MMT model, a one-dimensional nonlinear, dispersive wave model. b) Trajectory of an extreme event visualized when the wave field is projected in a spatially localized basis; the shaded region contains 97% of the probability mass. c) Energy transfers between different wavenumbers, occurring due to the transient instability. (Reproduced from Cousins & Sapsis (2014)).

where c_i are normalization constants while the center x_c is chosen close to the extreme event. The projection coefficients are denoted by $Y_i(t)$. The lengthscale L is associated with the instability. It is selected by solving a constrained optimization problem where the finite-time growth rate is maximized, but only within wavegroups that have finite-probability of occurrence, i.e. those that belong to the brown shaded region (Figure 3a). We refer to Cousins & Sapsis (2014) for a detailed discussion regarding the critical lengthscale in the MMT model, and to Cousins & Sapsis (2016) and Farazmand & Sapsis (2017b) for one-dimensional and two-dimensional water waves.

Using the localized basis, we can visualize the extreme wave as a large excursion in phase space. This trajectory is shown in Figure 4b together with a shaded surface that encloses a high-probability region: the 'core' of the attractor where trajectories of the system 'live' most of the time (this core accounts for 97% of the probability mass). The nonlinear character of the extreme event is also manifested by the energy of the projection coefficients (Figure 4c), where we can observe the nonlinear energy transfers between scales caused by the focusing instability. By identifying the most unstable modes that also have finite-probability of occurrence, one can formulate parsimonious short-term prediction schemes for water waves (Cousins & Sapsis 2016; Farazmand & Sapsis 2017b); see Cousins et al. (2019) for an experimental demonstration on one-dimensional water waves.

2.2.2. Extreme dissipation events in turbulent flows. Turbulent flows are unique for their complexity and richness of instabilities that include both persistent ones, i.e. associated with cascades of energy, but also intermittent bursts or transitions that occur less frequently but lead to important modifications of the flow. Examples of intermittent episodes range from large scale atmospheric blocking events (Tibaldi & Molteni 1990; Davini et al. 2012) to intermittent dissipation bursts in near wall turbulence (Jiménez & Moin 1991; Flores

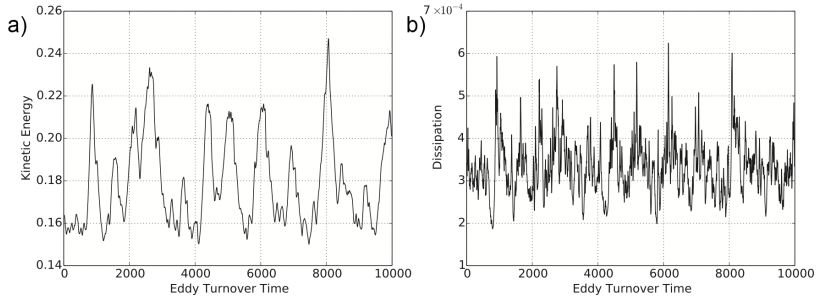

Figure 5

Time evolution of the kinetic energy (a) and the energy dissipation rate (b) for the near-wall minimal flow unit at $Re = 2200$. (Reproduced from Blonigan et al. (2019)).

& Jiménez 2010; Jiménez & Simens 2001). In contrast to the previous models, where the background stochasticity is formed due to dispersion, here it is the existence of persistent positive Lyapunov exponents that lead to chaotic dynamics. These instabilities form the random background out of which extreme dissipation events are randomly triggered. The transitions to extreme dissipation episodes and their precursors have been recently studied for the Kolmogorov flow (Farazmand & Sapsis 2017a) and the three-dimensional channel flow (Blonigan et al. 2019).

Turbulent channel flow has been a staple of numerical studies of turbulence for many years (Kim et al. 1987). The chaotic nature of these flows makes it difficult to analyze local spatiotemporal events and physical mechanisms in them, such as the formation and destruction of individual hairpin vortices in the near-wall region. To isolate these physical mechanisms work has been done to find "minimal flow units" for various regions of the channel. Jiménez & Moin (1991) found the minimal flow unit for near wall turbulence for several low-Reynolds-number flows by considering turbulent channel flow simulations with domains that were considerably smaller than conventional channel flow simulations. These smaller domains eliminate larger scale turbulent structures but accurately resolved the near-wall turbulent flow, matching turbulent flow statistics from experiments and prior numerical studies up to 40 wall units in the wall-normal direction.

The simulations of the near-wall minimal flow unit routinely show highly intermittent behavior. This observation led to numerous subsequent studies into the intermittent nature of near-wall turbulence using minimal flow units (see Jiménez (2012, 2018) for comprehensive reviews). The intermittent behavior of the flow at $Re = 2200$ can be seen in Fig. 5 from the spikes in kinetic energy $E(t)$ and dissipation $Z(t)$ (results reproduced from Blonigan et al. (2019)), where the kinetic energy and the energy dissipation rate $Z(t)$ are defined as

$$E(t) = \int_\Omega \rho u \cdot u dxdydz, \text{ and } Z(t) = \int_\Omega \mathrm{tr}(\tau \nabla u) dxdydz, \qquad 4.$$

with Ω being the flow domain, ρ is the fluid density, u is the flow velocity, and τ denotes the stress tensor, $\tau = \mu(\nabla u + \nabla u^\top)$, for an incompressible flow. The large spikes in $E(t)$ are the result of intermittent flow laminarization events.

Of course, the flow never completely reaches the laminar state, though it gets very close to it. Fig. 6 shows an example of a laminarization event where the flow undergoes the following stages (Blonigan et al. 2019). (I) The flow on the bottom wall laminarizes

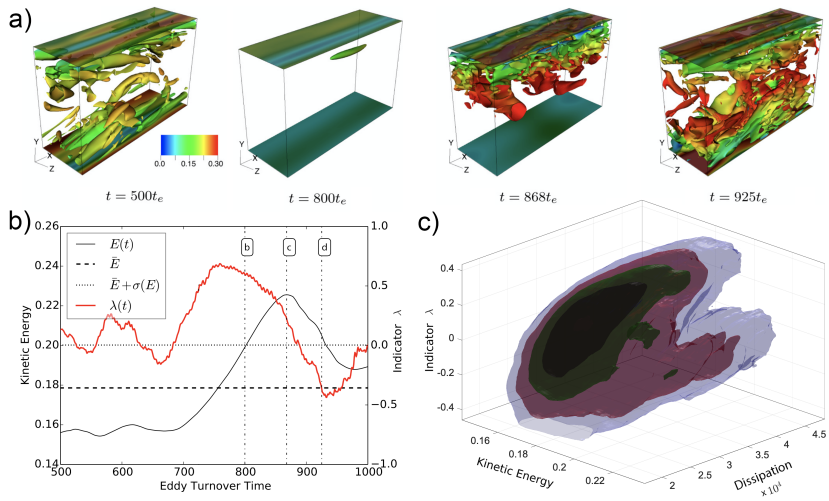

Figure 6

a) Different phases of an extreme laminarization event in a turbulent channel flow. b) The precursor, λ, measures how close the system state is to the transient instability. c) The joint pdf of the indicator, λ, the kinetic energy, E, and the dissipation rate, Z. (Reproduced from Blonigan et al. (2019)).

while the flow on the top wall becomes nearly laminar. (II) As the entire channel becomes nearly laminar, the streamwise velocity increases. (III) The higher velocities make the effective Reynolds number of the flow larger. This increases the likelihood of turbulent burst occurring since the flow is less stable to perturbations at a larger Reynolds number. (IV) A turbulent burst occurs on the top wall, which causes $Z(t)$ to increase rapidly and the bottom wall transitions to turbulence.

These flow laminarization events can be predicted by solving a constrained optimization problem looking for the state with the highest growth rate over finite time under the constraint that this state resides within the core of the attractor (Blonigan et al. 2019). By solving this optimization problem one can define a predictor (indicator λ) which measures the distance of the flow from that state. As shown in Fig. 6b the indicator λ identifies when the transient instability has been triggered, well before the kinetic energy becomes larger than its average value and of course before the dissipation rate burst which occurs after that. The predictive skill of the indicator λ can be seen by visualizing the joint pdf of (λ, E, Z), which parametrizes the large range of the kinetic energy and dissipation rate (including their extreme values) as it is shown in the pdf plot in Figure 6c.

2.2.3. Transitions between regular and chaotic regimes in VIV applications. Long, slender, flexible cylinders placed in a cross-flow, such as marine risers and mooring lines, exhibit high-frequency vortex-induced vibrations (VIV) accompanied by drag increase; this constitutes an important problem for ocean applications, especially for deep-water oil exploration and production. Several studies have relied on the hypothesis that the motion of the cylinder can be treated as a periodic phenomenon or as a statistically steady random response with a fundamental frequency close to the Strouhal frequency. However, it has recently been shown (Chasparis et al. 2009; Modarres-Sadeghi et al. 2011) that there are long pe-

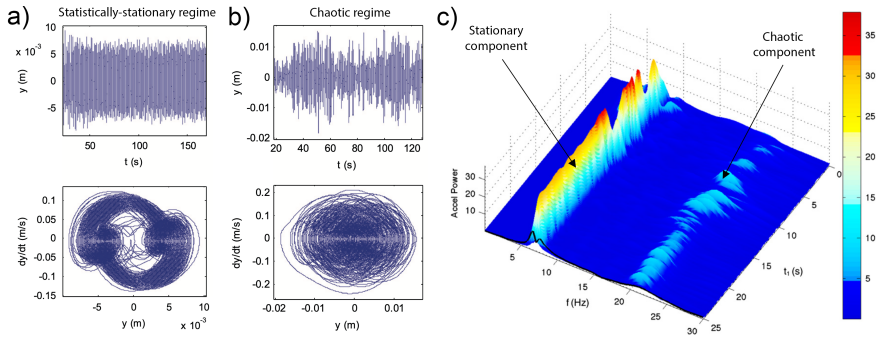

Figure 7

Random switching from a) stationary to b) non-stationary response in VIV. A time segment of the response in the two regimes is shown, as time series and in phasee portrait; c) Power spectrogram of the entire signal. (Reproduced from Modarres-Sadeghi et al. (2011)).

riods of VIV characterized by a statistically non-stationary response, with a rather wide power spectrum, which contains several frequencies in addition to the Strouhal frequency. Random transitions between stationary to non-stationary response is frequent and constitutes a principal feature of riser response. This random switching mechanism has important implications on the fatigue life of the structure (Modarres-Sadeghi et al. 2010) and therefore it is essential to quantify its effects probabilistically. In Figure 7 an example of such transitions for an experimentally measured VIV signal is shown (Modarres-Sadeghi et al. 2011). The signal consists of both statistically stationary segments with frequency content dominated by the Strouhal number (Fig. 7a), as well as chaotic segments (Fig. 7b). The transitional character of the signal can be observed clearly in the spectrogram (Fig. 7c). Similar transitions due to randomly triggered instabilities occur in other settings involving forced flows interacting with elastic structures, such as flutter in aeroelasticity of wings and flags (Dowell 2015; Zhang et al. 2000; Mandre & Mahadevan 2010).

2.3. Stable systems with weak nonlinearities

While transient instabilities characterize a wide range of systems exhibiting extreme events there are also problems characterized by stable dynamics and weak nonlinearities that lead to the formation of extreme events. Examples include fluid-structure interaction problems, as well as ship motions and loads. These systems are typically driven by stochasticity due to initial conditions or random excitations. The existence of nonlinearities in the dynamics essentially deforms trajectories resulting in larger values for the observable, q. In Figure 8 a schematic of such model is shown. The phase portrait of a weakly nonlinear system (blue curves) is given in Figure 8a, together with the corresponding phase portrait of the linearized system (red-dashed curves). The weakly non-linear system visibly differs only in the blue shaded regions where nonlinearity drives trajectories further away from the equilibrium. This type of 'softening' nonlinearity results in heavier tails for the state u_1, compared with the linear system that follows Gaussian statistics (Figure 8b). Below we provide a brief overview of several important models that exhibit extreme events due to stable, but weakly nonlinear dynamics.

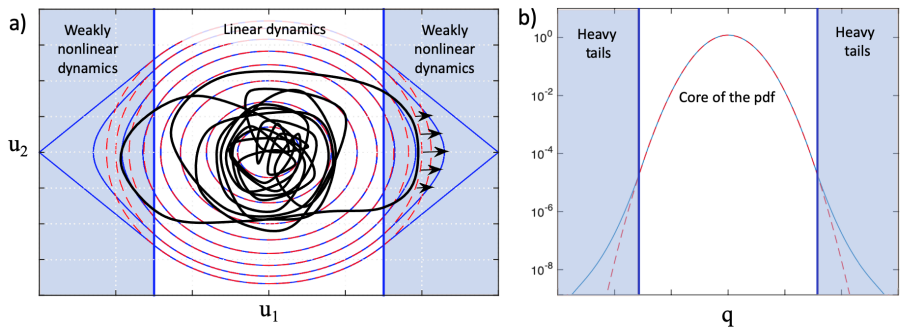

Figure 8

a) The phase portrait of a system with weak nonlinearity. The red curves correspond to the linear system while the blue curves represent the trajectories of the weakly nonlinear system. Nonlinearity is fully active in the shaded blue regions. A realization driven by random excitation is shown as a black solid curve. b) The pdf for the observable, $q \equiv u_1$.

2.3.1. Extreme events in ship motions and loads. The probabilistic quantification of extreme events for motions of and loads impaired to, ships subjected to random seas is an important open problem for ocean engineering. It is computationally challenging as it requires the solution of the full fluid-body interaction, i.e. the effect of the wave and flow field on the ship and vice versa. On the other hand, for many ocean engineering and naval applications it is critical to quantify events, such as extreme motions and loads that will occur once every century, a task that requires a vast amount of simulations.

In Figure 9 we present the non-trivial structure of the computed pdfs together with the hydrostatic restoring moments for pitch and roll motions of the ONR Tumblehome hull. Direct Monte-Carlo simulations were performed using the Large Amplitude Motion Program (LAMP), which is a time-domain simulation code for ship motions and wave loads incorporating a nonlinear (Froude-Krylov) calculation of the incident wave, hydrostatic restoring pressure forces and a three-dimensional potential flow panel solution of the wave-body disturbance (Shin et al. 2003). LAMP computes a pressure distribution on the wetted portion of the hull surface at each time increment, which is integrated to get the forces acting on the hull. Once forces are available, the 6-DOF equations of motion are solved for the next step. The hydrostatic moments are just one component out of many that need to be resolved in order to accurately capture the tails, although it plays a dominant role on the form of the pdf tails, especially for the roll motion (Belenky et al. 2019).

We observe that the tail of the roll pdf is heavier than Gaussian for roll angles around 10 degrees, a property that is connected with the softening character of the roll hydrostatic restoring moment (Fig 9b). However, for larger angles its heavy character disappears and it decays faster than the normal distribution. The vertical bending moment has an even more complicated form that is characterized by asymmetry (due to the longitudinal asymmetry of the hull) and a very heavy tail region for positive moments. This form is not the result of nonlinearity in the dynamics (in fact, pitch motion follows almost exactly Gaussian statistics) but rather it is associated with the nonlinearity of the observable, i.e. the vertical bending moment at the mid-ship section (Sapsis et al. 2020). Similar properties have been demonstrated in other areas, such as loads acting on offshore platforms and offshore wind turbines (Naess & Moan 2012).

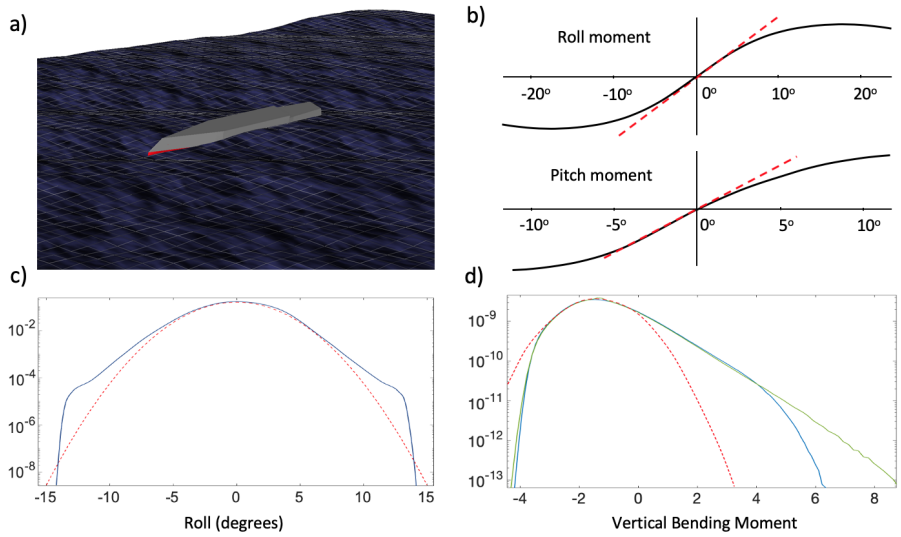

Figure 9

a) A 3D plot of the considered hull (ONR Tumblehome) subjected to random waves. b) The hydrostatic roll and pitch restoring moments for different angles. c) The computed pdf for the roll angle (blue curve) and its Gaussian approximation (red) obtained using the potential-flow wave-body hydrodynamics code LAMP. d) The corresponding pdf for the vertical bending moment at the midship section with (blue) and without considering the ship deck (green), as well as the Gaussian approximation (red).

2.3.2. Extreme events in weakly nonlinear waves. Extreme events can form either due to transient modulation instabilities or through second-order nonlinearities. While second-order interactions lead, in general, to less intense rogue waves compared with modulation instability, the resulting pdfs are also characterized by tails heavier than normal. For second-order waves in deep water, the surface elevation is approximated explicitly (Longuet-Higgins 1963; Forristall 2000)

$$\eta(x,y,t) = \sum_{n=1}^{N} A_n \cos(k_n \cos(\theta_n)x + k_n \sin(\theta_n)y - \omega_n t + \epsilon_n) \tag{5.}$$

$$\eta^{(2)}(x,y,t) = \frac{1}{4}\sum_{i=1}^{N}\sum_{j=1}^{N} A_i A_j \left(K^- \cos(\psi_i - \psi_j) + K^+ \cos(\psi_i + \psi_j)\right) \tag{6.}$$

where x, y are spatial coordinates, $k_n = |\mathbf{k}_n|$, $\mathbf{k}_n = (k_x, k_y)$ is the magnitude of the horizontal wavenumber and ω_n, θ_n are the radial frequency and directional angle of the nth Fourier component, respectively. The parameter ϵ_n is a phase uniformly distributed in the interval $[0, 2\pi]$. The frequency and the wavenumber are connected through the linear dispersion relation $\omega_n^2 = g|\mathbf{k}_n|$, where g is the acceleration of gravity. The coupling coefficients K^- and K^+ represent interaction coefficients, while $\psi_n = k_n \cos(\theta_n)x + k_n \sin(\theta_n)y - \omega_n t + \epsilon_n$. The resolved approximations obtained for the case of second order waves allow for the derivation of analytical expressions for the exceedance probability (Tayfun 1980; Huang et al. 1986). This methodology has also been employed to characterize the non-Gaussian properties of wave-induced structural loads in Sclavounos et al. (2019).

3. COMPUTATIONAL PROBABILISTIC ANALYSIS OF EXTREMES

We make a distinction between problems where i) an accurate model is available, from which we can obtain the value or the statistics of the quantity of interest for given values of the stochastic parameters, and ii) only data is available and possibly a prototype or reduced-order model that captures some relevant physics.

The first class involves most of the problems already described, where a good but possibly computationally intensive model exists. In this case, it is also important to know the statistics of the stochastic parameters, $r \in \mathbb{R}^M$. For problems where the stochasticity is created from exogenous causes or initial conditions (e.g., nonlinear waves or structures excited by waves or fluids) the statistics of the random parameters (r) is assumed to be known. For problems where the stochasticity is generated intrinsically (e.g., extreme dissipation events in turbulent flows or random transitions in VIV problems) one can calculate from short simulations the second-order statistics of the system state, $u \in \mathbb{R}^N$, and adopt a simple approximative model for its distribution, such as a Gaussian pdf. In this case the state vector is treated as a random parameter. Then the goal is to obtain the nonlinear map from the stochastic parameters to the quantity of interest $q = T(z), z \equiv (r, u) \in \mathbb{R}^{N+M}$, in a computationally efficient way.

The second class of methods involves problems where there is no accurate model that one can rely on, but rather a large data-set and possibly a reduced-order model that captures some relevant physics. The goal in this case is to extrapolate the statistics of certain quantities of interest in order to compute probabilities of extreme events. This class involves, for example, geophysical fluid flows over climate scales where reanalysis data is a reliable source of information. In contrast to standard statistical extrapolation methods (Lucarini et al. 2012, 2014) that rely on extreme value theorems to compute the pdf tails from data, the focus here is on statistical extrapolation methods that encode some information about the dynamics. This information can either be extracted directly from the data or by employing imperfect reduced-order models that qualitatively capture some aspects of the dynamics.

3.1. Model-based methods for extreme event statistics

For problems where an accurate model is available we review two methods that can significantly reduce the computational cost for the statistics of arbitrary quantities of interest. The first one, the output-weighted sequential sampling method, is dynamics-agnostic and relies on the solution of an optimization problem that provides the most informative samples for extreme event statistics. The second, the probabilistic decomposition synthesis, relies on the knowledge of the dynamical mechanism that is causing the extreme events, i.e. the domain of the phase space over which instabilities are triggered.

3.1.1. Output-weighted active sequential sampling. An important family of methods for quantifying rare events in complex systems is randomized sampling algorithms. These aim to guide the randomized sampling process towards rare events (see Webber et al. (2019) and Ragone & Bouchet (2019) for applications in fluids). On the other hand, we have active learning or optimal experimental design methods, where a sequence of vectors $\{\hat{z}_1, \hat{z}_2, ..., \hat{z}_{n-1}\}$ and the associated values of the observable $q_i = T(\hat{z}_i)$, are assumed to be known. Then, the goal is to sequentially identify the most informative vector $\hat{z}_n$ for which the computation of the observable $q_n = T(\hat{z}_n)$ will result in the most accurate representation of the map $q = T(z)$ through a surrogate function. In Stefanakis et al. (2014) an active

experimental design approach is formulated to study if small islands can protect nearby coasts from tsunamis. An active experimental design method using mutual information criteria is also employed in Beck & Guillas (2016) to efficiently compute the maximum elevation of a tsunami wave at the shoreline, for a wide range of scenarios.

In the context of extreme event statistics a more relevant problem is the estimation of the pdf of q. This cannot be achieved by just computing the vector z that results in the maximum value of q. On the other hand, computing the full map, $q = T(z)$, requires a very large number of samples even if optimal experimental design methods are employed. This is the aim of the output-weighted active sequential (OWAS) sampling method (Mohamad & Sapsis 2018; Sapsis 2020), where the goal is to sequentially identify the most informative vector $\hat{z}_n$ for which the computation of the observable $q_n = T(\hat{z}_n)$ will result in the fastest possible convergence for the statistics of the specific observable, q. The key idea is that for the statistics of the given observable not all the input parameters contribute equally and therefore the method aims to simultaneously identify the important input parameters and quantify their effect on the statistics of q. Therefore, in OWAS sampling method the samples depends directly on the chosen observable: different samples will be more effective in quantifying the statistics of different observables. The problem has been studied using criteria relying on mutual information theory or the Kullback-Leibler divergence (Chaloner & Verdinelli 1995; Pandita et al. 2019). However, as analyzed in Sapsis (2020) the employment of these criteria results in very poor performance that is comparable with random sampling. Moreover, the computation of mutual information criteria for high-dimensional parameters spaces, as it is the case for waves and fluid flows, is not practical.

OWAS sampling, introduced in Mohamad & Sapsis (2018), focuses on selecting samples in regions of the parameter space that i) have not been sampled previously; ii) have non-negligible probability to occur; and iii) are associated with degrees of freedom or parameters whose variation has important effect on the observable q. The combination of all three properties results in a sampling scheme that significantly accelerates the convergence of the statistics, especially in the tail region. The method consists of the following steps (see also Fig. 10 for a graphical description of the steps):

1. A Gaussian Process Regression (GPR) scheme is utilized, based on the existing set of samples $\mathcal{D}_{n-1} = \{\hat{z}_i, T(\hat{z}_i)\}_{i=1}^{n-1}$. To estimate T we place a Gaussian process (GP) prior over T and consider the function values as a realization of a GP. This gives the posterior mean $\bar{T}_{n-1}(z)$ and standard deviation, $\sigma_{n-1}(z)$.
2. A new sample $\hat{z}^*$ is hypothesized; this will be optimized according to the properties above. Since directly sampling the system before this new point is finalized is computationally expensive, it is assumed that the exact value of the map is given by the best linear unbiased estimator, using the $n-1$ points, i.e., $T(\hat{z}^*) \simeq \bar{T}_{n-1}(\hat{z}^*)$. The error map, $\sigma_n(z; \hat{z}^*)$, which takes into account the pair $(z^*, \bar{T}_{n-1}(\hat{z}^*))$, is computed.
3. Using the second-order statistics for the random vector z, the pdfs of the upper and lower bounds of the estimated map (denoted as $\rho_n^{\pm}(s; \hat{z}^*)$) are computed, i.e., the pdfs of $\bar{T}_{n-1}(z) \pm \sigma_n(z; \hat{z}^*)$. These pdfs take into account the hypothetical new point $\hat{z}^*$.
4. The criterion utilized to select the next-best point is based on the following L_1 distance between the pdfs

$$\hat{Q}(\hat{z}^*) \triangleq \frac{1}{2} \int |\log \rho_n^+(s; \hat{z}^*) - \log \rho_n^-(s; \hat{z}^*)| \, ds. \tag{7.}$$

The next sample point $\hat{z}_n$ is chosen so that $\hat{Q}$ is minimized.

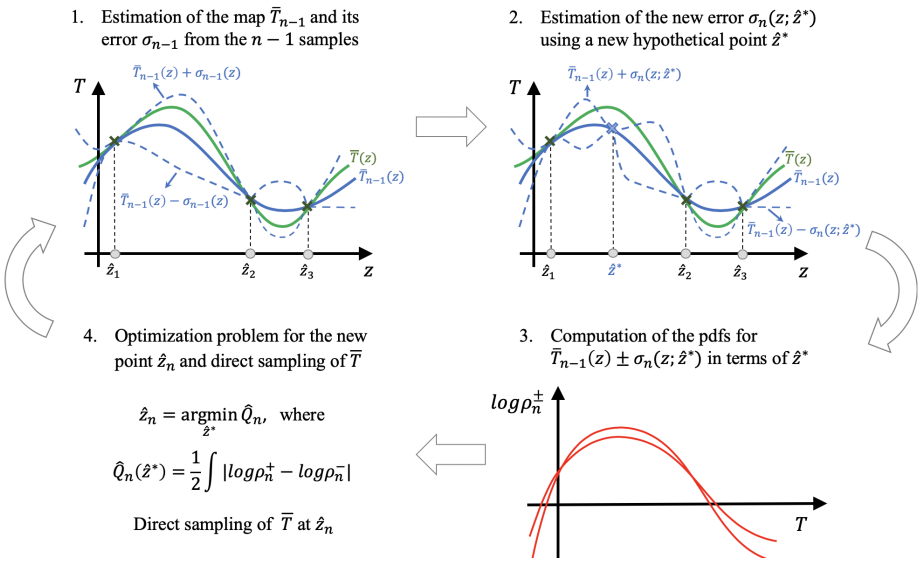

Figure 10

Summary of the active sequential sampling algorithm. The green curve represents the exact map $T(z)$ while the blue curves are the GPR estimates using the $n-1$ samples.

5. The true map T is evaluated at $\hat{z}_n$, and the pair $(\hat{z}_n, T(\hat{z}_n))$ is appended to the dataset, and the GPR model is updated.

After sufficient number iterations one can estimate the pdf for q using the approximate map, $q = \bar{T}(z)$, and the statistics of z. The logarithm utilized in the distance 7 provides more accuracy for the tail region. Despite its success in accelerating the convergence for extreme-event statistics, the minimization of the criterion 7 is a computationally expensive problem, especially for high dimensional inputs. To overcome this limitation an upper bound for the criterion 7 is derived in Sapsis (2020). This has the form:

$$\hat{Q}(z^*) = \int \frac{\rho_z(z)}{\rho_n(\bar{T}_{n-1}(z))} \sigma_n^2(z;z^*)dz. \qquad 8.$$

The new form of the $\hat{Q}$ criterion is more practical to compute, as its gradient can be analytically derived under mild conditions, allowing for the employment of gradient optimization methods. To this end, it is applicable in very high-dimensional spaces, such as fluids and waves, and as it is shown in Sapsis (2020) it results in orders of magnitude faster convergence compared with traditional optimal experimental design criteria.

We review results from Mohamad & Sapsis (2018) related to the estimation of the pdf describing the loads on an offshore platform in irregular seas. The response of the platform is quantified through direct, three-dimensional numerical simulations of Navier-Stokes utilizing the smoothed particle hydrodynamics (SPH) method (Crespo et al. 2015), an accurate but computationally very expensive method. The numerical setup parallels that of a physical wave tank experiment and consists of a wave maker on one end and a sloping 'beach' on the other end of the tank to quickly dissipate the energy of incident waves and avoid wave reflections (Fig. 11). Irregular seas follow a JONSWAP spectral density and

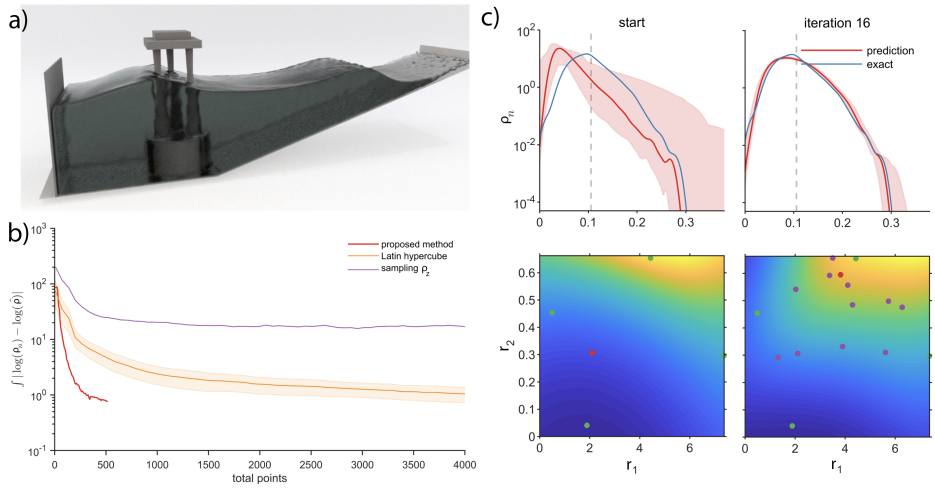

Figure 11

a) SPH simulation of an offshore platform, b) Convergence plots for different methods, c) Initial and subsequent ($n = 16$) estimation of the pdf (and the estimated map $\bar{T}_n(r_1, r_2)$); dots indicate the sampled points. (Reproduced by Mohamad & Sapsis (2018)).

they are parametrized by wavegroups containing two random parameters (Cousins & Sapsis 2016). To this end the input vector is two-dimensional, $z = (r_1, r_2)$.

The quantity of interest in this problem is the force acting on the platform. The incident wave propagates in the x direction (Fig. 11) and as such we consider the pdf of the force in the x direction, F_x:

$$q_f = \max_{t \in [0,\tau]} |F_x(t)|, \qquad 9.$$

where τ is a specified time length. In Fig. 11b,c we show the results for the force variable (details can be found in Mohamad & Sapsis (2018)). The experiment begins by arbitrary selecting four initial sample points from a Latin Hypercube sampling strategy and using those the next sample is selected by minimizing criterion 7. The lightly shaded red region in the pdf plots is a visualization of the uncertainty in the pdf. The figures demonstrate that 15 iterations are sufficient to approximate the pdf to good agreement with the pdf computed with a Monte-Carlo approach. Comparison with two other strategies, the Latin Hypercube and sampling according to the pdf of the parameters, is shown in Fig. 11b.

3.1.2. Probabilistic decomposition synthesis method. While active sampling methods are system-agnostic, one can obtain significant computational advantages when there is a known instability mechanism associated with extreme events. The idea behind the probabilistic decomposition-synthesis (PDS) method (Mohamad et al. 2016; Mohamad & Sapsis 2015) is to rely on an ergodicity assumption for the dynamics of the flow or wave field, and then employ i) a special treatment for the regimes associated with instabilities which are responsible for the heavy-tailed character of the pdf of the observable, and ii) second-order statistics to characterize the 'core' of the pdf. This method was first developed to model first-order prototype systems under multiplicative noise, modeling nonlinear energy transfers in turbulent flows (Mohamad & Sapsis 2015), and then extended for parametric

instabilities in ship rolling under irregular sea states (Mohamad & Sapsis 2016), as well as nonlinear wave equations modeling rogue waves (Mohamad et al. 2016).

We give a brief review of the method; details can be found in Mohamad et al. (2016). An ergodic dynamical system is considered, where it is assumed that the second-order statistics (i.e., a coarse representation of the chaotic background set) is known and that the instability region, R_e (green region in Fig. 3), is also known. Then, the conditional pdf of the observable is obtained, by direct computation of the system trajectories initiated in the unstable region. This conditional pdf, $\rho(q|$ extreme events), describes the statistics of the system during transient instabilities that lead to extreme events. The overall probability of occurrence of the unstable region, $\mathbb{P}_r = \mathbb{P}(\text{extreme events})$, i.e., how frequently the system exhibits extreme events due to the transient instabilities, is computed using the second-order statistics of the system state.

Next, the statistics of the system state u is employed away from the unstable region R_e to obtain the conditional statistics of the observable in the absence of extreme events. This provides the conditional pdf, $\rho(q|\text{regular events})$. Note that this pdf also includes rare occurrences of large magnitude which are not caused by instabilities. The final step is the probabilistic synthesis of this information with a total probability argument:

$$\rho(q) = \rho(q|\text{extreme events})\mathbb{P}_r + \rho(q|\text{regular events})\left(1 - \mathbb{P}_r\right). \qquad 10.$$

The first term expresses the contribution of extreme events due to internal instabilities and it is the heavy-tailed part of the distribution for q. The second term expresses the contribution of the background random set and accounts for the main probability mass in the pdf for q. The decomposition separates statistical quantities according to the total probability law through conditioning on different dynamical regimes. It provides with a partition in terms of the Gaussian 'core' due to the background random set or chaotic attractor and the heavy-tails caused by the intermittent bursts.

In Figure 12 the application of the PDS method to the modified NLS equation (Trulsen et al. 2000) modeling rogue waves is shown. The quantity of interest in this case is the local spatial maxima of the wave field. The wave field is represented as a set of wavegroups with random parameters, amplitude A, and length L. The statistics of these parameters (Fig. 12a) are computed directly from the wave spectrum (second-order information) (Cousins & Sapsis 2016; Farazmand & Sapsis 2017b). Next step is to identify the instability region; this is also done in terms of the two wave-group parameters (Adcock et al. 2012; Cousins & Sapsis 2015; Farazmand & Sapsis 2017b) (upper region defined by red curve in Fig. 12a). Several wavegroups from the instability region are directly simulated using the full equation and the information about extreme waves and regular waves is merged using the total probability law 10. The computed pdf is shown Fig. 12b and it is compared with the pdf obtained using Monte-Carlo sampling, over several standard deviations away from zero. The computational cost of the PDS method is orders of magnitude smaller than Monte-Carlo sampling but it requires an astute understanding of the instability mechanism.

3.2. Dynamics-constrained data-driven methods for extreme event statistics

Data-driven models for extreme events in fluids and waves related problems are significantly improved if information for the dynamics is employed. This information can be obtained either through data-driven methods or imperfect reduced-order models.

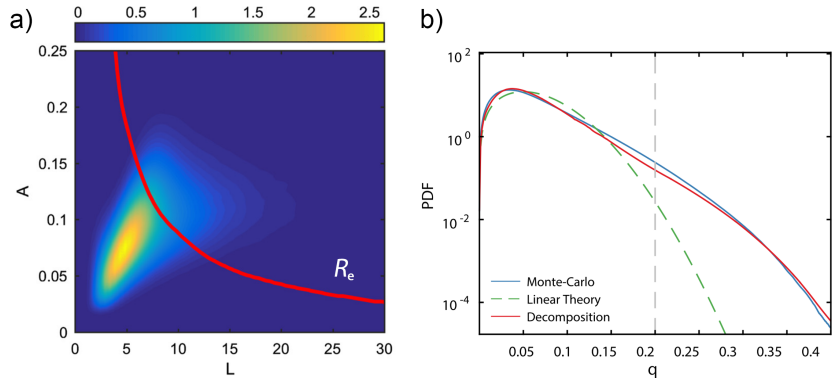

Figure 12

a) Joint pdf of amplitude A and lengthscale L of the random wave-groups that represent the background random set. The red line denotes the unstable region: wave-groups that undergo nonlinear focusing. b) Pdf of the wave elevation maxima using the PDS method and comparison with Monte-Carlo. The dashed line indicates the conditional pdf for regular events and the grey line indicates four standard deviations from the mean. (Reproduced from Mohamad et al. (2016)).

3.2.1. Dynamics represented with data-driven nonlinear maps and linear systems. Several works have focused on extracting dynamical information from data, (Kevrekidis et al. 2015; Wan & Sapsis 2017; Takeishi et al. 2017; Yeung et al. 2017; Li et al. 2017; Vlachas et al. 2018; Lusch et al. 2018; Otto & Rowley 2019; Brunton et al. 2020). Although many of these studies involve systems in fluid mechanics (see also the recent review (Brunton et al. 2020)), very few of these efforts have focused on assessing the capability to extrapolate the pdf tails of an observable, i.e. obtain information for the tails that is not present in the direct statistical estimation of the pdf with the available data. Instead, the focus is given on the short-term predictive capability, rather than the long-term behavior.

Here we review a recent framework that focuses on the statistical extrapolation of the tails by using the data both as a source of information for dynamics, as well as a source of statistics about the system behavior (Arbabi & Sapsis 2019). The core idea is to use optimal transport of probabilities (Parno & Marzouk 2018; Villani 2008) to map the non-Gaussian data distribution to a reference measure, typically chosen to be Gaussian, and then model each dimension separately using a simple stochastic differential equation that mimics the spectrum of the time series.

The first step in this framework is to find the mapping $S : \mathbb{R}^N \to \mathbb{R}^N$ that takes the state variable $u \sim \nu$ to another random variable, denoted as v, which has a normal distribution, that is,

$$v = S(u) \sim \pi, \qquad 11.$$

where π is the normal distribution. To find this change of variables the theory of optimal transport is employed, using the computational approach developed in Parno & Marzouk (2018). The map is assumed to be a lower triangular, differentiable and orientation-

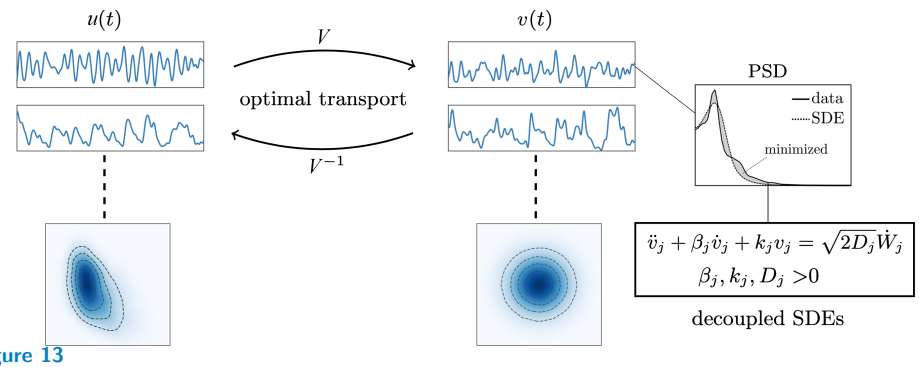

Figure 13

Data-driven framework based on optimal transport and a set of stochastic oscillators. (Reproduced by Arbabi & Sapsis (2019)).

preserving function:

$$S(u_1, u_2, \ldots, u_N) = \begin{bmatrix} S_1(u_1) \\ S_2(u_1, u_2) \\ \vdots \\ S_N(u_1, u_2, \ldots, u_N) \end{bmatrix}, \qquad 12.$$

where each component is approximated with a multivariate polynomial of a prescribed degree. Then the map S is found by minimizing the Kullback-Leibler divergence between the pullback of π under this map, $\tilde{\nu}(y) = \pi(\tilde{S}(y))|\det \nabla \tilde{S}(y)|$, and the data distribution ν:

$$S = \arg\min_{\tilde{S}} \mathbb{E}_\nu \left[\log \frac{\nu(y)}{\tilde{\nu}(y)} \right] = \arg\min_{\tilde{S}} \mathbb{E}_\nu \left[-\log \tilde{\nu}(y) \right], \qquad 13.$$

where $\mathbb{E}$ denotes expectation. This is approximated using the time average of data, assuming ergodicity. The key feature of this computational setup is that it leads to a convex optimization problem which can be solved for each S_i separately, thereby allowing an efficient computation for large dimensions and long time series.

After the map S is found, the next step is to fit a system of stochastic oscillators to the time series of the random variable $v = S(u)$. Since π is chosen to be a multiplicative measure, this fitting is done for each component of q independently. The following system of forced linear oscillators is considered:

$$\ddot{v}_j + \beta_j \dot{v}_j + k_j v_j = \sqrt{2D_j}\dot{W}_j, \quad \beta_j, k_j, D_j > 0, \quad j = 1, \ldots, N, \qquad 14.$$

where W_j's denote standard and mutually independent Brownian motions. Each oscillator admits a stationary density given by

$$\rho_j(v, \dot{v}) = c_j \exp\left\{ -\frac{\beta_j}{D_j} \left(\frac{\dot{v}^2}{2} + k_j \frac{v^2}{2} \right) \right\}, \quad j = 1, \ldots, N, \qquad 15.$$

with c_j being a normalization constant. By setting $D_j = k_j \beta_j$, the displacement of each oscillator follows a normal distribution with unit variance. In order to make the model replicate the *dynamics* of the time series, the free parameters, k_j, β_j are optimized for each system to match the power spectal density (PSD) of v.

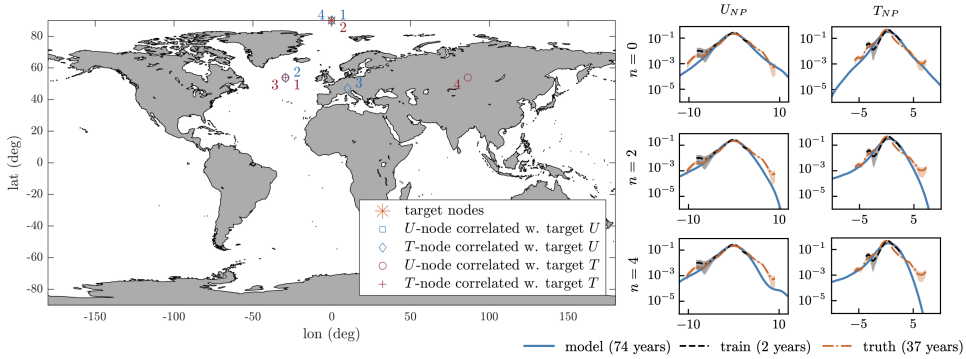

Figure 14

Extrapolation of tails for climate data: Location of covariates used to build models for the target random variables, i.e., u-velocity (U_{NP}) and temperature (T_{NP}) at the North Pole (left). Pdf for U_{NP} and T_{NP} using n of covariates (right). (Reproduced by Arbabi & Sapsis (2019)).

We review results from Arbabi & Sapsis (2019) on climate data based on 6-hourly reanalysis of velocity and temperature in the Earth atmosphere recorded at sigma level 0.95 from 1981 to 2017 (Berrisford et al. 2011). These global fields are expanded in a spherical wavelet basis described in Lessig (2019) and the objective is to extrapolate the pdf tails for the $u-$velocity U_{NP} and temperature T_{NP} at the North-Pole. A sequence of stochastic models is considered for each of the chaotic time series obtained from measurements at the North Pole. In each model, additional degrees of freedom are added from velocity and temperature wavelet coefficients at other locations to capture the statistical dependence between those variables and the target variable. These covariates are chosen from the set of all wavelet coefficients (> 8000 variables) in order of decreasing absolute linear correlation with the target variables.

The geographic position of the covariates is shown in Fig. 14. The training set consisted of the time series from 1981 and 1982, while the rest of the years were used for assessment of the statistical predictions. After discovering each model, a 74-year-long trajectory of the stochastic model was generated and transformed to the space of observations. The pdfs of the training data (2 years), the full reanalysis data (37 years) and the generated data (74 years) are shown in Fig. 14. Although the training data (only 2 years) contains very little information about the tails of the pdf, the reconstructed pdf has very good agreement with the pdf obtained using the full data set (37 years).

3.2.2. Dynamics obtained by exactly solvable reduced-order models or prototype systems. The last class of methods involves problems for which some data is available, as well as an imperfect reduced-order model or analytically solvable prototype system. The amount of data available is typically not sufficient to deduce statistical properties for extreme events and therefore the objective is to combine these two sources of information. The problem has been considered in various contexts involving fluid mechanics problems.

Passive tracer statistics in turbulent fluid flows such as tracer energy spectrum, tracer intermittency, and eddy diffusivity are examined in Qi & Majda (2018) using reduced-order models and data. The reduced-order models are formulated so that they encode the instability mechanism that leads to heavy tails (Majda & Kramer 1999), while their low-order statistical response is exactly solvable (Majda & Gershgorin 2013). This allows

for the calibration of the reduced-order model using a relatively small amount of data and empirical information theory, resulting in good skill on capturing the statistics of extreme events (Qi & Majda 2018). Methods for calibrating reduced-order models for turbulent systems based on nonlinear energy fluxes are discussed in Sapsis & Majda (2013c,b) and Sapsis (2013).

Analytically solvable, low-order, nonlinear models combined with small amounts of data have been successfully demonstrated for the characterization of extreme-event statistics in ship motions and loads (Belenky et al. 2019; Sapsis et al. 2020). In this case the low-order model incorporates information about the ship hull geometry, in particular hydrostatic properties which is an important source of nonlinearity, while data is employed to calibrate the second-order statistics of the obtained pdf. Stochastic models based on asymptotic expansions have also been demonstrated to work well with a small amount of data, on capturing the extreme-event statistics of wave-induced loads acting on offshore wind turbines (Sclavounos et al. 2019). An important open problem in this class involves the quantification of catastrophic events in climate systems, where very little data is available (typically a few decades) while models capture some aspects but they are far from perfect to be used alone.

4. CONCLUSIONS

We have reviewed some of the recent advances in the field of probabilistic quantification for extreme events in fluids and water waves. These include i) model-based methods were an accurate but very expensive model is available, and ii) data-driven methods where some data is available, typically not sufficient to rely on purely statistical methods, and they are combined with dynamical information about the system. These ideas have been demonstrated in several applications ranging from fluid-body interaction problems to rogue waves and turbulent fluids flows.

Several important challenges for extreme events in fluids and waves remain ahead. These include probabilistic quantification of extreme events over long periods (extreme event catalogs) in very large-scale systems, such as geophysical fluids flows over climate scales, where only a small data set exists and this has to be combined with models that are not always accurate. Several ideas involving machine-learning in fluids (Brunton et al. 2020) have shown good promise for such problems, as it has been illustrated in recent works (Wan et al. 2018; Guth & Sapsis 2019; Qi & Majda 2020). However, more advancements are essential in order to deal with the huge number of scales and computational cost. Finally, an important exciting direction involves the prompt prognosis of extreme events (Farazmand & Sapsis 2019b) and their efficient mitigation from the early development stages, i.e. before they have been fully developed. While the problem has been considered in specific setups (Farazmand & Sapsis 2019a) currently there is no general framework that can take into account the strongly transient character of the instabilities associated with these extreme events.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

I would like to thank former and current members of my group, Will Cousins, Mustafa Mohamad, Mohammad Farazmand, Antoine Blanchard, Hassan Arbabi, Zhong Yi Wan, and Stephen Guth for their hard work while working on these topics over the last years. I am particularly grateful to Antoine Blanchard for carefully reading the manuscript and providing several comments that led to improvements. I am indebted to Michael Triantafyllou and Nicholas Patrikalakis who suggested the topic of extreme events in ocean engineering as a research direction, as well as Andrew Majda who has provided important advice over the years. I appreciate several important discussions on these topics with Vadim Belenky and Boyko Dodov. I would like to thank Reza Malek-Madani from the Office of Naval Research (ONR), Samuel Stanton and Matthew Manson from the Army Research Office, Jean-Luc Cambier and Fariba Fahroo from the Air Force Office of Scientific Research who graciously supported my related research as program managers. I am also grateful to the ONR Summer Faculty Program, managed by Jack Price, which facilitated my summer visits at the Naval Surface Warfare Center in Carderock.

LITERATURE CITED

Adcock TAA, Gibbs RH, Taylor PH. 2012. The nonlinear evolution and approximate scaling of directionally spread wave groups on deep water. *Proc. R. Soc. A* 468:2704–2721

Arbabi H, Sapsis TP. 2019. Data-driven modeling of strongly nonlinear chaotic systems with non-Gaussian statistics. *arXiv* :1–32

Aubert A, McKinley R. 2011. Measurements of jet noise aboard US navy aircraft carriers, In *AIAA Centennial of Naval Aviation Forum "100 Years of Achievement and Progress"*

Beck J, Guillas S. 2016. Sequential design with mutual information for computer experiments (MICE): Emulation of a tsunami model. *SIAM-ASA Journal on Uncertainty Quantification* 4:739–766

Belenky V, Glotzer D, Pipiras V, Sapsis TP. 2019. Distribution tail structure and extreme value analysis of constrained piecewise linear oscillators. *Probabilistic Engineering Mechanics* 57:1–13

Belenky VL, Sevastianov NB. 2007. Stability and Safety of Ships: Risk of Capsizing. The Society of Naval Architects and Marine Engineers

Berrisford P, Dee DP, Poli P, Brugge R, Fielding M, et al. 2011. The ERA-Interim archive Version 2.0. *ECMWF Report*

Blonigan PJ, Farazmand M, Sapsis TP. 2019. Are extreme dissipation events predictable in turbulent fluid flows? *Physical Review Fluids* 4:044606

Brès GA, Ham FE, Nichols JW, Lele SK. 2017. Unstructured large-eddy simulations of supersonic jets, In *AIAA Journal*, vol. 55, pp. 1164–1184, American Institute of Aeronautics and Astronautics Inc.

Brunton S, Noack B, Koumoutsakos P. 2020. Machine Learning for Fluid Mechanics. *Ann. Rev. Fluid Mech.* 52:477–508

Chabchoub A, Hoffmann NP, Akhmediev N. 2011. Rogue wave observation in a water wave tank. *Physical Review Letters* 106:204502

Chaloner K, Verdinelli I. 1995. Bayesian experimental design: A review. *Statistical Science* 10:273–304

Chasparis F, Modarres-Sadeghi Y, Hover FS, Triantafyllou MS, Tognarelli M, Beynet P. 2009. Lock-In, Transient and Chaotic Response in Riser VIV, In *Volume 5: Polar and Arctic Sciences and Technology; CFD and VIV*, pp. 479–485, ASMEDC

Cousins W, Onorato M, Chabchoub A, Sapsis TP. 2019. Predicting ocean rogue waves from point measurements: An experimental study for unidirectional waves. *Physical Review E* 99:1–9

Cousins W, Sapsis TP. 2014. Quantification and prediction of extreme events in a one-dimensional nonlinear dispersive wave model. *Physica D* 280:48–58

Cousins W, Sapsis TP. 2015. The unsteady evolution of localized unidirectional deep water wave groups. *Physical Review E* 91:063204

Cousins W, Sapsis TP. 2016. Reduced order precursors of rare events in unidirectional nonlinear water waves. *Journal of Fluid Mechanics* 790:368–388

Crespo AJ, Domínguez JM, Rogers BD, Gómez-Gesteira M, Longshaw S, et al. 2015. DualSPHysics: Open-source parallel CFD solver based on Smoothed Particle Hydrodynamics (SPH). *Computer Physics Communications* 187:204–216

Davini P, Cagnazzo C, Gualdi S, Navarra A. 2012. Bidimensional diagnostics, variability, and trends of northern hemisphere blocking. *Journal of Climate* 25:6496–6509

DelSole T. 2004. Stochastic models of quasigeostrophic turbulence. *Surv. Geophys.* 25:107–149

Dematteis G, Grafke T, Vanden-Eijnden E. 2018. Rogue waves and large deviations in deep sea. *Proceedings of the National Academy of Sciences*

Dematteis G, Grafke T, Vanden-Eijnden E. 2019. Extreme event quantification in dynamical systems with random components. *SIAM-ASA Journal on Uncertainty Quantification* 7:1029–1059

Dowell EH. 2015. A modern course in aeroelasticity. Springer International Publishing Switzerland

Dysthe K, Krogstad HE, Müller P. 2008. Oceanic Rogue Waves. *Annu. Rev. Fluid Mech.* 40:287

Farazmand M, Sapsis TP. 2016. Dynamical indicators for the prediction of bursting phenomena in high-dimensional systems. *Physical Review E* 032212:1–31

Farazmand M, Sapsis TP. 2017a. A variational approach to probing extreme events in turbulent dynamical systems. *Science Advances* 3:e1701533

Farazmand M, Sapsis TP. 2017b. Reduced-order prediction of rogue waves in two-dimensional deep-water waves. *Journal of Computational Physics* 340:418–434

Farazmand M, Sapsis TP. 2019a. Closed-loop adaptive control of extreme events in a turbulent flow. *Physical Review E* 100:033110

Farazmand M, Sapsis TP. 2019b. Extreme Events: Mechanisms and Prediction. *Applied Mechanics Reviews* 71:1–19

Fedele F, Tayfun MA. 2009. On nonlinear wave groups and crest statistics. *Journal of Fluid Mechanics* 620:221–239

Flores O, Jiménez J. 2010. Hierarchy of minimal flow units in the logarithmic layer. *Physics of Fluids* 22:1–4

Forristall GZ. 2000. Wave Crest Distributions: Observations and Second-Order Theory. *Journal of Physical Oceanography* 30:1931–1943

Guth S, Sapsis TP. 2019. Machine learning predictors of extreme events occurring in complex dynamical systems. *Entropy* 21

Huang NE, Bliven LF, Long SR, Tung CC. 1986. An analytical model for oceanic whitecap coverage

Ishihara T, Kaneda Y, Yokokawa M, Itakura K, Uno A. 2007. Small-scale statistics in high-resolution direct numerical simulation of turbulence: Reynolds number dependence of one-point velocity gradient statistics. *Journal of Fluid Mechanics* 592:335–366

Iyer CO, Ceccio SL. 2002. The influence of developed cavitation on the flow of a turbulent shear layer. *Physics of Fluids* 14:3414–3431

Jiménez J. 2012. Cascades in wall-bounded turbulence. *Annual Review of Fluid Mechanics* 44:27–45

Jiménez J. 2018. Coherent structures in wall-bounded turbulence. *Journal of Fluid Mechanics* 842:P1

Jiménez J, Moin P. 1991. The minimal flow unit in near-wall turbulence. *Journal of Fluid Mechanics* 225

Jiménez J, Simens MP. 2001. Low-dimensional dynamics of a turbulent wall flow. *Journal of Fluid Mechanics* 435:81–91

Kevrekidis I, Rowley CW, Williams M. 2015. A kernel-based method for data-driven Koopman spectral analysis. *Journal of Computational Dynamics* 2:247–265

Kim J, Moin P, Moser R. 1987. Turbulence statistics in fully developed channel flow at low Reynolds number. *Journal of Fluid Mechanics* 177:133–166

Lessig C. 2019. Divergence free polar wavelets for the analysis and representation of fluid flows. *Journal of Mathematical Fluid Mechanics* 21:18

Lestang T, Bouchet F, Lévêque E. 2019. Rare-event sampling applied to the simulation of extreme mechanical efforts exerted by a turbulent flow on a bluff body. *arXiv*

Li Q, Dietrich F, Bollt EM, Kevrekidis IG. 2017. Extended dynamic mode decomposition with dictionary learning: A data-driven adaptive spectral decomposition of the Koopman operator. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 27:103111

Longuet-Higgins MS. 1963. The effect of non-linearities on statistical distributions in the theory of sea waves. *Journal of Fluid Mechanics* 17:459

Lucarini V, Faranda D, Wouters J. 2012. Universal behaviour of extreme value statistics for selected observables of dynamical systems. *Journal of Statistical Physics* 147:63–73

Lucarini V, Faranda D, Wouters J, Kuna T. 2014. Towards a general theory of extremes for observables of chaotic dynamical systems. *Journal of Statistical Physics* 154:723–750

Lusch B, Kutz JN, Brunton SL. 2018. Deep learning for universal linear embeddings of nonlinear dynamics. *Nature communications* 9:4950

Majda AJ, Abramov RV, Grote MJ. 2005. Information Theory and Stochastics for Multiscale Nonlinear Systems, vol. 25 of *CRM Monograph Series*. American Mathematical Society

Majda AJ, Branicki M. 2012. Lessons in Uncertainty Quantification for Turbulent Dynamical Systems. *Discrete and Continuous Dynamical Systems* 32:3133–3221

Majda AJ, Gershgorin B. 2013. Elementary models for turbulent diffusion with complex physical features: Eddy diffusivity, spectrum and intermittency. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 371

Majda AJ, Harlim J. 2012. Filtering Complex Turbulent Systems. Cambridge University Press

Majda AJ, Kramer PR. 1999. Simplified models for turbulent diffusion: Theory, numerical modelling, and physical phenomena. *Physics reports* 314:237–574

Majda AJ, McLaughlin DW, Tabak EG. 1997. A one-dimensional model for dispersive wave turbulence. *Journal of Nonlinear Science* 7:9–44

Majda AJ, Moore MN, Qi D. 2019. Statistical dynamical model to predict extreme events and anomalous features in shallow water waves with abrupt depth change. *Proceedings of the National Academy of Sciences of the United States of America* 116:3982–3987

Majda AJ, Qi D. 2019. Statistical Phase Transitions and Extreme Events in Shallow Water Waves with an Abrupt Depth Change. *Journal of Statistical Physics*

Majda AJ, Qi D, Sapsis TP. 2014. Blended particle filters for large-dimensional chaotic dynamical systems. *Proceedings of the National Academy of Sciences* 111:7511–7516

Mandre S, Mahadevan L. 2010. A generalized theory of viscous and inviscid flutter. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 466:141–156

Mezić I. 2013. Analysis of fluid flows via spectral properties of the Koopman operator. *Annual Review of Fluid Mechanics* 45:357–378

Modarres-Sadeghi Y, Chasparis F, Triantafyllou MS, Tognarelli M, Beynet P. 2011. Chaotic response is a generic feature of vortex-induced vibrations of flexible risers. *Journal of Sound and Vibration* 330:2565–2579

Modarres-Sadeghi Y, Mukundan H, Dahl JM, Hover FS, Triantafyllou MS. 2010. The effect of higher harmonic forces on fatigue life of marine risers. *Journal of Sound and Vibration* 329:43–55

Mohamad MA, Cousins W, Sapsis TP. 2016. A probabilistic decomposition-synthesis method for the quantification of rare events due to internal instabilities. *Journal of Computational Physics* 322:288–308

Mohamad MA, Sapsis TP. 2015. Probabilistic description of extreme events in intermittently unstable dynamical systems excited by correlated stochastic processes. *SIAM/ASA Journal on Uncertainty Quantification* 3:709–736

Mohamad MA, Sapsis TP. 2016. Probabilistic response and rare events in Mathieu's equation under correlated parametric excitation. *Ocean Engineering Journal* 120:289–297

Mohamad MA, Sapsis TP. 2018. Sequential sampling strategy for extreme event statistics in nonlinear dynamical systems. *Proceedings of the National Academy of Sciences* 115:11138–11143

Naess A, Moan T. 2012. Stochastic Dynamics of Marine Structures. Cambridge University Press

Ochi M, Wang S. 1976. Prediction of Extreme Wave-Induced Loads on Ocean Structures, In *Proceedings of the Symposium on Behavior of Offshore Structures*, pp. 170–186

Osborne AR. 2001. The random and deterministic dynamics of 'rogue waves' in unidirectional, deep-water wave trains. *Marine Structures* 14:275–293

Otto SE, Rowley CW. 2019. Linearly recurrent autoencoder networks for learning dynamics. *SIAM Journal on Applied Dynamical Systems* 18:558–593

Pandita P, Bilionis I, Panchal J. 2019. Bayesian Optimal Design of Experiments For Inferring The Statistical Expectation Of A Black-Box Function. *ASME J. Mechanical Design* 141:101404

Parno MD, Marzouk YM. 2018. Transport map accelerated Markov chain Monte Carlo. *SIAM/ASA Journal on Uncertainty Quantification* 6:645–682

Pedlosky J. 1998. Ocean Circulation Theory. Springer-Verlag

Qi D, Majda AJ. 2018. Predicting extreme events for passive scalar turbulence in two-layer baroclinic flows through reduced-order stochastic models. *Communications in Mathematical Sciences* 16:17–51

Qi D, Majda AJ. 2020. Using machine learning to predict extreme events in complex systems. *Proceedings of the National Academy of Sciences of the United States of America* 117:52–59

Ragone F, Bouchet F. 2019. Computation of extreme values of time averaged observables in climate models with large deviation techniques. *Journal of Statistical Physics* :1–29

Salmon R. 1998. Lectures on Geophysical Fluid Dynamics. Oxford University Press

Sapsis T, Pipiras V, Weems K, Belenky V. 2020. On Extreme Value Properties of Vertical Bending Moment, In *33rd Symposium on Naval Hydrodynamics, Osaka, Japan*

Sapsis TP. 2013. Attractor local dimensionality, nonlinear energy transfers, and finite-time instabilities in unstable dynamical systems with applications to 2D fluid flows. *Proceedings of the Royal Society A* 469:20120550

Sapsis TP. 2018. New perspectives for the prediction and statistical quantification of extreme events in high-dimensional dynamical systems. *Phil. Trans. R. Soc. Lond. A* 376:20170133

Sapsis TP. 2020. Output-weighted optimal sampling for Bayesian regression and rare event statistics using few samples. *Proc. R. Soc. A* 476:20190834

Sapsis TP, Majda AJ. 2013a. A statistically accurate modified quasilinear Gaussian closure for uncertainty quantification in turbulent dynamical systems. *Physica D* 252:34–45

Sapsis TP, Majda AJ. 2013b. Blending Modified Gaussian Closure and Non-Gaussian Reduced Subspace methods for Turbulent Dynamical Systems. *Journal of Nonlinear Science* 23:1039

Sapsis TP, Majda AJ. 2013c. Statistically accurate low-order models for uncertainty quantification in turbulent dynamical systems. *Proceedings of the National Academy of Sciences of the United States of America* 110

Schmid PJ. 2010. Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics* 656:5–28

Schmidt OT, Schmid PJ. 2019. A conditional space–time POD formalism for intermittent and rare events: example of acoustic bursts in turbulent jets. *Journal of Fluid Mechanics* 867:R2

Sclavounos PD, Zhang Y, Ma Y, Larson DF. 2019. Offshore wind turbine nonlinear wave loads and their statistics. *Journal of Offshore Mechanics and Arctic Engineering* 141

Secon H. 2020. California could go through all of February without a significant storm for the first time since 1864. A photo from space reveals why. *Business Insider*

Shin Y, Belenky V, Weems K, Lin W, Engle A. 2003. Nonlinear time domain simulation technology for seakeeping and wave-load analysis for modern ship design. *Transactions, Society of Naval Architects and Marine Engineers* 111

St Denis M, Pierson W. 1953. On the motion of ships in confused seas. *Transactions, Society of Naval Architects and Marine Engineers* 61

Stefanakis TS, Contal E, Vayatis N, Dias F, Synolakis CE. 2014. Can small islands protect nearby coasts from tsunamis? An active experimental design approach. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 470

Takeishi N, Kawahara Y, Yairi T. 2017. Learning Koopman invariant subspaces for dynamic mode decomposition, In *Advances in Neural Information Processing Systems*, pp. 1130–1140

Tayfun MA. 1980. Narrow-band nonlinear sea waves. *Journal of Geophysical Research* 85:1548

Tibaldi S, Molteni F. 1990. On the operational predictability of blocking

Trulsen K, Dysthe KB. 1996. A modified nonlinear Schrödinger equation for broader bandwidth gravity waves on deep water. *Wave motion* 24:281–289

Trulsen K, Kliakhandler I, Dysthe KB, Velarde MG. 2000. On weakly nonlinear modulation of waves on deep water. *Phys. Fluids* 12:2432

Tsuji Y, Fransson JH, Alfredsson PH, Johansson AV. 2007. Pressure statistics and their scaling in high-Reynolds-number turbulent boundary layers. *Journal of Fluid Mechanics* 585:1–40

Tsuji Y, Imayama S, Schlatter P, Alfredsson PH, Johansson AV, et al. 2012. Pressure fluctuation in high-Reynolds-number turbulent boundary layer: Results from experiments and DNS. *Journal of Turbulence* 13:1–19

Tsuji Y, Ishihara T. 2003. Similarity scaling of pressure fluctuation in turbulence. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics* 68:5

Varadhan SRS. 1984. Large Deviations and Applications. SIAM

Varadhan SRS. 2008. Special invited paper: Large deviations. *The Annals of Probability* 36:397

Villani C. 2008. Optimal transport: old and new, vol. 338. Springer Science & Business Media

Vlachas PR, Byeon W, Wan ZY, Sapsis TP, Koumoutsakos P. 2018. Data-driven forecasting of high-dimensional chaotic systems with long-short term memory networks. *Proceedings of the Royal Society A* 474:20170844

Wan Z, Sapsis TP. 2017. Reduced-space Gaussian Process regression for data-driven probabilistic forecast of chaotic dynamical systems. *Physica D: Nonlinear Phenomena* 345

Wan ZY, Vlachas PR, Koumoutsakos P, Sapsis TP. 2018. Data-assisted reduced-order modeling of extreme events in complex dynamical systems. *PLoS ONE* 13

Webber RJ, Plotkin DA, O'Neill ME, Abbot DS, Weare J. 2019. Practical rare event sampling for extreme mesoscale weather. *Chaos* 29

Xiao W, Liu Y, Wu G, Yue DKP. 2013. Rogue wave occurrence and dynamics by direct simulations of nonlinear wave-field evolution. *Journal of Fluid Mechanics* 720:357–392

Yeung E, Kundu S, Hodas N. 2017. Learning Deep Neural Network Representations for Koopman Operators of Nonlinear Dynamical Systems. *arXiv preprint arXiv:1708.06850*

Yeung PK, Zhai XM, Sreenivasan KR. 2015. Extreme events in computational turbulence. *Proceedings of the National Academy of Sciences* 112:12633–12638

Zakharov VE. 1968. Stability of periodic waves of finite amplitude on the surface of a deep fluid. *Journal of Applied Mechanics and Technical Physics* 9:190–194

Zhang J, Childress S, Libchaber A, Shelley M. 2000. Flexible filaments in a flowing soap film as a model for one-dimensional flags in a two-dimensional wind. *Nature* 408:835–839