

# *Applications of large deviation theory in geophysical fluid dynamics and climate science*

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Galfi, V. M., Lucarini, V. ORCID: <https://orcid.org/0000-0001-9392-1471>, Ragone, F. and Wouters, J. ORCID: <https://orcid.org/0000-0001-5418-7657> (2021) Applications of large deviation theory in geophysical fluid dynamics and climate science. *Rivista del Nuovo Cimento*, 44. pp. 291-363. ISSN 1826-9850 doi: 10.1007/s40766-021-00020-z Available at <https://centaur.reading.ac.uk/97624/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1007/s40766-021-00020-z>

Publisher: Springer

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online



# Applications of large deviation theory in geophysical fluid dynamics and climate science

Vera Melinda Gálfi<sup>1,2</sup> · Valerio Lucarini<sup>3,4</sup> · Francesco Ragone<sup>5,6</sup> · Jeroen Wouters<sup>3,4</sup>

Received: 19 February 2021 / Accepted: 19 April 2021  
© The Author(s) 2021

## Abstract

The climate is a complex, chaotic system with many degrees of freedom. Attaining a deeper level of understanding of climate dynamics is an urgent scientific challenge, given the evolving climate crisis. In statistical physics, many-particle systems are studied using Large Deviation Theory (LDT). A great potential exists for applying LDT to problems in geophysical fluid dynamics and climate science. In particular, LDT allows for understanding the properties of persistent deviations of climatic fields from long-term averages and for associating them to low-frequency, large-scale patterns. Additionally, LDT can be used in conjunction with rare event algorithms to explore rarely visited regions of the phase space. These applications are of key importance to improve our understanding of high-impact weather and climate events. Furthermore, LDT provides tools for evaluating the probability of noise-induced transitions between metastable climate states. This is, in turn, essential for understanding the global stability properties of the system. The goal of this review is manifold. First, we provide an introduction to LDT. We then present the existing literature. Finally, we propose possible lines of future investigations. We hope that this paper will prepare the ground for studies applying LDT to solve problems encountered in climate science and geophysical fluid dynamics.

---

✉ Vera Melinda Gálfi  
vera.melinda.galfi@geo.uu.se

<sup>1</sup> Department of Earth Sciences, Uppsala University, Uppsala, Sweden

<sup>2</sup> Meteorological Institute, CEN, University of Hamburg, Hamburg, Germany

<sup>3</sup> Department of Mathematics and Statistics, University of Reading, Reading, UK

<sup>4</sup> Centre for the Mathematics of Planet Earth, University of Reading, Reading, UK

<sup>5</sup> Georges Lemaître Centre for Earth and Climate Research, Earth and Life Institute, Université catholique de Louvain, Louvain-la-Neuve, Belgium

<sup>6</sup> Royal Meteorological Institute of Belgium, Brussels, Belgium

**Keywords** Large deviation theory · Climate system · Geophysical fluid dynamics · Low-frequency variability · Extreme events · Rare event algorithms · Metastable states · Noise-induced transitions · Instantons · Heatwaves · Cold Spells · Rogue Waves

## 1 Introduction and motivation

### 1.1 The climate crisis: extreme events in a changing climate

The climate is a forced and dissipative nonlinear heterogeneous system composed by several subdomains, namely the atmosphere, the hydrosphere, the cryosphere, the soil, and the biosphere. The climate evolves under the action of a primary forcing given by the incoming solar radiation and modulating factors such as the atmospheric composition, the optical properties of the surface of the planet, gravity, and the rotation of the Earth around its vertical axis. Each of these subsystems features complex nonlinear physical and chemical processes, and the various subsystems interact among themselves through exchanges of energy, momentum, and chemical species. As a result of the interplay between forcing, dissipation, and internal nonlinear dynamics, the climate system features variability of a vast range of spatial and temporal scales. The climate can be seen as a prominent example of non-equilibrium system where an approximate steady state is reached as the inhomogeneous absorption of solar radiation occurring throughout its domain is compensated by a variety of physical mechanisms, including thermal emission in the infrared and complex patterns of transport of sensible and latent heat. Lorenz [172] provided a first comprehensive theory of the dynamics and thermodynamics of climate able to bring together the main mechanisms of forcing, energy conversion, and dissipation. The large-scale flows of the ocean and of the atmosphere ultimately result from the conversion of available potential into kinetic energy performed by the climatic engine. The conversion takes place through various mechanisms of instability fuelled by the presence of spatial temperature gradients. Such instabilities allow for energy conversion between the background state and the fluctuations of the climatic field and lead to chaotic conditions that are associated with heterogeneous turbulence in the geophysical fluids. Additionally, these instabilities establish negative feedbacks, because they tend to reduce, via transport and mixing, the temperature gradients that support them. See [174,204] for an extensive discussion of these mechanism. We remark that an exact steady state is never achieved because of the fluctuations in the incoming solar radiation and in the processes, both natural and anthropogenic, that alter the atmospheric composition and the surface of the planet [104,250].

Improving our understanding of the dynamical and statistical properties of the climate system, of the links between its response to anthropogenic and natural forcings and of its natural variability is key to provide scientific tools for anticipating, predicting, and possibly addressing the ongoing *climate crisis*. The current popularity of the expression *climate crisis* as opposed to—the more usual one—*climate change* is motivated by the desire to focus on the understanding of how changing climate conditions will unfold as variations in the higher moments of the distribution of the

climatic variables, able to better capture the properties of extreme events and tipping points [104].

Indeed, the study of extreme events is essential for addressing the natural hazards associated with climate variability and climate change and affecting in a potentially catastrophic way human and environmental welfare. As the resilience of any system and the incurred damages due to an unusual, high-impact event change drastically when certain thresholds in the intensity and/or in the duration of the hazard are reached, it is clear that understanding the fate of extreme events in the context of the changing climate is essential for accurately factoring in future losses and damages and prepare for them [133]. High-impact events affecting human and environmental welfare are sometimes associated with the presence of long temporal persistence of a large anomaly in the field of interest, as resilience—the ability of any system to resist—against anomalous environmental conditions does not last indefinitely [76,146,214]. Meteo-climatic extremes characterised by persistence are usually referred to as slow onset events, as opposed to the fast onset ones, which, instead, are associated with fast processes. Prominent examples of slow onset events are droughts, heatwaves, and cold spells, while flash floods and intense snowfalls lead to hazards in the category of fast onset events [133]. We remark that it might be worth considering a revision of such a classical terminology, because of the multiscale nature of some of the so-called slow onset events: the onset of an atmospheric blocking responsible for a heatwave takes usually up to a couple of days, while its duration can be much much longer, and ranging up to several weeks. Also the exit from long lasting events can be quite rapid and, anyhow, considerably shorter than their active phase [254,255].

An important remark is also needed regarding the problem of defining persistent extreme events; we refer here, for the sake of clarity, to the specific case of heatwaves. While the general understanding is that a heatwave is a period of extreme and unusual warmth, there is no rigorous nor commonly accepted definition for it in terms of intensity and persistence of the anomalous weather conditions, despite several attempts in this direction; see, e.g. [223]. As noted in [206], *...it seems that almost, if not every climatological study that looks at heatwaves uses a different metric...* see also [183,243] for a discussion on the lack of consensus for a shared definition of heatwaves. The confusion around the definition of heatwave has serious implications as it hinders attempts at mitigating their impacts [38,195,216].

Changing climate conditions can lead to dramatic changes in the statistics of extreme events. As mentioned above, this is one of the main manifestations of the climate crisis. In the future climate, changes in the statistics of heatwaves are worrying, as more persistent and larger temperature fluctuations are possible as a result of changes in the properties of the low frequency variability of the atmosphere and of the properties of the soil. This effect compounds with the trend in the average temperature, leading to a greatly increased risk of such catastrophic events [50,210,238]. Climate change additionally leads to a reduced winter weather variability, as a result of the reduction in the temperature difference between low and high latitudes [250]. Therefore, the probability of occurrence of cold spells is likely to greatly decrease [242], even if structural changes in the dynamics of the atmosphere can rarely create special conditions that facilitate their occurrence [46,150]. Looking instead at flash floods, consensus exists that they will become more likely and more intense in the future because the higher

retention of water vapour in the atmosphere—made possible by warmer conditions as a result of the Clausius-Clapeyron relation—compounds with strengthened convective motions, possibly leading to disproportionately enhanced extreme precipitation events in specific locales [88,279]. However, we remark that, due to the complexity of the (microscopic and macroscopic) dynamical and thermodynamical processes involved, the precipitation is not distributed in space and time in direct proportion to the available precipitable water [278].

A somewhat separate research agenda tries, instead, to relate in a direct way climate change and *individual* extreme events, thus blurring the distinction between climate and weather, statistics and analysis of specific case studies. Following the landmark paper [5], considerable efforts have been directed at developing tools for assessing to what extent climate change has impacted and is impacting either the frequency, or the likelihood, or the intensity of *individual* extreme events, with the goal of providing the basis for science-based liability for the impacts of such extremes. The scientific debate around extreme events attribution has relevant implications in terms of climate adaptation, risk assessment, public policy, infrastructural design, insurance instruments design, international relations, and even migration policies [135,199,239,251].

## 1.2 Quest for universality of extreme events

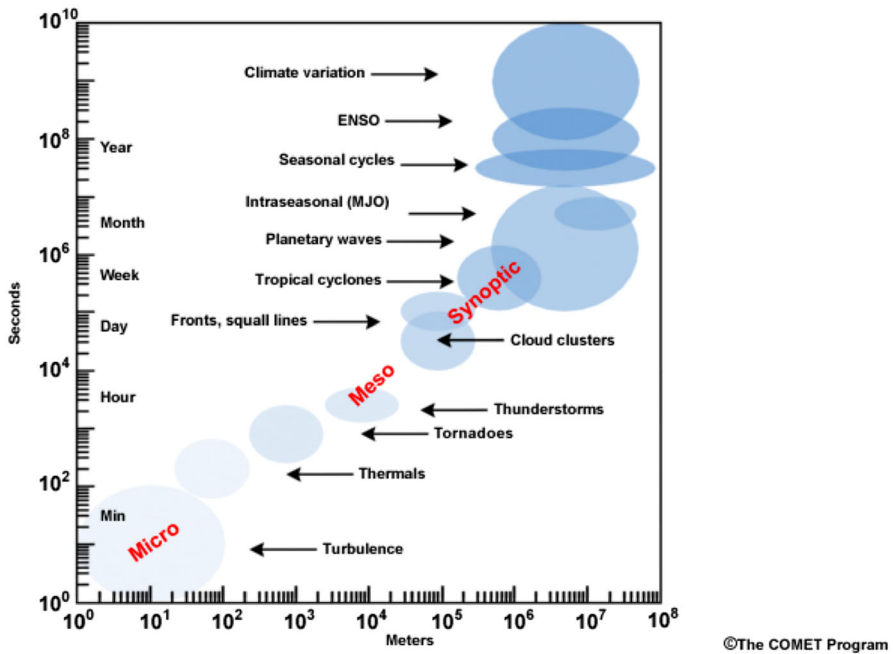
Empirical frequentist approaches aimed at the study of extremes applied to actual climate records or to the output of climate models are essential for keeping track of the observed events, but face the unavoidable problem of being unable to say anything about the probability of occurrence of out-of-sample events. In order to achieve predictive power in a statistical sense—i.e. being able to estimate the probability of occurrence of events that are more extreme than the observed ones—one needs to interpret data through mathematical approaches that provide some form of universality.

Extreme Value Theory (EVT) provides a powerful framework for studying extreme events in a multitude of applications. It is based on limit theorems mimicking the central limit theorem that allow one, under rather general hypotheses and taking suitable limits, to define universal laws describing the probability of occurrence of events generated according to a given stochastic process above a sufficiently high threshold. Alternatively, one can develop the theory in order to describe the distribution of the maximum of a set of independent and identically distributed stochastic variables in the limit of large sets [47]. The theory can be adapted for dealing with correlations between the variables [161] and for treating the outputs of chaotic dynamical systems [178], in such a way that deep connections emerge between the statistical properties of suitably defined extremes and the geometry of the attractor of the system [18,95,174,213]. EVT has received a great deal of attention in geosciences [83,91,106,265,270,273,280] and is extremely influential especially in hydrology [138,139,158]. In the context of climate dynamics, the analysis of extremes has proved very fruitful for providing a new viewpoint for understanding atmospheric predictability by looking at the recurrence of weather patterns [82,184]. Persistence, as mentioned above, is a key factor in determining the impact of large climatic fluctuations. EVT can deal

with time correlations in time series through the introduction of the extremal index, which allows one to quantify the average size of clusters of consecutive extreme events [87,187]. The extremal index encodes important information on the dynamics of the system [36].

A more direct line to attack the problem of studying persistent extreme events can be taken through the use of Large Deviation Theory (LDT) [263]. In a nutshell, in one of its most basic formulations, LDT aims at providing limit laws for the average of  $n$  (typically identically distributed) stochastic variables, where  $n$  is large. Similarly to the case of EVT, a unified approach for LDT can be used on stochastic processes and chaotic dynamical systems [141,283]. It is hard to overestimate the importance of LDT in contemporary physics and mathematics [62,63,257,267]. Establishing a large deviation principle for an observable—see below—leads to gaining predictive power of the process. While in EVT such a power is aimed at being able to predict the probability of occurrence of events larger than those already observed, in the case of LDT the predictive power is twofold, as it is directed towards predicting the property of occurrence of events that are larger and/or more persistent than the observed ones. Drawing an example from climate science, EVT is better suited for studying the probability of occurrence of extremely hot days, whereas LDT is better suited for studying the probability of occurrence of heatwaves.

A fascinating aspect of looking at the properties of long time-averages of climatic fields is the following. The theory of low-frequency variability of the atmosphere indicates that long temporal persistence and large spatial extent of the anomalous patterns go hand in hand [104,105]; see Fig. 1. In the mid-latitudes, it is customary and indeed scientifically meaningful to distinguish between synoptic variability, due to mid-latitude eastward-moving weather systems and associated with temporal scales of 3–7 days and spatial scales of the order of 1000 Km, and low-frequency variability, whose temporal and spatial scales are typically larger, amounting to 1–3 months and several thousands of Km, respectively. [105,246]. The main manifestations of low-frequency variability in the mid-latitudes are the so-called blocking events, which are persistent, large-scale departures from the approximately zonally symmetric flow associated with the presence of large-amplitude, almost-stationary pressure anomalies [79,105,105,169,180,246,254,255]. The difference between synoptic and low-frequency variability is clarified when performing a spectral analysis of the atmospheric fields: the former is associated with eastward propagating waves, while the latter is characterised by stationary or weakly propagating planetary waves [58]. Persistence is key to creating conditions conducive to long-lasting extreme events, and, indeed, it is well-known that the anomalies of the flow due to occurrence of blockings can lead to long-duration warm [67] as well as cold extreme events [34]. Given their long time duration and large spatial extent, blockings can lead, in a cascade process, to the onset of extreme events also at considerable geographic distance from the core of the blocking, as in the case of the summer 2010 floods in Pakistan resulting from the large scale flow associated with the blocking—and ensuing heatwave—in Russia [159]. Advancing our understanding of the low-frequency variability of the atmosphere would be very beneficial because, despite continuous improvements, our ability to perform accurate extended-range (beyond 7–10 days) weather forecast in the mid-latitudes is still limited [105], and because attaining a convincing representation



**Fig. 1** Idealised power spectra for the atmosphere indicating the relationship between the spatial and temporal scales of atmospheric flows. The source of this material is the COMET<sup>®</sup> Website at <http://meted.ucar.edu/> of the University Corporation for Atmospheric Research (UCAR), sponsored in part through cooperative agreement(s) with the National Oceanic and Atmospheric Administration (NOAA), U.S. Department of Commerce (DOC). ©1997–2017 University Corporation for Atmospheric Research. All Rights Reserved

of the statistical and dynamical properties of blocking events is still challenging for both numerical weather forecast models [85] and climate models [54].

The hope is that, by focusing on suitably defined large deviations of the atmospheric fields, one could distil information on the low-frequency variability of the atmosphere. Roughly speaking, as discussed below, it can be proven rigorously that *any large deviation is realised in the least unlikely of all the unlikely ways* [63]. Let's clarify this important concept using again an example drawn from climate science. Let's assume that we have established a large deviation law describing the probability of occurrence of heatwaves in a given location. In principle, the corresponding rare events can take place as a result of a variety of large scale atmospheric configurations; see a recent analysis of heatwaves in France [7]. Nonetheless, LDT imposes that, in fact, if we look at *true* extremes, with overwhelming probability the heatwaves we observe will take place, apart from small-scale spatio-temporal fluctuations, as a result of a well-defined large-scale atmospheric configuration, which is very rare in the standard statistics, but is *typical* if we consider the multitude of possible heatwaves with same intensity. By typical here we mean that the probability of the occurrence of a large scale atmospheric pattern that is very close to such a configuration, conditional on the occurrence of heatwave at the reference location, is very high, and gets closer to one as we consider more stringent criteria - in terms of intensity and duration - for



the occurrence of (rarer) heatwaves. In dynamical terms, one has that selecting events associated with large deviations amounts to considering a very small portion of the phase space. The property above implies that the (rarely occurring) approach to this very special region overwhelmingly occurs through a well-defined set of paths, that are singled out by LDT, even if much unlikelier paths are still possible.

Indeed, looking at the specific case of the catastrophic 2010 Russian heatwave, one does find that the observed extreme event is in some sense typical [67,94]. This does not exclude the possibility of more exotic atmospheric configurations on the scale of Eurasia, but their occurrence is much more unlikely than those, already extremely rare, described by LDT. These exotic events might be interpreted as *dragon kings* [245].

Of course the possibility of practically using LDT in a complex and multiscale system like the climate is far from being an obvious task for all possible climatic observables. The mathematical foundations for using LDT in the context of climate lay on taking into account, on the one hand, its chaoticity, and, on the other hand, the fact that stochastic effects emerge as a result of considering its coarse-grained evolution. Indeed, most of the results we present below are a natural extension of the scientific programme aimed at developing and analysing stochastic climate models pioneered by Hasselmann [125]; see later developments in [131,132,181,205,233]. Additionally, one needs to take into account that while most LDT results require stationarity of the time series, the climate system is only approximately stationary, because of the periodicity in the solar forcing and the natural and anthropogenic forcings to the atmospheric composition (e.g. change in greenhouse gases and in aerosols) and to the properties of the land surface (e.g. forest fires; agriculture; deforestation). Therefore, one might need to pre-process the data (e.g. removing the seasonal cycle; removing trends) before being able to apply LDT. Clearly, since the climate is a nonlinear system, the previous pre-processing aimed at removing part of the time-dependence is in principle partly arbitrary and definitely non uniquely defined. Nonetheless, one needs to resort to reasonable pragmatism in treating observational or model-generated data that do not conform exactly to the demands of the mathematical theory, and possibly derive nonetheless useful information, as often in fact done in physical sciences.

Another aspect to be kept in consideration is the presence of serial correlations in the time series of the observables. If one considers, for example, the serial correlation of the anomalies of the surface temperature (obtained after removing seasonal cycle and long-term trends) somewhere in the middle of a continent, like Central Europe, and the serial correlation of the same observable over an oceanic region, like the North Atlantic (not far away from the first location), one would notice that the strength of the serial correlation is much weaker and the auto-correlation function decays substantially faster over the continent as compared to the oceanic region. In the latest case, the decay of correlations will be slower than exponential (at least on a vast range of scales), as a result of the presence of long-term memory in the system. Large differences in the heat capacity of land surface vs water, and the dynamical link between surface waters and deeper levels of the ocean explain such a discrepancy between the two cases. The fact that the same climatic field - anomalies of the surface temperature - features such fundamentally different properties, in terms of stochasticity, depending on the geographical location of interest provides a good example of the complexity of the climate system. Note that, as we will discuss below, while in the former case one is able

to establish large deviation laws to describe accurately long and persistent temperature fluctuations behind heatwaves and cold spells, LDT will not apply in the latter case.

### 1.3 Paths and transitions

LDT can be used for different scopes than looking at persistent deviation of fields. Indeed, it provides tools for studying how such special configurations of the climate are dynamically realised. One can use a more general definition of events that encompasses trajectories in the phase space, and adapt LDT to study rare trajectories leading to target extreme events. In this settings the dynamical equations contain a small parameter, describing either a weak noise strength or an inverse time scale separation. Under such conditions the path probabilities collapse onto one single path as the small parameter goes to zero, either the deterministic zero-noise path for weak noise systems or the averaged equation for system with a time scale separation. Also here the principle holds that the unlikely event is reached in the least unlikely way. Such paths, called instantons, can be seen as minimisers of an action describing the cost of going against the natural tendency of the system [93]. Take, for example, a particle in a double well potential with weak noise. The particle can transition from one well to another, but in the weak noise limit such transitions will be rare. LDT then gives us not only an approximation of the transition probability, but also of the mean exit time and the transition trajectory.

Such a knowledge can furthermore be used to tackle challenges in numerically sampling unlikely paths to rare events. In rare event simulation methods, a model is dynamically driven in such a way that otherwise very rarely visited paths are overpopulated [29,228]. This can be done either by manipulating the dynamical equations of the system, or by implementing genetic algorithms on top of the system, which selectively kill and clone parallel realisations of the model. Hence, such trajectories become statistically tractable without resorting to ultra-long numerical integrations. Enriching the statistics, while retaining the correct dynamics, makes it possible to explore the dynamical processes behind the extreme event of interest.

The previously mentioned fact that LDT allows one to select *typical* extreme events is key for interpreting some recent results on so-called rogue waves in the ocean [2,66,241]. Rogue waves are extremely dangerous hazards impacting the marine and coastal environment, and manifest themselves as hard-to-predict surface waves that can have surprisingly high destructive power and that, apparently, materialise out of nothing [65,193]. A novel viewpoint has been recently proposed for finding a comprehensive theoretical framework on rogue waves, able to generalise earlier theories. The idea has been to use LDT to study the properties of the solutions of the one-dimensional nonlinear Schrödinger equation starting from suitably defined random initial conditions constructed in accordance with observations taken from an oceanographic campaign. Both numerical and experimental evidence strongly suggest that rogues waves can be seen as hydrodynamic instantons, whose precursors can be clearly identified, and that can be computed by minimising a suitably defined action [60,61].

A related area of investigation is the study of—rarely occurring—noise-induced transitions between metastable states associated with alternative configurations of

geophysical flows or actual competing climatic states. In this case, along the lines of the classical Freidlin–Wentzell theory [93], the target region in the phase space for the endpoints of the desired paths is a special portion in the basin boundary separating the competing basins of attraction, which corresponds to a saddle in the classical case of motion in an energy landscape.

The multistability of the climate system manifests itself both locally and globally. By local we mean that the difference between the competing metastable states is, in fact, geographically confined and associated with one of the so-called tipping elements [163], representing features of the climate system that can go through critical transitions if forced beyond the point of no return. These include the dieback of the Amazon forest [21], the shut-down of the thermohaline circulation of the Atlantic ocean [221], the methane release resulting from the melting of the permafrost [269], and the collapse of the atmospheric circulation regime associated to the Indian monsoon [166].

A hierarchically higher level of multistability is present in the Earth as our planet is well known to have at least two possible steady climatic states in the current and past astronomical configuration, the warm climate, and a frozen one, termed snowball, which features global glaciation, extremely low temperatures and limited climatic variability. This is confirmed by geological and paleomagnetic evidence [128,211] and well understood in terms of relevant dynamical processes [33,102,179,237]. Despite the presence of chaotic dynamics in the competing attractors and of a complex geometrical structure in the basin boundary [175], suitable generalisations of the Freidlin–Wentzell theory proposed in [115–117] allow one to establish large deviation laws able to describe in the weak noise limit the transitions between the competing metastable states. Indeed, one can define a generalised quasipotential, whose local minima correspond to the competing attractors, while the transition paths cross preferentially the basin boundaries in special locations, which are saddles also termed Melancholia states [19,175–177]. There are good reasons to believe that, in fact, the climate system allows for the presence of additional competing metastable states on top of the warm and snowball climate [1,32,167]. This leads to a more complex pattern of possible transition paths between them and requires a careful statistical examination when noise is added into the system [182]. Finally, one can interpret the localised tipping elements described above as being associated with smaller and localised minima and saddles, which define the multiscale nature of the quasi-potential. Therefore, an adequate use of LDT might be key for making a more careful assessment of the risk coming from irreversible transitions for present-day tipping elements, and then for more precisely evaluating the risk of going beyond the so-called global planetary boundaries [248,249].

## 1.4 This review

The goals of this paper are to provide an informal mathematical introduction to LDT and then to lead the reader to explore some relevant applications of the theory for analysing properties of geophysical flows and of the climate system. The range of topics covered by this paper is somewhat broader and more targeted to real-life applications as compared to the excellent and more theoretically inclined earlier contribution

by Bouchet and Vernaille on the statistical mechanics of two-dimensional and geophysical flows [30].

Depending on the observable and on the scales of interest, and specifically on the strength of correlations, one can rely on different stochastic models to approximate the behaviour of climatic observables: independent, identically distributed random variables, Markov chains, dependent sequences. The theoretical overview of LDT presented in Sect. 2 is organised according to this line of thoughts. Subsequently, Sect. 3 introduces the concept of coarse-graining for the dynamics of geophysical flows, presents the general framework of stochastic climate models, and discusses the establishment of large deviation laws in stochastic and deterministic dynamical systems. The analysis of large deviation laws for stochastic dynamical systems will provide key tools for understanding the dynamical and statistical properties of transition paths between competing metastable states and for studying rare paths, rather than just rare events. Instead, the results presented for deterministic dynamical systems will be useful for understanding the reason why Markov chain models are of general interest for modelling the statistical properties of chaotic dynamical systems. Section 4 will then present a range of applications of LDT in various areas of geophysical fluid dynamics and climate science. We will showcase its use for understanding persistent climatic fluctuations, for characterising the fluctuations of the predictability of geophysical flows on different time scales, for providing a unified viewpoint for the understanding of rogue waves in the ocean, as well as for explaining special dynamical features associated with transitions between competing metastable states, thus mirroring the theoretical framework presented in the previous sections. This section contains also novel, previously unpublished results. Finally, Sect. 5 presents our conclusions together with a discussion regarding opportunities and challenges for future applications of LDT in climate science

## 2 A summary of large deviation theory

In this section, we recapitulate the main elements of LDT for two stochastic models applied often successfully to geophysical data: independent, identically distributed (i.i.d.) random variables and Markov chains, or more generally dependent sequences. This summary is far from being complete and does not make use of much mathematical sophistication either. Hence readers experienced in mathematics are referred to [63], whereas readers versed in physics are referred to [257]. These are at the same time the main sources we follow.

### 2.1 Independent, identically distributed random variables

The first basic results of LDT is known as Cramér's theorem [51] and describes the large deviation behaviour of empirical sample averages  $\frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} S_n$ .

**Theorem 1** *Let  $(X_i)$  be i.i.d.  $\mathbb{R}$ -valued random variables with a finite moment generating function in a region around the origin, i.e.  $0 \in \text{int}(\mathcal{D}_\varphi)$  with  $\mathcal{D}_\varphi = \{t \in \mathbb{R} :$*

$\varphi(t) := \mathbb{E}[e^{tX_1}] < \infty\}$ , where  $\mathbb{E}[f(X_1)]$  is the expectation value of  $f(X_1)$ . Let  $S_n = \sum_{i=1}^n X_i$ . Then, for all  $a > \mathbb{E}[X_1]$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left( \frac{1}{n} S_n \geq a \right) = -I(a) \quad (1)$$

where

$$I(z) = \sup_{t \in \mathbb{R}} [zt - \log \varphi(t)] \quad (2)$$

According to (1), which can be written in the form  $\mathbb{P}(S_n/n \geq a) \asymp \exp(-nI(a))$ ,<sup>1</sup> the probability of empirical averages deviating from the mean decays exponentially with the averaging length  $n$ , as  $n$  increases. If this is the case, we say that we have found a large deviation principle. The speed of decay is described by the rate function  $I$ . The rate function in Theorem 1 has some important and useful properties, such as compact level sets, lower semi-continuity and convexity on  $\mathbb{R}$  as well as continuity, strict convexity and smoothness on the interior of  $\mathcal{D}_I = \{z \in \mathbb{R} : I(z) < \infty\}$ .  $I(z) \geq 0$  with equality if and only if  $z = \mu$ , with  $\mu = \mathbb{E}[X_1]$ . Thus, the minimum of the rate function is located at the expectation value of the random variable suggesting that the sample averages converge to the expected value, as stated by the law of large numbers. Furthermore,  $I''(\mu) = 1/\sigma^2$ , the second derivative of the rate function at its minimum is the inverse of the variance of the random variable  $X_1$ , which goes back to the central limit theorem.

As shown by (2), the rate function is the Legendre transform of the cumulant generating function  $\log \varphi$ . We will discuss this relationship in more detail below. Equations (1) and (2) describe two different methods to estimate the rate function in case of applications: a direct method based on the probability density function (pdf) of averages and an indirect one based on the cumulant generating function, as discussed in detail in Sect. 3.3.

Considering that the rate function is lower semi-continuous and convex, and attains its unique minimum at the expectation value  $\mu$ , if  $a > \mu$ , then  $I(z) \geq I(a)$  for all  $z \geq a$ . Thus, Eq. (1) can be rewritten for  $a > \mu$  as

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left( \frac{1}{n} S_n \in A \right) = - \inf_{z \in A} I(z) \quad \text{with } A = [a, \infty). \quad (3)$$

Similarly, if, instead,  $a < \mu$ , one obtains:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left( \frac{1}{n} S_n \in A \right) = - \inf_{z \in A} I(z) \quad \text{with } A = (-\infty, a]. \quad (4)$$

This indicates one of the basic principles of LDT that we have hinted in the introduction. The occurrence of a large deviation  $\{\frac{S_n}{n} \in A\}$  is closely associated with the specific event corresponding to the lowest value of the rate function  $I$  taken in  $A$ ,

<sup>1</sup> We have that  $a^\varepsilon \asymp b^\varepsilon$  if  $\lim_{\varepsilon \rightarrow 0} \frac{\ln(a^\varepsilon)}{\ln(b^\varepsilon)} = 1$ ; here  $1/\varepsilon = n$ .

as the probability of this event is exponentially larger than the probability of all the other events compatible with the conditions  $\{\frac{S_n}{n} \in A\}$ . The rate function can then be interpreted as a cost function, and we have that *any large deviation is done in the least unlikely of all the unlikely ways* [63].

In the following, we discuss some generalisations of Theorem 1 by going from large deviations of empirical averages to large deviations of empirical measures. From the more general setting of Cramér's theorem we go now to a finite state space, where the i.i.d. random variables  $X_1, X_2, \dots$  take values in a finite set  $X_i \in \Gamma = \{1, \dots, r\} \subset \mathbb{N}$  and obey the marginal law  $\rho = (\rho_s)_{s \in \Gamma}$ ,  $\rho_s > 0$ . The empirical measure  $L_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$  is a random probability measure on  $\Gamma$ . We denote the set of probability measures on  $\Gamma$  by  $M(\Gamma) = \{v = (v_1, \dots, v_r) \in [0, 1]^r : \sum_{s=1}^r v_s = 1\}$ , where the total variation distance between two measures  $\mu$  and  $v$  is defined as  $d(\mu, v) = \frac{1}{2} \sum_{s=1}^r |\mu_s - v_s|$ . The following theorem, which goes back to Sanov [234], contains a large deviation law of  $L_n$  with respect to  $\rho$ .

**Theorem 2** *Let  $(X_i)$  be i.i.d. random variables satisfying the conditions above, and  $L_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ . Then, for all  $a > 0$ ,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(L_n \in B_a^c(\rho)) = - \inf_{v \in B_a^c(\rho)} I_\rho(v), \quad (5)$$

where  $B_a(\rho) = \{v \in M(\Gamma) : d(v, \rho) \leq a\}$ ,  $B_a^c(\rho) = M(\Gamma) \setminus B_a(\rho)$ , and

$$I_\rho(v) = \sum_{s=1}^r v_s \log \left( \frac{v_s}{\rho_s} \right) := H(v|\rho) \quad (6)$$

When comparing (3) with (5), it becomes clear that Theorem 2 is nothing more than a higher dimensional version of Theorem 1. Instead of looking at deviations of the empirical averages away from the mean, we consider now deviations of the empirical measure  $L_n$  away from the true measure  $\rho$ . The rate functions depends in this case on the different measures  $v$  on  $\Gamma$  and on how similar they are to  $\rho$ . The quantity  $H(v|\rho)$  is the relative entropy of the measure  $v$  with respect to the measure  $\rho$  [152]. By applying Jensen's inequality to  $I_\rho(v) = - \sum_s v_s (\log(\rho_s/v_s))$ , we have that  $I_\rho(v) \geq - \log \sum_s v_s (\rho_s/v_s) = 0$ , with the equality being realised if and only if  $v = \rho$ .

In other terms, Sanov's theorem states that the exponential rate of decay of the probability of a large deviation of size  $\geq a$  between the empirical measure and the marginal distribution  $\rho$  is controlled by the element of all measures on  $\Gamma$  whose distance from  $\rho$  is  $\geq a$  that is closest to  $\rho$  in the sense of relative entropy.

The contents of Theorem 2 allow us to reinterpret and extend the results discussed in (3)–(4). Let's consider a function  $f$  with  $\sum_{s=1}^r f_s \rho_s = \mu_f \in \mathbb{R}$ . We define  $\Phi_{f,a} = \{\phi \in M(\Gamma) | \sum_s f_s \phi_s \geq a\}$ . We also define  $\Psi_{f,a} = \{\phi \in M(\Gamma) | \sum_s f_s \phi_s = a\}$ . Clearly, one has  $\Phi_{f,a} = \cup_{b \geq a} \Psi_{f,b}$ .

We have that  $\mathbb{P}(L_n \in \Phi_{f,a}) = \mathbb{P}\left(\frac{1}{n} \sum_{j=1}^n f(X_j) \geq a\right)$ , where  $a \geq \mu_f$  and we consider the empirical measure  $L_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$  introduced before. One then derives that:

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(L_n \in \Phi_{f,a}) &= \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\left(\frac{1}{n} \sum_{j=1}^n f(X_j) \geq a\right) \\ &= - \inf_{\nu \in \Phi_{f,a}} I_\rho(\nu) = - \inf_{z \geq a} \inf_{\nu \in \Psi_{f,z}} I_\rho(\nu) \end{aligned} \quad (7)$$

Let's now consider the case  $f(x) = x$ . The empirical average  $S_n/n$  is connected to the empirical measure  $L_n$  through the formula  $S_n/n = \sum_{s=1}^r s L_n(s)$ . The rate function in (3) can be obtained from (7) by

$$I(z) = \inf_{\nu \in \Psi_{f,z}} I_\rho(\nu). \quad (8)$$

Thus, the rate function of the empirical average  $z$  is equal to the infimum of the rate function for the empirical measure  $\nu$  if the infimum is taken over all the measures  $\nu$  with mean  $\mu = \sum_{s=1}^r s \nu_s = z$ . In other words, there is an equivalence between the large deviations of the empirical average  $z$  and the large deviations of the least unlikely empirical measure  $\nu$  with mean equal to  $z$ . This is an example of the contraction principle that we state now.

**Theorem 3** *Contraction Principle. Let  $A_n$  be a family of random variables such that*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(A_n \in \mathcal{A}) = - \inf_{z \in \mathcal{A}} I_A(z) \quad (9)$$

*and let's consider another family of random variables  $B_n = T(A_n)$  where  $T$  is a continuous function. It is possible to establish a large deviation principle for  $B_n$  as follows:*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(B_n \in \mathcal{B}) = - \inf_{z \in \mathcal{B}} I_B(z), \quad I_B(z) = \inf_{y=T^{-1}z} (I_A(y)). \quad (10)$$

Theorem 2 can be generalised further to large deviations of pair empirical measures as well as of measures with higher dimensions. Higher level large deviation laws imply the ones for lower levels, the downward link being provided by the contraction principle. The interested reader can find a short summary of the generalisations to higher dimension in Appendix A, for a detailed discussion of this topic we refer to [63].

## 2.2 Dependent sequences

We continue with a generalisation of Cramér's Theorem for random sequences that have a form of moderate dependence, which goes back to [100] and [80]. A rigorous

derivation of the Gärtner–Ellis (GE) theorem would go beyond the scope of this paper, thus we concentrate on the main results. As above, we follow here the work of [257].

We consider the sequence  $(Z_n)$  of random variables on the probability space  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mathbb{P})$ , where  $\mathcal{B}(\mathbb{R}^d)$  is the Borel sigma-field on  $\mathbb{R}^d$  with moment generating functions

$$\varphi_n(t) = \mathbb{E}[e^{\langle t, Z_n \rangle}], \quad t \in \mathbb{R}^d, n \in \mathbb{N} \quad (11)$$

with  $\langle \cdot, \cdot \rangle$  denoting the standard inner product. It can be useful to think of  $(Z_n)$  as an empirical average, but this doesn't have to be the case. We assume that the limit

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \varphi_n(nt) = \Lambda(t) \in [-\infty, \infty] \quad (12)$$

exists and

$$0 \in \text{int}(D_\Lambda), \text{ with } D_\Lambda = \{t \in \mathbb{R}^d : \Lambda(t) < \infty\}. \quad (13)$$

We also assume that  $\Lambda$  is convex and differentiable on  $\text{int}(D_\Lambda)$ . Furthermore, we assume that  $\Lambda$  is lower semi-continuous on  $\mathbb{R}$ , and either  $D_\Lambda = \mathbb{R}^d$  or  $\Lambda$  is steep at  $\partial D_\Lambda$ .

Let  $P_n(\cdot) = \mathbb{P}(Z_n \in \cdot)$ . Under the above conditions, the GE theorem states that  $(P_n)$  satisfies a large deviations principle on  $\mathbb{R}^d$  with rate  $n$  and with rate function

$$I(x) = \sup_{t \in \mathbb{R}^d} [\langle x, t \rangle - \Lambda(t)], \quad x \in \mathbb{R}^d. \quad (14)$$

Thus, the rate function  $I$  is the Legendre-transform of  $\Lambda$ , also called the scaled cumulant generating function. The rate function  $I$  is convex. Note that (14) is a generalised form of (2).

If  $Z_n = \frac{1}{n} \sum_{i=1}^n X_i$  with  $X_i$  a stationary random sequence, then conditions (12) and (13) can be interpreted as a kind of moderate dependence assumption on  $(X_i)$ . However, in case of strong dependence, the theorem would fail because the strict convexity of  $\Lambda$  would be violated.

We have seen that by using the GE theorem one obtains a large deviations principle under fairly mild regularity assumptions. As mentioned above, it is not necessary that  $Z_n$  represents sample averages. In fact, the large deviation principles presented in Sects. 2.1 and 2.2 for sample averages, empirical measures, pair empirical measures (see Appendix A), and so on, can all be obtained by following the route given by the GE theorem as well. Below, we derive based on [63, 257] the rate functions of sample averages for i.i.d. random variables and for Markov chains, by using the GE theorem.

1) Let  $(X_i)$  be i.i.d.  $\mathbb{R}$ -valued random variables satisfying  $\varphi(t) = \mathbb{E}[e^{tX_1}] < \infty$ , for all  $t \in \mathbb{R}$ . Let us consider the empirical average  $Z_n = \frac{1}{n} \sum_{i=1}^n X_i$ . Then,

$$\varphi_n(nt) = \mathbb{E}[e^{ntZ_n}] = \mathbb{E}\left[e^{t \sum_{i=1}^n X_i}\right] = [\varphi(t)]^n, \quad (15)$$



with  $\varphi$  the moment generating function of  $X_1$ . Hence  $\Lambda(t) = \log \varphi(t)$  and the GE theorem reduces to Cr  mer's theorem (Theorem 1).

2) Let  $(X_i)$  be a stationary  $\Gamma$ -valued Markov chain. Let  $Z_n = \frac{1}{n} \sum_{i=1}^n f(X_i)$ , where  $f : \Gamma \rightarrow \mathbb{R}^d$ ,  $d \geq 1$ . Then,

$$\begin{aligned} \varphi_n(nt) &= \mathbb{E}[e^{ntZ_n}] = \mathbb{E}\left[e^{t \sum_{i=1}^n f(X_i)}\right] \\ &= \sum_{x_1, \dots, x_n \in \Gamma} \pi(x_1) e^{tf(x_1)} P(x_2|x_1) e^{tf(x_2)} \cdots P(x_n|x_{n-1}) e^{tf(x_n)}, \end{aligned}$$

where  $\pi(x_1)$  denotes the probability of the initial state  $x_1$ , and  $P(x_i|x_{i-1})$  denotes the conditional probability of state  $x_i$  given  $x_{i-1}$ ,  $i = 1, \dots, n$ . By defining  $\pi_t(x_1) = \pi(x_1) e^{tf(x_1)}$  and  $P_t(x_i|x_{i-1}) = P(x_i|x_{i-1}) e^{tf(x_i)}$ , we have that

$$\varphi_n(nt) = \sum_{j \in \Gamma} (\Pi_t^{n-1} \pi_t)_j,$$

where  $\pi_t$  is the vector of probabilities for which  $(\pi_t)_i = \pi_t(x_1 = i)$ , and  $\Pi_t$  denotes the matrix with elements  $(\Pi_t)_{ji} = P_t(j|i)$ . Based on Perron–Frobenius theory for positive matrices we get that  $\lim_{n \rightarrow \infty} \log \varphi_n(nt) = \log \lambda(t)$ , with  $\lambda(t)$  denoting the unique largest eigenvalue of  $\Pi_t$ . Hence  $\Lambda(t) = \log \lambda(t)$ , and the rate function is given by the Legendre transform

$$I(z) = \sup_{t \in \mathbb{R}} [zt - \log \lambda(t)]. \quad (16)$$

Please note that (16) can be used to obtain the rate function only if  $\Pi$  has a unique stationary distribution  $\pi$ . If  $\Pi$  has several stationary distributions,  $\Lambda(t)$  exists, but depends on the initial distribution  $\pi(x_1)$ . If  $\Pi$  has no stationary distribution, generally no large deviation principle can be found and the law of large numbers does not even hold [257].

### 3 Large deviations in dynamical systems

At this point, we leave the idealised world of i.i.d. random variables and discrete time processes, and turn our attention to systems evolving continuously in time, as we want to look into mathematical models that are more relevant for capturing the dynamical properties of the climate system. Instead of empirical measures and sample averages, we consider in the following probabilities of trajectories or paths of deterministic dynamical systems and finite time averages along these trajectories. However, the main ingredients leading to large deviation results stay the same. One needs basically the attracting effect of an asymptotic limit leading to an exponential decay of probabilities of finite time estimates. By taking into consideration the dynamics in time and including the temporal dimension into the large deviation analysis, the methods presented below are directly relevant for geophysical applications. We will present some

basic results pertaining to stochastic and to deterministic chaotic dynamical systems, for the sake of completeness, and because the modelling of geophysical flows follows both dynamical paradigms.

First, we motivate the use of stochastic dynamics for investigating the properties of geophysical flows by introducing the concept of filtering and the development of evolution equations based on dynamical balances and specialised for specific scales of motion [104,142,247]. The introduction of stochastic parametrizations [16,92,104,143,277] is motivated through the use of the Mori–Zwanzig formalism [188,288]. When suitable limits are considered, the stochastic component, which provides a surrogate representation of the effects of the scales we are unable to describe explicitly, can be written as multiplicative white noise [202]. This provides the basis for a large class of stochastic climate models of very widespread use and great physical relevance [125,131,132,181,205,233]. Such stochastic models are amenable to being studied using the Freidlin–Wentzell theory [93], which allows to derive powerful large deviation results. Additionally, one should keep in mind that the climate undergoes actual stochastic forcing due to random fluctuations in the incoming solar radiation and other astrophysical factors. More in general, the use of stochastic dynamics for describing non-equilibrium statistical mechanical systems has reached a high level of popularity and has shown a great potential for deriving results of great theoretical and practical relevance [10,154,170,198].

In case of the Freidlin–Wentzell theory, the zero-noise limit of stochastic evolution law is given by its purely deterministic component. Hence, one obtains the probability of random paths deviating from the deterministic path in terms of large deviation laws. The probabilities of deviation of finite time averages from their asymptotic values can be obtained from the large deviation results for random paths using the contraction principle. A more general and pragmatic approach, however, which can be followed even in case of unknown model equations, is related to the fact that finite time averages of weakly correlated observables are (nearly) independent. Thus, one can model finite time averages of correlated observables as resulting from i.i.d. random variables or Markov chains. Consequently, the theorems presented in Sect. 2 can be applied in a similar way with the difference that the large deviations parameter  $n$  is now related to time. In Sect. 3.3 we discuss a modified version of the GE theorem (14) acting on time averaged observables.

Later on, we consider special chaotic dynamical systems, so-called Axiom A systems [31], and discuss the emerging large deviation laws for finite time averages of given observables. The framework of Axiom A systems—which are essentially the closest deterministic relatives of the *truly* stochastic systems—blurs the distinction between statistical mechanics and dynamical systems theory, mainly as a result of the fact that Axiom A systems possess a rather special ergodic invariant measure that has a clear physical interpretation [78]. Another remarkable property of Axiom A systems is that they admit a Markov partition, i.e. a partition of the attractor such that one can put in a one-to-one correspondence the actual orbit of the system with an infinite sequence of symbols describing the history of occupancy of the various elements of the partition. Accordingly, the original map can be associated with a shift map, i.e. a finite-state Markov chain describing the probability of transition between the various elements of the partition [97,229]. The possibility of establishing the so-called symbolic dynamics

guarantees that the results presented in Sect. 2.2 for finite-state Markov chain apply also for Axiom A systems. Nonetheless, there is no *free lunch*: it is in general far from trivial to actually construct the Markov partition. As discussed below, while Axiom A systems are very special dynamical objects, the chaotic hypothesis [97,98] makes them very relevant for providing a framework for studying large deviations laws in high-dimensional geophysical systems.

### 3.1 Stochastic climate models

The state of the climate system can be described using the continuum approximation, introducing field variables that depend on three spatial dimensions and time. The partial differential equations that describe the evolution of the field variables are based on the budget of mass (including different chemical species), momentum and energy. Since the climate system features variability on a vast range of spatial and temporal scales, as mentioned above, a key procedure one needs to apply, both on theoretical grounds and for reasons of defining efficient numerical models, is to specialise the evolution equations to a desired range of spatial and temporal scales of interest by the use of suitable approximations based on the validity of approximate dynamical balances [104,142,247]. Additionally, when constructing an actual numerical model, the three-dimensional fields are discretised on a lattice, either in the physical space, or in the reciprocal space via spectral projection, or in a suitable combination of the two. Hence, the impact of the physical processes occurring in the unresolved spatio-temporal scales on those taking place in the resolved ones can be represented only through approximate parametrizations [16,92]. The Mori–Zwanzig coarse-graining based on the projection operator [188,288] clarifies that such parametrizations have in general a deterministic, a stochastic, and a non-Markovian component [42,143,276,277].

Let us assume, for simplicity, that the *true* evolution equation for the climate system can be written as a system of autonomous ordinary differential equations<sup>2</sup> of the form

$$\frac{dz}{dt} = G(z) \quad (17)$$

where  $z \in \mathbb{R}^N$ . The procedure of coarse-graining, associated with specialising the equations for a specific range of time and spatial scales, implies that we rewrite the state vector  $z$  as  $z = (x, y)$ , where  $x \in \mathbb{R}^n$  and  $y \in \mathbb{R}^{N-n}$  and we aim at deriving approximate equations of the variables of interest  $x$ . It is reasonable to assume that  $n \ll N$ . Note that, alternatively,  $x$  can correspond to the variables describing the state of a portion of the climate system (e.g. the atmosphere), and  $y$  can instead describe the rest of the system. One does not need to assume *a priori* the presence of a very large time-scale separation between the dynamics of the  $x$  and  $y$  components. One can then rewrite (17) as:

<sup>2</sup> We are here neglecting the—very important—presence of explicit time-dependence and stochastic forcing in the dynamics; see [43,103,104] for a detailed discussion of this aspect.

$$\begin{aligned}\frac{dx}{dt} &= f(x) + \delta f_x(x, y) \\ \frac{dy}{dt} &= \frac{1}{\epsilon} g(y) + \frac{\delta}{\epsilon} g_y(x, y)\end{aligned}\quad (18)$$

where  $f$  and  $g$  define the autonomous dynamics of the  $x$  and  $y$  components, respectively,  $\delta$  is a constant controlling the intensity of the coupling, and  $\epsilon$  defines the time scale separation between the two sets of variables. The Mori–Zwanzig theory indicates that one can in general write the dynamics of the  $x$  variables in an implicit form as follows:

$$\frac{dx}{dt} = f_{\epsilon, \delta}(x) + \dot{\sigma}_{\epsilon, \delta}(x) + \int ds K_{\epsilon, \delta}(x, t-s)x(s), \quad (19)$$

where the three terms of the right hand side correspond to the deterministic drift, to a noise contribution, and to the memory term. In the weak-coupling limit ( $\delta \rightarrow 0$ ), it is possible to derive via perturbative approach an explicit expression for these three terms that is valid up to order  $\delta^2$  [276, 277]; see a practical implementation of this theory for the development of parametrizations in geophysical fluid dynamical models in [59, 264, 275]. Note that one can derive an expression for the Mori–Zwanzig projected dynamics using data-driven approaches [42, 143]. Very recently, it has been shown [120] that the data-driven and the top-down approach presented in [276, 277] are fundamentally equivalent.

Instead, if the two sets of variables  $x$  and  $y$  have an infinite time scale separation ( $\epsilon \rightarrow 0$ ), the dynamics of the variable  $x$  converges to a deterministic averaged equation (for more details, see Sect. 3.2 below). Via homogenisation theory [202] deviations from this averaged equation can be modelled by a stochastic differential equation without memory, and with multiplicative white noise, so that the evolution of the  $x \in \mathbb{R}^n$  variables is controlled by:

$$dx_t = F(x_t)dt + \Sigma(x_t)dW_t, \quad (20)$$

where it is possible to derive explicit formulas for the renormalised drift term  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ , and the diffusion matrix  $\Sigma : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times m}$ , while  $W_t$  is an  $m$ -dimensional Brownian motion. Here and in what follows we assume the Itô convention for stochastic differential equations. In a nutshell, the impact of the neglected scales of motions corresponding to the  $y$  variables is twofold: it leads to (a) a change in the deterministic contribution to the evolution of the  $x$  variables; and to (b) the inclusion of a random forcing. Stochasticity is essentially due to the lack of information on the state of the  $y$  variables in the projected  $x$  space; see a detailed discussion of this in [42].

Equation (20) is at the basis of stochastic climate models, whose investigation was initiated by Hasselmann [125]; see a comprehensive analysis of this viewpoint and further developments in [131, 132, 181, 205, 233]. Traditionally, the deterministic component of (20) features one or more fixed points, and the noise allows for the system to explore regions of the phase space far from the deterministic solutions, and to perform transitions between competing metastable states. We will provide a broader view point on this in Sect. 4.5, where we will consider more general competing asymptotic

states. Stochastic climate models have been key for discovering fundamental physical processes like stochastic resonance [14,15,99,191,192], and have provided key insights for studying the transitions between different weather regimes in the atmosphere [13,40,134,144,186,231]; see discussion in Sect. 4.6. Equation (20) is probably the most convenient starting point for discussing the use of LDT in geophysical flows even if, as shown in Sect. 3.4, LDT can be introduced also in the context of deterministic chaos.

### 3.2 Dynamical systems perturbed by weak noise

We now focus on the stochastic climate models introduced in the previous subsection and aim at deriving large deviation laws. The Freidlin–Wentzell theory [263,274], allows one to study the convergence of probability measures on the path-space of a stochastic differential equation  $X$  in  $\mathbb{R}^n$

$$dX_t^\varepsilon = b(X_t^\varepsilon)dt + \sqrt{\varepsilon}\sigma(X_t^\varepsilon)dW_t \quad X_0^\varepsilon = x, t \geq 0. \quad (21)$$

where, as in (20)  $b : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is a deterministic drift,  $\sigma : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times m}$  is the diffusion function,  $W_t$  is an  $m$ -dimensional Brownian motion, and we introduce here the parameter  $\varepsilon > 0$  that controls the intensity of the stochastic forcing.

For bounded and Lipschitz  $b$  and  $\sigma$ , it can be shown that as the noise intensity goes to zero ( $\varepsilon \rightarrow 0$ ), the distribution of paths of  $X_t^\varepsilon$  converges to the deterministic path determined by  $dx_t = b(x_t)dt$  [274]. For all  $T > 0$  and  $\delta > 0$

$$\lim_{\varepsilon \rightarrow 0} \mathbb{P} \left\{ \max_{0 \leq t \leq T} |X_t^\varepsilon - x_t| > \delta \right\} = 0.$$

We may wonder, of course, about the probability of observing a given path  $f(t) \neq x_t$  when  $\varepsilon \neq 0$ . It can be shown that a large deviation principle holds for  $X_t^\varepsilon$ , with a rate function or action functional

$$I_T(f) = \frac{1}{2} \int_0^T \left\langle \dot{f}(t) - b(f(t)), a^{-1}(f(t))(\dot{f}(t) - b(f(t))) \right\rangle dt, \quad (22)$$

where  $a(x) = \sigma(x)\sigma(x)^T$  is the noise covariance. We have that

$$\mathbb{P} \left\{ \sup_{t \in [0, T]} |X_t^\varepsilon - f(t)| < \delta \right\} \asymp \exp \left( -\frac{1}{\varepsilon} I_X(f) \right). \quad (23)$$

Similarly as integrals of the form  $\int_a^b e^{-\frac{1}{\varepsilon} h(x)} k(x) dx$  are dominated by the minimum  $x_0$  of  $h(x)$  in Lagrange's methods, as  $\varepsilon \rightarrow 0$ , the probability in a set  $F \subset C[0, T]$  of trajectories concentrates on the trajectory  $f^*$  with the smallest rate function  $I_X$ :

$$I_T(f^*) = \inf_{f \in F} I_T(f). \quad (24)$$

Such path is called the minimum action path or instanton.

*The exit problem* In the limit  $\varepsilon \rightarrow 0$ , the dynamics of (21) is determined by the drift field  $b$ . When  $b$  has an attractor, the trajectory will never escape from it in the absence of noise. The situation is markedly different when noise is added. The system can make excursions away from the attractor, can exit from its surrounding, and can possibly perform transitions to another attractor.

LDT provides a way to describe the exit from regions containing an attractor, e.g. if  $\Omega$  is a bounded set containing a stable fixed point  $\bar{x}$  of  $dx_t = b(x_t)dt$ , then the exit from the domain will happen close to the point minimising the action (22). For more details see [93].

*Instanton calculation* The minimisation of the action functional for problems of interest in geophysics can usually not be done analytically. In such cases the instanton needs to be calculated numerically.

Arguably the most direct way of finding the instanton is by minimising the action (22). In the minimum action method [75], the instanton  $f^*$  on a finite time interval  $[0, T]$  is approximated by  $f^*(t_i)$  on a discrete temporal grid, a discrete approximation to the action is derived and a quasi-Newton method is then applied to minimise the discretised action.

Another fairly simple method of numerically finding the instanton is solving the Hamilton equation connected to this minimisation problem. A difficulty arises here in that we are often looking for a minimisation with fixed start and end points for  $f^*$  at  $t = 0$  and  $t = T$ . To solve the Hamilton equation we need to specify initial values for the coordinates and their conjugate momenta, however. A shooting method can be applied to find the initial values of the momenta, but this is in general difficult to apply in high dimensions.

Both these methods can only be applied to finite time intervals, while in many cases we will want to allow for infinite time lengths of transition. In the special case where the drift term is a gradient, i.e.  $b = -\nabla U$ , and  $\sigma$  is the identity, these problems can be circumvented by using the string method [74] which uses that the instanton is always parallel to the drift. The method alternates relaxation along the drift with a redistribution of the discretisation points along the instanton curve.

This principle has been further generalised to non-gradient systems in the geometric minimum action method [261]. Here the action is reformulated in a geometric way that does not involve the time parametrization of the instanton. In this way the problem of infinite transition times can be circumvented.

An overview of numerical methods to calculate the instanton is given in [114].

*Systems with a time scale separation* As mentioned above, in some geophysical settings, we may be interested in the evolution of a number of slowly evolving variables  $x$  in interaction with other variables  $y$  that evolve on a much faster time scale. We then consider a slightly modified version of (18):

$$dx = f(x, y)dt \quad (25)$$

$$dy = \frac{1}{\varepsilon}g(x, y)dt + \frac{1}{\sqrt{\varepsilon}}\sigma(x, y)dW. \quad (26)$$

where we introduce a white noise forcing term for the fast (as we consider  $\epsilon \rightarrow 0$ ) variables  $y$ . Note that  $\mathbb{E}((W(t) - W(s))^2) = (t - s)$ , hence the scaling with  $\sqrt{\epsilon}$ . Intuitively speaking, in the limit  $\epsilon \rightarrow 0$ , in any time interval of order 1, no matter how short, the  $y$  variable will explore the invariant measure of the equation for  $y$  for fixed  $x$  determined by

$$d\tilde{y} = g_x(\tilde{y})dt + \sigma_x(\tilde{y})dW, \quad (27)$$

where  $g_x(y) = g(x, y)$  and  $\sigma_x(y) = \sigma(x, y)$ . As a result, we get a law-of-large-numbers-like result for the slow variable  $x$ . As  $\epsilon \rightarrow 0$ , the path  $x(t)$  converges to  $X(t)$ , the solution of

$$dX = F(X)dt,$$

where  $F(X) = \int f(X, y)\tilde{\mu}_x(dy)$  with  $\tilde{\mu}_x(dy)$  the invariant measure of (27).

As with the law of large numbers for averages of i.i.d. random variables, we may expect a large deviation result to hold here as well. To derive, at a heuristic level, the rate function for the path probabilities of the slow variable  $x$  we consider a discrete time approximation of (25)–(26). We approximate  $x(t)$  for  $t \in [0, T]$  by discrete  $x_i$  at times  $i\Delta t_x$  with  $i \in \{0, \dots, N_x\}$  with  $N_x = \lfloor T/\Delta t_x \rfloor$ .

Since  $x(t + t_x) = x(t) + \int_t^{t+t_x} f(x(\tau), y(\tau))d\tau$  we approximate the increment of  $x$  between two subsequent discrete times by

$$\begin{aligned} x_{i+1} - x_i &= \int_t^{t+\Delta t_x} f(x_i, \tilde{y}_{x_i}^{(\epsilon)}(\tau))d\tau, \\ &= \epsilon \int_0^{\Delta t_x/\epsilon} f(x_i, \tilde{y}_{x_i}(\tau))d\tau, \end{aligned}$$

which we can express in terms of a time average as

$$\frac{x_{i+1} - x_i}{\Delta t_x} = \frac{\epsilon}{\Delta t_x} \int_0^{\Delta t_x/\epsilon} f(x_i, \tilde{y}_{x_i}(\tau))d\tau.$$

The probability of the slow process going from some given value  $\varphi_i$  at time  $i\Delta t_x$  to  $\varphi_{i+1}$  at time  $(i+1)\Delta t_x$  can therefore be estimated via the large deviations of the time average of  $f(x_i, \tilde{y}_{x_i}(t))$  as

$$\begin{aligned} \mathbb{P}(x_{i+1} = \varphi_{i+1} | x_i = \varphi_i) &= \mathbb{P}\left(\frac{\Delta x_i}{\Delta t_x} = \frac{\Delta \varphi_i}{\Delta t_x} \middle| x_i = \varphi_i\right) \\ &= \mathbb{P}\left(\frac{\epsilon}{\Delta t_x} \int_0^{\Delta t_x/\epsilon} f(x_i, \tilde{y}_{x_i}(\tau))d\tau = \frac{\Delta \varphi_i}{\Delta t_x} \middle| x_i = \varphi_i\right) \\ &\approx e^{-\frac{\Delta t_x}{\epsilon} \Lambda_{\varphi_i}^*\left(\frac{\Delta \varphi_i}{\Delta t_x}\right)} \end{aligned}$$

where  $\Lambda_\varphi^*$  is the rate function for the time averages of  $f(\varphi, \tilde{y}_\varphi(t))$ , the Legendre–Fenchel transform of the scaled cumulant generating function

$$\lambda_\varphi(\theta) = \lim_{T \rightarrow \infty} \frac{1}{T} \ln \mathbb{E} \left( \exp \left( \theta \int_0^T f(\varphi, \tilde{y}_\varphi(\tau)) d\tau \right) \right).$$

Hence, assuming Markovianity for  $x_i$  in the limit  $\epsilon \rightarrow 0$  due to rapid decorrelation of the  $y$  process, the path probability for  $x$  can be approximated as

$$\begin{aligned} \mathbb{P}(x(t) = \varphi(t) \text{ for } t \in [0, T]) &\approx \mathbb{P}(x_0 = \varphi_0, \dots, x_{N_x} = \varphi_{N_x}) \\ &= \mathbb{P}(x_0 = \varphi_0) \mathbb{P}(x_1 = \varphi_1 | x_0 = \varphi_0) \dots \mathbb{P}(x_{N_x} = \varphi_{N_x} | x_{N_x-1} = \varphi_{N_x-1}) \\ &\approx \mathbb{P}(x_0 = \varphi_0) \prod_{i=0}^{N_x-1} \exp \left( -\frac{\Delta t_x}{\epsilon} \Lambda_{\varphi_i}^* \left( \frac{\Delta \varphi_i}{\Delta t_x} \right) \right) \\ &\rightarrow \mathbb{P}(x_0 = \varphi_0) \exp \left( -\frac{1}{\epsilon} \int_0^T \Lambda_{\varphi(\tau)}^*(\dot{\varphi}(\tau)) d\tau \right). \end{aligned}$$

From this very non-rigorous derivation we can expect that the rate function for the slow process  $x$  as  $\epsilon \rightarrow 0$  is  $\int_0^T \Lambda_{\varphi(\tau)}^*(\dot{\varphi}(\tau)) d\tau$ . The same result has been derived in a more rigorous manner in [25].

### 3.3 Time averaged observables

In this section we consider large deviation results for time averages of observables of dynamical systems. The large deviation parameter is in this case the time length  $T$  over which the average is taken. To illustrate the main results, let us consider a Markov process  $X(t) \in \mathbb{R}^n$ , and an observable  $A : \mathbb{R}^n \rightarrow \mathbb{R}$ . We have a large deviation principle for the time average  $a = \frac{1}{T} \int_0^T A(X(t)) dt$  if its probability distribution scales for large  $T$  as

$$\rho(a) \underset{T \rightarrow \infty}{\asymp} e^{-TI(a)} \quad (28)$$

with rate function  $I(a)$ . Similarly to what discussed in the previous sections, one can define the scaled cumulant generating function

$$\lambda(k) = \lim_{T \rightarrow +\infty} \frac{1}{T} \log \mathbb{E} \left[ e^{k \int_0^T A(t) dt} \right], \quad (29)$$

and the Gärtner–Ellis theorem relates rate function  $I(a)$  and scaled cumulant generating function  $\lambda(k)$  through Legendre transformation. In particular, when the rate function  $I(a)$  is convex and differentiable, or equivalently when  $\lambda(k)$  is differentiable, the Legendre transform can be inverted, and the rate function can be computed as solution of the variational problem as  $I(a) = k(a)a - \lambda(k(a))$ , where  $k(a)$  is given by  $a = \lambda'(k(a))$ .



Large deviation results of this kind hold in general for mixing dynamics and for observables for which the tails of the distribution decay sufficiently fast. Mathematically, sufficient conditions are given by [68–71]; see also the discussion in [257]. In most applications to geophysical fluid dynamics or climate sciences we either use stochastic models for which either condition is satisfied, or we consider deterministic chaotic systems of sufficient complexity that we expect the conditions to hold (see discussion in the introduction of Sect. 3 and more in detail in Sect. 3.4). However, it is important to keep in mind that the existence of a large deviation result is in general not guaranteed, and must be proved or validated empirically.

Several physical systems have been reported featuring anomalous large deviation scalings [190], i.e. scalings where the large deviation parameter is non-standard, and, specifically  $T^\alpha$  with  $\alpha \neq 1$  in the present case. Typically, these are systems featuring non-Markovian dynamics or long-range correlations [48,64,72,111,124,149,173,190,232,235,252,285]; see a detailed treatment of the problem in the case of deterministic dynamical systems in [41,222]. However, it has been shown that even in the case of a system as simple and well-behaved as the Ornstein–Uhlenbeck process, simply by considering as observable the third or higher moment of the state of the system one obtains anomalous large deviation scalings [41,190,222]. It is therefore important to proceed carefully when testing large deviation scalings in complex systems like the ones typically analysed in climate science.

If valid, a large deviation result for the time average of an observable gives an extension of the central limit theorem that allows to take into considerations fluctuations of order  $T$  rather than  $\sqrt{T}$ . For ergodic systems the time average of an observable  $A$  converges in the limit of large  $T$  to the ergodic average  $\mu = \mathbb{E}[A]$ . Under mixing hypotheses, the central limit theorem gives that for large  $T$ , typical fluctuations of the time average are of order  $\sqrt{T}$  and Gaussian distributed, that is  $(a - \mu)/\sqrt{T} \sim N(0, \sigma^2 \tau_c)$ , where  $\sigma^2$  is the variance of  $A$  and  $\tau_c$  its integral autocorrelation time  $\tau_c = \sigma^{-2} \int_{-\infty}^{+\infty} C(\tau, 0) d\tau$ , with  $C(t, s) = \mathbb{E}[(A(t) - \mu)(A(s) - \mu)]$  the covariance of  $A$ . However, in certain applications it is of interest to consider fluctuations that are more rare than those handled by the central limit theorem, and that instead scale with  $T$ . A large deviation result allows to have a limit distribution for these large fluctuations.

Being a more general result, the large deviation scaling allows to obtain information about the Gaussian fluctuations directly from the knowledge of  $I(a)$ . Let us first note that in the large deviation limit of large  $T$  the distribution function concentrates around the most probable value  $a_m$  for which  $I'(a_m) = 0$ , and that in the limit of large  $T$  this value corresponds also to the average, that is  $a_m = \mu = \mathbb{E}[A]$ . Expanding the rate function in  $a$  around this value, one finds that in the large deviation limit neglecting terms  $O((a - \mu)^3)$  the distribution is  $\rho(a) \asymp e^{-T(a-\mu)I''(\mu)/2}$ , that is a Gaussian distribution with variance given by the curvature of the rate function around the most probable value  $1/I''(\mu)$ .

The specific form of the distribution can be obtained expanding the scaled cumulant generating function for small values of  $k$  and using the Gärtner–Ellis theorem (see Appendix B), which results in the following quadratic form for the rate function

$$I(a) \approx \frac{(a - \mu)^2}{2\sigma^2\tau_c}. \quad (30)$$

We see that this corresponds to the Gaussian scaling predicted by the central limit theorem for the time average of a correlated process. A necessary condition for the approximation to hold is  $|a - \mu|/\sqrt{2\sigma^2\tau_c} < 1/\sqrt{T}$ , which is consistent with the expected scaling of Gaussian fluctuations.

For an observable that is Gaussian distributed the quadratic form above is exact. For more general processes however the rate function contains more information than just the Gaussian fluctuations. The higher order derivatives of the rate function correspond to higher order cumulants, and describe fluctuations beyond the Gaussian approximation. The most interesting aspect of studying the rate function is the reconstruction of the tails beyond the Gaussian bulk. Equation 30 however can still be of interest, as discussed in Sect. 4 and for different applications in [25,94,96,217,219].

If the process under study is an ergodic Markov process, the large deviation functions can be computed using the Donsker–Varadhan theory of additive Markov processes, essentially extending to continuous time the results presented in Sect. 2. However, for most applications in geophysical fluid dynamics or in the climate sciences a very careful analysis is required, as the system typically has an extremely complex (deterministic) dynamics, whose equations are in some cases not even known (e.g., in the case of real world climatic observations, or even for climate models data, given the complexity and relative opacity of the code of these numerical models). Part of the problem can be bypassed by taking the assumptions discussed in the introduction of Sect. 3 and more in details in Sect. 3.4, and treating the output of the system as an effective ergodic Markov process. However, it is still necessary to understand how one can compute the large deviation functions empirically, in the (frequent) cases when this is the only alternative.

Here we give a summary of a possible procedure, presented in more details in [217,224]. Let us assume that we have a time series of an observable  $A(t)$  from time 0 to time  $T$ . The idea is to proceed with a block-averaging approach, and divide the time series in  $N_b = T/\tau_b$  blocks of length  $\tau_b \gg \tau_c$ . Since the length of the block is much larger than the auto-correlation time, the time averages

$$a_b^j = \frac{1}{\tau_b} \int_{(j-1)\tau_b}^{j\tau_b} A(t) dt, \quad j = 1, \dots, N_b \quad (31)$$

can be considered as sum of  $\tau_b/\tau_c$  independent values, possibly leading to convergence to a large deviation result. Additionally, the  $N_b$  values  $a_b^j$  can be considered as independent realisations of the process  $a_b = \frac{1}{\tau_b} \int_0^{\tau_b} A(t) dt$ , and they can be used to compute the expectation values in the definitions of the large deviation functions as ensemble averages, and to study the convergence to the limit for  $\tau_b \rightarrow +\infty$ .

The large deviation functions for large but finite values of  $\tau_b$  can be computed in two ways. One way is to attempt to estimate directly the rate function by computing

$$I_b(a) = -\frac{\ln \rho(a_b = a)}{\tau_b}, \quad (32)$$

that gives an estimate of the rate function for finite  $\tau_b$ , up to an additive constant due to the prefactor in the large deviation scaling. The convergence of  $I_b(a)$  to a limit function  $I(a)$  can then be evaluated for each value of  $a$  by increasing  $\tau_b$  until the value reaches a plateau up to a given tolerance error. This approach has the advantage of being extremely straightforward. However, it requires a large amount of data to obtain a relative error that is constant with  $\tau_b$  [224], and it suffers from the drawback that it is not easy to study precisely the convergence of the estimators, as it is based on the estimate of a probability density.

A second way consists of computing the scaled cumulant generating function first, and then use the Gärtner–Ellis theorem to obtain the rate function [217,224]. This method has the advantage of not involving the estimate of a probability distribution and of requiring less data to achieve a similar precision. The method however suffers from the problem of only being able to define upper bounds to the statistical errors on the values of  $I(a)$ .

In typical climate applications both methods require a substantial amount of data to go beyond the Gaussian fluctuations [96,217]. In the direct method the estimate of the non-Gaussian tails of the rate function is corrupted by the inability of properly computing the probability density function in ranges dominated by sparse data and outliers. In the indirect method the estimate of the tails of the scaled cumulant generating function becomes artificially linear for large values of  $k$ . For these values the estimate of the generating function is dominated by the contribution of the outliers of  $a$ . Both methods therefore fail to provide reliable estimates more or less for the same range of values of the fluctuations. A solution to the problem of estimating the tails of the large deviation functions in numerical models is given by the use of rare event algorithms, as discussed in [217] and Sect. 4.3.

Regardless of the method one chooses, there is always a delicate interplay between the auto-correlation time of the process, the mixing time, the time scale of the block averaging, and the time necessary to converge to the large deviation limit, that has to be considered. First of all, for a simple process with exponential auto-correlation function, the integral auto-correlation time and the mixing time are of the same order of magnitude. However, other cases the picture can be more complicated, and the integral auto-correlation time may not be a good “time unit” to estimate the time scales of convergence to the large deviation limit [217]. Secondly, for time series of finite length there is a practical trade off between  $\tau_b$  and  $N_b$ . Larger values of  $\tau_b$  mean better convergence to the large deviation limit, but a smaller number of samples  $N_b$  and larger statistical errors. On the other hand, larger values of  $N_b$  mean good statistics and small statistical errors, but poor convergence (if at all) to the limit values. This mirrors the difficulties usually encountered in the statistical analysis of extreme events using EVT [47,178].

It is therefore important when performing these analysis to provide a systematic study of the convergence and of the statistical errors, to identify the best compromise and assess the robustness of the statistical estimators used. As a general rule it would be probably better to use both approaches side by side, as suggested by [155]. See

Sect. 4 for a discussion on different analysis performed on climate data by [94,96,217,219].

### 3.4 Large deviation laws in chaotic systems

In the previous subsections we have shown how large deviation laws emerge when looking at the statistical properties of stochastic dynamical systems. A different point of view on the problem suggests that it is possible to establish foundations for the study of non-equilibrium systems by taking advantage of the framework of chaotic dynamics [97,230]. More specifically, the idea is that non-equilibrium ensembles can be described by the Sinai–Ruelle–Bowen (SRB) measure supported by the attractors of Axiom A systems [78,229]. These concepts are briefly and informally recapitulated below.

Let's consider a flow on a smooth compact manifold  $\mathcal{M}$  of dimension  $n$  such that  $S^t x_0$  is the evolution at time  $t$  of the  $t_0 = 0$  initial condition  $x_0 \in \mathcal{M}$ . Such evolution can be represented in differential form as  $\dot{x} = b(x)$ , where we have removed the stochastic component from (21). We assume that the flow has a compact invariant set  $\Lambda$  such that  $S^t \Lambda = \Lambda$  for all  $t \geq 0$ . We also assume that  $\Lambda$  is not decomposable in two sets that are also invariant and that there is a neighbourhood  $U$  of  $\Lambda$  such that  $U \supset \Lambda$ ,  $S^t U \subset U \forall t \geq 0$  and  $\Lambda = \bigcap_{t \geq 0} S^t U$ .  $U$  is also called the forward isolating neighbourhood of  $\Lambda$ . We assume that  $0 < m_n(U) < \infty$ , where  $m_n$  is the  $n$ –Lebesgue measure. At practical level, one can think  $U$  as a finite-precision approximation of the true attractor  $\Lambda$  and is the asymptotic set that is de facto experimentally accessible in numerical simulations and experiments. Indeed, one can assume that, if an orbit is initialised in the basin of attraction of  $\Lambda$  (the union of all orbits which converge towards  $\Lambda$ ), its forward evolution after a possibly long transient enters the set  $U$ .  $U$  is contained in the basin of attraction.

We now assume that on  $\Lambda$  the flow is hyperbolic, which means that in  $\Lambda$  we can continuously split the tangent space as the sum of three nontrivial subspaces  $T_\Lambda \mathcal{M} = E^s + E^u + E^n$ , where there are constant  $c$  and  $\lambda$  such that  $DS^t(v) \leq c \exp(-\lambda t)|v|$  if  $v \in E^s$  and  $DS^{-t}(v) \geq c \exp(-\lambda t)|v|$  if  $v \in E^u$ ; additionally, the dimensionality of  $E^s$  and  $E^u$  is constant in  $\Lambda$ . In simpler terms, infinitesimal perturbations grow if initialised along  $E^u$  (unstable component) and shrink if initialised along  $E^s$  (stable component). Finally, we assume that  $E^n$ —the neutral space—is one-dimensional and associated with the direction of the flow. No contraction nor expansion takes place along  $E^n$ . We finally assume that  $\Lambda$  is densely populated by (unstable) periodic orbits.<sup>3</sup> We then have that  $\Lambda$  is an Axiom A attractor and the evolution law  $S^t$  defines an Axiom A system. Note that if the flow is on the average contractive ( $\nabla b < 0$ ), the Hausdorff dimension of the  $\Lambda$  is strictly smaller than  $n$  [78]. Therefore, choosing an initial condition randomly (with respect to the natural Lebesgue measure) in the set  $U$ , there is zero probability to choose a point belonging to  $\Lambda$ . This further clarifies the experimental relevance of  $U$ .

<sup>3</sup> This last hypothesis, which seems unnatural, has important consequences both at dynamical and statistical level, see [52].

In general, any invariant measure  $\nu$  is such that

$$\int \psi(y) \nu(dy) = \int \psi(S^t y) \nu(dy) = \int \psi(y) \Pi^t \nu(dy) \quad (33)$$

where  $\Pi^t$  is the so-called transfer operator, which pushes measures forward in time [12]. An invariant measure is a fixed point of the transfer operator for all  $t \geq 0$ :  $\Pi^t \nu = \nu$ . For Axiom A systems one can define a special SRB ergodic measure  $\mu_{SRB}$  with support on  $\Lambda$  such that for almost all (with respect to the measure  $m_n$ )  $x \in U$  and for each continuous observable  $\psi$ , we have that

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \psi(S^\tau x) d\tau = \int \psi(y) \mu_{SRB}(dy) = \mu_{SRB}(\psi). \quad (34)$$

In other terms, long time averages computed from initial conditions in  $U$  give the expectation value computed according to the invariant measure  $\mu_{SRB}$  on  $\Lambda$ . It is very important to note that we are not requesting that the initial condition is on the attractor  $\Lambda$ , but in its neighbourhood  $U$ , which has finite measure, and that is, physically speaking, experimentally accessible. The previous equation implies that after a certain transient almost any trajectory initialised in  $U$  explores the attractor  $\Lambda$  according to invariant measure  $\mu_{SRB}$ . This measure is, indeed, the one that is selected by any finite-precision operation on the system. The physical relevance of  $\mu_{SRB}$  is further supported by the fact that it coincides with the zero-noise limit of the invariant measure realised when one consider stochastic perturbations of the system above [229].

The mathematical setting given above of strange Axiom A attractors gives a possible (yet restrictive) setting for studying chaotic systems. Chaos is usually associated with the *negative* property that divergence of nearby trajectories leads to having a limited time horizon of deterministic prediction. This is the celebrated butterfly effect first discussed by Lorenz [171]. The limits posed by the butterfly effect provide the fundamental reason why improving the skill of a numerical weather forecast system is excruciatingly difficult; see [137] for an example of application of the splitting between stable, unstable, and neutral portions of the tangent space in the context of atmospheric predictability. On the other hand, (34) shows that chaos makes it possible to reconstruct ensemble averages even if we start outside the attractor (but, clearly, within its basin of attraction). Hence, we are able to collect the statistical properties of the system without knowing where precisely its attractor is. Therefore, chaos makes it possible to define at all the climate as the set of statistical properties of the climate system, and makes it operationally feasible to run climate models and interpret their results [104].

Unfortunately, Axiom A systems are far from being generic or even typical, as more general, weaker notions of hyperbolicity have to be used to deal with real-life chaotic systems. Recently, it has been shown that much larger classes of dynamical system possess SRB measure [45,283], thus providing further support

to the so-called *chaotic hypothesis* [97,98], which states, roughly speaking, that a chaotic system with many degrees of freedom de facto behaves as an Axiom A system and in particular possesses a physically relevant SRB-like invariant measure.

### 3.4.1 Large deviation laws for Axiom A systems

We are now able to formulate more precisely the problem of estimating the probability of large deviations of a smooth observable  $\psi$  for the system defined above. We want to study the rate of convergence of the average  $\psi_{T,x} = \frac{1}{T} \int_0^T \psi(S^\tau x) d\tau$  for  $x \in U$ . In other terms, we want to understand the probability of deviations of finite time averages with the respect to the asymptotic result given in (34) which is valid for almost all  $x \in U$ . We adapt below—in a very simplified way—for the case of flows the treatment of the problem presented in [41,282] for maps.

We choose a value  $a \in \mathbb{R}$  and define the set  $B_{T,\psi}^{a,+} = \{x \in U | \psi_{T,x} \geq a\}$ . From (34), we derive that  $\lim_{T \rightarrow \infty} \frac{1}{T} \log \mathbb{P}(B_{T,\psi}^{a,+}) = 0 \quad \forall a > \mu_{SRB}(\psi)$ , where  $\mathbb{P} = (1/m_n(U))m_n$  is a conditional probability measure on  $U$ . It is indeed possible to establish in general a large deviation principle for the finite-time averages of  $\psi$ . We obtain:

$$\lim_{T \rightarrow \infty} \frac{1}{T} \log \left( \mathbb{P}(B_{T,\psi}^{a,+}) \right) = - \inf_{z \in A} I_\psi(z) \quad \text{with } A = [a, \infty). \quad (35)$$

where  $I_\psi(z)$  is the rate function. If, instead, we set  $a < \mu_{SRB}(\psi)$  and define the set  $B_{T,\psi}^{a,-} = \{x \in U | \psi_{T,x} \leq a\}$ , we obtain:

$$\lim_{T \rightarrow \infty} \frac{1}{T} \log \left( \mathbb{P}(B_{T,\psi}^{a,-}) \right) = - \inf_{z \in A} I_\psi(z) \quad \text{with } A = (-\infty, a]. \quad (36)$$

These two results closely mirror what presented in (3)–(4). Note that the functional form of the rate function is known but is very non-trivial, as it must take into account the complex nonlinear correlations of the time evolving value of  $\psi$  resulting from the chaotic dynamics. One can derive the following expression for the rate function:

$$I_\psi(z) = - \sup_{v \in \mathcal{N}, v(\psi)=z} \left( h(v) - \Sigma_\lambda^+(v) \right). \quad (37)$$

where  $\mathcal{N}$  is the set of the invariant measures of the system (if  $v \in \mathcal{N}$  then  $\Pi^t v = v \quad \forall t$ ). Note that we impose as a constraint that the expectation value of  $\psi$  computed using the measure  $v$  must be equal to  $z$ . In the previous expression  $h(v)$  is the Kolmogorov–Sinai entropy, which measures the rate of creation of information, while  $\Sigma_\lambda^+(v)$  indicates the sum of the positive Lyapunov exponents w.r.t.  $v$ . The positive Lyapunov exponents measure the possible asymptotic rates of stretching of infinitesimal perturbations aligned along the unstable manifold [78], see discussion in Sect. 4.2. Note that  $I_\psi(z) \geq 0$  because for all invariant measures  $h(v) \leq \Sigma_\lambda^+(v)$ . The rate function attains its unique minimum for  $z = \mu_{SRB}(\psi)$ , which is realised when  $v = \mu_{SRB}$ .

Indeed, in this case one has that the Pesin identity is verified:  $h(\mu_{SRB}) = \Sigma_{\lambda}^+(\mu_{SRB})$  [229]. Note that the Pesin identity is implicitly assumed as valid in most numerical applications where one wants to evaluate the Kolmogorov–Sinai entropy.

Summarising, the large deviation law defined with respect to initial condition in the neighbourhood of the attractor  $U$  is determined by all the non-SRB invariant measures supported on the attractor  $\Lambda$ . While in the stochastic case the rate function is determined by the cost of deviating from the deterministic trajectory defined by the zero noise limit, in the deterministic case the rate function is, in some sense, determined by the cost of deviating from the reference SRB measure. In geophysical terms, such alternative invariant measures correspond to exceedingly unlikely, yet possible, exotic climates.

Following [41], an alternative way to derive a more direct and practically accessible definition of the rate function relies on using the scaled cumulant generating function described in Sect. 3.3. One defines:

$$\lambda_{\psi}(k) = \lim_{T \rightarrow +\infty} \frac{1}{T} \log \int \left[ e^{k \int_0^T dt \psi(S^t x)} \right] \mu_{SRB}(dx), \quad (38)$$

Taking advantage of the Gärtner–Ellis theorem, whose hypotheses apply in the case of Axiom A systems, one derives the rate function  $I_{\psi}(\Delta)$  as Legendre transform of  $\lambda_{\psi}(k)$  as follows:  $I_{\psi}(z) = \sup_{k \in \mathbb{R}} [zk - \lambda_{\psi}(k)]$ . The possibility of using the same construction for the rate function in both stochastic and Axiom A system clarifies the fundamental similarity, at statistical level, between the two.

Near the minimum of the rate function  $z_0$ , the large deviation law describes the central limit theorem. The results mirror precisely what shown in the previous section. Indeed, mirroring (30), for small values of  $z$  one has  $I_{\psi}(z) \approx (z - z_0)^2 / (2\sigma_{\psi}^2 \tau_{\psi}^c)$ , where  $\sigma_{\psi}^2 = \int \mu_{SRB}(dy) (\psi(y) - \mu_{SRB}(\psi))^2$  is the variance of  $\psi$  and  $\tau_{\psi}^c = \int_{-\infty}^{\infty} dt C_{\psi}(t)$ , where  $C_{\psi}(t) = 1/\sigma_{\psi}^2 \int \mu_{SRB}(dy) (\psi(y) - \mu_{SRB}(\psi))(\psi(S^t(y)) - \mu_{SRB}(\psi))$  is the time-lagged correlation of the observable  $\psi$  and  $\tau_{\psi}^c$  is the integrated correlation time. We can interpret  $\tau_{\psi}^c$  as a normalising factor for time in such a way that  $M$  consecutive observations of the observable  $\psi$  correspond to  $M/\tau_{\psi}^c$  approximately independent stochastic variables.

Note that the possibility of establishing large deviation laws for Axiom A systems is intimately related to the fact that for such systems one observes a rapid decay of correlations for observables (loss of memory being another characterisation of chaotic dynamics). Additionally, under certain conditions it has been possible to prove the existence of large deviation laws also for systems obeying weaker notions of hyperbolicity with respect to the Axiom A case. Such large deviation laws might diverge from the exponential form described above in the case the system has slow decay of correlations [41, 222].



## 4 Applications of large deviation theory to geophysical systems

### 4.1 Large deviation of time averaged observables and rare events

The previous section clearly indicates that we can study persistent extremes in geophysical flows obliviously to the fact of whether we are considering a deterministic or stochastic framework for the dynamics. This is a very important point both on practical and epistemological grounds.

The basic idea that connects a persistent extreme event to large deviations is that the sample mean recorded during the persistent event can be regarded as a large deviation from the long term mean. Nonetheless, to be able to answer the question whether the respective sample mean, besides of being large, represents indeed a large deviation, one has to check the convergence to the large deviation limit based on the rate functions, as described in detail in Sect. 3.3.

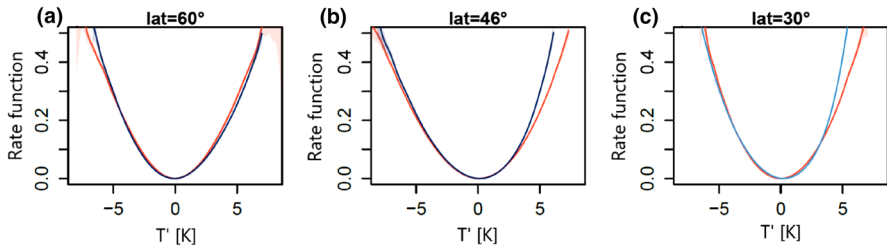
Geophysical observables are usually correlated in time and space, which should be considered in the computation of the rate functions (see also Sect. 3.3). In case of weakly correlated observables (i.e. two values have an exponentially decreasing correlation if they are far enough from each other, in time or in space), it is recommended to re-normalise the rate function in (32) additionally by the integrated auto-correlation  $\tau$  [96]

$$I_n(a) = - \frac{\ln p(A_n = a)}{n/\tau}. \quad (39)$$

Thus, we take into consideration that the averaging block length  $n$  consists of  $n/\tau$  (nearly) independent data points. Please note that the notation in (39) is slightly different from the one used for (32), with  $n$  replacing  $\tau_b$ . The normalisation by  $n/\tau$  is useful especially for comparing rate functions of observables with different characteristic scales.

Gálfi et al. [96] compared rate functions of near-surface air temperature obtained based on (39) and found that by looking locally in space at long time averages agrees with what is obtained, instead, by looking locally in time at large spatial averages along the latitude. They used the simplified General Circulation Model (GCM) of the atmosphere PUMA [90] without orography and performed simulations in a non-equilibrium steady state. These results suggest that, in case of homogeneous statistics in both time and space, the apparent discrepancy between temporal and spatial large deviations is only due to the difference between temporal and spatial scales. If one normalises the rate functions based on the number of independent data, one finds a universal function describing both temporal and spatial large deviations (Fig. 2). Hence, the correspondence between properly scaled temporal and spatial large deviations extends the universality emerging from the asymptotic nature of the large deviation principle, discussed in Sect. 1.2, by a further level, connecting the dimensions of time and space. This connection can be useful for the analysis of observational data, for example. Let us consider the common situation when only time series at sporadic locations are available. If the universality between time and space is satisfied, one can



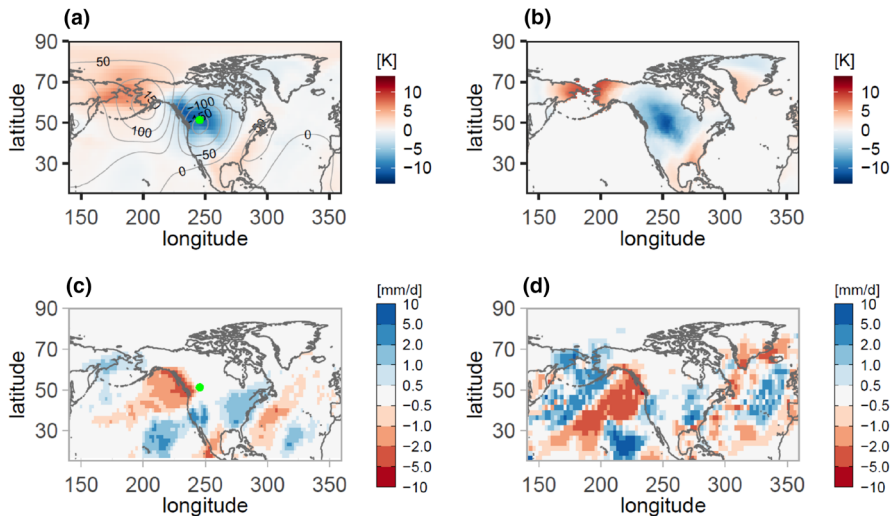


**Fig. 2** The rate functions obtained according to (39) based on averaging the near-surface grid point temperature in time (red) and along a latitudinal band (blue) at latitudes **a** 60°, **b** 46° and **c** 30° in the GCM PUMA illustrate the equivalence between temporal and spatial large deviations, as well as the increasing anisotropy as one goes from North to South, shown by a slight deterioration of the equivalence. From [96]

derive the spatial large deviations based on the information about temporal averages and the characteristic spatial scale.

The above connection between spatial and temporal large deviations is valid for asymptotic scales, i.e. if the rate functions have converged to the asymptotic one. This is fulfilled in the atmospheric model used by [96] for temporal and spatial scales of around  $20\tau$ , where  $\tau$  is either the temporal or the spatial integrated auto-correlation. Obtaining the spatial large deviation law can represent a problem though, due to heterogeneous orography or even due to the limited size of the Earth, which is in certain cases simply not large enough to reach asymptotic levels. Hence, the possibility to make use of the asymptotic connection between temporal and spatial large deviations can be inhibited by a strong anisotropy between the temporal and spatial dimensions. This is illustrated by Fig. 2 showing that going from North towards the Equator, the correspondence between the temporal and spatial rate functions deteriorates slightly, due to the increase of characteristic spatial scales (391 km at 60°, 732 km at 45°, 1292 km at 30°)—and consequently a slower convergence in space—while the temporal scales do not change substantially (1.32 days at 60°, 1.05 days at 45°, 1.61 days at 30°). Besides the purely spatial and temporal large deviations of local temperature observables, the universal rate function describes also large deviations in time of spatially averaged observables, as long as the spatial averaging is performed along the same latitudinal band, and one disregards the effect of orography. However, the spatial averaging length is crucial here. Due to strong spatial correlations on certain synoptic scales ( $\approx 2000$ – $4000$  km), the asymptotic rate function of spatial temperature averages is wider than the universal function, showing that large deviations on these spatial scales are more probable than on any other scales. Interestingly, these spatial scales correspond approximately with the ones of persistent synoptic disturbances, leading to high-impact heatwaves.

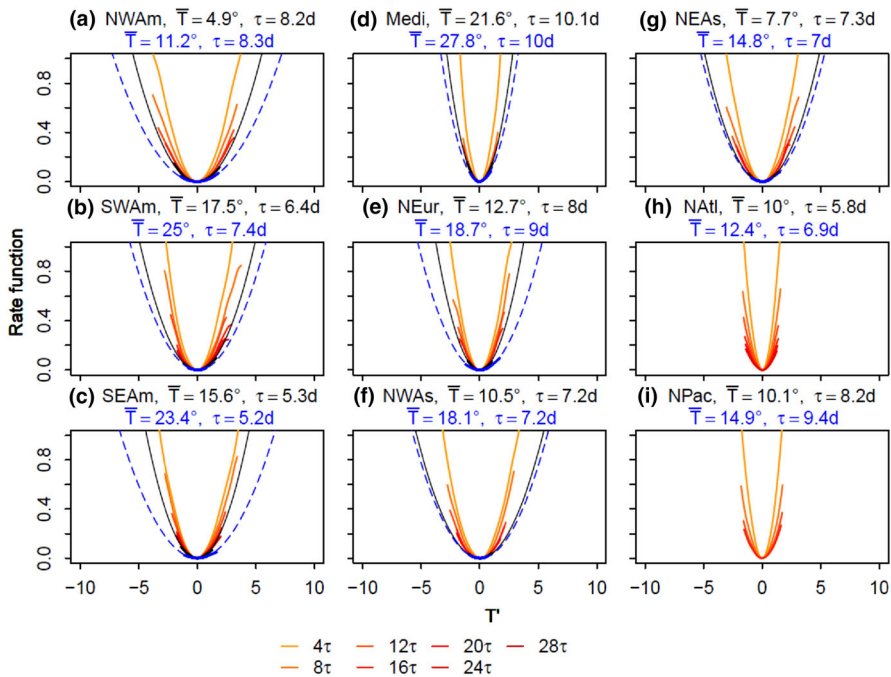
Gálfi and Lucarini [94] studied the connection between large deviations of surface air temperature and persistent temperature events, like heatwaves or cold spells, more thoroughly. They used CMIP6 simulations [81] performed with the state-of-the-art Earth System Model (ESM) MPI-ESM-LR [110] with seasonal cycle and orography. Although, the universality in time and space of large deviations does not hold in the presence of orography, large deviations in time of local temperature are still connected to anomaly fields extended in space and persistent in time. They show that large deviations in time of surface temperature at a single, suitably selected grid point, located



**Fig. 3** Composites of winter **a** surface air temperature and 500 hPa geopotential height and **c** precipitation anomaly fields corresponding to surface temperature anomalies of  $-10.5$  K lasting 30 days in the locale indicated by the green dot from pre-industrial CMIP6 simulations performed with the MPI-ESM-LR model. Observed **b** surface temperature (CRU-TS 4.04) and **d** precipitation (GPCP v2.3) anomalies for February 2019. The isolines in **a** indicate 500 hPa geopotential height anomalies

in the core of the event of interest, are related to spatially extended anomaly patterns (temporal averages of surface temperature and 500 hPa geopotential height anomaly fields), which resemble anomaly patterns from reanalysis data during two high-impact persistent events, the 2010 Russian Heatwave and the 2010 Mongolian Dzud. For the model results, [94] use non-equilibrium steady state simulations with pre-industrial  $\text{CO}_2$  concentration, thus pointing out that both kind of events are manifestations of the natural variability of the climate system. Furthermore, these events seem to be typical persistent events from the perspective of LDT. The fact that by using LDT one is able to capture the dynamical features of typical events—within the class of the very unlikely ones—is further explored later in Sect. 4.4; see also discussion in [112]. Thus, it seems that, based on large deviations, we can identify spatial structures of typical persistent events of our system, and, at the same time, obtain probability estimates for their occurrence. We note that the term “typical” does not refer to the magnitude or severity of the event. Furthermore, precipitation anomaly patterns selected based on large deviations of temperature are also similar to the one during the 2010 Russian Heatwave thus supporting the conclusion that the selection method based on large deviations is meaningful from a dynamical point of view. Remote effects are captured as well, as shown by the positive precipitation anomalies over the Indian subcontinent and Pakistan corresponding to the devastating floods in that region during the Russian Heatwave.

Another case study related to the 2019 North-American cold spell yields as well a noticeable similarity for both temperature and precipitation anomaly fields between the large deviation based selection of modelled (same model as in [94]) fields and reanalysis data for February 2019 (Fig. 3). For the composites in Fig. 3a, c we used



**Fig. 4** Summer rate functions for surface air temperature for increasing averaging windows (see legend) for **a** Northwest, **b** Southwest, and **c** Southeast America, **d** the Mediterranean, **e** North Europe, **f** Northwest and **g** Northeast Asia, **h** the North Atlantic, and **i** the North Pacific. The black (blue dashed) line represents rate functions obtained via multi-seasonal averages for the pre-industrial (quadruple CO<sub>2</sub>) run. On top, the mean surface temperature and integrated auto-correlation,  $\tau$ , for the control (black) and quadruple CO<sub>2</sub> (blue) runs. From [94]

CMIP6 pre-industrial control runs performed with the MPI-ESM-LR model, while Fig. 3b, c represents observed ST (precipitation) anomalies from February 2019 with respect to the 1981–2010 long-term monthly mean based on the CRU-TS 4.04 [123] (GPCP v2.3 [3]) data set.

As we have seen, LDT provides the probability of large sample averages, which then can be related to high-impact persistent events. Figure 4 shows rate function estimates from [94] of summer surface air temperature for different regions over the Northern Hemisphere for pre-industrial and quadruple CO<sub>2</sub> concentration experiments (MPI-ESM-LR model). By comparing the best estimates of the pre-industrial (black lines) and quadruple CO<sub>2</sub> (blue dashed line) experiments, we notice that the rate function becomes wider over North-American and European regions as an effect of the increased CO<sub>2</sub> concentrations, suggesting that heatwaves in summer become more frequent and longer lasting. Opposite results are found for cold spells during winter [94]. Although these results are unsurprising considering the effects of global warming, they show the utility of large deviation rate functions to quantify the changing probability of the considered persistent events.

To a good approximation, rate functions of temperature found in [94,96] are symmetric and close to a parabolic form. Consequently, Gálfi and Lucarini [94] propose to

use the formula for the Gaussian approximation to estimate the probability of observing an average anomaly of temperature  $T_n$  of amplitude  $a$  over a period of  $n$  days:

$$\log(p(T_n = a)) \approx -nI_T(a), \quad I_T(a) = \frac{a^2}{2\tau_T\sigma_T^2} \quad (40)$$

where  $\sigma_T^2$  is the daily variance of the field and  $\tau_T$  is the integrated auto-correlation of the variable  $T$  in units of days, and  $n \geq 10\tau_T$ . This approximate formula, can be useful for estimating the probability of occurrence of persistent large temperature deviations, i.e. heat wave or cold spells, from easily accessible statistical properties of the temperature fields. We note that the temporal resolution of the time series does not have to be daily, the formula can be used for any arbitrary resolution and for averaged series as well. However, one expects to obtain more robust results by using:

$$p(T_{n'} = a') \approx p(T_n = a) \exp\left(\frac{na^2 - n'a'^2}{2\tau_T\sigma_T^2}\right). \quad (41)$$

Based on (41) one can estimate the probability of occurrence of events of amplitude  $a' > a$  and length  $n' \geq n$  from the knowledge of events of amplitude  $a$  and length  $n$ . These results can be reframed in terms of average return periods of the events, which are simply the inverse of the occurrence probabilities.

When dealing with finite sized data, it is advisable to define an optimal averaging block size  $n^*$ . This is usually the minimal block size, for which the rate function belongs to the asymptotic regime, i.e.  $I(n^*) \approx I(n > n^*)$ . In case of the GCM used in [96],  $n^* \approx 20\tau$ , where  $\tau$  is the integrated auto-correlation. In case of the ESM from [94], the optimal averaging length is between 12–20  $\tau$  (1–2 months) for land areas, depending on the geographic region and the considered season. Over the oceans convergence, if any, takes longer than the duration of a season. As mentioned above, [96] and [94] estimate rate functions directly based on the probability density functions (pdf's) of sample averages according to (39), which is a slightly modified version of (32). [217] follow a different strategy and estimate the rate function of spatially averaged temperature over Europe as Legendre transform of the scaled cumulant generating function (29). They rely on an intermediate complexity climate model, run under perpetual summer conditions, and find a convergence to the large deviation limit at  $n^* = 3$  years. Moreover, they find an asymmetric rate function, whereas the ones found by [96] and [94] are generally symmetric.

The studies conducted by [96] and [94], on the one side, and [217], on the other side, show some contradictory results although they both look at temperature observables. Reasons for the discrepancy could be related to differences in the used models, estimation methods, and spatial averaging regions. Let us shortly discuss the possible reasons one by one. Both [96] and [94] find that the convergence to the large deviation limit is quite fast, i.e. shorter than the length of a season, although the used models are extremely different in terms of model physics and complexity. Consequently, the model differences are probably not the main reason for the different results, unless the simplified ocean dynamics in the model used by [217] does not slow down the evolu-

tion of the system too much leading to slow convergence. Despite the fact that the rate function estimation methods based on pdf's and Legendre transforms of the scaled cumulant generating function (see Sect. 3.3) can lead to slightly different results, they should provide similar rate functions [155]. The most feasible reason we can think of is related to the different spatial averaging areas. [217] average over Europe until longitude  $\sim 25^\circ$  E, leaving out the most continental part of Europe and considering all coastal regions, which can be influenced by the slow ocean dynamics. The slowly decaying serial correlation of this spatially averaged observable can lead to the slow convergence of the rate functions estimates.

## 4.2 Large deviations of finite time Lyapunov exponents

After discussing extremely persistent atmospheric states based on large deviations of temperature, we explore, in what follows, the finite time dynamics of the system at a more intrinsic level, based on large deviations of finite time Lyapunov exponents. The goal is to explore whether LDT can be of help for improving our ability to understand the predictability of geophysical flows [105,151,201].

Lyapunov Exponents (LEs) describe the asymptotic growth or decay of infinitesimal perturbation acting on the trajectory of a dynamical system. Their finite time estimates, the so-called Finite Time Lyapunov Exponents (FTLEs), refer to stability properties of a specific state of the system with respect to a predefined predictability horizon [22, 137,203,258,262]. Large deviations of FTLEs point out extremely stable or unstable states of the system [156] and provide relevant information on its predictability on time scales that are intermediate between the one given by the inverse of the first LE and ultra-long ones [155,180].

We briefly and informally introduce the LEs and their finite-time version below. We consider the autonomous deterministic dynamical system where evolution takes place in an  $n$ -dimensional compact manifold  $\mathcal{M}$  obtained by removing the stochastic component from (21):

$$\frac{dx}{dt} = b(x), \quad (42)$$

where  $x \in \mathbb{R}^n$  is the state vector and  $b : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is a smooth drift. We assume that the system possesses an invariant measure  $\rho(dx)$  supported on a compact attractor  $\Omega$ . We define the orbit  $x(t, x_0) = S^t x_0$  as the result of the evolution of the system after a time  $t$  starting from the initial condition  $x_0$  inside the basin of attraction of  $\Omega$ . An infinitesimal perturbation  $\delta x \in \mathbb{R}^n$  evolves along the orbit according to the linearised equation

$$\frac{d}{dt} \delta x(t) = \nabla b \Big|_{x=x(t, x_0)} \delta x. \quad (43)$$

The propagation of the infinitesimal perturbation between time  $t_0$  and  $t$  is described by

$$\delta x(t) = M(t - t_0, x_0) \delta x(t_0), \quad (44)$$

where  $M : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}$  is the so-called tangent linear propagator or resolvent matrix. It follows from (44) that the growth in time of the Euclidean norm of tangent vectors is controlled by the matrix  $M^T M$ . If the system is ergodic (as always assumed in this context), the limit

$$\lim_{t \rightarrow \infty} W(t - t_0, x_0) = \lim_{t \rightarrow \infty} [M(t - t_0, x_0)^T M(t - t_0, x_0)]^{\frac{1}{2(t-t_0)}} \quad (45)$$

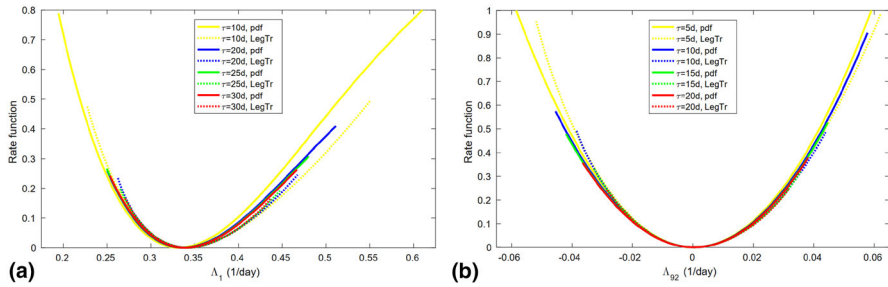
exists [196] and does not depend on  $x_0$ . The logarithm of eigenvalues of  $W$  are called (forward) Lyapunov exponents  $\lambda_i$ ,  $i = 1, \dots, K \leq n$ . Typically (in absence of symmetries), one has  $K = n$ ; additionally, in time-continuous systems one of the LEs has vanishing value as it corresponds to the direction of the flow. If only one vanishing LE is present, the system falls in the special category identified by Pesin as an extremely relevant case of (in general) nonuniform hyperbolicity [207], which provides a natural generalisation to the stricter properties characterising the splitting between the stable and unstable manifold for Axiom A systems, described earlier in Sect. 3.4 [229]. Indeed, the number of positive (negative) LEs, taken with their multiplicity if different from one, defines the dimension of the unstable (stable) manifold; see also discussion in Sect. 3.4. When sorted in a non-increasing order, the LEs form the so-called Lyapunov spectrum. The proposal by Pesin has had enormous importance for giving solidity and strong foundations to the exploration of the tangent space for many systems of practical relevance, and especially so in a geophysical context [137].

In case one is interested in finite time stability properties instead of asymptotic growth/decay rates, one has to consider estimates over a finite time  $\tau$  given by the FTLEs  $\Lambda_i(\tau, x_0)$ , which are the logarithms of eigenvalues of  $W(\tau, x_0)$ . By definition, we have that  $\lim_{\tau \rightarrow \infty} \Lambda_i(\tau, x_0) = \lambda_i$ . Additionally, the long-time averages of the FTLEs computed along the trajectory converge to the corresponding global LEs:

$$\lambda_i = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{t=0}^T \Lambda_i(\tau, S^t x_0) dt = \int \rho(dx) \Lambda_i(\tau, x) \quad (46)$$

where we have used ergodicity. Unlike the LEs, the FTLEs are norm-dependent, thus they depend on the computation method used to obtain them, which can be, for example, based on forward integration (as presented above), backward integration, or following the intersections of Oseledec subspaces [109, 153]. (46) shows that averages in time of FTLEs converge to the global LEs, thus suggesting that finite time and global LEs can be connected by a large deviation law. This is indeed true for Axiom A systems, and, assuming the chaotic hypothesis, for a wide range of chaotic systems which are not Axiom A, as explained in Sect. 3.4.1.

A systematic study of large deviations of FTLEs has been performed on geophysical fluid systems with different degrees of complexity [55, 155, 262]. Using a three-layer quasi-geostrophic (QG) atmospheric model with  $n = 700$ , in [155], a relatively fast convergence was found to the large deviation limit for all LEs. Convergence was slightly slower and rate functions were asymmetric for the strongly positive/negative LEs (Fig. 5a), whereas convergence was very fast and rate functions were symmetric in case of the near-zero and weakly positive/negative LEs (Fig. 5b). [262] studied the



**Fig. 5** Rate function estimates of **a** the first (most positive) FTLE and **b** the 92nd (zero) FTLE in an intermediate complexity atmospheric model. The colours represent the averaging block length, according to the legend. Continuous lines refer to probability density-based estimates, whereas the dashed lines to Legendre-transform based estimates. From [155]

convergence of the FTLE in a coupled ocean-atmosphere QG model with  $n = 36$ . They found a convergence and quadratic rate functions for positive and negative LEs, whereas very slow or even no convergence was found for near-zero LEs.

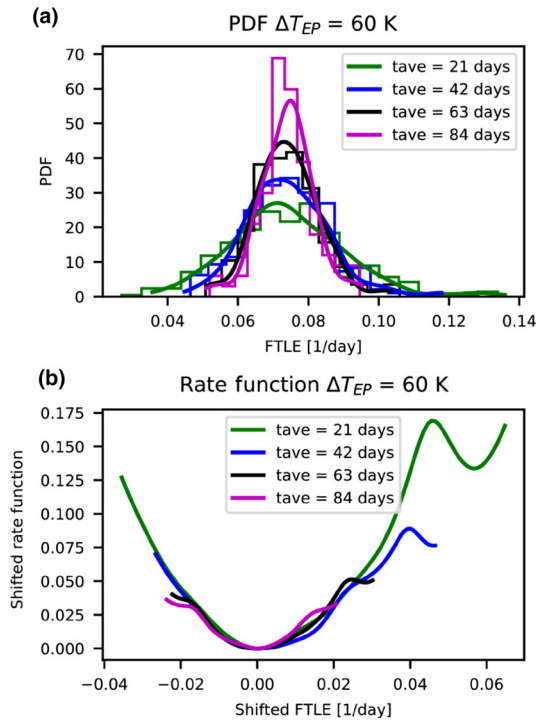
The apparently contradictory results of [155] and [262] show that convergence properties of FTLEs strongly depend on the dynamical properties of the system. The Lyapunov spectra of the two systems clearly show the dynamical differences. Whereas the atmospheric model has a monotonically decreasing Lyapunov spectrum, the spectrum of the coupled ocean-atmosphere model has an extended range of slightly positive and negative LEs, which are indistinguishable from zero, corresponding to the so-called slow manifold, which appears due to the slow oceanic dynamics and the interactions between slow oceanic modes and atmospheric ones.

Generally, for a physical observable, the convergence of finite time averages to the large deviation limit depends on two factors: the strength of the serial correlations and the asymmetry of the probability distribution. Due to serial correlations, larger averaging blocks are needed to reach the asymptotic limit, as compared to i.i.d random variables. Additionally, an eventual asymmetry (or departure from Gaussianity) of the parent distribution delays the convergence of the small Gaussian deviations around the minimum of the rate function, as required by the central limit theorem [96,155]. Accordingly, in case of the mentioned atmospheric model, on the one hand, serial correlations are so weak that the speed of convergence to the large deviation limit depends dominantly on the degree of non-Gaussianity of the FTLE's parent distribution. On the other hand, in case of the ocean-atmosphere model the strong, long-term correlations dominate the convergence (or non-convergence) to the large deviation limit, including the convergence of the FTLE's to the global LEs.

In [55] it was shown that a large deviation law can be found for FTLEs in case of the primitive equations model PUMA [90] run in a standard setting allowing for the establishment of turbulent, Earth-like conditions for the atmospheric flow. The primary injection of energy for PUMA occurs by imposing a fixed Equator-to-Pole temperature difference, as typically done for the so-called dynamical cores of atmospheric models [127]. When such temperature difference is set to the fairly realistic value of 60 K (symmetry is taken between the two hemisphere), one obtains, in the the setting



**Fig. 6** Statistics of the the largest FTLEs in a primitive equation model of the atmosphere. **a** Probability density functions of the largest FTLE for different averaging times. **b** Corresponding estimates of the rate function expressed in terms of the anomaly with respect to the asymptotic value of the first LE. From [55]



described in [55], 68 positive LEs which correspond to a very highly-dimensional attractor (Kaplan–Yorke dimension [78] of 172.6). Figure 6 shows the statistics of the largest FTLE computed with averaging times ranging from 21 to 64 days. One observes a good convergence to a well-defined rate function when longer averaging times are considered. Comparably good properties of convergence have been found for all the FTLEs [55].

Conversely, in [55] no convergence was found in a higher resolution version of the coupled ocean-atmosphere model mentioned above [262]. The authors conclude that the length of the simulations with the coupled model (614 years) is not sufficient to obtain convergence due to the long-term correlations in the system.

#### 4.3 Rare event sampling algorithms based on large deviation theory

Large deviations provide us with valuable insights into the probability with which and the way in which rare events occur. The theory is only valid in specific limits, for example taking the noise intensity in a stochastic differential equation to zero, or the length of a time average to infinity. Nevertheless the results are in many cases still useful when this limit has not been reached, which is the realistic setting we are really interested in.



In this pre-limit cases we may want to estimate quantities that are not available from the instanton, for example

$$\mathbb{P}(x(t) \in B | x(T) \in A) \quad \text{for } 0 < t < T$$

the probability of entering some set  $B$  at an intermediate time  $0 < t < T$ , conditioned on arriving in the rare event set  $A$  at the final time. To obtain such information numerical estimation is usually the only possible approach.

However, we face a difficult challenge in numerically sampling rare events. By definition any brute force Monte Carlo simulation will only contain a small sample of relevant events. This leaves us with large uncertainties on the quantities we want to estimate.

Numerical sampling methods known as rare event simulation (RES) methods have been successful in circumventing these sampling issues in many applications. Although many RES algorithms exist, the approaches are based on the same principle: we want to preferentially sample rare events, in such a way that it is possible to a posteriori determine how much more or less likely we have made the paths we have sampled. This information allows us to estimate probabilities of the system as though we had not interfered with its path distribution.

RES algorithms have been developed since the 50s [136] for a variety of applications in applied mathematics and statistical physics, and have been subject of recent mathematical interest [56]. In recent years there has been several attempts to apply these methods for geophysical applications. They have been applied to Lorenz models [275], partial differential equations [226], turbulence problems [77, 113, 160, 164, 165], geophysical fluid dynamics [28], heatwaves in general circulation models [217–219] and data-based stochastic weather generators [281], and tropical cyclones in regional climate models [212, 272]. Here we give an overview of methods and their applications that have been used to study problems directly related to the dynamics of planetary atmospheres and making use of concepts from LDT. Other applications, including approaches through minimum action methods also related to LDT, are reviewed in [114].

One of the applications we have discussed concerns the statistics of time averages of surface temperature and the study of heatwaves. In order to study long-lasting extremes of surface temperature, [217, 219] have applied a genealogical algorithm to the intermediate complexity climate model PlaSim, and recently to the more complex, state-of-the-art model CESM [218]. The algorithm was specifically designed in [107, 108, 162] to study large deviation functions of time averages, and is very similar to a class of methods described by [57] and whose mathematical properties are studied in details in [56].

The algorithm consists in running an ensemble simulation of  $N$  trajectories with a numerical model starting from different initial conditions. At constant intervals of a fixed resampling time  $\tau$ , each trajectory is assigned a weight, which determines if that trajectory is killed or if it continues its evolution generating copies of itself. The weight  $w_i^n$  of trajectory  $n$  at time  $t_i = i\tau$  is computed in such a way that the weight is large when an appropriately chosen score function is large. Choosing well the score function in order to define the selection criteria is critical. In [217–219] the

score function used was the time integral of the surface temperature over Europe. The weights are defined in such a way that, after a total running time  $T_a$ , the probability  $\mathbb{P}_k(\{\mathbf{X}(t)\})$  of observing a trajectory in the ensemble generated by the algorithm is related to the probability  $\mathbb{P}_0(\{\mathbf{X}(t)\})$  of observing the same trajectory in a normal ensemble simulation with no resampling as

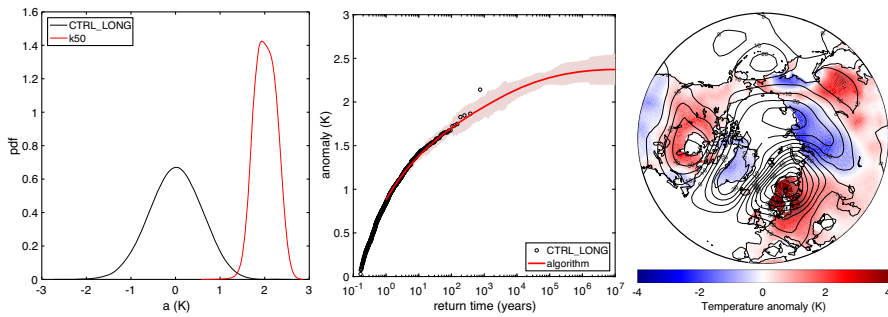
$$\mathbb{P}_k(\{\mathbf{X}(t)\}) = \frac{e^{k \int_0^{T_a} A(u) du}}{Z} \mathbb{P}_0(\{\mathbf{X}(t)\}), \quad (47)$$

where  $Z$  is a normalisation term,  $k$  is a biasing parameter of the algorithm that determines how strong is the tilt of the probability distribution, and the equation is valid in the limit of large  $N$  with relative errors on the computation of expectation values of the order of  $1/\sqrt{N}$  [56].

In ensemble simulations performed with the rare event algorithm, trajectories featuring a large value of the time average of the control observable over the entire simulation period are thus much more likely to be observed than in a normal simulation, and *vice-versa*. This is called *importance sampling*. It is important to note that Eq. 47 applies to the probability of trajectories, not of the score function itself. It can thus be used to compute any statistical property of the system by averaging over the trajectories in the tilted ensemble and reweighting the contribution of each trajectory by the inverse of the factor that multiplies  $\mathbb{P}_0(\{\mathbf{X}(t)\})$  in Eq. (47). Because of the tilting, statistical estimators based on (47) of quantities conditional on the occurrence of extremes of the time average of  $A$  have statistical uncertainties orders of magnitude smaller than with direct sampling, for a given computational cost.

This method as developed by [107,108,162] gives optimal estimates of the scaled cumulant generating function of the time average of the control observables for the value of  $k$  used as biasing parameter. In fact the normalisation term is the generating function  $Z = \mathbb{E}[e^{k \int_0^T A(t) dt}]$ , and an estimate of it is computed directly with a dedicated estimator based on the statistics of the weights responsible for the cloning of the trajectories. See [217] for a full analysis on this type of application. It is worth noting that the same method has been used by [253] to compute the large deviation properties of the finite time Lyapunov exponents of a selection of simple dynamical system, highlighting peculiar properties of the chaotic dynamics associated to these systems.

The method however proves extremely useful also to study the statistics of time averages not in the large deviation limit, as shown in [218,219]. Simulations performed with the algorithm allowed to perform importance sampling of seasonal anomalies of the European surface temperature (left panel of Fig. 7), in a condition in which convergence to the large deviation limit was far from being achieved. This allowed to simulate and estimate precisely the return times of rare and ultra rare anomalies corresponding to seasonal heatwaves with return times up to millions of years, with computational costs 3 order of magnitude smaller than what would have been necessary with direct sampling (centre panel of Fig. 7). The access to the dynamical trajectories leading to rare heatwaves allowed to compute precisely composite statistics of anomalies of surface temperature and 500 hPa geopotential height conditional on the occurrence of heatwaves with return time larger than 1000 years. This shows how this very rare and

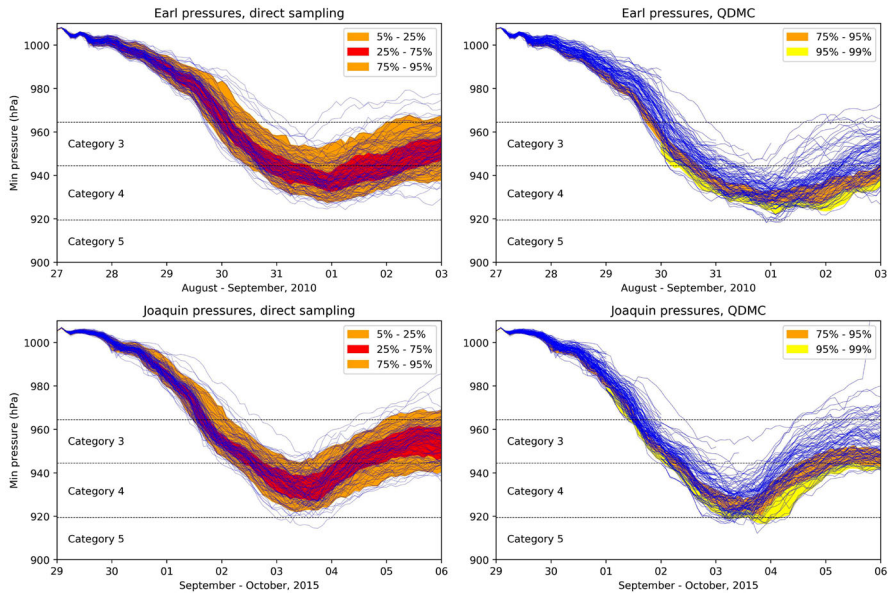


**Fig. 7** Left: importance sampling of average European surface temperature anomalies in ensemble simulations with a large deviation rare event algorithm applied to the general circulation model PlaSim, with the distribution obtained with the algorithm (red) shifted to the tail of the distribution obtained with a direct Monte Carlo approach (black), from [219]. Centre: return times of simulated average European surface temperature anomalies from the same experiments, showing how the application of the rare event algorithm allows to simulate rare and ultra rare events (red) compared with what possible with a direct Monte Carlo approach (black), from [219]. Right: low wavenumber teleconnection pattern of anomalies of surface temperature (red) and 500 hPa geopotential height (contours) in the composite averages of rare heatwaves (return time large than 1000 years) obtained with the large deviation rare event algorithm in the same experiments, from [219]

intense events are related to the occurrence of a hemispheric teleconnection pattern of low wavenumber between 3 and 4 (right panel of Fig. 7). The presence of these stationary patterns (although with varying wavenumbers) seems ubiquitous in analysis of extreme heatwaves with different approaches (see discussion in Sect. 4), which makes rare event algorithms a very promising tool to investigate the dynamic constituents and drivers of these events, thanks to the much better sampling they can provide. This approach has also been applied successfully to stochastic weather generators based on circulation analogues to simulate persistent rare European heatwave [281]. Indeed, the fact that the algorithm is able to select the correct dynamical patterns that generate extremes of a target observable, makes its pairing with analogue classification and selection analysis potentially extremely interesting.

A similar method has been used by [212,272] to study the intensification of tropical cyclones in the high resolution, non-hydrostatic regional climate model WRF. In this case the score function has been taken as the surface pressure anomaly at the centre of the cyclone. Additionally, the trajectory selection and cloning procedure has been modified by mapping at each resampling step the distribution of the weights on a Gaussian distribution on an equal-quantile basis [272]. This procedure helps to avoid that for large values of  $k$  the trajectory with the largest weight in the ensemble dominates all the others in the cloning rate, thus reducing the degree of degeneracy of the trajectories in the resampled ensemble. Applying this method [212,272] were able to obtain a large number of trajectories of extremely intense tropical cyclones in simulations featuring boundary conditions corresponding to two historical tropical cyclones (Fig. 8), thus showing the potential of this approach even with some of the most challenging applications in climate modelling.

It is worth noting that the resampling methods used in [212,217–219,272] belong to the same family of methods used in particle filtering by the data assimilation commu-

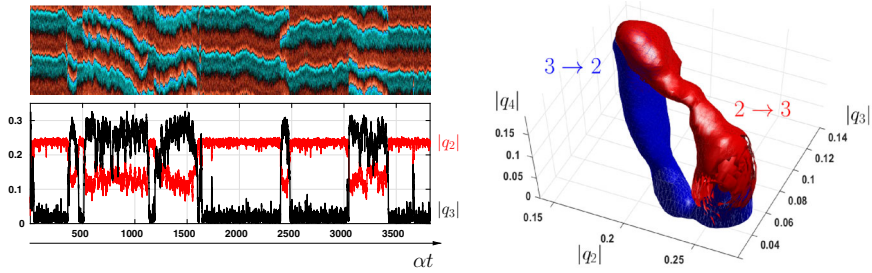


**Fig. 8** Intensity of tropical cyclones Earl (top) and Joaquin (bottom) in direct Monte Carlo simulations with the regional climate model WRF (left) and in ensemble simulations with a quantile Diffusion Monte Carlo rare event algorithm applied to the same model, showing the shift of the trajectories to higher percentiles of the intensity range thanks to the rare event algorithm. Reproduced with permission from [272]

nity [37,73,260]. Particle filtering is a notoriously difficult problem in data assimilation [244] due to the difficulty to target specific trajectories in a very high dimensional space. Particle filtering has nonetheless shown promising advances in recent years [260]. Although not directly related to extreme events or LDT, it would be interesting to see how developments in particle filtering and in the emerging field of the application of RES algorithms to climate modelling could possibly inform and support each other.

Another method that has been successfully applied to problems of interest for the geophysical and climate community is the Adaptive Multilevel Splitting (AMS) algorithm [39,53]. This method is particularly well suited to study rare transitions between two metastable states or attractors A and B. A score function or reaction coordinate is defined to measure the distance of a trajectory from say the target set B. An ensemble of  $N$  trajectories is then simulated starting from the set A, until all of them end in either A or B. Then the worst performing trajectory (the one falling in A that was the furthest from B in its evolution) is replaced by a new trajectory whose initial condition is taken on one of the other better performing  $N - 1$  trajectories. This step is the resampling step of the algorithm, and is repeated  $K$  times. The final ensemble of trajectories will have probability  $(1 - 1/N)^K$ , and it will be populated by many extremely rare transitions from A to B, which can be used to compute unbiased estimates of their probability. See [28,226,227] and references therein for more details.

The AMS algorithm has been used among other applications to study transitions to turbulence [225], the forces acting on a solid object in a turbulent flow [164,165],



**Fig. 9** Left: transitions between attractors with 2 and 3 jets in a stochastic barotropic beta-plane quasi-geostrophic model of Jupiter atmosphere: Hovmoeller diagram of zonal mean vorticity (top) and time series of wavenumber 2 and 3 Fourier components of zonal mean vorticity (bottom). Right: reactive tubes showing the concentration of the transition paths around instantons for the transitions between 2 and 3 jets states in the model of Jupiter atmosphere, obtained with the Adaptive Multilevel Splitting rare event algorithm. Reproduced with permission from [28]

and transitions between metastable states in geophysical fluid dynamics [28]. For example, [28] studied the transitions between metastable states with two and three jets in a stochastic barotropic beta-plane quasi-geostrophic model of the atmosphere of Jupiter (left panel of Fig. 9). Thanks to the application of the AMS algorithm, [28] were able to obtain thousands of trajectories representative of the rare transitions between the two turbulent attractors, showing how they cluster around preferential paths close to an instanton (right panel of Fig. 9). This was the first demonstration in a numerical simulation of a turbulent flow of this type of phenomenology. The AMS algorithm has thus proved to be a powerful tool to study rare transitions between different chaotic attractors, that could be applied also to climate studies where such transitions occur (see Sect. 4.5).

As said several other RES methods exist, tailored to different types of studies. The application of RES algorithms to geophysical and climate problems is an emerging field of research with a very strong potential. The close connection between large deviations and rare event simulation, both providing information on rare event probabilities, begs the question whether they can be usefully combined to provide insights on each other in this context. In particular, it would be extremely interesting to understand for a wider class of processes and applications on one hand if RES algorithms can be useful in estimating large deviation rate functions in systems and models of the complexity of the ones typically involved in applications, and on the other hand if knowing the instanton to a rare event set could be useful to set up an efficient RES algorithm to study the corresponding physical process.

#### 4.4 Rogue waves

Rogue waves are unusually large and virtually unpredictable surface waves that pose grave hazards to boats as well as to naval and coastal infrastructures, and to people living in coastal areas, as they can lead in open sea to tens of meters high water walls. They are traditionally defined as deep-water waves whose crest-to-trough height is at least twice as large as the significant wave height, which in turn is defined as four times

the standard deviation of the ocean surface elevation [2,65,193,241]. Traditionally, the understanding of such rogue waves, which are found with similar phenomenology in systems as different from the ocean as optical fibres [4] and plasma [11], has been approached according to two separate viewpoints. One sees rogue waves as emerging from linear superposition effects [168], the other one, instead, sees rogue waves as eminently nonlinear phenomena [17,66] where a solitonic structure emerges via nonlinear focusing [256].

Recently, a unified theory of rogue waves based on LDT has been proposed [60,61]. The fundamental idea is that rogue waves can be seen as instantons computed by minimising the action associated with the dynamics described by a one-dimensional nonlinear Schrödinger equation (NLSE) with random initial data. The initial data are chosen in such a way that their spectrum agrees with the spectrum of ocean wave height measured in an oceanographic observational campaign in the North sea, namely the Joint North Sea Wave Project (JONSWAP) [126]. The idea is to study the probability of having a wave that increases to a height larger than a given threshold  $z$  over a time  $T$  from one of the possible initial conditions:

$$P_T(z) = \mathbb{P}(F(u(T)) > z) \quad (48)$$

where  $[0, L]$  is the domain of the system,  $F(u(T)) = \max_{x \in [0, L]}(S^T u_0)$ ,  $\mathbb{P}$  is the probability computed over the distribution defining the initial data  $u_0$ , and  $S^T$  is the evolution operator up to time  $T$  defined by the NLSE. The next step is to look at the tail of the surface height distribution and write  $P_T(z)$  as a large deviation law of the form  $P_T(z) \approx \exp(-I_T(z))$  and define, in a way analogous to what discussed in Sect. 3.2<sup>4</sup> the instanton as the trajectory minimising the rate function  $I_T(z)$ .

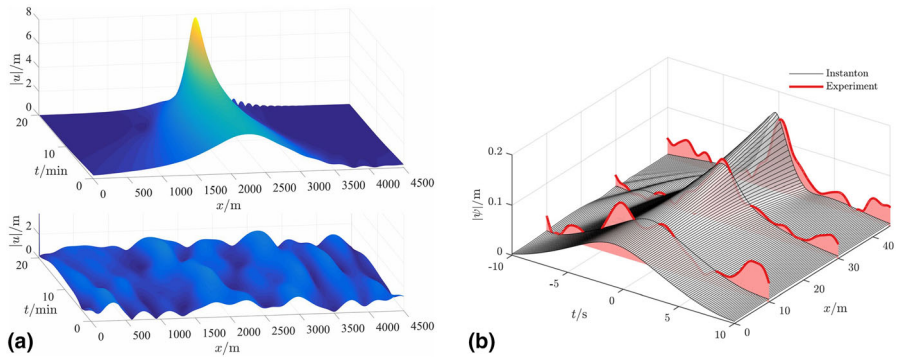
Numerical simulations clearly indicate that rogues waves emerge from fields—precursors—that are typical in terms of intensity but special in terms of spatial pattern—see Fig. 10a. Especially impressive is the fact that experimental results obtained in large water tanks suggest that instantonic solutions do resemble well actual observed rogue waves—see Fig. 10b. This latter result is—conceptually—in agreement with what shown in Sect. 4.1 regarding the good correspondence between observed heatwaves and cold spells and LDT-tailored climate variability generated by an Earth system model, see Fig. 3. Again, one can interpret such predictive power in terms of the universality resulting from the fact that LDT describes the most likely among events that are extreme and overall very unlikely.

#### 4.5 Metastability and noise-induce transitions across melancholia states

As mentioned earlier in the paper, a key problem in geosciences is the investigation of metastable systems and of the properties of transitions between the competing modes of operation. And, indeed, there is a complex phenomenology of multistability for the Earth system at involving different temporal and spatial scales. We describe below how methods and ideas related to LDT can help us better understand such features.

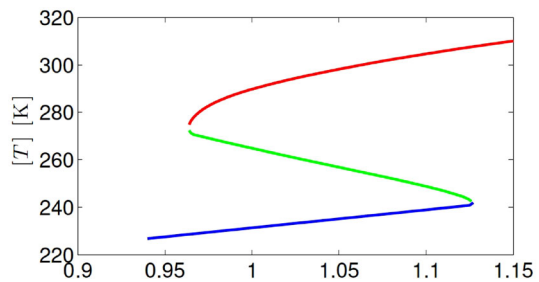
<sup>4</sup> But note that here the noise is included as random fluctuations of the initial conditions, whereas the evolution defined by the NLSE is fully deterministic.





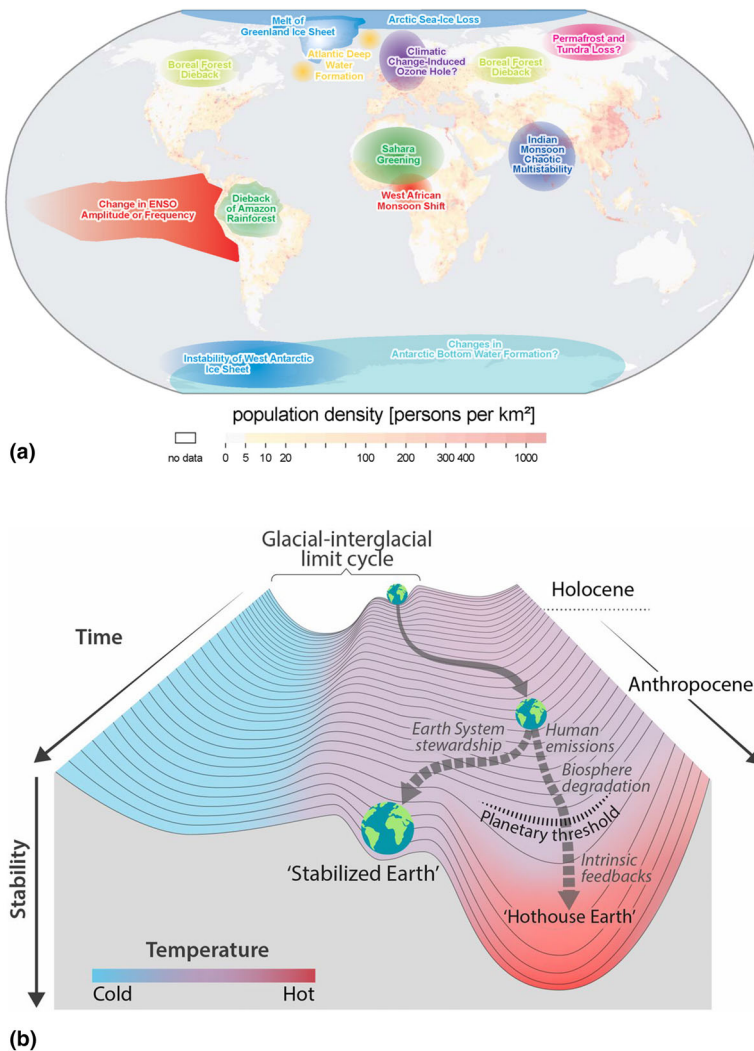
**Fig. 10** **a** The rogue wave precursors are wave patterns of regular height, but with a very specific shape. Upper panel: precursor leading to a rogue wave. Bottom panel: regular wave field. From [61]. **b** The instanton computed via LDT (grey surface) resembles well individual rogue waves (light red lines) observed in a large water tank experiment. From [60]

**Fig. 11** Bifurcation diagram of the Ghil–Sellers model [102] as a function of the ratio between the solar irradiance and the present-day one. The order parameter is the globally averaged surface temperature. The stable W and SB states and the unstable M state are indicated by the red, blue, and green solid line, respectively. Reproduced with permission from [20]



The current astronomical and astrophysical conditions support the co-existence of (at least) two competing states, corresponding to the snowball (SB) state and warm (W) state. Such states are so different that they seem to correspond, counter-intuitively, to two entirely different planets. The multistability is, by and large, made possible by the competition of two feedbacks: the negative Boltzmann feedback (a warmer planet emits more radiation to space) and the positive ice-albedo feedback (a colder planet stores more water in ice form; the surface ice reflects more radiation to space). Figure 11 portrays a bifurcation diagram taken from [20] where multistability is given in terms of globally averaged surface temperature, while the control parameter is the ratio between the solar irradiance and its present value. The W solution is separated from the SB solution by an unstable solution—a saddle in the phase space, which we will refer to as Melancholia (M) state [19,175–177]. The bifurcations of the system take place when one of the stable states meet the M state, according to the scenario of basin crisis [197].

On a less dramatic scale, modulations in the parameters of the climate system can lead to the occurrence of smaller scale critical transitions that have very strong impacts on specific climatic subsystems, the tipping points [163], see Fig. 12a. Recently, it has been proposed that future scenarios of climate change (*trajectories of the Anthropocene*) might be loosely seen as dynamically determined by a motion on some energy



**Fig. 12** **a** Tipping elements of the climate system. From [163]. **b** Cartoon of how the so-called *stability landscape* (here: quasi-potential) changes as a result of anthropogenic forcings. From [249]

potential, which might experience multiple local minima corresponding to competing metastable states [249], see Fig. 12b.

As we are considering a very high-dimensional system that cannot be reduced to a one-dimensional ordinary differential equation, linking minima of some dynamically-relevant potential and competing metastable states requires some nontrivial mathematical framework, which is close in spirit to Waddington's *epigenetic landscape* metaphor in evolutionary biology [9,86,130,268].

We take the point of view proposed by the Hasselmann programme [125] discussed in Sect. 3.1 and assume that the coarse-grained evolution of the climate system can



be written as in (21). We also assume that the deterministic evolution law  $\dot{x} = b(x)$  obtained by switching off the noise can be seen as given a smooth flow taking place in a compact  $n$ -dimensional manifold. If stochastic forcing is absent, the initial condition  $x_0$  determines the asymptotic state of its orbit. If several asymptotic states, defined by the attractors  $\Omega_j$ ,  $j = 1, \dots, J$ , are present the system is multistable. The phase space is divided between the basins of attraction  $B_j$  of the attractors  $\Omega_j$  and the basin boundaries  $\partial B_l$ ,  $l = 1, \dots, L$ , which possess a set of saddle points  $\Pi_{lk}$ ,  $l = 1, \dots, L$ ,  $k = 1, \dots, k_{\max} \geq 1$ . Such saddle points—the M states—attract initial conditions on the basin boundaries [119,157,266] and can be computed using the so-called edge tracking algorithm [240].

We now switch on the stochastic forcing. The presence of noise makes it possible for transitions between the competing metastable states to take place [93,118,122]. Under fairly general conditions, in the limit  $\varepsilon \rightarrow 0$ , one can propose the following ansatz for the invariant measure in the form of a large deviation law

$$\rho_\varepsilon(x) \sim Z(x) \exp\left(-\frac{2\Phi(x)}{\varepsilon}\right), \quad (49)$$

where  $\Phi(x)$ , a non-equilibrium generalisation of the notion of free energy, plays the role of a rate function, while  $Z(x)$  is a pre-exponential factor.  $\Phi$  obeys the following Hamilton–Jacobi equation [101,286]:

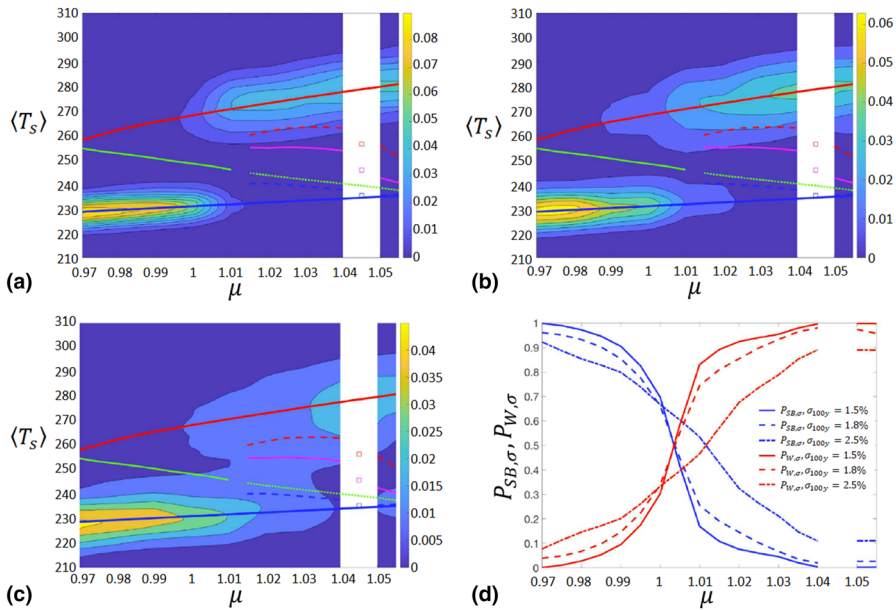
$$F_i(x)\partial_i\Phi(x) + a_{ij}(x)\partial_i\Phi(x)\partial_j\Phi(x) = 0. \quad (50)$$

where  $a(x) = \sigma(x)\sigma(x)^T : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}$  is the noise covariance matrix. Additionally,  $\Phi$  is a Lyapunov function whose decrease describes the convergence of an orbit to the attractor. Specifically,  $\Phi(x)$  has local minima at the deterministic attractors  $\Omega$ 's, and has a saddle behaviour at the saddles  $\Pi$ 's. If an attractor (saddle) is chaotic,  $\Phi$  has constant value over its support, which can then be a strange set [116,121]. Note that these qualitative properties do not depend on spatial patterns of the noise. Instead, the simple fact of adding some noise to the otherwise deterministic dynamics (thus setting  $\sigma > 0$ ) allows for a global exploration of the phase space of the system.

The instantons, which give, in the zero-noise limit, the most probable way to exit an attractor [118,140], can be constructed as minimisers of the Freidlin–Wentzell actions similarly to what has been shown in Sect. 3.2.<sup>5</sup> The instanton is intimately connected to the quasipotential  $\Phi(x)$  in that the local quasipotential  $\Phi_\Omega(x)$  within the basin of attraction of  $\Omega$  is equal to the action for the instanton between  $\Omega$  and  $x$  [24,112,113]. To recover the global quasipotential  $\Phi(x)$ , one needs to resort to a pruning-and-stitching strategy, gluing together the local portions  $\Phi_{\Omega_j}$ ,  $j = 1, \dots, J$ , see [116] and the careful description recently provided by [287]. A separate view on this problem, based upon a different interpretation of the noise has been proposed in [8,284].

Escapes from an attractor  $\Omega$  through a saddle  $\Pi$  into a neighbouring basin are Poisson-distributed events, where the probability that an orbit does not transition up

<sup>5</sup> Some nontrivial issues emerge when the noise matrix is singular; see [182].



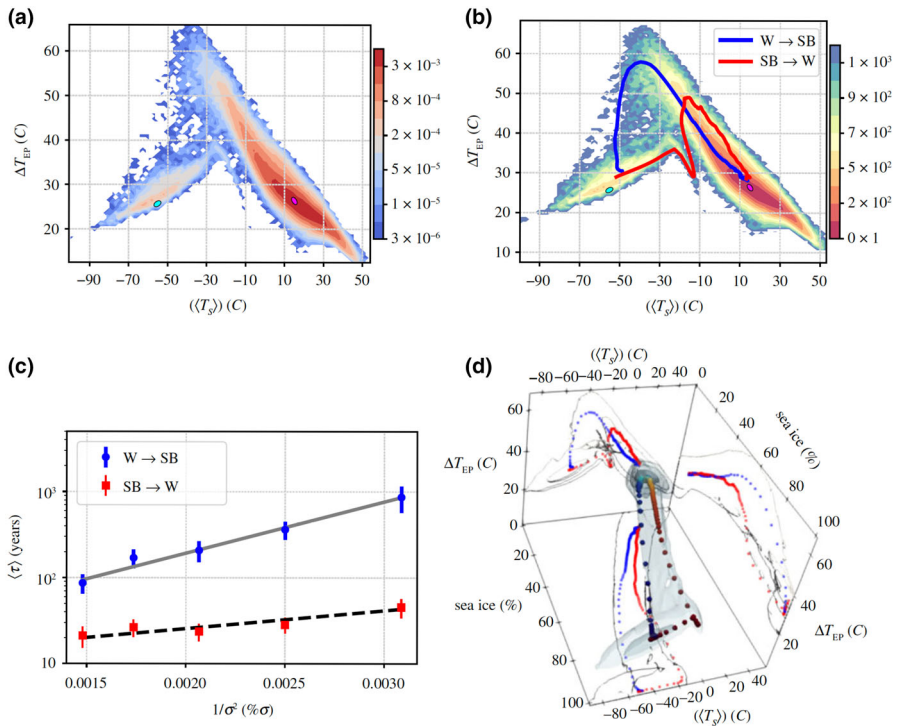
**Fig. 13** Effect of changing the intensity of the noise on the invariant measure of a climate model. **a–c** y-axis: globally averaged surface temperature; x-axis: ratio between the solar irradiance and the present-day. Solid lines: Bifurcation diagram of the deterministic system (W state: red line; SB state: blue line; and M state: green line are of relevance here). Shading: projection of the invariant measure. Noise increases from **a** to **c**. **d** Population of the W state (red lines) and of the SB state (blue lines) for different noise intensity. Here  $\sigma^2 = \varepsilon$ . From [177]

to time  $t$  is, similarly to the classic Kramers' law [148], given by:

$$P(t) = \frac{1}{\bar{\tau}_\varepsilon} \exp\left(-\frac{t}{\bar{\tau}_\varepsilon}\right), \quad \text{with } \lim_{\varepsilon \rightarrow 0} \varepsilon \log \bar{\tau}_\varepsilon = 2\Delta\Phi_{\Omega \rightarrow \Pi}, \quad (51)$$

being the expected escape time and  $\Delta\Phi_{\Omega \rightarrow \Pi} = \Phi_\Omega(\Pi) - \Phi_\Omega(\Omega)$  is the quasipotential barrier height at the relevant saddle [157]. Unfortunately, due to the global stitching procedure, one cannot in general simply read off the barrier height  $\Delta\Phi_{\Omega \rightarrow \Pi}$  from the  $\Phi(x)$  of Eq. 49. While the global quasipotential  $\Phi(x)$  yields information about the relative probability of attractors, and is available e.g. through global sampling of the system, the local notion of potential barriers,  $\Delta\Phi_{\Omega \rightarrow \Pi}(x)$  is relevant for the time-scale of transition events, and can be obtained e.g. by looking at transition times between attractors. Equivalence between the information provided by the local and global quasipotentials is realised if the system is an equilibrium one or, more generally, if only two competing states are present with a single saddle embedded in the boundary between the two basins of attraction (Fig. 13) [176,177].

Margazoglou et al. [182] have recently performed a thorough investigation of the noise-induced transitions between the competing SB and W states of the open-source climate model PlaSim [89,179], see Fig. 14. Panel (b) shows the probability density function in the projected space given by the globally averaged surface temperature



**Fig. 14** Properties of the  $W \rightarrow SB$  and  $SB \rightarrow W$  transitions in a stochastically perturbed version of PlaSim [89,179]. **a** Quasi-potential projected in the two-dimensional space given by the globally averaged surface temperature ( $x$ -axis) and difference between surface temperature at low and high latitudes ( $y$ -axis). **b** Projection of the probability density function obtained for 1.8% per century random fluctuation of the solar irradiance and estimates of the weak-noise limits of the  $W \rightarrow SB$  and  $SB \rightarrow W$  transition paths. **c** Change in the  $W \rightarrow SB$  and  $SB \rightarrow W$  transition times as a function of the intensity of the noise (note the logarithmic scale in the  $y$ -axis). **d** Three-dimensional representation of the  $W \rightarrow SB$  and  $SB \rightarrow W$  transition paths, where the sea-ice percentage acts as third dimension. Here  $\sigma^2 = \varepsilon$ . From [182]

and by the difference between surface temperature at low and high latitudes, and the best estimates of the transition paths in the weak-noise limit, while panel (a) shows the estimate of the projected quasi-potential  $\Phi$ . Additionally, panel (c) shows the exponential dependence of the average transitions times with respect to the inverse of the variance of the noise, as indicated in (51). Note that in this case, because of the complex structure of the basin boundary, the  $W \rightarrow SB$  and  $SB \rightarrow W$  transitions take place through separate M states, as can be seen in panel (d) where a third dimension, corresponding to the sea ice percentage, is included.

Let's now look specifically at the case of bistable systems, where we have two attractors  $\Omega_1$ ,  $\Omega_2$ , and one saddle  $\Pi$ . We can then express the average transitions times as follows:

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \log \tau_{\varepsilon}^{i \rightarrow j} = 2(\Phi(\Pi) - \Phi(\Omega_i)), \quad i \neq j = 1, 2 \quad (52)$$

so that

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \log \left( \frac{\tau_{\varepsilon}^{1 \rightarrow 2}}{\tau_{\varepsilon}^{2 \rightarrow 1}} \right) = 2(\Phi(\Omega_2) - \Phi(\Omega_1)). \quad (53)$$

This implies that, in the weak noise limit, both escape times diverge, but the escape time out of the attractor corresponding to the lower value of the quasi-potential diverges faster. By mass balance, at steady state the populations  $P_{1,\varepsilon}$ ,  $P_{2,\varepsilon}$  of the neighbourhood of the two attractors obey the following relation:

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \log \left( \frac{P_{1,\varepsilon}}{P_{2,\varepsilon}} \right) = \lim_{\varepsilon \rightarrow 0} \varepsilon \log \left( \frac{P_{1,\varepsilon}}{1 - P_{1,\varepsilon}} \right) = 2(\Phi(\Omega_2) - \Phi(\Omega_1)) \quad (54)$$

Equation (54) implies that in the zero-noise limit only one of the two deterministic attractors will be populated, and specifically the one where the quasi-potential has lower value.

Equation (54) can be obtained by integrating the invariant measure given in (49) in the neighbourhood of the attractors and taking a saddle point approximation. Hence, the statement above holds true for an arbitrary number of competing multistable states: in the zero-noise limit, as a result of the large deviation law, the measure will be concentrated only on the attractor corresponding to the lowest value of the quasi-potential. Note that, nonetheless, individual stochastic trajectories could be trapped for a very long time in secondary metastable states corresponding to the other attractors because low levels of noise could make it very time-consuming to reach the global minimum. We also note that two different noise laws differing for the correlation matrix  $C$  acting on top of the same drift field will define two different quasi-potentials, see (50). As a result of that, they will in general feature a different selection of the limit measure in the zero noise limit.

Panels (a)–(c) of Fig. 13 provide an illustrative example of the limit behaviour of the measure for a multistable system undergoing stochastic forcing. The data are taken from the output of a simplified stochastic climate model having  $\mathcal{O}(10^4)$  degrees of freedom constructed by coupling an atmosphere described by primitive equations with an energy balance model—indeed, adapted from [20,102]—that describes in a parsimonious way the large scale heat transport performed by the global ocean [176, 177]. The deterministic version of this model features multistability of values of the ratio between the solar irradiance and the present-day one ( $\mu$  in Fig. 13) ranging from about 0.97 to about 1.06 [175]. Within this range, the system is bistable, except between 1.04 and 1.05, where a third competing state is present. For sake of simplicity—see [175–177] for further details—we focus on the W and SB states (indicated by the red and blue solid lines), which are separated by the unstable M state indicated by the green solid line, similarly to what shown in Fig. 11. The shading in panels (a)–(c) of Fig. 13 indicates for each value of  $\mu$  the density of the 1D invariant measure projected on the globally averaged surface temperature. Going from (c) to (a), the intensity of the noise—here introduced as a yearly fluctuating value of the solar irradiance around the value indicated by  $\mu$ —decreases.

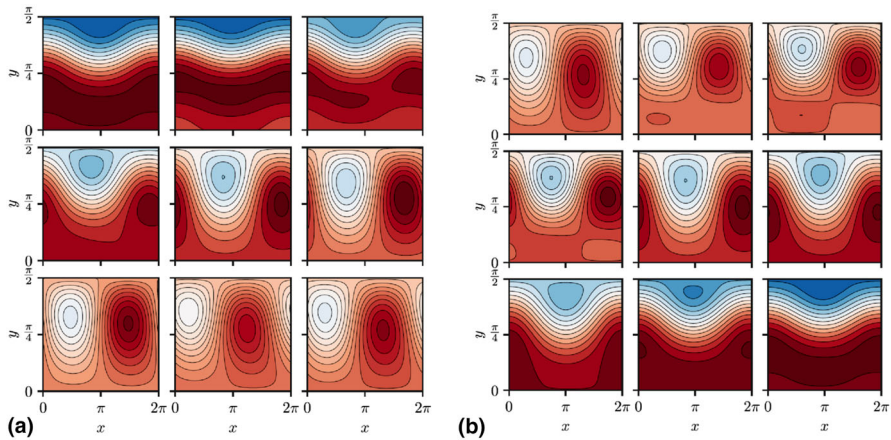
We observe that as the noise becomes weaker the measure peaks around the W (SB) attractor for  $\mu$  larger than about 1.01 (smaller than about 0.99), with a changeover around  $\mu = \mu_{crit} \approx 1.005$ . Panel (d) shows the population corresponding to the W and SB states for various levels of noise. We conclude that for  $\mu < \mu_{crit}$  the quasi-potential has a lower value at the SB attractor than at the W attractor, while the opposite occurs for  $\mu > \mu_{crit}$ . We conclude that for  $\mu = \mu_{crit}$  the system undergoes a non-equilibrium phase transition.

#### 4.6 Transitions between zonal flow and blockings

What has been presented here seems also relevant for investigating a separate, extremely relevant aspect of geophysical fluid dynamics, namely the existence in the atmosphere of different regimes of operation, which define the presence of substantial low-frequency variability on sub-seasonal time scales [104,105]. This boils down to the fact that, at coarse-grained level, due to extreme dynamical heterogeneity [180], one is practically looking at a multistable system [13,40,134,144,186,231], where one can define and detect transitions between different metastable states [26,27]; see discussion in Sect. 3.1. In particular, the fundamental dichotomy for the mid-latitude atmosphere is between the *standard* zonal flow and the blocked state [79,105,169,246], which is characterised by large spatial extent and long persistence. As discussed in Sects. 1 and 4.1, blockings are sometimes the leading cause of heatwaves and cold spells. In the coarse grained setting, the transition between the zonal and the blocked state are made possible by the stochastic forcing associated with higher frequency synoptic variability [13,186]. An accurate analysis of noise-induced transitions between the zonal and the blocked state in the celebrated Charney and DeVore [40] minimal model of the atmosphere using the LDT framework has recently been presented in [114]. There, the authors have been able to compute the optimal paths for both zonal-to-blocking and blocking-to-zonal transitions, and have elucidated (a) that the two paths are different, as we are dealing with a non-equilibrium system; and (b) that the two paths meet at the Melancholia state embedded in the boundary separating the two basins of attraction, according to the scenario discussed in Sect. 4.5. This is illustrated in Fig. 15, which portrays various snapshots of the stream function, which, in the Charney–DeVore model, is intended to approximate the 500 hPa geopotential height field.

### 5 Conclusions and perspectives

Extreme Value Theory (EVT) has shown its potential for providing information on the fundamental properties of the dynamical system generating suitably defined extreme events [178]. As an example, this feature is now being extensively used for providing a fresh outlook on the problem of understanding and characterising the predictability of the atmosphere [18,82,184], going beyond the more standard use of EVT for studying (rigorously) the tails of the distribution of meteo-climatic fields of interest [106,138]. When applying EVT to dynamical systems, universality emerges through



**Fig. 15** Transitions between the zonal flow and atmospheric blocking in the Charney–DeVore model [40]. **a** From top-left to bottom right: snapshots of the stream function along the transition from zonal flow to blocked state. **b** Same as **a**, but for the reversed transition. The central panel is the same for both paths and corresponds to the Melancholia state. Reproduced with permission from [114]

the procedure of looking at smaller and smaller portions of the attractor, which makes it possible to analyse accurately the properties of the physical measure supported on it [35,95,174,213].

Along similar lines, LDT provides powerful tools for addressing the complexity of geophysical flows and of the climate as a whole, the basic idea being that by allowing one to focus on specific, non-standard events selected according to a—possibly universal—statistical procedure, it makes it possible to elucidate fundamental properties of the system under investigation. Our viewpoint goes along the line of the scientific programme aimed at the development and interpretation of stochastic climate models proposed by Hasselmann [125], as discussed in detail in Sect. 3.1. In this review we have provided a summary of some of the key ingredients of LDT and have described the emergence and significance of large deviation laws in stochastic and deterministic chaotic systems. In a nutshell, since the procedure of computing large deviation laws relies on estimating the probability of occurrence of large anomalies in the finite size averages of stochastic variables with respect to their asymptotic values, one explores the combined effect of the static (invariant measure) and dynamic (statistics of correlations) features of the system of interest. LDT, by definition, captures the least unlikely of all the unlikely ways a given large and persistent fluctuation can take place [63]. Therefore, if we are in the correct asymptotic limit, LDT will define *typical extreme events* (but, by definition, very *atypical standard events*). This does not exclude the possibility of—quantitatively much unlikelyer—freak events or *dragon kings* [245], which live outside the approximation associated with (finite size) LDT. The key presence of such a dynamic component marks the conceptual difference between LDT- and EVT-based approaches to the study of extremes. Indeed, EVT can deal with the problem of persistence of extreme events only in a rather indirect way, through the introduction of the so-called extremal index, which measures the inverse of the characteristic cluster length of consecutive extreme events [87,178,187].



Hence, the ideas and methods proposed in this paper suggest new ways to look at the dynamics behind persistent and large fluctuations of the meteo-climatic field, in such a way that the underlying basic mechanisms behind them can be singled out and better understood. This paves the way for detailed analyses of the climate system starting both from observed data and reanalyses up to numerical models. Indeed, it seems urgent to specifically target audit and inter-comparison studies of Earth system models to such special events, in order to highlight possible deficiencies that could hardly be seen by looking at standard statistical properties. This applies to the study of individual events like heatwaves and cold spells, to the - closely related - investigation of the transitions between different modes of operation of the atmosphere responsible for its low-frequency variability, as well as to the analysis of the tipping elements of the climate system by looking at the noise-induced transitions between competing states.

It has been found that concurrent persistent extreme events, like heatwaves, droughts, or floods, are often related to specific almost stationary atmospheric wave patterns [49,208,236]. This has been shown for the summer of 2018, when several heat waves and rainfall extremes occurred almost simultaneously over different regions of the Northern Hemisphere [147]. Persistent weak phases of the stratospheric polar vortex have been related to cold extremes in Northern continental regions, especially in central Asia [150]. Interactions between persistent stationary mid-latitude wave patterns and tropical variability is poorly understood, but is suspected to cause unusual persistent extremes, like the consecutive extreme flooding, heat wave and typhoon in Japan in 2018 [271]. As explained in Sect. 1.2 when discussing the low-frequency variability of the atmosphere, local persistent anomalies of smooth observables, like temperature, are related to spatially extended large anomaly fields, and furthermore to persistent large-scale atmospheric fluctuations, i.e. quasi-stationary waves often associated with blocking events. An explanation for the phenomena of recurrent wave patterns linked to persistent extreme events from the perspective of LDT follows the principle already stated several times above representing one of the key principles of this theory: *there is one least unlikely way from all the unlikely ways that leads to a persistent extreme configuration of the atmosphere, which seems to manifest itself in the form of rather similar quasi-stationary wave patterns*. Thus, an interesting area of application would be to analyse persistent states of atmospheric circulation patterns, for example, as large deviations of jet indices, blocking indices, or of spatially averaged high-level wind fields.

An impressive example of the relevance of the principle above can be found in the studies dealing with rogue waves, which have been briefly summarised in Sect. 4.4. LDT allows to make sense of a very complex phenomenology and of competing theories by providing a (possibly) universal characterisation of rogue waves as instantons that minimise an action, in such a way that individual rogue waves do resemble the instantonic solutions, and that precursors can be identified [60,61]. This seems an excellent meeting point between theoretical results and extremely important applications in the evaluation and anticipation of natural hazards.

A natural consequence of the connection between quasi-stationary large-scale wave patterns and persistent extreme events discussed above is that several different extreme events can appear at the same time. Thus we arrive at the concept of compound

extremes, which received a lot of attention in the climate science community in the last years due to their devastating effect on both nature and human society [23,129,147,159]. Following the computation of multivariate rate functions presented in [155], large deviation rate functions could be used directly to study compound extreme events. However, the computation of multivariate rate functions would require a drastically larger amount of data than the already large amount of data needed to obtain univariate rate functions. In reference [94] it is shown that the compound nature of the persistent event can be captured also by a simple composite approach based on large deviations. We note that a high-impact persistent extreme event is usually a compound extreme event at the same time, as the probability that it triggers the occurrence of other extreme events is very high.

Summer weather persistence has been shown to increase as an effect of global warming [209,210]. As shown in [94], an increasing persistence of heatwaves in summer has been found in many regions of the Northern Hemisphere due to increasing CO<sub>2</sub> concentrations, based on a large deviation analysis. We point out that LDT provides a more natural way to define and analyse persistence than empirical approaches based on counting subsequent values above or below a certain pre-defined threshold. If the large deviations limit is reached at a certain averaging time scale, one can obtain the probability of every persistent event with duration equal to or longer than the respective averaging time, as explained in Sect. 3.3 and 4.1. Furthermore, it provides the probability of events of arbitrary intensity, within the limits set by the finite size of the available data. We further remark that LDT provides a macroscopic view on persistent events, in the sense that it focuses on deviations of the sample mean over a finite time period from the long term mean assumed to be equal to the expectation value, disregarding the amplitudes of individual instantaneous fluctuations. If one is interested to study instantaneous extremes instead of persistent ones, methods of EVT would be more appropriate to tackle the problem. We refer here also to the discussion about slow onset and fast onset events in Sect. 1.

The studies mentioned in Sects. 4.1 and 4.3 analysing persistent atmospheric extreme events, focus on persistent events of air temperature, i.e. heatwaves or cold spells. Nonetheless, by using the same large deviation based methodologies, one can study obviously in a similar way persistent events related to other geophysical observables too. Events like droughts, persistent rainfall events, floods, wind and solar lulls are just a few examples. However, the choice of the right observable is very important in order to be able to obtain a large deviation principle, as explained in Sect. 3.3. In case of some meteo-climatic observables the large deviation approaches presented in this article might not work due to long-term correlations. As an example, in Ref. [94] it is found that large deviation rate functions do not converge, at least on time scales shorter than several years, in case of air temperature anomalies over oceanic regions. However, it is possible that anomalous scaling laws [41,64,111,124,190,222], mentioned in Sect. 3.3, act in some cases in which the standard large deviation scaling fails due to the presence of some form of long-term memory. In a very recent paper [6], a nonlinear reparametrisation of the scaled cumulant generating functions is proposed to properly compute instantons in case of observables with heavy tailed distributions.

Besides serial correlations, there are also other factors that can hinder the applicability of LDT to geophysical time series. Non-stationarity is a serious issue, which



confines the application of LDT to steady state model simulations or to simulations with many ensemble members, unless one follows a pragmatic approach and pre-processes the data, for example by removing trends and the seasonal cycle, as discussed in Sect. 1.2. Strong heterogeneity of the data, i.e. the existence of several different parent distributions, can be another problem with respect to obtaining a large deviation principle. Nonetheless, LDT can be extremely helpful for several geophysical applications. Computing large deviation rate functions to obtain probabilities of persistent events could be useful for attribution studies, which try to answer the question whether and to which extent it is possible to attribute the probability of occurrence of individual extreme events to climate change, and rely at the moment mainly on empirical frequentist approaches [5,67,199,200,220]. Large ensemble climate model simulations [185,259] could be used to study the change in time of persistent event probabilities based on large deviations.

In case the data availability is poor for a proper computation of rate functions, one could rely on rare event sampling algorithms, discussed in Sect. 4.3. Rare events simulation techniques have a long history and have been applied in several fields of applied mathematics and statistical physics in the past decades. The application to problems specific to geophysical fluid dynamics and climate science are however very recent. Their potential in the field of climate modelling in particular is huge, as it allows to tackle two crucial problems emerging in the study of extreme events in the climate system, either from observations or numerical simulations.

First, extreme events are rare, which means that statistical errors are large, and robust analysis are very hard to perform, in particular if one is interested in understanding the dynamical properties of the events. This is one of the reasons for which dynamical properties are often analysed on individual case studies. This is particularly problematic when one wants to understand the impact of climate change on the statistics and dynamics of extremes (it is difficult enough to obtain estimates for a given condition, to compute variations is even worse). This however is one of the main concerns and areas of interest in the climate change debate, both at the scientific and policy level, and in the communication of the issue to the general public.

A second problem is that, given the length of the observational records (which is what it is) and given the length of the numerical experiments that is possible to perform with a given amount of computational power (which again, it is what it is), one can only study the events that have been sampled. As we discussed in this review, techniques like EVT and LDT can be used to extend the estimates of return periods or other statistical properties to events beyond what has been observed, making use of limit theorems in their corresponding areas of application. However, they do not provide the specific dynamics in the full phase space of the very rare, unobserved events, which may be relevant to study their predictability, climatic drivers and impacts. Also, the reliability of the extensions provided by EVT and LDT is dependent on the robustness of the statistics of the observed extremes. If the available statistics of extreme events is just barely enough to use them to fit the limit distributions, the uncertainties on the estimates of the properties of the unobserved events may be so large to make these estimates effectively useless. Worse, the very convergence to the limit distributions may be not properly obtained with the available samples, this may be difficult to

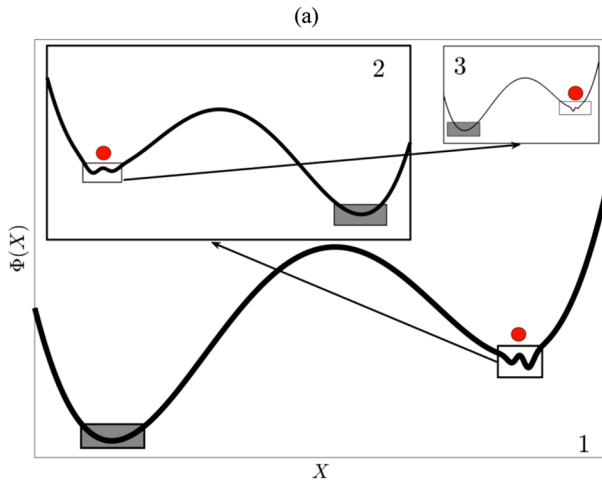
assess, and the application of these limit theories on data not properly approaching the asymptotic regime could give misleading results.

Rare event simulations techniques can provide solutions to these issues in numerical simulations, by improving drastically the statistics of extreme events, and allowing to explore ranges of events that would be simply impossible to observe in direct numerical simulations. Promising results have been obtained with genealogical algorithms for heatwaves (and in general persistent events, also not in the LDT regime) in general circulation models and data-based stochastic weather generators, and for the intensification of tropical cyclones in regional climate models. The application to more general problems will require different methods, which will be an exciting area of research in the future years. It is worth noticing that some of the methods used so far share a similar DNA with the methods used in particle filtering by the data assimilation community. Therefore, they also share similar problems and challenges when applied to complex, high dimensional systems (the *curse of dimensionality* problem). It is thus very encouraging that positive results have already been obtained with general circulation and regional climate models of state-of-the-art, nearly operational level of complexity.

The events we are able to investigate using LDT, while rare, are of disproportionately great relevance in terms of impacts on human and environmental welfare and need to be carefully taken into account when planning long-term infrastructures. Additionally, in many cases, the presence of a correspondence between long temporal persistence and large spatial coherence means that such events can pose situations of systemic risk, as their impacts can be very relevant for large regions. Hence, the results contained in this paper might be of extreme practical use for addressing climate risk, because LDT provides robust ways to estimate return times of yet unobserved events if one is in the right asymptotic regime. The possibility of establishing accurate climatologies of, e.g., heatwaves and cold spells in various regions of the globe seems particularly relevant for climate service centres, public agencies and private actors active, for example, in the energy and in agricultural sectors as well as in finance, and re-insurances. Indeed, in order to achieve full real-life applicability in the context of the presence of the seasonal cycle and of a changing climate, the results presented in this paper should be extended in such a way to incorporate the case of non-stationary time series of observables of non-autonomous dynamical systems, along the lines of what has been proposed in the context of EVT [47,84,145,189,194]. This is another great example of a very fruitful meeting point between applied and basic research.

LDT is also extremely useful for investigating some key dynamical properties of geophysical flows. As discussed in Sect. 4.2, it provides a way to frame the very important problem of evaluating the fluctuations of the predictability of the atmosphere [105,151,201] on different temporal scales by computing rate functions of various finite time Lyapunov exponents [55]. Indeed, one finds confirmation that the atmosphere is extremely *heterogeneous* in terms of its predictability [180].

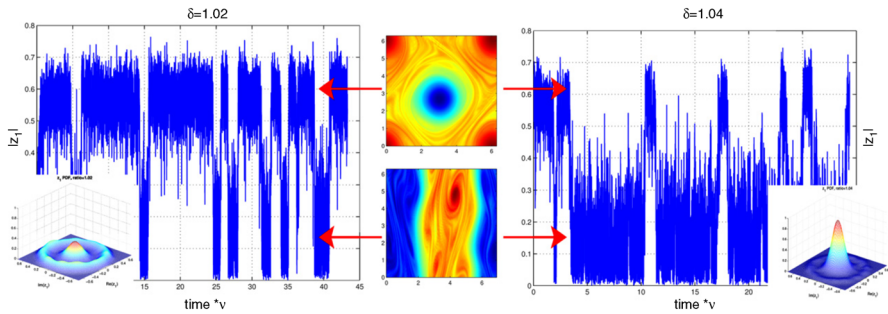
LDT is also helpful for quantifying predictability in a statistical sense and on climatic time scales. While tools like response theory allows one to study the impact of perturbations on the statistical properties of a complex system like the climate [104], such an exploration is, by definition, a local one. In order to understand the global stability properties of the climate system one needs to resort to a different view-



**Fig. 16** Cartoon showing an idealised quasi-potential where multistability occurs at multiple scales. Going from (1) to (3) one looks at smaller and smaller scales. Local maxima, minima, and saddles decorate the quasi-potential at all scales. From [182]

point. We have shown that LDT is a key tool for studying noise-induced transitions between competing metastable states for the climate system as a whole. Indeed, one can introduce a rate function, the quasi-potential, which condenses information from the deterministic vector field—the drift—and the properties of the noise. This viewpoint might prove of great usefulness for better understanding the dynamics of tipping points. This is another task of great urgency for Earth science, as the unravelling of critical transitions associated with tipping points poses great danger for human and environmental welfare. Indeed, one can represent the global stability properties of the climate system as being given by a multistable and multiscale quasi-potential—see Fig. 16. The quasi-potential features troughs, saddles, and ridges at different scales and will be characterised by a relatively few large basins (e.g. corresponding to the warm and snowball states), separated by high barriers, decorated by smaller local minima corresponding to hierarchically lower multistability features, e.g. associated to the tipping elements shown in Fig. 12a. As a result of the existence of large deviation laws describing the invariant measure and the transition probabilities between the competing metastable states, the presence of multiscale features in the quasi-potential makes it hard to perform an accurate exploration of the dynamical landscape of the system for any given choice of the noise intensity: some large scale feature of the dynamical landscape might be unattainable in any reasonable time because the noise is too weak to allow for the system to escape for a given region, whereas, conversely, other smaller scale features might be entirely washed out because the noise is too strong.

Previous investigations performed in relatively simple yet extremely relevant near-equilibrium physical systems like the stochastically perturbed two-dimensional Navier–Stokes equations have shown that LDT makes it possible to predict the various competing metastable states [30]; see Fig. 17. Along these lines, one hopes to be able to make the quasi-potential formalism a more constructive tool, rather than a descrip-



**Fig. 17** Phase transitions in the stochastic 2D Navier–Stokes equations in a doubly periodic domain: time series and probability density functions of the wavenumber 1 Fourier component in the  $y$  direction of vorticity, showing transitions between dipoles and unidirectional flows. Reproduced with permission from [30]

tive one, in order to be able to use it to predict at qualitative and quantitative level the multistability properties of the climate system and of its subsystems. At this regard, one would like to extend the very promising results [114] obtained on the transitions between zonal and blocked state in the Charney–DeVore model reported in Fig. 15 to more complex models of the atmosphere. A limitation for the numerical calculation of quantities such as minimum action paths and the quasi-potential for complex climate models is the lack of availability of derivatives of the deterministic fields determining the model evolution. Such derivatives are required for example when implementing Hamilton’s equation for the instanton, as discussed in Sect. 3.2. The development of climate models in flexible and high-performance programming languages that provide automatic differentiation, as pursued by the SciML [215] and ClimateMachine.jl [44] projects in the Julia programming language, may soon provide new ways of tackling this problem.

**Acknowledgements** VL acknowledges the support received from the EPSRC project EP/T018178/1 and from the EU Horizon 2020 project TiPES (Grant no. 820970). This review is TiPES contribution no. 79. VMG and VL acknowledge the support of the Collaborative Research Centre TRR 181 ‘Energy Transfer in the Atmosphere and Ocean’ funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, project number 274762653). The authors wish to thank F. Bouchet, D. Faranda, G. Gallavotti, M. Ghil, T. Grafke, A. Laio, B. Hoskins, G. Margazoglou, G. Messori, C. Penland, S. Vaienti, and S. Vannitsem for many scientific exchanges on the topics covered in this paper. The authors wish to thank the anonymous reviewer for their constructive comments and criticisms.

**Funding** Open access funding provided by Uppsala University.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## A Generalisations of large deviation laws: from pair empirical measures to empirical processes

In the following, we summarise some generalisations of Theorem 1 and Theorem 2 to higher dimensions, based on [63]. We consider the same situation as in case of Theorem 2 where  $X_i$  takes values in a finite set  $\Gamma = \{1, \dots, r\} \subset \mathbb{N}$ , and look at large deviations of pair empirical measures recording two successive values of  $X_1, X_2, \dots$  at each instant of time,  $L_n^2 = \frac{1}{n} \sum_{i=1}^n \delta_{(X_i, X_{i+1})}$ . The random measure  $L_n^2$  belongs now to the set of two dimensional probability measures  $M(\Gamma \times \Gamma)$ , the total variation distance is reformulated accordingly:  $d(\mu, \nu) = \frac{1}{2} \sum_{s,t} |\mu_{st} - \nu_{st}|$ . The analogue of Theorem 2 describes the large deviations of the pair empirical measure  $L_n^2$  away from the true measure  $\rho^2 = \rho \times \rho$ , whose componentwise expression is  $\rho_{st}^2 = \rho_s \rho_t$ :

**Theorem 4** *Let  $(X_i)$  be i.i.d. random variables satisfying the conditions above, and  $L_n^2 = \frac{1}{n} \sum_{i=1}^n \delta_{(X_i, X_{i+1})}$  with periodic boundary conditions. Then, the family  $(P_n^X)$  defined by  $P_n^X(\cdot) = \mathbb{P}^X(L_n^2 \in \cdot)$  satisfies the large deviation principle with rate  $n$  and with rate function*

$$I_\rho^2(\nu) = \sum_{s,t} \nu_{st} \log \left( \frac{\nu_{st}}{\bar{\nu}_s \rho_t} \right) := H(\nu | \bar{\nu} \times \rho), \quad (55)$$

with  $\bar{\nu}_s = \sum_t \nu_{st}$ .

Equation (55) tells us that the rate function  $I_\rho^2(\nu)$  of the pair empirical measure  $\nu$  is the relative entropy  $H(\nu | \bar{\nu} \times \rho)$  of  $\nu$  with respect to  $\bar{\nu} \times \rho$ .

Large deviations of pair empirical measures are especially useful if one considers Markov sequences. We stay in the finite state space setting with random variables taking values on a finite set  $X_i \in \Gamma = \{1, \dots, r\} \subset \mathbb{N}$ , but this time  $X_1, X_2, \dots$  is Markov with transition matrix  $P = (P_{st})_{s,t \in \Gamma}$ ,  $P_{st} > 0$  for all  $s, t \in \Gamma$ . We denote by  $\pi = (\pi_s)$  the unique stationary distribution of the Markov chain satisfying  $\pi_s > 0$ , for all  $s \in \Gamma$ .

**Theorem 5** *Let  $(X_i)$  be a Markov chain satisfying the conditions above, and  $L_n^2 = \frac{1}{n} \sum_{i=1}^n \delta_{(X_i, X_{i+1})}$  with periodic boundary conditions. Then, the family  $(P_n^X)$  defined by  $P_n^X(\cdot) = \mathbb{P}^X(L_n^2 \in \cdot)$  satisfies the large deviation principle with rate  $n$  and with rate function*

$$I_P^2(\nu) = \sum_{s,t} \nu_{st} \log \left( \frac{\nu_{st}}{\bar{\nu}_s P_{st}} \right) := H(\nu | \bar{\nu} \otimes P), \quad (56)$$

with  $\bar{\nu}_s = \sum_t \nu_{st}$ .

By comparing Theorem 4 with Theorem 5 becomes clear that large deviations of pair empirical measures are strongly related to those of pair dependence in Markov chains. In particular, note that if we have that  $P_{st} = \pi_t \forall s, t \in \Gamma$ , i.e. all the rows of the transition matrix are identical and are equal to the invariant measure, the Markov chain

describes a sequence of i.i.d. random variables distributed according to the measure  $\pi$ . In this case, indeed, (56) takes the form of (55). Additionally, one can deduce directly large deviation properties of Markov chains from those of i.i.d. sequences based on the Radon–Nikodym formula

$$\frac{d\mathbb{P}^X}{d\mathbb{P}^Y}[\cdot] = O(1)e^{nF(L_n^2[\cdot])}, \quad (57)$$

with

$$F(v) = \sum_{s,t} v_{st} \log \left( \frac{P_{st}}{\pi_t} \right), \quad v \in M(\Gamma \times \Gamma), \quad (58)$$

where  $X$  represents the Markov chain and  $Y$  denotes i.i.d.  $\Gamma$ -valued random variables with distribution  $\pi$ . The rate function in (56) can be obtained directly from the rate function for the pair empirical measure of  $Y$  by using  $F(v)$ :  $I_P^2(v) = I_\pi^2(v) - F(v)$ .

We have started the discussion of large deviations of Markov sequences with the pair empirical measure since pairs are innate for Markov chains. Nonetheless, the empirical measure  $L_n$  can still be relevant since it shows how often the Markov chain visits different points of the state space. Hence, it is sometimes termed as the occupation measure. The large deviation principle for the empirical measure can be derived based on the pair empirical measure by contraction, similarly to the case of i.i.d. random variables presented above in Sect. 2.1.

We assume that  $I_P^2$  is finite, continuous and strictly convex. Let  $P_n^X(\cdot) = \mathbb{P}(L_n \in \cdot)$ . Then  $(P_n^X)$  satisfies a large deviation principle on  $\mathcal{M}(\Gamma)$  with rate  $n$  and with rate function

$$I_P(\mu) = \inf_{v \in \mathcal{M}(\Gamma \times \Gamma): \bar{v} = \mu} I_P^2(v). \quad (59)$$

At the end of this section, we present another strategy to derive rate functions for Markov sequences based on the Gärtner–Ellis theorem.

One can pursue the generalisation of the above large deviation laws further in a quite straight-forward way by extending the length of the successive values of  $X_1, X_2, \dots$  to  $N$  successive values. The rate functions obtained in this finite state space setting have the very convenient properties of being finite, continuous and strictly convex. Furthermore, by letting  $N \rightarrow \infty$  one can derive the rate function of the empirical process. Due to  $N \rightarrow \infty$ , the mathematically rigorous way to formulate a large deviation principle involves the use of upper and lower limits. It is further possible to relax the condition of the finite state space to countable state space. In that case,  $\Gamma = \mathbb{N}$ , and the rate function loses the everywhere finiteness and continuity properties. It is clear that Theorem 4 is a generalisation of Theorem 2, and the later one is a generalisation of Theorem 1. Higher level large deviations laws imply the laws of lower levels, thus one can follow the link back provided by the contraction principle.

## B Quadratic approximation of rate functions of time averaged observables

Here we provide a step by step derivation of the approximate form of the rate function for Gaussian fluctuations. Let us consider the scaled cumulant generating function and auto-correlation time of a process with mean  $\mu$ , variance  $\sigma^2$  and temporal covariance  $C(t, s) = \mathbb{E}[(A(t) - \mu)(A(s) - \mu)]$

$$\lambda(k) = \lim_{T \rightarrow +\infty} \frac{1}{T} \log \mathbb{E} \left[ e^{k \int_0^T A(t) dt} \right], \quad \tau_c = \frac{1}{\sigma^2} \int_{-\infty}^{+\infty} C(\tau, 0) d\tau. \quad (60)$$

The scaled cumulant generating function can be rewritten as

$$\lambda(k) = \mu k + \lim_{T \rightarrow +\infty} \frac{1}{T} \log \mathbb{E} \left[ e^{k \int_0^T (A(t) - \mu) dt} \right]. \quad (61)$$

Expanding the exponential around  $k = 0$  and neglecting terms  $O(k^3)$  we have

$$\lambda(k) \approx \mu k + \lim_{T \rightarrow +\infty} \frac{1}{T} \log \mathbb{E} \left[ 1 + k \int_0^T (A(t) - \mu) dt + \frac{k^2}{2} \left( \int_0^T (A(t) - \mu) dt \right)^2 \right]. \quad (62)$$

Using  $\mathbb{E}[A(t)] = \mu$  and rewriting the last term as a double integral on two variables we have

$$\lambda(k) \approx \mu k + \lim_{T \rightarrow +\infty} \frac{1}{T} \log \left[ 1 + \frac{k^2}{2} \int_0^T \int_0^T \mathbb{E}[(A(t) - \mu)(A(s) - \mu)] dt ds \right] \quad (63)$$

Assuming again  $k$  small enough and expanding the logarithm we obtain

$$\lambda(k) \approx \mu k + \frac{k^2}{2} \lim_{T \rightarrow +\infty} \frac{1}{T} \int_0^T \int_0^T C(t, s) dt ds. \quad (64)$$

Since the process is stationary and the covariance is a symmetric function of  $t - s$  we have

$$\lim_{T \rightarrow +\infty} \frac{1}{T} \int_0^T \int_0^T C(t, s) dt ds = \int_{-\infty}^{+\infty} C(\tau, 0) d\tau = \sigma^2 \tau_c, \quad (65)$$

and the scaled cumulant generating function is eventually

$$\lambda(k) \approx \mu k + \frac{\sigma^2 \tau_c}{2} k^2. \quad (66)$$

Taking the Legendre transform we obtain the equivalent approximation for the rate function

$$I(a) \approx \frac{(a - \mu)^2}{2\sigma^2\tau_c}. \quad (67)$$

From the way we derived the approximation it is easy to see that  $\sigma^2\tau_ck^2/2 < 1/T$  is a necessary condition on the values of  $k$  given a value of  $T$  for the approximation to hold. The corresponding condition on the values of  $a$  can be obtained using the approximate form of  $\lambda(k)$  in the solution of the variational problem involved in the Legendre transformation,  $a = \lambda'(k)$ , that gives  $a = \mu + \sigma^2\tau_ck$ . The condition then becomes

$$\frac{|a - \mu|}{\sqrt{2\sigma^2\tau_c}} < \frac{1}{\sqrt{T}}, \quad (68)$$

consistently with the expected behaviour of Gaussian fluctuations.

## References

1. D.S. Abbot, A. Voigt, D. Koll, The Jormungand global climate state and implications for Neoproterozoic glaciations. *J. Geophys. Res. Atmos.* **116**, D18 (2011)
2. T.A.A. Adcock, P.H. Taylor, The physics of anomalous ('rogue') ocean waves. *Rep. Prog. Phys.* **77**(10), 105901 (2014)
3. R. Adler, G. Huffman, A. Chang, R. Ferraro, P. Xie, J. Janowiak, B. Rudolf, U. Schneider, S. Curtis, D. Bolvin, A. Gruber, J. Susskind, P. Arkin, The version 2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979–present). *J. Hydrometeorol.* **4**, 1147–1167 (2003)
4. N. Akhmediev, J.M. Dudley, D.R. Solli, S.K. Turitsyn, Recent progress in investigating optical rogue waves. *J. Opt.* **15**(6), 060201 (2013)
5. M. Allen, Liability for climate change. *Nature* **421**, 891–892 (2003)
6. M. Alqahtani, T. Grafke, Instantons for rare events in heavy-tailed distributions. *J. Phys. A Math. Theor.* **57**(17), 175001 (2021)
7. M.C. Alvarez-Castro, D. Faranda, P. Yiou, Atmospheric dynamics leading to west European summer hot temperatures since 1851. *Complexity* **2018**, 2494509 (2018)
8. P. Ao, Potential in stochastic differential equations: novel construction. *J. Phys. A Math. Gen.* **37**(3), L25–L30 (2004)
9. P. Ao, Global view of bionetwork dynamics: adaptive landscape. *J. Genet. Genomics* **36**(2), 63–73 (2009)
10. M. Baiesi, C. Maes, An update on the nonequilibrium linear response. *New J. Phys.* **15**(1), 013004 (2013)
11. H. Bailung, S.K. Sharma, Y. Nakamura, Observation of peregrine solitons in a multicomponent plasma with negative ions. *Phys. Rev. Lett.* **107**, 255005 (2011)
12. V. Baladi, *Positive Transfer Operators and Decay of Correlations* (World Scientific, Singapore, 2000)
13. R. Benzi, P. Malguzzi, A. Speranza, A. Sutera, The statistical properties of general atmospheric circulation: observational evidence and a minimal theory of bimodality. *Q. J. R. Meteorol. Soc.* **112**(473), 661–674 (1986)
14. R. Benzi, G. Parisi, A. Sutera, A. Vulpiani, Stochastic resonance in climatic change. *Tellus* **34**(1), 10–15 (1982)
15. R. Benzi, A. Sutera, A. Vulpiani, The mechanism of stochastic resonance. *J. Phys. A Math. Gen.* **14**(11), L453–L457 (1981)
16. J. Berner, U. Achatz, L. Batté, L. Bengtsson, Adl Cámara, H.M. Christensen, M. Colangeli, D.R.B. Coleman, D. Crommelin, S.I. Dolaptchiev, C.L.E. Franzke, P. Friederichs, P. Imkeller, H. Järvinen, S.



- Juricke, V. Kitsios, F. Lott, V. Lucarini, S. Mahajan, T.N. Palmer, C. Penland, M. Sakradzija, J.-S. von Storch, A. Weisheimer, M. Weniger, P.D. Williams, J.-I. Yano, Stochastic parameterization: toward a new view of weather and climate models. *Bull. Am. Meteorol. Soc.* **98**(3), 565–588 (2017)
17. M. Bertola, A. Tovbis, Universality for the focusing nonlinear Schrödinger equation at the gradient catastrophe point: rational breathers and poles of the tritonquée solution to painlevé i. *Commun. Pure Appl. Math.* **66**(5), 678–752 (2013)
18. T. Bódaí, Extreme value analysis in dynamical systems: two case studies, in *Nonlinear and Stochastic Climate Dynamics*, ed. by C.L.E. Franzke, T.J. O’Kane (Cambridge University Press, Cambridge, 2017), pp. 392–429
19. T. Bódaí, V. Lucarini, Rough basin boundaries in high dimension: can we classify them experimentally? *Chaos Interdiscip. J. Nonlinear Sci.* **30**(10), 103105 (2020)
20. T. Bódaí, V. Lucarini, F. Lunkeit, R. Boschi, Global instability in the Ghil–Sellers model. *Clim. Dyn.* **44**(11–12), 3361–3381 (2015)
21. N. Boers, N. Marwan, H.M.J. Barbosa, J. Kurths, A deforestation-induced tipping point for the south American monsoon system. *Sci. Rep.* **7**(1), 41489 (2017)
22. G. Boffetta, M. Cencini, M. Falcioni, A. Vulpiani, Predictability: a way to characterize complexity. *Phys. Rep.* **356**(6), 367–474 (2002)
23. R. Boschi, V. Lucarini, Water pathways for the Hindu-Kush-Himalaya and an analysis of three flood events. *Atmosphere* **10**(9), 489 (2019)
24. F. Bouchet, K. Gawedzki, C. Nardini, Perturbative calculation of quasi-potential in non-equilibrium diffusions: a mean-field example. *J. Stat. Phys.* **163**(5), 1157–1210 (2016)
25. F. Bouchet, T. Grafke, T. Tangarife, E. Vanden-Eijnden, Large deviations in fast-slow systems. *J. Stat. Phys.* **162**(4), 793–812 (2016)
26. F. Bouchet, J. Laurie, O. Zaboronski, Langevin dynamics, large deviations and instantons for the quasi-geostrophic model and two-dimensional Euler equations. *J. Stat. Phys.* **156**(6), 1066–1092 (2014)
27. F. Bouchet, C. Nardini, T. Tangarife, Stochastic averaging, large deviations and random transitions for the dynamics of 2D and geostrophic turbulent vortices. *Fluid Dyn. Res.* **46**, 061416 (2014)
28. F. Bouchet, J. Rolland, E. Simonnet, Rare event algorithm links transitions in turbulent flows with activated nucleations. *Phys. Rev. Lett.* **122**(7), 074502 (2019)
29. F. Bouchet, J. Rolland, J. Wouters, Rare event sampling methods. *Chaos* **8**, 080402 (2019)
30. F. Bouchet, A. Venaille, Statistical mechanics of two-dimensional and geophysical flows. *Phys. Rep.* **515**, 227–295 (2012)
31. R. Bowen, D. Ruelle, The ergodic theory of axiom a flows. *Invent. Math.* **29**, 181–202 (1975)
32. M. Brunetti, J. Kasparian, C. Vérard, Co-existing climate attractors in a coupled aquaplanet. *Clim. Dyn.* **53**(9), 6293–6308 (2019)
33. M.I. Budyko, The effect of solar radiation variations on the climate of the earth. *Tellus* **21**(5), 611–619 (1969)
34. T. Buehler, C.C. Raible, T.F. Stocker, The relationship of winter season north Atlantic blocking frequencies to extreme cold or dry spells in the era-40. *Tellus A* **63**(2), 212–222 (2011)
35. T. Caby, D. Faranda, G. Mantica, S. Vaienti, P. Yiou, Generalized dimensions, large deviations and the distribution of rare events. *Phys. D Nonlinear Phenom.* **400**, 132143 (2019)
36. T. Caby, D. Faranda, S. Vaienti, P. Yiou, On the computation of the extremal index for time series. *J. Stat. Phys.* **179**, 1666–1697 (2019)
37. A. Carrassi, M. Bocquet, L. Bertino, G. Evensen, Data assimilation in the geosciences: an overview of methods, issues, and perspectives. *WIREs Clim Change* **9**, e535 (2018)
38. A. Casanueva, A. Burgstall, S. Kotlarski, A. Messeri, M. Morabito, A.D. Flouris, L. Nybo, C. Spirig, C. Schwierz, Overview of existing heat-health warning systems in Europe. *Int. J. Environ. Res. Public Health* **16**, 15 (2019)
39. F. Cérou, A. Guyader, Adaptive multilevel splitting for rare events analysis. *Stoch. Anal. Appl.* **25**, 417–443 (2007)
40. J.G. Charney, J.G. DeVore, Multiple flow equilibria in the atmosphere and blocking. *J. Atmos. Sci.* **36**, 1205–1216 (1979)
41. J.-R. Chazottes, Fluctuations of observables in dynamical systems: from limit theorems to concentration inequalities, in *Nonlinear Dynamics New Directions: Theoretical Aspects*, ed. by H. González-Aguilar, E. Ugalde (Springer International Publishing, Cham, 2015), pp. 47–85

42. M. Chekroun, H. Liu, S. Wang, Approximation of stochastic invariant manifolds: stochastic manifolds for nonlinear SPDEs I. Springer Briefs in Mathematics (Springer International Publishing, 2015)
43. M.D. Chekroun, E. Simonnet, M. Ghil, Stochastic climate dynamics: random attractors and time-dependent invariant measures. *Phys. D Nonlinear Phenom.* **240**(21), 1685–1700 (2011)
44. Climate Modeling Alliance. *ClimateMachine.jl*. <https://github.com/ClimateMachine.jl> (2021)
45. V. Climenhaga, S. Luzzatto, Y. Pesin, The geometric approach for constructing Sinai–Ruelle–Bowen measures. *J. Stat. Phys.* **166**(3), 467–493 (2017)
46. J. Cohen, K. Pfeiffer, J.A. Francis, Warm arctic episodes linked with increased frequency of extreme winter weather in the United States. *Nat. Commun.* **9**(1), 869 (2018)
47. S. Coles, *An Introduction to Statistical Modeling of Extreme Values* (Springer, New York, 2001)
48. I. Corwin, P. Ghosal, A. Krajenbrink, P.L. Doussal, L.-C. Tsai, Coulomb-gas electrostatics controls large fluctuations of the KPZ equation. *Phys. Rev. Lett.* **121**, 060201 (2018)
49. D. Coumou, V. Petoukhov, S. Rahmstorf, S. Petria, H. Schellnhuber, Quasi-resonant circulation regimes and hemispheric synchronization of extreme weather in boreal summer. *PNAS* **111**(34), e993 (2014)
50. D. Coumou, A. Robinson, S. Rahmstorf, Global increase in record-breaking monthly-mean temperatures. *Clim. Change* **118**(3), 771–782 (2013)
51. H. Cramér, Sur un nouveau théorème-limite de la théorie des probabilités. *Act. Sci. Ind.* **736**, 5–23 (1938)
52. P. Cvitanović, Invariant measurement of strange sets in terms of cycles. *Phys. Rev. Lett.* **61**, 2729–2732 (1988)
53. F. Cérou, A. Guyader, M. Rousset, Adaptive multilevel splitting: historical perspective and recent results. *Chaos Interdiscip. J. Nonlinear Sci.* **29**(4), 043108 (2019)
54. F. D’Andrea, P. Drobinski, M. Stefanon, European heat waves: the effect of soil moisture, vegetation, and land use. in *Dynamics and Predictability of Large-Scale, High-Impact Weather and Climate Events* (Cambridge University Press, 2016), pp. 185–197
55. L. De Cruz, S. Schubert, J. Demaeyer, V. Lucarini, S. Vannitsem, Exploring the Lyapunov instability properties of high-dimensional atmospheric and climate models. *Nonlinear Process. Geophys.* **2018**, 387–412 (2018)
56. P. Del Moral, *Feynman–Kac formulae genealogical and interacting particle systems with applications* (Springer, New York, 2004)
57. P. Del Moral, J. Garnier et al., Genealogical particle analysis of rare events. *Ann. Appl. Probab.* **15**(4), 2496–2534 (2005)
58. A. Dell’Aquila, V. Lucarini, P.M. Ruti, S. Calmanti, Hayashi spectra of the northern hemisphere mid-latitude atmospheric variability in the NCEP–NCAR and ECMWF reanalyses. *Clim. Dyn.* **25**(6), 639–652 (2005)
59. J. Demaeyer, S. Vannitsem, Stochastic parametrization of subgrid-scale processes in coupled ocean–atmosphere systems: benefits and limitations of response theory. *Q. J. R. Meteorol. Soc.* **143**(703), 881–896 (2017)
60. G. Dematteis, T. Grafke, M. Onorato, E. Vanden-Eijnden, Experimental evidence of hydrodynamic instants: the universal route to rogue waves. *Phys. Rev. X* **9**, 041057 (2019)
61. G. Dematteis, T. Grafke, E. Vanden-Eijnden, Rogue waves and large deviations in deep sea. *Proc. Natl. Acad. Sci.* **115**(5), 855–860 (2018)
62. A. Dembo, O. Zeitouni, *Large Deviations Techniques and Applications. Applications of Mathematics* (Springer, New York, 1998)
63. F. den Hollander, *Large Deviations* (American Mathematical Society, Providence, 2000)
64. B. Derrida, Z. Shi, Slower deviations of the branching Brownian motion and of branching random walks. *J. Phys. A* **50**, 344001 (2017)
65. E. Didenkulova, Catalogue of rogue waves occurred in the world ocean from 2011 to 2018 reported by mass media sources. *Ocean Coast. Manag.* **188**, 105076 (2020)
66. I. Didenkulova, E. Pelinovsky, Rogue waves in nonlinear hyperbolic systems (shallow-water framework). *Nonlinearity* **24**(3), R1–R18 (2011)
67. R. Dole, M. Hoerling, J. Perlwitz, J. Eischeid, P. Pegion, T. Zhang, X.-W. Quan, T. Xu, D. Murray, Was there a basis for anticipating the 2010 Russian heat wave? *Geophys. Res. Lett.* **38**, 6 (2011)
68. M.D. Donsker, S.R.S. Varadhan, Asymptotic evaluation of certain Markov process expectations for large time. *I. Commun. Pure Appl. Math.* **28**, 1–47 (1975)

69. M.D. Donsker, S.R.S. Varadhan, Asymptotic evaluation of certain Markov process expectations for large time. II. Commun. Pure Appl. Math. **28**, 279–301 (1975)
70. M.D. Donsker, S.R.S. Varadhan, Asymptotic evaluation of certain Markov process expectations for large time. III. Commun. Pure Appl. Math. **29**, 389–461 (1976)
71. M.D. Donsker, S.R.S. Varadhan, Asymptotic evaluation of certain Markov process expectations for large time. IV. Commun. Pure Appl. Math. **36**, 183–212 (1983)
72. P.L. Doussal, S.N. Majumdar, G. Schehr, Large deviations for the height in 1D Kardar–Parisi–Zhang growth at late times. Europhys. Lett. **113**, 60004 (2016)
73. S. Dubinkina, H. Goosse, An assessment of climate state reconstructions obtained using particle filtering methods. Clim. Past **9**, 1141–1152 (2013)
74. E.W. Ren, W.E. Vanden-Eijnden, String method for the study of rare events. Phys. Rev. B **66**(5), 052301 (2002)
75. E.W. Ren, W.E. Vanden-Eijnden, Minimum action method for the study of rare events. Commun. Pure Appl. Math. **57**(5), 637–656 (2004)
76. D.R. Easterling, G.A. Meehl, C. Parmesan, S.A. Changnon, T.R. Karl, L.O. Mearns, Climate extremes: observations, modeling, and impacts. Science **289**(5487), 2068–2074 (2000)
77. L. Ebener, G. Margazoglou, J. Friedrich, L. Biferale, R. Grauer, Instanton base importance sampling for rare events in stochastic PDES. Chaos Interdiscip. J. Nonlinear Sci. **29**(6), 063102 (2019)
78. J.P. Eckmann, D. Ruelle, Ergodic theory of chaos and strange attractors. Rev. Mod. Phys. **57**, 617–656 (1985)
79. J. Egger, Dynamics of blocking highs. J. Atmos. Sci. **35**(10), 1788–1801 (1978)
80. R.S. Ellis, Large deviations for a general class of random vectors. Ann. Probab. **12**, 1–12 (1984)
81. V. Eyring, S. Bony, G.A. Meehl, C.A. Senior, B. Stevens, R.J. Stouffer, K.E. Taylor, Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. Geosci. Model Dev. **9**, 10539–10583 (2016)
82. D. Faranda, G. Messori, P. Yiou, Dynamical proxies of North Atlantic predictability and extremes. Sci. Rep. **7**(1), 41278 (2017)
83. M. Felici, V. Lucarini, A. Speranza, R. Vitolo, Extreme value statistics of the total energy in an intermediate-complexity model of the midlatitude atmospheric jet. Part I: stationary case. J. Atmos. Sci. **64**(7), 2137–2158 (2007)
84. M. Felici, V. Lucarini, A. Speranza, R. Vitolo, Extreme value statistics of the total energy in an intermediate-complexity model of the midlatitude atmospheric jet. part II: trend detection and assessment. J. Atmos. Sci. **64**(7), 2159–2175 (2007)
85. L. Ferranti, S. Corti, M. Janousek, Flow-dependent verification of the ECMWF ensemble over the Euro-Atlantic sector. Q. J. R. Meteorol. Soc. **141**(688), 916–924 (2015)
86. J.E. Ferrell, Bistability, bifurcations, and Waddington’s epigenetic landscape. Curr. Biol. **22**, R458 (2012)
87. C.A.T. Ferro, J. Segers, Inference for clusters of extreme values. J. R. Stat. Soc. Ser. B (Stat. Methodol.) **65**(2), 545–556 (2003)
88. H.J. Fowler, G. Lenderink, A.F. Prein, S. Westra, R.P. Allan, N. Ban, R. Barbero, P. Berg, S. Blenkinsop, H.X. Do, S. Guerreiro, J.O. Haerter, E.J. Kendon, E. Lewis, C. Schaer, A. Sharma, G. Villarini, C. Wasko, X. Zhang, Anthropogenic intensification of short-duration rainfall extremes. Nat. Rev. Earth Environ. **2**, 107–122 (2021)
89. K. Fraedrich, H. Jansen, E. Kirk, U. Luksch, F. Lunkeit, The planet simulator: towards a user friendly model. Meteorol. Z. **14**(3), 299–304 (2005)
90. K. Fraedrich, E. Kirk, F. Lunkeit, Portable university model of the atmosphere. Technical Report 16 Deutsches Klimarechenzentrum (1998)
91. C.L.E. Franzke, Extremes in dynamic-stochastic systems. Chaos Interdiscip. J. Nonlinear Sci. **27**(1), 012101 (2017)
92. C.L.E. Franzke, T.J. O’Kane, J. Berner, P.D. Williams, V. Lucarini, Stochastic climate theory and modeling. Wiley Interdiscip. Rev. Clim. Change **6**(1), 63–78 (2015)
93. M.I. Freidlin, A.D. Wentzell, *Random Perturbations of Dynamical Systems. Grundlehren Der Mathematischen Wissenschaften* (Springer, 192)
94. V.M. Gálfi, V. Lucarini, Fingerprinting heatwaves and cold spells and assessing their response to climate change using large deviation theory (2020). [arXiv:2010.08272](https://arxiv.org/abs/2010.08272)
95. V.M. Gálfi, T. Bódi, V. Lucarini, Convergence of extreme value statistics in a two-layer quasi-geostrophic atmospheric model. Complexity **2017**, 5340858 (2017)

96. V.M. Gálfi, V. Lucarini, J. Wouters, A large deviation theory-based analysis of heat waves and cold spells in a simplified model of the general circulation of the atmosphere. *J. Stat. Mech. Theory Exp.* **2019**(3), 033404 (2019)
97. G. Gallavotti, *Nonequilibrium and Irreversibility* (Springer, New York, 2014)
98. G. Gallavotti, E. Cohen, Dynamical ensembles in stationary states. *J. Stat. Phys.* **80**, 931–970 (1995)
99. L. Gammaitoni, P. Hänggi, P. Jung, F. Marchesoni, Stochastic resonance. *Rev. Mod. Phys.* **70**, 223–287 (1998)
100. J. Gärtner, On large deviations from the invariant measure. *Theory Probab. Appl.* **22**, 24–39 (1977)
101. P. Gaspard, Trace formula for noisy flows. *J. Stat. Phys.* **106**(1–2), 57–96 (2002)
102. M. Ghil, Climate stability for a sellers-type model. *J. Atmos. Sci.* **33**(1), 3–20 (1976)
103. M. Ghil, A mathematical theory of climate sensitivity or, How to deal with both anthropogenic forcing and natural variability?, in *Climate Change: Multidecadal and Beyond*, ed. by C.P. Chang, M. Ghil, M. Latif, J.M. Wallace (World Scientific Publishing Co./Imperial College Press, Singapore, 2015), pp. 31–51
104. M. Ghil, V. Lucarini, The physics of climate variability and climate change. *Rev. Mod. Phys.* **92**, 035002 (2020)
105. M. Ghil, A.W. Robertson, “Waves” vs. “particles” in the atmosphere’s phase space: a pathway to long-range forecasting? *Proc. Natl. Acad. Sci. U.S.A.* **99**(Suppl. 1), 2493–2500 (2002)
106. M. Ghil, P. Yiou, S. Hallegatte, B.D. Malamud, P. Naveau, A. Soloviev, P. Friederichs, V. Keilis-Borok, D. Kondrashov, V. Kossobokov, O. Mestre, C. Nicolis, H.W. Rust, P. Shebalin, M. Vrac, A. Witt, I. Zaliapin, Extreme events: dynamics, statistics and prediction. *Nonlinear Process. Geophys.* **18**(3), 295–350 (2011)
107. C. Giardina, J. Kurchan, V. Lecomte, J. Tailleur, Simulating rare events in dynamical processes. *J. Stat. Phys.* **145**(4), 787–811 (2011)
108. C. Giardina, J. Kurchan, L. Peliti, Direct evaluation of large-deviation functions. *Phys. Rev. Lett.* **96**, 120603 (2006)
109. F. Ginelli, P. Poggi, A. Turchi, H. Chaté, R. Livi, A. Politi, Characterizing dynamics with covariant Lyapunov vectors. *Phys. Rev. Lett.* **99**, 130601 (2007)
110. M.A. Giorgetta, J. Jungclaus, C.H. Reick, S. Legutke, J. Bader, M. Böttinger, V. Brovkin, T. Crueger, M. Esch, K. Fieg, K. Glushak, V. Gayler, H. Haak, H.-D. Hollweg, T. Ilyina, S. Kinne, L. Kornbluh, D. Matei, T. Mauritsen, U. Mikolajewicz, W. Mueller, D. Notz, F. Pithan, T. Raddatz, S. Rast, R. Redler, E. Roeckner, H. Schmidt, R. Schnur, J. Segschneider, K.D. Six, M. Stockhause, C. Timmreck, J. Wegner, H. Widmann, K.-H. Wieners, M. Claussen, J. Marotzke, B. Stevens, Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the coupled model intercomparison project phase 5. *J. Adv. Model. Earth Syst.* **5**(3), 572–597 (2013)
111. G. Gradenigo, A. Sarracino, A. Puglisi, H. Touchette, Fluctuation relations without uniform large deviations. *J. Phys. A Math. Theor.* **46**, 335002 (2013)
112. T. Grafke, String method for generalized gradient flows: computation of rare events in reversible stochastic processes. *J. Stat. Mech. Theory Exp.* **2019**(4), 043206 (2019)
113. T. Grafke, R. Grauer, T. Schäfer, The instanton method and its numerical implementation in fluid mechanics. *J. Phys. A Math. Theor.* **48**(33), 333001 (2015)
114. T. Grafke, E. Vanden-Eijnden, Numerical computation of rare events via large deviation theory. *Chaos Interdiscip. J. Nonlinear Sci.* **29**(6), 063118 (2019)
115. R. Graham, Macroscopic potentials, bifurcations and noise in dissipative systems, in *Fluctuations and Stochastic Phenomena in Condensed Matter*, ed. by L. Garrido (Springer, Berlin, 1987), pp. 1–34
116. R. Graham, A. Hamm, T. Tél, Nonequilibrium potentials for dynamical systems with fractal attractors or repellers. *Phys. Rev. Lett.* **66**, 3089–3092 (1991)
117. R. Graham, T. Tél, Nonequilibrium potential for coexisting attractors. *Phys. Rev. A* **33**, 1322–1337 (1986)
118. P. Grassberger, Noise-induced escape from attractors. *J. Phys. A Math. Gen.* **22**(16), 3283–3290 (1989)
119. C. Grebogi, E. Ott, J.A. Yorke, Fractal basin boundaries, long-lived chaotic transients, and unstable-unstable pair bifurcation. *Phys. Rev. Lett.* **50**, 935–938 (1983)
120. M.S. Gutiérrez, V. Lucarini, M.D. Chekroun, M. Ghil, Reduced-order models for coupled dynamical systems: Koopman operator and data-driven methods. *Chaos* **31**, 053116 (2021). <https://doi.org/10.1063/5.0039496>

121. A. Hamm, T. Tél, R. Graham, Noise-induced attractor explosions near tangent bifurcations. *Phys. Lett. A* **185**(3), 313–320 (1994)
122. P. Hänggi, Escape from a metastable state. *J. Stat. Phys.* **42**(1–2), 105–148 (1986)
123. I. Harris, T. Osborn, P. Jones, D. Lister, Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Sci. Data* **7**, 109 (2020)
124. R.J. Harris, H. Touchette, Current fluctuations in stochastic systems with long-range memory. *J. Phys. A Math. Theor.* **42**, 342001 (2009)
125. K. Hasselmann, Stochastic climate models, part I. Theory. *Tellus* **28**(6), 473–485 (1976)
126. K. Hasselmann, D. Olbers, Measurements of wind-wave growth and swell decay during the joint north sea wave project (JONSWAP). *Ergänzung zur Deut. Hydrogr. Z. Reihe A* (8) **12**, 1–95 (1973)
127. I. Held, M. Suarez, A proposal for the intercomparison of the dynamical cores of atmospheric general circulation models. *Bull. Am. Meteorol. Soc.* **75**, 1825–1830 (1994)
128. P.F. Hoffman, A.J. Kaufman, G.P. Halverson, D.P. Schrag, A Neoproterozoic snowball earth. *Science* **281**(5381), 1342–1346 (1998)
129. C.-C. Hong, H.-H. Hsu, N.-H. Lin, H. Chiu, Roles of European blocking and tropical–extratropical interaction in the 2010 Pakistan flooding. *Geophys. Res. Lett.* **38**, 13 (2011)
130. S. Huang, The molecular and mathematical basis of Waddington’s epigenetic landscape: a framework for post-Darwinian biology? *BioEssays* **34**, 149 (2012)
131. P. Imkeller, A.H. Monahan, Conceptual stochastic climate models. *Stoch. Dyn.* **02**(03), 311–326 (2002)
132. P. Imkeller, J.S. von Storch, *Stochastic Climate Models* (Birkhauser, Basel, 2001)
133. IPCC, in *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change*, ed. by C.B. Field, V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, P.M. Midgley (Cambridge University Press, 2012)
134. H. Itoh, M. Kimoto, Multiple attractors and chaotic itinerancy in a quasigeostrophic model with realistic topography: implications for weather regimes and low-frequency variability. *J. Atmos. Sci.* **53**(15), 2217–2231 (1996)
135. A. Jézéquel, V. Dépoues, H. Guillemot, M. Trolliet, J.-P. Vanderlinden, P. Yiou, Behind the veil of extreme event attribution. *Clim. Change* **149**(3), 367–383 (2018)
136. H. Kahn, T. Harris, Estimation of particle transmission by random sampling. *Natl. Bur. Stand. Appl. Math. Ser.* **12**, 27–30 (1951)
137. E. Kalnay, *Atmospheric Modeling, Data Assimilation and Predictability* (Cambridge University Press, Cambridge, 2003)
138. R. Katz, G. Brush, M. Parlange, Statistics of extremes: modeling ecological disturbances. *Ecology* **86**(5), 1124–1134 (2005)
139. R.W. Katz, M.B. Parlange, P. Naveau, Statistics of extremes in hydrology. *Adv. Water Resour.* **25**(8), 1287–1304 (2002)
140. R. Kautz, Activation energy for thermally induced escape from a basin of attraction. *Phys. Lett. A* **125**(6), 315–319 (1987)
141. Y. Kifer, Large deviations in dynamical systems and stochastic processes. *Trans. Am. Math. Soc.* **321**(2), 505–524 (1990)
142. R. Klein, Scale-dependent models for atmospheric flows. *Annu. Rev. Fluid Mech.* **42**(1), 249–274 (2010)
143. D. Kondrashov, M.D. Chekroun, M. Ghil, Data-driven non-Markovian closure models. *Phys. D* **297**, 33–55 (2015)
144. D. Kondrashov, K. Ide, M. Ghil, Weather regimes and preferred transition paths in a three-level quasigeostrophic model. *J. Atmos. Sci.* **61**(5), 568–587 (2004)
145. E. Konzen, C. Neves, P. Jonathan, Modeling nonstationary extremes of storm severity: Comparing parametric and semiparametric inference. *Environmetrics*, e2667 (2021)
146. C. Koppe, R. Sari Kovats, B. Menne, G. Jendritzky, World Health Organization. Regional Office for Europe and London School of Hygiene and Tropical Medicine and European Commission. Energy, Environment and Sustainable Development and Deutscher Wetterdienst. Heat-waves: risks and responses. Health and global environmental change, No. 2, WHO Regional Office for Europe, Copenhagen (2004)

147. K. Kornhuber, S. Osprey, D. Coumou, S. Petri, V. Petoukhov, S. Rahmstorf, L. Gray, Extreme weather events in early summer 2018 connected by a recurrent hemispheric wave-7 pattern. *Environ. Res. Lett.* **14**(5), 054002 (2019)
148. H.A. Kramers, Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **7**, 284 (1940)
149. P.L. Krapivsky, K. Mallick, T. Sadhu, Large deviations in single-file diffusion. *Phys. Rev. Lett.* **113**, 078101 (2014)
150. M. Kretschmer, D. Coumou, L. Agel, M. Barlow, E. Tziperman, J. Cohen, More-persistent weak stratospheric polar vortex states linked to cold extremes. *Bull. Am. Meteorol. Soc.* **99**(1), 49–60 (2018)
151. V. Krishnamurthy, Predictability of weather and climate. *Earth Space Sci.* **6**(7), 1043–1056 (2019)
152. S. Kullback, R.A. Leibler, On information and sufficiency. *Ann. Math. Stat.* **22**(1), 79–86 (1951)
153. P. Kuptsov, U. Parlitz, Theory and computation of covariant Lyapunov vectors. *J. Nonlinear Sci.* **22**, 727–762 (2012)
154. J. Kurchan, Fluctuation theorem for stochastic dynamics. *J. Phys. A Math. Gen.* **31**(16), 3719–3729 (1998)
155. F. Kwasniok, Fluctuations of finite-time Lyapunov exponents in an intermediate-complexity atmospheric model: a multivariate and large-deviation perspective. *Nonlinear Process. Geophys.* **26**(3), 195–209 (2019)
156. T. Laffargue, K.-D.N.T. Lam, J. Kurchan, J. Tailleur, Large deviations of Lyapunov exponents. *J. Phys. A Math. Theor.* **46**(25), 254002 (2013)
157. Y.-C. Lai, T. Tél, *Transient Chaos: Complex Dynamics on Finite Time Scales*, vol. 173 (Springer, New York, 2011)
158. A. Langousis, A.A. Carsteanu, R. Deidda, A simple approximation to multifractal rainfall maxima using a generalized extreme value distribution model. *Stoch. Environ. Res. Risk Assess.* **27**(6), 1525–1531 (2013)
159. W.K.M. Lau, K.-M. Kim, The 2010 Pakistan flood and Russian heat wave: teleconnection of hydrometeorological extremes. *J. Hydrometeorol.* **13**(1), 392–403 (2012)
160. J. Laurie, F. Bouchet, Computation of rare transitions in the barotropic quasi-geostrophic equations. *New J. Phys.* **17**, 015009 (2015)
161. M.R. Leadbetter, H. Rootzén, On extreme values in stationary random fields. in *Stochastic Processes and Related Topics, Trends Math.* (Birkhäuser Boston, Boston, 1998) pp. 275–285
162. V. Lecomte, J. Tailleur, A numerical approach to large deviations in continuous time. *J. Stat. Mech. Theory Exp.* **2007**, P03004 (2007)
163. T.M. Lenton, H. Held, E. Kriegler, J.W. Hall, W. Lucht, S. Rahmstorf, H.J. Schellnhuber, Tipping elements in the earth’s climate system. *Proc. Natl. Acad. Sci.* **105**(6), 1786–1793 (2008)
164. T. Lestang, F. Bouchet, E. Lévêque, Numerical study of extreme mechanical force exerted by a turbulent flow on a bluff body by direct and rare-event sampling techniques. *J. Fluid Mech.* **895** (2020)
165. T. Lestang, F. Ragone, C.-E. Bréhier, C. Herbert, F. Bouchet, Computing return times or return periods with rare event algorithms. *J. Stat. Mech. Theory Exp.* **2018**(4), 043213 (2018)
166. A. Levermann, J. Schewe, V. Petoukhov, H. Held, Basic mechanism for abrupt monsoon transitions. *Proc. Natl. Acad. Sci.* **106**(49), 20572–20577 (2009)
167. J.P. Lewis, A.J. Weaver, M. Eby, Snowball versus slushball earth: dynamic versus nondynamic sea ice? *J. Geophys. Res. Oceans* **112**, C11 (2007)
168. G. Lindgren, Local maxima of Gaussian fields. *Arkiv för Matematik* **10**(1), 195–218 (1972)
169. R.S. Lindzen, Stationary planetary waves, blocking, and interannual variability. in *Advances in Geophysics*, vol. 29 (Elsevier, 1986) pp. 251–273
170. R. Livi, P. Politi, *Nonequilibrium Statistical Physics: A Modern Perspective* (Cambridge University Press, Cambridge, 2017)
171. E.N. Lorenz, Deterministic nonperiodic flow. *J. Atmos. Sci.* **20**, 130 (1963)
172. E.N. Lorenz, *The Nature and Theory of the General Circulation of the Atmosphere* (World Meteorological Organization, Geneva, 1967)
173. O. Loidor, W. Perkins, Large deviations for the empirical distribution in the branching random walk. *Electron. J. Probab.* **20**, 18 (2015)
174. V. Lucarini, R. Blender, C. Herbert, F. Ragone, S. Pascale, J. Wouters, Mathematical and physical ideas for climate science. *Rev. Geophys.* **52**(4), 809–859 (2014)



175. V. Lucarini, T. Bódai, Edge states in the climate system: exploring global instabilities and critical transitions. *Nonlinearity* **30**(7), R32 (2017)
176. V. Lucarini, T. Bódai, Transitions across melancholia states in a climate model: reconciling the deterministic and stochastic points of view. *Phys. Rev. Lett.* **122**, 158701 (2019)
177. V. Lucarini, T. Bódai, Global stability properties of the climate: melancholia states, invariant measures, and phase transitions. *Nonlinearity* **33**(9), R59–R92 (2020)
178. V. Lucarini, D. Faranda, A.C.G.M.M. de Freitas, J.M.M. de Freitas, M. Holland, T. Kuna, M. Nicol, M. Todd, S. Vaienti, *Extremes and Recurrence in Dynamical Systems* (Wiley, New York, 2016)
179. V. Lucarini, K. Fraedrich, F. Lunkeit, Thermodynamic analysis of snowball earth hysteresis experiment: efficiency, entropy production and irreversibility. *Q. J. R. Meteorol. Soc.* **136**(646), 2–11 (2010)
180. V. Lucarini, A. Gritsun, A new mathematical framework for atmospheric blocking events. *Clim. Dyn.* **54**(1), 575–598 (2020)
181. A.J. Majda, I. Timofeyev, E. Vanden Eijnden, A mathematical framework for stochastic climate models. *Commun. Pure Appl. Math.* **54**(8), 891–974 (2001)
182. G. Margazoglou, T. Grafke, A. Laio, V. Lucarini, Dynamical landscape and multistability of a climate model. *Proc. R. Soc. A* 20210019 (2021). <https://doi.org/10.1098/rspa.2021.0019>
183. M. McCarthy, L. Armstrong, N. Armstrong, A new heatwave definition for the UK. *Weather* **74**(11), 382–387 (2019)
184. G. Messori, R. Caballero, F. Bouchet, D. Faranda, R. Grotjahn, N. Harnik, S. Jewson, J. Pinto, G. Rivière, T. Woollings, P. Yiou, An interdisciplinary approach to the study of extreme weather events: large-scale atmospheric controls and insights from dynamical systems theory and statistical mechanics. *Bull. Am. Meteorol. Soc.* **99**(5), ES81–ES85 (2018)
185. S. Milinski, N. Maher, D. Olonscheck, How large does a large ensemble need to be? *Earth Syst. Dyn.* **11**(4), 885–901 (2020)
186. K. Mo, M. Ghil, Statistics and dynamics of persistent anomalies. *J. Atmos. Sci.* **44**(5), 877–902 (1987)
187. N. Moloney, D. Faranda, Y. Sato, An overview of the extremal index. *Chaos Interdiscip. J. Nonlinear Sci.* **29**(2), 022101 (2019)
188. H. Mori, Transport, collective motion, and Brownian motion. *Prog. Theor. Phys.* **33**(3), 423–455 (1965)
189. J. Nakajima, T. Kuniyama, Y. Omori, S. Frühwirth-Schnatter, Generalized extreme value distribution with time-dependence using the ar and ma models in state space form. *Comput. Stat. Data Anal.* **56**(11), 3241–3259 (2012). 1st issue of the *Annals of Computational and Financial Econometrics Sixth Special Issue on Computational Econometrics*
190. D. Nickelsen, H. Touchette, Anomalous scaling of dynamical large deviations. *Phys. Rev. Lett.* **121**, 090602 (2018)
191. C. Nicolis, Solar variability and stochastic effects on climate, in *Physics of Solar Variations*, ed. by V. Domingo (Springer Netherlands, Dordrecht, 1981), pp. 473–478
192. C. Nicolis, Stochastic aspects of climatic transitions—response to a periodic forcing. *Tellus* **34**(3), 277–308 (1982)
193. I. Nikolkina, I. Didenkulova, Rogue waves in 2006–2010. *Nat. Hazards Earth Syst. Sci.* **11**(11), 2913–2924 (2011)
194. X.-F. Niu, Extreme value theory for a class of nonstationary time series with applications. *Ann. Appl. Probab.* **7**(2), 508–522 (1997)
195. W.M. Organization, W.H. Organization, *Heatwaves and Health: Guidance on Warning–System Development* (WMO, 2015)
196. V. Oseledec, A multiplicative ergodic theorem. characteristic Ljapunov exponents of dynamical systems. *Trans. Moscow Math. Soc.* **19**, 179–210 (1968)
197. E. Ott, *Chaos in Dynamical Systems* (Cambridge University Press, Cambridge, 2002)
198. H.C. Öttinger, M.A. Peletier, A. Montefusco, A framework of nonequilibrium statistical mechanics. I. Role and types of fluctuations. *J. Non-Equilib. Thermodyn.* **46**(1), 1–13 (2021)
199. F. Otto, and the members of the Climate Science Communications Group, Attribution of extreme weather events: how does climate change affect weather?. *Weather* **74**(9), 325–326 (2019)
200. F.E.L. Otto, N. Massey, G.J. van Oldenborgh, R.G. Jones, M.R. Allen, Reconciling two approaches to attribution of the 2010 Russian heat wave. *Geophys. Res. Lett.* **39**, 4 (2012)
201. T.N. Palmer, L. Zanna, Singular vectors, predictability and ensemble forecasting for weather and climate. *J. Phys. A Math. Theor.* **46**(25), 254018 (2013)

202. G. Pavliotis, A. Stuart, *Multiscale Methods. Texts in Applied Mathematics: TAM* (Springer, New York, 2008)
203. D. Pázó, J.M. López, A. Politi, Universal scaling of Lyapunov-exponent fluctuations in space-time chaos. *Phys. Rev. E* **87**, 062909 (2013)
204. J.P. Peixoto, A.H. Oort, *Physics of Climate* (AIP Press, New York, 1992)
205. C. Penland, A stochastic approach to nonlinear dynamics: a review. *Bull. Am. Meteorol. Soc.* **84**(7), ES43–ES52 (2003)
206. S.E. Perkins, A review on the scientific understanding of heatwaves—their measurement, driving mechanisms, and changes at the global scale. *Atmos. Res.* **164**, 242–267 (2015)
207. Y.B. Pesin, Characteristic Lyapunov exponents and smooth ergodic theory. *Russ. Math. Surv.* **32**(4), 55–114 (1977)
208. V. Petoukhov, S. Rahmstorf, S. Petri, H.J. Schellnhuber, Quasiresonant amplification of planetary waves and recent northern hemisphere weather extremes. *Proc. Natl. Acad. Sci. U.S.A.* **110**(4), 5336–5341 (2013)
209. P. Pfleiderer, D. Coumou, Quantification of temperature persistence over the northern hemisphere land-area. *Clim. Dyn.* **51**, 627–637 (2018)
210. P. Pfleiderer, C.-F. Schleussner, K. Kornhuber, D. Coumou, Summer weather becomes more persistent in a 2 °C world. *Nat. Clim. Change* **9**(9), 666–671 (2019)
211. R. Pierrehumbert, D. Abbot, A. Voigt, D. Koll, Climate of the neoproterozoic. *Annu. Rev. Earth Planet. Sci.* **39**(1), 417–460 (2011)
212. D.A. Plotkin, R.J. Webber, M.E. O'Neill, J. Weare, D.S. Abbot, Maximizing simulated tropical cyclone intensity with action minimization. *J. Adv. Model. Earth Syst.* **11**(4), 863–891 (2019)
213. F.M.E. Pons, G. Messori, M.C. Alvarez-Castro, D. Faranda, Sampling hyperspheres via extreme value theory: Implications for measuring attractor dimensions. *J. Stat. Phys.* **179**, 1698–1717 (2020)
214. M. Poumadère, C. Mays, S. Le Mer, R. Blong, The 2003 heat wave in France: dangerous climate change here and now. *Risk Anal.* **25**(6), 1483–1494 (2005)
215. C. Rackauckas, Y. Ma, V. Dixit, X. Guo, M. Innes, J. Revels, J. Nyberg, V. Ivaturi, A Comparison of automatic differentiation and continuous sensitivity analysis for derivatives of differential equation solutions, arXiv preprint. [arXiv:1812.01892](https://arxiv.org/abs/1812.01892) (2018)
216. V. Radović, I. Iglesias, *Extreme Weather Events: Definition, Classification and Guidelines towards Vulnerability Reduction and Adaptation Management* (Springer International Publishing, Cham, 2019), pp. 1–13
217. F. Ragone, F. Bouchet, Computation of extreme values of time averaged observables in climate models with large deviation techniques. *J. Stat. Phys.* **179**, 1637–1665 (2020)
218. F. Ragone, F. Bouchet, Rare event algorithm study of extreme warm summers and heat waves over Europe (2021). [arXiv:2009.02519](https://arxiv.org/abs/2009.02519)
219. F. Ragone, F. Bouchet, Rare event algorithm study of extreme warm summers and heatwaves over Europe. *Geophys. Res. Lett.* **48**, e2020GL091197. <https://doi.org/10.1029/2020GL091197>
220. S. Rahmstorf, D. Coumou, Increase of extreme events in a warming world. *Proc. Natl. Acad. Sci.* **108**(44), 17905–17909 (2011)
221. S. Rahmstorf, M. Crucifix, A. Ganopolski, H. Goosse, I. Kamenkovich, R. Knutti, G. Lohmann, R. Marsh, L.A. Mysak, Z. Wang, A.J. Weaver, Thermohaline circulation hysteresis: a model intercomparison. *Geophys. Res. Lett.* **32**, 23 (2005)
222. L. Rey-Bellet, L.-S. Young, Large deviations in non-uniformly hyperbolic dynamical systems. *Ergod. Theory Dyn. Syst.* **28**(2), 587–612 (2008)
223. P.J. Robinson, On the definition of a heat wave. *J. Appl. Meteorol.* **40**(4), 762–775 (2001)
224. C.M. Rohwer, F. Angeletti, H. Touchette, Convergence of large-deviation estimators. *Phys. Rev. E* **92**, 052104 (2015)
225. J. Rolland, Extremely rare collapse and build-up of turbulence in stochastic models of transitional wall flows. *Phys. Rev. E* **97**, 023109 (2018)
226. J. Rolland, F. Bouchet, E. Simonnet, Computing transition rates for the 1-D stochastic Ginzburg–Landau–Allen–Cahn equation for finite-amplitude noise with a rare event algorithm. *J. Stat. Phys.* **162**, 277–311 (2016)
227. J. Rolland, E. Simonnet, Statistical behaviour of adaptive multilevel splitting algorithms in simple models. *J. Comput. Phys.* **283**, 541–558 (2015)
228. G. Rubino, B. Tuffin (eds.), *Rare Event Simulation Using Monte Carlo Methods* (Wiley, Chichester, 2009)



229. D. Ruelle, *Chaotic Evolution and Strange Attractors* (Cambridge University Press, Cambridge, 1989)
230. D. Ruelle, Smooth dynamics and new theoretical ideas in nonequilibrium statistical mechanics. *J. Stat. Phys.* **95**(1), 393–468 (1999)
231. P.M. Ruti, V. Lucarini, A. Dell’Aquila, S. Calmanti, A. Speranza, Does the subtropical jet catalyze the midlatitude atmospheric regimes? *Geophys. Res. Lett.* **33**, 6 (2006)
232. T. Sadhu, B. Derrida, Large deviation function of a tracer position in single file diffusion. *J. Stat. Mech.* P09008 (2015)
233. B. Saltzman, *Dynamical Paleoclimatology: Generalized Theory of Global Climate Change* (Academic Press, New York, 2001)
234. I.N. Sanov, On the probability of large deviations of random variables (in Russian. English translation in *Sel. Trans. Math. Statist. Probab.*, pp 213–244, 1961). *Mat. Sbornik* **42**, 11–44 (1957)
235. P. Sasorov, B. Meerson, S. Prolhac, Large deviations of surface height in the 1+1-dimensional Kardar–Parisi–Zhang equation: exact long-time results for  $\lambda h < 0$ . *J. Stat. Mech.* 063203 (2017)
236. J. Screen, I. Simmonds, Amplified mid-latitude planetary waves favour particular regional weather extremes. *Nat. Clim. Change* **4**(22), 704–709 (2014)
237. W.D. Sellers, A global climatic model based on the energy balance of the earth-atmosphere system. *J. Appl. Meteorol.* **8**(3), 392–400 (1969)
238. S.I. Seneviratne, D. Lüthi, M. Litschi, C. Schär, Land–atmosphere coupling and climate change in Europe. *Nature* **443**(7108), 205–209 (2006)
239. S. Sippel, D. Mitchell, M.T. Black, A.J. Dittus, L. Harrington, N. Schaller, F.E. Otto, Combining large model ensembles with extreme value statistics to improve attribution statements of rare events. *Weather Clim. Extremes* **9**, 25–35 (2015). The World Climate Research Program Grand Challenge on Extremes—WCRP-ICTP Summer School on Attribution and Prediction of Extreme Events
240. J.D. Skufca, J.A. Yorke, B. Eckhardt, Edge of chaos in a parallel shear flow. *Phys. Rev. Lett.* **96**, 174101 (2006)
241. A. Slunyaev, I. Didenkulova, E. Pelinovsky, Rogue waters. *Contemp. Phys.* **52**(6), 571–590 (2011)
242. E.T. Smith, S.C. Sheridan, Where do cold air outbreaks occur, and how have they changed over time? *Geophys. Res. Lett.* **47**(13), e2020GL086983 (2020)
243. T.T. Smith, B.F. Zaitchik, J.M. Gohlke, Heat waves in the united states: definitions, patterns and trends. *Clim. Change* **118**(3–4), 811–825 (2013)
244. C. Snyder, T. Bengtsson, P. Bickel, J. Anderson, Obstacles to high-dimensional particle filtering. *Mon. Weather Rev.* **136**(12), 4629–4640 (2008)
245. D. Sornette, G. Ouillon, Dragon-kings: mechanisms, statistical methods and empirical evidence. *Eur. Phys. J. Spec. Top.* **205**(1), 1–26 (2012)
246. A. Speranza, Deterministic and statistical properties of the westerlies. *Paleogeophysics* **121**, 511–562 (1983)
247. A. Speranza, V. Lucarini, Environmental science, physical principles and applications, in *Encyclopedia of Condensed Matter Physics*, ed. by F. Bassani, G.L. Liedl, P. Wyder (Elsevier, Oxford, 2005), pp. 146–156
248. W. Steffen, K. Richardson, J. Rockström, S.E. Cornell, I. Fetzer, E.M. Bennett, R. Biggs, S.R. Carpenter, W. de Vries, C.A. de Wit, C. Folke, D. Gerten, J. Heinke, G.M. Mace, L.M. Persson, V. Ramanathan, B. Reyers, S. Sörlin, Planetary boundaries: guiding human development on a changing planet. *Science* **347**, 6223 (2015)
249. W. Steffen, J. Rockström, K. Richardson, T.M. Lenton, C. Folke, D. Liverman, C.P. Summerhayes, A.D. Barnosky, S.E. Cornell, M. Crucifix, J.F. Donges, I. Fetzer, S.J. Lade, M. Scheffer, R. Winkelmann, H.J. Schellnhuber, Trajectories of the earth system in the anthropocene. *Proc. Natl. Acad. Sci.* **115**(33), 8252–8259 (2018)
250. T.F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, P.M. Midgley et al., Climate change 2013: the physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change, vol. 1535 (2013)
251. D.L. Swain, D. Singh, D. Touma, N.S. Diffenbaugh, Attributing extreme events to climate change: a new frontier in a warming world. *One Earth* **2**(6), 522–527 (2020)
252. K.M.T. Imamura, T. Sasamoto, Large deviations of a tracer in the symmetric exclusion process. *Phys. Rev. Lett.* **118**, 160601 (2017)
253. J. Tailleur, J. Kurchan, Probing rare physical trajectories with Lyapunov weighted dynamics. *Nat. Phys.* **3**, 203–207 (2007)
254. S. Tibaldi, F. Molteni, On the operational predictability of blocking. *Tellus A* **42**(3), 343–365 (1990)

255. S. Tibaldi, F. Molteni, Atmospheric blocking in observation and models (2018)
256. A. Tikan, C. Billet, G. El, A. Tovbis, M. Bertola, T. Sylvestre, F. Gustave, S. Randoux, G. Genty, P. Suret, J.M. Dudley, Universality of the peregrine soliton in the focusing dynamics of the cubic nonlinear Schrödinger equation. *Phys. Rev. Lett.* **119**, 033901 (2017)
257. H. Touchette, The large deviation approach to statistical mechanics. *Phys. Rep.* **478**, 1–69 (2009)
258. J.C. Vallejo, M.A.F. Sanjuán, Predictability of orbits in coupled systems through finite-time Lyapunov exponents. *New J. Phys.* **15**(11), 113064 (2013)
259. K. van der Wiel, N. Wanders, F.M. Selten, M.F.P. Bierkens, Added value of large ensemble simulations for assessing extreme river discharge in a 2°C warmer world. *Geophys. Res. Lett.* **46**(4), 2093–2102 (2019)
260. P. van Leeuwen, H. Kunsch, L. Nerger, R. Potthast, S. Reich, Particle filters for high-dimensional geoscience applications: a review. *Q. J. R. Meteorol. Soc.* **145**, 2335–2365 (2019)
261. E. Vanden-Eijnden, M. Heymann, The geometric minimum action method: a least action principle on the space of curves. *Commun. Pure Appl. Math.* **61**(8), 1052–1117 (2008)
262. S. Vannitsem, V. Lucarini, Statistical and dynamical properties of covariant Lyapunov vectors in a coupled atmosphere-ocean model—multiscale effects, geometric degeneracy, and error dynamics. *J. Phys. A Math. Theor.* **49**, 224001 (2016)
263. S.R.S. Varadhan, *Large Deviations and Applications*. No. 46 in CBMS-NSF Regional Conference Series in Applied Mathematics (Society for Industrial and Applied Mathematics, Philadelphia, 1984)
264. G. Vissio, V. Lucarini, A proof of concept for scale-adaptive parametrizations: the case of the Lorenz '96 model. *Q. J. R. Meteorol. Soc.* **144**(710), 63–75 (2018)
265. R. Vitolo, P.M. Ruti, A. Dell'Aquila, M. Felici, V. Lucarini, A. Speranza, Accessing extremes of mid-latitude wave activity: methodology and application. *Tellus A* **61**(1), 35–49 (2009)
266. J. Vollmer, T.M. Schneider, B. Eckhardt, Basin boundary, edge of chaos and edge state in a two-dimensional model. *New J. Phys.* **11**(1), 013040 (2009)
267. A. Vulpiani, F. Cecconi, M. Cencini, A. Puglisi, D. Vergni (eds.), *Lecture Notes in Physics*, vol. 885 (Springer, Berlin, 2014)
268. C. Waddington, H. Kacser, *The Strategy of the Genes: A Discussion of Some Aspects of Theoretical Biology* (Allen & Unwin, Sydney, 1957)
269. K.M. Walter, S.A. Zimov, J.P. Chanton, D. Verbyla, F.S. Chapin, Methane bubbling from Siberian thaw lakes as a positive feedback to climate warming. *Nature* **443**(7107), 71–75 (2006)
270. J. Wang, Y. Han, M.L. Stein, V.R. Kotamarthi, W.K. Huang, Evaluation of dynamically downscaled extreme temperature using a spatially-aggregated generalized extreme value (GEV) model. *Clim. Dyn.* **47**(9), 2833–2849 (2016)
271. S. Wang, H. Kim, D. Coumou, J.-H. Yoon, L. Zhao, R. Gillies, Consecutive extreme flooding and heat wave in Japan: are they becoming a norm? *Atmos. Sci. Lett.* **20**, 12331–12336 (2019)
272. R.J. Webber, D.A. Plotkin, M.E. O'Neill, D.S. Abbot, J. Weare, Practical rare event sampling for extreme mesoscale weather. *Chaos Interdiscip. J. Nonlinear Sci.* **29**(5), 053109 (2019)
273. M. Wehner, P. Gleckler, J. Lee, Characterization of long period return values of extreme daily temperature and precipitation in the cmip6 models: Part 1, model evaluation. *Weather Clim. Extremes* **30**, 100283 (2020)
274. A.D. Wentzell, M.I. Freidlin, *Random Perturbations of Dynamical Systems*, 2nd edn. (Springer, New York, 1998)
275. J. Wouters, F. Bouchet, Rare event computation in deterministic chaotic systems using genealogical particle analysis. *J. Phys. A Math. Theor.* **49**(37), 374002 (2016)
276. J. Wouters, V. Lucarini, Disentangling multi-level systems: averaging, correlations and memory. *J. Stat. Mech. Theory Exp.* **2012**(03), P03003 (2012)
277. J. Wouters, V. Lucarini, Multi-level dynamical systems: connecting the Ruelle response theory and the Mori–Zwanzig approach. *J. Stat. Phys.* **151**, 5 (2013)
278. J.-I. Yano, A. Manzato, “Moisture control” of the precipitation: a probabilistic perspective. in *Precipitation Science* (2021)
279. J. Yin, P. Gentine, S. Zhou, S.C. Sullivan, R. Wang, Y. Zhang, S. Guo, Large increase in global storm runoff extremes driven by climate and anthropogenic changes. *Nat. Commun.* **9**(1), 4389 (2018)
280. P. Yiou, K. Goubanova, Z.X. Li, M. Nogaj, Weather regime dependence of extreme value statistics for summer temperature and precipitation. *Nonlinear Process. Geophys.* **15**(3), 365–378 (2008)
281. P. Yiou, A. Jézéquel, Simulation of extreme heat waves with empirical importance sampling. *Geosci. Model Dev.* **13**(2), 763–781 (2020)

282. L.-S. Young, Some large deviation results for dynamical systems. *Trans. Am. Math. Soc.* **318**(2), 525–543 (1990)
283. L.-S. Young, What are SRB measures, and which dynamical systems have them? *J. Stat. Phys.* **108**(5), 733–754 (2002)
284. R. Yuan, Y. Tang, P. Ao, SDE decomposition and a-type stochastic interpretation in nonequilibrium processes. *Front. Phys.* **12**(6), 120201 (2017)
285. O. Zeitouni, Random walks in random environments. *J. Phys. A Math. Theor.* **39**, R433 (2006)
286. J.X. Zhou, M.D.S. Aliyu, E. Aurell, S. Huang, Quasi-potential landscape in complex multi-stable systems. *J. R. Soc. Interface* **9**(77), 3539–3553 (2012)
287. P. Zhou, T. Li, Construction of the landscape for multi-stable systems: potential landscape, quasi-potential, a-type integral and beyond. *J. Chem. Phys.* **144**(9), 094109 (2016)
288. R. Zwanzig, Memory effects in irreversible thermodynamics. *Phys. Rev.* **124**(4), 983–992 (1961)