Edge states in the climate system: exploring global instabilities and critical transitions

Lucarini, V. ORCID: https://orcid.org/0000-0001-9392-1471 and Bodai, T. (2017) Edge states in the climate system: exploring global instabilities and critical transitions. Nonlinearity, 30 (7). R32-R66. ISSN 1361-6544 doi: https://doi.org/10.1088/1361-6544/aa6b11 Available at https://centaur.reading.ac.uk/70134/

Edge states in the climate system: exploring global instabilities and critical transitions

2017 Nonlinearity 30 R32

(http://iopscience.iop.org/0951-7715/30/7/R32)

OPEN ACCESS
IOP Publishing | London Mathematical Society Nonlinearity
Nonlinearity **30** (2017) R32–R66 https://doi.org/10.1088/1361-6544/aa6b11

Invited Article

Edge states in the climate system: exploring global instabilities and critical transitions

Valerio Lucarini[1,2,3] **and Tamás Bódai**[1,2]

[1] Department of Mathematics and Statistics, University of Reading, Reading, United Kingdom
[2] Centre for the Mathematics of Planet Earth, University of Reading, Reading, United Kingdom
[3] CEN, Meteorological Institute, University of Hamburg, Hamburg, Germany

E-mail: v.lucarini@reading.ac.uk

Received 12 November 2015, revised 6 March 2017
Accepted for publication 4 April 2017
Published 2 June 2017

Recommended by Professor Bruno Eckhardt

Abstract

Multistability is a ubiquitous feature in systems of geophysical relevance and provides key challenges for our ability to predict a system's response to perturbations. Near critical transitions small causes can lead to large effects and—for all practical purposes—irreversible changes in the properties of the system. As is well known, the Earth climate is multistable: present astronomical and astrophysical conditions support two stable regimes, the warm climate we live in, and a snowball climate characterized by global glaciation. We first provide an overview of methods and ideas relevant for studying the climate response to forcings and focus on the properties of critical transitions in the context of both stochastic and deterministic dynamics, and assess strengths and weaknesses of simplified approaches to the problem. Following an idea developed by Eckhardt and collaborators for the investigation of multistable turbulent fluid dynamical systems, we study the global instability giving rise to the snowball/warm multistability in the climate system by identifying the climatic edge state, a saddle embedded in the boundary between the two basins of attraction of the stable climates. The edge state attracts initial conditions belonging to such a boundary and, while being defined by the deterministic dynamics, is the gate facilitating noise-induced transitions between competing attractors. We use a simplified yet Earth-like intermediate complexity climate model constructed by

coupling a primitive equations model of the atmosphere with a simple diffusive ocean. We refer to the climatic edge states as Melancholia states and provide an extensive analysis of their features. We study their dynamics, their symmetry properties, and we follow a complex set of bifurcations. We find situations where the Melancholia state has chaotic dynamics. In these cases, we have that the basin boundary between the two basins of attraction is a strange geometric set with a nearly zero codimension, and relate this feature to the time scale separation between instabilities occurring on weather and climatic time scales. We also discover a new stable climatic state that is similar to a Melancholia state and is characterized by non-trivial symmetry properties.

Keywords: edge state, climate response, critical transitions, climate science, bifurcations, symmetry breaking, multistability
Mathematics Subject Classification numbers: 34C28, 37C70, 37D45, 37G35, 76U05, 76E20, 86A10

S Supplementary material for this article is available online

(Some figures may appear in colour only in the online journal)

1. Introductory remarks

Providing a comprehensive framework for climate dynamics is one of the grand challenges of contemporary science [1–4]. While classical approaches to the definition of climate sensitivity are based on concepts borrowed from the analysis of systems at or near equilibrium, a more general point of view is indeed needed [5, 6]. The issue has relevance both for the understanding of (a) the history of our planet, the relationship between climatic conditions and the development of terrestrial life across the geological epochs, and the pressing issue of anthropogenic climate change; and (b) for the possibility of having an encompassing view of planetary circulations, a much coveted goal in view of the extraordinary achievements of contemporary planetary science. Many planets are being discovered by the day, and the quest for finding atmospheres able to support life (as we know it) is ongoing; see the NASA exoplanet archive (http://exoplanetarchive.ipac.caltech.edu) for some useful information in this regard.

The climate is a complex system featuring variability of a vast range of spatial and temporal scales, resulting from instabilities fundamentally fuelled by the inhomogeneous absorption of solar radiation, and from the corresponding fluxes of mass, chemical species, and energy that, acting as negative feedbacks, allow for the establishment of steady state conditions. Interestingly, one can interpret the presence of organized motions of the geophysical fluids as the result of mechanical work, transforming available potential energy into kinetic energy, which is eventually dissipated by friction [7, 8]. Altogether, the climate can be seen as a thermal engine able to transform heat into mechanical energy with a given efficiency, and featuring many different irreversible processes that make it non-ideal [3, 9, 10]. An overarching theory of climate dynamics, able to encompass instabilities, re-equilibrating feedbacks, and large scale dynamical structures is still far from having been achieved, despite a lot of progress in the last decades.

Nonequilibrium systems like the climate can be essentially characterized as being in contact with at least two thermostats with different temperatures (or, e.g., chemical potentials) [11]. Nonequilibrium systems exhibit an extremely complex phenomenology: while a lot is known for (near-)equilibrium systems, our understanding of nonequilibrium systems is comparatively poor and limited, in spite of the wealth of phenomena occurring out of equilibrium.

Obviously, much of the interest on the understanding of the response of the climate system to forcings comes from the evidence of the complex interplay between climate conditions and life. Since the early 1990s the Intergovernmental Panel on Climate Change (IPCC) has started collecting and analyzing in a systematic way the scientific literature on the topic of climate dynamics, looking at the distant and near paleoclimatological conditions, but with a special focus on global warming and on the socio-economic impacts of anthropogenic climate change [12–14]. An enormous scientific effort is aimed at improving our understanding of climate change, and climate-related investigations account for some of the most extensive data collection campaigns and computational exercises across all disciplines.

The response of the climate system to perturbations can be qualitatively divided into two different yet coexisting forms. In some cases, such a response is smooth with respect to the perturbation parameters. In other cases, the smoothness is lost and abrupt changes can result from small perturbations. In some previous works, we have studied using a statistical mechanical approach the regime of smooth response [3, 4, 15, 16] and the conditions under which such smoothness is lost [17]. In other previous works, we have, instead, used thermodynamical concepts to better understand under which conditions small perturbations can lead to qualitative changes in the properties of the climate system [18–20]. In this work we wish to further advance the understanding of the global instabilities of the climate system by taking advantage of some powerful concepts and methods borrowed from the theory of dynamical systems.

In section 2 we review the two scenarios of smooth versus nonsmooth climatic response to perturbations, in order to provide a meaningful context for the results contained in this paper and for the motivations behind our work. We introduce the concept of *critical transitions* (*tipping points*) in geosciences [21, 22], and provide basic concepts and models for understanding the multistability of the Earth climate coming from the coexistence of the so-called snowball and warm climate states in a wide range of boundary conditions [23–28].

In section 3 we present an overview of some methods discussed in the literature for investigating the critical transitions. We briefly summarize the rationale behind using stochastic models for studying climate dynamics [27, 29, 30] and discuss how large deviations theory [31] can be relevant, within the Freidlin–Wentzell framework [32] for explaining the transitions of the system between coexisting regimes [33, 34]. Following the work of Eckhardt and collaborators in the context of the fluid dynamics of turbulent flows, and sticking to the paradigm of deterministic dynamics, we introduce the concept of *edge state*, the unstable saddle separating different regimes of motions and discuss how it is possible to find it using the edge tracking algorithm [35–37]. We also briefly discuss the application we have previously presented [38] on a simple geophysical model [25].

In section 4 we introduce PUMA-GS, a new simple yet potentially relevant climate model constructed by coupling a modified version of the model introduced in [25] with PUMA [39], a simple open-source primitive equations model of the atmospheric circulation. We then clarify how using a suitably adapted version of the edge tracking algorithm it is possible to identify the edge states of the system, i.e. unstable climatic states living on the boundary between the basins of attraction of the snowball and warm climate states[4].

In section 5 we study the geometrical properties of the boundary between competing basins of attraction, along the lines of [37, 40–43], the properties of the climatic edge states,

[4] We have decided to use for the edge states of the climate system the name of *Melancholia states*, as a homage to the 2011 movie *Melancholia* directed by Lars von Trier, where a masterful and accurate portrait of the challenges of prediction in the vicinity of critical transitions is given. The movie itself has provided a fundamental inspiration for this investigation. See www.melancholiathemovie.com.

investigating whether they correspond to fixed points, periodic points or chaotic saddles featuring a limited horizon of predictability, and the bifurcations between such regimes, including a process of symmetry breaking. We also show how the methodology proposed here allows for identifying a new stable climate regime that would hardly be discoverable through standard direct numerical integration.

In section 6 we summarize the main ideas presented in this paper and present our conclusions and perspectives for future work.

2. Response of the climate system to perturbations

2.1. Response theory and climate change

In (near-)equilibrium systems, we have formidable tools for relating forced and free fluctuations of a system. Kubo [44, 45] constructed a mathematically nonrigorous but physically extremely powerful theory able to predict how a system whose statistics is described by the canonical ensemble responds to external perturbations. Kubo's response theory also includes the fluctuation-dissipation theorem, which basically shows that the linear response of a system to forcings can be derived from its natural fluctuation through suitably defined linear operators. This result has been extremely influential in many areas of the physical and chemical sciences. See a comprehensive review in [46] and a recent extension to the nonlinear case in [47].

The lack of mathematical rigour of the response theory was later addressed by Ruelle [48–50], who showed that if one considers Axiom A dynamical systems, and limits oneself to considering sufficiently smooth observables, it is possible to prove the differentiability of the invariant measure of the system with respect to small parameters describing the change in the dynamics, and to provide explicit formulas for computing the change in the measure. Axiom A systems are chaotic and uniformly hyperbolic on the attractor, where the asymptotic dynamics takes place.

Moreover, modern methods of spectral theory have provided different and elegant proofs of Ruelle's results. The response theory can be developed by comparing the Perron-Frobenius transfer operator [51] of the unperturbed and of the perturbed system. The spectrum of the transfer operator describes, among other things, how an initial probability distribution converges to—in the case of mixing systems—a unique invariant measure. In this way, one can study directly the change in the invariant measure resulting from the applied perturbation [52]. Such a point of view has also allowed the extension of Ruelle's results to the study of the response for smooth observables in more general dynamical systems [53, 54] or of non-smooth observables in Axiom A systems [55].

Axiom A systems are indeed far from being typical dynamical systems, but, according to the *chaotic hypothesis* of Gallavotti and Cohen [56], they can be taken as effective models for chaotic physical systems with many degrees of freedom. Specifically, this means that when looking at macroscopic observables in *sufficiently* chaotic (to be intended in a qualitative sense) high-dimensional systems, it is extremely hard to distinguish their properties from those of an Axiom A system, including some degree of structural stability. One can interpret the chaotic hypothesis as the possibility of constructing robust physical properties for the system under investigation. Therefore, providing results for Axiom A systems can be thought of as being of rather general physical relevance.

Axiom A systems corresponding to equilibrium physical systems possess an invariant measure that is absolutely continuous with respect to Lebesgue because the phase space does not contract nor expand, as the flow is nondivergent. Indeed, in this case the fluctuation-dissipation theorem applies.

Axiom A systems corresponding to nonequilibrium conditions feature a phase space that contracts on the average, so that the asymptotic dynamics takes place on a strange attractor with dimension smaller than the dimension of the phase space. As a result of the singularity of the invariant measure, the usual form of the fluctuation-dissipation theorem does not hold. Ruelle [48–50] provides a mathematical explanation of this property, while a physical interpretation is given in e.g. [57–59]. This implies that for nonequilibrium systems there is no obvious relationship between free and forced fluctuations of the system, as already suggested by Lorenz [60].

Various proposals have been put forward to avoid the problem of the lack of smoothness of the invariant measure in nonequilibrium statistical mechanical systems. Some authors consider adding some stochastic forcing on top of the deterministic dynamics, in order to take into account the effect of unresolved scales [61], having in mind the Mori-Zwanzig theory [62, 63].

Interestingly, while on one side there have been positive examples of applications of the fluctuation-dissipation theorem for studying the climate, it is clear that the quality of the obtained response operator depends substantially on the chosen observable [64–66]. The difficulties in constructing *ab initio* the response operator using Ruelle's explicit formulas come from the extremely different mathematical properties of the contributions coming from the unstable and stable manifold [61] plus from the possible presence, in the case of high-dimensional chaos, of rather diverse time scales. Algorithms based on adjoint methods seem to partially ease these issues [67, 68]. A possible way to sort out the problem of estimating the response operator relies on projecting the perturbation flow on the covariant Lyapunov vectors, which allow for constructing a covariant base spanning the stable and unstable manifolds [69], or exploiting the reconstruction of the invariant measure obtained using unstable periodic orbits [70, 71]. In fact, through a combined use of the formalism of covariant Lyapunov vectors and of unstable periodic orbits, it has been recently possible to elucidate interesting mathematical and physical aspects behind the violation of the fluctuation-dissipation theorem in a simple atmospheric model [16].

Convincingly good results in terms of climate prediction performed using the linear response theory have instead been obtained by bypassing the problem of constructing the response operator and using, instead, the formal properties of the Green's function [3, 15, 59]. Note that the response theory is able to predict not only the response of globally averaged quantities to an increase in the CO_2 concentration, but also the spatial patterns of climate change [4].

2.2. Where the response theory fails: critical transitions of the climate system

The response theory is based on a perturbative expansion of the invariant measure, so its intrinsic limitation comes from the fact that the radius within which the expansion is valid could be extremely small. Of course, one might think of covering an extended range of parametric changes by repeating the expansion in neighbouring intervals. However, the procedure will fail in the vicinity of critical transitions, where one can find a qualitative change in the dynamics of the system resulting from a crisis of the attractor.

Bifurcations involving non-chaotic attractors have been very accurately studied, usually taking the point of view of analysing how the existence and the stability properties of invariant sets of the considered system change when one parameter is changed [72]. The resulting bifurcations can be classified according to universal classes [73], which has proved of immense relevance for studying a large variety of physical systems, and indeed successfully in the context of climate dynamics [1, 74].

We remark that Ashwin *et al* [75] have recently emphasized the need for widening the usual treatment of bifurcations in dynamical systems in order to accommodate for more general scenarios of critical transitions. In addition to the usual bifurcation scenario described above (*B-tipping*), they have studied the transitions between competing basins of attraction depending on the presence of a finite rate of change of one or more parameters of the system (*R-tipping*), or on the presence of stochastic forcing of finite amplitude (*N-tipping*). See in [76, 77] an earlier discussion of the role of time-dependent forcings in the context of another geophysically relevant critical transition, i.e. the collapse of the large scale oceanic circulation [78].

The situation is much more complex in the case of chaotic (and possibly high-dimensional) attractors, where critical transitions are associated to a large variety of mathematical mechanisms, including boundary crises, where an attractor is destroyed by contact with its basin boundary; interior crises, where the shape of the attractor undergoes a sudden change; synchronization and symmetry breaking (and restoring) processes, whereby different subsystems exhibit (or lose) special temporal and/or spatial correlations; and many others; see comprehensive treatments of these fascinating issues in, e.g. [79], and a survey of recent advances on synchronization phenomena (which have been initially studied in the simpler context of non-chaotic attractors) in [80].

For reasons that will be clear in section 3, in the rest of the paper our main (but not exclusive) focus will be on boundary crises, which result into the disappearance of an attractor. The vicinity of such critical transitions is accompanied, on the one hand, by a rough dependence of the system's properties on the parameter to be tuned, as a result of the Ruelle–Pollicott resonances, and, on the other hand, by the presence of a slow decay of correlations (*critical slowing-down*) of the system's observables, with the decorrelation times diverging when the control parameter reaches its critical value [17, 81, 82]. Looking at the Ruelle response formulas, it is clear that the presence of a slow decay of correlations leads to a lack of convergence in the estimate of the Green's function, even though the relationship between the presence of sufficiently fast decay of correlations and the validity of response theory can be clarified more easily using spectral theory [52].

The climate system is characterized by such critical transitions, which in the geophysical literature are usually dubbed as *tipping points*. Relevant examples of tipping points include the dieback of the Amazon forest, the shut-down of the thermohaline circulation of the Atlantic ocean, the methane release resulting from the melting of the permafrost, and the collapse of the atmospheric circulation regime associated to the Indian monsoon; see [21, 22] for a useful perspective. Obviously, not all tipping points are equally *critical*, both in terms of global climatic impacts, and, more technically, in the sense that the lack of smoothness in the response might be practically detectable only when looking at specific observables. As an example, one can expect that non-smoothness in the response associated to the dieback of the Amazon forest is easier to detect when looking at climatic fields in South America rather than in Siberia or Australia. Understanding the qualitative and quantitative properties of climatic tipping points is another great challenge of climate science.

2.3. Snowball and warm climate states

Probably the most impressive example of critical transition in the Earth system can be found when considering that the current astronomical and astrophysical conditions support at least two possible steady climate states, one being the one we are experiencing, and the other one characterized by much lower temperatures, almost total absence of water vapour in the

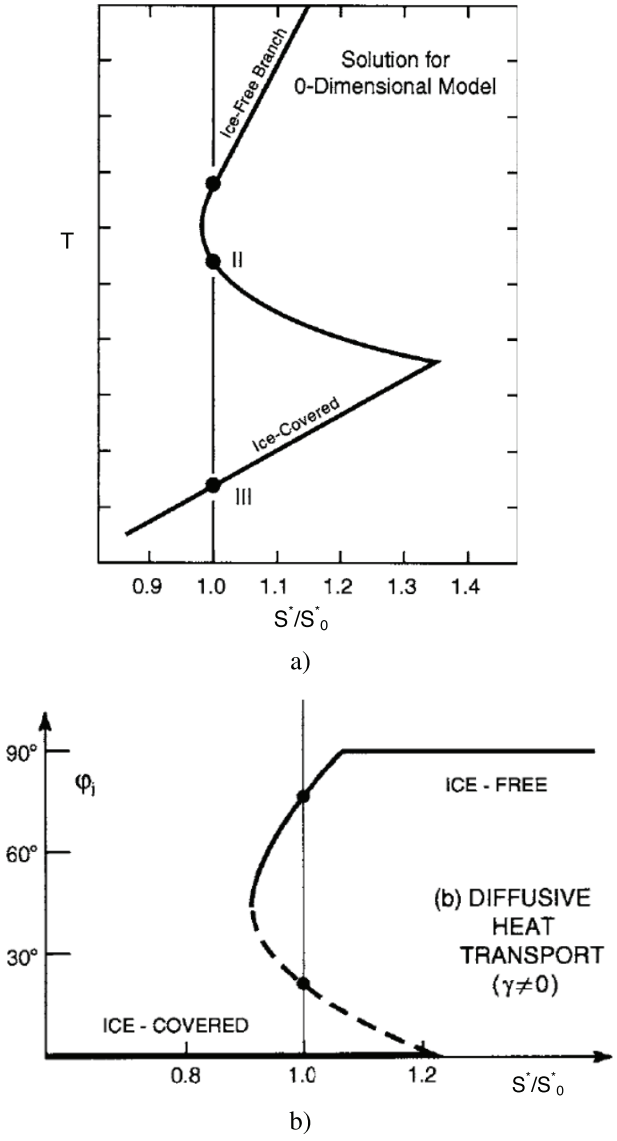

Figure 1. Multistability of the climate system corresponding to the coexistence of warm and snowball conditions. (a) Bifurcation diagram obtained from the 0D EBM as given in equation (1). The y-axis represents the temperature T, the control parameter in the abscissa is S^*/S_0^*, where S^* is the solar constant and $S_0^* \approx 1360$ Wm^{-2} is its present-day value. (b) Bifurcation diagram obtained using a 1D EBM as described in equation (2), where γ controls the strength of the diffusion operator. The y-axis shows the ice-line latitude φ_i, the control parameter in the abscissa is S^*/S_0^*. Adapted from [27], Copyright (2001), with permission from Elsevier.

atmosphere, and global ice-cover, referred to as *snowball* Earth. Models of different levels of complexity support such a statement, and, more importantly, geological evidence suggests that during the Neoproterozoic (the period spanning from about 1000 million to about 540 million years ago), the Earth suffered two of its most severe periods of glaciation [26, 83] and entered into snowball climate states.

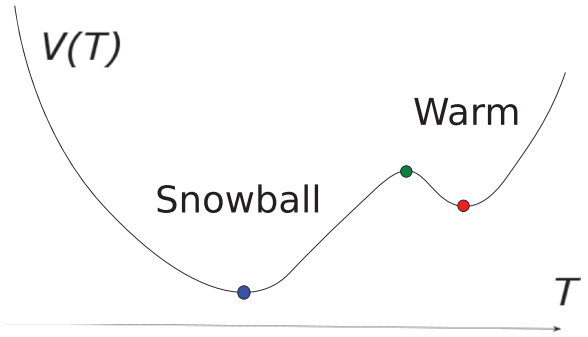

Figure 2. Scalar double-well potential function $V(T)$; the warm and the snowball states correspond to the two attractors of the system, separated by the saddle point. The Gaussian white noise triggers the transition from one basin of attraction to the other with a mean exit time described by equation (7).

The onset and decay of snowball conditions are associated to changes in the value of the solar constant and in the composition of the atmosphere, with the ensuing variation of the greenhouse effect. For a fixed atmospheric composition one finds a range of values of the solar constant S^* where the system features multistability: the warm climate can exist for all values of $S^* > S^*_{\mathrm{W}\to\mathrm{SB}}$ and the snowball state can exist for all values of $S^* < S^*_{\mathrm{SB}\to\mathrm{W}}$, with $S^*_{\mathrm{W}\to\mathrm{SB}} < S^*_{\mathrm{SB}\to\mathrm{W}}$. The critical value $S^* = S^*_{\mathrm{W}\to\mathrm{SB}}$ ($S^* = S^*_{\mathrm{SB}\to\mathrm{W}}$) is such that if we are in the warm (snowball) state and slowly reduce (increase) the value of S^*, for $S^* = S^*_{\mathrm{W}\to\mathrm{SB}}$ ($S^* = S^*_{\mathrm{SB}\to\mathrm{W}}$) the stability of the warm (snowball) climate is lost and the system tips to the snowball (warm) state.

The critical transitions associated with the boundary crisis happen in conditions where the stabilising negative feedbacks, to be described later, are overcome by the positive ice-albedo feedback. The positive feedback works as follows: colder (warmer) temperatures lead to an increase (a decrease) in the ice cover, which, by increasing (decreasing) the albedo (the fraction of incoming solar radiation elastically scattered back to space), leads in turn to reduced (enhanced) heat absorption [27]. When the ice-albedo feedback dominates, the temperature decreases (increases) substantially until a new balance sets in, corresponding to the cold (warm) stable state realised after the critical transition.

We wish to remark that some theoretical investigations and observational evidence point at the possible existence of a third possible stable climate state, which is characterized by cold conditions (yet not as extreme as in the snowball), where the equatorial belt is ice-free and characterized by vigorous oceanic and atmospheric circulations and a non-trivial hydrological cycle. Two main variants of the third climate state have been proposed, the so-called *slushball Earth* [84] and the *Jormungand* state [85]. In the rest of the paper, we will only comment briefly on this—indeed relevant—geophysical problem.

A simple mathematical formulation of the physical processes responsible for the snowball-warm multistability of the Earth system can be given by energy balance models (EBMs), which provide a simplified yet relevant representation of the energy exchanges in the climate system [27, 86]. Reducing the Earth to a 0-dimensional (0D) system (in physical space) where the only relevant quantity is the global temperature indicator T (which can be heuristically interpreted as a globally averaged effective temperature), one can construct a 0D EBM by modelling the energy budget as follows [27]:

$$C\frac{\mathrm{d}}{\mathrm{d}t}T(t) = I(1-\alpha(T)) - O(T) \to \frac{\mathrm{d}}{\mathrm{d}t}T(t) = -\frac{\mathrm{d}}{\mathrm{d}T}V(T), \tag{1}$$

where t is time, C is an effective average heat capacity per unit area, $I = S^*/4$ is the average incoming solar radiation per unit area, S^* is the solar constant (the factor 4 emerges looking at the geometry of the Earth–Sun system; see [27]), α is the albedo, which is parameterised as a non-increasing function of T, and O is the outgoing radiation per unit area. $O(T)$ is a monotonically increasing function of T, i.e. an increasing surface temperature leads to an increase in the outgoing radiation, which is the basic mechanism behind the Boltzmann (negative) radiative feedback.

By suitable (and reasonable) choices of α and O, one indeed finds bistability as a function of S^*; see the discussion in [27] and figure 1(a), where the two stable (warm and snowball) solutions are separated by the unstable solution *II*.

Note that the time derivative of the temperature in equation (1) can be written as minus the derivative of a potential $V(T)$, and in the case of bistability the two local minima of V correspond to the stable solutions, and the local maximum of V corresponds to the unstable solution; see figure 2 for a qualitative description.

EBMs can be extended in such a way as to include a latitudinal dependence of the Earth's temperature, as in the case of the celebrated models by Budyko [23], Sellers [24], and Ghil [25]. The time evolution of the temperature for one-dimensional EBMs (1D-EBMs) can be written in terms of the following partial differential equation:

$$C(\phi)\partial_t T(t,\phi) = I(\phi)(1-\alpha(\phi,T)) - O(T) - D[T,\phi], \tag{2}$$

where, as opposed to equation (1), the heat capacity C, the incoming radiation I, and the albedo α are explicitly dependent on the latitude ϕ, and we have to include in the energy budget the divergence of energy transport performed by the geophysical fluids, represented as the diffusion operator $D[T,\phi]$. It is important to note that the diffusion operator describes in a very simple yet efficient way the action of the negative feedbacks associated to the meridional temperature gradient, which acts through the meridional heat transport facilitated by atmospheric and oceanic motions. On weather time scales (few days), the feedback is powered by the baroclinic instability: the stronger the large scale meridional temperature gradient, the higher the intensity of the midlatitude atmospheric eddies and of their related counter-gradient heat transport [8, 87]. On climatic time scales (several years) the ocean acts in a similar way (though by different specific physical mechanisms), by dampening the large scale temperature differences through the transport of heat from warm to cold regions [28].

With a suitable choice for the parameters controlling the terms responsible for the energy budget in equation (2), we obtain a range of values of S^* where bistability is found; see figure 1(b), where the unstable solution sits in between the two stable solutions[5].

The existence of multistability for the Earth's climatic conditions is a robust feature across a modelling hierarchy, ranging from the simple EBMs described above to global climate models: see a detailed analysis of acting feedbacks as well as glaciation/deglaciation scenarios and mechanisms in [28]. Global climate models feature a range of bistability $[S^*_{\mathrm{W}\to\mathrm{SB}},\ S^*_{\mathrm{SB}\to\mathrm{W}}]$ that includes the present value of the solar constant S^*_0; see figure 3(a) for an example. While in simple EBMs the stable solutions correspond to fixed points in the phase space, in global climate models the attracting solutions correspond to states featuring, in general, chaotic behaviour, where the invariant measure is supported on a strange attractor [3, 18]. The temperature profiles of the stable climates obtained with the 1D-EBM given in equation (2) with a suitable choice of the value of the parameters are in broad agreement with what is obtained with complex general circulation models; see [38].

[5] It is historically remarkable to note that, after the Budyko and Sellers models were introduced, it became apparent that a *Nuclear Winter*, by reducing the incoming solar radiation as a result of increased albedo due to dramatic increases in the particulate matter in the atmosphere, could potentially trigger an even greater disaster for life on Earth than the nuclear war itself. This was indeed influential in supporting the reduction of the size of the nuclear arsenals at the end of the Cold War. It is comforting to note that, somewhat ironically, the two contributions came almost simultaneously from scientists belonging to the two opposing geopolitical blocks.

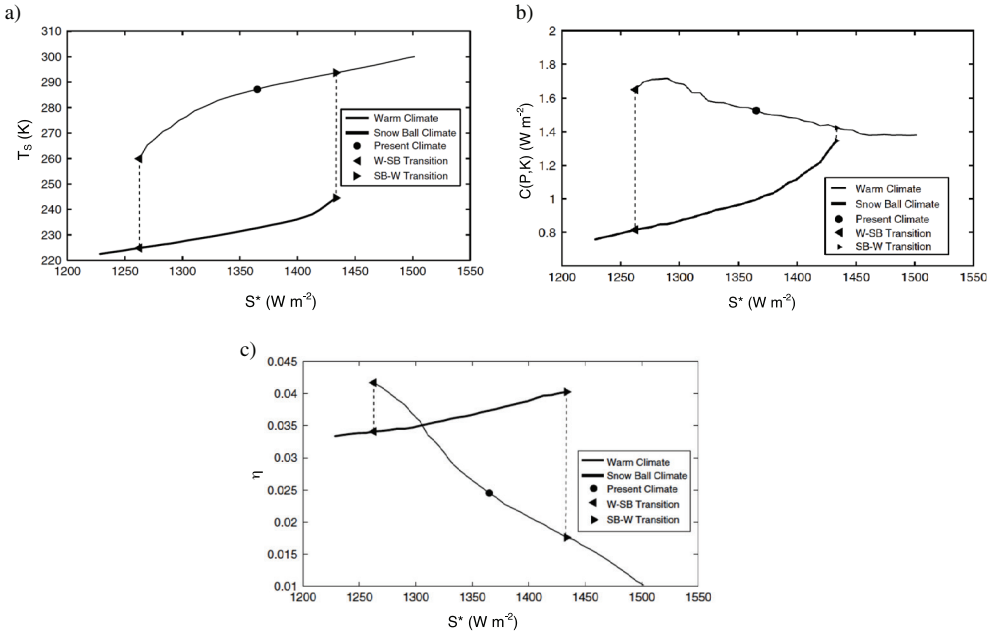

Figure 3. Multistability of the climate system corresponding to the coexistence of warm and snowball conditions. (a) Globally averaged surface temperature T_S. (b) Average intensity of the Lorenz energy cycle $C(P, K)$. (c) Efficiency η of the climate system. W-SB (SB-W) indicates the occurrence of transitions from warm to snowball conditions (from snowball to warm conditions). These results are obtained using the open source climate model PlaSim [88]. Adapted from [18] John Wiley & Sons. Copyright © 2010 Royal Meteorological Society.

For a given value of S^* within the range of bistability, the warm climate is characterized by a very active atmosphere whose variability is sustained by the presence of a large meridional temperature gradient, which supports the existence of eddies as a result of the baroclinic conversion between potential and kinetic energy, whereas the snowball climate has a much lower level of atmospheric variability, with a correspondingly low meridional temperature gradient. Figure 3(b) shows that the intensity of the Lorenz energy cycle $C(P, K)$, given by the average rate of conversion of potential into kinetic energy, which, at steady state, is also equal to the average rate of dissipation of kinetic energy due to friction [7, 8, 10], is substantially higher in the warm than in the snowball state. See [28] for a comprehensive discussion of the atmospheric circulation in the two regimes.

Thermodynamics provides additional insight to the critical transitions. One can define a Carnot-like climate *efficiency* $\eta = (\Theta^+ - \Theta^-)/\Theta^+$ in terms of the average temperatures inside the system where heating (Θ^+) and cooling (Θ^-) processes take place on average, respectively [10]. The efficiency is higher when there is stronger correlation between temperature fluctuations and heating rates. The tipping point is accompanied by a sudden decrease in the efficiency η, which can be interpreted as a result of the system getting closer to thermodynamic equilibrium; see figure 3(c) and the discussion in [3, 18–20][6].

[6] A larger (smaller) concentration of greenhouse gases leads to a shift of the range of multistability towards lower (higher) values of S^*. The presence of a gigantic concentration of CO_2 is deemed responsible for the exit of the Earth from the snowball state experienced in the Neoproterozoic [28]. Note also that the range of multistability is altered when one considers variations in the length of the day, with multistability being eventually lost and a unique climate emerging when the day and the year have similar length, and in particular in the 1:1 phase tidal locked condition [19, 20].

3. Critical transitions and edge states

3.1. Energy landscapes, large deviations, and transitions

We note that, far from the critical transitions indicated in figure 3, it is possible to jump from the warm attractor to the snowball attractor and *vice versa* by perturbing the system through suitably defined time-dependent deterministic and/or stochastic forcing. The transition from one basin of attraction to another one is most easily obtained through a Dirac's δ-like perturbation applied to the orbit (this might correspond, in physical terms, to the impact of an asteroid on our planet).

Nonetheless, arguably the most interesting scenario of forcing is given by the presence of stochastic perturbations. Climate science has indeed been one of the first areas of natural sciences where the paradigm of stochastically perturbed dynamical systems has been extensively used, following the Hasselmann programme [29]; see also later discussions in [27, 30, 89]. The basic idea is the following: if one considers separately slow and fast modes of variability of the climate system, in the limit of infinite time scale separation, the impact of the fast modes on the slow modes can be treated, to a first approximation, as a stochastic correction to the deterministic dynamics of the slow modes. This point of view is closely related to the mathematical theory of the *averaging method* for performing the coarse-graining of the dynamics of multiscale dynamical systems [90, 91]. We further comment on this point of view below in the text.

The Freidlin–Wentzell theory [32] provides a comprehensive framework in multistable systems for studying the probability of transitions outside one of the basins of attraction triggered by stochastic forcing. The Freidlin–Wentzell theory is most easily understood in the case of a system whose dynamics takes place in an energy landscape (plus additive white noise), such that:

$$\mathrm{d}y = -\nabla_y V(y)\mathrm{d}t + \epsilon \mathrm{d}W, \tag{3}$$

where $y \in \mathbb{R}^n$, $V : \mathbb{R}^n \to \mathbb{R}$ is sufficiently smooth, and $\mathrm{d}W$ is a vector of n independent increments of normalised Brownian motions (the Itô or Stratonovich conventions are equivalent, in this case), and ϵ determines the strength of the noise. As is well known, one can associate to the stochastic differential equation given by equation (3) a Fokker–Planck equation describing the evolution of the probability density function (pdf) $p(y,t)$ of an ensemble of trajectories obeying the stochastic differential equation as follows:

$$\frac{\partial p(y,t)}{\partial t} = \nabla_y \cdot (\nabla_y V(y) p(y,t)) + \frac{\epsilon^2}{2}\Delta p(y,t), \tag{4}$$

where, under suitable conditions of integrability and in the weak noise limit $\epsilon \to 0$, the stationary solution corresponding to the invariant measure is given by $\lim_{t\to\infty} p(y,t) = p(y)$, where at leading order $p(y) \propto \exp[-2V(y)/\epsilon^2]$. The local minima of the potential V (fixed-point attractors in the case of deterministic dynamics with $\epsilon = 0$) correspond to the local maxima of p, so that, e.g. a double-well potential corresponds to a bimodal pdf. In the limit of weak noise, trajectories starting near a local minimum of V typically wait a long time before moving to the neighborhood of a different local minimum of V, and the transitions take place most likely through the lowest energy saddle linking the initial basin of attraction to any other basin.

A comprehensive treatment of the critical transitions accounting for the effect of noise can be found in [92]. Let us now consider the simple case of $y \in \mathbb{R}$ and focus again on the problem of the transitions between the snowball and warm climates. If one includes stochastic forcing in the form of additive Gaussian white noise on the right hand side of equation (1), we have

$$\mathrm{d}T(t) = -\frac{\mathrm{d}}{\mathrm{d}T}V(T)\mathrm{d}t + \epsilon \mathrm{d}W \tag{5}$$

where $\mathrm{d}W$ is the increment of a Brownian motion and $\epsilon > 0$. In the limit of weak noise, the stationary distribution can be expressed at leading order as

$$p(T) \propto \exp[-2V(T)/\epsilon^2]. \tag{6}$$

Using large deviations theory [31], one can derive that, at leading order, the mean exit time for the transition from the basin of attraction of the stable solution A to the basin of attraction of the stable solution B through the unstable saddle U can be written as:

$$\tau_{A\to B} \propto \exp\left[2(V(U) - V(A))/\epsilon^2\right], \tag{7}$$

while the factor before the exponential function in equation (7) is specified by the celebrated Kramer's escape rate formula [93, 94]. See figure 2, where the states A and B can be though of as corresponding to the warm and snowball state, and the saddle U coincides with the boundary of the two basins of attraction.

In the geophysical literature tipping points are mostly studied by using equation (6) to construct an equivalent one-dimensional pseudo-potential from the pdf derived starting from the time series of a selected climate observable. Subsequently, Kramer's theory is used to estimate the probability of transition from one basin of attraction to another one within a given time frame [21, 22]. Nonetheless, inconsistencies in this method emerge from the fact that the operation of projecting the dynamics onto only one observable is very likely to cause errors in the evaluation of the time scales of the dynamics, thus breaking the simple and powerful relations existing between invariant density and escape probability from a local minimum of the effective potential; see the discussion in [95]. A different and promising point of view on the detection of tipping points based on extreme value theory has been recently proposed in [97].

When considering dynamical systems featuring multiple attractors and undergoing Gaussian stochastic forcing, large deviations theory [31] provides methods for constructing the so-called instantons, which are the most likely noise-driven trajectories leading to the transitions from one basin of attraction to the other one. An instanton is constructed as a minimizer of the related Freidlin–Wentzell action and can be interpreted as the *most efficient* way to exit a local potential minimum. See [96] for an introduction in the context of fluid dynamics.

Some geophysical fluid dynamical systems feature extremely rare transitions between different regimes of motions. These are rare excursions between regions of the attractor of the system that take place only on ultralong time scales, so that, performing an operation of coarse graining to the dynamics, one can describe them as noise-induced transitions between separate attracting sets for the deterministic part of the dynamics [33, 34].

The instanton method allows for treating more general cases than those of systems whose dynamics is governed by an energy landscape. What we find a bit unsatisfactory in this otherwise extremely powerful framework of noise-driven transitions is the fact that one is left with the question of what is the mathematical nature and the physical justification of the noise. Note that, given the lack of time-scale separation in the climate variability [8], it is a bit of a stretch to use standard arguments to justify the presence of white noise as resulting from fast atmospheric motions. In other terms, despite the huge insight of the Hasselmann program mentioned above, climate dynamics is in fact not the best setting for using the averaging method. As discussed in [98, 99] in the spirit of [62, 63] and confirmed in the mathematically more rigorous treatment in [100, 101], when no scale separation is present between the scales of motions we want to resolve and those we want to parameterise, the effect of the latter on the former entails the consideration of time-correlated noise and non-markovian effects.

As a final note we wish to remark that even in the context of systems of the form given in equation (3), if one assumes that the noise possesses slow decay of correlations, such as in the case of Lévy processes, the standard Freidlin–Wentzell picture is dramatically altered, with an entirely different dynamical scenario to be considered. While in the case of white noise the transition from one basin of attraction to the other occurs in the very unlikely case that many subsequent stochastic perturbations *conjure* to push the orbit across the saddle, in the case of Lévy noise one or few large kicks do the job [102, 103], so that the expression for the mean exit time is rather different from that reported in equation (7). This scenario seems to be relevant also in the context of climate dynamics [104–106].

3.2. Edge states

Following a classical point of view within statistical mechanics, we would like to be able to understand how noise-induced transitions between separate climatic attractors may take place by gaining insight into the underlying purely deterministic evolution laws. In this regard, we take advantage of the *edge tracking method* recently developed by Eckhardt and collaborators for studying fluid dynamical systems possessing multiple (quasi-)steady states [35, 36, 107].

The plane Couette flow or the pipe flow, in the case of sufficiently high Reynolds numbers, feature the coexistence of a (quasi-)attractor corresponding to the turbulent state and of an attracting fixed point corresponding to laminar flow. We remark that the turbulent state cannot be properly described as corresponding to a separate attractor, because turbulence is transient—even if with exceedingly long life time, which grows superexponentially fast with the Reynolds number [107–110]. But, using arguments of time-scale separation, we can say that the scenario is (almost) indistinguishable from that of two separate attractors, where the turbulent state is steady, so we will use an abuse of terminology in describing these results, and refer to the turbulent state as corresponding to a proper attractor, possessing its own basin of attraction.

The basic goal is to understand what lies in-between the two attractors in order to (a) have a global view on the properties of the phase space of the system, beyond the paradigm of looking at steady states; and (b) understanding which combinations of external forcings and internal processes might more easily lead to transitions between the steady states.

Clearly, in the case of a bistable system, there must be a separatrix, a repelling set which acts as boundary between the two basins of attraction. The natural question one might ask is: what is the evolution of an orbit starting nearby or on such a separatrix? In the case of an energy landscape, such a basin boundary is the *mountain crest* separating the two basins of attraction, and the evolution law would lead an orbit starting from a point on the boundary to the nearby local minimum *restricted* to the subset of the phase space corresponding to the basin boundary, where the evolution would stop. Clearly, the end point is in fact a saddle, because there is an (additional) unstable direction. Orbits departing near the basin boundary but not exactly on it, would end up in the attracting set corresponding to the basin they start from.

Eckhardt and collaborators showed that trajectories initialised on the boundary between the two basins of attraction converge, basically as a result of the contraction of the phase space, to a unstable saddle, which is a relative attractor with respect to the basin boundary (but obviously a repelling set when viewed globally). The scenario has been confirmed in the case of truly separate attractors in a toy model discussed in [37].

Such an unstable saddle, the so-called *edge state*, has been found to be either a fixed point, or a periodic orbit, or, in some cases, a chaotic solution taking place on a strange geometrical set. The relevance of the edge state lies in the fact that it is crucial for understanding the global

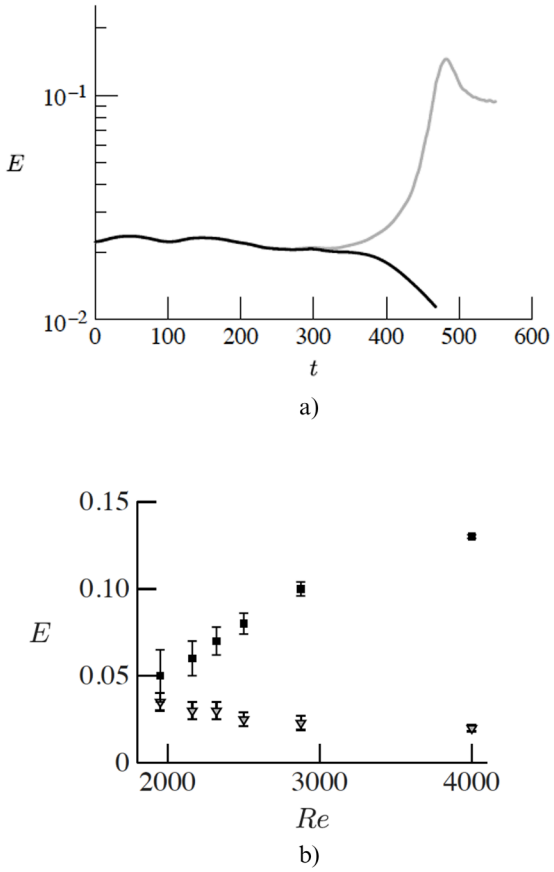

Figure 4. Edge state in the pipe flow. (a) Time evolution of the energy of the turbulent flow for two trajectories starting near the edge state but belonging to two different basins of attraction: one trajectory collapses to zero (laminar state), and another one increases its turbulent kinetic energy until the quasi-steady turbulent state is reached. (b) Turbulent kinetic energy of the steady turbulent state (square dots) and of the edge state (triangles) for various values of the Reynolds number. The turbulent energy of the other steady state, the laminar state, is zero by definition. Note that we are using an abuse of language in defining the turbulent state as corresponding to a true attractor (it is, instead, a very long transient; see text). Adapted from [36] with permission of The Royal Society of Chemistry.

properties of the system and that it provides the gate between the two attractors: trajectories resulting from (weak) Gaussian white random external forcings that are successful in achieving the critical transition have to pass nearby the edge state.

Of course, given its instability, one cannot find the edge state by direct forward numerical integration; instead, one can find it by a suitable algorithmic procedure named *edge tracking*. The basic idea is to find by bisections (and long forward integrations) two nearby points in the phase space that are on the two sides of the separatrix, follow their forward evolution for a given time (typically short), and then repeat the bisection in order to reduce the divergence between the two trajectories due to the global instability connected with the bistability of the system.

The bisection is obtained by interpolation to the midpoint (more general constructions of convex combinations are possible) between the initial conditions belonging to the two sides

of the basin boundary, checking, using a long run, whether the new initial condition leads to one or to the other attracting set, and repeating until two nearby (according to a prescribed criterion) initial conditions leading to different asymptotic dynamics are obtained. One retains short segments of the pair of control trajectories belonging to these most tightly bracketing initial conditions. The length of these segments are limited by a requirement on the maximal separation of the control trajectories. After a transient, any of the segments obtained iterating this procedure can be taken to represent the edge state, approximated by many discontinuous segments of the trajectory.

A key mathematical fact to be considered here is that the basins of attraction are invariant sets, which ensures that the procedure is well-defined. It is a crucial—in terms of robustness and efficiency—aspect of the procedure to choose a useful observable able to tell us unambiguously and soon enough whether the trajectory is ending up on one or on the other attractor; see figure 4(a). An efficient observable has a relatively small variability in the transient runs compared with the actual difference between the expectation values computed on the two separate attractors.

Additionally, it seems quite natural to investigate the geometrical properties of the boundary separating the basins of attraction, as it is not a-priori clear whether one should expect a manifold of co-dimension one or a more complex geometrical object. We will come back to this aspect in section 5.3.

In the case of the problems studied by Eckhardt and collaborators, the natural choice is to use the total turbulent kinetic energy of the flow. In the limit of an infinitely small separation between the two bracketing trajectories, one actually obtains the transient behaviour leading to the edge state, and, then, is able to reconstruct the dynamics on the edge state. While this procedure is rather intuitive, one might reasonably expect lots of technical difficulties related to the need of controlling strong instabilities in a high-dimensional dynamical system. Eckhardt and collaborators have convincingly shown the robustness of the edge tracking method [35–37].

In a previous paper [38] we have introduced in the geophysical community the edge tracking method to provide a fresh outlook on the properties of the Ghil-Sellers (GS) 1D-EBM [25], which can be written in the form of equation (2). Ghil [25] solved the appropriate boundary value problem and found for a wide range of values of the solar constant three coexisting stationary solutions, and carried out a stability analysis able to identify the stable warm and snowball climates, plus an unstable state. Such a state was identified by Ghil as the saddle point of a suitably constructed potential, and we proved that it coincides with the edge state found using the tracking procedure; see figure 5. Our investigation made it apparent that the most efficient (but definitely not unique) observable to be used for characterizing in the edge tracking process whether the trajectory ends up in the snowball or warm state is the globally averaged surface temperature, because warming and cooling are mainly controlled by the anomalies in radiation emission and changes in albedo, which are both to a good extent controlled by changes in the globally averaged temperature; see also equation (1). We also studied the physics of the three states, relating their instabilities to relevant macroscopic thermodynamical properties such as large scale temperature gradients and entropy production, using the conceptual framework introduced in [18–20].

3.3. Remark

Looking at the top right projection of the three-dimensional curve portrayed in figure 5, one sees that the edge state is characterized by the fact that lower average temperatures correspond to higher values of solar constant. One can interpret this feature as representative of the fact

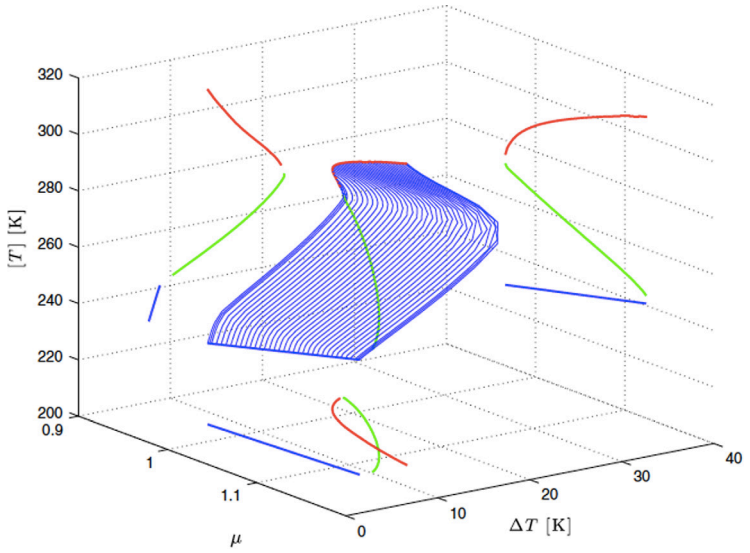

Figure 5. Bifurcation diagram of the Ghil-Sellers [25] model defining the constitutive relations between the normalised intensity of the incoming radiation ($\mu = S^*/S_0^* = 1$ corresponds to the current conditions), the globally averaged temperature [T] and the low-to-high-latitudes temperature difference ΔT. The thick red, green, and blue lines correspond to the warm, edge, and snowball states, respectively. The thin blue lines depict heteroclinic trajectories—the instantons. Adapted from [38] (2014) © Springer-Verlag Berlin Heidelberg 2014. With permission of Springer.

that the edge state has, effectively, a *negative* heat capacity, and is not thermodynamically stable. Making a *Gedankenexperiment* where the Earth system in the edge state is put in an isolated box with another body emitting and absorbing radiation and looking at the dynamics of fluctuations should clarify this point.

Interestingly, a qualitatively similar property can be found when looking at the case of turbulent flows analysed by Eckhardt and co. [36] and portrayed in figure 4(b): in the edge state, larger values of the Reynolds number, which is representative of the applied pressure gradient, correspond to lower values of the turbulent kinetic energy of the flow. This can be interpreted as corresponding to conditions of mechanical instability, where the viscosity is *negative*.

3.4. Melancholia states in a climate model

In this paper we want to show how the concept of the edge state can be used for characterizing globally the dynamical properties of a severely simplified yet Earth-like new climate model featuring multistable properties in a realistic range of values of the solar constant. As mentioned in section 1, we propose to call the climatic edge states *Melancholia states*.

Figure 6 clarifies how this work relates to previous analyses we have performed on the thermodynamics and statistical mechanics of climate. While most of the studies we consider here have been performed using the open source climate model PlaSim [88], in this specific investigation, given the complexity of the procedure of edge tracking, we resort to using the simpler model PUMA-GS, later described in detail in section 4.

In summary, the climate model PUMA-GS is constructed by coupling PUMA, an open source three-dimensional primitive equations spectral model of the atmosphere [39], with the surface described by a slightly modified version of the reaction-diffusion GS model [25],

extended symmetrically along the longitudinal direction, thus making it possible to couple it with the atmosphere aloft. The coupling is realised by imposing that the relaxation temperature profile that provides the baroclinic forcing to the atmospheric component is enslaved, through a simple representation of radiative convective adjustment, to the surface temperature field of the GS model, which basically plays the role of the ocean. This provides a natural and coherent separation between fast variables (atmosphere) and slow variables (ocean). The ocean exchanges energy with the atmosphere aloft as well as featuring absorption of shortwave radiation and emission of longwave radiation. This newly introduced model sits, in terms of complexity, between PUMA [39] and PlaSim [88].

PlaSim is comparatively simple with respect to the state-of-the-art climate models considered in the compilation of the latest IPCC report [14], but nonetheless includes a (simplified) treatment via parameterisations of important physical processes such as convection, sea ice formation, cloud formation, precipitation in liquid and solid forms, boundary layer processes, which are entirely absent in PUMA-GS, plus a much more advanced treatment of radiative processes. Additionally, PlaSim allows for a flexible configuration of the land surface mask and of the orography, with an ensuing large variety of configurations for the ocean. Nonetheless, the ocean component of the model, similarly to PUMA-GS, has merely the role of horizontally transporting heat via diffusion, while not featuring any description of actual ocean currents.

The challenge we address is threefold:

- We want to show that the edge tracking algorithm works also in the context of a system characterized by multiscale dynamics, a variety of positive and negative feedbacks, inhomogeneous physical and mathematical properties, and featuring intense fluxes across its subdomains. These features potentially result in a time-dependent edge state, possibly characterized by nontrivial dynamics and nontrivial geometrical properties, and that cannot be found as a solution to a boundary value problem. This suggests an increased potential and scope for the methodology originally proposed by Eckhardt and collaborators.
- We want to specifically use the edge tracking algorithm for identifying the edge states of the climate system separating the warm climate states from the snowball states. We want to study what lies in-between the two curves corresponding to the stable climates, portrayed in, e.g. figure 3(a), and derive the long term statistics of the climatic edge states, mirroring what has been shown in the case of edge states in turbulent flows in, e.g. figure 4.
- As we observe that edge states are relative (in the reduced space of the basin boundary) attractors, we want to complement the characterization of their climatology with the investigation of their dynamical properties, and in particular understand whether they feature trivial, quasi-periodic, or chaotic dynamics, in order to classify the properties of the weather of such special states. Note that in this way we separate the slow global climatic instability (driven by the ice-albedo feedback) leading to the presence of multiple steady states, from the fast baroclinic instability [87] responsible for the variability of the weather and acting as negative feedback.

4. The climate model PUMA-GS: formulation and edge-tracking

4.1. Model formulation

The atmospheric component of the PUMA-GS model is provided by PUMA [39]: the dry hydrostatic primitive equations on the sphere (mapped horizontally by the latitude ϕ and

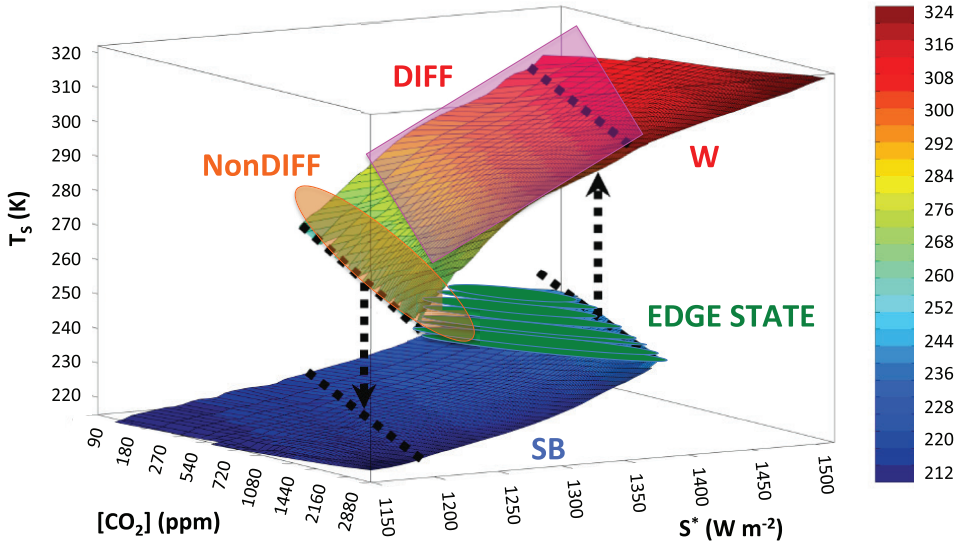

Figure 6. In [18–20, 111] we studied the macroscopic thermodynamic properties of the climate system in the W state and SB state. In [4, 15] we studied the differentiability of the invariant measure in the W state around the present climate conditions (purple shading, see DIFF). In [17] we studied the loss of differentiability of the invariant measure near the $W \to SB$ tipping point (orange shading, see NonDIFF). All of these analyses have been performed using the climate model PlaSim [88]. The goal of this paper is to charactacterize what is in-between the W and SB states, i.e. constructing and studying the edge states (green shading, see EDGE STATE). For technical reasons, we use a simplified yet Earth-like climate model, PUMA-GS. The 3D surface plot is adapted from [19] John Wiley & Sons. Copyright © 2013 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.

longitude λ) are solved by a spectral transform method (only linear terms are evaluated in the spectral domain, nonlinear terms are evaluated in grid point space). The equations for the prognostic state variables—the vertical component (with respect to the local surface) of the absolute vorticity $\zeta = \xi + 2\nu\Omega_E$ (where ξ is the vertical component of the relative vorticity, $\nu = \sin\phi$, and $\Omega_E = 2\pi$/day is the angular frequency of the Earth rotation), the (horizontal) divergence D, the (atmospheric) temperature $T_\mathrm{a} = \bar{T}_\mathrm{a} + T'_\mathrm{a}$ (separated into a time-independent arbitrary reference part $\bar{T}_\mathrm{a}$ and anomalies T'_a) and the logarithmic pressure (normalised by the surface pressure p_s) $\sigma = \ln p/p_s$—read as follows:

$$\partial_t\zeta = s^2\partial_\lambda F_v - \partial_\nu F_u - \tau_f^{-1}\xi - K\nabla^8\xi, \tag{8}$$

$$\partial_t D = s^2\partial_\lambda F_u + \partial_\nu F_v - \nabla^2[s^2(U^2+V^2)/2 + \Phi + T_\mathrm{a}\ln p_s] - \tau_f^{-1}D - K\nabla^8 D, \tag{9}$$

$$\partial_t T'_\mathrm{a} = s^2\partial_\lambda(UT'_\mathrm{a}) - \partial_\nu(VT'_\mathrm{a}) + DT'_\mathrm{a} - \dot\sigma\partial_\sigma T_\mathrm{a} + \kappa T_\mathrm{a}\omega/p + \tau_c^{-1}(T_\mathrm{R}(T_\mathrm{s}) - T_\mathrm{a}) - K\nabla^8 T'_\mathrm{a}, \tag{10}$$

$$\partial_t \ln p_s = -s^2\partial_\lambda \ln p_s - V\partial_\nu \ln p_s - D - \partial_\sigma\dot\sigma, \tag{11}$$

$$\partial_{\ln\sigma}\Phi = -T_\mathrm{a}, \tag{12}$$

where $s^2 = 1/(1-\nu^2)$, $F_u = V\zeta - \dot{\sigma}\partial_\sigma U - T'_a\partial_\lambda \ln p_s$, $F_v = -U\zeta - \dot{\sigma}\partial_\sigma V - T'_a s^{-2}\partial_\nu \ln p_s$ $\dot{\sigma}$ is the vertical velocity $U = u\cos\phi$, $V = v\cos\phi$, u, v are the horizontal wind velocity components, $\kappa = R/C_p$ (where R is the gas constant and C_p the specific heat capacity at constant pressure for dry air), and Φ is the geopotential height. Equations (8), (9) describe the budget of momentum in the horizontal direction, equation (10) describes the budget of internal energy, equation (11) expresses the conservation of mass, and equation (12) expresses the equation of state.

A number of simple parametrizations are adopted in order to improve the realism and the stability of the model. Firstly, the hyperdiffusion operator $K\nabla^8$ is added to the equations of vorticity, divergence and temperature, to represent *subgrid-scale* eddies. Secondly, *large-scale* dissipation of vorticity and divergence is facilitated by Rayleigh friction of time scale τ_f. Thirdly, the physics of diabatic heating due to radiative heat transport is parametrized by Newtonian cooling: the temperature field is relaxed (with a time scale τ_c) towards a reference or *restoration* temperature field T_R, which can be considered a radiative-convective equilibrium solution. We adopt the following simple expression for the restoration temperature [39]:

$$T_R = (T_R)_{tp} + \sqrt{[L(z_{tp} - z(\sigma))/2]^2 + S^2} + L(z_{tp} - z(\sigma))/2, \tag{13}$$

$$(T_R)_{tp} = [T_s] - \bar{L}z_{tp}, \tag{14}$$

$$L(\lambda,\phi) = \partial_z T_R = (T_s(\lambda,\phi) - (T_R)_{tp})/z_{tp}, \tag{15}$$

where $(T_R)_{tp}$ and z_{tp} are the temperature and height of the tropopause, respectively, L ($\bar{L}$) is the (average) lapse rate, $[T_s]$ is the globally averaged surface temperature, and $z(\sigma)$ is determined by an iterative procedure [39]. The above expressions indicate that the restoration temperature profile is *anchored* to the surface temperature T_s. However, as equation (14) indicates, T_R at any one point on the sphere is determined by not only the local (dynamical) surface temperature, but also the global average $[T_s]$.

The surface temperature is taken to be governed by the 2D version of the GS EBM [25], which is in the form of equation (2), while the precise expression can be found in [38]. Since land masses are absent in this configuration, we are in fact dealing with an aquaplanet climate model. Two changes are introduced to the GS EBM evolution equations with respect to our previous study [38]:

- We introduce a longitudinal component in the model (even if diffusion takes places only along the meridional direction); adding a second dimension is important because we introduce the following atmosphere-to-surface coupling term on the right-hand-side of the equation for the field of the surface temperature: $k_3(T_a(\sigma = 1) - T_s)$, where $T_a(\sigma = 1)$ depends on latitude, longitude, and time.
- We change the value of the original meridional diffusion coefficient, in order to make sure it represents the transport of the ocean only: in the original model the diffusion mimics the effects of both atmospheric and oceanic heat transports, but in PUMA-GS we have now a dynamical model for the atmosphere. We choose to reduce the value of the diffusion coefficient by a factor of 4 in order to make sure that in conditions resembling the present-day climate the ocean heat transport is few times weaker than the atmospheric one, in agreeement with observations and model outputs [3, 8, 112].

We note that $T_a(\sigma = 1)$ is obtained by linear extrapolation, according to $T_a(\sigma = 1) \approx T_a(\sigma = 0.9) + \eta(T_s - T_R(\sigma = 0.9))$, $0 < \eta < 1$. With $\eta = 1$ the coupling term is $k_3(T_a(\sigma = 1) - T_s) \approx k_3(T_a(\sigma = 0.9) - T_R(\sigma = 0.9))$. Generally $\overline{T_a(\sigma = 1) - T_s} \neq 0$ (laterally inhomogeneous heating), but $\overline{[T_a(\sigma = 1)]} = \overline{[T_s]}$, where the overbar denotes averaging with respect to time.

For our setup we choose: $K^{-1} = 0.25\,\mathrm{d}, \tau_c = 30\,\mathrm{d}, \tau_f = 1\,\mathrm{d}, \bar{L} = 0.0065\,\mathrm{K\,m^{-1}}, z_{tp} = 12\,000\,\mathrm{m}$. We also adopt a coarse resolution of T21 (i.e. the series of spherical harmonics are triangular-truncated at total wave number 21). As a result, the optimal number of Gaussian grid points is given by $N_{\mathrm{lon}} = 2N_{\mathrm{lat}} = 64$; furthermore, we choose as a number of vertical levels $N_{\mathrm{lev}} = 10$. No orography is defined, i.e. we have a zonally-symmetric configuration, and the empirical functions of the GS EBM, like $C(\phi)$, depend on the latitude only. The equations are integrated numerically using a $\Delta t = 1$ h time step size.

With such a setup we find bistability in a range of the relative solar strength $\mu = S^*/S_0^*$: the warm-to-cold (cold-to-warm) tipping point is found at $\mu_{\mathrm{W}\to\mathrm{SB}} = S^*_{\mathrm{W}\to\mathrm{SB}}/S_0^0 \approx 0.97$ ($\mu_{\mathrm{SB}\to\mathrm{W}} = S^*_{\mathrm{SB}\to\mathrm{W}}/S_0^0 \approx 1.055$). We take 18 equally spaced ($\Delta\mu = 0.005$) sample values for μ in order to study how the properties of the system change when μ is varied.

4.2. Algorithm for tracking the Melancholia states

We adapt the edge tracking algorithm so that in the initialising phase of the procedure the basin boundary between the warm and snowball states is closely bracketed by two initial conditions. See figure 7 for a cartoon representing the procedure. The initial ($j = 0$) bracketing can be obtained as follows. One starts with two arbitrary points in phase space, $x_{\mathrm{W},i,j}$ and $x_{\mathrm{SB},i,j}$, $i = 0, j = 0$ (small black disks in the schematics), belonging to the two different basins of attraction. These points need not to belong to the attractors themselves, which are marked by W and SB in figure 7 (the red and the blue object, respectively). By iterative bisecting the basin boundary (thick green line—a simple geometrical object for convenience of visualisation) is bracketed by points a distance ϵ_1 apart along a straight line in the phase space. After each bisection, a control simulation (thick black line) reveals on which side of the boundary the midpoint (indicated by the horizontal arrows) is situated. This is realised by checking the vicinity of which attractor (dashed black circles) the trajectory enters starting from the midpoint, possibly after a long waiting time. It is easier to consider a suitably defined *scalar* indicator quantity, to measure ε_1 in this single dimension, and to find scalar threshold values (indicated by horizontal dashed lines) that unmistakably indicate the outcome.

The best indicator to be used for all phases of the edge tracking algorithm is the globally averaged surface temperature $[T_{\mathrm{s}}]$, similarly to what was discussed in [38]. The basic reason for such a choice is simply the fact that at 0^{th} order the dynamics of the system can be approximated, as discussed above, by equation (1), where the energy landscape is shown in figure 2. In the case of such a simple model, for initial conditions near the saddle, the ice-albedo feedback pushes the system towards the W (SB) attractor through a monotonic increase (decrease) of temperature. In the case of the more complex model, monotonicity does not strictly hold, but its violation is restricted to small temperature scales: compare figure 9(a) to (b), and figure 9(c) to (d), respectively.

After the initialisation, we run the actual edge tracking algorithm; a cartoon is presented in figure 8. We launch the two simulations with initial conditions portrayed in figure 7 and stop them when the value of their globally averaged surface temperature is apart more than a given value ε_2. We then repeat the bisection procedure outlined in figure 7 and define two nearby initial conditions whose globally averaged surface temperature differs by less than ε_1. The two orbits originating from those two initial conditions end up in two different climates. But, since they are both initialised very close to the basin boundary, they remain near it for quite some time, as quantified by the Lyapunov exponent corresponding to the dominating unstable dimension leading to the bistability. In the limit of $\varepsilon_1, \varepsilon_2 \to 0$, we construct the dashed line

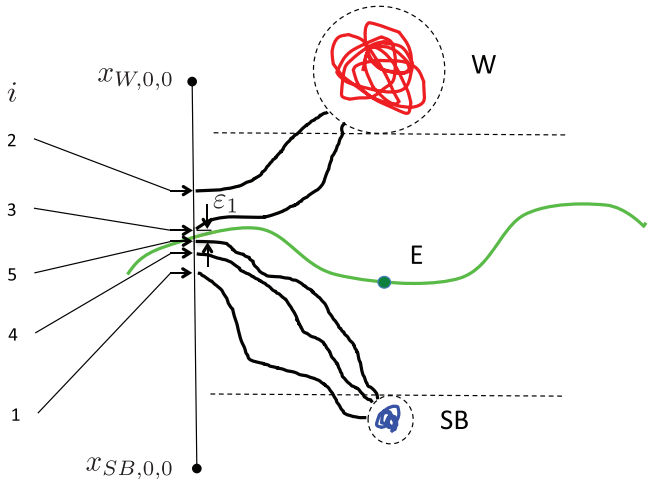

Figure 7. Scheme of the initialisation of the edge tracking procedure. By subsequent bisections we define two initial conditions near the basin boundary and near to each other such that they belong to the two different basins of attraction. The bisection is guided by testing whether our guessed initial conditions get close enough to either attractor. The snowball attractor SB, the warm attractor W and the edge state E are depicted. Details are in the text.

representing a trajectory on the basin boundary. We repeat the procedure until a statistically steady state is realised, corresponding to having reached the edge state indicated with E in figure 7. The procedure is then continued in order to be able to construct a trajectory long enough for reconstructing accurately enough the statistical properties of the edge state, and in particular for computing the maximum Lyapunov exponent (MLE) [79].

5. Results

5.1. Practical implementation of the edge tracking algorithm

Let us delve into the details of how edge tracking works in two selected cases. Results are presented in figure 9, where we plot the time series of the globally averaged surface temperature $[T_s]$ for the bracketing and control trajectories. The reason why atmospheric variables are not included in the choice of the observable has to do with the fact that they have much larger fluctuations than the surface temperature, because of the smaller heat capacity of the atmosphere with respect to the surface. As discussed above, in order to construct an efficient algorithm along the lines of what is depicted in figure 8, we want the short-term natural fluctuations of the observable to be much smaller than the difference of their climatological values in the two attractors.

After some testing we have chosen $\varepsilon_1 = 0.1$ K and, additionally, we have selected $\varepsilon_2 = 2\varepsilon_1$, which implies that a single bisection is taken per edge tracking $j-$cycle, as done in [38]. Such a bracket size is much larger than the fast variability of $[T_s]$, so that we have little risk of being influenced by spurious trends. The control trajectories are continued up to a threshold value chosen to be close to the mean value for the stable warm or cold climates.

In figure 9(a) we show how the procedure works for $\mu = 1.03$, with figure 9(b) providing a zoom: we are able to conclude that the procedure converges already in the $j = 1$ cycle almost completely (bracketing trajectories shown in red), an edge state having $[T_s] \approx 241.8$ K

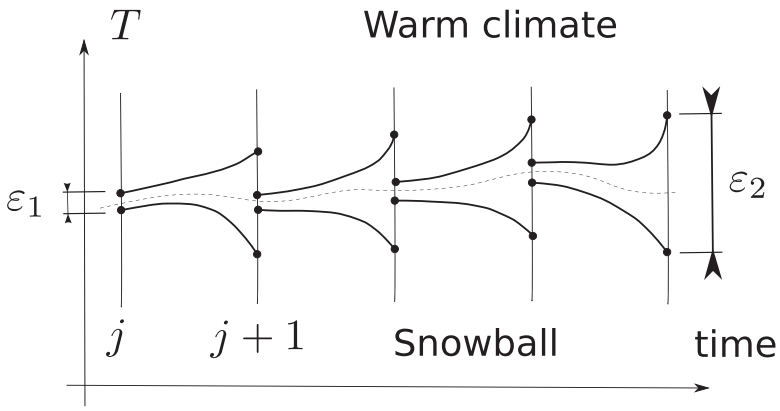

Figure 8. Scheme of the tracking method used to construct the edge state in the climate model. We let the two initial conditions defined in figure 7 evolve until the two globally averaged surface temperatures differ by ε_2. We then repeat the bisection as in figure 7 until a steady state is obtained. See details in the text.

is identified, and we observe a (closely quasi-) monotonic (in fact, approximately exponential) change of $[T_s]$ as trajectories diverge from the edge state, pulled apart by the ice-albedo feedback.

Somewhat more interesting results are shown in figure 9(c) and its zoomed-in version figure 9(d), where the edge tracking procedure is applied for $\mu = 0.98$. Figure 9(d) clarifies that the time series of $[T_s]$ of trajectories going to different attractors can, in fact, cross, so that in general no trivial monotonic increase/decrease of the chosen indicator should be expected when a complex model is considered. In other terms, the cartoon provided in figure 8 (and, *a fortiori*, ultra-simplified points of view as expressed by equation (1) and figure 2) should be taken with a grain of salt.

In the supplementary material, we include a movie showing the evolution of three orbits initialised near the Melancholia state realised for $\mu = 1.0$ towards the W state, the SB state, plus one trajectory held on the edge state through the edge tracking algorithm described here. See movie 1 (stacks.iop.org/Non/30/R32/mmedia) and movie 2 (also available on YouTube)[7] for the evolution of the atmospheric and surface temperature fields, respectively.

5.2. A plethora of states beyond the stable climates

In what follows we will show how much additional information one can draw from the climate model studied here beyond the usual bifurcation diagrams of the form shown in figure 3. We will be able to construct the Melancholia states, characterize their properties in terms of symmetry, variability, and degree of chaoticity, and prove that, near each tipping point, their properties converge to those of the stable climate that is in the process of losing stability.

Figure 10 provides a first comprehensive summary of our results and should be compared with figure 5 discussed in [38] and with figure 4(b) published in [36]. We portray the bifurcation diagram of our model, where $[T_s]$ for all the detected states is plotted versus the value of the control parameter μ. A similar bifurcation diagram can be obtained by considering the globally averaged temperature of the lowest atmospheric layer $[T_a(\sigma = 0.9)]$. The markers #1 and #2 indicated in the figure correspond to the W and SB attractors, and closely correspond to what is shown in figure 3. The marker #3 corresponds to the Melancholia state sitting

[7] Movie 1 is available at https://youtu.be/mLYZiyzO8c4. Movie 2 is available at https://youtu.be/OOYqUuG_VUE.

in-between the W and SB attractor and features decreasing values of $[T_s]$ with increasing values of μ, similarly to what is shown in figure 5 for the simple model studied in [38].

Figure 11 shows that the state corresponding to the warm attractor—#1— features a positive MLE, i.e. it has a limited horizon of predictability, while the cold state—#2—is a fixed point. What we obtain in terms of qualitative properties of the snowball state is partially unsatisfactory. Previous studies performed using comprehensive atmospheric models feature much weaker (yet non-vanishing) atmospheric variability in the snowball state than in the warm state; see e.g. [18]. In our case the atmosphere is too *quiet*, as a result of the very weak atmospheric meridional temperature gradient which is below (even if close to) the critical value needed for having sufficient baroclinicity supporting unstable motions. The main culprits for this behaviour we can indicate are the fact that (a) the PUMA-GS has no continents and no orography, which makes it impossible to have geographically limited regions where higher baroclinicity can be present, thus leading to the generation of cyclones (as preferentially happens in our planet in the westward portions of the Atlantic and Pacific storm tracks [87]); and (b) the simple ocean model we use is slightly too effective in transporting heat from the equator to the polar regions, thus reducing too much the meridional temperature gradient. Such—in our opinion, minor—limitation of our investigation could be overcome by repeating the analysis using the full PlaSim model with a setup as in [18].

Most interestingly, we are able to *identify chaoticity in the Melancholia state (*state #3*)*: we are here describing the properties of the dynamics restricted to the basin boundary between the W and SB attractors, so that the global instability due to the ice-albedo feedback is filtered out, and weather-related processes are instead retained. In other terms, we have that the Melancholia state features a dynamics of weather qualitatively analogous to what is found in the W state, including full life-cycles for mid-latitude disturbances and a non-trivial active Lorenz energy cycle. We present additional evidences of this in the movie included in the supplementary material. See the evolution of the atmospheric temperature field in the edge state belonging to $\mu = 1.0$ in Movie 1 (middle panel).

One finds—compare with figure 12—that, broadly speaking (one cannot expect a one-to-one correspondence), the value of the MLE of states #1, #2, and #3 is controlled by the intensity of the large scale meridional temperature gradient at surface, as suggested by the basics of baroclinic instability theory [87]. We note that regions of strong meridional temperature gradient for the surface temperature correspond by and large to where the variability of the atmospheric temperature near surface is more pronounced, as a result of baroclinic disturbances (not shown).

Let us go back to figure 10. As μ is increased between 1.010 and 1.015, a rather interesting phenomenon appears. The Melancholia state shown with marker #3 goes through a symmetry breaking according to the following pattern. After an extremely long transient of the order of 100 years or more (marker #4), a large longitudinal modulation of the atmospheric and surface fields suddenly appears. Movie 3 and Movie 4, available in the supplementary material and on YouTube[8] show the evolution of the atmospheric and surface temperature fields, respectively, for $\mu = 1.025$. About two-thirds of the planet warms up (marker #7 shows the meridional average at the *warmest longitude*, following the revolution of the field) and is in conditions relatively similar to those of the co-existing W state, while the rest of the planet cools down and reaches conditions relatively similar to those of the co-existing SB state (marker #9 shows the meridional average at the *coldest longitude*). The longitudinal modulation impacts all latitudes and rotates with a constant and extremely low angular velocity, and one can define the average properties of the Melancholia state (marker #5) without the need

[8] Movie 3 is available at https://www.youtube.com/watch?v=u4--tRnBBS8. Movie 4 is available at https://youtu.be/Q4YAbU9O15U.

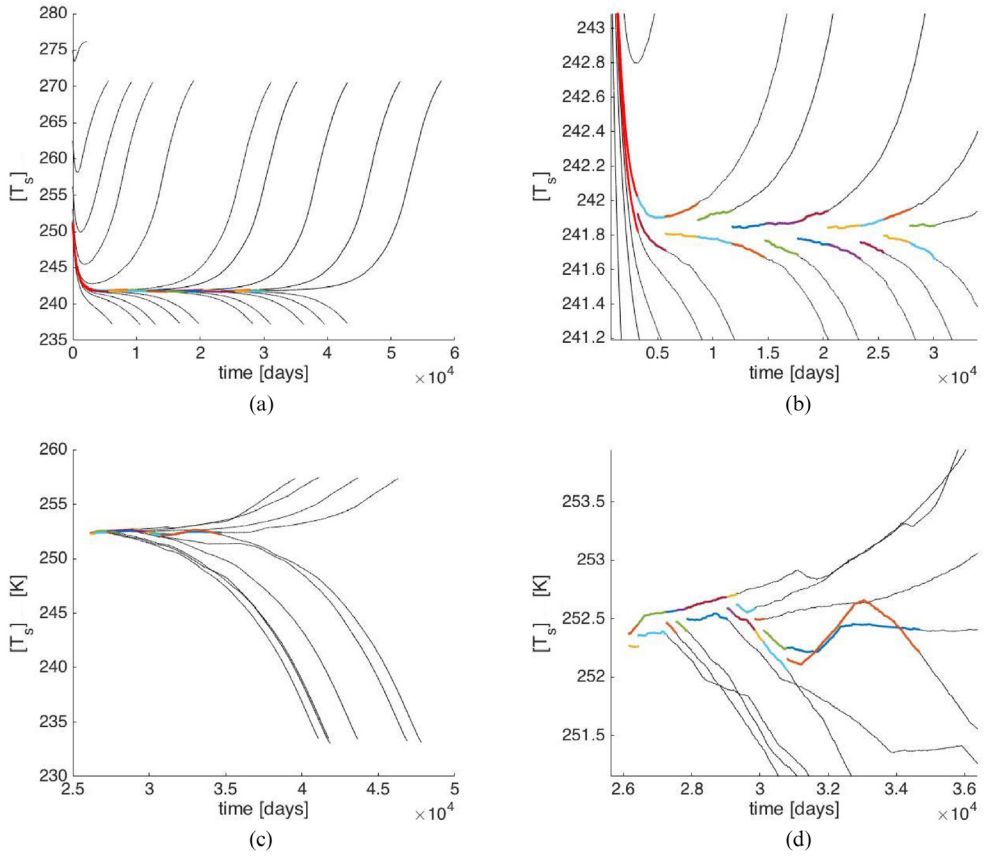

Figure 9. How the edge tracking procedure works: bracketing and control trajectories in colour and black, respectively. We plot the time series of the globally averaged surface temperature $[T_s]$. (a) $\mu = 1.03$; (b) Zoomed-in version of panel (a); (c) $\mu = 0.98$; (d) Zoomed-in version of panel (c). Details are in the text.

for considering ultralong time scales (order of 10^3 years). In order to define the properties of the system in the states #7 and #9 one needs to compute averages in a reference frame co-rotating with the temperature wave. Note that even the direction of rotation is not the same for all values of μ.

In the supplementary material (see Movie 3) one can also see that the warm portion of the domain features a large temperature gradient and supports the growth of baroclinic disturbances, which tend to decay as they reach the cold region, where the low temperature gradient does not provide sufficient baroclinic forcing. In a loose sense, the Melancholia states realised as symmetry breaking of the transient state #4 resemble some sort of chimera states [113], yet unstable ones.

An additional aspect of our results should be highlighted. For $\mu = 1.045$ we have a fundamentally different phase portrait for our model, which seems to be specific to a small neighborhood of this value of μ. We find that the system features three stable climates, which include, apart from the W and SB climate described elsewhere, the result of the bifurcation of the Melancholia state described by the markers #5, #7, and #9 into a stable state (see

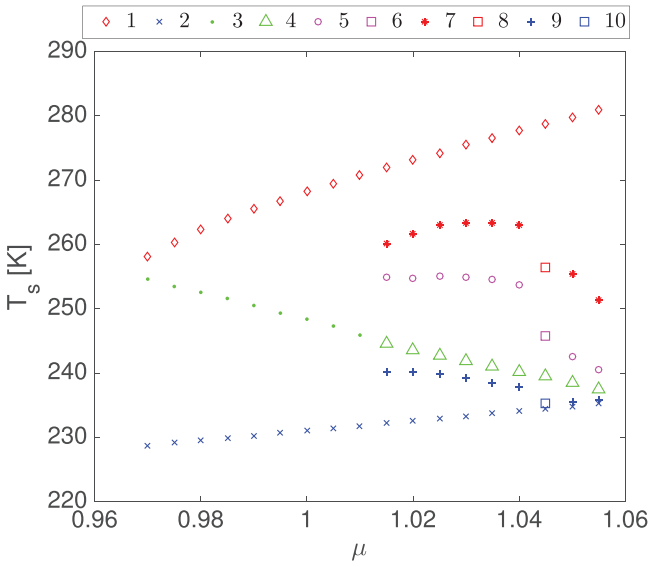

Figure 10. Bifurcation diagram for the surface temperature. The markers in the legend are number-coded and correspond to qualitatively different states in the diagram. The markers #1–#6 refer to the globally averaged surface temperature $[T_s]$. For markers #7–#10 a different averaging for the temperature field is performed. See text for details.

markers #6, #8, and #10, respectively), which displays similar features in terms of the presence of a semi-stationary pattern of strong longitudinal gradients of temperature. We are unaware of any other study reporting the existence of such an exotic climate state.

One is naturally bound to expect the presence of additional Melancholia states sitting in-between the three stable climates found for this value of $\mu = 1.045$. We do not pursue this investigation, because it would lead us to what we see at this stage as unnecessary complication. It is important to note that it would have been extremely unlikely to find such an additional *attracting climatic steady state*, had we not used the edge tracking algorithm, because the width of parametric window where it exists is small, and rather limited is also the size of its basin of attraction.

The presence of a large time scale separation between the rotation and all the other dynamical processes allows for constructing the statistics embodied in the markers #7 and #9, which are computed following the slow rotating wave, in order to guarantee homogeneity in the results.

Additionally, for $\mu \approx \mu_{\mathrm{W}\to\mathrm{SB}}$, the properties of the Melancholia state get very close to those of the W state, in agreement with the fact that at the tipping point the edge state and the W attractor touch, thus leading to the disappearance of one attractor. For $\mu \approx \mu_{\mathrm{SB}\to\mathrm{W}}$, we have that the properties of the Melancholia state (marker #5), transient Melancholia state (marker #4), and SB state tend all to converge, as the tipping point is reached.

By considering the argument of time scale separation, it is also possible to construct a (pseudo-)MLE for state #4; we cannot obviously take an infinite time horizon to compute it, because of the symmetry breaking process in action. It is also possible to compute the MLE for state #5. As shown in figure 11, one finds that the values of the (pseudo-)MLE obtained for state #4 follow closely those of the actual Melancholia state #3 (and obeys

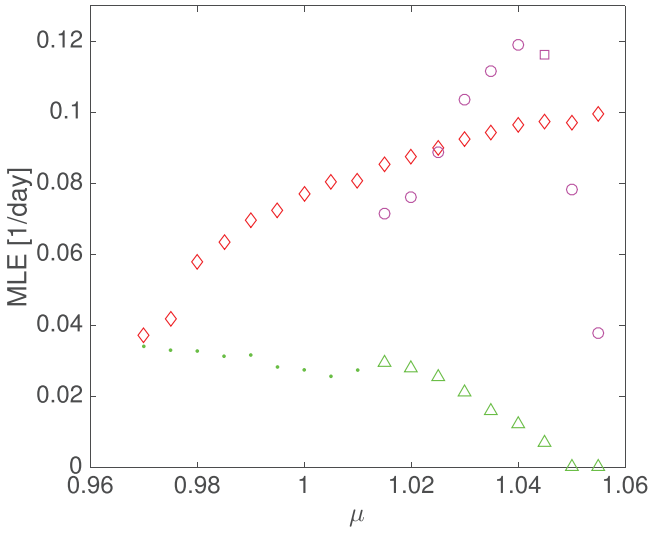

Figure 11. Bifurcation diagram of the maximum Lyapunov exponents (MLE). When the MLE is larger than zero we have a limited horizon of predictability due to chaotic dynamics. Note that while the MLE can be rigorously defined for the states #1, #3, and #5, a word of caution must be used in the case of states #4. Same number coding as in figure 10. Details in the text.

the previously discussed relationship between the value of the MLE and the strength of the meridional temperature gradient), while the MLE values obtained for state #5 are much larger (even larger than those realised in state #1), despite the presence of a weaker average meridional temperature gradient; see figure 12. This can be explained by considering that in state #5, in addition to the presence of a meridional temperature gradient, the system experiences also a large longitudinal temperature gradient between the warm and cold portions of the planet (states #7 and #9, respectively). Baroclinic conversion is very powerful in this region of large temperature gradients. The strongly nonlinear dynamics of cyclones when transiting from the relatively cold to the relatively warm regions (see Movie 3 in the supplementary material) is a manifestation of this effect.

A further characterization of the dynamics of the model can be obtained by looking at the power spectrum of an atmospheric variable, keeping in mind that a broadband spectrum is a signature of chaotic dynamics. The power spectra for $T_a(\sigma = 0.9)$ in a grid point situated at 30° *N* are presented in figure 13. We have that, in agreement with what is reported in figure 11, the states corresponding to the W attractor feature a broadband spectrum for all values of μ, while no variability is present for the SB states (not shown).

As for states #3 and #4, we have that for increasing μ the total intensity *S* and the breadth of the spectrum decreases as well as featuring an overall shift to lower frequencies, as a result of the decrease of the meridional temperature gradient (see figure 12), and, by thermal wind relation [87], of the intensity of the jet, which defines the time scales of the mid-latitude. For very large values of μ, the system reaches a quasi-periodic behaviour, where one or few frequencies are present in the spectrum.

When constructing the power spectra of states #7–#10 (note that no information on these states could be obtained by looking at the MLE), we make sure to discard the effect of the longitudinally slowly-moving temperature wave. The power spectra of the cold regions of the longitudinally-modulated climate, corresponding to states #9 and #10, are extremely weak,

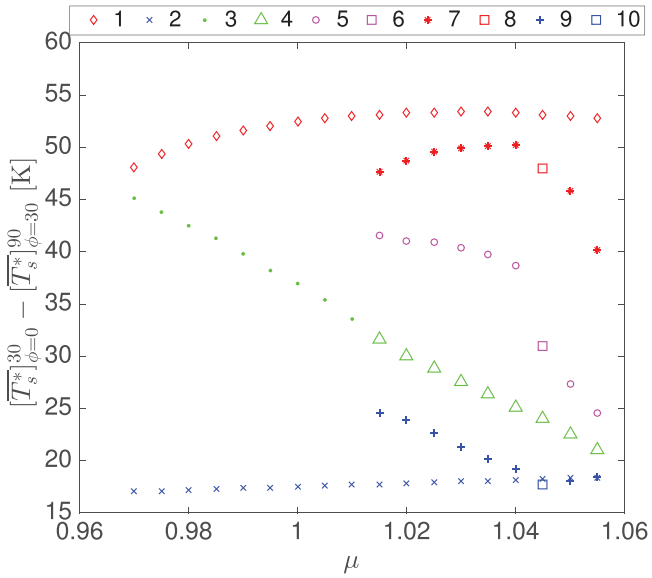

Figure 12. Bifurcation diagram for a large scale meridional surface temperature gradient defined as the difference between low-latitude and high-latitude 'boxes', as done in [38]. Same number coding as in figure 10. Details in the text.

corresponding to the fact that the variability is only inherited from the active dynamics taking place in the warm regions corresponding to the states #7 and #8. The variability is intense with broad spectrum for states #7 when $\mu < 1.04$, in agreement with the fact that a large horizontal temperature gradient is found across the mid-latitudes.

5.3. The geometry of the basin boundary

A very nontrivial aspect of the problem we are studying is the characterization of the geometrical properties of the boundary between the basins of attraction of the stable climates. The cartoons shown in figures 7–8 suggest that the two basins of attraction are separated by a simple manifold of co-dimension one. Clearly, the existence of two separate *true* attractors (as opposed to the case of transient turbulence studied by Eckhardt and collaborators) implies that the co-dimension of the boundary can be at most one, because otherwise one would automatically have holes allowing for transitions between the two basins of attraction, in contrast with their property of being invariant. Nonetheless, nothing prevents in principle the boundary from being a more complex geometrical object of co-dimension smaller than one, such that the intersection between such a boundary and a line connecting the two competing attractors has a non-trivial Cantor-like structure.

This issue has been studied in simple mathematical models in, e.g. [37, 40–43], and has practical relevance as the presence of a fractal boundary makes the prediction of the final state given the knowledge of the initial conditions of the system (with finite precision) a very non-trivial task. In other terms, it controls what Lorenz called the *predictability of the second kind* [7, 8].

As discussed in, e.g. [114], the presence of a chaotic edge state is a necessary (but not sufficient) condition for the existence of a geometrically complex basin boundary. In fact, the geometrical complexity of the boundary requires that that the MLE characterizing the separation

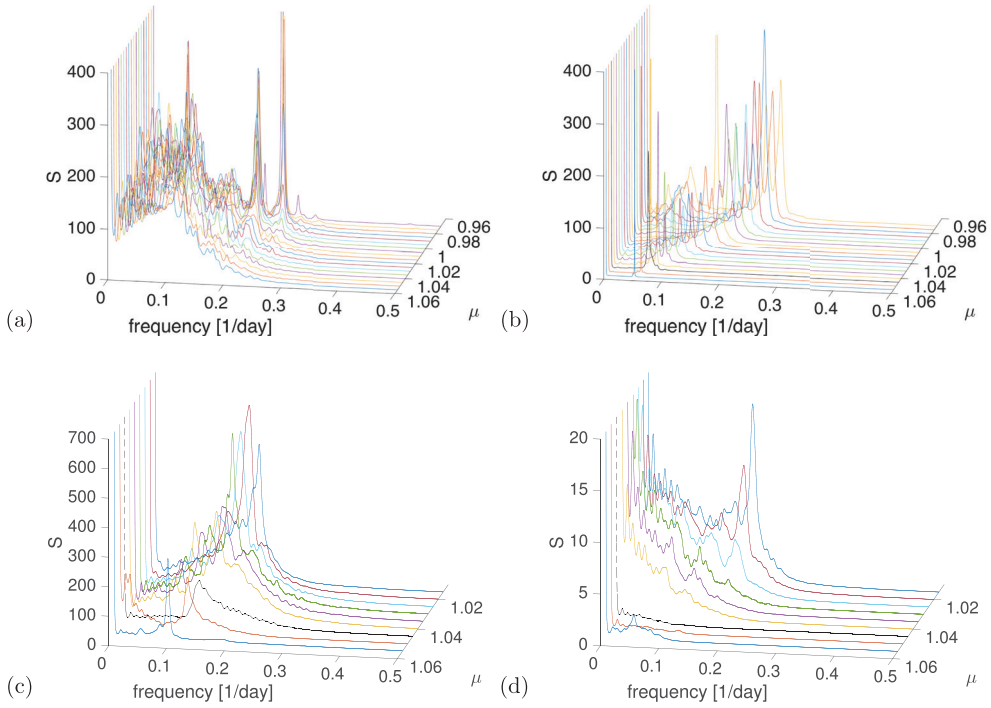

Figure 13. Power spectra of $T_a(\sigma = 0.9)$ for grid points situated at ≈30° *N* latitude.The panels describe the properties of states coded by numbered markers in figure 10; (a): state #1; (b): states #3, and #4; (c): states #7 and #8 (black line); (d): states #9 and #10 (black line). Note the different ranges of the total power *S* measured on the vertical axis. Details in the text.

of trajectories inside the basin boundary is larger than the Lyapunov exponent describing the separation of trajectories in the unstable direction (and leading to orbits ending up in either attractor) [37, 40].

Following a suggestion by A Pikovsky during a public presentation of some preliminary results then included in the present paper, we have decided to investigate the geometrical properties of the basin boundary in which the Melancholia states are embedded. We considered the case $\mu = 0.98$, for which we have already found a chaotic Melancholia state, and have considered two nearby initial conditions belonging, respectively, to the basin of attraction of the warm and of the snowball state. These two initial conditions are represented by the first bracket seen in figure 9(d). The two initial conditions are very similar in terms of dynamical and thermodynamical properties and differ by only 0.5 K in terms of the globally averaged surface temperature. We have then considered additional 1022 initial conditions obtained by equispaced convex linear interpolations of the two original initial conditions, which results in the fact that their corresponding globally averaged surface temperatures are ordered and equispaced, so that the globally averaged surface temperature of two neighbouring initial conditions differ by about 5×10^{-4} K. We then integrated forward each of these initial conditions and have labelled by 1 the orbits ending in the warm state and by 0 the orbits ending in the snowball state. The results are shown in figure 14(a). Against intuition, one finds an extremely complex geometric structure across the boundary. Very often, nearby initial conditions belong

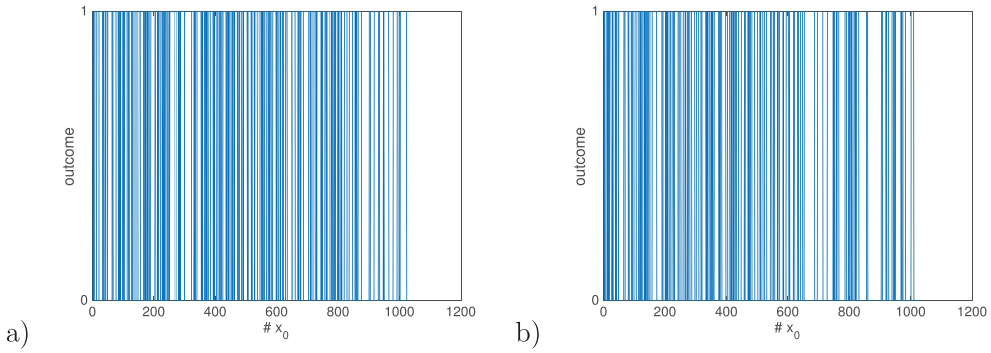

Figure 14. (a) Outcome of the forward integration of 1024 very similar initial conditions, populating a straight line in phase space that traverses the basin boundary. The difference between the globally averaged surface temperature of the 1st and of the 1024th is 0.5 K. Outcome 1 indicates that the trajectory ends in the warm state, and outcome 0 indicates that the trajectory ends in the snowball state. (b) Same as (a), but for a different set of initial conditions. Details in the text.

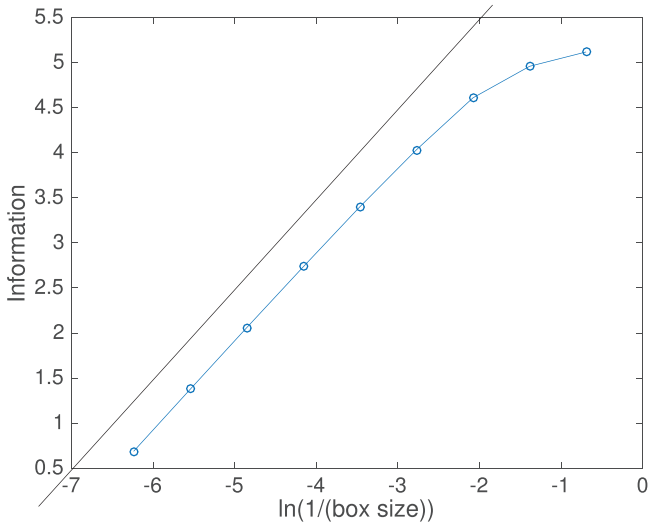

Figure 15. Computation of the fractal dimension along the transversal direction of the basin boundary. The slope of the straight line is about 0.98.

to different basins of attraction. Similar results are obtained by repeating the investigation in another region of the phase space: see figure 14(b), where we follow the procedure detailed above starting from the fifth bracket in figure 9(d). Clearly, at finite precision one cannot in principle exclude the possibility that the basin boundary is a regular—yet very tightly folded—manifold, rather than being a strange geometrical object. Further analysis is shown in figure 15, where we have computed the fractal dimension of the intersection set between the segment including the 1024 initial conditions above and the basin boundary using the data shown in figure 14(a). Entirely compatible results are obtained using, instead, the data in figure 14(b). We find a fractal dimension of about 0.98, which indicates that—at least within a certain region—the boundary has almost full dimension.

As a result, our ability to predict the final state of the system (warm or snowball) given initial conditions near the basin boundary is extremely low. Such an uncertainty can be measured using the uncertainty exponent [79]. Going back to the mathematical conditions discussed in [37, 40], we must have that the instability inside the basin boundary is (much) stronger than the instability in the unstable direction. And here physical intuition comes to our help: the instability occurring inside the basin boundary is associated to weather processes, and corresponds to the baroclinic processes (having characteristic time scales of tens of days) responsible for the generation of eddies in the Melancholia state; see figure 11. Instead, the instability in the transversal direction results from the ice-albedo feedback, and acts on time scales of years [25, 38]. In conclusion, the strangeness of the basin boundary comes from the large time scale separation between weather and climatic unstable processes.

A natural question one may ask is why, despite such a geometrical complexity, the edge tracking algorithm still works well. The reason is that the bisection algorithm removes the motions in the unstable direction associated to the ice-albedo feedback. Therefore, the constructed pseudo-trajectory along the basin boundary is constrained to explore the saddle set with no regard to its geometry and to be eventually attracted to the edge state, as a result of the fact that the phase space contracts.

6. Summary and conclusions

In this paper we have attempted to provide a review of some ideas concerning the study of critical transitions in the context of climate science. We have first briefly recapitulated under which conditions one can expect a regular response of the climate system to forcings, and have clarified that the response theory introduced by Ruelle gives the mathematical framework for addressing comprehensively the problem of predicting how the statistical properties of the climate system change as a result of (possibly time-dependent) perturbations to its dynamics. We have then clarified that the lack of regularity in the response of the system to perturbations, associated to a slowing-down in the rate at which the system decorrelates, flags being near to conditions where critical transitions take place. VL (in collaboration with Alexis Tantet) is currently exploring this problem using the formalism of transfer operator.

In order to have a more complete picture of climate dynamics, one needs to take a global point of view in the study of the phase space of the climate system and consider the possibility of the existence of multiple steady states. We have briefly discussed the literature concerning the existence of multistability in the Earth system, with a special focus on the key scientific issue regarding the co-existence of snowball and warm climate conditions given the current astrophysical and astronomical conditions. In this regard, we have presented a hierarchy of conceptual to more realistic mathematical models commonly used for studying multistability in a geophysical context and have underlined virtues and limitations of using a stochastic dynamical point of view.

While typically one wants to study the properties of the co-existing climate states and of the probability of noise-induced jumping from one to the other, the question we find most interesting and indeed novel regarding the specific climatic problem we focus on is to define what—dynamically—lies in-between the multiple (say, two) co-existing steady states of the climate system. In other terms: is there a special dynamical configuration that generalises the saddle point in an energy landscape and acts as gate facilitating the noise-induced transitions between the two attractors?

The edge state introduced by Eckhardt and collaborators for studying turbulent fluid systems has proven rather useful for our study. The edge state is the attracting set embedded in and relative to the boundary between the two basins of attraction of the co-existing attractors. The edge state attracts, in fact, almost all the initial conditions belonging to the boundary of

the basins of attraction. The edge state cannot be found by direct numerical integration, but can be constructed through a control algorithm—the edge tracking algorithm—that proceeds by bracketing the basin boundary by nearby trajectories evolving on either side.

In recent years we have followed a scientific programme aiming at a comprehensive view on climate response that has led us in the last years to study—using the same climate model—(a) the macroscopic thermodynamical properties of the climate system associated to the critical transitions responsible for the snowball/warm bistability; (b) the smooth climate response to perturbations using the Ruelle response theory; and (c) the breakdown of the smooth response near the critical transitions due to slow decay of correlations using a transfer operator method in a reduced space.

In this paper we have aimed at completing the puzzle by taking a global point of view with the analysis of dynamical regimes that do not correspond to attracting steady states of the climate, but rather constitute the edge states sitting in-between co-existing climate states. We refer to these states as Melancholia states. We have decided to stick to the point of view of statistical mechanics that proposes to investigate complex systems using a purely deterministic dynamics, with the hope of being able to understand more deeply the effect of adding noise.

By adapting the edge tracking algorithm to a climate model that is somewhat simpler than what was used in previous investigations (yet Earth-like), we have been able to identify a plethora of climate states and study the properties of the Melancholia states existing in the region of bistability where snowball and warm climate attractors coexist.

Our control parameter is μ, the ratio between the considered and the present-day solar constant. The tipping points are in fact reached when one of the climate attractors collides with the Melancholia state. We provide ample diagnostics for characterizing such states, including some movies that are available as supplementary material. The Melancholia states are obviously unstable with respect to the ice-albedo feedback, and, once such an instability is kept under control, they exhibit in some parametric range the kind of variability usually associated to regular climate conditions. The Melancholia states feature, *e.g.*, mid-latitude cyclones growing as a result of baroclinic instability fuelled by the meridional temperature gradient, and have a chaotic behaviour associated to a limited horizon of predictability.

As μ is increased from the value corresponding to the tipping point from the warm to the snowball state, one finds that the Melancholia states undergo a symmetry breaking bifurcation, after which they are characterized by strongly zonally non-symmetric conditions, in the form of a slowly rotating temperature anomaly wave. This regime is a sort of mix of the co-existing warm and snowball attractors, and feature a complex dynamics at the interface between the cold and warm regions. In a small window of values of μ, one finds three stable climate attractors, but this regime is not further explored in this paper. After such a window, the asymmetric Melancholia state is recovered and converges to the cold attractor at the tipping point associated to the transition from the snowball to the warm state.

The Melancholia states discussed in this paper correspond to climatic regimes that *cannot* be realised in nature nor can be obtained by a forward integration of a numerical model. We would like to suggest two arguments supporting the idea that the Melancholia states have, nonetheless, great physical relevance.

First, they can be used to flag *potentially dangerous* forcings acting on systems originally living in one of the coexisting attractors. Obviously, forced transitions between the two basins of attraction can take place also bypassing the Melancholia state, but, surely, if the forced system resembles more and more a Melancholia state, the risk of catastrophic changes taking place if the forcing is not stopped is very real. Clearly, the risk is much higher when the unperturbed system lives in an attractor close to the Melancholia state—so in a parametric range near the critical transition.

Second, as we have discussed in the latter part of our paper, Melancholia states might morph into actual stable yet very nontrivial stable climate states for small changes in the value of the parameters. These exotic climate states might be extremely hard to discover through direct numerical integration, given the limited size of their basin of attraction and the small parametric window where they actually exist.

The results presented provide a first example of a successful reconstruction of the global properties of a climate model possessing multiple steady states and complex regimes of motion, and hope to provide a positive stimulation in the direction of having a more thorough understanding of the properties of the climate system. It seems then relevant to extend the study of Melancholia state for the Earth and for other planets, as they may be key to understanding the evolution and the statistics of observed planetary atmospheres.

Additionally, we have shown that the boundary between the basins of attraction has a complex geometrical structure, as a result of the very different time scales associated to weather-like (baroclinic processes) and climatic instabilities (ice-albedo feedback). Obtaining a rather detailed geometrical description in such a high-dimensional system as the one analysed here seems quite an achievement. The presence of a *thick* basin boundary with a complex geometry makes the prediction of the final state given a finite precision knowledge of the initial conditions extremely challenging, thus indicating a reduced predictability of the second kind.

As a future investigation, we aim at relating the dynamical properties of the edge state to the statistics of noise-induced transitions between competing attractors, starting from simple up to more complex multistable models. As the edge state is *the* gate controlling the noise-induced transitions between the two co-existing attractors, its dynamical properties are crucial in assessing how likely a random forcing might lead to a catastrophic transition to another basin of attraction. Therefore, one could hope to derive estimates or semi-quantitative statements relating the properties of the edge state to the escape rates one can study via the Freidlin–Wentzell theory. Additionally, we wish to analyse in detail the geometry of the basin boundary and investigate the relationship between the escape rate and Lyapunov exponent along the unstable direction, along the lines indicated in [114].

Additionally, we remark that the existence of more than two stable states leads to the existence of a potentially more complex partition of the phase space in different basins of attraction, and of an ensuing more complex population of edge states sitting in-between different pairs of attractors.

In terms of specific geophysical relevance, it seems relevant to go deeper in the properties of the Melancholia states found in this study and analyse in greater detail (a) more specific physical reasons behind the presence of such a complex basin boundary (e.g. the role of localised instabilities in the physical space); (b) the physical mechanisms behind the symmetry break found in the upper range of bistability, and, (c) indeed, the apparent change of stability that morphs a Melancholia state into a new stable climate state and *vice versa* for small modulations of the solar constant.

The properties and relevance of the newly found stable climate state characterized by the break-up of the zonal symmetry and by intense horizonal temperature gradients seem indeed worth investigating. As far as we are aware, no previous studies found any qualitatively analogous stable state.

A minor pitfall of the model used in this study is that it features a stable atmospheric circulation with no fluctuations in the snowball states. Instead, the snowball states obtained using more comprehensive models have weak yet non-vanishing atmospheric variability, as baroclinic instability is active in the atmosphere, despite the low meridional temperature gradient.

Hence, repeating our analysis with a climate model like the one used in [18] seems a useful future exercise.

As discussed earlier in the paper, a more sophisticated analysis of the mechanisms behind the multistability of the Earth climate using models able to resolve accurately the oceanic time scales suggests the possibility of an additional climate state co-existing with the warm and the snowball state. A complete analysis of the Melancholia states situated in-between the possible climate states would definitely provide a more complete picture of the global dynamical properties of the Earth system and possibly lead to identifying additional, nontrivial stable climates of relevance for paleoclimate and for the study of exoplanetary systems.

Acknowledgments

The authors wish to express gratitude to B Eckhardt for multiple intellectual stimulations on the theory and intuitive aspects of the edge state and on the details of the edge tracking method; to A Pikovsky for pointing out the importance of looking at the geometry of the basin boundary; to C Grebogi, A Politi, and J Yorke for various very inspiring discussions; to two anonymous reviewers for providing insightful comments on the manuscript; and to L von Trier for directing and producing the movie *Melancholia*. The authors also wish to thank F Lunkeit and E Kirk for the great support on many computational aspects of this paper. The authors acknowledge the support of the FP7-ERC StG Grant NAMASTE, of the Horizon2020 Project CRESCENDO, and of the DFG Cluster of Excellence CliSAP. VL acknowledges the support of the DFG project MERCI. VL wishes to thank GGG for many things, which include forcing him to watch *Melancholia*.

References

[1] Ghil M and Childress S 1987 *Topics in Geophysical Fluid Dynamics: Atmospheric Dynamics, Dynamo Theory, and Climate Dynamics* (Heildelberg: Springer)
[2] Chekroun M D, Simonnet E and Ghil M 2011 *Phys. D: Nonlinear Phenom.* **240** 1685–700
[3] Lucarini V, Blender R, Herbert C, Ragone F, Pascale S and Wouters J 2014 *Rev. Geophys.* **52** 809–59
[4] Lucarini V, Ragone F and Lunkeit F 2017 *J. Stat. Phys.* **166** 1036–64
[5] Ghil M 2013 *J. Atmos. Sci.* **33** 3–20
[6] Ghil M 2015 A mathematical theory of climate sensitivity or, How to deal with both anthropogenic forcing and natural variability? *Climate Change: Multidecadal and Beyond* ed C-P Chang *et al* (Dordrecht: Kluwer) pp 31–51
[7] Lorenz E 1967 *The Nature and Theory of the General Circulation of the Atmosphere* (Geneva: World Meteorological Organization)
[8] Peixoto J P and Oort A H 1992 *Physics of Climate* (New York: American Institute of Physics)
[9] Kleidon A and Lorenz R (ed) 2005 *Non-Equilibrium Thermodynamics and the Production of Entropy* (Berlin: Springer)
[10] Lucarini V 2009 *Phys. Rev.* E **80** 021118
[11] Gallavotti G 2014 *Nonequilibrium and Irreversibility* (New York: Springer)
[12] Houghton J *et al* (ed) 2001 Intergovernmental panel on climate change *IPCC Third Assessment Report: Working Group I Report 'The Physical Science Basis'* (Cambridge: Cambridge University Press)
[13] Solomon S *et al* (ed) 2007 Intergovernmental panel on climate change *Climate Change 2007—The Physical Science Basis: Working Group I Contribution to the Fourth Assessment Report of the IPCC* (Cambridge: Cambridge University Press)
[14] Stocker T *et al* (ed) 2014 Intergovernmental panel on climate change *Climate Change 2013—The Physical Science Basis IPCC Working Group I Contribution to AR5* (Cambridge: Cambridge University Press)
[15] Ragone F, Lucarini V and Lunkeit F 2015 *Clim. Dyn.* **46** 1459–71

[16] Gritsun A and Lucarini V 2017 *Physica* D **349** 62–76
[17] Tantet A, Lucarini V, Lunkeit F and Dijkstra H A 2015 arXiv:1507.02228
[18] Lucarini V, Fraedrich K and Lunkeit F 2010 *Q. J. R. Met. Soc.* **136** 2–11
[19] Lucarini V, Pascale S, Boschi V, Kirk E and Iro N 2013 *Astron. Nachr.* **334** 576–88
[20] Boschi R, Lucarini V and Pascale S 2013 *Icarus* **226** 1724–42
[21] Lenton T M, Held H, Kriegler E, Hall J W, Lucht W, Rahmstorf S and Schellnhuber H J 2008 *Proc. Natl Acad. Sci.* **105** 1786–93
[22] Scheffer M *et al* 2012 *Science* **338** 344–8
[23] Budyko M 1969 *Tellus* **21** 611–9
[24] Sellers W 1969 *J. Appl. Meteorol.* **8** 392–400
[25] Ghil M 1976 *J. Atmos. Sci.* **33** 3–20
[26] Hoffman P, Kaufman A, Halverson G and Schrag D 1998 *Science* **281** 1342–46
[27] Saltzman B 2001 *Dynamical Paleoclimatology* (New York: Academic)
[28] Pierrehumbert R T, Abbot D, Voigt A and Koll D 2011 *Ann. Rev. Earth Planet. Sci.* **39** 417
[29] Hasselmann K 1976 *Tellus* **28** 473–85
[30] Arnold L 2001 Hasselmann's program revisited: the analysis of stochasticity in deterministic climate models *Stochastic Climate Models eds* Imkeller and von Storch (Basel: Birkhäuser) pp 141–57
[31] Touchette H 2009 *Phys. Rep.* **478** 1–69
[32] Freidlin M I and Wentzell A D 1998 *Random Perturbations of Dynamical Systems* (New York: Springer)
[33] Bouchet F, Laurie J and Zaboronski O 2014 *J. Stat. Phys.* **156** 1066–92
[34] Laurie J and Bouchet F 2015 *New J. Phys.* **17** 015009
[35] Skufca J D, Yorke J A and Eckhardt B 2006 *Phys. Rev. Lett.* **96** 174101
[36] Schneider T M and Eckhardt B 2009 *Phil. Trans. R. Soc.* A **367** 577–87
[37] Vollmer J, Schneider T M and Eckhardt B 2009 *New J. Phys.* **11** 013040
[38] Bódai T, Lucarini V, Lunkeit F and Boschi R 2014 *Clim. Dyn.* **44** 3361–81
[39] Fraedrich K, Kirk E and Lunkeit F 1998 PUMA: Portable university model of the atmosphere *Technical Report* Deutsches Klimarechenzentrum Hamburg
[40] Grebogi C, Ott E and Yorke J A 1983 *Phys. Rev. Lett.* **50** 935–8
[41] Kaplan J L, Mallet-Paret J and Yorke J A 1984 *Ergod. Theor. Dyn. Syst.* **4** 261–81
[42] McDonald S W, Grebogi C, Ott E and Yorke J A 1985 *Phys. D: Nonlinear Phenom.* **17** 125–53
[43] Grebogi C, Ott E and Yorke J A 1987 *Science* **238** 632–8
[44] Kubo R 1957 *J. Phys. Soc. Japan* **12** 570–86
[45] Kubo R, Toda M and Hashitsume N 1988 *Statistical Physics II* Nonequilibrium Statistical Mechanics (Heidelberg: Springer)
[46] Marconi U M B, Puglisi A, Rondoni L and Vulpiani A 2008 *Phys. Rep.* **461** 111
[47] Lucarini V and Colangeli M 2012 *J. Stat. Mech.* P05013
[48] Ruelle D 1997 *Commun. Math. Phys.* **187** 227–41
[49] Ruelle D 1999 *J. Stat. Phys.* **95** 393–468
[50] Ruelle D 2009 *Nonlinearity* **22** 855–70
[51] Baladi V 2000 *Positive Transfer Operators and Decay of Correlations* (Singapore: World Scientific)
[52] Butterley O and Liverani C 2007 *J. Mod. Dyn.* **1** 301–22
[53] Baladi V 2008 *Nonlinearity* **21** T81
[54] Baladi V 2014 arXiv:1408.2937
[55] Baladi V, Kuna T and Lucarini V 2017 *Nonlinearity* **30** 1204–20
[56] Gallavotti G and Cohen E G D 1995 *J. Stat. Phys.* **80** 931–70
[57] Lucarini V 2008 *J. Stat. Phys.* **131** 543–58
[58] Lucarini V 2009 *J. Stat. Phys.* **134** 381–400
[59] Lucarini V and Sarno S 2011 *Nonlinear Process. Geophys.* **18** 7–28
[60] Lorenz E 1979 *J. Atmos. Sci* **36** 1367–76
[61] Abramov R V and Majda A J 2007 *Nonlinearity* **20** 2793–821
[62] Zwanzig R 1961 *Phys. Rev.* **124** 983–92
[63] Mori H 1965 *Prog. Theor. Phys.* **33** 423–55
[64] Gritsun A and Branstator G 2007 *J. Atmos. Sci.* **64** 2558–75
[65] Gritsun A, Branstator G and Majda A 2008 *J. Atmos. Sci.* **65** 2824–41
[66] Cooper F C and Haynes P H 2011 *J. Atmos. Sci.* **68** 937–53
[67] Eyink G L, Haine T W N and Lea D J 2004 *Nonlinearity* **17** 1867

[68] Wang Q 2013 *J. Comput. Phys.* **235** 1–13
[69] Ginelli F, Poggi P, Turchi A, Chaté H, Livi R and Politi A 2007 *Phys. Rev. Lett.* **99** 130601
[70] Cvitanovíc P 1988 *Phys. Rev. Lett.* **61** 2729–32
[71] Cvitanovíc P and Eckhardt B 1991 *J. Phys. A: Math. Gen.* **24** L237
[72] Guckenheimer J and Holmes P 1983 *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields* (New York: Springer)
[73] Arnold V 1992 *Catastrophe Theory* 3d edn (Berlin: Springer)
[74] Dijkstra H 2013 *Nonlinear Climate Dynamics* (Cambridge: Cambridge University Press)
[75] Ashwin P, Wieczorek S, Vitolo R and Cox P 2012 *Phil. Trans. R. Soc. Lond.* A **370** 1166–84
[76] Lucarini V, Calmanti S and Artale V 2005 *Clim. Dyn.* **24** 253–62
[77] Lucarini V, Calmanti S and Artale V 2007 *Russ. J. Math. Phys.* **14** 224–31
[78] Rahmstorf S *et al* 2005 *Geophys. Res. Lett.* **32** L23605
[79] Ott E 2002 *Chaos in Dynamical Systems* (New York: Cambridge University Press)
[80] Pikovsky A, Rosenblum M and Kurths J 2001 *Synchronization: a Universal Concept in Nonlinear Sciences* (Cambridge: Cambridge University Press)
[81] Chekroun M D, Neelin D J, Kondrashov D, McWilliams J C and Ghil M 2014 *Proc. Natl Acad. Sci.* **111** 1684–90
[82] Tantet A, van der Burgt F R and Dijkstra H A 2015 *Chaos* **25** 036406
[83] Hoffman P F and Schrag D P 2002 *Terra Nova* **14** 129
[84] Lewis J P, Weaver A J and Eby M 2007 *J. Geophys. Res.: Oceans* **112** C11014
[85] Abbot D S, Voigt A and Koll D 2011 *J. Geophys. Res.* **116** D18103
[86] North G R, Cahalan R F and Coakley J A 1981 *Rev. Geophys.* **19** 91–121
[87] Holton J 2004 *An introduction to Dynamic Meteorology* (New York: Academic)
[88] Fraedrich K, Jansen H, Kirk E, Luksch U and Lunkeit F 2005 *Meteorol. Z.* **14** 299–304
[89] Monahan A H and Culina J 2011 *J. Clim.* **24** 3068–88
[90] Arnold L 1988 *Random Dynamical Systems* (New York: Springer)
[91] Kifer Y 1992 *Inventiones Math.* **110** 337–70
[92] Kuehn C 2011 *Phys. D: Nonlinear Phenom.* **240** 1020–35
[93] Kramers H A 1940 *Physica* **7** 284–304
[94] Hänggi P, Talkner P and Borkovec M 1990 *Rev. Mod. Phys.* **62** 251
[95] Lucarini V, Faranda D and Wouters J 2012 *J. Stat. Phys.* **147** 63–73
[96] Grafke T, Grauer R and Schäfer T 2015 *J. Phys. A: Math. Theor.* **48** 333001
[97] Faranda D, Lucarini V, Manneville P and Wouters J *Chaos, Solitons & Fractals* **64** 26–35
[98] Wouters J and Lucarini V 2012 *J. Stat. Mech.* P03003
[99] Wouters J and Lucarini V 2013 *J. Stat. Phys.* **151** 850–60
[100] Chekroun M, Liu H and Wang S 2015 *Approximation of Stochastic Invariant Manifolds: Stochastic Manifolds for Nonlinear SPDEs I* (*Springer Briefs in Mathematics*) (Berlin: Springer)
[101] Chekroun M, Liu H and Wang S 2015 *Approximation of Stochastic Invariant Manifolds: Stochastic Manifolds for Nonlinear SPDEs II* (*Springer Briefs in Mathematics*) (Berlin: Springer)
[102] Ditlevsen P D 1999 *Phys. Rev.* E **60** 172–9
[103] Imkeller P and Pavlyukevich I 2006 *Stoch. Process. Appl.* **116** 611–42
[104] Ditlevsen P D 1999 *Geophys. Res. Lett.* **26** 1441–4
[105] Ditlevsen P D and Johnsen S J 2010 *Geophys. Res. Lett.* **37** L19703
[106] Gairing J, Imkeller P, Hein C and Pavlyukevich I 2011 Mathematisches forschungsinstitut oberwolfach *Technical Report* N. 40/2011
[107] Schneider T M, Gibson J F, Lagha M, De Lillo F and Eckhardt B 2008 *Phys. Rev.* E **78** 037301
[108] Grossmann S 2000 *Rev. Mod. Phys.* **72** 603–18
[109] Hof B, Westerweel J, Schneider T M and Eckhardt B 2006 *Nature* **443** 59–62
[110] Hof B, de Lozar A, Kuik D J and Westerweel J 2008 *Phys. Rev. Lett.* **101** 214501
[111] Lucarini V, Fraedrich K and Lunkeit F 2010 *Atmos. Chem. Phys* **10** 9729–37
[112] Lucarini V and Ragone F 2011 *Rev. Geophys.* **49** RG1001
[113] Panaggio M J and Abrams D M 2015 *Nonlinearity* **28** R67
[114] Tél T and Gruiz M 2006 *Chaotic Dynamics* (Cambridge: Cambridge University Press)