Global stability properties of the climate: melancholia states, invariant measures, and phase transitions

INVITED ARTICLE • OPEN ACCESS

Global stability properties of the climate: Melancholia states, invariant measures, and phase transitions

OPEN ACCESS
IOP Publishing | London Mathematical Society
Nonlinearity

Nonlinearity **33** (2020) R59–R92 https://doi.org/10.1088/1361-6544/ab86cc

Global stability properties of the climate: Melancholia states, invariant measures, and phase transitions

Valerio Lucarini[1,2,3,6] and **Tamás Bódai**[1,2,4,5]

[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, University of Hamburg, Hamburg, Germany
[4] Pusan National University, Busan, Republic of Korea
[5] Center for Climate Physics, Institute for Basic Science, Busan, Republic of Korea

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

Received 19 March 2019, revised 17 March 2020
Accepted for publication 6 April 2020
Published 20 July 2020

Abstract

For a wide range of values of the intensity of the incoming solar radiation, the Earth features at least two attracting states, which correspond to competing climates. The warm climate is analogous to the present one; the snowball climate features global glaciation and conditions that can hardly support life forms. Paleoclimatic evidences suggest that in the past our planet flipped between these two states. The main physical mechanism responsible for such an instability is the ice-albedo feedback. In a previous work, we defined the Melancholia states that sit between the two climates. Such states are embedded in the boundaries between the two basins of attraction and feature extensive glaciation down to relatively low latitudes. Here, we explore the global stability properties of the system by introducing random perturbations as modulations to the intensity of the incoming solar radiation. We observe noise-induced transitions between the competing basins of attraction. In the weak-noise limit, large deviation laws define the invariant measure, the statistics of escape times, and typical escape paths called instantons. By constructing the instantons empirically, we show that the Melancholia states are the gateways for the noise-induced transitions. In the region of multistability, in the zero-noise limit, the measure is supported only on one of the competing attractors. For low (high) values of the solar

[6] Author to whom any correspondence should be addressed.

irradiance, the limit measure is the snowball (warm) climate. The changeover between the two regimes corresponds to a first-order phase transition in the system. The framework we propose seems of general relevance for the study of complex multistable systems. Finally, we put forward a new method for constructing Melancholia states from direct numerical simulations, which provides a possible alternative with respect to the edge-tracking algorithm.

Keywords: climate dynamics, global stability properties, invariant measure, stochastic dynamics, Melancholia states, phase transitions, basins of attraction

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

1. Introduction

In the late 1960's and in the 1970's, Budyko, Sellers, and Ghil [1–3] proposed the idea that the Earth, in the current astrophysical and astronomical configuration, supports two co-existing attractors, the warm (W) state, which is analogous to the one we live in, and the so-called snowball (SB) state, which is characterised by global glaciation and a globally averaged surface temperature of about 200–220 K. Using parsimonious yet physically meaningful energy balance models, they indicated that the bistability of the climate system is the result of the competition between the positive ice-albedo feedback (a glaciated surface reflects the incoming radiation more effectively) and the negative Boltzmann radiative feedback (a warmer surface emits more radiation to space). The relevance of the theoretical ansatz became apparent when paleoclimatic data showed that, indeed, our planet has been flipping in and out of states of global glaciations corresponding to the predicated SB states during the Proterozoic, about 650 Mya [4–6]. According to these energy balance models, the Earth's climate is bistable for a substantial range of values of the solar irradiance S^*, which include the present day value. Below the critical value $S^*_{W\rightarrow SB}$, only the SB state is permitted, whereas above the critical value $S^*_{SB\rightarrow W}$, only the W state is permitted. Such critical values, which determine the boundaries of the region in parametric space where bistability is realised, are defined by bifurcations that occur when, roughly speaking, the strength of the positive, destabilising feedbacks becomes as strong as the negative, stabilizing feedbacks. Indeed, models of different levels of complexity ranging up to the state-of-the-art Earth system models currently used for climate projections agree on predicting the existence of multistability in the climate system and point to the fundamental mechanisms described above as responsible for it, as well as providing values for $S^*_{W\rightarrow SB}$ that are in broad agreement with those obtained using simple models [7–9]. We remark that both the concentration of greenhouse gases and the position of the continents have an impact on the values of $S^*_{W\rightarrow SB}$ and $S^*_{SB\rightarrow W}$ and on the properties of the W and SB states [10]. Extremely high values of the concentration of CO_2 seem to be needed to deglaciate from an SB state [11].

Improving our understanding of the critical transitions associated to such a bistability is a key challenge of geosciences and has strong implications also in terms of the quest for understanding or anticipating planetary habitability. Planets in the habitable zone have astronomical and astrophysical configurations that allow, in principle, the presence of water at surface. Therefore, $S^*_{W\rightarrow SB}$ defines the cold boundary of the habitable zone. Clearly, an exoplanet in the habitable zone can be in the regime of bistability: if the planet is in the SB state, it will have very hard time supporting life[7]. Additionally, astronomical parameters such as the obliquity of the planet [12–14], eccentricity [12], or the length of the year [15, 16] can have a dramatic effect on the properties of multistability of a planet in the habitable zone, up to the point of erasing it altogether. In particular, planets with Earth-like atmospheres and high seasonal variability can have ice-free areas at much larger distance from the host star than planets without seasonal variability, which leads to a substantial expansion of the outer edge of the habitable zone. Additionally, one expects that tidally locked planets with an active carbon cycle can never be found in an SB state [17].

1.1. Multistability of the climate system

Further investigations have proposed the possibility of the existence of alternative cold states with respect to the SB one. Such states are characterised by the existence of a thin strip of ice-free region near the equator. Clearly, this possibility has key implications in terms of habitability and the evolution of life on Earth. Different physical mechanisms have been proposed for explaining the existence of such a state, based either on the role of a dynamical ocean [18] or the specific properties of the albedo of sea ice [19]. In fact, in a previous work [20] we have found that, indeed, a third co-existing stable state, intermediate between the SB and the W climate, can be found in a climate model featuring a very simplified representation of the oceanic heat transport, so that one might expect that the existence of more than two competing attractors could be a rather robust property of the climate system.

Indeed, the dynamical landscape of the Earth system might be even more complex than what is usually expected. A recent study [21], performed using a rather sophisticated climate model run using aquaplanet boundary conditions (without continents), indicates the existence of at least five competing climate states, ranging from a snowball to a very warm state without sea ice.

Additionally, the physics and the chemistry of the climate system feature further complexities when one considers even warmer conditions. For sufficiently large values of S^*, the W climate state loses its stability as a result of the dramatic strengthening of the positive feedback associated to the presence of water vapour in the atmosphere. Indeed, warm conditions favour, through the thermodynamic effect associated with the Clausius–Clayperon relation, the water vapour retaining capacity of the atmosphere. The water vapour is a powerful greenhouse gas, as it is active in the infrared radiation. As a result, when the concentration of water vapour is sufficiently high, the planet performs a transition to either the so-called moist greenhouse state or the so-called runaway greenhouse state (associated to a complete evaporation of the oceans) [22, 23]. This defines the warm (or inner) edge of the habitable zone [24].

In what follows, we consider the simpler—yet extremely relevant—scenario where the only relevant co-existing climates are the SB state and the W state. As discussed in [9], the physics of the system is especially interesting when the critical transitions are approached. As the solar irradiance S^* nears the critical value $S^*_{\mathrm{W}\rightarrow\mathrm{SB}}$ with the system being in the W state, the climatic engine becomes more efficient, because larger temperature gradients are realised inside the domain. Such an increased efficiency leads to a stronger atmospheric circulation, which is fuelled by temperature gradients and tends to reduce them by transporting heat from warm to cold regions, acting as a non-trivial diffusion process. Such a nonlinear equilibration mechanism acts as a negative feedback and, broadly speaking, is a macroscopic manifestation

[7] The project EDEN (http://project-eden.space/) combines ideas and methods in astrophysics, planetary sciences, and astrobiology to search for and characterize nearby habitable worlds.

of the second law of thermodynamics. One of the results of the heat transport performed by the atmospheric circulation is the stabilization of the ice-line. When $S^* = S^*_{\mathrm{W}\to\mathrm{SB}}$, the ice-albedo feedback becomes as strong as the negative feedbacks of the system, and the system flips to SB state with the ice-line reaching the equator. Similar mechanisms are in place when the system is in the SB state and S^* nears, instead, $S^*_{\mathrm{SB}\to\mathrm{W}}$. When $S^* = S^*_{\mathrm{SB}\to\mathrm{W}}$, the ice begins to melt near the equator, leading to a rapid poleward retreat of the ice line.

At the critical transitions the climate system is not anymore able to dampen fluctuations due to (infinitesimal) external forcings. Using methods borrowed from transfer operator theory [25], the investigation of the behaviour of the same model used in [9] has indeed shown that when S^* nears $S^*_{\mathrm{W}\to\mathrm{SB}}$, the spectral gap—defined as the absolute value of the (negative) imaginary part of the subdominant Ruelle–Pollicott pole [26, 27]—of the transfer operator constructed in a suitably defined reduced phase space vanishes. As a result, exponential decay of correlation is lost and the system experiences what is often referred to as *critical slowing down* [28]; see also references [29, 30].

Far from critical transitions, it has been shown [31–33] that it is possible to perform climate change projections resulting from a time-dependent CO_2 forcing using Ruelle's response theory [34]. In the case of perturbations not depending explicitly on time, response theory allows one to describe how the measure of the system changes differentiably with respect to small changes in the dynamics of the system. In the case of time-dependent perturbations, response theory makes it possible to reconstruct the measure supported on the pullback attractor [35–37] (see also the closely related concept of snapshot attractor [38, 39]) of the non-autonomous system through a perturbative approach around a reference state, which, in the case of the climate studies referred to here, corresponds to the pre-industrial conditions. When we are nearing a critical transition, it is reasonable to expect a monotonic decrease of the spectral gap [40, 41]. Hence, in the vicinity of the critical transitions the presence of a vanishing spectral gap leads to having a vanishing radius of expansion for response theory [42]. Indeed, it is expected that the (near) closure of the spectral gap is associated to a strongly enhanced sensitivity of the system's statistics to perturbations [43]. See [44] for a thorough discussion on the various regimes of climatic response to forcings and of the relationship between climate change and climate variability across various temporal scales.

1.2. Melancholia states of the climate system

A key question is what lies in-between the stable co-existing climates within the region of the parameters space where bistability is found. In simple models, it is often possible to identify unstable solutions sitting in-between the two stable climates. Such unstable solutions are embedded in the boundary between the two basins of attraction and, are roughly speaking, ice-covered up to the mid-latitudes. These solutions are saddles because they attract orbits starting from initial conditions on the basin boundary but are not stable, as small generic perturbations push the orbit outside the basin boundary with probability one and then lead to the system falling eventually into one of the competing attractors [3, 45].

When studying more comprehensive climate models featuring chaotic dynamics, things are, as described in the next section, considerably more complex from a mathematical point of view, and the individuation of the unstable saddle solutions is much harder [46–49]. Since these solutions are unstable, they cannot be found by direct numerical simulation. In a previous investigation [20], we adapted the edge tracking algorithm [50, 51] introduced for constructing the edge states, i.e. the special solutions separating laminar from long-lived turbulent regimes of motion in a fluid dynamical setting[8]. We used a recursive technique of bisection on the initial conditions[9] for shadowing trajectories on the basin boundary separating the two co-existing W and SB states, and managed to populate the corresponding saddles.

The analysis was performed using an intermediate complexity climate model with $O(10^4)$ degrees of freedom. The saddles we found had the remarkable property of featuring chaos (we found evidence that the first Lyapunov exponent was positive), and were named as Melancholia (M) states. Figures 1(a) and (b) summarize the main properties of the system, by showing the long term averages of the globally averaged surface temperature $\langle T_\mathrm{S} \rangle$ and of the temperature difference between low and high latitudes ΔT_S as a function of the relative solar irradiance $\mu = S^*/S_0^*$, where S_0^* is the present day value. By focusing on the M states we have been able to show the existence of much richer than previously thought dynamical landscape.

As discussed in [20], up to $\mu \sim 1.01$, the M state is characterised by longitudinal symmetry in its statistical properties, just as the boundary conditions of the system, are, indeed, longitudinally symmetric. The chaotic dynamics manifests itself as weather variability in a form not too dissimilar from the usual one observed in stable climates. Nonetheless, on long time scales, orbits initialised near the M states drift to either the W or the cold SB state, as a result of the dominating positive ice-albedo feedback. For $\mu \sim 1.01$, the symmetric M state becomes transient, evolving very slowly (on a time scale much longer than the other ones typical of the system) into a symmetry-broken state, where very cold and very warm conditions co-exist, separated by two regions of very strong *longitudinal* temperature gradient. The two regions feature rather different dynamical behaviour and the boundary between them rotates very slowly in time. The nontrivial bifurcation associated to such a symmetry break leads to dynamical regimes that resembles *chimera states* in extensive systems [54, 55]. The third climate state mentioned before exists in a small parametric window near $\mu \sim 1.045$ [20].

We remark that the dynamical systems viewpoint clarifies that the critical transitions for the W (snowball) state occurring when S^* approaches the critical value $S^*_{\mathrm{W}\to\mathrm{SB}}$ ($S^*_{\mathrm{SB}\to\mathrm{W}}$) are associated with the collision between the M state and the W (SB) climate, according to the dynamical scenario of boundary crisis [48]. The system's reduced ability to dampen fluctuations near the tipping points and the associated shrinking of the spectral gap described above can be seen, dynamically, as the result of the fact that the attractor *attracts orbits less effectively* in its immediate neighbourhood because of the presence of a nearby M state [28].

1.3. This paper: goals and main results

Studying multistable systems in general, and the climate system in particular, using deterministic autonomous dynamical systems faces two important issues, both resulting from the fact that the phase space is partitioned into disjoint invariant sets—the various basins of attractions and their respective boundaries. First, it is not possible to account for transitions between the co-existing basins of attraction. Instead, transitions between distinct *regimes of motion* are observed in many systems of interest. Additionally, one cannot establish an ergodic, physically relevant invariant measure, as the co-existing attractors are disjoint. Assigning a weight to each of them is, indeed, a highly arbitrary operation. Hence, one cannot answer the question

[8] Note that in such systems, technically, there are no competing attractors, because the only true attractor is the laminar state.

[9] A different approach based on control theory aims at finding unstable solutions by a feedback loop involving changes in the value of the control parameter defining the region of bistability [52, 53].

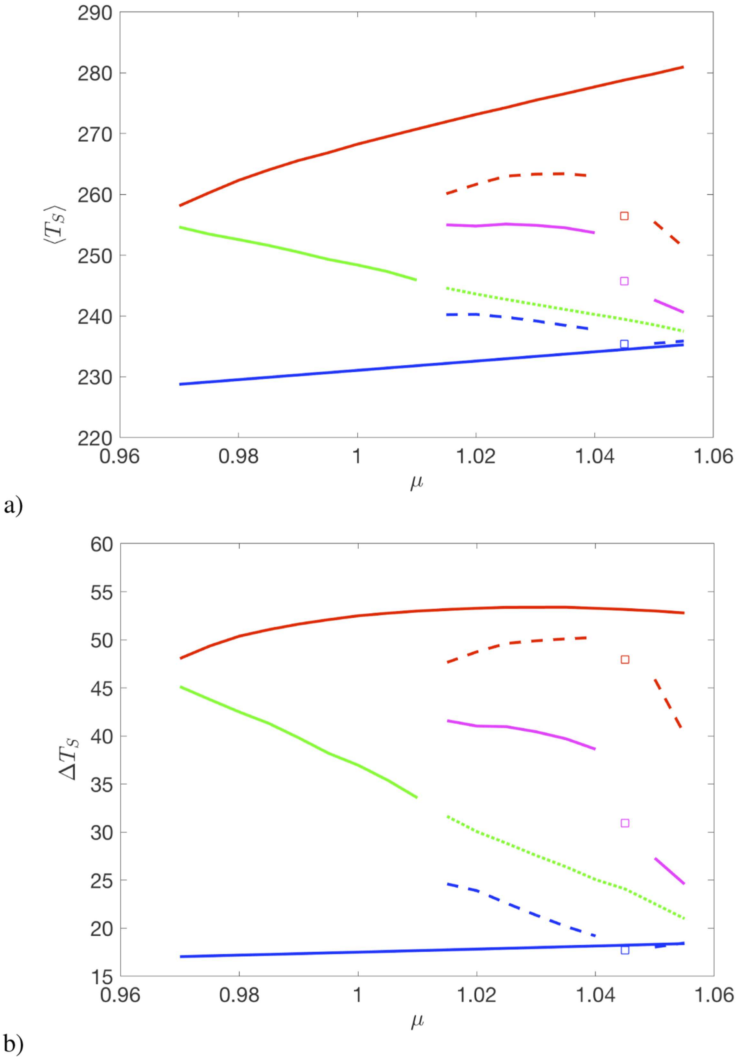

Figure 1. Bifurcation diagram for the coupled climate model studied in [20]. Panel (a) globally averaged ocean temperature $\langle T_S \rangle$ vs μ. Bistability is found for a large range of values of μ. Red continuous line: W attractor; blue continuous line: SB attractor; green continuous line: chaotic M state; purple continuous line: mean properties of the symmetry-broken chaotic M state: red dashed line: warm side of the symmetry-broken chaotic M state; blue dashed line: cold side of the symmetry-broken chaotic M state; green dotted line; transient symmetric chaotic M state; empty squares: warm side (red), cold side (blue) and average properties of the third attractor. The W-to-SB tipping point is located at $\mu \sim 0.965$; the SB-to-W tipping point is located at $\mu \sim 1.055$. See [20].

of how likely it is for the system of interest to be found at a given time in a specific regime of motion. Instead, all the statistical properties of an orbit are conditional on which invariant set its initial condition belongs to. Indeed, building upon the preliminary results reported in a short communication [56] and giving them a much broader scope and setting them in a more robust mathematical framework, we want to address these points in the present paper. For the benefit of the reader, we report below the main goals of our investigation and the main findings presented in this paper.

We will introduce a stochastic forcing to the model studied in [20] as a Gaussian perturbation with variance σ^2 modulating the intensity of the incoming solar radiation. This impacts the radiative budget of the planet in a very nontrivial way and, thanks to the ice-albedo feedback, acts as a multiplicative noise. Additionally, we conjecture, supported by physical and qualitative mathematical arguments, that the combination of noise and nonlinear deterministic dynamics leads to a hypoelliptic diffusion process, i.e. the noise propagates to all degrees of freedom of the system [57]. The presence of a random forcing allows the system to perform transitions between the neighbourhoods of the deterministic attractors, crossing the basin boundaries that, in the unperturbed case, are impenetrable; see some classical results in [58–60]. The consideration of a random forcing will allow us to construct the ergodic invariant measure of the system by performing long integrations, and investigate the properties of noise-induced transitions.

Following [49, 61–63], we show that, in the weak-noise limit and under suitable conditions, the invariant measure can be written as a large deviation law with the following notable properties. The exponent is given by minus the quasi-potential function divided by $\sigma^2/2$. We will show how to compute the quasi-potential from the drift and volatility fields, and show that the drift field can be decomposed in two contributions having radically different dynamical meaning. Additionally, the quasi-potential is a Lyapunov function and provides a clear picture of the evolution of the system in the absence of noise. In the region of bistability, the quasi-potential has local minima associated to the attractors and has a saddle behaviour at the M states of the deterministic system. In the region where only one stable state is realised in the deterministic case, the quasi-potential has just one local (and global) minimum, corresponding to the unique attractor.

In the case of multistability, the logarithm of the average permanence time in a basin of attraction increases linearly with the product of $2/\sigma^2$ times the difference between the value of the quasi-potential at the M state through which the transitions takes place and its value at the corresponding attractor. We also show that the stochastically averaged exit trajectories connect the attractor with the M state, which is indeed the most likely exit point from the deterministic basin of attraction. In the weak-noise limit, such paths correspond to the instantons [59, 60, 64–66]. In our simulations, we show that such an identification becomes more accurate as the intensity of the noise is decreased. Exploiting this property, we also propose a new method for constructing the M states via *direct* numerical integration of stochastic differential equations.

We discover that, since generally the local minima of the quasi-potential corresponding to the two attractors have different values, in the zero-noise limit only one attractor—the one corresponding to the global minimum of the quasi-potential—is populated. Nonetheless, an individual trajectory may in fact persist near the competing metastable state for a very long time, as the permanence time in the corresponding basin of attraction also diverges (but at a slower rate than for the asymptotic state). As indicated by physical intuition, for low values of S^*, the noise selects as limit measure the SB state, while for high values of S^*, the limit measure is the W attractor. The changeover takes place at a critical value of the relative solar irradiance $\mu = \mu_{\text{crit}}$, where $\mu_{\text{crit}} \approx 1.005$. For such a value of μ, the equivalent of a first-order phase transition for equilibrium systems takes place in the system. The order parameter that more obviously captures the transition is the globally averaged surface temperature.

The paper is structured as follows. In section 2 we summarise the main mathematical concepts and ideas that we have used to frame our investigation, to perform the data analysis, and to interpret our results. In section 3 we report the modelling suite we have used, describe the data we have produced, and give information on how they have been processed. In section 4 we present and discuss our results. In section 5 we propose and test numerically a new method for constructing the M states by looking at the intersections of instantons constructed by taking into account three different noise laws superimposed on to the same deterministic dynamics. In section 6 we summarize and interpret our findings, describe the limitations of our work, and propose ideas for future investigations. In appendix A, we speculate on the fact that the framework presented in the paper is potentially suitable for clarifying the role played in complex historical processes by *contingency*, as discussed by Gould [67] in the context of evolutionary biology.

2. Mathematical background

2.1. Geometry of the phase space: attractors and M states

The investigation of systems possessing multiple steady states is an extremely active area of interdisciplinary research, encompassing mathematics, natural sciences, and social sciences. The recent review by Feudel *et al* [68], introducing a special journal issue on the topic, gives a rather complete overview of the state-of-the-art of the ongoing activities on this topic and provides several interesting examples. In the context of Earth sciences, it is more common to refer to critical transitions in multistable systems using the expression *tipping points*, which has received a great popularity following a paper by Lenton *et al* [69].

A possible way to introduce the mathematical background for multistable systems is the following. We consider a smooth autonomous continuous-time deterministic dynamical system defined on a smooth finite-dimensional compact manifold $\mathcal{M}$ (in what follows, a subset of $\mathbb{R}^N$). We assume that the dynamical system is dissipative, so that the phase space continuously contracts and the Lebesgue measure is not conserved. The orbit evolves from an initial condition $\mathbf{x}_0 \in \mathcal{M}$ at time $t = 0$. We define $\mathbf{x}(t, \mathbf{x}_0) = S^t(\mathbf{x}_0)$ as the orbit at a generic time t, where S^t denotes the evolution operator. The corresponding set of ordinary differential equations written componentwise is:

$$\dot{x}_i = F_i(\mathbf{x}), \quad i = 1, \ldots, N, \tag{1}$$

where $\mathbf{F}(\mathbf{x}) = \mathrm{d}/\mathrm{d}\tau S^\tau(\mathbf{x})|_{\tau=0}$ is a smooth N-dimensional vector field. We define such a dynamical system as multistable if it features more than one asymptotic states. Specifically, we mean that there are two or more attractors Ω_j, $j = 1, \ldots, J$, each possessing a corresponding basin of attraction of a finite Lebesgue measure. Each of the attractors is an invariant set and is the support of an invariant measure of the system. The asymptotic state where the trajectory falls into is determined by its initial conditions, and the phase space is partitioned between the basins of attraction B_l of the various attractors Ω_j and the boundaries ∂B_l, $l = 1, \ldots, L$ of such basins.

We assume, for simplicity, that within each basin of attraction an ergodic measure can be defined as the limit of the empirical measure constructed by averaging over infinitely long forward trajectories for Lebesgue almost all initial conditions in the basin of attraction. In other terms, all the invariant measures of the system can be written as a convex sum of its $j = 1, \ldots, J$ ergodic components, each supported on the corresponding attractor $\Omega_j, j = 1, \ldots, J$.

We also assume that an orbit initialized on the basin boundary ∂B_l, $l = 1, \ldots, L$, is attracted towards one of the invariant saddles Π_{l_k}, k=1,...,K$_l$: the basin boundary ∂B_l contains K$_l \geqslant 1$ such saddles. We assume that all saddles feature only one unstable direction. Such instability repels trajectories initialised near the saddle towards either of the competing attractors.

In general, one can be in the situation where the asymptotic dynamics on one or more of the competing attractors can be chaotic (meaning here that at least one Lyapunov exponent is positive and unstable periodic orbits are dense). In some systems, chaotic dynamics can be realised on one or more of the saddles embedded in the basin boundaries [46–48, 70]. These are what we refer to as M states. M states are non-trivial geometrical objects that support a typically nontrivial measure [49].

A fascinating aspect of multistable systems is the following. If the first Lyapunov exponent of a chaotic saddle is larger than the inverse of the life time of the saddle state itself (which measures the rate at which orbits initialised near the saddle state are repelled towards the two nearby asymptotic states), then the basin boundary separating the basin of attraction of the two asymptotic sets has co-dimension strictly smaller than one. In two-dimensional maps, it has been proved that in this case, and provided that the first Lyapunov exponent is not the cross-boundary one, the basin boundary is not a manifold, but rather a *rough fractal* [46, 49, 70]. The authors have recently proposed a generalisation of this result to the case of *N*-dimensional maps [71]. In [20], despite high-dimensionality of the system, one could detect that the co-dimension of the basin boundary is strictly smaller than one. In fact, such a co-dimension was found to be very close to zero, as a result of time scale difference between the most relevant instabilities: the climatic instability due to the ice-albedo feedback acting near the M state is much slower than the fast weather-like instability associated to baroclinic processes occurring on the M state. This implies that near the basin boundary there is basically no predictability of the second kind in the sense of Lorenz [72]. The basin boundary is, *de facto*, a grey zone, and it is difficult to assess where orbits initialised near the boundary will end up.

2.2. Impact of stochastic perturbations: invariant measure and noise-induced transitions

The goal of our investigation here is to analyse the impact of imposing a random forcing on the deterministic dynamics discussed above. Processes that can be described as a noise-induced escape from an attractor have long been studied in the natural sciences; see [58–60]. We generalise equation (1) by adding a stochastic component. We then consider a stochastic differential equation (SDE) in Itô form written as

$$\mathrm{d}x_i = F_i(\mathbf{x})\mathrm{d}t + \sigma s(\mathbf{x})_{ij}\mathrm{d}W_j, \tag{2}$$

where $\mathbf{x}, \mathbf{F} \in \mathbb{R}^N$, $\mathbf{F}$ and $s_{ij} \in \mathbb{R}^{N\times N}$ are smooth, $\sigma \geqslant 0$, $\mathrm{d}W_j$ is the increment of an $N-$dimensional brownian motion, and $C_{ij}(\mathbf{x}) = s_{ik}(\mathbf{x})s_{jk}(\mathbf{x})$ is the noise covariance matrix. We consider the case of a hypoelliptic diffusion process. This amounts to assuming that while the covariance matrix of noise can be singular, the drift term $\mathbf{F}(\mathbf{x})$ modified according to the Stratonovich convention and the columns of the volatility matrix $\boldsymbol{s}$ satisfy the so-called Hörmander condition, i.e., the Lie algebra generated by them has dimension N everywhere [73]. As a result of this, a smooth invariant density with respect to Lebesgue is realised because the noise is propagated to all the coordinates through the drift term; see [57].[10] We discuss below in section 3.2 why such a mathematically important assumption can be heuristically justified in the specific case studied here.

Taking inspiration from the Freidlin–Wentzell [61] theory and modifications thereof [49, 62, 63], in the weak-noise limit $\sigma \to 0$ we seek a special functional form for the invariant

[10] Assuming an elliptic diffusion process is extremely restrictive because it requires all degrees of freedom to be driven by Gaussian noise.

measure. Indeed, we look for a large deviation law:

$$\Pi_\sigma(\mathbf{x}) \sim \exp\left(-\frac{2\Phi(\mathbf{x})}{\sigma^2}\right), \tag{3}$$

where the rate function $\Phi(\mathbf{x})$ is referred to as quasi-potential, and we have neglected the pre-exponential term. Specifically, the symbol $\sim$ in equation (3) implies that $\Phi(\mathbf{x}) = -1/2 \lim_{\sigma\to 0}\sigma^2 \log \Pi_\sigma(\mathbf{x})$. The function $\Phi(\mathbf{x})$ can be obtained as follows. We solve the stationary Fokker–Planck equation corresponding to equation (2):

$$\partial_j J_j(\mathbf{x}) = 0, \quad J_j(\mathbf{x}) = -F_j(\mathbf{x})\Pi_\sigma(\mathbf{x}) + \sigma^2\partial_i\left(C_{ij}(\mathbf{x})\Pi_\sigma(\mathbf{x})\right), \tag{4}$$

where $\mathbf{J}$ is the probability current. We then consider the weak-noise limit, and use as ansatz the expression given in equation (3). We obtain the following Hamilton–Jacobi equation [74]:

$$F_i(\mathbf{x})\partial_i\Phi(\mathbf{x}) + C_{ij}(\mathbf{x})\partial_i\Phi(\mathbf{x})\partial_j\Phi(\mathbf{x}) = 0. \tag{5}$$

This equation allows one to express Φ in terms of the drift and volatility fields introduced in equation (2). The quasi-potential Φ can also be computed by solving the variational problem associated with the Freidlin–Wentzell action [75]. Finally, alternative routes for computing Φ have been proposed in [76, 77].

The explicit computation of Φ is far from trivial, yet of great interest in many applications; see e.g. [78] for the case of biological systems. Brackston *et al* [79] have recently proposed an algorithm for estimating Φ in the case that the governing equations are polynomial and involves solving an optimization over the coefficients of a polynomial function. Instead, Tang *et al* [80] proposed a variational method for estimating in stochastically forced system the fraction of probability corresponding to each deterministic attractor without resorting to the computation of the invariant measure.

Following [62, 63], we now describe the dynamical meaning of Φ. Indeed, solving the previous Hamilton–Jacobi equation corresponds to the fact that it is possible to write the drift vector field as the sum of two vector fields:

$$F_i(\mathbf{x}) = R_i(\mathbf{x}) - C_{ij}(\mathbf{x})\partial_j\Phi(\mathbf{x}) \tag{6}$$

that are mutually orthogonal, so that $R_i(\mathbf{x})\partial_i\Phi(\mathbf{x}) = 0$. In the case equation (2) describes a thermodynamical system near equilibrium, $\mathbf{R}$ defines the time-reversible dynamics, while $\mathbf{F} - \mathbf{R}$ defines the irreversible, dissipative dynamics [81]. One finds that

$$\mathrm{d}\Phi(\mathbf{x})/\mathrm{d}t = -C_{ij}(\mathbf{x})\partial_i\Phi(\mathbf{x})\partial_j\Phi(\mathbf{x}) + R_i(\mathbf{x})\partial_i\Phi(\mathbf{x}) = -C_{ij}(\mathbf{x})\partial_i\Phi(\mathbf{x})\partial_j\Phi(\mathbf{x}). \tag{7}$$

As a result of this, Φ has the role of a Lyapunov function whose decrease describes the convergence of an orbit to the attractor. Specifically, $\Phi(\mathbf{x})$ has local minima at the deterministic attractors Ω_j, $j = 1, \ldots, J$, and has a saddle behaviour at the saddles Π_{l_k}, $l = 1, \ldots, L$, k=1, …,K$_\mathrm{l}$. If an attractor (saddle) is chaotic, Φ has constant value over its support, which can then be a strange set [62, 63].

Note that in the standard case of dynamics taking place in an energy landscape defined by a (confining) potential $U(\mathbf{x})$ and noise correlation matrix proportional to the identity (obtained by setting $F_i(\mathbf{x}) = -\partial_i U(\mathbf{x})$ and $C_{ij}(\mathbf{x}) = \mathbf{1}$ in the previous equations) one has $\Phi = U$. Additionally, one derives $\dot{U}(\mathbf{x}) = -\partial_i U(\mathbf{x})\partial_i U(\mathbf{x}) < 0$ and $U(\mathbf{x})$ is a Lyapunov function, and one recovers an equilibrium state, where detailed balance applies and, by definition, the current vanishes ($\mathbf{J} = 0$)[11].

We remark that the function $\Phi(\mathbf{x})$ is defined globally but is not, in general, twice differentiable everywhere. Indeed, discontinuities in its first derivatives are present if (a) the Hamiltonian associated with the Hamilton–Jacobi equation given in equation (5) is not integrable (non-integrability being the typical situation), and (b) if the system features more than one co-existing attractors. These latter discontinuities are of little practical relevance because they appear only for values of Φ larger than those at the saddles [82], for reasons that will become apparent below.

2.3. *Noise-induced escape from the attractor*

The quasi-potential Φ is key for determining the statistics of noise-induced escape from a given attractor. Indeed, the probability that an orbit with initial condition in B_j does not escape from it over a time p decays as:

$$P_{j,\sigma}(p) \approx \frac{1}{\tau_{j,\sigma}} \exp\left(-\frac{p}{\tau_{j,\sigma}}\right), \quad \tau_{j,\sigma} \propto \exp\left(\frac{2\Delta\Phi_j}{\sigma^2}\right), \tag{8}$$

where $\tau_{j,\sigma}$ is the expected escape time and $\Delta\Phi_j = \Phi(\Pi_l) - \Phi(\Omega_j)$ is the lowest quasi-potential barrier height [49], i.e. Φ has the lowest value in Π_l compared to all the other saddles neighbouring Ω_j. In general, one may need to add a correcting prefactor to $P_{j,\sigma}(p)$ [49].

Note that $\tau_{j,\sigma}$ given in equation (8) does not contain the pre-exponential factor. Reference [66] provided an expression for such pre-exponential factor for general non-equilibrium diffusion processes under the assumption that attractors and saddles are simple points, thus generalising what is given in [83]. As we aim at treating also a more general setting for the geometry of attractors and saddles, we pay below the price of having to take the pre-exponential factors as phenomenological parameters that one can find from experiments or numerical simulations [84]. We also remark that, in the zero-noise limit, the transition paths outside a basin of attraction follow the instantons. Instantons are defined as solutions of

$$\mathrm{d}x_i/\mathrm{d}t = \tilde{F}_i(\mathbf{x}) = R_i(\mathbf{x}) + C_{ij}(\mathbf{x})\partial_j\Phi(\mathbf{x}) \tag{9}$$

that connect a point in Ω_j to a point in Π_l. Instantonic trajectories have a reversed component of the gradient contribution to the vector field compared to regular—relaxation—trajectories.

Let us now take the simple case of bistable systems where we have two attractors Ω_1, Ω_2, and one saddle Π_1. We can then express the average transitions times as follows:

$$\tau_\sigma^{1\to 2} \propto \exp\left(\frac{2(\Phi(\Pi_1) - \Phi(\Omega_1))}{\sigma^2}\right), \tag{10}$$

$$\tau_\sigma^{2\to 1} \propto \exp\left(\frac{2\,(\Phi(\Pi_1) - \Phi(\Omega_2))}{\sigma^2}\right), \tag{11}$$

so that

$$\frac{\tau_\sigma^{1\to 2}}{\tau_\sigma^{2\to 1}} \propto \exp\left(\frac{2\,(\Phi(\Omega_2) - \Phi(\Omega_1))}{\sigma^2}\right). \tag{12}$$

[11] Note that, in general, we can have an equilibrium state if and only if the drift term has a gradient structure with respect to the metric defined by the noise covariance tensor [30].

This implies that, in the weak-noise limit, both escape times diverge, but the escape time out of the attractor corresponding to the lower value of the quasi-potential diverges faster. Note that one can expect the proportionality constant in equation (12) to be $O(1)$. Taking a maximally coarse-grained view on the problem, where we consider the state as represented by the populations $P_{1,\sigma}$, $P_{2,\sigma}$ of the neighbourhood of the two attractors, we can write the following master equation:

$$\dot{P}_{1,\sigma} = -\frac{P_{1,\sigma}}{\tau_\sigma^{1\to 2}} + \frac{P_{2,\sigma}}{\tau_\sigma^{2\to 1}}, \tag{13}$$

$$\dot{P}_{2,\sigma} = -\frac{P_{2,\sigma}}{\tau_\sigma^{2\to 1}} + \frac{P_{1,\sigma}}{\tau_\sigma^{1\to 2}}. \tag{14}$$

The master equation above makes sense if one assumes the presence of clear timescale separation between the relaxation motions near each attractor and those across the saddle, which depends critically on the presence of weak noise [85, 86]. At steady state, we obtain that

$$\frac{P_{1,\sigma}}{P_{2,\sigma}} = \frac{\tau_\sigma^{1\to 2}}{\tau_\sigma^{2\to 1}} \propto \exp\left(\frac{2(\Phi(\Omega_2)-\Phi(\Omega_1))}{\sigma^2}\right). \tag{15}$$

Equation (15) could also be obtained by integrating the invariant measure given in equation (3) in the neighborouhood of the attractors and taking a saddle point approximation. Additionally, equation (15) implies that in the weak-noise limit only one of the two deterministic attractors will be populated, and specifically the one where the quasi-potential has lower value. We remark that two different noise laws differing for the correlation matrix C_{ij} acting on top of the same drift field will define two different quasi-potentials, see equation (5). As a result of that, they will in general feature a different selection of the dominating population in the zero noise limit. One can easily extend the master equation defined above to the case where multiple states and multiple paths of transitions are present. Finally, note that in [87] the mathematical framework described in this section has been used to study stochastic resonance for general non-equilibrium systems.

3. Numerical modelling

The climate model considered here is constructed by coupling the primitive equations atmospheric model PUMA [88] with the Ghil–Sellers energy balance model [3], the latter describing succinctly the meridional oceanic heat transport. It has been already presented in [20] with the name of PUMA-GS, but we report here again its formulation in order to elucidate the role of stochastic forcing, which was absent in the previous version. The stochastic forcing is added as a fluctuating term modulating the value of the incoming radiation determining the energy input into the system.

3.1. The atmospheric component

The atmospheric component of the PUMA-GS model is provided by PUMA [88], which consists of a dynamical core: the dry hydrostatic primitive equations on the sphere (mapped laterally by the latitude ϕ and longitude λ), 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 $\zeta = \xi + 2\nu\Omega_{\mathrm{E}}$ (where ξ is the vertical component of the relative vorticity, $\nu = \sin\phi$, and $\Omega_{\mathrm{E}} = 2\pi/\mathrm{day}$ is the angular frequency of the Earth rotation) the (horizontal) divergence of the velocity field 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 (normalized by the surface pressure p_{s}) $\sigma = \ln p/p_{\mathrm{s}}$, read as follows:

$$\partial_t \zeta = s^2 \partial_\lambda F_v - \partial_\nu F_u - \tau_{\mathrm{f}}^{-1} \xi - K\nabla^8 \xi, \tag{16}$$

$$\begin{aligned}\partial_t D = {} & s^2 \partial_\lambda F_u + \partial_\nu F_v - \nabla^2 [s^2 (U^2 + V^2)/2 + \Psi + T_{\mathrm{a}} \ln p_{\mathrm{s}}] \\ & - \tau_{\mathrm{f}}^{-1} D - K\nabla^8 D,\end{aligned} \tag{17}$$

$$\begin{aligned}\partial_t T'_{\mathrm{a}} = {} & s^2 \partial_\lambda (U T'_{\mathrm{a}}) - \partial_\nu (V T'_{\mathrm{a}}) + D T'_{\mathrm{a}} - \dot{\sigma} \partial_\sigma T_{\mathrm{a}} \\ & + \kappa T_{\mathrm{a}} \omega / p + \tau_{\mathrm{c}}^{-1} (T_{\mathrm{R}}(T_{\mathrm{S}}) - T_{\mathrm{a}}) - K\nabla^8 T'_{\mathrm{a}},\end{aligned} \tag{18}$$

$$\partial_t \ln p_{\mathrm{s}} = -s^2 \partial_\lambda \ln p_{\mathrm{s}} - V \partial_\nu \ln p_{\mathrm{s}} - D - \partial_\sigma \dot{\sigma}, \tag{19}$$

$$\partial_{\ln \sigma} \Psi = -T_{\mathrm{a}}, \tag{20}$$

where $s^2 = 1/(1 - \nu^2)$, $F_u = V\zeta - \dot{\sigma}\partial_\sigma U - T'_{\mathrm{a}} \partial_\lambda \ln p_{\mathrm{s}}$, $F_v = -U\zeta - \dot{\sigma}\partial_\sigma V - T'_{\mathrm{a}} s^{-2} \partial_\nu \ln p_{\mathrm{s}}$, $U = u\cos\phi$, $V = v\cos\phi$, u, v being respectively the horizontal and vertical wind velocity components, and Ψ is the geopotential height. Equations (16), (17), and (19) express the conservation of momentum, equation (18) expresses the conservation of energy, and equation (20) is 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 [88]:

$$T_{\mathrm{R}} = (T_{\mathrm{R}})_{\mathrm{tp}} + \sqrt{[L(z_{\mathrm{tp}} - z(\sigma))/2]^2 + S^2} + L(z_{\mathrm{tp}} - z(\sigma))/2, \tag{21}$$

$$(T_{\mathrm{R}})_{\mathrm{tp}} = \langle T_{\mathrm{S}} \rangle - \bar{L} z_{\mathrm{tp}}, \tag{22}$$

$$L(\lambda, \phi) = \partial_z T_{\mathrm{R}} = (T_{\mathrm{S}}(\lambda, \phi) - (T_{\mathrm{R}})_{\mathrm{tp}})/z_{\mathrm{tp}}, \tag{23}$$

where $(T_{\mathrm{R}})_{\mathrm{tp}}$ and z_{tp} are the temperature and height of the tropopause, respectively, L ($\bar{L}$) is the (average) lapse rate, $\langle T_{\mathrm{S}} \rangle$ is the globally averaged surface temperature, and $z(\sigma)$ is determined by an iterative procedure [88]. The above expressions indicate that the restoration temperature profile is *anchored* to the surface temperature T_{S}. However, as equation (22) 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 $\langle T_{\mathrm{S}} \rangle$. We note that $T_{\mathrm{a}}(\sigma = 1)$ is obtained by linear extrapolation, according to $T_{\mathrm{a}}(\sigma = 1) \approx T_{\mathrm{a}}(\sigma = 0.9) + \eta(T_{\mathrm{S}} - T_{\mathrm{R}}(\sigma = 0.9))$, $0 < \eta < 1$. With $\eta = 1$ the coupling term is $k_3(T_{\mathrm{a}}(\sigma = 1) - T_{\mathrm{S}}) \approx k_3(T_{\mathrm{a}}(\sigma = 0.9) - T_{\mathrm{R}}(\sigma = 0.9))$.

Generally $\overline{T_{\mathrm{a}}(\sigma=1)-T_{\mathrm{S}}} \neq 0$ (laterally inhomogeneous heating), but $\overline{\langle T_{\mathrm{a}}(\sigma=1)\rangle} = \overline{\langle T_{\mathrm{S}}\rangle}$, where the overbar denotes averaging with respect to time.

For our setup we choose: $K^{-1} = 0.25$ days, $\tau_{\mathrm{c}} = 30$ days, $\tau_{\mathrm{f}} = 1$ day, $\bar{L} = 0.0065$ K/m, $z_{\mathrm{tp}} = 12\ 000$ m, and $k_3 = 10^{-4}$. We also adopt a coarse resolution of T21 (i.e., the series of spherical harmonics are triangular-truncated at total wave number 21). This implies the optimal number of Gaussian grid points: $N_{\mathrm{lon}} = 2N_{\mathrm{lat}} = 64$. Finally, we consider $N_{\mathrm{lev}} = 10$ vertical layers and consider a vanishing orography, so that we have a zonally-symmetric configuration. The equations are integrated numerically using a $\Delta t = 1$ [hour] time step size.

3.2. The ocean component and the stochastic forcing

The surface temperature T_{S} is taken to be governed by the a 2D version of the GS EBM [3, 45]. This model includes a simple yet effective representation of the ice-albedo feedback, and basically defines the slow manifold of the coupled atmosphere–ocean system. The partial differential equation describing the evolution of the ocean surface temperature field $T_{\mathrm{S}} = T_{\mathrm{S}}(t, \phi, \lambda)$ is:

$$\partial_t T_{\mathrm{S}}(t,\phi,\lambda) = \mu \frac{I(\phi)}{C(\phi)} \frac{S_0^*}{4}(1-\alpha(\phi,T_{\mathrm{S}})) - \frac{O(T_{\mathrm{S}})}{C(\phi)} - \frac{\bar{D}_\phi[T_{\mathrm{S}}]}{C(\phi)} + \frac{\chi[T_{\mathrm{S}},T_{\mathrm{A}}]}{C(\phi)} + \mathrm{s.f.}, \quad (24)$$

where S_0^* is the present solar irradiance[12], $\mu = S^*/S_0^*$ as introduced in section 1, while the heat capacity $C(\phi)$ and the geometrical factor $I(\phi)$ are explicitly dependent on ϕ only, thus enforcing zonally-symmetric boundary conditions. The albedo α depends on ϕ and, critically, on T_{S}, with a rapid transition from strong albedo for low values of T_{S} ($\alpha_{\max} = 0.6$) to weak albedo for $T_{\mathrm{S}} \gtrsim 260$ K ($\alpha_{\min} = 0.2$), which fuels the positive ice-albedo feedback. Additionally, O is the outgoing radiation per unit area, expressed as a monotonically increasing function of T_{S} (this is responsible for the negative Boltzmann feedback, taking into account also the greenhouse effect), $\bar{D}_\phi$ is a diffusion operator parametrizing the meridional heat transport, and χ describes the heat exchange with the atmosphere. See [20, 45] for further details.

Finally, the last term on the right-hand side s.f. is the stochastic forcing, which is introduced as a random modulation of the solar irradiance given by μS_0^*. Hence, we have:

$$\mathrm{s.f.} = \sigma s(T_{\mathrm{S}},\phi,\lambda)\frac{\mathrm{d}W}{\mathrm{d}t} = \sigma\mu\frac{I(\phi)}{C(\phi)}\frac{S_0^*}{4}(1-\alpha(\phi,T_{\mathrm{S}}))\frac{\mathrm{d}W}{\mathrm{d}t}, \quad (25)$$

where σ controls the intensity of the noise, s defines the noise law, and $\mathrm{d}W$ is the increment of a one-dimensional Wiener process. Since s depends explicitly on T_{S} via the term $\alpha(\phi, T_{\mathrm{S}})$, we are dealing with a multiplicative noise law. We consider the Itô convention for noise, so that our (discretised) equations are in the form of equation (2). See a discussion on the relevance of the chosen convention in section 4.5. Adding a Gaussian random variable of variance σ at each time step Δt (1 h) of the model amounts to considering that, on the time scale $\tau = N \times \Delta t$, the relative fluctuation of the solar irradiance scales as $\sigma_\tau = \sigma/\sqrt{N}$.

As mentioned in section 2.2, our approach requires assuming the validity of the hypoellipticity condition. In order to prove this, we should test the Hörmander condition for the evolution equations of the model. This is of great relevance but is beyond the specific scope of this paper, while indeed deserving a separate and accurate investigation. However, as discussed below, we can heuristically understand why, indeed, it is reasonable to assume that stochastic forcing acting on the oceanic surface temperature propagates to all degrees of freedom of the coupled

[12] The factor 4 emerges as a result of the geometry of the Earth–Sun system [89].

system, as usually implicitly assumed in basically any numerical study of stochastically forced geophysical flows[13].

We can first approach this problem by looking at the structure of the evolution equations. The stochastic forcing given in equation (25) impacts directly the T_S field, as shown in equation (24). The T_S fields determines the restoration temperature T_R, see equations (21)–(23). In turn, the restoration temperature impacts the anomaly of the atmospheric temperature field T'_{a} (see equation (18)), which in turn affects the vorticity field ζ—equation (16)—and divergence field D—equation (17). Finally, anomalies in D impact the surface pressure p_s, as is clear from equation (19). The nonlinear terms corresponding to advective processes on the right hand side of equations (16)–(19), which contain two more of the above mentioned fields, make sure that noise propagates across all scales in each field. This latter point could be better understood by taking a truncated Fourier representation of equations (16)–(19), which, in fact, closely corresponds to the actual formulation of the numerical solving method [88] implemented here. Concluding, no dynamical or thermodynamical field and no scale within each field is insulated from the noise, even if the covariance matrix of the noise law is *extremely* singular, as noise impacts directly only a small fraction of the degrees of freedom of the system.

On more physical grounds, one can observe that the ocean surface temperature drives the restoration profile of the atmospheric temperature, and that fluctuations of such a profile modulate the atmospheric instabilities, whose energy cascades down to the smallest scales resolved by the model; see [44, 90, 91] for a detailed treatment of the energetics of the climate system.

4. Results

We first treat in detail three cases inside the region of bistability depicted in figure 1, namely, $\mu = 0.98$ (close to the tipping point $\mu_{\mathrm{W}\to\mathrm{SB}}$), $\mu = 1.0$ (corresponding to present-day solar irradiance), and $\mu = 1.02$ (in the parametric region where the M state undergoes a symmetry-break bifurcation). We then construct the weak-noise limit of the invariant measures for all the values of μ in the region of bistability. We remark that the results shown in figures 2, 3, and 5(a) have already been reported in the short communication [56], but we deem extremely useful to present them here as well, because they are now part of a more complex, coherent, and detailed narrative.

4.1. Escapes from basins of attraction and instantons

In the case of $\mu = 0.98$, we first perform a set of simulations with noise of different intensity ranging from $\sigma_\tau = 0.5\%$ to $\sigma_\tau = 1.4\%$, with $\tau = 100$ years (y). For each value of the noise intensity, we initialise 50 trajectories in the basin of attraction of the W climate and study the statistics of the escape time to the SB attractor. When a transition takes place, we stop the integrations. We observe (not shown) that for each value of σ_τ the escape times are to a good approximation exponentially distributed, thus obeying equation (8); the process of transition behaves like a Poisson process. The results on the expectation value of the transition times are presented in figure 2, where we show that, indeed, τ_σ agrees with the prediction of equation (8). Hence, it is possible to define the difference between the value of the quasi-potential Φ at the M state and at the W attractor as the slope of the straight line. For reference, we have that for $\sigma_{100\mathrm{y}} = 0.5\%$ the average escape time is about 5.2×10^3 y. We can predict that the escape

[13] Obviously, numerical truncation introduces some additional noise on all degrees of freedom of the system.

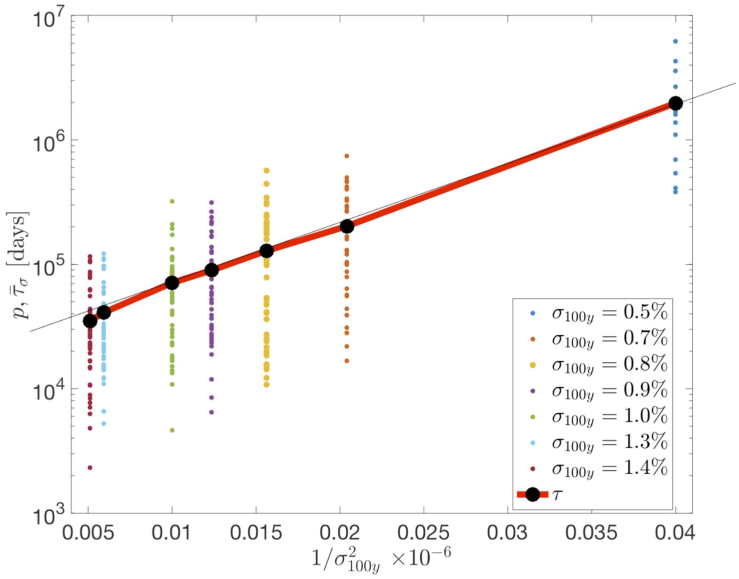

Figure 2. Statistics of the escape times p for the noise-induced W → SB transitions for various noise strengths. Each coloured dot corresponds to an observed escape time p. The estimate of the expected escape times $\bar{\tau}_\sigma$ are indicated by the black dots connected by the red line. The slope of the straight line fit gives the potential difference described in equation (10). See [84] for an optimal algorithm for estimating the potential difference. Reproduced from [56]. CC BY 4.0.

rate increases to about 1.5×10^7 y when $\sigma_{100\mathrm{y}} \sim 0.3\%$. We have that the slope of the straight line in figure 2 gives $\sim 2(\Phi(M) - \Phi(W))$. Therefore, we show that it is indeed possible to estimate quantitatively the properties of the quasi-potential also in a very high-dimensional dynamical system like the one considered here. The operation can be repeated for all the other values of μ in the range of multistability and for the processes of escape from the SB attractor, but we do not pursue here a systematic study of this.

We then wish to look at the paths corresponding to the transitions. Following the discussion in [20, 45], we choose to consider the reduced phase space spanned by the globally averaged surface ocean temperature $\langle T_\mathrm{S} \rangle$ and by the meridional temperature difference ΔT_S, defined as the difference between the spatially averaged ocean temperature field between the equator and 30°N and between 30°N and the north pole. This reduced phase space provides a minimal yet physically informative viewpoint on the problem, because it is directly linked with the main physical processes occurring in the climate model:

- The average surface temperature $\langle T_\mathrm{S} \rangle$ is directly associated to the positive ice-albedo feedback and the negative Bolzmann radiative feedback;
- The meridional temperature difference ΔT controls the meridional heat transport performed by the ocean, as a result of the diffusive law we insert into its evolution equation;
- The meridional temperature difference ΔT also controls the meridional heat transport performed by the atmosphere, as a result of the mechanism of baroclinic instability [92].

Figure 3 depicts, for the case $\sigma_\tau = 1.0\%$, the transient two-dimensional distribution function $\tilde{\rho}$ constructed using a frequentist approach using the 50 simulations described above,

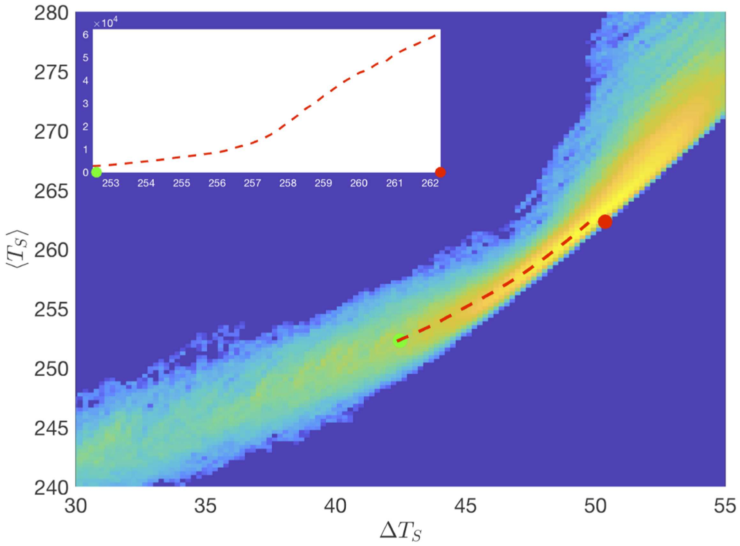

Figure 3. Main graph: logarithm of the transient density $\tilde{\rho}$ in the reduced space $(\Delta T_\mathrm{S}, \langle T_\mathrm{S} \rangle)$ (in units of K on both axes), with indication of the actual position of the W attractor (red dot) and the M state (green dot) for $\mu = 0.98$. We have used $\sigma_{100\mathrm{y}} = 1\%$. The W → SB approximate instanton is indicated. Top left inset: pdf along the path of the instanton ($\langle T_\mathrm{S} \rangle$ on the x-axis). Reproduced from [56]. CC BY 4.0.

where the statistics is collected only until the W → SB transition is realised. The distribution we obtain cannot be interpreted as an approximation of the invariant measure, because the integrations are stopped after the transitions. Nonetheless, it is apparent that the transitions take prominently place in a very narrow band linking the W attractor and the M state[14]. In order to obtain a better understanding of the transition paths, we construct an estimate of the instanton linking the W attractor to the M state and associated to the W → SB transitions by conditionally averaging the trajectories according to the value of $\langle T_\mathrm{S} \rangle$. To a good approximation, the instanton connects the W attractor to the M state, and follows a path of decreasing density. We do not find evidence of different paths for the trajectories leading to an escape and the relaxation trajectories, which is, instead, a typical signature of non-equilibrium [93]. This can be explained by considering [45], where it is shown that the ocean model evolve to a good approximation in an energy landscape. See also section 5.

4.2. Construction of the invariant measure

The information contained in figure 3 is limited because we are studying only W → SB escape processes, and we do not allow for the establishment of an invariant measure. The problem lies in the fact that the quasi-potential minimum associated to the SB attractor is much deeper than the one associated to the W attractor, so that the average escape time associated to SB → W transitions is prohibitively long for the range of (rather weak) noise intensities used for constructing figure 3. In fact, in order to be able to construct the invariant measure of the

[14] Note that, even if the W attractor and the M state look like dots, they have, in fact, a finite (yet very small) size, because they are both chaotic (see caption of figure 1). Here we are considering oceanic variables, which feature a very small variability in the deterministic chaotic case.

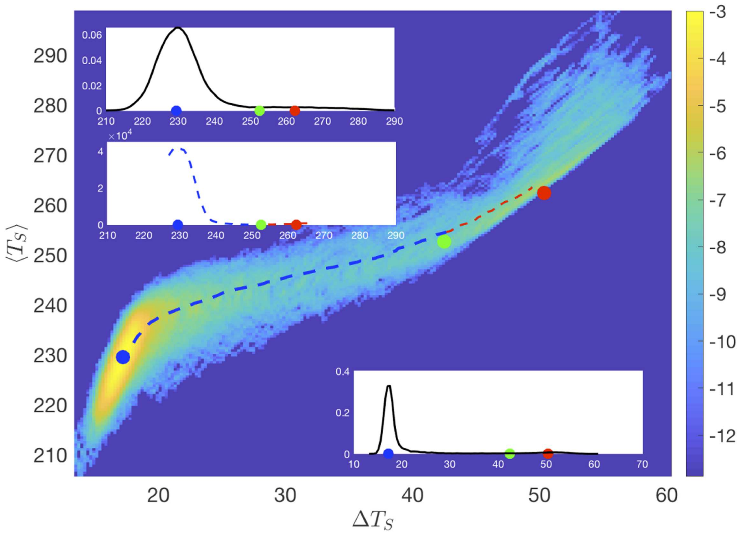

Figure 4. Main graph: density in the projected phase space (ΔT_S, $\langle T_\mathrm{S} \rangle$) (in units of K on both axes), with indication of the actual position of the W attractor (red dot), SB attractor (blue dot), M state (green dot) for $\mu = 0.98$. The W $\rightarrow$ SB and SB $\rightarrow$ W approximate instantons are also indicated. We have used $\sigma_{100\mathrm{y}} = 1.5\%$. Top left inset: marginal pdf with respect to $\langle T_\mathrm{S} \rangle$. Bottom right inset: marginal pdf with respect to ΔT_S. Center left inset: pdf along the path of the two instantons.

system, we need to observe a sufficient number of W $\rightarrow$ SB and SB $\rightarrow$ W transitions, to be sure that we have collected a satisfactory statistics; see also the master equations for the populations given in equations (13) and (14). We next increase the noise intensity by setting $\sigma_{100\mathrm{y}} = 1.8\%$, so that, in an integration lasting about 1.0×10^4 y, we observe 10 SB $\rightarrow$ W and W $\rightarrow$ SB transitions. The number of transitions is low because it is extremely hard to escape from the SB state.

Our results are shown in figure 4. We portray the logarithm of the projection of the invariant measure on the (ΔT_S, $\langle T_\mathrm{S} \rangle$) plane; we refer to this also as the two-dimensional probability distribution function (pdf). We have that the peaks of the pdf are in good agreement with the position of the W and SB attractors, as implied by the large deviation result presented in equation (3). The agreement is even clearer when considering the two marginal pdfs, constructed by projecting the invariant measure on the one dimensional spaces defined by ΔT_S and $\langle T_\mathrm{S} \rangle$. Note that the peak over the W state is hardly noticeable because the occupation of the state is extremely low (less than 5% of the total); compare with the cases studied below where higher values of μ are considered. We are also able to construct both the W $\rightarrow$ SB and the SB $\rightarrow$ W instantons, whose starting points agree remarkably well with the position of the W and the SB attractors, respectively, while their final points are in good agreement with the position of the M state. By constructing the pdf along the instantons, we find that they follow a path of monotonic descent (indeed, they follow closely the crests of the pdf), with the minimum located at the M state. Again, this last property can be hardly visualised in figure 3 for the W $\rightarrow$ SB instanton (see inset), because the population near the W state is quite small.

Next, we repeat the analysis for $\mu = 1$; results are shown in figures 5(a) and (b). In panel (a) we consider an integration with $\sigma_{100\mathrm{y}} = 1.5\%$ lasting about 2.9×10^4 y and characterised by 41 SB → W and W → SB transitions. We have an occupation of about 30% for the W basin of attraction, and of about 70% for the SB basin of attraction; additionally, we have $\tau_{\sigma}^{\mathrm{W}\rightarrow\mathrm{SB}} \sim 210\,\mathrm{y}$ and $\tau_{\sigma}^{\mathrm{SB}\rightarrow\mathrm{W}} \sim 460\,\mathrm{y}$. Also in this case, the projection of the invariant measure in the $(\Delta T_{\mathrm{S}}, \langle T_{\mathrm{S}}\rangle)$ plane shows that there is good agreement between the position of the peaks of the pdfs and the attractors, and that the estimates of the instantons connect attractors and M states with a good precision. It is also clear that the instantons follow a path of descent in terms of probability, as shown by the central inset.

In panel (b) we show the results of repeating the analysis for $\sigma_{100\mathrm{y}} = 1.8\%$. In this case the simulation lasts about 2.7×10^4 y and we obtain 73 SB → W and W → SB transitions, and we can draw similar conclusions as in panel (a) regarding the relative position of the attractors, of the M state, and of the instantons. The marginal pdfs are clearly less peaked than in panel (a); the occupancy rate changes slightly with respect to the previous case: it is about 35% for the W basin of attraction, and of about 65% for the SB. Instead, the average escape times change more substantially, and can be estimated as $\tau_{\sigma}^{\mathrm{W}\rightarrow\mathrm{SB}} \sim 160\,\mathrm{y}$ and $\tau_{\sigma}^{\mathrm{SB}\rightarrow\mathrm{W}} \sim 300\,\mathrm{y}$. These last two results indicate that the difference between the value of the quasi-potential at the two competing attractors is relatively small. We will explore this matter in section 4.4, where we will try to deduce where the quasi-potential reaches its absolute minimum for each value of μ in the bistable region.

It is worth looking more in detail at how the estimate of the instantons is impacted by the intensity of the noise used in the simulation. As instantons are defined in the weak-noise limit, we would expect that one achieves higher precision when weaker noise is used. This is confirmed by the results shown in figure 6: we have that the estimates of the instantons obtained using lower noise intensity come closer to the attractors and to the M state. Nonetheless, also the instantons obtained for very strong noise are relatively accurate.

4.3. Instantons and transitions across the symmetry-broken Melancholia state

We next examine the noise-induced transitions for $\mu = 1.02$. This case is quite interesting because, as discussed in [20] and reported in figure 1, the longitudinally-symmetric M state is transient with a very long life time, and slowly evolves into a symmetry broken M state featuring a relatively cold and a relatively warm region, separated by two small regions with large longitudinal temperature gradients at all latitudes. It seems relevant to test whether noise-induced transitions take place through the transient, symmetric M state or the true, symmetry-broken one. Results are shown in figure 7, where we use data from a simulation lasting about 10^4 y with $\sigma_{100\mathrm{y}} = 1.5\%$. We observe only 6 transitions in both directions. This time, as opposed to the case of $\mu = 0.98$, the figure is so low because it is extremely hard to escape from the W state. We estimate the escape times as $\tau_{\sigma}^{\mathrm{W}\rightarrow\mathrm{SB}} \sim 1400\,\mathrm{y}$ and $\tau_{\sigma}^{\mathrm{SB}\rightarrow\mathrm{W}} \sim 140\,\mathrm{y}$. We portray the logarithm of the invariant measure of the system in the usual projected space, and the estimates of the instantons. We first observe that in this case most of the density is concentrated around the W attractor, and a nontrivial relation exists between the paths of the instantons and the dynamical structures on the basin boundary. It is clear that the transient M state plays the role of the gateway for transitions as seen in the previous cases, despite its transient nature. The noise-induced transitions do not go through the actual M state (magenta square), while they seem to go thorough states resembling the properties of the warm (red square) and cold (blue square) sectors in the asymptotic M state. This feature might result from the consideration of noise with finite (and not infinitesimal) strength: the quasi-potential near the transient M state might be just barely higher than that of the asymptotic M state (and possibly with a more favourable pre-exponential factor), so that a small but finite noise perturbation might push an orbit near the transient M state into the other basin of attraction. In order to address

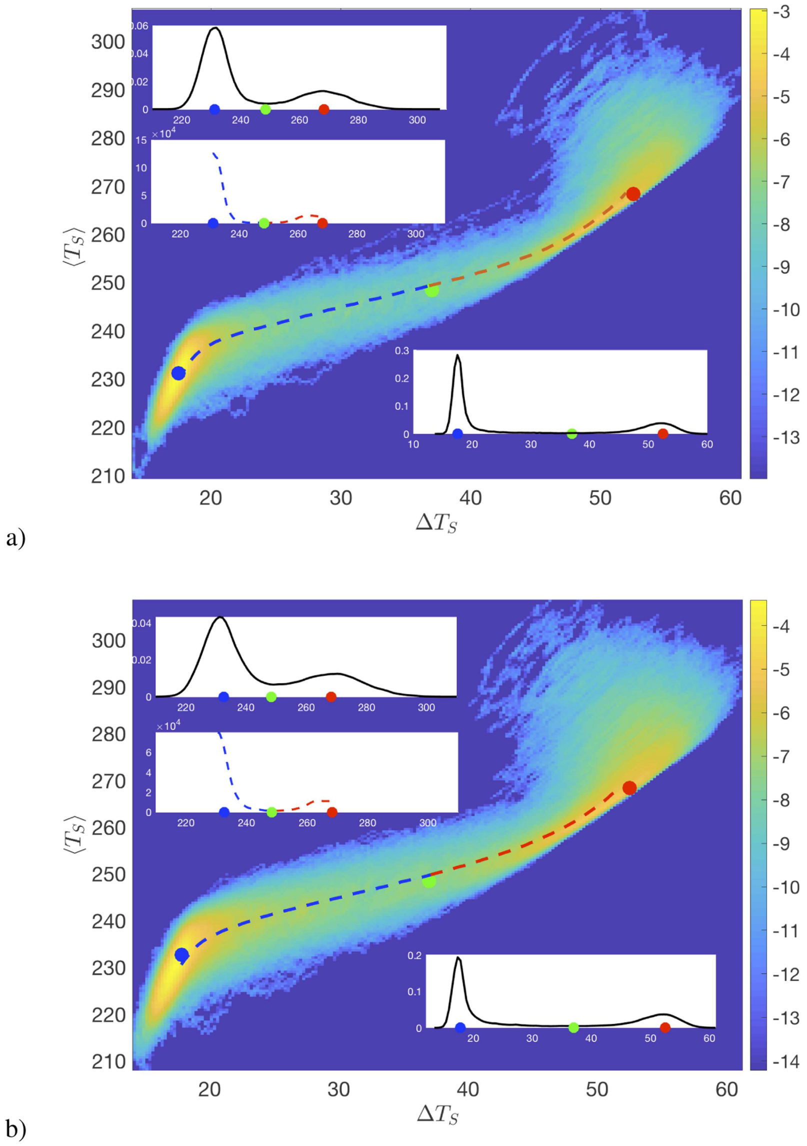

Figure 5. Panel (a) main graph: density in the projected phase space $(\Delta T_S, \langle T_S \rangle)$ (in units of K on both axes), with indication of the actual position of the W attractor (red dot), SB attractor (blue dot), M state (green dot) for $\mu = 1$. The W → SB and SB → W approximate instantons are also indicated. We have used $\sigma_{100y} = 1.5\%$. Top left inset: marginal pdf with respect to $\langle T_S \rangle$. Bottom right inset: marginal pdf with respect to ΔT_S. Center left inset: pdf along the path of the two instantons. Reproduced from [56]. CC BY 4.0. Panel (b): same as panel (a), with $\sigma_{100y} = 1.8\%$.

this point and find instantonic paths connecting the attractors with the true M states, one might need to resort to using more sophisticated numerical techniques. In particular, one should consider using rare events algorithms [94, 95] to rigorously construct instantonic trajectories [96]. This is beyond the current abilities of the authors but definitely deserves attention in future studies.

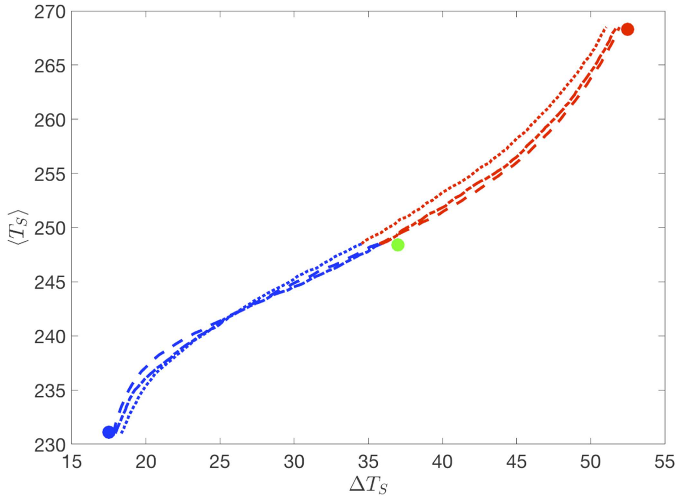

Figure 6. Estimate of the instantons for $\mu = 1$ obtained using $\sigma_{100\text{y}} = 1.5\%$ (dashed lines), $\sigma_{100\text{y}} = 1.8\%$ (dash-dotted lines), and $\sigma_{100\text{y}} = 2.5\%$ (dotted lines) (units of K on both axes). The red (blue) lines show the estimates for the W $\rightarrow$ SB (SB $\rightarrow$ W) instanton. The dots indicate the actual position of the W attractor (red dot), SB attractor (blue dot), and M state (green dot). The estimate of the instanton improves as the intensity of the noise is reduced.

4.4. Selection of the limit measure in the weak-noise limit and first-order phase transition

Equation (15) indicates that, in the weak-noise limit, all of the measure will be concentrated on the attractor featuring the lowest value for the quasi-potential Φ. Since for $\mu < S^*_{\text{W}\rightarrow\text{SB}}/S^*_0$ only the SB state is realised, physical intuition suggests that for low values of μ within the range of bistability, the SB attractor should contain all the mass in the limit of weak noise, as one can also anticipate it by looking at figure 3. Conversely, one expects that for high values of μ within the range of bistability, the W should be dominant in the weak-noise limit. It is reasonable (yet far from obvious or rigorous) to expect that there should be a critical value of $\mu = \mu_{\text{crit}}$ separating the two regimes. Following [20], we consider 18 equally spaced values of μ ($\Delta\mu = 0.005$) within the multistable regime. For each of these values of μ (excluding the case of $\mu = 1.045$, where three stable states are realised) we determine the fraction of the population residing within the basin of attraction of the deterministic W attractor $P_{\text{W},\sigma}(\mu)$ and its complement, residing in the basin of attraction of the SB attractor $P_{\text{SB},\sigma}(\mu)$. Related results are shown in figure 8. We show how the probability distribution of the variable $\langle T_{\text{S}} \rangle$ depends on μ for three different noise levels: $\sigma_{100\text{y}} = 1.5\%$ (panel a), $\sigma_{100\text{y}} = 1.8\%$ (panel b), and $\sigma_{100\text{y}} = 2.5\%$ (panel c). In these panels we superimpose the bifurcation diagram reported in figure 1(a). We observe that as the noise is reduced, for all values of μ the distributions are (a) more peaked around the attractors, and (b) one of the attractors becomes clearly dominant.

In panel (d) we plot for each value of μ the integral of the pdfs reported in the three panels (a)–(c) up to the values of $\langle T_{\text{S}} \rangle$ corresponding to the M state (green continuous and dotted lines). To a very good degree of approximation, this corresponds to the integral of the invariant measure over the support of the deterministic basin of attraction of the SB climate. We obtain that for decreasing values of the noise intensity, the emerging invariant measure converges to the deterministic measure supported on the SB attractor for $\mu \leqslant \mu_{\text{crit}} \approx 1.005$, while the invariant measure converges to the deterministic measure supported on the W attractor for

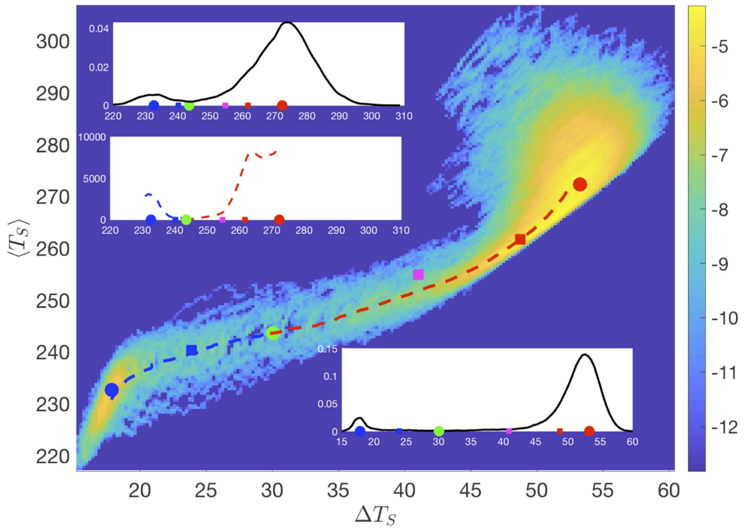

Figure 7. Main graph: density in the projected phase space (ΔT_S, $\langle T_\mathrm{S}\rangle$), with indication of the position of the W attractor (red dot), SB attractor (blue dot), transient M state (green dot) for $\mu = 1.02$ (units of K in both axes). The squares indicate the properties of the asymptotic M state: W sector (red square); cold sector (blue square); average (magenta square). The estimates of the W $\rightarrow$ SB and SB $\rightarrow$ W instantons are also indicated as the red and the blue line, respectively. We have used $\sigma_{100\mathrm{y}} = 1.5\%$. Top left inset: marginal pdf with respect to $\langle T_\mathrm{S}\rangle$. Bottom right inset: marginal pdf with respect to ΔT_S. Center left inset: pdf along the path of the two instantons.

$\mu \geqslant \mu_\mathrm{crit} \approx 1.005$. The absolute minimum of the quasi-potential Φ is realised in the W attractor for $\mu \geqslant \mu_\mathrm{crit}$ and in the SB attractor for $\mu \leqslant \mu_\mathrm{crit}$. The changeover is, curiously, quite close to the reference case $\mu = 1$, for which the weak-noise limit of the measure is given by the SB state.

We add a note on the uncertainty associated to the figures reported in figure 8(d). We remark that for all values of μ and σ we have used simulations lasting at least 10^4 y. For very low ($\leqslant 0.98$) and very large ($\geqslant 1.02$) values of μ in the considered range, the simulation length does not allow for observing more than a few transitions for the two lowest considered noise levels. Therefore, according to the fact that the transitions occur following a Poisson law, one expects in this range an uncertainty on the figures reported in figure 8 of the order of the values of the smaller between $P_{\mathrm{SB},\sigma}(\mu)$ and $P_{\mathrm{W},\sigma}(\mu)$. This corresponds, in fact, to a low uncertainty, because most of the mass is concentrated near one of the two deterministic attractors. The uncertainty is quite small also for $0.98 \leqslant \mu \leqslant 1.03$, because, in all cases, we observe relatively many transitions. The uncertainty in this range can be safely estimated to be below 5%. Summarising, while the values reported in figure 8(d) might have some non-negligible uncertainties, it seems that the estimate of μ_crit is quite robust.

Finally, we have verified that for all values of μ the escape time $\tau_\sigma^{\mathrm{W}\rightarrow\mathrm{SB}}$ and $\tau_\sigma^{\mathrm{SB}\rightarrow\mathrm{W}}$ grow rapidly with decreasing values of the intensity of the noise for all values of μ. In agreement with what is shown in figure 8, we have that $\tau_\sigma^{\mathrm{SB}\rightarrow\mathrm{W}}$ grows more rapidly than $\tau_\sigma^{\mathrm{W}\rightarrow\mathrm{SB}}$ for $\mu \leqslant \mu_\mathrm{crit}$ (and vice versa for $\mu \geqslant \mu_\mathrm{crit}$). As a result, in the weak-noise limit an individual trajectory might be trapped for a very long time in the metastable, non-asymptotic state. We remark that we have not performed here a systematic evaluation of the exponential relationship between the escape times and σ, as instead done for $\mu = 0.98$ and shown in figure 2, also because, as

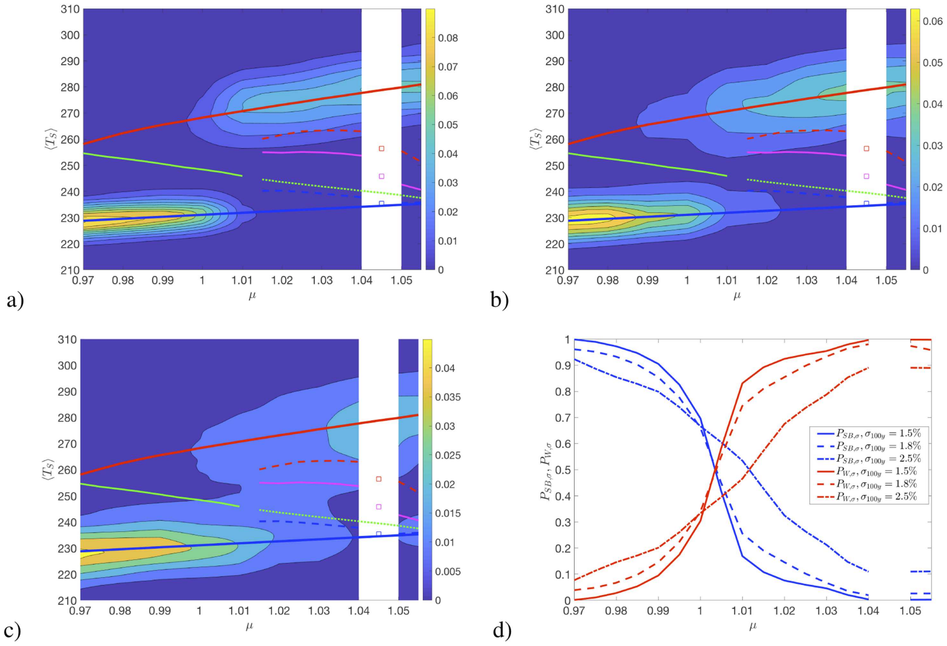

Figure 8. Projection of the measure on the variable $\langle T_{\mathrm{S}} \rangle$ (units of K) for $\sigma_{100\mathrm{y}} = 1.5\%$ (panel a), $\sigma_{100\mathrm{y}} = 1.8\%$ (panel b), and $\sigma_{100\mathrm{y}} = 2.5\%$ (panel c). The spacing of the isolines is the same in the three panels. Panel (d) fraction of the measure supported in the basin of attraction of the SB state as a function of μ and of the noise intensity. As the noise decreases, we observe a fast transition between SB- and W-dominated populations for $\mu = \mu_{\mathrm{crit}} \approx 1.005$, which corresponds to a first-order phase transition.

for the reasons explained before, this would require computational resources that are beyond what has been allocated for this study. We remark that this would allow for evaluating for each value of μ the difference between the value of the quasi-potential realised at the attractors and at the M state. The algorithm proposed in [84] can be very useful to reduce the computational burden.

We can say that for $\mu = \mu_{\mathrm{crit}}$ our system exhibits a behaviour that is reminiscent of a first-order phase transition for near-equilibrium statistical mechanical systems, like a liquid–gaseous transition. In our case, μ is the control parameter (corresponding to the temperature in the equilibrium case), the quasi-potential Φ is the equivalent of a thermodynamic potential, the scaling factor for the noise intensity σ is the equivalent of (square root of) the temperature, and the globally averaged surface temperature $\langle T_{\mathrm{S}} \rangle$ is the natural order parameter, e.g. density. The discontinuous change in the properties of the system for $\mu = \mu_{\mathrm{crit}}$ is associated to the change in the amount of absorbed and emitted radiation, as a result of the macroscopic change in the albedo of the planet due to the discontinuity in the position of the ice-line. We remark that choosing a different noise law would in general lead to a different value of μ_{crit}, as a result of the fact that the functional form of Φ would be different.

4.5. Relevance of the choice of the Itô convention for the noise

A reasonable question to ask concerns the extent to which our results are sensitive to the fact that we have chosen the Itô convention for the noise, which provides the starting point of the results discussed in sections 2.2 and 2.3. We argue that choosing other conventions would not alter essentially our findings because, to a first approximation, the stochastic forcing we have introduced can be treated as one corresponding to perturbing the system with additive noise of different strengths near the cold and W attractors, plus a transition region between the two attractors (which is evidently very sparsely populated by the system), where the effective intensity of the noise decreases with the globally averaged surface temperature $\langle T_S \rangle$ and the multiplicative nature of the noise is more evident.

In the phase space region near the cold attractor, we have that $1 - \alpha(\phi, T_S) \sim 0.4$, because the temperature T_S is extremely low and the planet is fully glaciated (or almost entirely so), so that $\alpha(\phi, T_S)$ is virtually constant, with $\alpha(\phi, T_S) \sim \alpha_{\min}$. Near the W attractor, the properties of the field $\alpha(\phi, T_S)$ are slightly more complex, because part of the planet is glaciated and part of it is ice-free. Nonetheless, to a first approximation, the ratio of the variance of the noise in the SB attractor vs W attractor is of the order $((1 - \alpha_{\min})/(1 - \alpha_W))^2 \sim 3$. Loosely speaking, the competing W and SB climate states have different statistical mechanical, microscopic—as well as thermodynamical, macroscopic—temperatures.

5. An alternative construction of the M states using stochastic perturbations

As discussed above, the construction of saddles for multistable systems is far from being a trivial task, and requires the use of the edge tracking algorithm introduced in [50, 51] and used also by the authors in [20, 45]. We wish to provide here a proof of concept of an alternative procedure for constructing the saddles—especially relevant when they are complex M states—without resorting to such an algorithm, but rather using only direct numerical simulations. The procedure discussed below might be useful when the edge tracking algorithm is hard to implement. As an example, this could be the case when the presence of similar time scales associated to the instability along and across the basin boundary might hinder an accurate computation of the saddles. Alternatively, it can be seen as a way to test the results obtained from the study of the deterministic dynamics.

The idea is to exploit the fact that, as discussed in section 2, under rather general conditions on the noise law, the saddle, in the weak-noise limit, acts as the gate for noise-induced transitions between the competing attractors. We then propose to proceed as follows. Let us consider the following SDEs:

$$\mathrm{d}\mathbf{x} = \mathbf{F}(\mathbf{x})\mathrm{d}t + \sigma \mathbf{s}_k(\mathbf{x})\mathrm{d}\mathbf{W}, \quad k = 1, \ldots, K. \tag{26}$$

We consider the possibility of perturbing the deterministic flow with K different noise laws, defined by the K functions $\mathbf{s}_k(\mathbf{x})$, each leading to a noise with a different covariance matrix $C_{ij,k}(\mathbf{x})$. We assume, for simplicity, that the deterministic system defined by $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x})$ is bistable. We choose K noise laws such that they obey the hypotheses discussed in sections 2.2 and 2.3.

In the weak-noise limit $\sigma \to 0$, the invariant measure of the kth SDE can be written as $\Pi_{\sigma,k} \sim \exp\left(-\frac{2\Phi_k(\mathbf{x})}{\sigma^2}\right)$. Clearly, for each noise law the quasi-potential $\Phi_k(\mathbf{x})$ is different. Nonetheless, as discussed above, in all cases $\Phi_k(\mathbf{x})$ has a local minimum (and is constant) on the support of the two deterministic attractors, and it is a saddle with constant value on the saddle separating the two attractors. We here have to assume that either the saddle is unique, or that all the quasi-potentials $\Phi_k(\mathbf{x})$ select the same saddle as the one with the lowest quasi-potential.

Additionally, for each of the K SDEs the drift flow is split differently between the gradient-like component and the rest (see equation (6)). Therefore, as suggested by equation (9), the instantonic paths are also different; yet, they connect the same attractors to the same saddle.

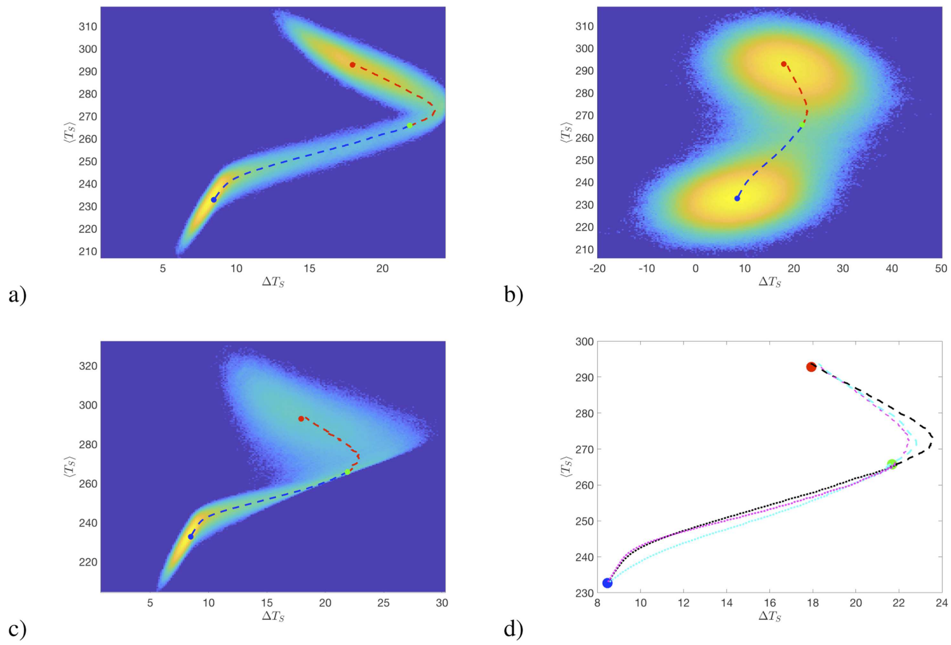

Figure 9. Projected measure in the $(\Delta T_{\mathrm{S}}, \langle T_{\mathrm{S}}\rangle)$ space (units of K on both axes) for selected stochastically perturbed simulations of the Ghil–Sellers models and related instantons. All results have been obtained with simulations lasting 1.6×10^6 model years. (a) Additive noise s_1: Logarithm of the pdf; W $\rightarrow$ SB instanton (red dashed line); SB $\rightarrow$ W instanton (blue dashed line); W attractor (red dot); SB attractor (blue dot); saddle (green dot). The results have been obtained with $\sigma = 0.4$. (b) Same as in (a), for additive noise s_2 (results obtained setting $\sigma = 2$. (c) Same as in (a), for the multiplicative noise used in the rest of the paper (results obtained setting $\sigma = 1.0$). (d) The instantons from (a) (black line), (b) (cyan line) and (c) (magenta line) are plotted together. Dashed (dotted) lines correspond to the W $\rightarrow$ SB (SB $\rightarrow$ W) instantons. They cross at the saddle (green dot) and at the W attractor (red dot) and at the SB attractor (blue dot).

Assuming that the attractors and the saddles are points or at least small sets given in a coarse-grained description of the phase space, we can identify the saddle as the only point in space where the instantons corresponding to all the K noise laws will intersect.

In order to show that this approach does indeed work, we investigate the properties of $K = 3$ variants of the Ghil–Sellers diffusive model we studied in [45], differing with respect to the law of the stochastic perturbation impacting the energy balance of the climate system. The Ghil–Sellers diffusive model can be written by removing the atmosphere-ocean coupling term in equation (24). In order to conform to equation (2) (note that we treat below the numerical discretization of a stochastically perturbed partial different equation), we express the three stochastically perturbed models as follows:

$$\frac{\partial T_{\mathrm{S}}(\phi,t)}{\partial t} = \mu \frac{S_0^*}{4}\frac{I(\phi)}{C(\phi)}(1-\alpha(\phi,T_{\mathrm{S}})) - \frac{O(T_{\mathrm{S}})}{C(\phi)} - \frac{D_\phi[T_{\mathrm{S}}]}{C(\phi)} + \sigma s_k(\phi,T_{\mathrm{S}})\frac{\mathrm{d}W}{\mathrm{d}t}, \quad k = 1,2,3, \tag{27}$$

where dW is the increment of a one-dimensional Wiener process, and we have:

$$s_1(\phi, T_S) = \mu \frac{S_0^*}{4} \frac{I(\phi)}{C(\phi)}, \tag{28}$$

$$s_2(\phi, T_S) = 1, \tag{29}$$

$$s_3(\phi, T_S) = \mu \frac{S_0^*}{4} \frac{I(\phi)}{C(\phi)}, (1 - \alpha(\phi, T_S)). \tag{30}$$

Specifically, we have that $s_1(\phi, T_S) = s_1(\phi)$ and $s_2(\phi, T_S) = s_2(\phi)$ correspond to additive noise laws, which feature different diffusion matrices. Instead, $s_3(\phi, T_S)$ is a multiplicative noise law as a result of the temperature-dependence of the albedo and is closely related to what has been studied in the rest of the paper; see equation (24). We construct for these three SDEs the invariant measure and, by stochastically averaging, we estimate the instantons connecting the attractors and the saddles. These sets are simple points. We make sure that the instantons are estimated using very weak noise amplitudes. Results are presented in figure 9, where we show the projections on the $(\Delta T_S, \langle T_S \rangle)$ space. Panels (a)–(c) show the invariant measures and the instantons constructed for the noise laws s_1 (using $\sigma = 0.4$), s_2 (using $\sigma = 2.0$), and s_3 (using $\sigma = 1.0$), respectively. Note that the instanton constructed with the multiplicative noise law looks qualitatively different from what is shown in figures 5 and 6, as a result of the lack of atmospheric motions in the simpler model discussed here. Panel (d) portrays the instantons constructed for the three noise laws. Indeed, we have a confirmation that all of them are different, as a result of the different noise laws of the three SDEs, and intersect at the attractors and at the saddle. The position of the saddle in the projected phase space can be identified through this geometric procedure, which is based exclusively on direct numerical simulations. Considering additional noise laws can be helpful in resolving possible geometrical degeneracies due to the use of projections. Projecting in more than two dimensions could also serve a similar scope and provide a better understanding of the alternative transition paths.

6. Conclusions

The goal of this paper has been the investigation of the properties of the noise-induced transitions across the multiple basins of attractions in an intermediate complexity climate model with $O(10^4)$ degrees of freedom, describing the coupled evolution of atmospheric (fast) and oceanic (slow) variables. The model features the co-existence of W and SB attractors for a fairly broad range of values of the solar irradiance. In a previous investigation, we had been able to construct the full phase portrait of the deterministic version of the climate model considered here, and had constructed, beside the attractors, the M states of the climate system in the region of bistability [20].

The stochastic forcing is introduced here as a random modulation of the incoming solar radiation, and leads to a nontrivial multiplicative noise law, because the radiative forcing is affected by the albedo of the surface, which, in turn, depends on the surface temperature. The noise, by allowing transitions between the deterministic basins of attraction, allows for establishing an (apparently) ergodic invariant measure of the system. The theory of SDEs indicates that for systems obeying the hypoellipticity condition, and for a suitable class of noise laws one can write fairly generally the invariant measures in terms of a large deviation law, where the rate function can be identified with the quasi-potential. We have clarified how to compute the quasi-potential from the drift and volatility fields of the SDE and explained its property of being a Lyapunov function. The quasi-potential has local minima on the deterministic attractors, and has a saddle behaviour at the M states. Additionally, in the weak-noise limit, transitions take place along special paths called instantons, which link the deterministic attractors and the M states. While, for a given deterministic dynamics, instantons corresponding to different noise laws follow different paths, they all link the same deterministic attractors to the same M states. We have shown how this property can be exploited to geometrically construct M states directly from direct numerical simulations of stochastic systems. Indeed, while the edge tracking algorithm applied to the deterministic system is *a priori* the preferred choice for finding M states, it might become nontrivial to implement in complex numerical models where the intermediate states constructed by bisection might correspond to regions where the model is numerically unstable, possibly because the realised physical fields are extremely exotic or non-realisable.

We have studied in detail the noise-induced transitions between the deterministic basins of attraction in the range of multistability, extending the results presented in a short communication [56]. We have shown that by studying how the average escape time depends on the intensity of the noise it is possible to estimate the difference between the value of the quasi-potential at the M state and at the attractor that the trajectories are escaping from. The estimates of the instantons are shown to become more precise as weaker noise is used in the simulations. The instanton, in a case of special interest where the M state was shown to undergo a symmetry-break process, selects as optimal point of passage between the SB and the W climate the transient M state instead of the asymptotic one, possibly as a result of the finite amplitude of the noise.

Finally, by studying how the populations of the W and SB climate change as a function of the intensity of the noise, and using the large deviation law for the measure predicted by the theory, we find an estimate of a critical value of $\mu = \mu_{\rm crit} \approx 1.005$, such that for $\mu \geqslant \mu_{\rm crit}$ the zero-noise limit of the invariant measure is supported on the W deterministic attractor, while for $\mu \leqslant \mu_{\rm crit}$ the weak-noise limit of the invariant measure is supported on the SB attractor. The asymptotic state corresponds to the attractor featuring the lowest value of the quasi-potential.

These results obtained here indicate that, as soon as noise—in some form—is added to the system, multistability is factually lost in the weak-noise limit, as the noise law is responsible for selecting, for each value of the control parameter (here μ), a specific asymptotic state. Changing the value of the control parameter, one will find one or more abrupt transitions in the statistical properties of the system (here realised at $\mu_{\rm crit}$), i.e., in other terms discontinuities in the response of the system to changes in the control parameter. What happens in our model at $\mu = \mu_{\rm crit}$ is mathematically analogous to a first-order phase transition occurring in a near-equilibrium statistical mechanical system. We remark that, since the escape time away from either attractor grows exponentially with the inverse of the parameter controlling the variance of the noise, an individual trajectory might be trapped for very long time in a metastable state.

In collaboration with C Kilic (Bern) and F Lunkeit (Hamburg) the authors have started some preliminary simulations where stochastic forcing is added to PLASIM [97], a much more complex climate model than the one used in this study (yet missing some essential ocean dynamical processes). As shown in figure 10, the first results we have obtained are encouraging in indicating that the findings of this paper might be relevant for more realistic model configurations. For the future, we aim to obtain detailed information on the large scale properties of the flow configurations leading to the noise-induced transitions, taking the thermodynamic lens we originally explored in [9]. An important question of specific relevance for paleoclimatic and planetary science studies is to understand whether the third climatic state found in [20] for $\mu = 1.045$ is recovered also for more realistic model configurations. Additionally, the complexity of the dynamical landscape of the climate system discussed in [21] suggests the

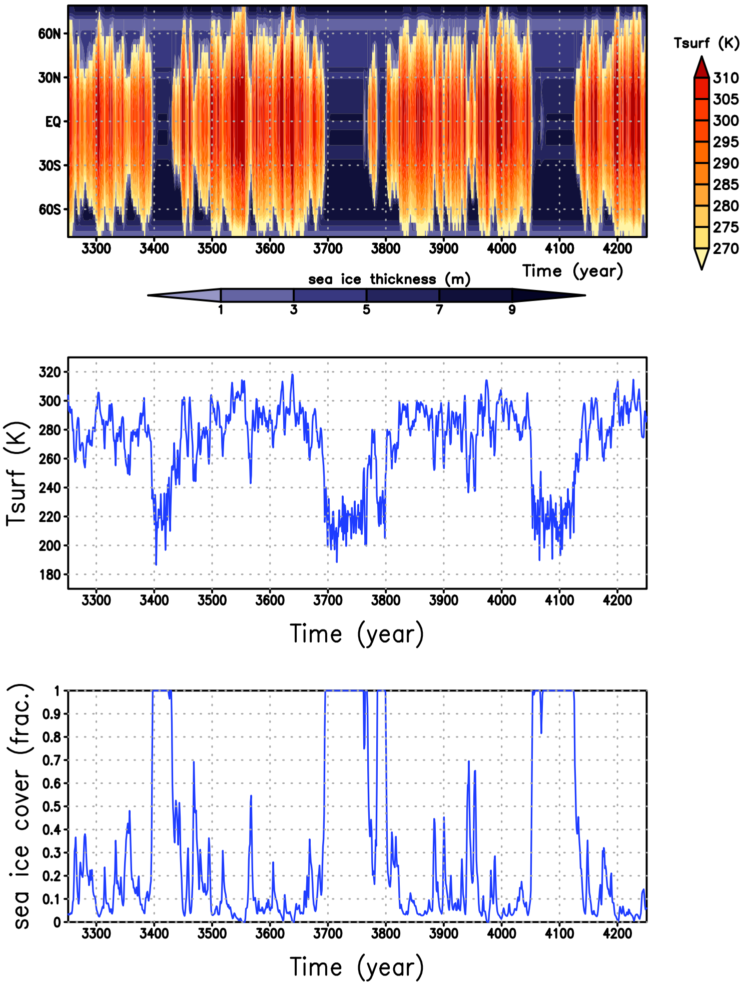

Figure 10. Example of noise-induced transitions between SB and W climate state for present-day solar constant ($\sigma_{100\mathrm{y}} = 2\%$) for the climate model PLASIM [97] in a simulation of 10^3 y (reproduced with permission from F Lunkeit). From top to bottom: zonal averages of the temperature field and of the sea ice cover; globally averaged surface temperature; fraction of ice-covered ocean. The characteristic escape times is comparable with what is obtained in the simpler model PUMA-GS used in this work.

existence of a possibly topologically non-trivial network of transition paths between the many competing attractors, each crossing an M state. Maybe the itinerancy between possibly many competing attractors might be a way to explaining the ultralow frequency variability of the climate. The computational needs of a naive approach to these issues seem prohibitive, so that one should definitely take advantage of rare events algorithms to construct instantonic trajectories [94–96].

The approach presented here, based upon combining the knowledge of the dynamical landscape of a multistable deterministic dynamical system with the analysis of the impacts of stochastic perturbations, seems of more general interested than the specific problem we have studied. Along these lines, in appendix A we speculate on the possible relevance of M states in the context of the theory of biological evolution and of synthetic models of evolution. Specifically, the idea is that the presence of qualitatively diverse historical paths of evolution might result from the presence of M states in suitable defined dynamical landscapes, exactly because M states hinder predictability of the second kind in the sense of Lorenz.

In the specific case of the geosciences, we believe that it can be key for addressing the challenge of understanding tipping points in the Earth system [69] as well as providing insights to a large class of multistable systems [68]. We will dedicate future efforts exactly to exploring these research lines, in particular looking at the Atlantic meridional overturning circulation, an element of the global ocean circulation that is well-know to have more than one competing *modes of operation*. The occurrence of one or of the other state has important implications for the global climate, and rather dramatic ones for the regional climate of the north Atlantic sector; see [98] for a review of this topic. Multistability has been recently reported in the von Karman turbulent flow in [99]: the methods proposed in this paper could elucidate paths and mechanisms underlying the transitions between the asymptotic states.

Acknowledgments

VL and TB acknowledge the support received by the EU Horizon2020 projects Blue-Action (Grant No. 727852) and CRESCENDO (Grant No. 641816). VL acknowledges the support of the EU Horizon 2020 project TiPES (Grant No. 820970), of the DFG SFB/Transregio project TRR181, and of the Royal Society (Grant No. IEC/R2/170001). TB acknowledges the support received from the Institute for Basic Science (IBS), Republic of Korea, under IBS-R028-D1. The authors wish to thank T Tél for inspiring exchanges. VL wishes to thank S Brusatte for suggesting some fascinating literature on evolutionary biology; M Ghil and J Yorke for great encouragement in the course of several cheerful discussions; T Kuna, T Lelievre, X-M Li, G Pavliotis, and N Zagli for providing insights to stochastic dynamics. This paper has been partly prepared during the Advanced Workshop on *Nonequilibrium Systems in Physics, Geosciences, and Life Sciences* held at ICTP, Trieste, on May 15-24 2018 and during the Scientific Programme on the *Mathematics of Climate and the Environment* held at the Institut H. Poincaré, Paris, on September–December 2019. The authors wish to thank F Lunkeit for kindly sharing figure 10. The authors wish to dedicate this paper to the memory of Bruno Eckhardt (1960–2019).

Appendix A. An interdisciplinary outlook: evolutionary biology

The concepts discussed in this paper might provide a conceptual and mathematical framework of possible interest for thinking at evolutive processes in biology and for constructing synthetic models of evolution. In the speculative discussion below, specifically, we want to highlight the role of M states in making possible the existence of multiple possible yet vastly different paths of historical development of biological systems.

In the influential book discussing the Cambrian fossils from the Burgess shale and their importance for explaining the mechanisms of evolution, Gould [67] presents some ideas on the scientific methodology inherent to historical sciences, and, specifically, to evolutionary biology. He clarifies that historical scientific explanations take the form of a narrative, whereby subsequent phenomena follow in a specific order[15]: the time-ordering and causal links between such phenomena can be discovered and convincingly explained when multiple, independent

[15] The related *storyline* approach is being currently proposed for studying weather and climate phenomena [100].

sources provide indication for the same historical pattern of change. Gould argues that this method is fundamentally different from the classic—which he calls *stereotypical*—scientific method *á la Galileo* whereby one specific experiment can be repeated in idealised conditions, and the final outcomes of the experiment can be predicted using the fundamental laws of nature. The latter is the—indeed too simplified—version he gives of how *hard sciences* work. An obvious difference between the two methods comes from the fact that certain experiments—like the evolution of the climate and of the biosphere of a specific planet—cannot be repeated. Another difference comes from the fact—Gould argues—that historical processes are dominated by *contingency*: while we observe a specific path of historical development, many others of such paths are indeed compatible with the laws of nature, and could have been realised had the system been forced in a slightly different way in the past. We are able to understand the mechanisms of historical development, but not to predict accurately the specific path. Gould argues that the evolution in our planet could have led to fundamentally different forms of life, had the contingencies been different in the distant past. The existence of our own species is the—*a priori* very unlikely—result of such contingencies. In general, in systems dominated by contingency, *if we could run again the movie the outcome would be vastly different*. Gould's views on evolution have been criticised by other authors, as Conway Morris, proposing that *convergence*, rather than contingency, is the main mechanism of evolution, so that evolution is seen as a—mostly—deterministic path of change, of which nowadays we see the unavoidable outcome [101]; see the debate in [102]. Some authors have proposed that both mechanisms are indeed in action, yet dominant at different scales of diversification of the organisms [103].

We argue that Gould's view can be put in the context of the mathematical framework discussed in this paper. Stochastically forced complex systems evolve in a phase space where an individual trajectory corresponds to a historical realisation of the system. The historical realisations, even starting from the same initial condition, can be vastly—even qualitatively—different if the deterministic dynamics supports the existence of multiple attractors, because different realisations can be trapped for very long times in very distant regions of the phase space. As discussed here, stochastic forcing allows for the system to jump across the various basins of attraction, and the mechanism defining the evolution of the trajectory are defined by the differential equations. The predictability of the system, both of the first and second kind in the sense of Lorenz, is finite but non-zero. Finally, we can interpret the M states as the true agents of the contingency discussed by Gould. It is not the presence of a forcing, however strong, that changes radically the future history of the system, but rather the existence of special regions of the phase space—near the M states—where even small perturbations can force two nearby trajectories towards qualitatively different future histories. What Conway Morris proposes is associated with a scenario where M states are either absent—so that the system is not multistable—or the noise is so weak that the probability of getting close to an M state is exceedingly low, despite the presence of stochastic perturbations, the system will be (almost) always quite close to a specific deterministic attractor, whose basin of attraction the initial condition belongs to. In other terms even *if we could run again the movie*, i.e., run another simulation, the outcome would be very similar.

The interpretation given above receives some support from recent results obtained on synthetic models of evolution. Using the so-called "tangled nature" model, which is conjectured to be a prototypical example of evolution, evolutive processes are interpreted as orbits of stochastic systems in a complex dynamical landscape featuring two or more competing metastable states [104, 105]. This viewpoint, which has been openly inspired by Gould's ideas, is in close correspondence with what has been discussed in this paper. While in these works the language and methodology are eminently of statistical mechanical nature and they aim at detecting and studying the metastables states, the viewpoint proposed in this paper has a stronger emphasis on understanding the transitions paths between such competing attractors through the M states.

ORCID iDs

Valerio Lucarini https://orcid.org/0000-0001-9392-1471
Tamás Bódai https://orcid.org/0000-0002-3049-107X

References

[1] Budyko M I 1969 The effect of solar radiation variations on the climate of the Earth *Tellus* **21** 611–9

[2] Sellers W D 1969 A global climatic model based on the energy balance of the Earth atmosphere system *J. Appl. Meteorol.* **8** 392–400

[3] Ghil M 1976 Climate stability for a sellers-type model *J. Atmos. Sci.* **33** 3–20

[4] Kirschvink J L 1992 Late proterozoic low-latitude global glaciation: the snowball Earth *The Proterozoic Biosphere: A Multidisciplinary Study* ed J W Schopf and C Klein (Cambridge: Cambridge University Press) pp 91–2

[5] Hoffman P F and Schrag D P 2002 The snowball Earth hypothesis: testing the limits of global change *Terra Nova* **14** 129–55

[6] Pierrehumbert R T, Abbot D S, Voigt A and Koll D 2011 Climate of the Neoproterozoic *Annu. Rev. Earth Planet. Sci.* **39** 417–60

[7] Hyde W T, Crowley T J, Baum S K and Peltier R W 2000 Neoproterozoic ‘snowball earth’ simulations with a coupled climate/ice-sheet model *Nature* **405** 425–9

[8] Voigt A and Marotzke J 2010 The transition from the present-day climate to a modern snowball Earth *Clim. Dyn.* **35** 887–905

[9] Lucarini V, Fraedrich K and Lunkeit F 2010 Thermodynamic analysis of snowball Earth hysteresis experiment: efficiency, entropy production and irreversibility *Q. J. R. Meteorol. Soc.* **136** 2–11

[10] Boschi R, Lucarini V and Pascale S 2013 Bistability of the climate around the habitable zone: a thermodynamic investigation *Icarus* **227** 1724–42

[11] Crowley T J, Hyde W J and Richard Peltier W 2001 CO2 levels required for deglaciation of a ‘near-snowball’ Earth *Geophys. Res. Lett.* **28** 283–6

[12] Linsenmeier M, Pascale S and Lucarini V 2015 Climate of Earth-like planets with high obliquity and eccentric orbits: implications for habitability conditions *Planet. Space Sci.* **105** 43–59

[13] Kilic C, Raible C C and Stocker T F 2017 Multiple climate states of habitable exoplanets: the role of obliquity and irradiance *Astrophys. J.* **844** 147

[14] Kilic C, Frank L, Raible C C and Stocker T F 2018 Stable equatorial ice belts at high obliquity in a coupled atmosphere–ocean model *Astrophys. J.* **864** 106

[15] Lucarini V, Pascale S, Boschi R, Kirk E and Iro N 2013 Habitability and multistablility in Earth-like plantets *Astron. Nachr.* **334** 576–88

[16] Abbot D S, Bloch-Johnson J, Checlair J, Farahat N X, Graham R J, Plotkin D, Popovic P and Spaulding-Astudillo F 2018 Decrease in hysteresis of planetary climate for planets with long solar days *Astrophys. J.* **854** 3

[17] Checlair J, Menou K and Abbot D S 2017 No snowball on habitable tidally locked planets *Astrophys. J.* **845** 132

[18] Lewis J P, Weaver A J and Eby M 2007 Snowball versus slushball Earth: dynamic versus nondynamic sea ice? *J. Geophys. Res.* **112** C11014

[19] Abbot D S, Voigt A and Koll D 2011 The Jormungand global climate state and implications for Neoproterozoic glaciations *J. Geophys. Res.* **116** D18103

[20] Lucarini V and Bódai T 2017 Edge states in the climate system: exploring global instabilities and critical transitions *Nonlinearity* **30** R32

[21] Brunetti M, Kasparian J and Vérard C 2019 Co-existing climate attractors in a coupled aquaplanet *Clim. Dyn.* **53** 6293–308

[22] Gómez-Leal I, Kaltenegger L, Lucarini V and Lunkeit F 2018 Climate sensitivity to carbon dioxide and the moist greenhouse threshold of Earth-like planets under an increasing solar forcing *Astrophys. J.* **869** 129

[23] Gómez-Leal I, Kaltenegger L, Lucarini V and Lunkeit F 2019 Climate sensitivity to ozone and its relevance on the habitability of Earth-like planets *Icarus* **321** 608–18

[24] Kasting J F, Whitmire D P and Reynolds R T 1993 Habitable zones around main sequence stars *Icarus* **101** 108–28

[25] Baladi V 2000 *Positive Transfer Operators and Decay of Correlations* (Singapore: World Scientific)

[26] Pollicott M 1985 On the rate of mixing of Axiom A flows *Invent. Math.* **81** 413–26

[27] Ruelle D 1986 Resonances of chaotic dynamical systems *Phys. Rev. Lett.* **56** 405–7

[28] Tantet A, Lucarini V, Lunkeit F and Dijkstra H A 2018 Crisis of the chaotic attractor of a climate model: a transfer operator approach *Nonlinearity* **31** 2221

[29] Shiino M 1987 Dynamical behavior of stochastic systems of infinitely many coupled nonlinear oscillators exhibiting phase transitions of mean-field type: H theorem on asymptotic approach to equilibrium and critical slowing down of order-parameter fluctuations *Phys. Rev.* A **36** 2393–412

[30] Pavliotis G A 2014 Stochastic processes and applications: diffusion processes, the Fokker-Planck and Langevin equations *Texts in Applied Mathematics* (New York: Springer)

[31] Ragone F, Lucarini V and Lunkeit F 2016 A new framework for climate sensitivity and prediction: a modelling perspective *Clim. Dyn.* **46** 1459–71

[32] Lucarini V, Ragone F and Lunkeit F 2017 Predicting climate change using response theory: global averages and spatial patterns *J. Stat. Phys.* **166** 1036–64

[33] Lembo V, Lucarini V and Ragone F 2020 Beyond forcing scenarios: predicting climate change through response operators in a coupled general circulation model *Sci. Rep.* **10** 8668

[34] Ruelle D 2009 A review of linear response theory for general differentiable dynamical systems *Nonlinearity* **22** 855–70

[35] Ghil M, Chekroun M and Simonnet E 2008 Climate dynamics and fluid mechanics: natural variability and related uncertainties *Physica* D **237** 2111–26

[36] Chekroun M, Simonnet E and Ghil M 2011 Stochastic climate dynamics: random attractors and time-dependent invariant measures *Physica* D **240** 1685–700

[37] Carvalho A N, Langa J and Robinson J C 2013 The pullback attractor *Attractors for Infinite-Dimensional Non-Autonomous Dynamical Systems* Applied Mathematical Sciences vol 182 (New York: Springer) pp 3–22

[38] Romeiras F J, Grebogi C and Ott E 1990 Multifractal properties of snapshot attractors of random maps *Phys. Rev.* A **41** 784

[39] Drótos G, Bódai T and Tél T 2015 Probabilistic concepts in a changing climate: a snapshot attractor picture *J. Clim.* **28** 3275–88

[40] Dembo A and Deuschel J-D 2010 Markovian perturbation, response and fluctuation dissipation theorem *Ann. inst. Henri Poincare* B **46** 822–52

[41] Assaraf R, Jourdain B, Lelièvre T and Roux R 2018 Computation of sensitivities for the invariant measure of a parameter dependent diffusion *Stochast. PDE: Anal. Comput.* **6** 125–83

[42] Lucarini V 2016 Response operators for markov processes in a finite state space: radius of convergence and link to the response theory for axiom a systems *J. Stat. Phys.* **162** 312–33

[43] Chekroun M, Neelin J D, Kondrashov D, McWilliams J C and Ghil M 2014 Rough parameter dependence in climate models and the role of Ruelle-Pollicott resonances *Proc. Natl Acad. Sci. USA* **111** 1684–90

[44] Ghil M and Lucarini V 2020 The physics of climate variability and climate change *Rev. Mod. Phys.* accepted

[45] Bódai T, Lucarini V, Lunkeit F and Boschi R 2014 Global instability in the Ghil–sellers model *Clim. Dyn.* **44** 3361–81

[46] Grebogi C, Ott E and Yorke J A 1983 Fractal basin boundaries, long-lived chaotic transients, and unstable-unstable pair bifurcation *Phys. Rev. Lett.* **50** 935–8

[47] Robert C, Alligood K T, Ott E and Yorke J A 2000 Explosions of chaotic sets *Physica* D **144** 44–61

[48] Ott E 2002 *Chaos in Dynamical Systems* (Cambridge: Cambridge University Press)

[49] Lai Y-C and Tél T 2011 *Transient Chaos* (New York: Springer)

[50] Skufca J D, Yorke J A and Eckhardt B 2006 Edge of chaos in a parallel shear flow *Phys. Rev. Lett.* **96** 174101

[51] Schneider T M, Eckhardt B and Yorke J A 2007 Turbulence transition and the edge of chaos in pipe flow *Phys. Rev. Lett.* **99** 034502

[52] Barton D A W and Sieber J 2013 Systematic experimental exploration of bifurcations with noninvasive control *Phys. Rev.* E **87** 052916

[53] Sieber J, Omel'chenko O E and Wolfrum M 2014 Controlling unstable chaos: stabilizing chimera states by feedback *Phys. Rev. Lett.* **112** 054102

[54] Abrams D M and Strogatz S H 2004 Chimera states for coupled oscillators *Phys. Rev. Lett.* **93** 174102

[55] Omel'chenko O E 2018 The mathematics behind chimera states *Nonlinearity* **31** R121

[56] Lucarini V and Bódai T 2019 Transitions across melancholia states in a climate model: reconciling the deterministic and stochastic points of view *Phys. Rev. Lett.* **122** 158701

[57] Bell D R 2004 *Stochastic Differential Equations and Hypoelliptic Operators* (Boston, MA: Birkhäuser) pp 9–42

[58] Hanggi P 1986 Escape from a metastable state *J. Stat. Phys.* **42** 105–48

[59] Kautz R L 1987 Activation energy for thermally induced escape from a basin of attraction *Phys. Lett.* A **125** 315–9

[60] Grassberger P 1989 Noise-induced escape from attractors *J. Phys. A: Math. Gen.* **22** 3283

[61] Freidlin M I and Wentzell A D 1984 *Random Perturbations of Dynamical Systems* (New York: Springer)

[62] Graham R, Hamm A and Tél T 1991 Nonequilibrium potentials for dynamical systems with fractal attractors or repellers *Phys. Rev. Lett.* **66** 3089–92

[63] Hamm A, Tél T and Graham R 1994 Noise-induced attractor explosions near tangent bifurcations *Phys. Lett.* A **185** 313–20

[64] Kraut S and Feudel U 2002 Multistability, noise, and attractor hopping: the crucial role of chaotic saddles *Phys. Rev.* E **66** 015207

[65] Beri S, Mannella R, Luchinsky D G, Silchenko A N and McClintock P V E 2005 Solution of the boundary value problem for optimal escape in continuous stochastic systems and maps *Phys. Rev.* E **72** 036131

[66] Bouchet F and Reygner J 2016 Generalisation of the Eyring–Kramers transition rate formula to irreversible diffusion processes *Ann. Henri Poincaré* **17** 3499–532

[67] Gould S J 1989 *Wonderful Life: The Burgess Shale and the Nature of History* (New York: W. W. Norton)

[68] Feudel U, Pisarchik A N and Showalter K 2018 Multistability and tipping: from mathematics and physics to climate and brain: minireview and preface to the focus issue *Chaos* **28** 033501

[69] Lenton T M, Held H, Kriegler E, Hall J W, Lucht W, Rahmstorf S and Schellnhuber H J 2008 Tipping elements in the Earth's climate system *Proc. Natl Acad. Sci.* **105** 1786–93

[70] Vollmer J, Schneider T M and Eckhardt B 2009 Basin boundary, edge of chaos and edge state in a two-dimensional model *New J. Phys.* **11** 013040

[71] Bódai T and Lucarini V 2020 Rough basin boundaries in high dimension: can we classify them experimentally? arXiv:2001.08871

[72] Lorenz E N 1975 Climate predictability *The Physical Basis of Climate and Climate Modelling* WMO GARP Publ. Series 16 (Geneva: WMO) pp 132–6

[73] Hairer M 2011 On Malliavin's proof of Hörmander's theorem *Bull. Sci. Math.* **135** 650–66 Special issue in memory of Paul Malliavin

[74] Gaspard P 2002 Trace formula for noisy flows *J. Stat. Phys.* **106** 57–96

[75] Bouchet F, Gawedzki K and Nardini C 2016 Perturbative calculation of quasi-potential in non-equilibrium diffusions: a mean-field example *J. Stat. Phys.* **163** 1157–210

[76] Ao P 2004 Potential in stochastic differential equations: novel construction *J. Phys. A: Math. Gen.* **37** L25–30

[77] Yin L and Ao P 2006 Existence and construction of dynamical potential in nonequilibrium processes without detailed balance *J. Phys. A: Math. Gen.* **39** 8593–601

[78] Zhou J X, Aliyu M D S, Aurell E and Huang S 2012 Quasi-potential landscape in complex multi-stable systems *J. R. Soc. Interface* **9** 3539–53

[79] Brackston R D, Wynn A and Stumpf M P H 2018 Construction of quasipotentials for stochastic dynamical systems: an optimization approach *Phys. Rev.* E **98** 022136

[80] Tang Y, Yuan R, Wang G, Zhu X and Ao P 2017 Potential landscape of high dimensional nonlinear stochastic dynamics with large noise *Sci. Rep.* **7** 15762

[81] Graham R 1987 Macroscopic potentials, bifurcations and noise in dissipative systems *Fluctuations and Stochastic Phenomena in Condensed Matter* ed L Garrido (Berlin: Springer) pp 1–34

[82] Graham R and Tél T 1986 Nonequilibrium potential for coexisting attractors *Phys. Rev.* A **33** 1322–37

[83] Bovier A, Eckhoff M, Gayrard V and Klein M 2004 Metastability in reversible diffusion processes I. Sharp asymptotics for capacities and exit times *J. Eur. Math. Soc.* **6** 399–424

[84] Bódai T 2020 An efficient algorithm to estimate the potential barrier height from noise-induced escape time data *J. Stat. Phys.* accepted (https://doi.org/10.1007/s10955-020-02574-4)

[85] Lelièvre T 2015 Accelerated dynamics: mathematical foundations and algorithmic improvements *Eur. Phys. J. Spec. Top.* **224** 2429–44

[86] Di Gesù G, Lelièvre T, Le Peutrec D and Nectoux B 2019 Sharp asymptotics of the first exit point density *Ann. PDE* **5** 5

[87] Lucarini V 2019 Stochastic resonance for nonequilibrium systems *Phys. Rev.* E **100** 062124

[88] Frisius T, Lunkeit F, Fraedrich K and James I N 1998 Storm-track organization and variability in a simplified atmospheric global circulation model *Q. J. R. Meteorol. Soc.* **124** 1019–43

[89] Saltzman B 2001 *Dynamical Paleoclimatology* (New York: Academic)

[90] Peixoto J P and Oort A H 1992 *Physics of Climate* (New York: American Institute of Physics)

[91] Lucarini V, Blender R, Herbert C, Ragone F, Pascale S and Wouters J 2014 Mathematical and physical ideas for climate science *Rev. Geophys.* **52** 1–51

[92] Holton J R 2004 *An Introduction to Dynamic Meteorology* 4th edn International Geophysics Series (Burlington, MA: Elsevier)

[93] Zinn-Justin J 1996 *Quantum Field Theory and Critical Phenomena* (Oxford: Oxford University Press)

[94] Rubino G and Tuffin B 2009 *Rare Event Simulation Using Monte Carlo Methods* (New York: Wiley)

[95] Ragone F, Wouters J and Bouchet F 2018 Computation of extreme heat waves in climate models using a large deviation algorithm *Proc. Natl Acad. Sci.* **115** 24–9

[96] Grafke T, Grauer R and Tobias S 2013 Instanton filtering for the stochastic Burgers equation *J. Phys. A: Math. Theor.* **46** 062002

[97] Fraedrich K, Jansen H, Kirk E, Luksch U and Lunkeit F 2005 The Planet Simulator: towards a user friendly model *Meteorol. Z.* **14** 299–304

[98] Kuhlbrodt T, Griesel A, Montoya M, Levermann A, Hofmann M and Rahmstorf S 2007 On the driving processes of the Atlantic meridional overturning circulation *Rev. Geophys.* **45** RG2001

[99] Faranda D, Sato Y, Saint-Michel B, Wiertel C, Padilla V, Dubrulle B and Daviaud F 2017 Stochastic chaos in a turbulent swirling flow *Phys. Rev. Lett.* **119** 014502

[100] Shepherd T G *et al* 2018 Storylines: an alternative approach to representing uncertainty in physical aspects of climate change *Clim. Change* **151** 555–71

[101] Morris S C 1998 *The Crucible of Creation: The Burgess Shale and the Rise of Animals* (Oxford: Oxford University Press)

[102] Gould S J and Morriss S C 1999 Showdown on the Burgess shale *Nat. Hist. Mag.* **107** 48–55

[103] Losos J B 2017 *Improbable Destinies: Fate, Chance, and the Future of evolution* (New York: Riverhead Books)

[104] Jones D, Jensen H J and Sibani P 2010 Tempo and mode of evolution in the tangled nature model *Phys. Rev.* E **82** 036121

[105] Jensen H J 2018 Tangled nature: a model of emergent structure and temporal mode among co-evolving agents *Eur. J. Phys.* **40** 014005