Exploring sudden stratospheric warmings with transition path theory

JUSTIN FINKEL*

Committee on Computational and Applied Mathematics, University of Chicago

ROBERT J. WEBBER

Department of Computing and Mathematical Sciences, California Institute of Technology

EDWIN P. GERBER

Courant Institute of Mathematical Sciences, New York University

DORIAN S. ABBOT

Department of the Geophysical Sciences, University of Chicago

JONATHAN WEARE

Courant Institute of Mathematical Sciences, New York University

ABSTRACT

Transition Path Theory (TPT) is used for a comprehensive analysis of sudden stratospheric warming (SSW) events in a highly idealized wave-mean flow interaction system with stochastic forcing. TPT is a statistical mechanics framework that explicitly considers rare events as an ensemble, and provides relationships between short-term forecasting and long-term climatology. We use the probability current, a central TPT quantity, to build a picture of critical altitude-dependent interactions between waves and the mean flow that fuel SSW events, both average behavior and variability across the SSW ensemble. We find that the rapid deceleration of zonal wind tends to be preceded by a gradual, halting decay in wind strength and a steady increase in meridional heat flux, which conspire to precondition the vortex for collapse. The ensemble-level description allows us to identify the signal of an oncoming SSW emerging from background variability during preconditioning, well before the sudden collapse. To circumvent the costly approach of extensive direct simulation of the full rare event ensemble, we implement a highly parallel computational method that launches a large collection of short simulations from many initial conditions, estimating long-timescale rare event statistics from short-term tendencies.

1. Introduction

Extreme weather events, by definition, are exceptional and occupy the fringes of the atmosphere's behavior distribution. Nevertheless, extreme events play an important role in atmospheric circulation. Large storms and changes in circulation are responsible for rapid movement of heat and moisture through the atmosphere. From a human perspective, weather is inconsequential when it follows mean behavior; it is the anomalies that challenge society (Lesk et al. 2016; Kron et al. 2019). Extreme weather is taking an increasing toll on ecosystems, economies, and human life, due to both a changing climate and increasing reliance on weather-susceptible infrastructure (e.g., Mann et al. 2017; Frame et al. 2020).

Significant efforts have gone toward estimating rare event probabilities and forecasting them with as much lead time as possible (e.g., Stephenson et al. 2008; Kim et al. 2019; Vitart and Robertson 2018). Earth system models are growing ever more powerful, and there is increasing interest in measuring their fidelity on extreme events, beyond just mean behavior (Hu et al. 2019). Capturing extremes is arguably the most important task of climate modeling, and part of our goal here is to motivate a set of quantities that a good model should reproduce.

Transition path theory (TPT), introduced in E and Vanden-Eijnden (2006), is a statistical mechanics framework to describe rare event dynamics. TPT has been applied within numerous studies of conformational change in biomolecules (e.g., Noé et al. 2009a; Meng et al. 2016; Liu et al. 2019; Thiede et al. 2019; Strahan et al. 2021), but has only recently been applied to geophysical dynamics. Miron et al. (2021) used TPT to map out garbage transport paths across the two-dimensional ocean, and Finkel et al. (2020) used TPT to understand rare stratospheric transitions in a highly reduced (just three variables) model of sudden stratospheric warming (SSW) events by Ruzmaikin et al.

*Corresponding author: Justin Finkel, jfinkel@uchicago.edu

(2003) and Birner and Williams (2008). Here, we explore a stochastically forced version of the classic Holton and Mass (1976) model, one of the first models to capture the key elements of a SSW. In the language of TPT, a SSW event is a trajectory that begins in a climatologically "normal" state (a strong polar vortex) and ends in an "extreme" state (a sudden warming, where the vortex has been broken down).

This paper complements our recent analysis of forecasting and predictability in the Holton-Mass model (Finkel et al. 2021), where we computed key forecasting functions—the *forward committor* and *lead time*—that give the probability of SSW and its expected arrival time, as a function of initial conditions. The TPT analysis we undertake here is related to the forecasting problem, but furthermore addresses the event's mechanism all the way from start to finish, not just forward in time from a fixed initial condition. Crucially, TPT distinguishes between the *onset* of an atmospheric disturbance (in our case study, a breakdown of the polar vortex from strong to weak) and the *persistence* of that disturbance (the "vacillation cycles" of an already weakened jet; Holton and Mass (1976)). In this paper, we use TPT to connect short-term weather forecast statistics, encoded by the committor and lead time, to the long-term climatology of SSW events, including their frequency, duration, and the distribution of pathways encoded by the *probability current*: the average tendency of the system conditioned on the occurrence of an SSW. By visualizing the probability current, we quantitatively assess the interaction between wave disturbances and zonal wind anomalies, and the extent to which they are uniquely associated with an SSW. TPT gives information about the *variability* of these processes, not just their mean behavior. In particular, we will show differences in the variability between successive stages of a SSW event. The preconditioning of the polar vortex manifests as a steady, predictable weakening of the lower-level zonal wind. The latter stage is an abrupt burst of heat flux and collapse of zonal wind that is much more variable in its timing and intensity. These are only a few deliverables of TPT, which can be adapted to probe many other weather phenomena.

Along with the TPT framework, we also advance an alternative computational strategy to direct numerical simulation (DNS), in which a model is integrated for a long time to produce many extreme events. In this paper, as in Finkel et al. (2021), we instead simulate many, short trajectories in parallel, and afterward combine information from all of them to compute rare event statistics without ever observing a complete event. (We use "DNS" to mean a single-threaded integration of a model, as opposed to a parallel integration from many initial conditions. This departs from the computational fluid dynamics usage, where it means "without subgrid closure".) While the fundamental strategy is the same as in Finkel et al. (2021), a full TPT analysis additionally requires *backward-in-time* forecasts to recover steady-state statistics from short-trajectories. The particular approach we use was introduced in Thiede et al. (2019) and Strahan et al. (2021) and extends work in the biophysics community over the last decade on approaches to analyze long timescale phenomena using short simulated trajectories (e.g., Jayachandran et al. 2006; Chodera and Noé 2014, and references therein). In particular, Noé et al. (2009b) combine an approach using short simulated trajectories similar to the one employed in this paper with TPT to study a protein folding event.

This paper is organized as follows. In section 2 we briefly summarize the dynamical model under study. In section 3, we visualize the evolution of SSW events through the probability current, and compare to the minimum action method. The resulting physical insight will motivate the more technical section 4, where we outline the computational approach, and the more thorough supplementary document. We assess future possibilities and conclude in section 5.

2. Model description

We use exactly the same prototype model for SSW events as analyzed in Finkel et al. (2021). We review the key features of the model here, but direct the reader to section 2b of Finkel et al. (2021) for more details.

Holton and Mass (1976) developed a minimal model for the variability of the winter stratospheric polar vortex, capturing the wave-mean flow interactions behind sudden stratospheric warming events. The model's prognostic variables consist of a zonally averaged zonal wind $\overline{u}(y,z,t)$ and a perturbation geostrophic streamfunction $\psi'(x,y,z,t)$ on a β-plane channel with a central latitude of $\theta = 60°$N and a meridional extent of 60°N. $\overline{u}$ and ψ' are projected onto a single zonal wavenumber $k = 2/(a\cos\theta)$ and a meridional wavenumber $\ell = 3/a$:

$$\overline{u}(y,z,t) = U(z,t)\sin(\ell y) \tag{1}$$

$$\psi'(x,y,z,t) = \mathrm{Re}\{\Psi(z,t)e^{ikx}\}e^{z/2H}\sin(\ell y), \tag{2}$$

where $a = 6370$ km $\approx$ the radius of Earth, and $H = 7$ km is the scale height. U (the mean flow) and Ψ (a complex-valued wave amplitude) evolve according to the projected primitive equations and the linearized quasi-geostrophic potential vorticity (QGPV) equation. The notation follows Christiansen (2000).

We use the same constant parameters and boundary conditions as Finkel et al. (2021), which give rise to two stable equilibria: a radiative equilibrium-like state, denoted **a**, and a disturbed state **b**, in which upward propagating stationary waves flux momentum down to the lower boundary, weakening zonal winds. Figure 1(a,b) depicts the zonal wind and streamfunction of these two equilibria. SSW events in this model are abrupt transitions from the region near **a** to the region near **b**. If a strong wave from below happens to catch the stratospheric vortex in a "vulnerable" configuration—e.g., measured by an index of refraction (Charney and Drazin 1961; Yoden 1987)—then a burst of wave activity can propagate upward, ripping apart the polar vortex and causing zonal wind to collapse. With certain parameters, the vortex can get stuck in repeated "vacillation cycles", in which the vortex begins to restore with the help of radiative forcing, only to be undermined quickly by the wave. The situation of two well-separated equilibria is highly idealized, and not a generic feature of climate phenomena; this system, with these parameters, serves to demonstrate qualitative features of SSW, not represent the real stratosphere quantitatively. It also gives a clear demonstration of our quite general method. We refer the reader to Holton and Mass (1976); Yoden (1987); Christiansen (2000); and Finkel et al. (2021) for complete model specification.

After discretizing to 27 vertical levels, we end up with a state space with a dimension of $d = 3 \times (27-2) = 75$, with a state vector

$$\mathbf{X}(t) = \Big[\mathrm{Re}\{\Psi(t)\}, \mathrm{Im}\{\Psi(t)\}, U(t)\Big] \in \mathbb{R}^{75} \tag{3}$$

each of the three entries representing a vector with 25 discrete altitudes. We thus obtain a system of 75 ODEs, $\dot{\mathbf{X}}(t) = \boldsymbol{v}(\mathbf{X}(t))$. We furthermore perturb the system by stochastic forcing to represent unresolved processes such as gravity waves, an idea originally put forward by Birner and Williams (2008) and used more recently by Esler and Mester (2019). It could also represent model error, e.g., the effects of smaller-scale waves that have been truncated. The forcing is white in time, giving an Itô diffusion

$$d\mathbf{X}(t) = \boldsymbol{v}(\mathbf{X}(t))\,dt + \boldsymbol{\sigma}(\mathbf{X}(t))\,d\mathbf{W}(t) \tag{4}$$

where $\mathbf{W}(t)$ is an m-dimensional white-noise process, and $\boldsymbol{\sigma} \in \mathbb{R}^{d\times m}$ is a matrix specifying the spatial structure of the noise. We use the exact same form of noise as in Finkel et al. (2021), and also explain it here for reference. At each timestep $\delta t = 0.005$ days, after incrementing the full system by $\delta\mathbf{X} = \boldsymbol{v}(\mathbf{X})\delta t$, we additionally increment the zonal wind profile by

$$\delta U(z) = \sigma_U \sum_{k=0}^{m} \eta_k \sin\left[\left(k+\frac{1}{2}\right)\pi\frac{z}{z_{\text{top}}}\right]\sqrt{\delta t} \tag{5}$$

where $\sigma_U = 1$ m s^{-1} day$^{-1/2}$, whose units reflect the quadratic variation of Brownian motion (e.g., Oksendal 2003). The numerical scheme is known as Euler-Maruyama (e.g., Pavliotis 2014, ch. 5). The vertical coordinate z ranges from 0 at the bottom of the domain (the tropopause) to 70 km at the top of the domain. Equation (5) determines the matrix $\boldsymbol{\sigma}$ in (4). This noise is smooth in space, consisting of $m = 2$ Fourier modes in the vertical. The specific choice of stochastic forcing does affect the transition path statistics, but our method can be applied to any stochastic forcing.

A *transition path* is defined as an unbroken segment, or trajectory, of the system that begins in a region A of state space (a neighborhood of **a**) and travels to another region B (a neighborhood of **b**) without returning to A. As in Finkel et al. (2021), we define A and B based on the zonal-mean zonal wind at $z = 30$ km:

$$A = \{\mathbf{x} \in \mathbb{R}^d : U(30\text{ km})(\mathbf{x}) \geq U(30\text{ km})(\mathbf{a}) = 53.8\text{ m/s}\} \tag{6}$$

$$B = \{\mathbf{x} \in \mathbb{R}^d : U(30\text{ km})(\mathbf{x}) \leq U(30\text{ km})(\mathbf{b}) = 1.75\text{ m/s}\} \tag{7}$$

A SSW event is then a transition from A to B, while the reverse, from B to A, represents the recovery of the vortex. The definition of B modifies the widely used definition of Charlton and Polvani (2007) in two ways. First, we use zonal wind at 30 km above the tropopause (in log-pressure coordinates, which is roughly twice as high as the 10 hPa standard) because 30 km is where the zonal wind profile of **b** reaches a minimum, and Christiansen (2000) used this same coordinate when studying the same model. We also modify the zonal wind thresholds order to ensure that $\mathbf{a} \in A$ and $\mathbf{b} \in B$. Our method could easily adapt to other definitions—bistability is not a requirement for doing TPT analysis—but the bistability in this model makes for a clear demonstration.

3. Transition path ensemble

Every SSW event, or transition path, is a sample from a high-dimensional distribution called the *transition path ensemble*, which refers to the infinite collection of paths one would obtain by running the model forever. We will first give an account of the transition path ensemble based on storylines of the few individual events shown Fig. 2. Subsequently, we will present the TPT analysis, which describes the distribution as a whole using a specific collection of functions including probability densities, committors, and currents.

a. SSW storylines

Fig. 1c shows a 3000-day model integration in a two-dimensional subspace consisting of zonal wind $U(30\text{ km})$ and vertically integrated eddy meridional heat flux, which is abbreviated IHF (integrated heat flux) and defined as

$$\mathrm{IHF}(30\text{ km}) = \int_0^{30\text{ km}} e^{-z/H}\,\overline{v'T'}(z)\,dz \tag{8}$$

IHF quantifies the heat being advected into the polar region associated with the sudden warming. In the Holton-Mass model, the integrand takes the form

$$e^{-z/H}\overline{v'T'}(z) = e^{-z/H}\frac{Hf_0}{R}\overline{\frac{\partial\psi'}{\partial y}\frac{\partial\psi'}{\partial z}} \propto |\Psi(z)|^2\frac{\partial\varphi}{\partial z}, \tag{9}$$

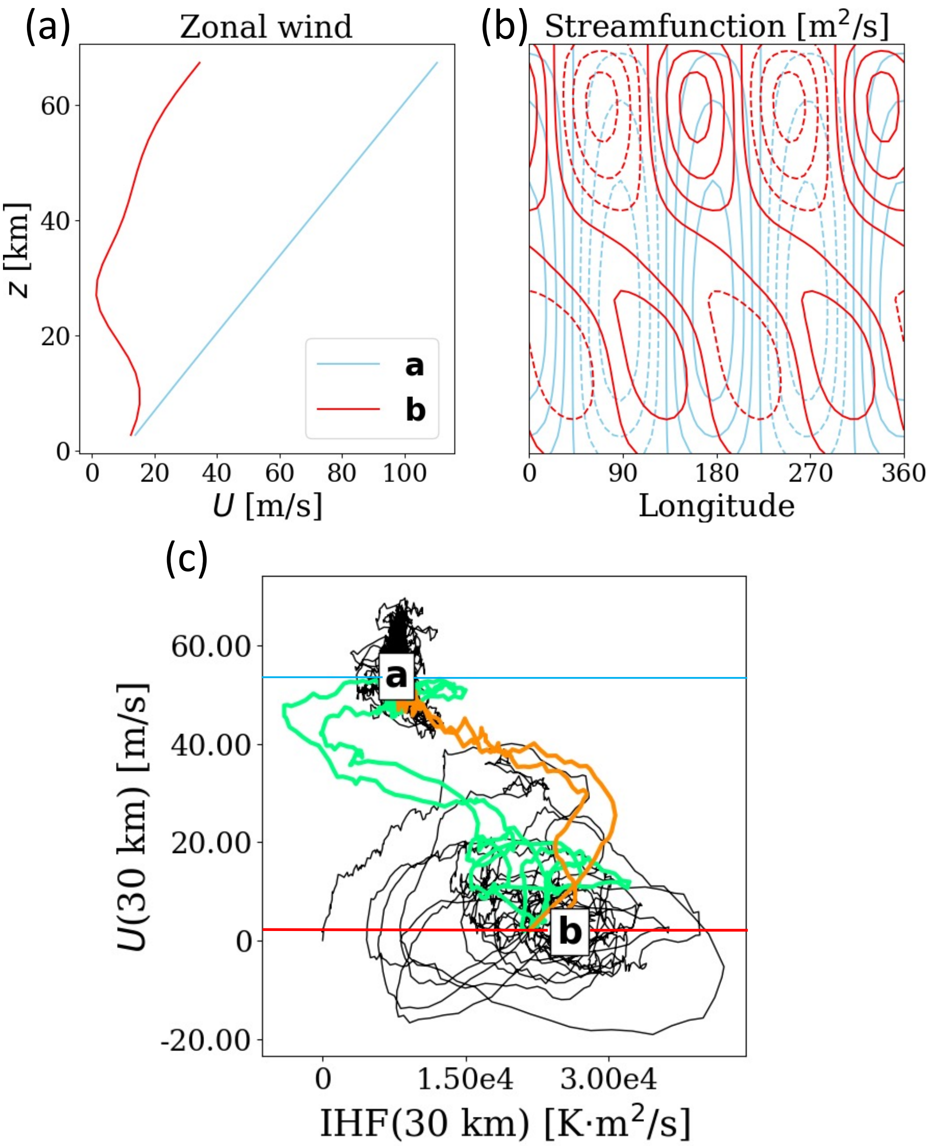

FIG. 1. **The two stable equilibria of the Holton-Mass model**. (a) Zonal-mean zonal wind $U(z)$ and (b) perturbation streamfunction $\psi'(z)$, with contour spacing of 1.5×10^7 m^2/s. Blue indicates the strong vortex equilibrium, **a**, and red indicates the weak vortex equilibrium, **b**, as in Eq. (6). (c) A 3000-day model integration in the subspace of IHF(30 km) and $U(30\text{ km})$. IHF(30 km) = the heat flux integrated from 0 to 30 km; see text for definition. The 3000-day integration contains two SSW events (transitions from $A\rightarrow B$, in orange) and two recovery events (transitions from $B\rightarrow A$, in green). A is the region above **a**, and B is the region below **b**, both delineated by horizontal lines. The figure is similar to Fig. 1 of Finkel et al. (2021).

where R is the ideal gas constant for dry air, and φ is the phase of Ψ. Hence the heat flux is related to the amplitude and phase tilt of the waves, both of which rise significantly during a SSW event. In Fig. 1c, the fixed point **b** has more than twice the IHF of **a**, and the $A\rightarrow B$ transitions (orange segments) begin with a simultaneous decrease in U and increase in IHF. The $B\rightarrow A$ transitions (green segments) do not retrace the same route backward, but rather linger in the vicinity of B before gaining zonal wind strength and decreasing in IHF, which even dips slightly negative in the late stages of vortex recovery.

The same two variables, U and IHF, are plotted over time in Fig. 2(a,c), with transition paths highlighted in the same colors. The neighborhoods A and B are clearly metastable: the system tends to linger in one of the regions for an extended period before quickly switching to the

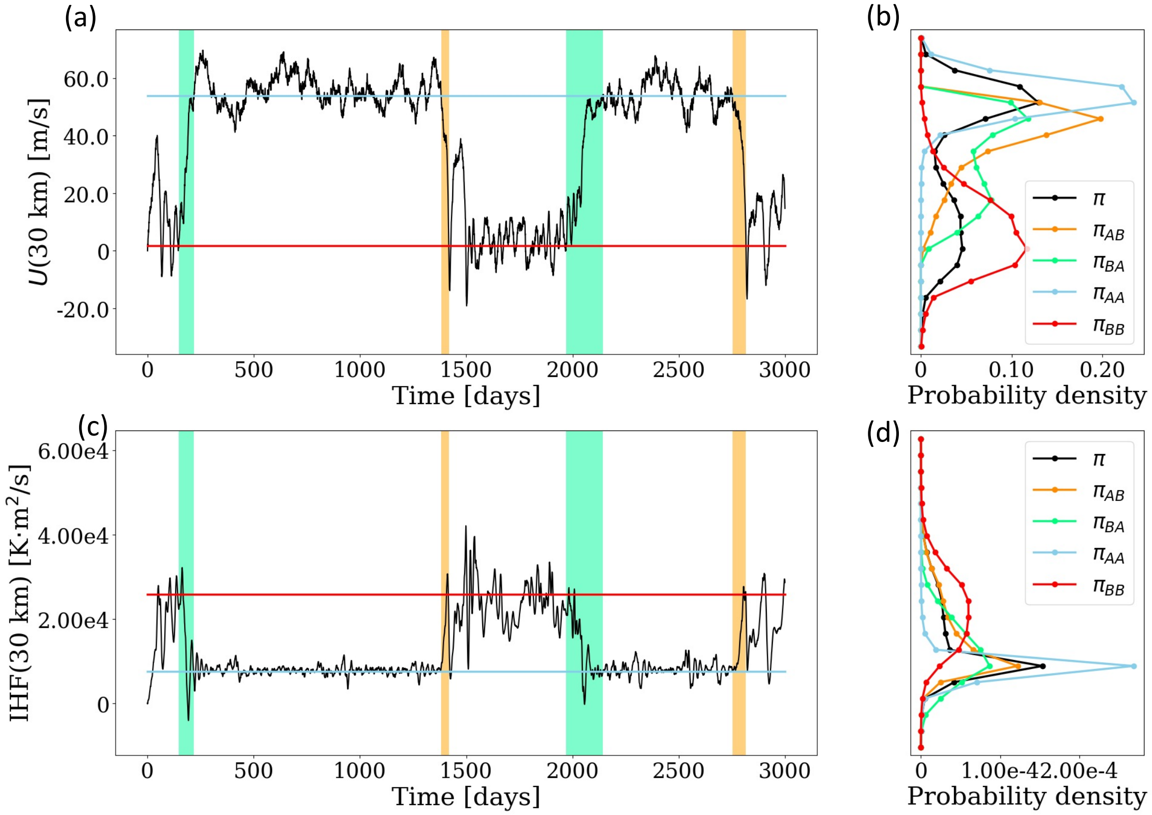

FIG. 2. **Bistable time series**. (a) Zonal wind at 30 km over time, with $A \rightarrow B$ transitions (SSWs) highlighted in orange and $B \rightarrow A$ transitions highlighted in green. (b) Conditional probability distributions of each of the four phases. (c-d) Same as a-b but with integrated heat flux up to 30 km plotted instead of zonal wind at 30 km. Blue and red lines show the position of the two fixed points, **a** and **b**, along these two observables.

other. We can also see bistability by looking at the steady-state probability density, denoted $\pi(\mathbf{x})$, which is plotted as black curves in Fig. 2(b,d). The curve is bimodal over U(30 km). Over IHF(30 km), $\pi(\mathbf{x})$ is sharply peaked over A but low and flat over B, reflecting persistent fluctuations, the "vacillation cycles" of Holton and Mass (1976), in the weak-vortex regime.

We can decompose the distribution more explicitly into four separate "phases" induced by the presence of sets A and B. (i) In the $A \rightarrow B$ phase, marked by orange, the vortex is breaking down, en route from A to B. (ii) In the $B \rightarrow A$ phase, marked by green, the vortex is recovering from the vacillating regime back to the radiatively driven regime. (iii) In the $A \rightarrow A$ phase, the vortex is strong and remaining strong for the time being, either inside set A or taking a brief excursion before returning back to A. (iv) In the $B \rightarrow B$ phase, the vortex is weak, caught in ongoing vacillation cycles in the vicinity of B. We denote the corresponding probability densities as π_{AB}, π_{BA}, π_{AA}, and π_{BB}, and plot them in Fig. 2(b,d) along with the overall density π. Concretely, π_{AB} can be obtained from DNS by running a long simulation, extracting only the $A \rightarrow B$ transition paths, and plotting a histogram of those states. The other phases are obtained analogously. (The supplement explains the alternative short-trajectory computation.) The two peaks in $\pi(\mathbf{x})$, over both observables U(30 km) and IHF(30 km), are seen to come from two unimodal distributions, π_{AA} and π_{BB}. In both panels (b) and (d) the peak over A is narrow and tall compared to the low, wide peak over B, indicating a higher degree of variability associated with vacillation cycles.

When the system is en route from A to B, we say it is (AB)-*reactive*, using a term from chemistry literature where the passage from A (reactant) to B (product) models a chemical reaction. Therefore we refer to π_{AB} and π_{BA} as (AB and BA)-*reactive densities*, which reveal structure hidden from view within the sparsely-populated region between A and B. Along U(30 km), π_{AB} is peaked near A and falls off rapidly toward B, suggesting that transition paths spend much of their time slowly crawling away from

A before speeding up later on. π_{BA} has two peaks in the transition region, suggesting that the system takes a long time to escape from B, and also a long time to re-enter A. This asymmetry is not so clear over the observable IHF(30 km), in which π_{AB} and π_{BA} look quite similar, underscoring the need to examine multiple subspaces to distinguish the phases.

The two events in Figs. 1c and 2(a,b) are only samples from the full transition path ensemble. Any small sample of events cannot fully represent the whole ensemble of transition paths (for example, in the real world, SSWs have two distinct types: split and displacement). How should we describe this complicated ensemble faithfully? The distributions π_{AB} and π_{BA} tell us where transition paths tend to linger, on average, but not much about their detailed movement through state space. A standard approach is to average together multiple events to obtain a composite evolution, which can reveal important features of SSW climatology (e.g., Charlton and Polvani 2007; Albers and Birner 2014; Mitchell et al. 2011). However, lining up multiple time series with different durations requires some arbitrary choices. Conventionally, the "central date" of the warming—when zonal wind first reverses—is used as a reference point, but this may obscure the initial seeds of SSW that happen at different times in advance.

The issue is illustrated in Fig. 3. Panel (a) shows zonal wind over time for 300 observed transition events leading up the warming. Three of these paths are colored, only in between the last-exit time from A (denoted τ_A^-) and the first-entrance time to B (denoted τ_B^+), to illustrate some of the variability between transition paths. The red curve sinks steadily downward until accelerating into a SSW, while the black curve spends a long time trapped in a partially weakened vortex state before its ultimate decline. The cyan pathway does something in between. The remaining gray trajectories include several deep dives and partial recoveries of zonal wind before ultimately descending into B. Panel (b) shows the composite evolution of these 300 trajectories: at every point in time, the black curve shows the median, while the three red envelopes show the middle 20th, 50th, and 90th percentile ranges. (We include in this average the timeseries that have not yet left set A, although the definitions to follow will exclude these early segments from the analysis). The composite evolution successfully captures the sharp nosedive in zonal wind at the end of the transition pathway, but misses the large meanders that some paths, including the black path, go through before the precipitous decline. A comprehensive account of the transition path ensemble should include the stagnations as well. In order to capture these initial stages, we have defined SSW in such a way that the full process takes ~ 80 days, much longer than the ~ 10 days time horizon that traditionally comprises a SSW event. This model, like the true atmosphere, sees the most dramatic zonal wind collapse only in the last few days; however, we will show that most of the probabilistic progress occurs during the longer preceding "preconditioning" stage.

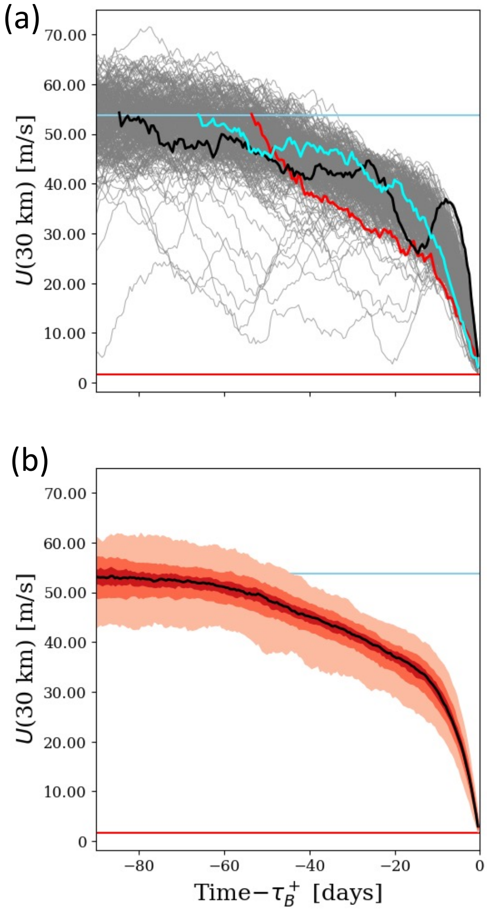

FIG. 3. **SSW ensemble and composites.** (a) 100 SSW realizations in gray in terms of U(30 km), aligned by the central date of the warming when zonal wind dips below 1.75 m/s. Three of the realizations are colored in between their last-exit time from A (τ_A^-) and their next-hitting time to B (τ_B^+). (b) Composite evolution of U(30 km). The black curve shows the pointwise median, and the three red-orange envelopes show the middle 20, 50, and 90 percentile ranges.

The TPT approach averages trajectories together in a different way, aligning them by their position in state space rather than by the time until SSW (which is itself a random variable). This new kind of composite evolution is the essence of the probability current, which highlights the sequence of events that must happen between A and B regardless of the time horizon. In the rest of this section, we define and visualize probability currents, starting with their basic ingredients: committor functions.

b. Committors, densities, and currents

Let us fix an initial condition $\mathbf{X}(t_0) = \mathbf{x}$ with a vortex that is neither strong nor fully broken down, so $\mathbf{x} \notin A \cup B$.

$\mathbf{X}(t)$ will soon evolve into either A or B, since both are attractive. The probability of hitting B first is called the *forward committor* (to B):

$$q_B^+(\mathbf{x}) = \mathbb{P}_\mathbf{x}\{\mathbf{X}(\tau_{A\cup B}^+(t_0)) \in B\} \tag{10}$$

where the subscript $\mathbf{x}$ denotes a conditional probability given $\mathbf{X}(t_0) = \mathbf{x}$, and $\tau_S^+(t_0)$ is the *first hitting time* after t_0 to a set $S \subset \mathbb{R}^d$:

$$\tau_S^+(t_0) = \min\{t > t_0 : \mathbf{X}(t) \in S\}. \tag{11}$$

Here, S is the union of A and B, i.e., the trajectory has returned to a metastable state. The probability of hitting A first instead—the "forward committor to A"—is $q_A^+(\mathbf{x}) = 1 - q_B^+(\mathbf{x})$. Unless specified otherwise, we call q_B^+ the forward committor, as the SSW event is our main interest. Committors are deterministic functions of state space involving ensemble averages of $\mathbf{X}(t)$, whereas hitting times are random variables depending on the realization of $\mathbf{X}(t)$. Our system is autonomous, with no external time-dependent forcing, so we can set $t_0 = 0$ and drop the argument from $\tau_{A\cup B}^+$ without loss of generality. The autonomous assumption can be relaxed, either by augmenting $\mathbf{x}$ with a periodic variable for time (e.g., to include the seasonal cycle) or by augmenting A and B to include initial and terminal times (e.g., to examine climate change effects). Periodic- and finite-time TPT has been presented in Helfmann et al. (2020), and we plan to utilize this framework in a forthcoming paper using state-of-the-art ensemble forecasts. As a conceptual demonstration, the autonomous Holton-Mass model makes for a clearer exposition.

While the forward committor is a central quantity for forecasting, it does not distinguish the $A \to B$ phase from the $B \to B$ phase, i.e., it tells us nothing about the past of $\mathbf{X}(t)$ for $t < t_0$. For this we also need to introduce the *backward committor* (to A):

$$q_A^-(\mathbf{x}) = \mathbb{P}_\mathbf{x}\{\mathbf{X}(\tau_{A\cup B}^-(t_0)) \in A\} \tag{12}$$

where $\tau_S^-(t_0)$ is the *most recent hitting time*

$$\tau_S^-(t_0) = \max\{t < t_0 : \mathbf{X}(t) \in S\} \tag{13}$$

The backward-in-time probabilities refer specifically to the process $\mathbf{X}(t)$ *at equilibrium*, allowing us once again to set $t_0 = 0$. The backward committor to B is $q_B^-(\mathbf{x}) = 1 - q_A^-(\mathbf{x})$. Again, the phrase "backward committor" will refer to q_A^- unless stated otherwise.

The forward and backward committors are shown in Fig. 4(a,b). In this and later figures, the white regions of state space have insignificant probability. Note that q_B^+ and q_A^- have very different contour structures, a sign of irreversible behavior (in a stochastic system with detailed balance, i.e., a reversible system, $q_A^- = 1 - q_B^+$). Both q_B^+ and q_A^- are large in the upper-right flank of state space, meaning that whenever medium-strength zonal wind and large IHF are observed together, chances are high that the system both came from A and will next hit B. In other words, a SSW is underway. Compare to the middle-left flank of state space, where q_B^+ is large but q_A^- is small: there, the system is likely headed toward B, *from* B, which does not count as a SSW event.

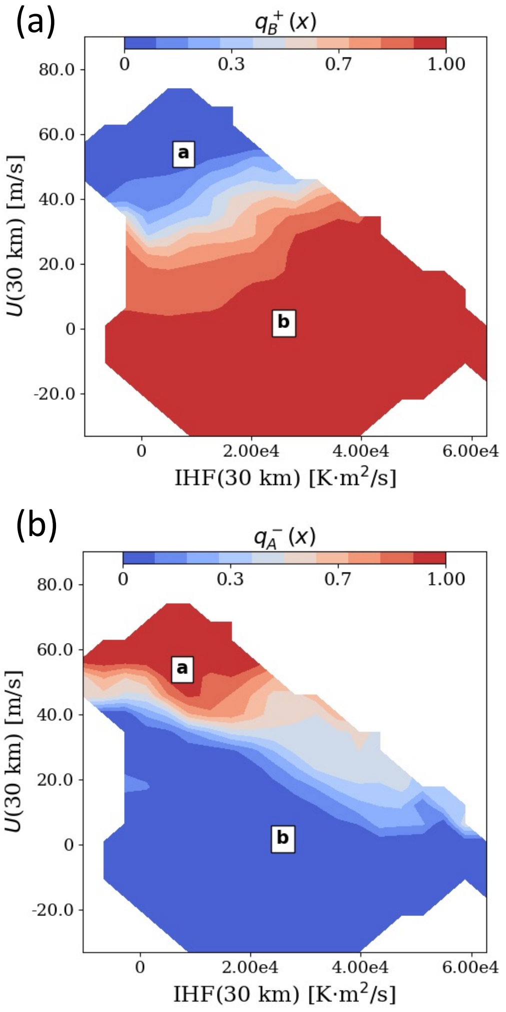

FIG. 4. **Committors.** (a) Forward committor $q_B^+(\mathbf{x})$, the probability to hit B next starting from $\mathbf{x}$, and (b) backward committor $q_A^-(\mathbf{x})$, the probability to have come from A last given the current state $\mathbf{x}$. The committors are projected on a two-dimensional space (IHF(30 km),U(30 km)).

With committor functions, we can now formally define the transition probability density π_{AB} (and π_{BA} as well, just by swapping A and B in the formulas to follow).

$$\pi_{AB}(\mathbf{x}) = \frac{1}{Z_{AB}}\pi(\mathbf{x})q_A^-(\mathbf{x})q_B^+(\mathbf{x}) \tag{14}$$

where Z_{AB} is a normalizing constant such that the right-hand side integrates to one.

Each probability density (π, π_{AB}, π_{BB}, etc.) is associated with a *probability current* ($\mathbf{J}$, $\mathbf{J}_{AB}$, $\mathbf{J}_{BB}$, etc.). The steady-state current $\mathbf{J}(\mathbf{x})$ is a vector field that describes the probability mass flux through $\mathbf{x}$. It is related to the deterministic flow $\dot{\mathbf{X}}(t) = \boldsymbol{v}(\mathbf{X}(t))$, but differs by a factor of $\pi(\mathbf{x})$ to account for density variations and a diffusion term to account for the stochastic perturbations. For a diffusion process of the form (4), these currents have the explicit form

$$\mathbf{J}(\mathbf{x}) = \pi\boldsymbol{v} - \nabla\cdot(\mathbf{D}\pi), \tag{15}$$

$$\mathbf{J}_{AB}(\mathbf{x}) = q_A^- q_B^+ \mathbf{J} + \pi\mathbf{D}\big[q_A^-\nabla q_B^+ - q_B^+\nabla q_A^-\big], \tag{16}$$

where the diffusion matrix $\mathbf{D}(\mathbf{x}) = \frac{1}{2}\boldsymbol{\sigma}(\mathbf{x})\boldsymbol{\sigma}(\mathbf{x})^\top$, and ∇ represents the gradient operator over state space. One can substitute A and B for other symbols to single out the phase of interest. Dependence on $\mathbf{x}$ has been suppressed throughout. Unlike the deterministic flow field $\boldsymbol{v}(\mathbf{x})$, $\mathbf{J}(\mathbf{x})$ is divergence-free, reflecting the steady-state property that every region of state space has a constant probability mass. (See Vanden-Eijnden (2006) and Metzner et al. (2006) for a thorough mathematical treatment, or Finkel et al. (2020) for an application to a simpler SSW model.) Fig. 5a overlays $\mathbf{J}(\mathbf{x})$ (black arrows) atop $\pi(\mathbf{x})$ (orange logarithmic color scale). The vector field lives in $\mathbb{R}^{75}$, but we have projected it into two dimensions using a visualization procedure due to Strahan et al. (2021) and described in section 2 of the supplement. The two black curves in Fig. 2 are the two marginals of the orange density in Fig. 5. The two probability peaks around A and B are seen as dark blobs, each of which is surrounded by strong probability currents and separated by a region of weaker current.

To understand this vector field, we make a fluid-dynamical analogy. If A and B are two coherent eddies in a body of water, a tracer particle spends most of its time trapped in one of the two, but is occasionally ejected from one eddy and entrained in the other. The equilibrium current is thus dominated by the velocity fields of the two eddies, but the smaller filaments that connect them are responsible for occasional transition events, which of course are our primary interest. To single out the dynamics of each phase, we decompose $\mathbf{J}(\mathbf{x})$ just as we decomposed $\pi(\mathbf{x})$, conditioning on the past and future of $\mathbf{X}(t)$ as it passes through $\mathbf{x}$. $\mathbf{J}_{AB}(\mathbf{x})$, shown in Fig. 5b, is the average flow of trajectories moving from A to B through $\mathbf{x}$; $\mathbf{J}_{AA}(\mathbf{x})$, shown in Fig. 5c, is the flow from A back to A through $\mathbf{x}$, etc. The background colors are the probability densities for the corresponding phase. For example, panel (c) shows $\pi_{AB}(\mathbf{x})$, the probability of finding a trajectory at $\mathbf{x}$ given that it is en route from A to B.

By visualizing transition pathways as static vector fields in state space, we switch from a Lagrangian to an Eulerian reference frame and fulfill our promise to "align transition paths by their position in state space." The averaging choices in Fig. 3 were challenging because each "particle" (ensemble member) approaches B through a different pathway. The probability currents protray the global behavior of transitions, as opposed to "case studies" provided by individual trajectories.

Let us examine the characteristics of each phase. The current $\mathbf{J}_{AA}$ is disorderly and suggests that typical fluctuations around A are usually extinguished swiftly by the restoring force of radiative equilibrium. On the other hand, $\mathbf{J}_{BB}$ is a highly organized "eddy" around $\mathbf{b}$. This reflects the vacillation cycles seen in the time series of Fig. 2, and offers a dynamic perspective not available from the stationary distribution $\pi_{BB}(\mathbf{x})$. Each cycle consists of a slow buildup of zonal wind driven by radiative cooling, wave enhancement allowed by the growing PV gradient, and subsequent collapse of zonal wind. Mathematically, the linearized system near $\mathbf{b}$ is stable with complex eigenvalues; $\mathbf{b}$ is an attracting fixed point, and without noise the oscillations would die out eventually. Stochastic forcing injects enough energy to excite the system off of the fixed point, and a nearby limit cycle beyond a Hopf bifurcation directs this energy into sustained oscillations (Yoden 1987).

Comparing Fig. 5(a,b,e), we see that the steady-state current is approximately the sum of $\mathbf{J}_{AA}$ and $\mathbf{J}_{BB}$, two coherent eddies separated by a barrier at U(30 km) ≈ 35 m/s. The occasional $A \longrightarrow B$ transition breaches this barrier in a way described by $\mathbf{J}_{AB}$ in Fig. 5c. $\mathbf{J}_{AB}$ emerges from set A with gradually increasing IHF and decreasing zonal wind. At first $\mathbf{J}_{AB}$ matches approximately with $\mathbf{J}_{AA}$, extending out of the lower-right corner of A, but at U(30 km) ≈ 40 m/s, $\mathbf{J}_{AB}$ separates decisively into its own unique stream. Down to U(30 km) ≈ 30 m/s, $\mathbf{J}_{AB}$ remains strong and localized in a narrow tube going downward and rightward. Subsequently, $\mathbf{J}_{AB}$ weakens and spreads out as it turns downward for its final descent into B, indicating that pathways tend to meander more widely through this late stage of a SSW in the Holton-Mass system.

To corroborate the representation of transition pathways by $\mathbf{J}_{AB}$, we have also plotted five realized transition paths from the reference simulation in blue. True to the vector field, the transition paths stay tightly clustered together as zonal wind slackens and the streamfunction begins to tilt, but scatter widely when they dip below U(30 km) ≈ 30 m/s, and enter B with a range of IHF values between 2×10^4 and 5×10^4 K·m^2/s.

As a second point of comparison, we have also plotted the minimum-action pathways (both from $A \longrightarrow B$ and $B \longrightarrow A$) with thick cyan lines, representing the most likely transition path in the low-noise limit (e.g., Freidlin and Wentzell 1970; E et al. 2004; Forgoston and Moore 2018). The pathway solves an optimization problem, deviating as minimally as possible from the deterministic dynamics

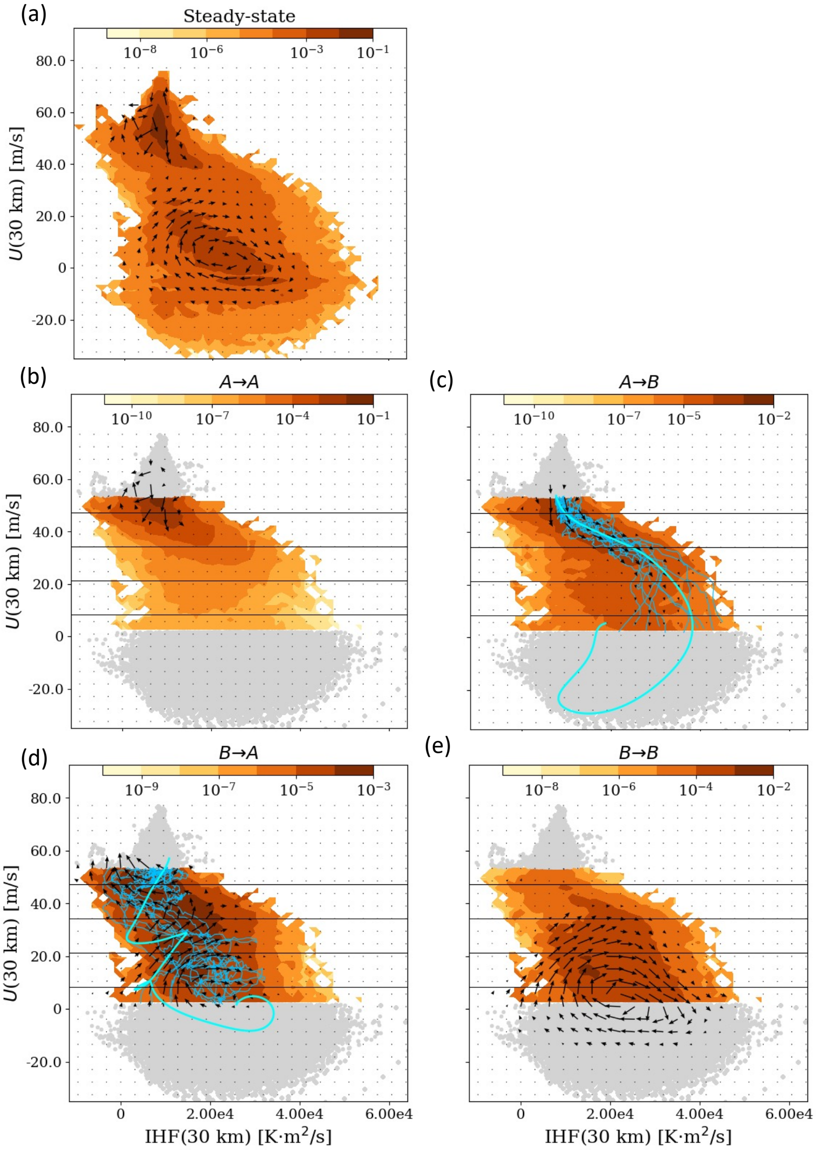

Fig. 5. **Densities and currents.** (a) shows the equilibrium density $\pi(\mathbf{x})$ and equilibrium current $\mathbf{J}(\mathbf{x})$. (b-e) show the reactive densities and currents for $A \longrightarrow A$, $A \longrightarrow B$, $B \longrightarrow A$, and $B \longrightarrow B$ transitions, respectively. For example, (c) shows the reactive current $\mathbf{J}_{AB}(\mathbf{x})$ overlaid on the reactive $\pi_{AB}(\mathbf{x})$, illustrating the most common pathways of SSW trajectories from the strong to weak vortex state. Thick cyan curves in (c) and (d) mark the minimum-action pathways from $A \longrightarrow B$ and $B \longrightarrow A$, respectively, while thin blue curves show a few sampled realized transition pathways. Gray dots are data points inside states A and B.

while still bridging the gap all the way from A to B. We use sequential quadratic programming to approximate the minimum-action path following Plotkin et al. (2019), and describe our procedure further in section 3 of the supplement. (In fact we solve for the minimum-action pathway almost all the way to the fixed point $\mathbf{b}$; up to the boundary of B, this makes no difference.) As the stochastic forcing shrinks to zero, we expect $\mathbf{J}_{AB}$ to collapse into a single streamline following the minimum-action path (but becoming increasingly unlikely as we approach this limit). The finite-noise transition path ensemble, however, departs significantly from it. In the initial stages of transition in Fig. 5c, the minimum-action path tracks right down the center of $\mathbf{J}_{AB}$, suggesting this feature is stable with noise. At the end of the process, widening of current streamlines makes it impossible for the minimum-action path to represent the full ensemble meaningfully.

After a SSW event and ensuing vacillation cycles, the vortex eventually recovers, returning from B back to A, which is encoded by the current $\mathbf{J}_{BA}$ in Fig. 5d. The $B \longrightarrow A$ current is very different from a reversed $A \longrightarrow B$ current. After many loops around B, $\mathbf{J}_{BA}$ emerges upward out of B just as in any other vacillation cycle, with a partial restoration of wind. The current then bifurcates: one branch continues its upward creep in zonal wind while reversing course in the IHF direction, eventually rebuilding a strong enough polar vortex to inhibit wave propagation and allowing radiative relaxation to take over, drawing it back into A. Meanwhile, the other branch of current continues to track with $\mathbf{J}_{BB}$ halfway through the wave amplification phase, as if about to execute another loop around **b**. But $\mathbf{J}_{BA}$ stalls in the middle of the wave amplification phase, near IHF(30 km) = 3×10^4 K·m^2/s. Where does the current go from there? Fig. 5(d,e) indicates that the eddy is centered slightly above the top of B, allowing some room for small vacillation cycles to proceed without technically re-entering B. This is the likely fate of some trajectories along the second branch of $\mathbf{J}_{BA}$, which finally achieve "escape velocity" the second time around.

The minimum-action path from B to A captures some of the tortuous nature of this transition, with several setbacks and subsequent regrouping events. However, it differs significantly from $\mathbf{J}_{BA}$ overall. Because $\mathbf{J}_{BA}$ flows over a wide channel, any single path (even the minimum-action path) cannot reasonably be expected to represent the ensemble meaningfully.

c. Stages of a SSW from probability current

We can analyze SSW progression more systematically and quantitatively using the following property of reactive currents. Let C be a closed hypersurface in $\mathbb{R}^d$ which encloses A and is disjoint with B; we call this a *dividing surface*. Then we have

$$\oint_C \mathbf{J}_{AB} \cdot \mathbf{n}\, d\sigma = \text{Transition rate} \tag{17}$$

where $\mathbf{n}$ is an outward unit normal from C, σ is a surface element, and the transition rate is the average number of $A \longrightarrow B$ transitions (SSW events) per unit time, or equivalently the inverse return period. The stochastic Holton-Mass model has a rate of $\sim (1700 \text{ days})^{-1}$, which changes with parameters such as noise strength. The integral relationship (17) holds for any dividing surface, implying that the current is divergence-free outside of A and B, but has a source in A and a sink in B (vice versa for $\mathbf{J}_{BA}$). The integrand $\mathbf{J}_{AB} \cdot \mathbf{n}$, which we will henceforth call the $\mathbf{J}_{AB}$-*flux density* (not to be confused with heat flux or IHF) can be interpreted as a quasi-probability density, which is normalized to integrate to a constant (the transition rate) but may take on negative values for some choices of dividing surfaces. Because the number of $A \longrightarrow B$ transitions per unit time must equal the number of $B \longrightarrow A$ transitions per unit time, Eq. (17) must also hold when $\mathbf{J}_{AB}$ is replaced by $\mathbf{J}_{BA}$ and $\mathbf{n}$ is replaced by $-\mathbf{n}$. The reactive current essentially decomposes the rate among a continuum of possible pathways, which is much more dynamically insightful than the numerical value of the rate itself.

We visualize the progression of SSW events as $\mathbf{J}_{AB}$-flux densities through dividing surfaces, for two different families of dividing surfaces (zonal wind strengths and committor levels) to illustrate different aspects of the process. We will then quantify how SSW progresses over time.

1) Surfaces of constant zonal wind

The simplest choice of dividing surfaces is a series of hyperplanes with constant U(30 km), represented as horizontal black lines in Fig. 5(b-e). To get from A to B, a trajectory must pass downward once through each threshold. It may also cross down, then up, then down; or three times down and two times up, etc., as long as the *net* number of downward crossings is one for each surface. The $\mathbf{J}_{AB}$-flux density element $\mathbf{J}_{AB}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x})\, d\sigma(\mathbf{x})$ can be interpreted as the long-term average number of net crossings through the surface at $\mathbf{x}$. Note that in the $A \longrightarrow B$ direction, $\mathbf{n}$ points in the direction of negative U(30 km), i.e., $\mathbf{n} = -\nabla U(30 \text{ km})/\|\nabla U(30 \text{ km})\|$.

Fig. 6 shows the $\mathbf{J}_{AB}$-flux densities (a) and $\mathbf{J}_{BA}$-flux densities (b) across each surface. The horizontal axis is IHF(30 km), as in Fig. 5, which instantiates the $\mathbf{J}_{AB}$-flux density element as

$$\mathbf{J}_{AB} \cdot \mathbf{n}\, d\sigma = \mathbf{J}_{AB} \cdot \left(-\frac{\nabla U(30 \text{ km})}{\|\nabla U(30 \text{ km})\|} \right) d\big[\text{IHF}(30 \text{ km})\big]$$

Here, the differential $d[\text{IHF}(30 \text{ km})]$ is shorthand for $\int dx_1 \dots \int dx_{73}\, d\big[\text{IHF}(30 \text{ km})\big]$, where $x_1, \dots, x_{73}$ are the 73 dimensions of state space orthogonal to both the IHF(30 km) and the U(30 km) axes. Accordingly, the vertical axis of Fig. 6 has the units needed to normalize the integrals to a transition rate in days^{-1}. At the first $A \longrightarrow B$ threshold U(30 km) = 47.3 m/s, the flux distribution has a tall, narrow, negative spike, where $\mathbf{J}_{AB}$ points downward across the surface. There is also a small positive spike to the left due to a small amount of backflow where transition paths temporarily regain a bit of the lost zonal wind—not enough to re-enter A—before weakening again. This backflow corresponds to the small wiggles early in the black and cyan time series in Fig. 3. Moving from blue to red curves, as zonal wind drops further, we see the negative spike widen and slightly flatten, while the positive spike shrinks and disappears. By the last threshold U(30 km) = 8.3 m/s, the $\mathbf{J}_{AB}$-flux density appears entirely negative, consistent with the sharp downturn into B seen in both Figs. 3 and 5c. It also covers a wider range of integrated heat flux, consistent with the weaker current magnitude pointing into B in Fig. 3c. The $\mathbf{J}_{BA}$-flux density somewhat mirrors the

$\mathbf{J}_{AB}$-flux density, but with a larger backflow spike relative to the forward flow: in the early stages of vortex recovery (red and orange curves in panel (b)), a strengthening zonal wind at low values of IHF is accompanied by weakening zonal wind at higher value of IHF. This is consistent with the winding, branching character of $\mathbf{J}_{AB}$ in Fig. 5d, which inherits some clockwise circulation from $\mathbf{J}_{BB}$. In other words, the early $B \longrightarrow A$ transition stages experience residual vacillation cycles, which ultimately dampen and die by the time zonal wind has reached 47.3 m/s (there is no noticeable negative dip in the dark blue curve in Fig. 6b).

These flux densities trace out a simpler version of the "transition tubes" defined in Vanden-Eijnden (2006). The distributions cannot be interpreted as the path of a single event, but rather as the flow of SSW "traffic" through a sequence of thresholds, indicating the most frequently traveled paths. Another important caveat is that a single-signed $\mathbf{J}_{AB}$-flux density (such as the red curve in Fig. 6a) does not imply strictly monotonic changes in zonal wind across that surface: it only means that the backflow, if present, is not systematically displaced from the forward flow along the IHF axis, as it is in the red curve in panel (b). However, a different choice of horizontal axis might reveal more coherent cyclical behavior. In general, reactive currents generally contain much more information that can be queried by slicing it along in different dimensions, which should be chosen with some physical intuition.

2) Surfaces of constant committor

Zonal wind, the defining coordinate for A and B, is an obvious measure of progress which we have used in Fig. 6. However, in some ways it is not the most natural. First, the presence of backflow, while it does reveal some interesting dynamics of transition paths, suggests that a particular zonal wind level might be associated with forward or backward progress depending on other variables. Second, by the time a typical transition path reaches the halfway point of $U(30\text{ km}) \approx 25$ m/s, its committor probability has risen to nearly 100% (cf. Figs. 4 and 5c; "typical" means along the main channel of $\mathbf{J}_{AB}$). The subsequent collapse of zonal wind is locked in by that point. The committor itself is a more balanced metric of progress toward B, and can be used the same way to find transition routes. A committor level set $\{\mathbf{x} : q_B^+(\mathbf{x}) = q_0\}$, i.e., all states with equal likelihood q_0 of SSW, is a dividing surface just like a level set of U, and thus supports a $\mathbf{J}_{AB}$-flux density similar to those in Fig. 6. We will see that this flux density is almost uniformly positive.

In Fig. 7, we plot a larger collection of $\mathbf{J}_{AB}$-flux densities, represented by gray histograms, across 15 level sets of the committor. The $\mathbf{J}_{AB}$-flux density elements for panels (a) and (b), are, respectively,

$$\mathbf{J}_{AB} \cdot \left(\frac{\nabla q_B^+}{\|\nabla q_B^+\|}\right) d\big[U(30\text{ km})\big] \tag{18}$$

$$\mathbf{J}_{AB} \cdot \left(\frac{\nabla q_B^+}{\|\nabla q_B^+\|}\right) d\big[\text{IHF}(30\text{ km})\big] \tag{19}$$

We also display the minimum-action path with a dashed black curve for comparison. Panel (a) confirms that the

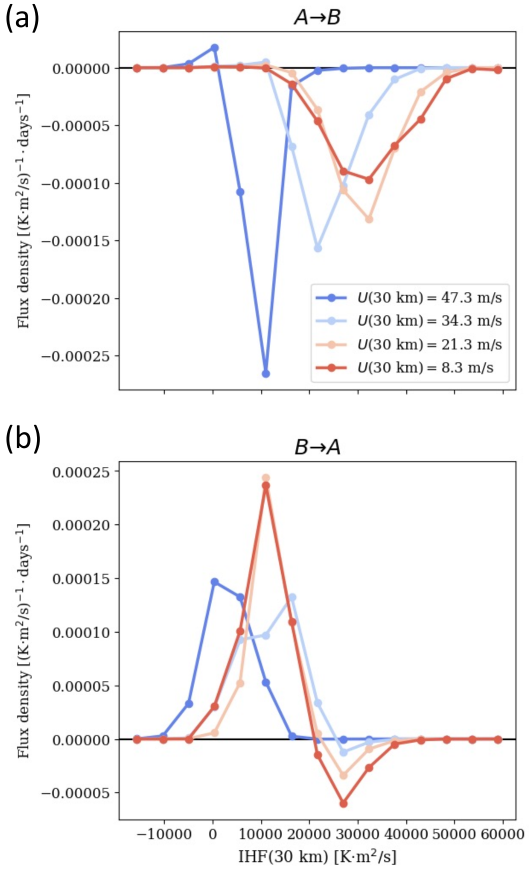

Fig. 6. **$\mathbf{J}_{AB}$-flux density (a) and $\mathbf{J}_{BA}$-flux density (b) as a function of IHF(30 km), over four different level sets of U (30 km).** These cross sections of the reactive current from A to B and B to A illustrate the mean direction of trajectories crossing different zonal wind thresholds as a function the IHF. For an SSW (a), the progression marches from high winds (blue curves) to low winds (red) with increasing mean and variabilty of the IHF, while for the recovery of the vortex (b), the main progresssion is up toward higher wind, albeit with more substantial cycling down at higher values of IHF. Each density should have the same integral (in absolute value), equal to the rate. Due to numerical error, the integrals can vary and the rate is calculated by an averaging procedure (see section 3 of the supplement). For visual clarity, we have normalized each curve to have the same integral. To integrate to a rate, in days^{-1}, the vertical axis must have units of $[\text{K}\cdot\text{m}^2/\text{s}]^{-1}\text{days}^{-1}$. This unit depends on the orientation of the dividing surface in state space, as well as the coordinates along that surface chosen for projection.

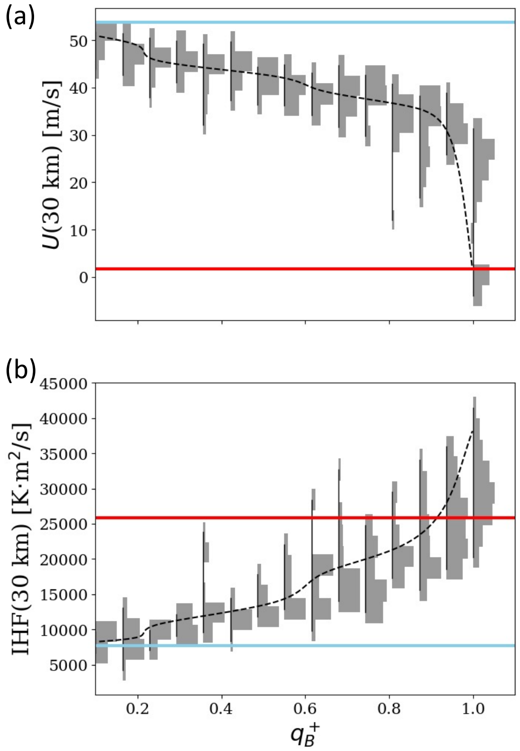

FIG. 7. **Minimum-action paths and path distributions**. At a series of level sets in the committor q_B^+, gray histograms indicate the $\mathbf{J}_{AB}$-flux density of (a) zonal wind U(30 km) and (b) integrated heat flux IHF(30 km). Dashed curves show the minimum-action pathway in the same space. The minimum-action path tracks the mean of the full ensemble except very near SSW (q_B^+ near 1), where the jet breaks down more rapidly, accompanied by an extreme heat flux. The more extreme nature of the minimum-action path was also observed in Figure 5c, where it tracks along the rightmost envelope of more typical trajectories.

zonal wind-committor relationship is nonlinear: approximately half of the total decline in zonal wind happens after the committor has surpassed 80% probability. The $\mathbf{J}_{AB}$-flux density widens across U in the late transition stages, past $q_B^+ \sim 0.7$. This indicates, somewhat puzzlingly, that the system may "commit" to B at a range of zonal wind strengths, even though B itself is defined by a fixed threshold $U(30\text{ km}) \leq 1.75$ m/s. Fig. 4a offers some explanation: as the committor increases towards 1, the level sets become increasingly tilted across this two-dimensional state space. The last visible level set (the boundary between dark orange and red) spans the approximate range 5 m/s $\lesssim U(30\text{ km}) \lesssim$ 30 m/s, depending on the value of IHF(30 km) along the horizontal. A large heat flux carries the promise of imminent SSW by sending waves into the stratosphere that will deposit enough negative momentum to surely destroy the vortex, even if the vortex is still persisting for the time being. If heat flux is weak, on the other hand, zonal wind must also be very weak to ensure the same degree of SSW certainty. Thus, the spread in zonal wind is closely tied with the spread in heat flux. This is consistent with Fig. 7b which shows the integrated heat flux distribution across each level set of q_B^+. Indeed, the distribution widens progressively from $q_B^+ \approx 0.5$ until the end of the path, consistent with the diffusing $\mathbf{J}_{AB}$ vector field and the diverging sample paths in 5c, as well as the broadening flux distributions in Fig. 6a. An interesting difference between the flux distributions and minimum-action path is that the latter decisively chooses the high-heat flux route, far outstripping the bulk of the flux distribution in Fig. 7b and hugging the right end of state space in Fig. 5c. We speculate that because stochastic forcing only acts on zonal wind, rather than the streamfunction (which determines heat flux), the minimum-action path recruits the heat flux mechanism to do the "heavy lifting" of decelerating the zonal wind, thereby achieving SSW with a lower cost. An interesting future experiment would be to vary the form of stochasticity (5) and explore the consequences for flux distributions and minimum-action paths. TPT may thus offer an important rare event-oriented calibration tool for stochastic parameterization of climate models.

We have so far focused on observables at a fixed altitude of $z = 30$ km (or integrated up to 30 km), but the vertical structure of zonal wind and heat flux is essential to understand the physical processes of SSW onset. Every altitude z has a separate observable $U(z)$, with its own $\mathbf{J}_{AB}$-flux density $\mathbf{J}_{AB} \cdot \nabla U(z)/\|\nabla U(z)\|$ of the same kind as Figs. 6 and 7. We visualize this z-indexed family of distributions in Fig. 8a by plotting their medians (solid curves) as functions of z, for five different committor level sets from 0 to 1. The background shading covers the interquartile range (25th-75th percentiles) of the $\mathbf{J}_{AB}$-flux density. There is essentially zero "backflow" across these surfaces, so the $\mathbf{J}_{AB}$-flux densities are ordinary nonnegative probability densities. Blue and red dashed curves represent the fixed points **a** and **b**. Fig. 8b shows the same construction, but with z-dependent meridional heat flux $\overline{v'T'}(z)$ as the independent variable. Together, these profiles give an idea of the joint evolution of propagating waves and weakening mean flow during the course of SSW.

As the committor increases from 0 to 0.6 (blue to yellow), the zonal wind profile slackens most noticeably at a low altitude range of 10-20 km, and the interquartile range remains narrow, suggesting that transitions are constrained to play out along a range of pathways with low variability. At the same time, meridional heat flux develops a positive bulge at the same low altitude range, indicating some upward flux of wave activity emanating from the troposphere. Later, as the committor increases to 1.0 (yellow to red), the wind profile stagnates at altitudes below 20 km, and above that continues weakening gradually. Most noticeably, the *variability*, both in zonal wind and heat flux, increases at higher altitudes of 30-50 km. At $q_B^+ = 0.95$, the distribution

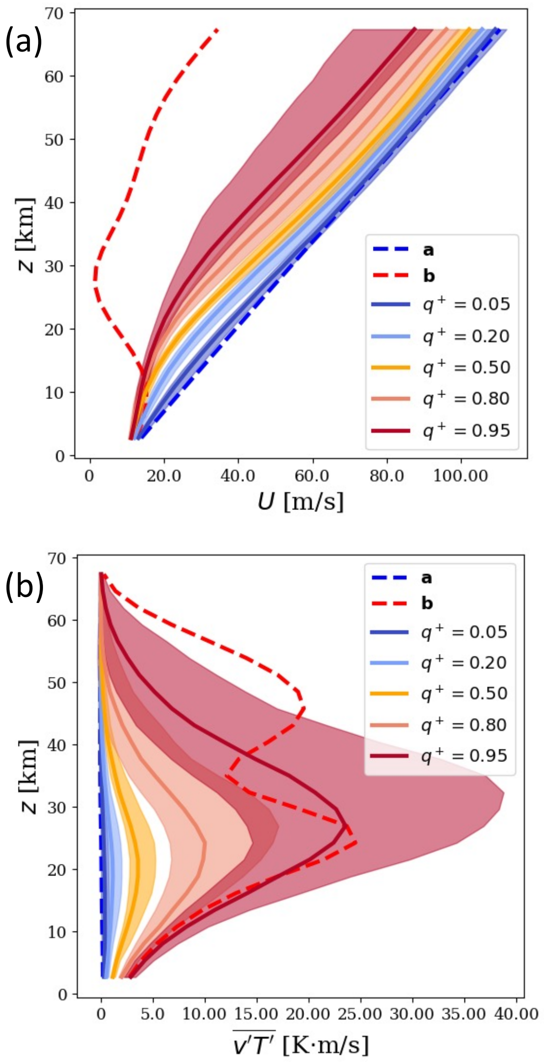

Fig. 8. **Typical transition states and variability**. For a sequence of five committor ranges, we plot (a) the zonal wind profile and (b) the meridional heat flux profile that is most typical of that committor range in the sense of reactive current flux density. Shading represents the 25th-75th percentile range of the flux distribution. Blue and red dashed curves represent the profiles for the fixed points **a** and **b**, respectively. The widening of the distribution of both winds and IHF at high committor values (close to the SSW) highlights the diversity in late stage events which is lost in a composite approach (as in Figure 3) that pins all events together by the point of the vortex reversal. Even at a committor value of 0.95, the vortex is still largely intact above 15 km, emphasizing the importance of preconditioning the low level winds.)

of zonal wind at high altitudes begins to skew sharply toward weak winds. Meanwhile, the distribution of heat flux profiles grows and widens, and the bulge moves slightly upward toward 30 km. This is consistent with the broadening of $\mathbf{J}_{AB}$ in IHF space in the final transition stages (Fig. 5c), and indicates a continued upward flow of wave activity. A slight change in zonal wind belies a substantial increase in SSW probability, which will eventually bring about an abrupt breakdown and explosion of variability in zonal wind.

3) Evolution over time

We have now measured SSW progress by two different coordinates, $U(30\text{ km})$ and $q_B^+(\mathbf{x})$, and visualized its composite evolution in both spaces. What neither of them captures directly is time: how long does SSW take to complete, and how long is each stage? We wish to produce a TPT version of the composite evolution shown in Fig. 3. To do this, we replace the hitting time τ_B^+ (a random variable) with its conditional expectation, the *lead time*,

$$\eta_B^+(\mathbf{x}) = \mathbb{E}_{\mathbf{x}}\left[\tau_{A\cup B}^+ | \mathbf{X}(\tau_{A\cup B}^+) \in B\right], \tag{20}$$

in other words, the average time from $\mathbf{x}$ to B conditional on hitting B before A. The composites in Fig. 3b parameterize the SSW process by τ_B^+ itself, which varies randomly from path to path, whereas $\eta_B^+(\mathbf{x})$ is the average value of τ_B^+ over all possible paths and hence a deterministic function of state space. We used η_B^+ as a forecast function in Finkel et al. (2021), and we display it here in Fig. 9 over the same two-dimensional subspace, along with several committor level sets for comparison. $\eta_B^+(\mathbf{x})$ is uniformly zero on set B, increases farther away from B, and becomes undefined on set A. Fig. 9b gives an idea of how the certainty of SSW is related to the time until it happens. It turns out that along transition paths, the committor increases at an approximately linear rate with respect to time. Both the flux distributions and the minimum-action path indicate that the lead time drops by ~8 days for every additional ~10% in the likelihood of SSW. In particular, this means that the ultimate collapse of zonal wind in Fig. 7 is not only "sudden" with respect to the committor, but also with respect to the lead time. The final 20 days of the transition path (as measured by η_B^+) corresponds to the final ~ 5% of probability needed to achieve SSW, from 95% to 100%, and yet this same interval sees approximately 30 m/s reduction in zonal wind—the entire second half of its journey from A to B. This is the sense in which the pre-sudden part of SSW constitutes most of the probabilistic progress. Dynamically, it seems that this half-weakened polar vortex has been accompanied by "irreversible" changes in the flow field, the Holton-Mass version of the threshold behavior found in Nakamura et al. (2020).

To visualize the time dependence of transition paths more directly, we can construct $U(30\text{ km})$ (or any other observable) as a function of time implicitly by considering the *joint* distribution of $U(30\text{ km})$ and η_B^+ across different committor level sets, according to the $\mathbf{J}_{AB}$-flux density. The corresponding infinitesimal element is

$$\mathbf{J}_{AB}\cdot\left(\frac{\nabla q_B^+}{\|\nabla q_B^+\|}\right) d\left[\eta_B^+\right] d\left[U(30\text{ km})\right] \tag{21}$$

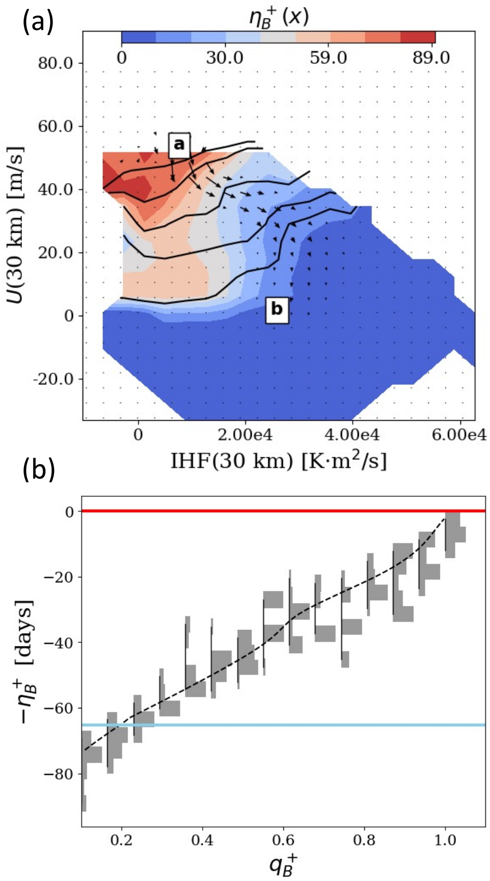

Fig. 9. **Lead time-committor relationship.** (a) Background color shows η_B^+, the expected time to reach B from initial condition $\mathbf{x}$, conditional on hitting B next. Note that the contour structure is very different from that of the forward committor, whose level sets $q_B^+ = 0.1, 0.2, 0.5, 0.8$, and 0.9 are shown in solid black lines (cf. Fig. 4). Notable differences are in the light red region where the wind is approximately 20 m/s and IHF near 10^4 K· m/s: SSW events rarely occur from these initial conditions, and are associated with long trajectories (lead time of about 60 days) that often cycle back towards state A before swinging down to state B. Probability current $\mathbf{J}_{AB}$ is overlaid, the same as in Fig. 5c. (b) The distribution of lead time across a series of level sets of the committor, the same level sets as in Fig. 7.

whose two-dimensional integral is, again, the transition rate. For a sequence of 30 committor level surfaces, Fig. 10a shows quantiles of U(30 km) (a) and IHF(30 km) (b) vs. the median lead time η_B^+ in the horizontal. These "TPT composites" resemble the traditional composite of Fig. 3b. but differ in several important ways. The traditional composite narrows toward the end, by construction, since the entrance to B is defined by a single value of U(30 km). In contrast, the TPT composite widens toward the end before the final narrowing: as Fig. 9a demonstrates, the level sets of q_B^+ and η_B^+ closest to B both cover a range of U(30 km) values. The final collapse of zonal wind, which typically happens in the lower-right corner of state space, is so sudden that the lead time hardly changes, and so inevitable that the committor hardly changes. Of course, formally $\eta_B^+ = 0$ and $q_B^+ = 1$ if and only if $U(30\text{ km}) \leq 0$, a boundary condition we have enforced in Fig. 10a (see section 2 of the supplement for details). From the TPT perspective, however, the process is essentially complete.

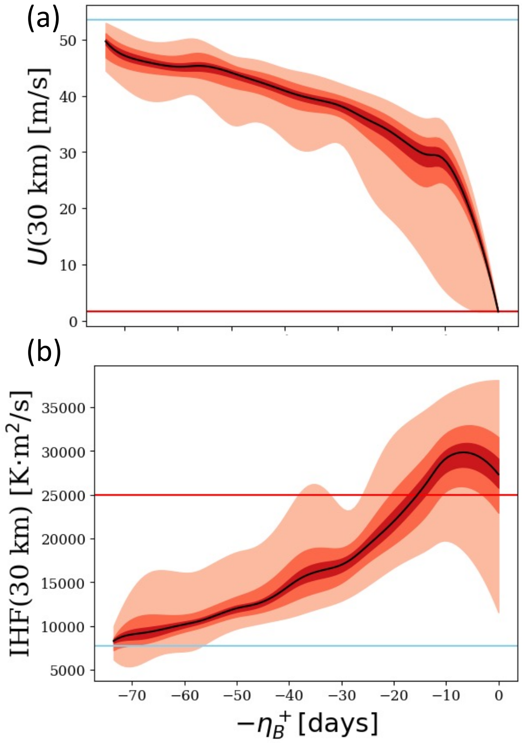

Fig. 10. **TPT composite evolution vs. time.** For 15 committor level sets (the same as in Figs. 7 and 9b) we approximate the joint distribution of (a) lead time and zonal wind, and (b) lead time and integrated heat flux, according to the flux density of $\mathbf{J}_{AB} \cdot \mathbf{n}$ through the committor level surface. The three red-orange envelopes represent the middle 20%, 50%, and 90% percentile ranges. Black curves connect the medians. Unlike the traditional SSW composite shown in Figure 3, the variability in trajectories is more uniform in lead time, actually increasing near the event. This is due to use of the committor as the ordering coordinate, which aligns paths by the future predictability of an event. The widening at near $\eta_B^+ = 10$ days reflects the diversity of model states when a SSW is approximately 95% likely to occur, as seen in Figure 8. All of these states are equally likely to move to an SSW with an expected lead time of 10 days, but there is a distribution of actual lead times which contributes to the spread in winds and heat flux.

The TPT composite also has a wavy character not captured by the traditional composite. The individual samples in Fig. 3a do seem to proceed in pulses of steady downward progress punctuated by brief, partial recoveries. Because these partial recoveries are staggered in time between paths, the traditional composite in Fig. 3b cannot capture them. However, these wiggles may correspond robustly to various level sets of committor or lead time, which would suggest the waviness of the TPT composite is indeed capturing this same phenomenon. Some of the gray transition paths in Fig. 3 go through even larger oscillations after approaching close to B, which may correspond to the rapid expansion of the outer envelope (middle 90 percentile) in Fig. 10a at $\eta_B^+ = 30$ days. The probability currents in the lower left corners of Fig. 5(a,d,e) indicate, indeed, that this region is associated with *increasing* zonal wind strength, which of course is only temporary if the trajectory is bound for B. These partial recoveries may be interpreted as minor warmings preceding the major warming. Nevertheless, the individual pathways are only case studies, and their detailed correspondence with the TPT composite is speculative. A more refined DGA discretization, or a large-scale time series statistical analysis, would confirm or deny the robustness of these oscillatory features, but such analysis is beyond the scope of this paper.

4. Numerical method

The results above can in principle be computed by direct numerical simulation (DNS). To demonstrate that TPT analysis can scale to more complex models, we have instead used the dynamical Galerkin approximation (DGA) which avoids the need to simulate trajectories on the timescale of the SSW return time.

DGA is detailed in the supplement and in previous papers (Thiede et al. 2019; Strahan et al. 2021; Finkel et al. 2021), but we briefly sketch the procedure here. The key observation underpinning DGA is that unknown "forecast functions" of interest — $q_B^+(\mathbf{x}), q_A^-(\mathbf{x}), \eta_B^+(\mathbf{x}), \pi(\mathbf{x})$, etc — can be expressed as solutions to equations involving only short-time evolution of $\mathbf{X}$. For example, the committor, q_B^+, solves the equation

$$\begin{aligned} q_B^+(\mathbf{x}) &= \mathbb{E}_{\mathbf{x}}\big[q_B^+(\mathbf{X}(\Delta t))|\mathbf{X}(0)=\mathbf{x}\big], \quad \mathbf{x} \notin A \cup B \qquad (22) \\ q_B^+(\mathbf{x}) &= 1,\ \mathbf{x} \in B \quad \text{and} \quad q_B^+(\mathbf{x}) = 0,\ \mathbf{x} \in A \end{aligned}$$

for $\mathbf{x}$ outside of A and B. In this equation we interpret evolution of $\mathbf{X}(t)$ to stop upon entrance to A or B. The user-chosen parameter Δt limits the length of the simulated trajectories. Crucially, Eq. (22) identifies q_B^+ exactly for any choice of Δt.

To approximately solve Eq. (22) and similar equations for other quantities of interest, we first generate a data set by sampling many points $\mathbf{X}_n(0)$ from all over state space according to some *sampling measure*, μ, and then launching a short trajectory from each one, yielding a data set $\{\mathbf{X}_n(t) : 0 \le t \le \Delta t\}_{n=1}^N$. This sampling measure, the number $N = 3\times10^5$ trajectories, and the length $\Delta t = 20$ days, are key parameters of the method. The trajectories are significantly shorter than the typical ~ 80-day duration of SSW. As in Finkel et al. (2021), the initial conditions are resampled from a long (2×10^5 days) control simulation to be uniformly distributed on the space $(|\Psi|(30\,\text{km}), U(30\,\text{km}))$. With a more complex (expensive) model we would not be able to rely on a long control simulation to seed the initial points. Optimizing this procedure is, therefore, a crucial step for future research, and should draw on existing rare event sampling strategies such as those presented in Ragone et al. (2018); Webber et al. (2019); Simonnet et al. (2021); Abbot et al. (2021) and others, perhaps with a combination of surrogate and high-fidelity models.

After generating the data, we expand unknown functions of interest in basis sets informed by the data, and then solve matrix equations for the expansion coefficients. For the forward committor we write

$$q_B^+(\mathbf{x}) \approx \sum_{j=1}^{M} w_j(q_B^+)\phi_j(\mathbf{x}) \qquad (23)$$

with analogous expansion coefficients $w_j(q_A^-)$ and $w_j(\pi)$ for the backward committor and steady-state density, respectively. There is a wide range of choices for constructing basis functions, and in fact different bases may be optimal to compute different quantities of interest. In this work, we simply use indicator (or characteristic) functions. To construct the basis sets, we divide state space $\mathbb{R}^d$ into a partition of disjoint sets $\{S_1, \dots, S_M\}$ and discretize the continuous-space process $\mathbf{X}(t) \in \mathbb{R}^n$ into an index process $S(t) \in \{1, \dots, M\}$, where $S(t) = j$ if $\mathbf{X}(t) \in S_j$. The corresponding basis functions are

$$\phi_j(\mathbf{x}) = \mathbb{1}_{S_j}(\mathbf{x}) := \begin{cases} 1 & \mathbf{x} \in S_j \\ 0 & \text{otherwise.} \end{cases} \qquad (24)$$

The sets $\{S_1, \dots, S_M\}$ are found by clustering the complete set of states in our short-trajectory data set using K-means clustering as implemented in the `scikit-learn` Python library (Pedregosa et al. 2011) along with the hierarchical adjustment described in Finkel et al. (2021), with $M = 1500$ clusters. The choice of a basis of indicator functions found by data clustering is borrowed from a well-studied class of coarse-grained models known as Markov state models (MSMs) (Noé et al. 2009b; Chodera and Noé 2014), and with this choice our estimates of the committors and steady-state density are nearly identical (up to details related to boundary conditions) to those obtained by the MSM approach with the same clusters.

The Galerkin method proceeds by inserting the expansion in Eq. (23) into the short-trajectory equation solved by the quantity of interest (Eq. (22) in the case of q_B^+) and then integrating both sides against a test function ϕ_i, also from the basis. The result is in an $M \times M$ linear system. With an indicator basis as in Eq. (24), the matrix elements yield a Markov transition probability matrix

$$P_{ij} = \mathbb{P}_\mu\{\mathbf{X}(\Delta t) \in S_j\,|\,\mathbf{X}(0) \in S_i\},\ \ i, j \in \{1, \dots, M\}. \quad (25)$$

where the subscript μ indicates that $\mathbf{X}(0)$ is drawn from the sampling measure μ, restricted to S_i. The matrix entries are expectations over both the initial conditions $\mathbf{X}_n(0)$ and the final conditions $\mathbf{X}_n(\Delta t)$ and are estimated by sample averaging using our short trajectory data set, i.e. by

$$P_{ij} = \frac{\#\{n : \mathbf{X}_n(0) \in S_i, \mathbf{X}_n(\Delta t) \in S_j\}}{\#\{n : \mathbf{X}_n(0) \in S_i\}}, \qquad (26)$$

Given the transition matrix P_{ij}, the committor coefficient vector obeys a discrete version of Eq. (22):

$$\begin{aligned} w_i(q_B^+) &= \sum_{j=1}^{M} P_{ij} w_j(q_B^+), \quad S_i \not\subseteq A \cup B \qquad (27) \\ w_i(q_B^+) &= 1,\ S_i \subseteq B \quad \text{and} \quad w_i(q_B^+) = 0,\ S_i \subseteq A \end{aligned}$$

We have assumed that A, B, and $(A \cup B)^c$ are partitioned separately, meaning each S_i is either completely inside A, completely inside B, or disjoint from both, which we ensure in the clustering step. As in Eq. (22), in Eq. (27) we interperet evolution of $\mathbf{X}_n(t)$ to be stopped upon entrance to A or B.

The coefficients of the steady-state density obey another linear equation:

$$\begin{aligned} w_i(\pi) &= \sum_{j=1}^{M} w_j(\pi) P_{ji} \qquad i = 1, \ldots, M \qquad (28) \\ & \sum_{j=1}^{M} w_j(\pi) = 1. \end{aligned}$$

Note that in this case the equation involves the transpose of P instead of P itself and does not come with any boundary conditions.

The backward committor obeys a similar equation to (27), but with two differences. First, P_{ij} is replaced by $\widetilde{P}_{ij} = \frac{w_j(\pi)}{w_i(\pi)} P_{ji}$, which represents the process under time reversal at steady-state. Second, for q_A^-, the boundary conditions are flipped from those of q_B^+: $w_i(q_A^-) = 1$ for $S_i \subseteq A$ and $w_i(q_A^-) = 0$ for $S_i \subseteq B$. Because the time-reversed matrix depends on the steady-state density, q_A^- must be solved after π.

The lead time η_B^+ solves a similar, but slightly more intricate, equation. We postpone that formula to the supplement, where we also provide complete details for the the derivations presented in this section as well as numerical validation of the DGA procedure.

With approximations to the committors and steady-state density provided by DGA (or any other means), TPT provides recipes to assemble approximations of the transition path statistics examined in this paper. For example, the reactive density π_{AB} can be computed directly from its definition in (14). The transition rate and projections of the reactive current $\mathbf{J}_{AB}$ are estimated by more involved procedures presented in detail in section 3 of the supplement.

5. Conclusion

Extreme weather events are a central challenge of climate modeling. We need to be able to characterize changes in flooding, heat waves, cold spells, and other natural disasters. While many existing techniques are being developed to simulate and diagnose rare events, there is an overall lack of standard language and benchmarks for comparison. A related computational problem is that rare events take a long time to appear, let alone produce a significant statistical distribution, in both models and observations.

We have advocated two ideas to advance extreme weather modeling. First, we have presented a transition path theory (TPT) analysis of a prototypical extreme event, sudden stratospheric warmings (SSW) in the Holton-Mass model. TPT provides a set of summary statistics that encapsulate important features of rare events, including rates (or inverse return times), precursors, and onset behavior. Probability densities and currents tell us how the system evolves through state space to an SSW event, including the interplay between momentum and heat transfers. The minimum-action method provides a useful but limited point of comparison, as it provides no information about the *variablity* of transitions. Second, we have demonstrated the numerical ability to use short simulations to estimate rare event statistics, which has great potential as a parallelizable alternative to running long simulations. This was shown in Finkel et al. (2021) for the narrow goal of forecasting SSW events in the Holton-Mass model; here we have used the same computational method to ask more intricate statistical questions about the evolution of SSWs from start to finish.

We have shown that transition paths in the Holton-Mass model generally evolve through two distinct phases: (i) a gradual, halting decline in zonal wind strength in tandem with a slowly increasing meridional heat flux over a period of approximately 2 months, followed by (ii), a rapid burst of heat flux and deceleration of zonal wind in the last 10 days. The sudden breakdown of the vortex in the second stage encompasses the classic synoptic evolution of an SSW, but from a predictability standpoint, it is changes in the precondition phase that are most critical, allowing one to forecast a warming before the event is already in motion. Our key conclusion is the SSW committor probability rises the most during the preconditioning phase. The committor signals an upcoming SSW before changes in the vortex (as quantified by just the zonal mean zonal wind) can be clearly identified above the noise in an individual trajectory.

A judicious choice of the "climatological state" A is essential to maximize predictive and dynamical understanding of the rare event's origin when using the TPT framework. In defining A relative to winds in the strong vortex meta-stable state, we were able to fully include stage (i). This lengthened the window over which we could tracked SSW trajectories to seasonal time scales. Extending this work to the atmosphere, where the climatological state is itself evolving on comparable time scales, remains a challenge.

Our work is an early application of TPT to atmospheric science. We believe it holds potential as a framework for forecasting, risk analysis, and uncertainty quantification. Thus far, it has been used mainly to analyze protein folding in molecular dynamics, but is now being applied in diverse fields such as social science (Helfmann et al. 2021), as well as ocean and atmospheric science (Finkel et al. 2020; Helfmann et al. 2020; Lucente et al. 2021). A potential limitation of TPT is that it cannot easily quantify long-term correlations between successive rare events. For example, a large earthquake might release enough tectonic stress to make the next one less severe. The approach will require further extensions to address such issues.

Significant challenges also remain for deploying DGA at scale to state-of-the-art climate models. The numerical pipeline used in this paper is far from optimal, as we have focused on basic deliverables of TPT. One important limitation is the data generation step. We used a long ergodic trajectory to sample the attractor, which served the double purpose of seeding initial data points for short trajectories (i.e., defining the sampling measure μ) and providing a ground truth for validating the accuracy of DGA. In a real application where DGA is advantageous, this data set would not be available, and more advanced sampling methods would be required. One promising strategy is splitting: starting from initial points in A and B, simulate forward for a short time, and replicate trajectories that explore new regions of state space. Efficient sampling is an active research area, with recent work including Hoffman et al. (2006); Weare (2009); Bouchet et al. (2011, 2014); Vanden-Eijnden and Weare (2013); Chen et al. (2014); Yasuda et al. (2017); Farazmand and Sapsis (2017); Dematteis et al. (2018); Mohamad and Sapsis (2018); Dematteis et al. (2019); Ragone et al. (2018); Webber et al. (2019); Bouchet et al. (2019a,b); Plotkin et al. (2019); Simonnet et al. (2021); Ragone and Bouchet (2020); Sapsis (2021); Abbot et al. (2021). We will draw upon these developing methods when scaling DGA up to more realistic models and data.

Acknowledgments. J.F. is supported by the U.S. DOE, Office of Science, Office of Advanced Scientific Computing Research, Department of Energy Computational Science Graduate Fellowship under Award Number DE-SC0019323. During the time of writing, R.J.W. was supported by New York University's Dean's Dissertation Fellowship and by the Research Training Group in Modeling and Simulation funded by the NSF via grant RTG/DMS-1646339. E.P.G. acknowledges support from the NSF through grant AGS-1852727. This work was partially supported by the NASA Astrobiology Program, grant No. 80NSSC18K0829 and benefited from participation in the NASA Nexus for Exoplanet Systems Science research coordination network. J.W. acknowledges support from the Advanced Scientific Computing Research Program within the DOE Office of Science through award DE-SC0020427 and from the NSF through award DMS-2054306. The computations in the paper were done on the high-performance computing cluster at New York University.

We thank John Strahan, Aaron Dinner, and Chatipat Lorpaiboon for helpful methodological advice.

Data availability statement. The code to produce the data set and results, either on the Holton-Mass model or on other systems, is publicly available at `https://github.com/justinfocus12/SHORT`. Interested users are encouraged to contact J.F. for more guidance on usage of the code.

References

Abbot, D. S., R. J. Webber, S. Hadden, and J. Weare, 2021: Rare event sampling improves mercury instability statistics. 2106.09091.

Albers, J. R., and T. Birner, 2014: Vortex preconditioning due to planetary and gravity waves prior to sudden stratospheric warmings. *Journal of the Atmospheric Sciences*, **71 (11)**, 4028–4054, doi:10.1175/JAS-D-14-0026.1.

Birner, T., and P. D. Williams, 2008: Sudden stratospheric warmings as noise-induced transitions. *Journal of the Atmospheric Sciences*, **65 (10)**, 3337–3343, doi:10.1175/2008JAS2770.1.

Bouchet, F., J. Laurie, and O. Zaboronski, 2011: Control and instanton trajectories for random transitions in turbulent flows. *Journal of Physics: Conference Series*, **318 (2)**, 022 041, doi:10.1088/1742-6596/318/2/022041, URL https://doi.org/10.1088%2F1742-6596%2F318%2F2%2F022041.

Bouchet, F., J. Laurie, and O. Zaboronski, 2014: Langevin dynamics, large deviations and instantons for the quasi-geostrophic model and two-dimensional euler equations. *Journal of Statistical Physics*, **156**, 1066–1092, doi:10.1007/s10955-014-1052-5, URL https://doi.org/10.1007/s10955-014-1052-5.

Bouchet, F., J. Rolland, and E. Simonnet, 2019a: Rare event algorithm links transitions in turbulent flows with activated nucleations. *Physical Review Letters*, **122 (7)**, 074 502, doi:10.1103/PhysRevLett.122.074502.

Bouchet, F., J. Rolland, and J. Wouters, 2019b: Rare event sampling methods. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **29 (8)**, 080 402, doi:10.1063/1.5120509.

Charlton, A. J., and L. M. Polvani, 2007: A new look at stratospheric sudden warmings. part i: Climatology and modeling benchmarks. *Journal of Climate*, **20 (3)**, 449–469, doi:10.1175/JCLI3996.1.

Charney, J. G., and P. G. Drazin, 1961: Propagation of planetary-scale disturbances from the lower into the upper atmosphere. *Journal of Geophysical Research (1896-1977)*, **66 (1)**, 83–109, doi:10.1029/JZ066i001p00083, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/JZ066i001p00083, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/JZ066i001p00083.

Chen, N., D. Giannakis, R. Herbei, and A. J. Majda, 2014: An mcmc algorithm for parameter estimation in signals with hidden intermittent instability. *SIAM/ASA Journal on Uncertainty Quantification*, **2 (1)**, 647–669, doi:10.1137/130944977, URL https://doi.org/10.1137/130944977, https://doi.org/10.1137/130944977.

Chodera, J. D., and F. Noé, 2014: Markov state models of biomolecular conformational dynamics. *Current Opinion in Structural Biology*, **25**, 135 – 144, doi:https://doi.org/10.1016/j.sbi.2014.04.002, URL http://www.sciencedirect.com/science/article/pii/S0959440X14000426, theory and simulation / Macromolecular machines.

Christiansen, B., 2000: Chaos, quasiperiodicity, and interannual variability: Studies of a stratospheric vacillation model. *Journal of the Atmospheric Sciences*, **57 (18)**, 3161–3173, doi:10.1175/1520-0469(2000)057<3161:CQAIVS>2.0.CO;2.

Dematteis, G., T. Grafke, M. Onorato, and E. Vanden-Eijnden, 2019: Experimental evidence of hydrodynamic instantons: The universal route to rogue waves. *Phys. Rev. X*, **9**, 041 057, doi:10.1103/PhysRevX.9.041057, URL https://link.aps.org/doi/10.1103/PhysRevX.9.041057.

Dematteis, G., T. Grafke, and E. Vanden-Eijnden, 2018: Rogue waves and large deviations in deep sea. *Proceedings of the National Academy of Sciences*, **115 (5)**, 855–860, doi:10.1073/pnas.1710670115.

E, W., W. Ren, and E. Vanden-Eijnden, 2004: Minimum action method for the study of rare events. *Communications on Pure and Applied Mathematics*, **57 (5)**, 637–656, doi:https://doi.org/10.1002/cpa.20005, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.20005, https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpa.20005.

E, W., and E. Vanden-Eijnden, 2006: Towards a Theory of Transition Paths. *Journal of Statistical Physics*, **123 (3)**, 503, doi:10.1007/s10955-005-9003-9, URL https://doi.org/10.1007/s10955-005-9003-9.

Esler, J. G., and M. Mester, 2019: Noise-induced vortex-splitting stratospheric sudden warmings. *Quarterly Journal of the Royal Meteorological Society*, **145 (719)**, 476–494, doi:https://doi.org/10.1002/qj.3443, URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3443, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3443.

Farazmand, M., and T. P. Sapsis, 2017: A variational approach to probing extreme events in turbulent dynamical systems. *Science Advances*, **3 (9)**, doi:10.1126/sciadv.1701533, URL https://advances.sciencemag.org/content/3/9/e1701533, https://advances.sciencemag.org/content/3/9/e1701533.full.pdf.

Finkel, J., D. S. Abbot, and J. Weare, 2020: Path Properties of Atmospheric Transitions: Illustration with a Low-Order Sudden Stratospheric Warming Model. *Journal of the Atmospheric Sciences*, **77 (7)**, 2327–2347, doi:10.1175/JAS-D-19-0278.1, URL https://doi.org/10.1175/JAS-D-19-0278.1, https://journals.ametsoc.org/jas/article-pdf/77/7/2327/4958190/jasd190278.pdf.

Finkel, J., R. J. Webber, D. S. Abbot, E. P. Gerber, and J. Weare, 2021: Learning forecasts of rare stratospheric transitions from short simulations. 2102.07760.

Forgoston, E., and R. O. Moore, 2018: A primer on noise-induced transitions in applied dynamical systems. *SIAM Review*, **60 (4)**, 969–1009.

Frame, D. J., S. M. Rosier, I. Noy, L. J. Harrington, T. Carey-Smith, S. N. Sparrow, D. A. Stone, and S. M. Dean, 2020: Climate change attribution and the economic costs of extreme weather events: a study on damages from extreme rainfall and drought. *Climatic Change*, **162 (2)**, 781–797.

Freidlin, M. I., and A. D. Wentzell, 1970: *Random perturbations of dynamical systems*. Springer.

Helfmann, L., J. Heitzig, P. Koltai, J. Kurths, and C. Schütte, 2021: Statistical analysis of tipping pathways in agent-based models. *The European Physical Journal Special Topics*, 1–23.

Helfmann, L., E. Ribera Borrell, C. Schütte, and P. Koltai, 2020: Extending transition path theory: Periodically driven and finite-time dynamics. *Journal of Nonlinear Science*, doi:10.1007/s00332-020-09652-7.

Hoffman, R. N., J. M. Henderson, S. M. Leidner, C. Grassotti, and T. Nehrkorn, 2006: The response of damaging winds of a simulated tropical cyclone to finite-amplitude perturbations of different variables. *Journal of the Atmospheric Sciences*, **63 (7)**, 1924 – 1937, doi:10.1175/JAS3720.1, URL https://journals.ametsoc.org/view/journals/atsc/63/7/jas3720.1.xml.

Holton, J. R., and C. Mass, 1976: Stratospheric vacillation cycles. *Journal of the Atmospheric Sciences*, **33 (11)**, 2218–2225, doi:10.1175/1520-0469(1976)033<2218:SVC>2.0.CO;2.

Hu, G., T. Bódai, and V. Lucarini, 2019: Effects of stochastic parametrization on extreme value statistics. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **29 (8)**, 083 102, doi:10.1063/1.5095756, URL https://doi.org/10.1063/1.5095756, https://doi.org/10.1063/1.5095756.

Jayachandran, G., V. Vishal, and V. S. Pande, 2006: Using massively parallel simulation and markovian models to study protein folding: Examining the dynamics of the villin headpiece. *The Journal of Chemical Physics*, **124 (16)**, 164 902, doi:10.1063/1.2186317, URL https://doi.org/10.1063/1.2186317, https://doi.org/10.1063/1.2186317.

Kim, S., H. Kim, J. Lee, S. Yoon, S. E. Kahou, K. Kashinath, and M. Prabhat, 2019: Deep-hurricane-tracker: Tracking and forecasting extreme climate events. *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 1761–1769, doi:10.1109/WACV.2019.00192.

Kron, W., P. Löw, and Z. W. Kundzewicz, 2019: Changes in risk of extreme weather events in europe. *Environmental Science & Policy*, **100**, 74–83, doi:https://doi.org/10.1016/j.envsci.2019.06.007, URL https://www.sciencedirect.com/science/article/pii/S146290111930142X.

Lesk, C., P. Rowhani, and N. Ramankutty, 2016: Influence of extreme weather disasters on global crop production. *Nature*, **529 (7584)**, 84–87, doi:10.1038/nature16467, URL https://doi.org/10.1038/nature16467.

Liu, Y., D. P. Hickey, S. D. Minteer, A. Dickson, and S. Calabrese Barton, 2019: Markov-State Transition Path Analysis of Electrostatic Channeling. *The Journal of Physical Chemistry C*, **123 (24)**, 15 284–15 292, doi:10.1021/acs.jpcc.9b02844, URL https://doi.org/10.1021/acs.jpcc.9b02844, publisher: American Chemical Society.

Lucente, D., C. Herbert, and F. Bouchet, 2021: Committor functions for climate phenomena at the predictability margin: The example of el niño southern oscillation in the jin and timmerman model. 2106.14990.

Mann, M. E., S. Rahmstorf, K. Kornhuber, B. A. Steinman, S. K. Miller, and D. Coumou, 2017: Influence of anthropogenic climate change on planetary wave resonance and extreme weather events. *Scientific Reports*, **7 (1)**, 45 242.

Meng, Y., D. Shukla, V. S. Pande, and B. Roux, 2016: Transition path theory analysis of c-Src kinase activation. *Proceedings of the National Academy of Sciences*, **113 (33)**, 9193–9198, doi:10.1073/pnas.1602790113, URL http://www.pnas.org/lookup/doi/10.1073/pnas.1602790113.

Metzner, P., C. Schutte, and E. Vanden-Eijnden, 2006: Illustration of transition path theory on a collection of simple examples. *The Journal of Chemical Physics*, **125 (8)**, 1–2, doi:10.1063/1.2335447.

Miron, P., F. Beron-Vera, L. Helfmann, and P. Koltai, 2021: Transition paths of marine debris and the stability of the garbage patches. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, accepted for publication.

Mitchell, D. M., A. J. Charlton-Perez, and L. J. Gray, 2011: Characterizing the variability and extremes of the stratospheric polar vortices using 2d moment analysis. *Journal of the Atmospheric Sciences*, **68 (6)**, 1194 – 1213, doi:10.1175/2010JAS3555.1, URL https://journals.ametsoc.org/view/journals/atsc/68/6/2010jas3555.1.xml.

Mohamad, M. A., and T. P. Sapsis, 2018: Sequential sampling strategy for extreme event statistics in nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, **115 (44)**, 11 138–11 143, doi:10.1073/pnas.1813263115, URL https://www.pnas.org/content/115/44/11138, https://www.pnas.org/content/115/44/11138.full.pdf.

Nakamura, N., J. Falk, and S. W. Lubis, 2020: Why are stratospheric sudden warmings sudden (and intermittent)? *Journal of the Atmospheric Sciences*, **77 (3)**, 943 – 964, doi:10.1175/JAS-D-19-0249.1, URL https://journals.ametsoc.org/view/journals/atsc/77/3/jas-d-19-0249.1.xml.

Noé, F., C. Schütte, E. Vanden-Eijnden, L. Reich, and T. R. Weikl, 2009a: Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proceedings of the National Academy of Sciences*, **106 (45)**, 19 011–19 016, doi:10.1073/pnas.0905466106, URL https://www.pnas.org/content/106/45/19011, https://www.pnas.org/content/106/45/19011.full.pdf.

Noé, F., C. Schütte, E. Vanden-Eijnden, L. Reich, and T. R. Weikl, 2009b: Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proceedings of the National Academy of Sciences*, **106 (45)**, 19 011–19 016, doi:10.1073/pnas.0905466106, URL https://www.pnas.org/content/106/45/19011, https://www.pnas.org/content/106/45/19011.full.pdf.

Oksendal, B., 2003: *Stochastic Differential Equations: An Introduction with Applications*. Springer.

Pavliotis, G. A., 2014: *Stochastic processes and applications*. Springer.

Pedregosa, F., and Coauthors, 2011: Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, **12**, 2825–2830.

Plotkin, D. A., R. J. Webber, M. E. O'Neill, J. Weare, and D. S. Abbot, 2019: Maximizing simulated tropical cyclone intensity with action minimization. *Journal of Advances in Modeling Earth Systems*, **11 (4)**, 863–891, doi:https://doi.org/10.1029/2018MS001419, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018MS001419, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2018MS001419.

Ragone, F., and F. Bouchet, 2020: Computation of extreme values of time averaged observables in climate models with large deviation techniques. *Journal of Statistical Physics*, **179 (5)**, 1637–1665, doi:10.1007/s10955-019-02429-7, URL https://doi.org/10.1007/s10955-019-02429-7.

Ragone, F., J. Wouters, and F. Bouchet, 2018: Computation of extreme heat waves in climate models using a large deviation algorithm. *Proceedings of the National Academy of Sciences*, **115 (1)**, 24–29, doi:10.1073/pnas.1712645115, URL https://www.pnas.org/content/115/1/24, https://www.pnas.org/content/115/1/24.full.pdf.

Ruzmaikin, A., J. Lawrence, and C. Cadavid, 2003: A simple model of stratospheric dynamics including solar variability. *Journal of Climate*, **16**, 1593–1600, doi:10.1175/2007JCLI2119.1.

Sapsis, T. P., 2021: Statistics of extreme events in fluid flows and waves. *Annual Review of Fluid Mechanics*, **53 (1)**, 85–111, doi:10.1146/annurev-fluid-030420-032810, URL https://doi.org/10.1146/annurev-fluid-030420-032810, https://doi.org/10.1146/annurev-fluid-030420-032810.

Simonnet, E., J. Rolland, and F. Bouchet, 2021: Multistability and rare spontaneous transitions in barotropic beta-plane turbulence. *Journal of the Atmospheric Sciences*, **78 (6)**, 1889 – 1911, doi:10.1175/JAS-D-20-0279.1, URL https://journals.ametsoc.org/view/journals/atsc/78/6/JAS-D-20-0279.1.xml.

Stephenson, D. B., B. Casati, C. A. T. Ferro, and C. A. Wilson, 2008: The extreme dependency score: a non-vanishing measure for forecasts of rare events. *Meteorological Applications*, **15 (1)**, 41–50, doi:https://doi.org/10.1002/met.53, URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/met.53, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/met.53.

Strahan, J., A. Antoszewski, C. Lorpaiboon, B. P. Vani, J. Weare, and A. R. Dinner, 2021: Long-time-scale predictions from short-trajectory data: A benchmark analysis of the trp-cage miniprotein. *Journal of Chemical Theory and Computation*, **17 (5)**, 2948–2963, doi:10.1021/acs.jctc.0c00933, URL https://doi.org/10.1021/acs.jctc.0c00933, pMID: 33908762, https://doi.org/10.1021/acs.jctc.0c00933.

Thiede, E., D. Giannakis, A. R. Dinner, and J. Weare, 2019: Approximation of dynamical quantities using trajectory data. *arXiv:1810.01841 [physics.data-an]*, 1–24, doi:1810.01841.

Vanden-Eijnden, E., 2006: *Transition Path Theory*, 453–493. Springer Berlin Heidelberg, Berlin, Heidelberg, doi:10.1007/3-540-35273-2_13, URL https://doi.org/10.1007/3-540-35273-2_13.

Vanden-Eijnden, E., and J. Weare, 2013: Data assimilation in the low noise regime with application to the kuroshio. *Monthly Weather Review*, **141 (6)**, 1822–1841, doi:10.1175/MWR-D-12-00060.1.

Vitart, F., and A. W. Robertson, 2018: The sub-seasonal to seasonal prediction project (s2s) and the prediction of extreme events. *npj Climate and Atmospheric Science*, **1 (1)**, 3.

Weare, J., 2009: Particle filtering with path sampling and an application to a bimodal ocean current model. *Journal of Computational Physics*, **228 (12)**, 4312 – 4331, doi:https://doi.org/10.1016/j.jcp.2009.02.033.

Webber, R. J., D. A. Plotkin, M. E. O'Neill, D. S. Abbot, and J. Weare, 2019: Practical rare event sampling for extreme mesoscale weather. *Chaos*, **29 (5)**, 053 109, doi:10.1063/1.5081461.

Yasuda, Y., F. Bouchet, and A. Venaille, 2017: A new interpretation of vortex-split sudden stratospheric warmings in terms of equilibrium statistical mechanics. *Journal of the Atmospheric Sciences*, **74 (12)**, 3915–3936, doi:10.1175/JAS-D-17-0045.1.

Yoden, S., 1987: Dynamical Aspects of Stratospheric Vacillations in a Highly Truncated Model. *Journal of the Atmospheric Sciences*, **44 (24)**, 3683–3695, doi:10.1175/1520-0469(1987)044<3683:DAOSVI>2.0.CO;2, URL https://doi.org/10.1175/1520-0469(1987)044<3683:DAOSVI>2.0.CO;2.