Data-driven transition path analysis yields a statistical understanding of sudden stratospheric warming events in an idealized model

Justin Finkel*

Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology

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

**Corresponding author*: Justin Finkel, ju26596@mit.edu

ABSTRACT

Atmospheric regime transitions are highly impactful as drivers of extreme weather events, but pose two formidable modeling challenges: predicting the next event (weather forecasting), and characterizing the statistics of events of a given severity (the risk climatology). Each event has a different duration and spatial structure, making it hard to define an objective "average event." We argue here that transition path theory (TPT), a stochastic process framework, is an appropriate tool for the task. We demonstrate TPT's capacities on a wave-mean flow model of sudden stratospheric warmings (SSWs) developed by Holton and Mass (1976), which is idealized enough for transparent TPT analysis but complex enough to demonstrate computational scalability. Whereas a recent article (Finkel et al. 2021) studied near-term SSW predictability, the present article uses TPT to link predictability to long-term SSW frequency. This requires not only forecasting forward in time from an initial condition, but also *backward in time* to assess the probability of the initial conditions themselves. TPT enables one to condition the dynamics on the regime transition occurring, and thus visualize its physical drivers with a vector field called the *reactive current*. The reactive current shows that before an SSW, dissipation and stochastic forcing drive a slow decay of vortex strength at lower altitudes. The response of upper-level winds is late and sudden, occurring only after the transition is almost complete from a probabilistic point of view. This case study demonstrates that TPT quantities, visualized in a space of physically meaningful variables, can help one understand the dynamics of regime transitions.

1. Introduction

Many features of the atmosphere-ocean system's large-scale variability can be viewed as transitions between qualitatively different regimes. Examples include blocking, monsoons, El Niño, and Sudden Stratospheric Warming events (SSWs, the subject of this paper), all of which are associated with extreme weather. From a scientific perspective, regime transitions are handles by which to probe the climate's nonlinear, non-equilibrium dynamics. They expose novel physics and push us to qualitatively expand our physical understanding. From a human perspective, these relatively rare anomalies pose major societal challenges (Lesk et al. 2016; Kron et al. 2019), especially with a changing climate and increasing reliance on weather-susceptible infrastructure (e.g., Mann et al. 2017; Frame et al. 2020).

Regime transitions are used as benchmarks for model development across a hierarchy, from state-of-the-art Earth system models with billions of variables (e.g., Stephenson et al. 2008; Lengaigne and Vecchi 2010; Vitart and Robertson 2018) to conceptual low-order models with fewer than ten variables (e.g., Charney and DeVore 1979; Timmermann et al. 2003; Ruzmaikin et al. 2003; Crommelin et al. 2004; Thual et al. 2016). In Finkel et al. (2021), we addressed near term forecasting of regime transitions in the context of an idealized sudden stratospheric warming (SSW) model constructed by Holton and Mass (1976), which possesses two metastable states: a strong-vortex regime and a weak-vortex regime. The present paper's chief goal is to address questions about the long-term climate statistics of rare events by way of a case study on SSW-like regime transitions in the Holton-Mass model: how often do they occur, what are their typical development pathways, and how variable are those pathways between events?

We will use the framework of transition path theory (TPT; E and Vanden-Eijnden 2006), which offers a concise set of quantities to answer these questions. An SSW event is represented as a *transition path* from the strong vortex regime, which we denote state A, to the weak vortex regime, state B. The main quantity of interest will be the *reactive current* $\mathbf{J}_{AB}$, defined in section 3, which specifies the flow of probability density through state space *conditioned on an* $A \to B$ *transition event being underway*. To properly implement that conditional statement, we will need two auxiliary quantities. First, the *forward committor* $q_B^+(\mathbf{x})$ gives the probability that the system, initialized in a state $\mathbf{x}$, next reaches B before A. This is a measure of progress toward SSW: what is the probability of observing a SSW before returning to the strong vortex climatology? Second, the *backward committor* $q_A^-(\mathbf{x})$ gives the probability, looking backward in time, that the system visited A more recently than B, i.e., the model was last in the meta-stable strong vortex climatology, as opposed to just recovering from a recent SSW.

The forward committor itself was a primary focus of Finkel et al. (2021), where we pursued forecasting as the main objective. Committor probabilities are generally gaining traction as a metric for weather prediction; see Tantet et al. (2015) for an application to atmospheric blocking, Lee et al. (2018) for an application to tropical cyclone downscaling, Lucente et al. (2022) for an application to El Niño, and Miloshevich et al. (2022) for an application to heat waves. However, in the present paper we are pursuing climatological statistics rather than forecasting probabilities, using the committor only as an intermediate calculation for the reactive current, which characterizes the full transition process from A to B rather than its "forward half" from $\mathbf{x}$ to B.

Some previous studies (Crommelin 2003; Tantet et al. 2015) have visualized what are essentially reactive currents for blocking events in an observable subspace of leading EOFs. However, these studies were not couched in the language of TPT, a formalism that brings more quantitative results. Namely, the reactive current $\mathbf{J}_{AB}$ provides a direct estimate of the SSW rate, decomposing it over a continuous probability distribution of pathways. Formal TPT has not yet been widely taken up by the atmosphere-ocean science community, besides a few exceptions (Finkel et al. 2020; Miron et al.

2021, 2022). Part of our goal here is to encourage a common quantitative language for discussing regime transitions, which could help to organize several existing lines of research.

$\mathbf{J}_{AB}$, like q_B^+, can be expressed as a function of any observable subspace for visual exploration, with the complementary subspace treated as random variables. It is most enlightening to use observables with concrete physical meaning. A recent article Miloshevich et al. (2022) exploited this property to interpret a neural-network-learned committor for heat waves in terms of geopotential height and soil moisture, thus quantifying their predictive power at various lead times. In Finkel et al. (2021), we visualized the committor and expected lead time in a two-dimensional subspace consisting of zonal wind U, an index for polar vortex strength, and vertically integrated heat flux (IHF), which roughly measures the amplitude and phase tilt of vortex-disrupting planetary waves. Here we continue to use those coordinates, but also introduce a new subspace based on the zonal-mean meridional potential vorticity (PV) gradient and eddy enstrophy. These two quantities obey a conservation law in the absence of dissipation and stochastic forcing, a slight variation of the Eliassen-Palm relation. This allows us to diagnose more precisely the crucial roles of dissipation and stochastic forcing in driving the transition process, an important step toward understanding their causal relationship. Other kinds of atmospheric regime transitions will have different relevant physical diagnostics, any of which can be seen as an independent variable for the committor function and reactive current.

This paper is organized as follows. In section 2 we review the dynamical model. In section 3 we visualize the evolution of SSW events using the probability current, and introduce the key quantities for TPT—committors, densities, and currents—along with a brief summary of the method to compute them, which is more thoroughly explained in the supplementary document. In section 4, we use reactive current to construct a composite SSW evolution, and compare this to the standard composite method. In section 5, we change coordinates to better examine the dynamics of SSW events. We assess future directions and conclude in section 6.

2. A stochastically forced Holton-Mass model of SSW dynamics

We use exactly the same model as in Finkel et al. (2021), which is presented here for completeness.

a. Model specification

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, a meridional extent of 60°, and a height of 70 km, with the coordinate z ranging from 0 at the bottom of the domain (the tropopause) to 70 km at the top of the domain. $\overline{u}$ 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 is 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. A non-dimensionalized version of the equations is as follows, rearranged slightly from Finkel et al. (2021). The mean flow

$U(z,t)$ satisfies

$$\frac{2}{(\varepsilon\ell)^2}\partial_t\Big[\mathcal{G}^2\beta+\varepsilon\big(\mathcal{G}^2\ell^2U+U_z-U_{zz}\big)\Big] \tag{3a}$$
$$=\frac{2}{\varepsilon\ell^2}e^z\partial_z\big[e^{-z}\alpha\partial_z(U-U^R)\big]$$
$$+ke^z\mathrm{Im}\{\Psi^*\Psi_{zz}\}$$

with boundary conditions

$$U(z=0)=U^R(z=0)=10\,\mathrm{m/s}$$

$$U_z(z=z_{\mathrm{top}})=U_z^R(z=z_{\mathrm{top}})=\gamma/1000$$

while the perturbation streamfunction amplitude $\Psi(z,t)$ satisfies

$$(\partial_t+ik\varepsilon U)\Big[-\mathcal{G}^2(k^2+\ell^2)-\frac{1}{4}+\partial_z^2\Big]\Psi \tag{3b}$$
$$+ik\Psi\Big[\mathcal{G}^2\beta+\varepsilon\big(\mathcal{G}^2\ell^2U+U_z-U_{zz}\big)\Big]$$
$$=-\left(\partial_z-\frac{1}{2}\right)\left[\alpha\left(\partial_z+\frac{1}{2}\right)\Psi\right]$$

with boundary conditions

$$\Psi(z=0)=\frac{gh}{f_0}$$

$$\Psi(z=z_{\mathrm{top}})=0.$$

We have defined the nondimensional parameter $\mathcal{G}^2 := H^2N^2/(f_0^2L^2)$, where f_0 is the coriolis parameter at 60°N, $N^2 = 4\times10^{-4}$ is the the stratification, and $L = 2.5\times10^5$ km is a horizontal length scale chosen to make non-dimensionalized U and Ψ variables have similar climatological variances. The linear relaxation towards $U^R(z) = 10\,\mathrm{m/s} + (\gamma/1000)z$ on the right-hand side of Eq. (3a) is the force that maintains the typically strong polar vortex. Here $\gamma = 1.5$ m s^{-1} km^{-1}. The relaxation is mediated by a Newtonian cooling profile $\alpha(z)$, which is plotted in Fig. 1a, in its original dimensional units. Meanwhile, the lower boundary condition on Ψ comes from a bottom topography $h\cos(kx)$, where $h = 38.5$ m. This serves as a source of planetary waves.

There are two differences from Finkel et al. (2021), besides rearrangement. First, Finkel et al. (2021) had an erroneous but inconsequential negative sign in front of U^R_{zz} (their Eq. 3) which is corrected in Eq. (3a). Second, the left side of Eq. (3b) has two terms, $\pm ik\varepsilon\mathcal{G}^2\ell^2 U\Psi$, which could be cancelled out; we have retained them both to maintain a term-by-term correspondence with the original QGPV equation,

$$(\partial_t + \overline{u}\partial_x)q' + v'\partial_y\overline{q} = \text{ sources} - \text{sinks}, \tag{4}$$

$$\text{where } q' = \nabla^2\psi' + \frac{f_0^2}{N^2}e^{z/H}\partial_z(e^{-z/H}\psi') \tag{5}$$

$$\text{and } v' = \partial_x\psi' \tag{6}$$

which will be important when deriving the enstrophy budget in section 5.

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

$$\mathbf{X}(t) = \Big[\mathrm{Re}\{\Psi(t)\}, \mathrm{Im}\{\Psi(t)\}, U(t)\Big] \in \mathbb{R}^{75} \tag{7}$$

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 smaller-scale Rossby and gravity waves, initial condition uncertainties, and sources of model error, an approach originally put forward by Birner and Williams (2008) and used more recently by Esler and Mester (2019). The forcing is white in time, giving an Itô diffusion

$$d\mathbf{X}(t) = \boldsymbol{v}(\mathbf{X}(t))\,dt + \boldsymbol{\sigma}(\mathbf{X}(t))\,d\mathbf{W}(t) \tag{8}$$

where $\boldsymbol{v}(\mathbf{x})$ (not to be confused with meridional wind velocity, v) is the drift function determined by Eqs. (3). $\mathbf{W}(t)$ is an $(m+1)$-dimensional white-noise process, and $\boldsymbol{\sigma} \in \mathbb{R}^{d\times(m+1)}$ is a matrix specifying the spatially smooth structure of the noise as Fourier modes in the vertical. $\boldsymbol{\sigma}$ could depend on the state vector $\mathbf{X}$, but for simplicity we fix it to a constant, defined as follows. 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

$$\delta U(z) = \sigma_U \sum_{k=0}^{m} \eta_k \sin\left[\left(k + \frac{1}{2}\right)\pi \frac{z}{z_{\mathrm{top}}}\right]\sqrt{\delta t} \tag{9}$$

where $\sigma_U = 1$ m s^{-1} day$^{-1/2}$, whose units reflect the quadratic variation of Brownian motion. The numerical scheme is known as Euler-Maruyama (see, e.g., Pavliotis 2014, ch. 5). Equation 9 fully defines the matrix $\boldsymbol{\sigma}$. For $k = 0, \dots, m$, the kth column starts with 50 zeros, since there is no forcing on $\mathrm{Re}\{\Psi\}$ or $\mathrm{Im}\{\Psi\}$. The last 25 entries are evenly spaced samples of the sinusoidal factor in Eq. (9), all times σ_U.

The specific choice of stochastic forcing does affect the transition path statistics, but our method can be applied to any stochastic forcing. Because of the nonlinear coupling between $U(z)$ and $\Psi(z)$ in Eqs. (3a) and (3b), the noise injected to U feeds to Ψ after a single timestep.

b. Diagnostics

Until section 5, we use two main diagnostics for visualization, the same as in Finkel et al. (2021). The first is zonal wind strength $U(z)$, an index for vortex strength which is used to define regimes A and B. The second is the meridional eddy heat flux $\overline{v'T'}(z)$, which quantifies the heat being advected into the polar region associated with the sudden warming, and in the quasi-geostrophic limit, the vertical propagation of Rossby waves. In the Holton-Mass model, this takes the form

$$\overline{v'T'}(z) = \frac{Hf_0}{R}\overline{\frac{\partial\psi'}{\partial x}\frac{\partial\psi'}{\partial z}} \propto e^{z/H}|\Psi(z)|^2\frac{\partial\varphi}{\partial z}, \tag{10}$$

where R is the ideal gas constant for dry air and φ is the phase of the complex-valued streamfunction 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. We also use the density-weighted vertical integral of heat flux,

$$\mathrm{IHF}(z) := \int_0^z e^{-z/H}\overline{v'T'}(z')\,dz' \tag{11}$$

which varies more smoothly than $\overline{v'T'}$ at any single altitude.

c. Bistability

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. Detailed bifurcation analysis by Yoden (1987a) and Christiansen (2000) found a range of values for bottom topography h that create bistability. Figure 1(b,c) 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, then a burst of wave activity can propagate upward, ripping apart the polar vortex and causing zonal wind to collapse (Charney and Drazin 1961; Yoden 1987b). 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. Holton and Mass (1976); Yoden (1987b); Christiansen (2000), and Finkel et al. (2021) contain further details.

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:

$$A = \{\mathbf{x} \in \mathbb{R}^d : U(30\text{ km})(\mathbf{x}) \geq 53.8\text{ m/s}\} \tag{12a}$$

$$B = \{\mathbf{x} \in \mathbb{R}^d : U(30\text{ km})(\mathbf{x}) \leq 1.75\text{ m/s}\} \tag{12b}$$

where the velocity thresholds correspond to the vortex strength at 30 km for the fixed points **a** and **b**, respectively.

An 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), because 30 km is where the zonal wind profile of **b** reaches a minimum; Christiansen (2000) used this same coordinate when studying the same model. (The standard 10 hPa pressure level would correspond to $z = -7\,\text{km} \times \log(10/1000) - 10\,\text{km} \approx 22$ km above the troposphere in this model.) We also modify the zonal wind thresholds order to ensure that $\mathbf{a} \in A$ and $\mathbf{b} \in B$.

An important consequence of our A and B definitions is that the $A \to B$ transition path takes ~ 80 days. By design, this includes the slow initial *preconditioning* stage of vortex breakdown in advance of the ~ 10-day time horizon that traditionally comprises an SSW event. In this paper, 'SSW event' should be interpreted as both the preconditioning and the ensuing vortex collapse.

Figure 2 shows timeseries of U and $\overline{v'T'}$ at several different altitudes as the system goes through several transition paths in a long simulation. As in Fig. 2 of Finkel et al. (2021), orange strips denote $A \to B$ transitions while green strips denote $B \to A$ transitions. The long periods in between, which we call the $A \to A$ and $B \to B$ phases, demonstrate the bistable nature of regimes A and B. The fleeting $A \to B$ phase, however, is what we seek to understand. 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. The following section will introduce the *reactive density* $\pi_{AB}(\mathbf{x})$ and associated *reactive current* $\mathbf{J}_{AB}(\mathbf{x})$ which help us visualize the transition as a path distribution through state space and make the foregoing observations more quantitative.

3. The reactive density and reactive current: A distribution over transition paths

We consider the long-time behavior of our stochastic Holton-Mass model $\mathbf{X}(t)$ undergoing transitions between states A and B. Aggregating together statistics from only the transition paths yields a probability distribution called the *reactive density* $\pi_{AB}(\mathbf{x})$, defined such that

$$\begin{aligned}\pi_{AB}(\mathbf{x})\,d\mathbf{x} = \mathbb{P}\{&\mathbf{X}(t) \in d\mathbf{x} | \mathbf{X}(t) \text{ is in} \\ &\text{transition from } A \text{ to } B\}\end{aligned} \tag{13}$$

where $d\mathbf{x}$ is a small region about $\mathbf{x}$. One could estimate π_{AB} by binning samples from a long simulation, but including only those samples in transit directly from A to B. Associated to π_{AB} is a vector field called the *reactive current* $\mathbf{J}_{AB}(\mathbf{x})$, which quantifies the probability flux passing through $\mathbf{x}$ per unit time only during transition paths. Roughly speaking, π_{AB} specifies where transition paths go, and $\mathbf{J}_{AB}$ specifies how they move. Below we define them formally, but Fig. 3(a-c) gives some intuition by projecting them on the subspace (U, IHF) at $z = 10$, 20, and 30 km. Background shading indicates the strength of π_{AB}, and arrows indicate the magnitude and direction of $\mathbf{J}_{AB}$. Overlaid in thin blue lines are ten randomly sampled transition paths from the long ergodic simulation. These sample paths cluster in the same regions of state space identified as high-probability under π_{AB}, and on average flow along the arrows, corroborating qualitatively that $\pi_{AB}(\mathbf{x})$ and $\mathbf{J}_{AB}$ describe the location and evolution of the model in state space.

The transition path ensemble shows marked differences between altitudes. At z =10 km, the vortex strength (U) of states **a** and **b** is about the same, but the IHF is very distinct. The reactive current aligns with the IHF axis. Mathematically, this reflects the lower boundary condition $U(z = 0) = U^R(z = 0)$. Physically, this means that the heat flux due to the wave is the dominant physical process, with only small changes in zonal wind strength. The higher altitude of z = 30 km, by contrast, exhibits a large reduction in zonal wind strength, but only in the late stages of the process. In fact, the pattern of reactive density π_{AB} at z = 30 km (panel c) tells us that this final deceleration is quite sudden: the magnitude of π_{AB} is large near A, meaning transition paths linger there for a long time and only slowly crawl downward and to the right. But at the point IHF(30 km) $\approx 2.5\times10^4$ K·m/s, U(30 km) $\approx$ 30 m/s (the region marked by a dotted circle in panels c and f), π_{AB} reduces in magnitude and the reactive current spreads out widely as it turns downward toward set B. This is a signal that the transition paths are becoming both faster and more variable.

As a further point of comparison with $\mathbf{J}_{AB}$, we have plotted the minimum-action pathway from A to B with thick cyan lines (section 3 of the supplement specifies the numerical method). This represents the most likely transition path in the low-noise limit (e.g., Freidlin and Wentzell 1970; E et al. 2004; Forgoston and Moore 2018), and indeed it follows the direction of reactive current. With finite noise, however, the transition path ensemble spreads significantly around the minimum-action pathway, especially at the higher altitude of 30 km in the late stage of the transition process. Because of this, it is not possible for *any* single pathway, mininimum-action or not, to meaningfully represent the full ensemble.

We will show that the slow, initial phase of SSW involves *preconditioning* of the vortex: gradual erosion of the wind field by the stochastic forcing into a configuration that is especially susceptible to wave propagation. Once the wave burst is triggered, it imparts swift changes to the entire zonal wind profile. However, the bulk of SSW progress, probabilistically speaking, occurs in the preconditioning phase. Below we make this qualitative description precise by relating the reactive current to the forecast functions from Finkel et al. (2021): the committor and expected lead time metrics.

a. Mathematical relationship between current, committor, density, and rate

To formalize the description above and interpret the current rigorously, some definitions are in order, including a brief recap of the quantities from Finkel et al. (2021). 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):

$$q_B^+(\mathbf{x}) = \mathbb{P}_\mathbf{x}\{\mathbf{X}(\tau_{A\cup B}^+(t_0)) \in B\} \tag{14}$$

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 $S \subset \mathbb{R}^d$:

$$\tau_S^+(t_0) = \min\{t > t_0 : \mathbf{X}(t) \in S\}. \tag{15}$$

Like the expected lead time introduced below, the committor (under various aliases) predates TPT as an object of interest in the study of rare events (Du et al. 1998; Bolhuis et al. 2002). However, as we will see below, it is a key ingredient in any TPT analysis.

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 better examine climate change effects). Periodic- and finite-time TPT has been presented formally in Helfmann et al. (2020), and we have applied it to a dataset of state-of-the-art ensemble forecasts in Finkel et al. (2022). As a conceptual demonstration, however, the autonomous Holton-Mass model makes for a clearer exposition.

While $\tau^+_{A\cup B}$ itself is a random variable, one can take its expectation to obtain the *expected lead time* (to B),

$$\eta^+_B(\mathbf{x}) := \mathbb{E}_{\mathbf{x}}[\tau^+_{A\cup B}|\tau^+_B < \tau^+_A], \tag{16}$$

in other words, the expected time of arrival to B conditional on hitting B first. Finkel et al. (2021) described q^+_B and η^+_B in detail, as they are central quantities for forecasting, and visualized them in their Figs. 2c,d and 3c in the observable subspace (U, IHF). We do the same here, but additionally we overlay the reactive current. In Fig. 3(d,e,f), background shading represents the expected lead time and black contours represent committor level sets of 0.1, 0.2, 0.5, 0.8, and 0.9.

The committor's contour structure differs a lot between altitude levels. At 10 and 30 km (panels d and f), the contours have kinks. Depending on the initial condition, either a fluctuation in U or IHF might have a greater effect on the committor. The intermediate altitude of 10 km seems special in having committor contours that align with the IHF axis along the main channel of reactive current. In other words, $q^+_B(\mathbf{x})$ is well-approximated by a linear function of U(20 km), which is consistent with the finding in Finkel et al. (2021) that the 21.5-km altitude holds the most predictive power for q^+_B.

$\mathbf{J}_{AB}$ is related to q^+_B, generally flowing up the committor gradient. But $\mathbf{J}_{AB}$ contains some key information that the committor does not. As a *fore*cast function, the committor does not distinguish $A \to B$ transitions from $B \to B$ transitions, where the system leaves state B (beginning to recover), but then falls back to the weak-vortex state. To isolate the transition events from A to B, we need to introduce the *backward committor* (to A):

$$q^-_A(\mathbf{x}) = \mathbb{P}_{\mathbf{x}}\{\mathbf{X}(\tau^-_{A\cup B}(t_0)) \in A\} \tag{17}$$

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

$$\tau_S^-(t_0) = \max\{t < t_0 : \mathbf{X}(t) \in S\} \tag{18}$$

Intuitively, $q_A^-(\mathbf{x})$ is the probability of the system at point $\mathbf{x}$ last came from A, not B. The backward-in-time probabilities refer specifically to the process $\mathbf{X}(t)$ *in steady-state*, allowing us once again to set $t_0 = 0$. In other words, $q_A^-(\mathbf{x})$ depends explicitly on the *steady-state probability density* $\pi(\mathbf{x})$, where $\pi(\mathbf{x})\,d\mathbf{x} = \mathbb{P}\{\mathbf{X}(t) \in d\mathbf{x}\}$ is the long-term (climatological) probability of finding the system in a small region $d\mathbf{x}$ about $\mathbf{x}$.

Having defined both forward and backward committors, we can express the reactive density as

$$\pi_{AB}(\mathbf{x}) = \frac{1}{Z_{AB}}\pi(\mathbf{x})q_A^-(\mathbf{x})q_B^+(\mathbf{x}) \tag{19}$$

where Z_{AB} is a normalizing constant such that the right-hand side integrates to one. The associated reactive current can in turn be expressed

$$\mathbf{J}_{AB}(\mathbf{x}) = q_A^- q_B^+\big[\pi\boldsymbol{v} - \nabla\cdot(\mathbf{D}\pi)\big] \tag{20}$$

$$+\pi\mathbf{D}\big[q_A^-\nabla q_B^+ - q_B^+\nabla q_A^-\big], \tag{21}$$

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.

Eq. (21) is a specific expression for the current of a diffusion process of the form (8), which is the same general formulation as our model. But a more illuminating and general definition is its connection to the *rate*, or inverse return time, of the event (approximately (1700 days)$^{-1}$ for the Holton-Mass model with our chosen parameters). Let C be a closed hypersurface in $\mathbb{R}^d$ which encloses A and is disjoint with B; we call this a *dividing surface*. In the context of the diagrams in Fig. 3, C is any curve separating region A from region B. Then we have

$$\oint_C \mathbf{J}_{AB}\cdot\mathbf{n}\,dS = \text{Transition rate} \tag{22}$$

where $\mathbf{n}$ is an outward unit normal from C and dS is a surface area element. The integral relationship (22) 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 (see Vanden-Eijnden (2006) for a thorough mathematical explanation of $\mathbf{J}_{AB}$.) This constraint immediately implies a link between magnitude and width of $\mathbf{J}_{AB}$ streamlines. In Fig. 3(c,f), the strong magnitude of $\mathbf{J}_{AB}$ near $\mathbf{a}$ implies a thin central channel, and strict constraints on the mechanisms of early SSW onset. In other words, the initial preconditioning phase can only happen in a small number of ways. On the other hand, the subsequent weakening of $\mathbf{J}_{AB}$ between $q_B^+ = 0.5$ and $q_B^+ = 0.8$ (in the boxed region of Fig. 3c,f) implies that paths fan out across state space, becoming more variable. This spreading, or diversity of events, is only with respect to U and IHF at 30 km; at the lower altitudes, the current remains strong and narrow all the way through the transition process (Fig. 3, columns 1 and 2).

The reactive current and density characterize the transition path ensemble across the continuum of possible pathways, providing more information than the numerical value of the rate itself. Given any user-defined set of coordinates, the reactive current projection maps the transition paths in those coordinates, as a statistical ensemble with average behavior and variability. Below, following a brief note on the computational method, sections 4 and 5 demonstrate how to use reactive current and density to describe climatology and strengthen physical understanding of a rare transition event.

b. Computational method

The quantities presented in section 3, as well as the results to follow, could be computed directly by running a model for long enough to undergo a large number of SSW events and analyzing the statistics of those transitions. This procedure, which we call the "ergodic simulation" (ES) method, is possible in the 75-dimensional Holton-Mass model, and we have performed such a simulation of 10^6 days for validation purposes. However, this can be a major computational barrier in global climate models when the numerical integration is costly and the return period is long compared to the simulation timestep. Anticipating the need for fundamentally different techniques in high-dimensional state spaces, we have instead used the Dynamical Galerkin Approximation (DGA; Thiede et al. 2019; Strahan et al. 2021). A large collection of trajectories are launched in parallel with initial conditions distributed across state space, each one running for only a short time relative to the return period. Here we use 3×10^5 trajectories of length 20 days each, which is shorter than the 80-day duration of a single SSW event and much shorter than the 1700-day return period. Afterward, we assemble all these pieces together to estimate the quantities of interest, exploiting the Markov property. The total simulation time is not always reduced by this method—in our case, the short simulations total 6×10^6 days compared with the 1×10^6-day ES—but the format opens the door for many interesting possibilities, such as massive parallelization and adaptive sampling. In particular, as we show in Finkel et al. (2022), DGA is uniquely positioned to exploit large ensembles of short weather forecasts from high-fidelity operational models.

The basic DGA algorithm for rare event analysis has been described and tested in a recent series of articles (Thiede et al. 2019; Strahan et al. 2021; Finkel et al. 2021; Antoszewski et al. 2021). It is closely related to the "analogue Markov chain" approach of Lucente et al. (2021). Recently, an approach to learning neural network approximations of forecast functions using short trajectory data was introduced in Strahan et al. (2022). Due to the dependence on steady state and backward-in-time quantities, a full TPT analysis as carried out in this paper requires additional calculations beyond what is described in Finkel et al. (2021). We leave these details to the supplement in order to keep the focus on the results of our TPT analysis, which are robust with respect to algorithmic parameters.

4. SSW composites

Here we explain the traditional notion of a rare event ‘composite’ and contrast it with the composite intrinsically defined by TPT. The results are qualitatively similar, but the TPT description allows a rigorous mathematical connection to the reactive current and SSW rate.

The standard “composite” of an SSW event is a day-by-day aggregate of all the SSW events in a given dataset, aligned by the central warming date. This can include statistics, such as the mean and quantiles, of any observable function, such as the zonal-mean zonal wind or heat flux. Charlton and Polvani (2007) and Charlton et al. (2007) used this method to describe SSW climatology and establish benchmarks for stratosphere-resolving GCMs. We form a standard composite of U(30 km) from our Holton-Mass model in Fig. 4a, averaging together 300 events from a long ergodic simulation.

Here, we propose a complementary “TPT composite” based on reactive density. Instead of aligning events by the central warming date, we align the events by a general coordinate $f(\mathbf{x})$, which can be user-defined but must fulfill the minimal criterion of increasing from A to B, so it represents some objective notion of progress. At any progress level f_0, the TPT composite is defined by restricting the reactive density $\pi_{AB}(\mathbf{x})$ to the level set $\{\mathbf{x} : f(\mathbf{x}) = f_0\}$. Fixing $f = f_0$ is not the same as fixing the lead time τ_B^+, because the threshold might be crossed at different times by different transition paths. Note that $f(\mathbf{x})$ is a deterministic function of initial condition $\mathbf{x}$, unlike the hitting time τ_B^+, which is a random variable that changes between realizations launched from the same initial condition. Therefore, τ_B^+ cannot itself be used as a progress coordinate.

In Fig. 4b,c, we juxtapose alternative composites with the standard warming date coordinate $-\tau_B^+$. In panel b, we aggregate paths based on the negative expected lead time $-\eta_B^+$ defined above: the *expected* time until the central warming date. $-\eta_B^+$ is the deterministic progress function that is closest (in the mean-square sense) to the random progress function $t-\tau_B^+$ defining traditional composites. Panel c uses an altogether different progress metric, the committor q_B^+ itself, which increases from 0 on A to 1 on B.

The traditional and TPT composites are similar in shape, with an initially gradual decay in U(30 km) accelerating into a rapid decline in the final few days. As a function of $-\eta_B^+$, U(30 km) accelerates steadily through the whole transition, in both the traditional and TPT composites. But as a function of committor, U(30 km) decreases linearly at first and then accelerates downward between $q_B^+=0.6$ and $q_B^+=0.7$. According to the standard composite, U(30 km) becomes steadily less variable over time, with the whole ensemble collapsing into a single path by construction, as $t=0$ is the time of the event when U(30 km)$=0$. But when viewed as a function of expected lead time or committor, U(30 km) becomes more variable in the middle of the path, starting at $\eta_B^+\approx 50$ days or $q_B^+\approx 0.65$ and lasting until the end, when $\eta_B^+\to 0$ and $q^+\to 1$.

The same variability is reflected in Fig. 3c,f. In the boxed region, the reactive density weakens and the reactive current spreads out, some paths turning straight downward into B and others accumulating still more heat flux before making the plunge. The q_B^+ and η_B^+ contours in Fig. 3f convey geometrically how it is possible to have such wide variation in zonal wind strength even at a fixed expected lead time. Along the central channel of strong reactive current, where most of the transition paths flow, the committor and expected lead time have an approximately (negative) linear relationship. But in the weak-U flank of the current, especially in the boxed region, the q_B^+ level sets "unkink" to align with the IHF axis while the η_B^+ level sets turn downward to align with the U axis. The lowest visible level set of η_B^+ thus spans a range of vortex strengths of U(30 km).

Physically, the TPT composites are more variable than the traditional composite because $-\eta_B^+$, the expected lead time—a deterministic function—is a coarser description than $t-\tau_B^+$, a random variable. The former is an average over all realizations, while the latter takes on a specific value for each realization, which is not actually known until after the warming occurs. Given only information on the resolved variables $\Psi(z,t)$ and $U(z,t)$ at a given time, the TPT composite is the best one can do. The expected lead time quantifies SSW predictability, as established in Finkel et al. (2021). Here, we additionally incorporate the backward committor q_A^- via the reactive density π_{AB}, and so restrict focus to *transition* events—"major warmings"—from A to B.

As a loose analogy, a student's progress toward a degree can be measured objectively in course credits. On the other hand, first-year exams might weed out half of all students, which means that the *probabilistic* half-way point usually comes before half of required credits are done. A third metric, the time until graduation, can vary due to random effects like gap years and pandemics, which can cause a student to space their course load unevenly in time. Each cross-section of the student population—conditioning on a fixed number of credits completed, probability of graduation, or expected time until graduation—is a different statistical ensemble, each one conveying different information.

Going forward, we will use the committor as the progress coordinate of choice. That way, each point along the composite is an average over trajectories that are equally predictable in their probability to reach B, i.e., to proceed to an SSW. Often it is not just a singular coin toss that determines the fate of $\mathbf{X}(t)$, but a whole sequence of 'coin tosses'—random turns through state space—aligning in just such a way to navigate from A to B. With the committor as a progress coordinate, the 'coin tosses' are equidistributed along the horizontal axis, though they may not be equidistributed in time.

The same composite technique can be used to visualize the vertical wind structure at different stages. Fig. 5 plots $U(z)$ and $\overline{v'T'}(z)$ as altitude-indexed probability distributions at a series of committor level sets: $q_B^+ = 0.1$, 0.5, and 0.9. The widening variability with increasing committor is faintly visible at low altitudes, but increases dramatically above ~ 23 km, where at the $q_B^+ = 0.9$

level, the mean state (orange curve) falls well below the median state (central gray envelope.) This means the distribution of transition states is skewed left by a minority of paths with early collapse of upper-level winds. At the same committor range of 0.5-0.9, the vertical profile of meridional heat flux inflates dramatically. The altitude range of z = 20-25 km is the key transition region, below which zonal wind evolves relatively smoothly and with a symmetric distribution, and above which it varies rapidly with a skewed distribution. $\overline{v'T'}(z)$ is maximum near this altitude. We speculate that the underlying reason is the Newtonian cooling profile $\alpha(z)$, which has its own transition region centered at 25 km. It is not surprising that zonal wind just below, at 21.5 km, is an optimal linear predictor, as we found in Finkel et al. (2021).

5. A wave-mean flow interaction perspective

The previous section presented $\mathbf{J}_{AB}$ and π_{AB} as functions of two basic observables, zonal wind and integrated heat flux, and constructed a composite evolution of these observables. In this section, we incorporate more detailed physical knowledge to improve the interpretability of our TPT results. In particular, we manipulate the the dynamical equations to derive an enstrophy budget in the Holton-Mass model, which reveals a more natural set of coordinates that separates conservative from non-conservative processes. By visualizing the current in these coordinates, we identify physical drivers of each stage in the transition process. Our goal is twofold: first, to show how TPT can be formulated for any observables, and second, more narrowly in the context of this study, how the dynamics become more clear when those observables are well-chosen.

a. An eddy enstrophy formulation of the Holton-Mass model

A common diagnostic for wave-mean flow interaction systems is the wave activity, $\mathcal{A} = \rho_s\overline{q'^2}/(2\partial_y\overline{q})$, whose evolution is related to the Eliassen-Palm (EP) flux divergence (Andrews and McIntyre 1976). Yoden (1987b) used wave activity extensively to analyze the vacillating regime (our set B) of the Holton-Mass model, in particular the upward wave propagation that destabilizes the vortex. Below we derive a related set of equations for the eddy enstrophy, which enjoys a simpler balance equation and which we have found is better numerically suited for TPT analysis.

The first step in deriving the EP relation is to multiply the QGPV equation (4) by q' and take a zonal average, yielding

$$\partial_t\left(\frac{\overline{q'^2}}{2}\right)+\overline{v'q'}\partial_y\overline{q}=\overline{q'(\text{sources}-\text{sinks})} \tag{23}$$

We wish to work with the projected version of the equation, Eq. (3b), rather than the original PDE, to account for the approximation $\sin^2(\ell y)\approx\varepsilon\sin(\ell y)$ introduced by Holton and Mass (1976) for projecting quadratic nonlinearities. The procedure is summarized below, and spelled out more thoroughly in section 4 of the supplement.

Because of the ansatz (2), q' is represented in the projected equations by

$$\begin{aligned} q' &\longleftrightarrow \left[-\mathcal{G}^2(k^2+\ell^2)-\frac{1}{4}+\partial_z^2\right]\Psi \\ &=: (-\delta+\partial_z^2)\Psi \end{aligned} \tag{24}$$

where $\longleftrightarrow$ denotes correspondence between the full governing equations and the projected, non-dimensionalized equations in the Holton-Mass model. Recall that Ψ is the complex amplitude for the zonal-perturbation streamfunction $\psi'(x,y,z,t)$, in geostrophic balance with the wind (u,v).

As a general rule, the zonal average of the product of two wave quantities ψ_1' and ψ_2' of the form in Eq. (2).is found by the following formula:

$$\begin{aligned} \overline{\psi_1'\psi_2'} &= \overline{\mathrm{Re}\{\Psi_1 e^{ikx}\}\mathrm{Re}\{\Psi_2 e^{ikx}\}} \\ &= \mathrm{Re}\{\Psi_1^*\Psi_2\} \end{aligned} \tag{25}$$

Therefore, we multiply both sides of Eq. (3b) by the complex conjugate of (24) and take the real part to obtain

$$\partial_t \mathcal{E} + F_q \beta_e = D \tag{26a}$$

where

$$\begin{aligned} \mathcal{E} &= \frac{1}{2} e^z \left| \left(-\delta + \partial_z^2 \right) \Psi \right|^2 \\ &\longleftrightarrow \frac{1}{2} \overline{q'^2} \end{aligned} \tag{26b}$$

represents the eddy enstrophy;

$$\begin{aligned} F_q &= k e^z \mathrm{Im}\{\Psi^* \Psi_{zz}\} \\ &\longleftrightarrow \overline{v'q'} \end{aligned} \tag{26c}$$

represents the meridional eddy PV flux;

$$\begin{aligned} \beta_e &= \mathcal{G}^2 \beta + \varepsilon \left(\mathcal{G}^2 \ell^2 U + U_z - U_{zz} \right) \\ &\longleftrightarrow \partial_y \overline{q} \end{aligned} \tag{26d}$$

represents the meridional PV gradient; and

$$\begin{aligned} D = -\mathrm{Re}\Bigg\{ & e^z \left[\left(-\delta + \partial_z^2 \right) \Psi^* \right] \times \\ & \left(\partial_z - \frac{1}{2} \right) \left[\alpha \left(\partial_z + \frac{1}{2} \right) \Psi \right] \Bigg\} \\ \longleftrightarrow\ & \overline{q' (\text{sources} - \text{sinks})} \end{aligned}$$

represents the production and dissipation of enstrophy.

The standard EP relation would be found by dividing both sides by the meridional PV gradient β_e, as in Yoden (1987b). Instead, we next turn to the mean-flow equation (3a), which is an evolution equation for the PV gradient β_e rather than U directly. Multiplying through by β_e, we find

$$\partial_t \Gamma = R\beta_e + F_q\beta_e \tag{27a}$$

where

$$\Gamma := \left(\frac{\beta_e}{\varepsilon\ell}\right)^2 \tag{27b}$$

$$R := \frac{2}{\varepsilon\ell^2} e^z \partial_z \left[e^{-z} \alpha \partial_z (U - U_R) \right] \tag{27c}$$

The new quantity $\Gamma(z)$ is the squared meridional gradient of zonal-mean potential vorticity, which is highly correlated to zonal wind strength $U(z)$ in the Holton-Mass model. R is a relaxation coefficient for Γ, strengthening the vortex via radiative cooling.

The advantage of this alternative EP relation is now clear: adding together Eqs. (26) and (27), the meridional PV transport $F_q\beta_e$ cancels to give

$$\partial_t(\Gamma + \mathcal{E}) = R\beta_e + D. \tag{28}$$

In this form, all the dissipative effects are contained on the right-hand side via the cooling coefficient $\alpha(z)$, which appears both in D and R. $\Gamma + \mathcal{E}$ would conserved, at every altitude separately, in the absence of dissipation and stochastic forcing. In this limit, an increase in eddy enstrophy $\mathcal{E}$ can only occur at the expense of the mean PV gradient characterized by Γ. Of course, both non-conservative effects—dissipation and stochastic forcing—are critically important; vacillation cycles and transitions are possible only because the Holton-Mass model, like the full atmosphere, is an open system. The utility of Eq. (28) is to isolate those nonconservative effects as almost extrinsic inputs.

b. Using the reactive current to quantify the importance of non-conservative processes

Dissipation and forcing act to disrupt the conservation of $\Gamma+\mathcal{E}$, with a specific pattern shown in Fig. 6. The reactive current is shown at three altitudes, as in Fig. 3, but this time in the space $(\Gamma^{1/2},\mathcal{E}^{1/2})$ instead of (U, IHF). We take square roots because the visualizations are more clear, and the units of s^{-1} are more comparable with those of zonal wind $U(z)$ and radiative cooling $\alpha(z)$. (We note that the fixed point **b** in panel (d) appears to have committor < 1; this is possible when projecting out nonlinear coordinates because set B is defined based on the 30-km level, and the state-space regions that resemble **b** at 10 km may not resemble it at 30 km.) In the upper stratosphere, at $z=30$ km (panels c and f), the main channel of reactive current flows along a circular arc, approximately conserving $\Gamma+\mathcal{E}$, all the way through the $q_B^+=0.9$ surface: the evolution of an SSW is a nearly conservative interaction between waves and the mean flow right up to the end. Then, the current weakens in magnitude and spreads out, indicating the critical non-conservative processes at the end, where the breaking and dissipation of the anomalous waves cements the SSW event. Just as in the (U,IHF) space, the reactive density π_{AB} decreases along that circular arc, meaning the transition paths accelerate.

On the other hand, $\mathbf{J}_{AB}$ projected at $z=10$ km (panels a and d) shows that the dynamics are never conservative in the lower stratosphere: the initial motion points not along a circular arc but directly leftward, such that $\Gamma+\mathcal{E}$ is decreasing from the start. From the enstrophy budget (28), we conclude that a combination of dissipation and stochastic forcing acts strongly at 10 km to precondition the vortex. The next subsection shows that stochastic forcing plays the more decisive role.

Finally, consider the middle altitude of 20 km, where $\mathbf{J}_{AB}$ has a shape that is intermediate between the current at 10 and 30 km. It does not have distinctly positive or negative curvature, but flows along a straight channel from A to B. 20 km seems to be in just the right altitude range to feel significant dissipation and stochastic forcing—a feature of the lower boundary—but also to channel a good share of the loss of Γ to the gain of $\mathcal{E}$, a quasi-conservative property of the loftier 30 km. The resulting committor, expected lead time, and reactive current are approximately linear functions of $\Gamma^{1/2}(20\text{ km})$ and $\mathcal{E}^{1/2}(20\text{ km})$. Indeed, the wind and heat flux at 20 km were the most useful for prediction in (Finkel et al. 2021, their section 4).

Fig. 7a,b,c show the composite evolution of $\Gamma+\mathcal{E}$ in orange, along with Γ in blue and $\mathcal{E}$ in pink, at the same three altitudes 10, 20, and 30 km. All three altitudes show evidence of dissipation, with $\Gamma+\mathcal{E}$ weakening as the committor increases, but with distinct differences in the rates. The $\Gamma+\mathcal{E}$ composite is concave up at 10 km, implying dissipation is most important at the early stage, when the predictability of the event is limited. At 20 km, the composite is practically linear, implying that dissipation maintains a constant role in the event's evolution, gradually opening the valve to wave propagation at the last stage of the event. At 30 km, the composite is concave down: the flow is initially conservative, with exchange between mean flow and eddies at the onset of vortex breakdown, followed by strong dissipation of the waves when the event is all but assured.

At 20 and 30 km, the distribution of $\Gamma+\mathcal{E}$ begins symmetric, with the mean (orange) tracking the median (near the center of the dark gray band). Then between $q_B^+=0.6$ and 0.7, the lower tail of the distribution expands quickly, skewing the distribution negative. The distribution at 10 km maintains a slight negative skew for the entire transition path. The skewness reflects the occurrence of "minor warmings" preceding the SSW, when the vortex begins to break down, but partially recovers before the final event.

The composites, as well as the reactive currents, support the notion of the "typical" transition path as an initially non-conservative creep at low altitudes, opening up a valve to allow waves to propagate upward, finally yielding a very abrupt collapse at high altitudes follows after a long, mostly conservative phase. With the enstrophy budget (28), we can assess the importance of each term by plotting those composites as well. Fig. 7d,e,f show the composite evolution of each term at each altitude: $R\beta_e$ (the relaxation of the squared mean PV gradient, Γ) in blue, D (the dissipation of enstrophy, $\mathcal{E}$) in pink, and $\beta_e F_q$ (the transfer of enstrophy from Γ to $\mathcal{E}$) in black, all normalized by the total $\Gamma+\mathcal{E}$ at each level to account for the altitude-dependent differences in variability. This allows us to compare how strong each dissipative force is *relative* to the total budget. The sum $(R\beta_e+D)/(\Gamma+\mathcal{E})$—the normalized, deterministic tendency $\partial_t(\Gamma+\mathcal{E})/(\Gamma+\mathcal{E})$—is shown as a dashed orange curve. Note that this tendency is positive at 10 and 20 km even though $\Gamma+\mathcal{E}$ is actually decreasing. Without stochastic forcing, the system will always approach state **a** or **b**, depending on where the initial condition falls relative to the surface dividing the two attractors.

To quantify the critical role of stochastic forcing in effecting the transition at each committor level, we define the stochastic tendency of $\Gamma+\mathcal{E}$ along transition paths:

$$\mathcal{L}_{AB}[\Gamma+\mathcal{E}](\mathbf{x}) = \tag{29}$$

$$\lim_{\Delta t\to 0}\mathbb{E}\left[\frac{(\Gamma+\mathcal{E})(\mathbf{X}(t+\Delta t))-(\Gamma+\mathcal{E})(\mathbf{X}(t-\Delta t))}{2\Delta t}\right.$$
$$\left|\,\mathbf{X}(t)=\mathbf{x}\text{ and }\mathbf{X}(t)\text{ is in transition}\right] \tag{30}$$

which is related to the ordinary infinitesimal generator $\mathcal{L}$ (see Oksendal (2003) for mathematical background and the appendix of Finkel et al. (2021) for its application to the Holton-Mass model). The supplement describes the numerical procedure to approximate $\mathcal{L}_{AB}$ using short trajectories and a finite lag time. There, we show that $\mathcal{L}_{AB}f(\mathbf{x})$ is related to $\mathbf{J}_{AB}\cdot\nabla f(\mathbf{x})$ for any observable f, so it is appropriate to view the arrows in Fig. 3 and 6 as a proxy for the stochastic tendencies of the projected observables.

We introduce $\mathcal{L}_{AB}$ to compare with the deterministic tendency $\partial_t(\Gamma+\mathcal{E})(\mathbf{x})$, which for a diffusion process of the form (8) is simply $\boldsymbol{v}(\mathbf{x})\cdot\nabla(\Gamma+\mathcal{E})(\mathbf{x})$ by the chain rule. Their difference shows the impact of stochastic forcing responsible for transitions. More specifically, $\mathcal{L}_{AB}-\partial_t$ averaged over a committor level q_0 highlights the stochastic effects responsible for taking the system from q_0 to q_0+dq. Often it is not just a single coin flip that decides the fate of $\mathbf{X}(t)$, but a whole sequence of random turns through state space aligning in just such a way to navigate from A to B.

The role of stochasticity is most stark at 10 and 20 km (panels (d) and (e)) and for $q_B^+ < 0.5$, where $\mathcal{L}_{AB}(\Gamma+\mathcal{E})$ is negative while $\partial_t(\Gamma+\mathcal{E})$ is positive, due to a strong positive tug of radiative cooling versus the weak dissipation of enstrophy. As q_B^+ increases, the stochastic and deterministic tendencies grow closer together: the more likely the transition to B, the easier it is for deterministic drift to carry it out alone. At 30 km (panel f), all forms of dissipation and forcing start out *relatively* small compared to the magnitude of $\Gamma+\mathcal{E}$, but as the path progresses they all diverge away from zero. Most notably, the stochastic and deterministic tendencies never diverge very far; if anything, stochastic noise *slows* the collapse of U(30 km) at the end. It seems that to achieve the $A\to B$ transition, which is defined entirely in terms of U(30 km), the most common mechanism is a persistent negative push applied to lower altitudes, and this ultimately sets up the higher altitudes for more sudden, deterministic collapse after the "hard work" of eroding the vortex from below is mostly finished.

In summary, the TPT diagnostics have demonstrated that the SSW process begins with steady, significant decay of the PV gradient (here, its squared gradient, Γ) at lower altitudes, driven by the stochastic forcing, with only conservative changes taking place at higher altitudes. This preconditioning of the vortex opens up a valve to the mid-stratosphere. In the late stages of the transition, starting between $q_B^+ = 0.6$ and 0.7, the upper-level winds decline very suddenly. This begins conservatively as eddies grow, exchanging energy with the mean flow, and finishes non-conservatively, as friction dissipates the waves.

6. Conclusion

Transition path theory (TPT) is a mathematical framework that can be used to assess the near-term predictability and long-term climatology of anomalous weather events. The framework lends itself naturally to events associated with regime transitions, but it can be applied to more general anomalies. The key is to be able to define a suitable "reaction coordinate", or measure of progress, linking the event to the mean state. We have analyzed the statistical ensemble of Sudden Stratospheric Warmings (SSWs) in the idealized Holton-Mass model. Here, measures of the vortex strength (or the mean potential vorticity) and heat flux (eddy enstrophy) provide natural coordinates for applying the theory.

Probability densities and currents tell us how the system evolves through state space during a breakdown of the polar stratospheric vortex. The reactive current, $\mathbf{J}_{AB}$, allows one to condition dynamical tendencies on the occurrence of a rare event. By overlaying $\mathbf{J}_{AB}$ over observable sub-spaces at different altitudes in the stratosphere, we have identified the key roles of dissipation and stochastic forcing in driving SSWs in the Holton-Mass model. The stochastic driving represents the effects of unresolved Rossby and gravity waves that have been stripped from this highly truncated model. The action of these non-conservative processes, stochastic driving in particular, matter most at lower altitudes early in the transition process, conditioning the vortex, while the higher altitudes are shielded from significant dissipation. It is only late in the transition process, after the likelihood of the event has surpassed 60%, that the upper-level winds play a significant role in the dynamics.

This 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; Miron et al. 2021, 2022). TPT results are best interpreted when viewed in a physically meaningful observable subspace of variables. Utilizing physical knowledge and experience with the system allows one to gain the most from the methodology. With the rather simple Holton-Mass model, we identified such a subspace based on an enstrophy budget. In different versions of quasigeostrophic dynamics, the wave activity (Nakamura and Solomon 2010; Lubis et al. 2018) and other diagnostics based on the transformed-Eulerian-mean (Andrews and McIntyre 1976) are likely to be informative coordinates.

Significant challenges remain for deploying TPT analysis at scale to state-of-the-art climate models. We have used a Dynamical Galerkin Approximation (DGA) short trajectory analysis algorithm to compute TPT quantities. One important limitation of this computational pipeline is the data generation step. We used a long direct simulation to sample the background climatology, which served the double purpose of seeding initial data points for short trajectories and providing a ground truth for validating the accuracy of DGA. The former point is critical: one must cover the space of initial conditions to capture the dynamics of extreme events. In some cases, short trajectory data already exist, e.g., from the subseasonal-to-seasonal (S2S) database (Vitart and Robertson 2018), which we have used recently in Finkel et al. (2022) to estimate centennial-scale SSW rates from only 21 years of ensemble forecasts. In other cases, it is advantageous to generate fresh data in undersampled regions of state space, which would require more advanced sampling methods such as the adaptive sampling strategies proposed in Lucente et al. (2021) and Strahan et al. (2022), or rare event simulation schemes such as in Mohamad and Sapsis (2018), Ragone et al. (2018), Webber et al. (2019), and Ragone and Bouchet (2020).

Acknowledgments. During the time of writing, J.F. was 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 grants AGS-1852727 and OAC-2004572. 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 many helpful conversations and 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

Andrews, D. G., and M. E. McIntyre, 1976: Planetary waves in horizontal and vertical shear: The generalized eliassen-palm relation and the mean zonal acceleration. *Journal of Atmospheric Sciences*, **33 (11)**, 2031 – 2048, doi:10.1175/1520-0469(1976)033<2031:PWIHAV>2.0.CO;2, URL https://journals.ametsoc.org/view/journals/atsc/33/11/1520-0469_1976_033_2031_pwihav_2_0_co_2.xml.

Antoszewski, A., C. Lorpaiboon, J. Strahan, and A. R. Dinner, 2021: Kinetics of phenol escape from the insulin r6 hexamer. *The Journal of Physical Chemistry B*, **125 (42)**, 11 637–11 649, doi:10.1021/acs.jpcb.1c06544, URL https://doi.org/10.1021/acs.jpcb.1c06544, pMID: 34648712, https://doi.org/10.1021/acs.jpcb.1c06544.

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.

Bolhuis, P. G., D. Chandler, C. Dellago, and P. L. Geissler, 2002: Transition path sampling: Throwing ropes over mountain passes in the dark. *Annual Review of Physical Chemistry*, **53**, 291–318.

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.

Charlton, A. J., and Coauthors, 2007: A new look at stratospheric sudden warmings. part ii: Evaluation of numerical model simulations. *Journal of Climate*, **20 (3)**, 470–488, doi:10.1175/JCLI3994.1.

Charney, J. G., and J. G. DeVore, 1979: Multiple Flow Equilibria in the Atmosphere and Blocking. *Journal of the Atmospheric Sciences*, **36 (7)**, 1205–1216, doi:10.1175/1520-0469(1979)036<1205:MFEITA>2.0.CO;2, URL https://doi.org/10.1175/1520-0469(1979)036<1205:MFEITA>2.0.CO;2, https://journals.ametsoc.org/jas/article-pdf/36/7/1205/3420739/1520-0469(1979)036_1205_mfeita_2_0_co_2.pdf.

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.

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.

Crommelin, D. T., 2003: Regime transitions and heteroclinic connections in a barotropic atmosphere. *Journal of the Atmospheric Sciences*, **60 (2)**, 229 – 246, doi:10.1175/1520-0469(2003)060<0229:RTAHCI>2.0.CO;2, URL https://journals.ametsoc.org/view/journals/atsc/60/2/1520-0469_2003_060_0229_rtahci_2.0.co_2.xml.

Crommelin, D. T., J. D. Opsteegh, and F. Verhulst, 2004: A Mechanism for Atmospheric Regime Behavior. *Journal of the Atmospheric Sciences*, **61 (12)**, 1406–1419, doi:10.1175/1520-0469(2004)061<1406:AMFARB>2.0.CO;2, URL https://doi.org/10.1175/1520-0469(2004)061<1406:AMFARB>2.0.CO;2, https://journals.ametsoc.org/jas/article-pdf/61/12/1406/3472147/1520-0469(2004)061_1406_amfarb_2_0_co_2.pdf.

Du, R., V. S. Pande, A. Y. Grosberg, T. Tanaka, and E. S. Shakhnovich, 1998: On the transition coordinate for protein folding. *Journal of Chemical Physics*, **108 (1)**, 334–350.

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.

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., E. P. Gerber, D. S. Abbot, and J. Weare, 2022: Revealing the statistics of extreme events hidden in short weather forecast data. arXiv, URL https://arxiv.org/abs/2206.05363, doi:10.48550/ARXIV.2206.05363.

Finkel, J., R. J. Webber, E. P. Gerber, D. S. Abbot, and J. Weare, 2021: Learning forecasts of rare stratospheric transitions from short simulations. *Monthly Weather Review*, **149 (11)**, 3647 – 3669, doi:10.1175/MWR-D-21-0024.1, URL https://journals.ametsoc.org/view/journals/mwre/149/11/MWR-D-21-0024.1.xml.

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.

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.

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.

Lee, C.-Y., M. K. Tippett, A. H. Sobel, and S. J. Camargo, 2018: An environmentally forced tropical cyclone hazard model. *Journal of Advances in Modeling Earth Systems*, **10 (1)**, 223–241, doi:https://doi.org/10.1002/2017MS001186, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017MS001186, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2017MS001186.

Lengaigne, M., and G. A. Vecchi, 2010: Contrasting the termination of moderate and extreme el niño events in coupled general circulation models. *Climate Dynamics*, **35 (2)**, 299–313, doi:10.1007/s00382-009-0562-3, URL https://doi.org/10.1007/s00382-009-0562-3.

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.

Lubis, S. W., C. S. Y. Huang, and N. Nakamura, 2018: Role of finite-amplitude eddies and mixing in the life cycle of stratospheric sudden warmings. *Journal of the Atmospheric Sciences*, **75 (11)**, 3987 – 4003, doi:10.1175/JAS-D-18-0138.1, URL https://journals.ametsoc.org/view/journals/atsc/75/11/jas-d-18-0138.1.xml.

Lucente, D., C. Herbert, and F. Bouchet, 2022: Committor functions for climate phenomena at the predictability margin: The example of el niño southern oscillation in the jin and timmermann model. *Journal of the Atmospheric Sciences*, doi:10.1175/JAS-D-22-0038.1, URL https://journals.ametsoc.org/view/journals/atsc/aop/JAS-D-22-0038.1/JAS-D-22-0038.1.xml.

Lucente, D., J. Rolland, C. Herbert, and F. Bouchet, 2021: Coupling rare event algorithms with data-based learned committor functions using the analogue Markov chain. *arXiv preprint arXiv:2110.05050*.

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.

Miloshevich, G., B. Cozian, P. Abry, P. Borgnat, and F. Bouchet, 2022: Probabilistic forecasts of extreme heatwaves using convolutional neural networks in a regime of lack of data. arXiv, URL https://arxiv.org/abs/2208.00971, doi:10.48550/ARXIV.2208.00971.

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.

Miron, P., F. J. Beron-Vera, and M. J. Olascoaga, 2022: Transition paths of north atlantic deep water. *Journal of Atmospheric and Oceanic Technology*, **39 (7)**, 959 – 971, doi:10.1175/JTECH-D-22-0022.1, URL https://journals.ametsoc.org/view/journals/atot/39/7/JTECH-D-22-0022.1.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., and A. Solomon, 2010: Finite-amplitude wave activity and mean flow adjustments in the atmospheric general circulation. part i: Quasigeostrophic theory and analysis. *Journal of the Atmospheric Sciences*, **67 (12)**, 3967 – 3983, doi:10.1175/2010JAS3503.1, URL https://journals.ametsoc.org/view/journals/atsc/67/12/2010jas3503.1.xml.

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

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

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.

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.

Strahan, J., J. Finkel, A. R. Dinner, and J. Weare, 2022: Forecasting using neural networks and short-trajectory data. arXiv, URL https://arxiv.org/abs/2208.01717, doi:10.48550/ARXIV.2208.01717.

Tantet, A., F. R. van der Burgt, and H. A. Dijkstra, 2015: An early warning indicator for atmospheric blocking events using transfer operators. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **25 (3)**, 036 406, doi:10.1063/1.4908174, URL https://doi.org/10.1063/1.4908174, https://doi.org/10.1063/1.4908174.

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.

Thual, S., A. J. Majda, N. Chen, and S. N. Stechmann, 2016: Simple stochastic model for el niño with westerly wind bursts. *Proceedings of the National Academy of Sciences*,

113 (37), 10 245–10 250, doi:10.1073/pnas.1612002113, URL https://www.pnas.org/doi/abs/10.1073/pnas.1612002113, https://www.pnas.org/doi/pdf/10.1073/pnas.1612002113.

Timmermann, A., F.-F. Jin, and J. Abshagen, 2003: A nonlinear theory for el niño bursting. *Journal of the Atmospheric Sciences*, **60 (1)**, 152 – 165, doi:10.1175/1520-0469(2003)060<0152:ANTFEN>2.0.CO;2, URL https://journals.ametsoc.org/view/journals/atsc/60/1/1520-0469_2003_060_0152_antfen_2.0.co_2.xml.

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.

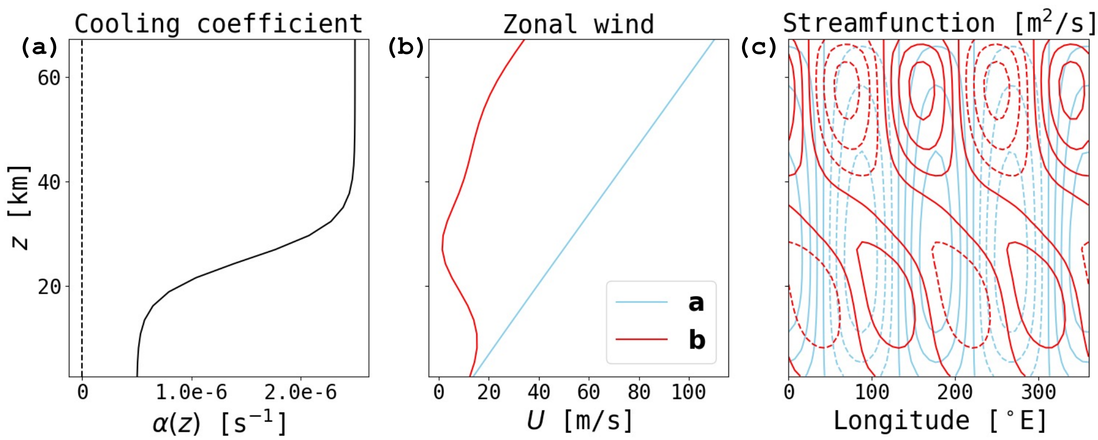
Vitart, F., and A. W. Robertson, 2018: The sub-seasonal to seasonal prediction project (s2s) and the prediction of extreme events. *npj Climate and Atmospheric Science*, **1 (1)**, 3.

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.

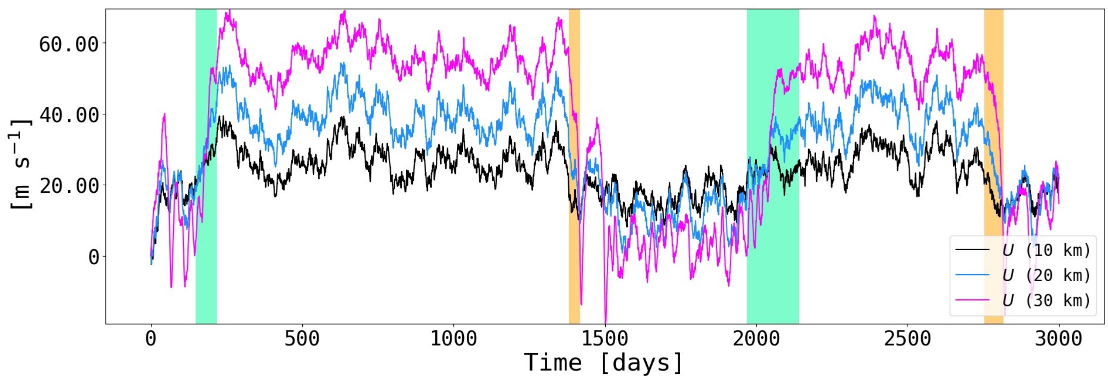
Yoden, S., 1987a: Bifurcation properties of a stratospheric vacillation model. *Journal of the Atmospheric Sciences*, **44 (13)**, 1723–1733, doi:10.1175/1520-0469(1987)044<1723:BPOASV>2.0.CO;2.

Yoden, S., 1987b: 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.

LIST OF FIGURES



859 FIG. 1. **Parameters and stable equilibria of the Holton-Mass model**. (a) The Newtonian cooling profile $\alpha(z)$. (b) Zonal-mean zonal wind $U(z)$ and (c) perturbation streamfunction $\psi'(x, 60°\text{N}, z)$, with contour spacing of 1.5×10^7 m^2/s. Dashed lines mean negative values. Blue indicates the strong vortex equilibrium, **a**, and red indicates the weak vortex equilibrium, **b**, as in Eqs. (12).



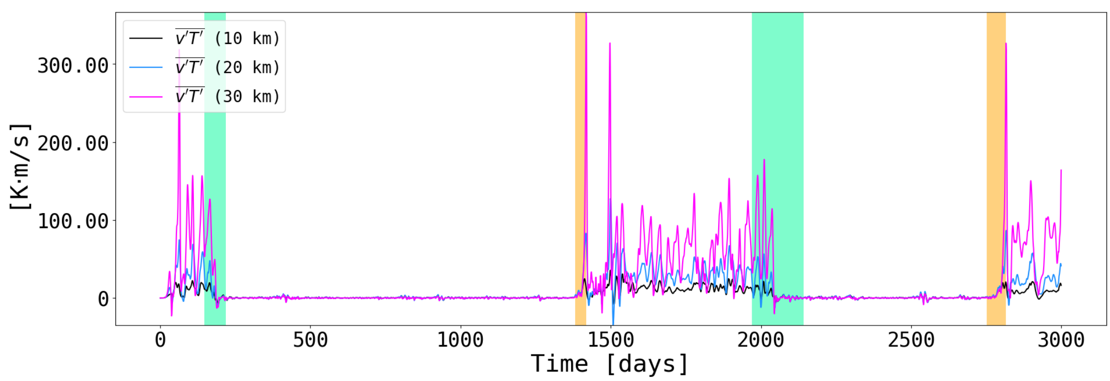


FIG. 2. **Regime transitions**. We plot (a) the zonal-wind strength U, and (b) the eddy heat flux $\overline{v'T'}$, over the first 3000 days of a long stochastic simulation. The quantities are evaluated at $z = 10, 20$, and 30 km. The time interval contains two transitions from A (a strong vortex) to B (a weak vortex) and back. $A \longrightarrow B$ transitions are highlighted in orange, and $B \longrightarrow A$ transitions are highlighted in green.

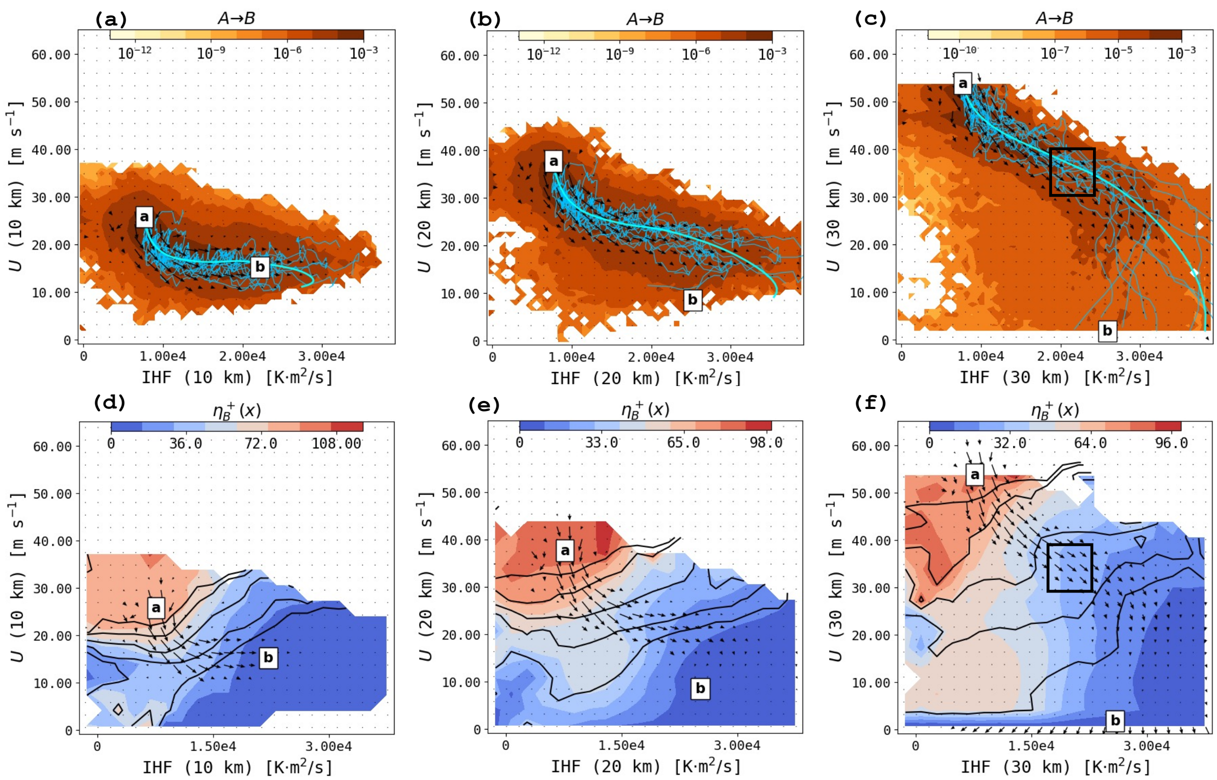


FIG. 3. **Currents, densities, committors, and expected lead times.** (a): Background shading is the reactive density π_{AB}, on a log scale. Thin blue lines are ten randomly selected transition paths from the long control simulation. Thick cyan curve is the minimum-action path from A to B. Also overlaid is a vector field representing reactive current $\mathbf{J}_{AB}$. The subspace is (U, IHF) evaluated at $z = 10$ km. Positions of the fixed points **a** and **b** are marked. Arrows represent $\mathbf{J}_{AB}$. (b, c): Same as (a), but at $z = 20$ and 30 km respectively. (d) The expected lead time η_B^+ is shaded as background color, and level sets of the committor q_B^+ 0.1, 0.2, 0.5, 0.8, and 0.9 are overlaid as black curves. (e, f): Same as (d), but at $z = 20$ km and 30 km respectively. A box marks a transition region between narrow, constrained current and wide, dispersed current. See text for a description.

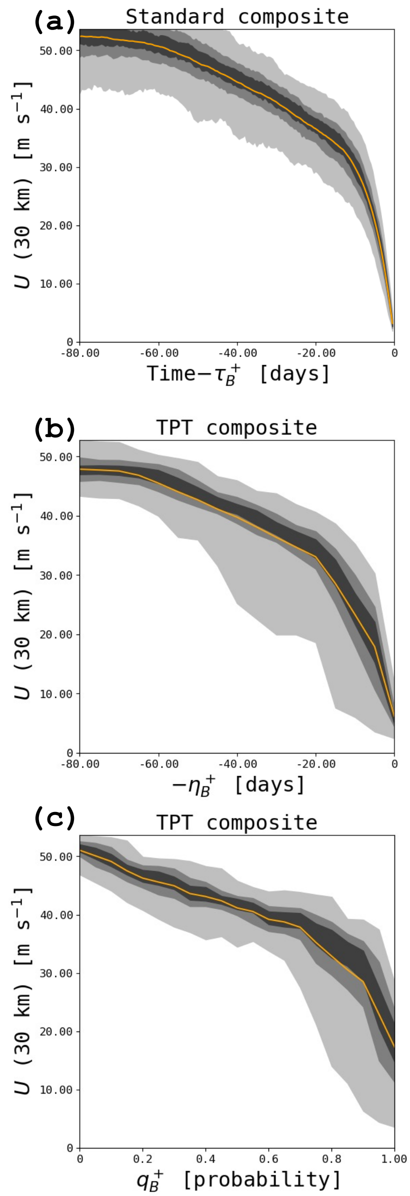


FIG. 4. **Composites evolution of SSW events**. Orange curves plot the mean value of U(30 km) at a given stage in the transition process; expanding gray envelopes show the middle 25-, 50-, and 90-percentile ranges. We use three different notions of progress: hitting time to B ($t-\tau_B^+$, panel a), expected hitting time to B ($-\eta_B$, panel b), and committor (q_B^+, panel c).

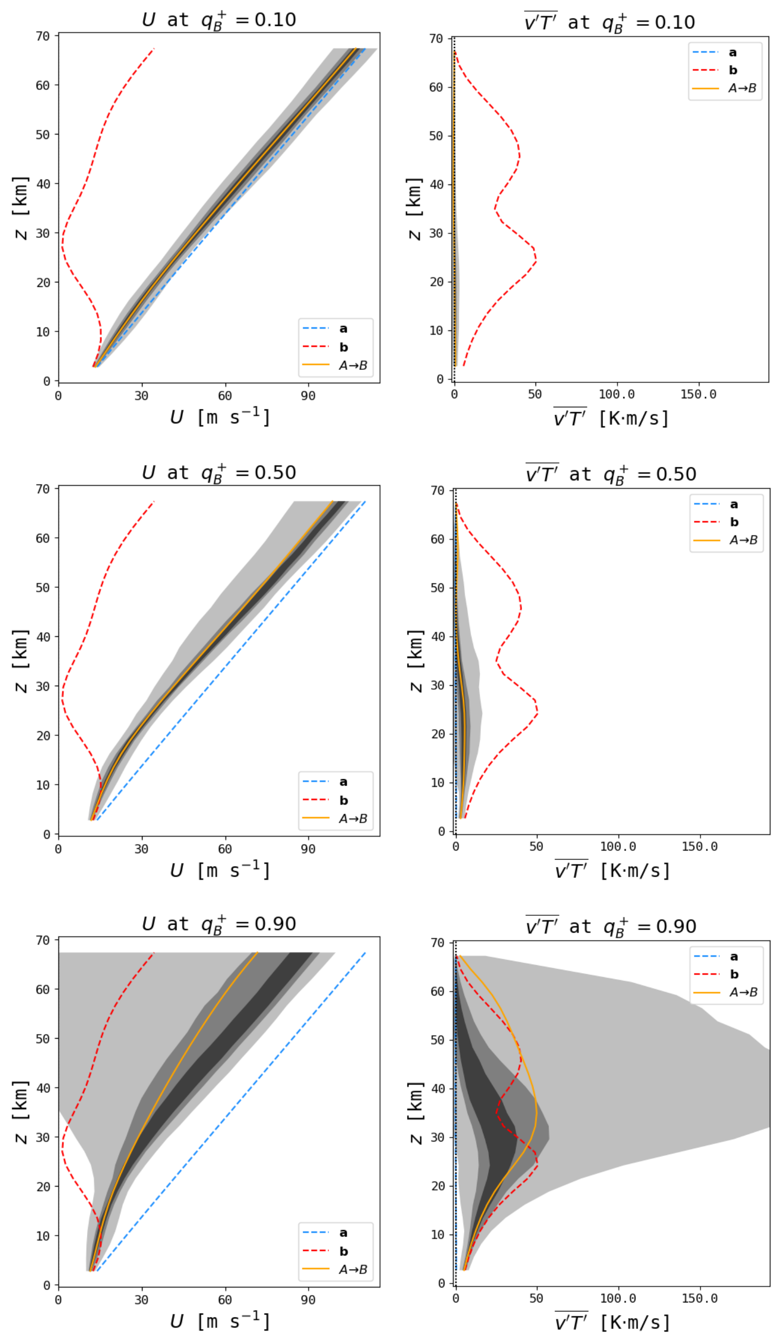
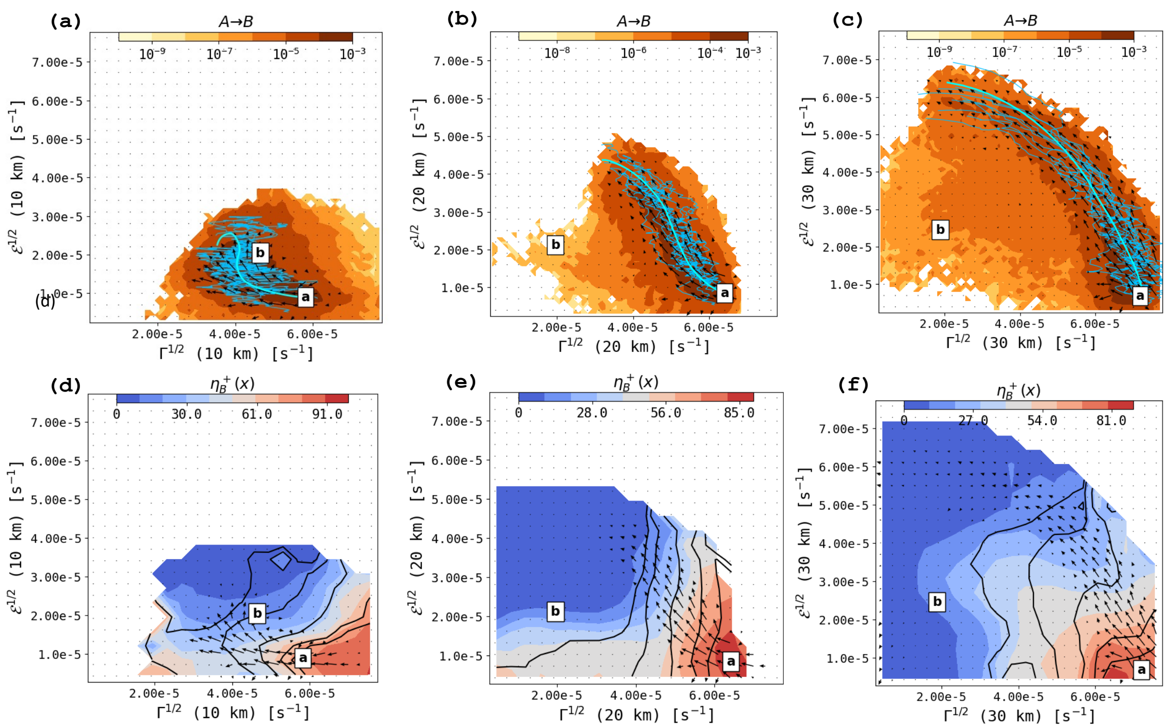


FIG. 5. **Vertical profiles of transition states and tendencies.** Left column: $U(z)$ averaged over $q_B^+ = 0.1$, 0.5, and 0.9. Orange curve is the mean, and gray envelopes represent the middle 25-, 50-, and 90-percentile ranges. Dashed blue and red curves represent $U(z)$ for the fixed points **a** and **b**. Right column: same as left, but for eddy meridional heat flux $\overline{v'T'}$.



883 FIG. 6. **Current in wave-mean flow coordinates.** Same as Fig. 3, but for a different observable subspace
884 $(\Gamma^{1/2}, \mathcal{E}^{1/2})$ instead of (U, IHF). See text for definitions. Eddies are characterized by RMS perturbation PV, $\mathcal{E}^{1/2}$,
885 and the mean flow by the zonal mean PV gradient, $\Gamma^{1/2}$.

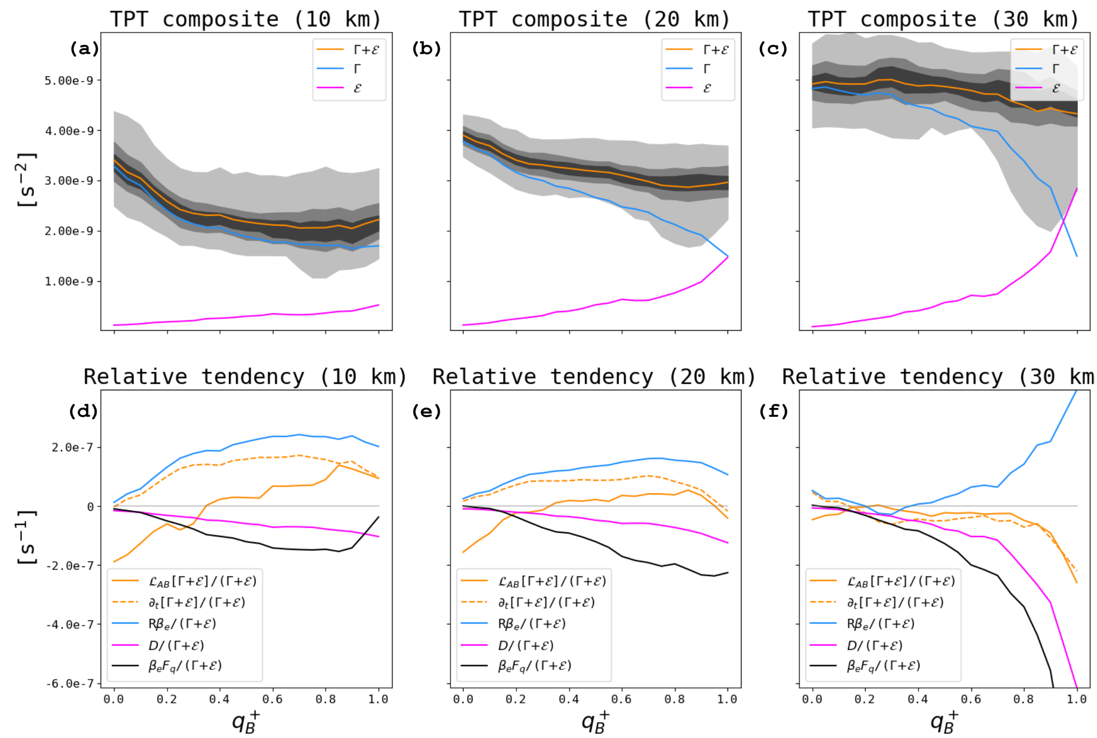


FIG. 7. **Enstrophy budget analysis through the $A \to B$ transition**. (a) Blue, pink, and orange curves represent mean values of Γ, $\mathcal{E}$, and their sum at $z = 10$ km, conditioned on the system being in a transition path and near a given committor level (which varies along the horizontal axis). Gray envelopes represent the middle 25, 50, and 90-percentile ranges of $\Gamma + \mathcal{E}$; when the orange curve is not at the center of the gray envelopes, the distribution is skewed. (b, c): same as (a), but at $z = 20$ and 30 km respectively. (d) Solid orange curve shows the expected tendency of $\Gamma + \mathcal{E}$ at 10 km, again conditioned on being in a transition path and near a given committor level. Dashed orange curve shows the deterministic tendency at the same committor levels; the difference between the two indicates the role of stochastic forcing. Blue curve shows the relaxation of Γ (the squared meridional PV gradient), pink curve shows the dissipation of enstrophy, and black curve shows the meridional transport of PV, $F_q\beta_e$, which when negative indicates a gain for $\mathcal{E}$ at the expense of Γ. The sum of the blue and pink curves gives the dashed orange curve. (e, f): same as (d), but at $z = 10$ and 20 km respectively. All tendencies are normalized by $\Gamma + \mathcal{E}$, as the legend shows, for a comparable vertical scale across altitudes.