

Exploring stratospheric rare events with transition path theory and short simulations

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

Extreme weather events are simultaneously the least likely and the most impactful features of the climate system, increasingly so as climate change proceeds. Extreme events are multi-faceted, highly variable processes which can be characterized in many ways: return time, worst-case severity, and predictability are all sought-after quantities for various kinds of rare events. A unifying framework is needed to define and calculate the most important quantities of interest for the purposes of near-term forecasting, long-term risk assessment, and benchmarking of reduced-order models. Here we use Transition Path Theory (TPT) for a comprehensive analysis of sudden stratospheric warming (SSW) events in a highly idealized wave-mean flow interaction system with stochastic forcing. TPT links together probabilities, dynamical behavior, and other risk metrics associated with rare events that represents their full statistical variability. At face value, fulfilling this promise demands extensive direct simulation to generate the rare event many times. Instead, we implement a highly parallel computational method that launches a large ensemble of short simulations, estimating long-timescale rare event statistics from short-term tendencies. We specifically investigate properties of SSW events including passage time distributions and large anomalies in vortex strength and heat flux. We visualize high-dimensional probability densities and currents, obtaining a nuanced picture of critical altitude-dependent interactions between waves and the mean flow that fuel SSW events. We find that TPT more faithfully captures the statistical variability between events as compared to the more conventional minimum action method.

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).

Historically, significant effort has already 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 heterogeneous, from fully coupled GCMs to regional climate models to machine learning parameterizations, 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 holistic and relevant set of quantities that a good model should reproduce.

Commonly used benchmarks are sensible and important, but necessarily ad hoc given the variety of extreme events. Quantiles of univariate distributions provide some information. For example, O’Gorman and Dwyer (2018)

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

use the 99.9th percentile of 3-hour rainfall, as a function of latitude, to validate a machine-learned convection parameterization. Qi and Majda (2020) incorporate tail probabilities explicitly into the loss function for a machine-learned shallow water wave model. Extreme value theory provides another mathematically rigorous way to quantify tails (Katz 2010; Lucarini et al. 2016). However, fidelity of different metrics can conflict with each other. Li et al. (2018) experimented with a regional climate model under increasing temperature, and successfully matched observed trends in spatial extent but failed to match observed trends in extreme rainfall. Moreover, the many scales and components of the climate system lead to ambiguous definitions of extremes. For example, atmospheric blocking is a recurrent large-scale feature driving many kinds of weather extremes, but a multitude of definitions hampers objective comparison across models (Woollings et al. 2018; Chan et al. 2019; Gálfi et al. 2019; Chattopadhyay et al. 2020).

The heterogeneity of rare weather events and their impacts calls for a unifying, but flexible, rubric for describing them. For a given event, however defined, we wish to quantify its precursors, its frequency, and its spatiotemporal structure. Most rare event studies, such as those cited above, address some subset of these metrics depending on the methodology, but rarely all together. Transition path theory (TPT), introduced in E and Vanden-Eijnden (2006), is a statistical mechanics framework that ties together all of these quantities by abstracting the rare event into a path through high-dimensional state space. 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 across a three-dimensional model state space due to Ruzmaikin et al. (2003) and Birner and Williams (2008). Here, we explore a stochastically forced version of the classic model of Holton and Mass (1976), one of the first models to capture important aspects of sudden stratospheric warming (SSW). In the language of TPT, a SSW event is a system 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).

TPT analysis is related to, but distinct from, the forecasting problem, whose importance is well-recognized. Every extra day of advance warning helps us prepare for the ensuing cold snaps following SSW. The subseasonal-to-seasonal (S2S) timescale on which SSW occurs is an outstanding challenge in state-of-the-art weather forecasting (Vitart and Robertson 2018). Many data-driven and probabilistic approaches are building on traditional ensemble forecasting methods. In a recent paper (Finkel et al. 2021),

we computed two key forecasting quantities for a prototype model of SSW, the Holton-Mass model—namely, the committor probability and lead time, to be defined below—as functions of initial condition, finding optimally predictive physical observables for SSW onset. Tantet et al. (2015) computed early warning signs of atmospheric blocking by estimating a committor (by a different name) in a reduced state space. Lucente et al. (2019) and Lucente et al. (2021) have also computed committor functions for simple models of El Niño, mathematically quantifying the so-called predictability barrier. Bayesian machine learning (Chen et al. 2021) and kernel forecasting (Wang et al. 2020) are also being investigated in the quest to forecast El Niño.

But near-term forecasting alone does not completely characterize long-term risk or the event’s mechanism from start to finish; this is what TPT provides. In the present paper, we use TPT to connect the short-term weather forecasting problem to the long-term climatology of SSW events, including their precursors, overall frequency, and distribution of severity (by several metrics). Furthermore, TPT offers detailed insight into the development of SSW through the *probability current*: the average tendency of the system conditioned on the occurrence of an SSW. By visualizing the current in various ways, we will quantitatively assess the interplay between wave disturbances, zonal wind anomalies, and heat flux during SSW, and the extent to which they are uniquely associated with an SSW. Crucially, TPT tells us the *variability* of these processes, not just their mean behavior. In particular, we will see a dichotomy between successive stages of an 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.

The statistical ensemble of transition pathways characterized by TPT can be compared to the single pathway found by minimizing the Freidlin-Wentzell action (Freidlin and Wentzell 1970). That pathway is representative of the rare event’s development in the low-noise limit and is computed by the *minimum-action method* (e.g., E et al. 2004). “Noise” here means unresolved processes that evolve quickly and unpredictably relative to the large-scale variables of interest, and are typically represented by stochastic forcing (e.g., Berner et al. 2017). The minimum-action method plays a central role in recent rare event analyses, such as Hoffman et al. (2006a) and Plotkin et al. (2019) for tropical cyclones, Bouchet et al. (2014) for two-dimensional fluid mechanics, and Dematteis et al. (2018) for rogue waves. The prominent role of one single trajectory is both a strength and weakness of this method: eliminating all variability among realizations of the rare event leads to a clear, but possibly biased, narrative.

The straightforward way to quantify the statistical ensemble of transition paths is by direct numerical simulation (DNS): integrate a model for a long time until many such events are observed, and then proceed with statistical analysis. Alternatively, in very low-dimensional models, one can compute the necessary quantities by fully discretizing state space, as in Finkel et al. (2020). Unfortunately, for rare events in high-dimensional models, discretization is impossible and DNS becomes too expensive to generate a data set with statistical power. In this paper, as in Finkel et al. (2021), we circumvent this problem by simulating very many, very short trajectories in parallel, and combining information from all of them to compute rare event statistics without ever observing a complete rare event. The particular approach we use was introduced in (Thiede et al. 2019; 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 in (Noé et al. 2009b) the authors 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 summarize the dynamical model under study. In section 3, we define TPT quantities of interest and visualize them on the Holton-Mass model. In particular, we use probability densities and currents to give a description of the geometry and dynamical behavior of transition pathways. We also examine the relationship between TPT and the minimum action method. The resulting physical insight will motivate the more technical section 4, where we outline the computational approach and tabulate some quantitative statistical properties of transition paths and their variability. 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 $\bar{u}(y, z, t)$ and a perturbation geostrophic streamfunction $\psi'(x, y, z, t)$ reduced down to a single wave mode. Two competing forces in the model bring about bistability. First, an altitude-dependent radiative cooling $\alpha(z)$ relaxes the zonal wind toward a strong polar winter state in thermal wind balance with a radiative equilibrium temperature field. This is the basic mechanism maintaining the winter polar vortex, which in the Holton-Mass model corresponds to a stable

equilibrium we denote state **a**. Second, a wave perturbation is forced at the lower boundary, the nominal tropopause, associated with stationary waves in the troposphere induced by topography and land-sea contrast. The second state in the model is a wave propagation regime, where upward propagating stationary waves flux momentum down to the lower boundary, creating a highly disturbed state with weak zonal winds. This is a second stable equilibrium we denote state **b**.

Sudden stratospheric warming events are abrupt transitions from the strong vortex state (**a**) to the disturbed, wave driven state (**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 coexistence of these two regimes is fundamental to the Holton-Mass model.

For reference we write down the PDE system here, but refer the reader to Holton and Mass (1976); Yoden (1987); Christiansen (2000) and Finkel et al. (2021) for complete explanations. The zonal wind $\bar{u}(y, z, t)$ and perturbation streamfunction $\psi'(x, y, z, t)$ are projected onto a single wavenumber $k = 2/(a \cos \theta)$ in the zonal direction and a single wavenumber $\ell = 3/a$ in the meridional direction (a = the radius of Earth), hence the ansatz

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

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

where H is the scale height, 7 km. In the resulting highly reduced model, the state space consists of $\Psi(z, t)$ (which is complex) and $U(z, t)$. We impose the boundary conditions according to a height parameter, h , representing the orographic forcing, and a background radiative zonal wind profile $U^R(z)$ that increases linearly with altitude:

$$\Psi(0, t) = \frac{gh}{f_0}, \quad \Psi(z_{top}, t) = 0, \quad (3)$$

$$U(0, t) = U^R(0), \quad \partial_z U(z_{top}, t) = \partial_z U^R(z_{top}).$$

We fix $h = 38.5 m$ in this study; g and f_0 are the gravitational acceleration and coriolis parameter, respectively.

The reduced state variables (2) are then inserted into the prediction equation for \bar{u} and the linearized quasi-geostrophic potential vorticity equation for ψ' . We nondimensionalize the equations with the parameter $\mathcal{G}^2 = H^2 N^2 / (f_0^2 L^2)$, where $T = 86400 s$ (1 day) and L is a tunable length scale. Note that $\mathcal{G} = L_d / L$, where $L_d = NH / f_0$ is the deformation radius. In order to make a data set homogeneously distributed in state space, we select a length

scale $L = 2 \times 10^5$ m and thus $\mathcal{G} = 4.4$, a choice motivated by numerics rather than dynamics. The resulting QGPV equation reads

$$\begin{aligned} & \left[-\left(\mathcal{G}^2(k^2 + \ell^2) + \frac{1}{4} \right) + \frac{\partial^2}{\partial z^2} \right] \frac{\partial \Psi}{\partial t} \\ &= \left[\left(\frac{\alpha}{4} - \frac{\alpha_z}{2} - i\mathcal{G}^2 k\beta \right) - \alpha_z \frac{\partial}{\partial z} - \alpha \frac{\partial^2}{\partial z^2} \right] \Psi \\ &+ \left\{ ik\varepsilon \left[\left(k^2 \mathcal{G}^2 + \frac{1}{4} \right) - \frac{\partial}{\partial z} + \frac{\partial^2}{\partial z^2} \right] U \right\} \Psi - ik\varepsilon \frac{\partial^2 \Psi}{\partial z^2} U, \end{aligned} \quad (4)$$

specifying the tendency of the wave Ψ . The mean flow prediction equation becomes

$$\begin{aligned} & \left(-\mathcal{G}^2 \ell^2 - \frac{\partial}{\partial z} + \frac{\partial^2}{\partial z^2} \right) \frac{\partial U}{\partial t} = [(\alpha_z - \alpha)U_z^R - \alpha U_{zz}^R] \\ & - \left[(\alpha_z - \alpha) \frac{\partial}{\partial z} + \alpha \frac{\partial^2}{\partial z^2} \right] U + \frac{\varepsilon k \ell^2}{2} e^z \text{Im} \left\{ \Psi \frac{\partial^2 \Psi^*}{\partial z^2} \right\} \end{aligned} \quad (5)$$

where $\alpha = \alpha(z)$ is the cooling coefficient and $\varepsilon = 8/(3\pi)$ is a mode projection coefficient. The notation follows Christiansen (2000), where the parameters are explained in more detail.

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

$$\mathbf{X}(t) = [\text{Re}\{\Psi(t)\}, \text{Im}\{\Psi(t)\}, U(t)] \in \mathbb{R}^{75} \quad (6)$$

each of the three entries representing a vector with 25 discrete altitudes.

SSWs are associated with both internal stratospheric variability, generated even with time-independent forcing from the troposphere (Scott and Polvani 2006; Matthewman and Esler 2011), and external time-dependent forcing (Sjoberg and Birner 2012; Scott et al. 2008). As a minimal demonstration of our computational approach, we consider only time-independent boundary conditions, but add stochastic perturbations to the zonal wind profile, as specified in Finkel et al. (2021), to drive transitions between the weak and strong vortex states. The noise is smooth in space, but white in time, consisting of two Fourier modes in the vertical. This stochasticity represents 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, i.e., variations in the flow associated with smaller scale waves that have been truncated. We emphasize that while this use of stochastic forcing affects the transition path statistics, it does not affect the method for computing them.

The two stable equilibria arising from these competing forces are depicted in Fig. 1 (a,b): **a** is the strong vortex state, with a linearly increasing zonal wind profile and almost barotropic streamfunction, while **b** is the disturbed

vortex state, with a weak zonal wind profile and a phase-tilted streamfunction. A *transition path* is defined as an unbroken segment, or trajectory, of the system that begins in a region A of state space and travels to another region B without returning to A . In this paper we define A and B as spherical neighborhoods (in non-dimensional space) about the two fixed points:

$$A = \{\mathbf{x} \in \mathbb{R}^d : |\mathbf{x} - \mathbf{a}| \leq r_A = 8\} \quad (7)$$

$$B = \{\mathbf{x} \in \mathbb{R}^d : |\mathbf{x} - \mathbf{b}| \leq r_B = 30\} \quad (8)$$

A larger “ball” around B is needed to capture equivalent populations in each state. An “SSW” is then a transition from A to B , while the reverse, from B to A , represents the recovery of the vortex. This departs slightly from Finkel et al. (2021), where we defined A and B based on zonal wind at a single altitude. The definition based exclusively on the zonal wind near 10 hPa was chosen to be consistent with the standard WMO definition. In this paper, we focus on dynamical insights into the *entire* process from A to B , and the transient dynamics are not over once $U(30 \text{ km})$ drops below an arbitrary threshold.

The second row of Fig. 1 illustrates the state space geometry. The 75-dimensional fixed points **a** and **b** are projected onto two-dimensional subspaces, along with plots of the system’s evolution from a long control simulation including two transition paths from $A \rightarrow B$ and back. The heavy clusters of black curves around **a** and **b** indicate that their neighborhoods A and B are metastable, the vortex tending to linger in one of the regions for an extended period before quickly switching to the other. The transition paths are highlighted in orange ($A \rightarrow B$) and green ($B \rightarrow A$), beginning precisely when the path last exits A and ending when it first enters B or vice versa. These two samples are only anecdotal, meant to give a general sense for the dynamics. The transition path ensemble refers to the infinite collection of paths, which vary significantly from the two samples shown.

The subspaces in Fig. 1(c,d) are dynamically relevant to SSW physics. The vertical axis is zonal wind at 30 km, approximately where the **b** profile reaches its minimum. Panel (c) has wave magnitude $|\Psi|$ in the horizontal, which is greatly enhanced during vacillation cycles and during the $A \rightarrow B$ transition. Panel (d) has vertically integrated heat flux up to 30 km, abbreviated IHF(30 km), in the horizontal, which quantifies both the magnitude and phase of the streamfunction across altitudes. In the Holton-Mass model,

$$\overline{v'T'}(z) = \frac{H f_0}{R} \frac{\partial \psi'}{\partial y} \frac{\partial \psi'}{\partial z} \propto e^{z/H} |\Psi(z)|^2 \frac{\partial \varphi}{\partial z} \quad (9)$$

f_0 is the Coriolis parameter at 60°N, 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

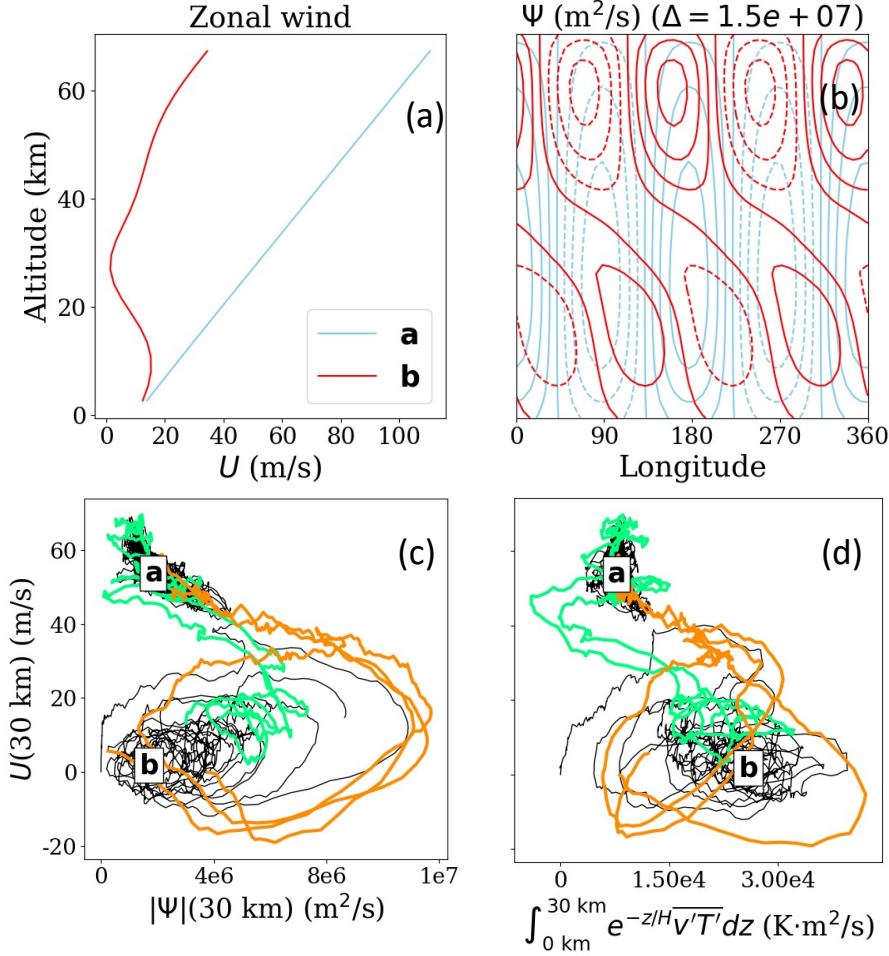


FIG. 1. **The two stable equilibria of the Holton-Mass model**, similar to Fig. 1 of Finkel et al. (2021). The upper two panels show the zonal wind $U(z)$ and the streamfunction $\psi'(z)$ for the two fixed points **a** (the strong vortex, in blue) and **b** (the weak vortex, in red). The lower panels project a model integration onto two subspaces: $(U, |\Psi|)$ and (U, IHF) , at 30 km. Several SSW events (transitions from $A \rightarrow B$, highlighted in orange) and recovery events (transitions from $B \rightarrow A$, highlighted in green), are observed.

streamfunction. The horizontal axis of (d) is also weighted by density before being integrated:

$$\begin{aligned} \text{IHF}(30\text{km}) &= \int_{0\text{km}}^{30\text{km}} e^{-z/H} \overline{v'T'}(z) dz \\ &\propto \int_{0\text{km}}^{30\text{km}} |\Psi(z)|^2 \frac{\partial \varphi}{\partial z} dz \end{aligned} \quad (10)$$

The right side measures the area swept out in the complex plane by $\Psi(z)$, $0\text{km} \leq z \leq 30\text{km}$. Whether interpreted dynamically or geometrically, this quantity is important for SSW and separates **a** and **b** more than the wave amplitude itself, as seen by the horizontal separation between **a** and **b** in panel (d).

Fig. 1 (c) and (d) illustrate some of the transitory dynamics we wish to capture quantitatively. For example, the orange $A \rightarrow B$ segments take an excursion to especially

large wave magnitude and negative zonal wind before approaching set **B** and the cluster of vacillation cycles. This motivates the more restrictive definitions of A and B in (7): here, the early and late transition stages are part of the path, rather than being obscured by the states. The green $B \rightarrow A$ segments take an altogether different route through state space, gradually increasing in zonal wind and decreasing in wave amplitude, with a late negative spike in IHF.

3. Describing transition paths: sample and mean behavior

We first point out some physically notable features of simulated transition paths. Motivated by anecdotal observations, we then visualize probability densities and currents from TPT, which characterize the statistical behavior of transition paths, providing a richer description than the minimum-action path.

a. Qualitative transition path properties

In the left column of Fig. 2, we plot several observable functions for the first 3000 days of a control simulation, the same data as shown in Fig. 1 (c) and (d). The functions are zonal wind, U , at 30 km, streamfunction magnitude, $|\Psi|$ at 30 km, heat flux, $\overline{v'T'}$, at 30 km, and the vertically integrated heat flux, IHF, up to 30 km. The sets A and B divide the dynamics into four separate “phases” delimited by highlights in the time series plots. (1) In the $A \rightarrow B$ phase, marked by orange, the vortex is breaking down, en route from A to B . (2) In the $B \rightarrow A$ phase, marked by green, the vortex is recovering from the vacillating regime back to the radiatively driven regime. (3) 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 (between the end of a green segment and the start of an orange segment). (4) In the $B \rightarrow B$ phase, the vortex is weak, caught in ongoing vacillation cycles in the vicinity of B (between the end of an orange segment and the start of a green segment).

Fig. 2 demonstrates that the system behaves very differently in each phase. Zonal wind is weaker overall during $B \rightarrow B$ than $A \rightarrow A$ (by definition), but moreover it oscillates at a higher frequency, similar to the “vacillation cycles” observed originally by Holton and Mass (1976). But the largest negative zonal wind spikes appear to occur during the SSW, i.e., during the $A \rightarrow B$ phase; see the two large dips in zonal wind at the first transition in Fig. 1, which corresponds to the orange transition path in Fig. 1 (c) that loops twice. A similar pattern appears in both $|\Psi|$ and $\overline{v'T'}$ time series, where the $B \rightarrow B$ phase supports regular oscillations that are overshadowed by the preceding $A \rightarrow B$ transition.

The vortex recovery phase $B \rightarrow A$ appears considerably tamer, at least in terms of zonal wind, which exhibits no obvious “extremal” behavior while recovering. Rather, the oscillations inherent to set B gradually weaken through the first half of the recovery, after which a smooth restoration to A ensues. However, the lower left panel of Fig. 2 shows unusually negative dips in IHF at 30 km midway through the $B \rightarrow A$ transitions. IHF is a proxy for streamfunction phase tilt and amplitude, which must decrease from B to A , but the overshoot, indicating wave reflection and downward propagation, is not necessary *a priori*.

These observations, though potentially insightful, are based on only two transitions. We start to evaluate the hypotheses quantitatively by describing each phase with a conditional probability distribution. TPT focuses specifically on the $A \rightarrow B$ and $B \rightarrow A$ phases, during which we say that \mathbf{X} is “reactive”, using a term from chemistry literature where the passage from A (reactant) to B (product) models a chemical reaction.

	DGA	DNS
% Time $A \rightarrow A$	49.0	52.4 ± 10.2
% Time $A \rightarrow B$	6.8	4.2 ± 1.0
% Time $B \rightarrow B$	37.5	38.1 ± 9.7
% Time $B \rightarrow A$	6.7	5.3 ± 2.0
Period (days)	1744	2058 ± 710

TABLE 1. Fraction of time that the system spends in each phase of the SSW lifecycle, computed both from DGA and empirically from the control simulation, plus or minus two standard deviations.

b. Steady-state and reactive probability densities

The right-hand column of Fig. 2 shows five probability distributions projected onto the vertical coordinate. The process as a whole exhibits a *steady-state density* $\pi(\mathbf{x})$, which describes the long-term probability that the system is in a state near \mathbf{x} . It can be computed simply by DNS, i.e., integrating the dynamics for a long time and binning data into a histogram. We have instead performed the Dynamic Galerkin Approximation (DGA) procedure (described in the supplement) with only short simulations. $\pi(\mathbf{x})$, shown in black in the right column of Fig. 2, is a weighted mixture of the other four phase-specific densities π_{AA} , π_{BB} , π_{AB} , and π_{BA} (the latter two are called *reactive densities*). The weights are given by the percentage of time spent in each phase, listed in Table 1 along with the average period of the cycle (or inverse rate). The table presents numbers according to both DGA and DNS. It is difficult to assess its uncertainty, but we can assess the uncertainty in the DNS estimate by dividing the long 2×10^5 -day simulation into 10 equal-sized chunks, calculating the appropriate statistic in each chunk, and taking the sample variance of the 10 sample means. Table 1 presents the 95% confidence interval of each number, i.e., two standard deviations in each direction. The interval includes the DGA point estimate in all cases except for the $A \rightarrow B$ time fraction. We discuss numerical error and how it can be reduced in the supplement.

Table 1 specifies the weight of each phase distribution π_{AA} , π_{AB} , etc., but their shapes reveal their unique characteristics. Each density is peaked in a different place: while π is strongly bimodal with peaks near \mathbf{a} and \mathbf{b} , the reactive densities have more probability at intermediate wind strength, which transition paths must cross through. π_{BA} even has a secondary peak around $U = 20$ m/s, suggesting that the recovery process tends to stall or slow down on its way back to A . Comparing with the $B \rightarrow A$ transition paths highlighted in green, the slowdown seems to coincide with the mid-recovery switch from weakening oscillations to steady recovery of zonal wind.

π_{AB} also has a secondary peak, not at intermediate strength but rather at a negative extreme. This confirms the observation above that zonal wind tends to plummet

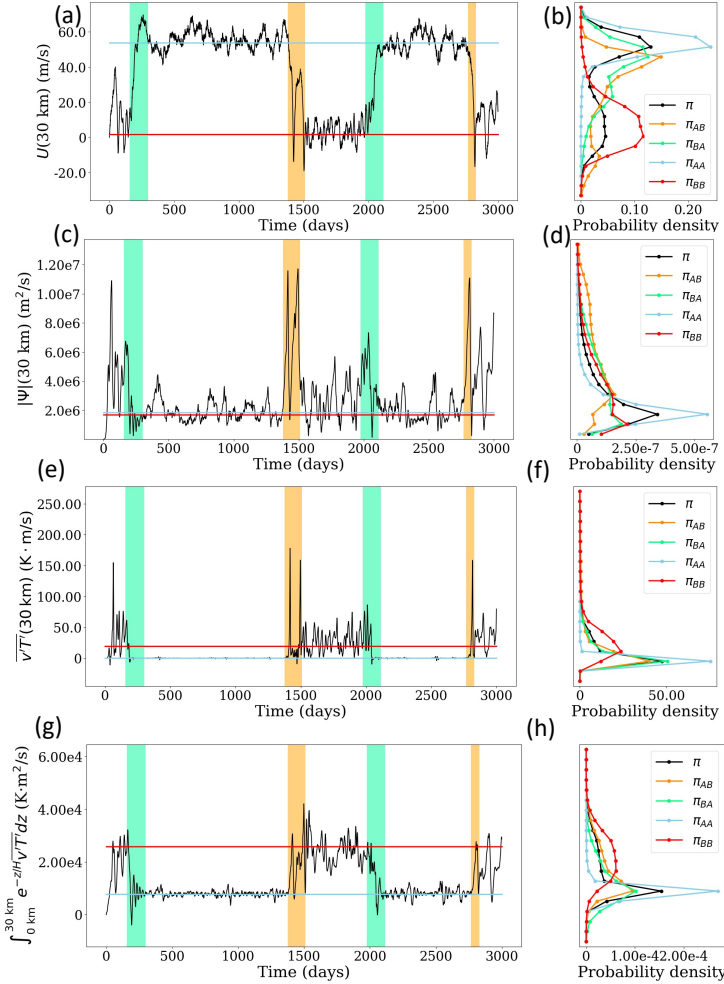


FIG. 2. **Observed transition paths and reactive densities.** Each row displays a timeseries of a different observable: zonal wind $U(30 \text{ km})$, wave magnitude $|\Psi|(30 \text{ km})$, meridional heat flux $\overline{v'T'}(30 \text{ km})$, and integrated heat flux $\int_{0 \text{ km}}^{30 \text{ km}} \rho \overline{v'T'} dz$, a proxy for phase tilt. Several SSW events (transitions from $A \rightarrow B$, highlighted in orange) and recovery events (transitions from $B \rightarrow A$, highlighted in green), are observed.

especially low during the vortex breakdown process. Meanwhile, the negative IHF anomalies during $B \rightarrow A$ are also borne out by the density π_{BA} in the lower right panel, which has a slightly heavier negative tail than π or π_{AB} . Over the variable $\overline{v'T'}(30 \text{ km})$, the large spikes observed in both of the two orange intervals suggest that π_{AB} has a heavier upper tail than the other distributions. The tail is so small compared to its peak that a visual assessment is not practical, and we confirm this observation below.

To corroborate these effects numerically, we display a selection of *lifetime averages* in Fig. 3. We have defined three “indicator functions” tailored to capture the anoma-

lies described above:

$$\Gamma(\mathbf{x}) = \mathbb{1}\{U(30 \text{ km}) < -20 \text{ m/s}\} \quad (11)$$

$$\Gamma(\mathbf{x}) = \mathbb{1}\{\overline{v'T'}(30 \text{ km}) > 60 \text{ K} \cdot \text{m/s}\} \quad (12)$$

$$\Gamma(\mathbf{x}) = \mathbb{1}\left\{\int_{0 \text{ km}}^{30 \text{ km}} e^{-z/H} \overline{v'T'}(z) dz < 0 \text{ K} \cdot \text{m}^2/\text{s}\right\} \quad (13)$$

i.e., $\Gamma(\mathbf{x})$ in (11) is one if \mathbf{x} is a state with $U(30 \text{ km}) < -20 \text{ m/s}$ and zero otherwise. Assuming ergodicity, the fraction of time spent with $U(30 \text{ km}) < -20 \text{ m/s}$ is

$$\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \Gamma(\mathbf{X}(t)) dt = \int_{\mathbb{R}^d} \Gamma(\mathbf{x}) \pi(\mathbf{x}) d\mathbf{x} =: \langle \Gamma \rangle_{\pi}, \quad (14)$$

or equivalently, the area under the black curve in Fig. 2(b) below the -20 m/s tick mark. The average can also be

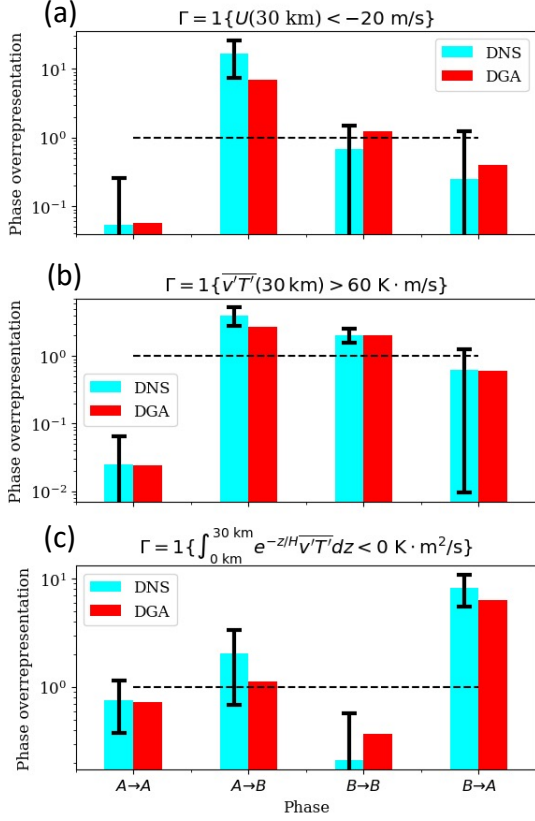


FIG. 3. **Lifecycle means.** For a given indicator function $\Gamma(\mathbf{x})$ representing extremal behavior, e.g., a negative wind anomaly in (a), the “overrepresentation” metric $\langle \Gamma \rangle_{\pi_{AB}} / \langle \Gamma \rangle_{\pi}$ says how much the $A \rightarrow B$ phase is responsible for that anomalous behavior relative to its time fraction. These are computed for each phase of the SSW lifecycle, both using DNS (cyan bars) and from DGA (red bars) for a variety of anomaly indicators Γ defined in the subplot titles. Error bars are computed for DNS by dividing the control simulation into 10 equal pieces.

restricted to any given phase. For example, $\langle \Gamma \rangle_{\pi_{AB}}$ is the fraction of time during $A \rightarrow B$ transitions that $U(30 \text{ km}) < -20 \text{ m/s}$ —in other words, the area under the orange curve below the same mark. The ratio $\langle \Gamma \rangle_{\pi_{AB}} / \langle \Gamma \rangle_{\pi}$ measures the “overrepresentation” of negative wind anomalies in the $A \rightarrow B$ phase.

Fig. 3 plots each of the three anomaly indicators for each phase in the lifecycle, according to both DGA and DNS, along with error bars for the latter (note the log scale makes them asymmetric). Consider the $A \rightarrow B$ phase in the top panel, indicating a ratio $\langle \Gamma \rangle_{\pi_{AB}} / \langle \Gamma \rangle_{\pi} \approx 10$. By default, we would expect about 6.8% of negative wind spikes to occur during $A \rightarrow B$ transitions, according to the DGA time fraction in Table 1. The factor of 10 suggests that instead, the $A \rightarrow B$ phase accounts for about 68% of such anomalies. In terms of positive heat flux anomalies, the $A \rightarrow B$ transition also punches above its weight, though not quite ten times. The lower panel confirms that negative

IHF is a largely unique trait of $B \rightarrow A$ transitions. The DGA/DNS matches are rough, but typically within the error bars and consistent in terms of ranking phases against each other.

These “reactive probability densities” can be projected onto any observable of interest, providing a simple but powerful tool to measure unique aspects of transition paths. However, reactive densities do not tell the whole story, because they are *static*. For example, the negative tail of π_{AB} over $U(30 \text{ km})$ says nothing about the shape of the vacillation cycles as they wind through state space. In other words, reactive densities say where transitions go and for how long they linger, but not the details of their route. In the next two subsections we explain and visualize *committors* and *reactive currents* (E and Vanden-Eijnden 2006), which go further to describe the *dynamic* behavior of transition pathways.

c. Committors

A necessary prerequisite to understand the motion of transition pathways is a suitable measure of progress from A to B , namely a committor function. Suppose an initial condition $\mathbf{X}(t_0) = \mathbf{x}$ is observed 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, and the respective probabilities define the committor function:

$$q^+(\mathbf{x}) = \mathbb{P}_{\mathbf{x}}\{\mathbf{X}(\tau^+(t_0)) \in B\} \quad (15)$$

where $\tau^+(t_0)$ is the *first hitting time* to set A or B ,

$$\tau^+(t_0) = \inf\{t > t_0 : \mathbf{X}(\tau^+(t_0)) \in A \cup B\} \quad (16)$$

(a random variable), and the subscript \mathbf{x} denotes a conditional probability given $\mathbf{X}(t_0) = \mathbf{x}$. We assume the system is autonomous, so we can safely set $t_0 = 0$ and drop the argument from τ^+ (however, the argument returns in the mathematical formulation of TPT presented in section 1 of the supplement). We will also use the intuitive notation $q^+(\mathbf{x}) = \mathbb{P}_{\mathbf{x}}\{\mathbf{x} \rightarrow B\}$, as q^+ simply tells how likely it is to next go to B , not A . The committor measures probabilistic progress toward B , and we argued for its utility as a forecast in Finkel et al. (2021). However, it does not distinguish the $A \rightarrow B$ phase from the $B \rightarrow B$ 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*:

$$q^-(\mathbf{x}) = \mathbb{P}_{\mathbf{x}}\{\mathbf{X}(\tau^-(t_0)) \in A\} =: \mathbb{P}_{\mathbf{x}}\{A \rightarrow \mathbf{x}\} \quad (17)$$

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

$$\tau^-(t_0) = \sup\{t < t_0 : \mathbf{X}(\tau^-(t_0)) \in A \cup B\} \quad (18)$$

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 forward and backward committors are shown in Fig. 4(a) and (b). Note that their contour structures are very different, a sign of irreversible behavior. In particular, the negative extreme of zonal wind has both large q^+ and large q^- , meaning that whenever a negative wind anomaly is observed, chances are high that the vortex has undergone an SSW and is on its way from a strong vortex (state A) to a weak vortex (state B). We see this more clearly by combining the two committors:

$$q^-(\mathbf{x})q^+(\mathbf{x}) = \mathbb{P}_{\mathbf{x}}\{A \rightarrow \mathbf{x}\}\mathbb{P}_{\mathbf{x}}\{\mathbf{x} \rightarrow B\} \quad (19)$$

$$=: \mathbb{P}_{\mathbf{x}}\{A \rightarrow B\} \quad (20)$$

$$(1 - q^-(\mathbf{x}))(1 - q^+(\mathbf{x})) = \mathbb{P}_{\mathbf{x}}\{B \rightarrow \mathbf{x}\}\mathbb{P}_{\mathbf{x}}\{\mathbf{x} \rightarrow A\} \quad (21)$$

$$=: \mathbb{P}_{\mathbf{x}}\{B \rightarrow A\} \quad (22)$$

We display these probabilities in Fig. 4(c,d) to highlight signatures of the $A \rightarrow B$ and $B \rightarrow A$ transitions. The red crescent along the right and bottom flank of state space confirms that large positive $|\Psi|$ spikes and negative U spikes are strongly associated with the $A \rightarrow B$ transition, meaning SSWs are responsible for a disproportionate share of these anomalies. This confirms the precedence of the $A \rightarrow B$ phase in Fig. 3(a), but also reveals details of the dynamical interaction between zonal wind and wave amplitude. Meanwhile, a small region of low $|\Psi|$ and medium U is associated with the $B \rightarrow A$ transition. A projection onto (IHF, U) space (which we omit) confirms the association between negative IHF anomalies and the $B \rightarrow A$ phase that we have observed in Fig. 2 and Fig. 3(c).

The maps in Fig. 4(c,d) are related to the reactive densities discussed in the last subsection, viz.

$$\pi_{AB}(\mathbf{x}) \propto q^+(\mathbf{x})q^-(\mathbf{x})\pi(\mathbf{x}) \quad (23)$$

$$= \mathbb{P}_{\mathbf{x}}\{A \rightarrow B\}\pi(\mathbf{x}) \quad (24)$$

and likewise for $B \rightarrow A$. The proportionality constants are for normalization. These maps, although static, hint at preferred routes through state space, which we confirm in the next subsection by visualizing the reactive current.

d. Reactive current and minimum-action path

In parallel with the three probability densities $\pi(\mathbf{x})$, $\pi_{AB}(\mathbf{x})$ and $\pi_{BA}(\mathbf{x})$ are three *probability currents*. The *equilibrium current* $\mathbf{J}(\mathbf{x})$ is a vector field over \mathbb{R}^d that indicates, roughly speaking, the “average” motion of the system as it passes through \mathbf{x} , regardless of the SSW phase. The *reactive current* $\mathbf{J}_{AB}(\mathbf{x})$ describes the average motion of $A \rightarrow B$ transition paths crossing through \mathbf{x} , and likewise $\mathbf{J}_{BA}(\mathbf{x})$ describes $B \rightarrow A$ transitions. We define these currents precisely in the supplement, but first make an analogy with fluid-dynamical currents. 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 reactive current highlights the smaller filaments that connect them.

This analogy helps to interpret Fig. 5, which maps out the equilibrium and reactive densities and currents in the same subspace $(|\Psi(30\text{km})|, U(30\text{km}))$ as shown in Fig. 1(c). The orange-scale color shading in Fig. 5(a) indicates the steady-state density $\pi(\mathbf{x})$, the same as the black curves in the right column of Fig. 2, but now projected onto two dimensions and mapped with a logarithmic scale. $\pi(\mathbf{x})$ is larger near \mathbf{a} and \mathbf{b} , indicating the system’s overall bimodality. Overlaid on $\pi(\mathbf{x})$ is the equilibrium current $\mathbf{J}(\mathbf{x})$, which is disorderly near \mathbf{a} , but highly organized as an “eddy” around \mathbf{b} . This reflects the vacillation cycles in the $B \rightarrow B$ phase seen in the time series of Fig. 2, and offers a *dynamical* perspective not available from the stationary distribution $\pi(\mathbf{x})$. 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} has imaginary eigenvalues; however, since \mathbf{b} is stable, without noise the oscillations would die out eventually. Stochasticity is essential to maintain those circulations at equilibrium. In a sense we can think of them as hinting at the ghost of the limit cycle lurking beyond a Hopf bifurcation. A similar effect termed “self-induced stochastic resonance” has been observed and analyzed in simple excitable systems (Muratov et al. 2005), and Weiss et al. (2020) recognized current loops, or “probability angular momentum,” as a ubiquitous feature in climate dynamics.

The equilibrium current in between A and B is weaker, but essential for carrying the system during the transitory phases. In panel (a), the horizontal slice in phase space near $U(30\text{km}) = 20$ m/s exhibits vector field arrows lying almost horizontally in the $(|\Psi|, U)$ plane, forming a probabilistic basin boundary between two eddies. For a transition to occur, that boundary has to be ruptured, and exactly where and how it breaks indicates a lot about the preferred transition mechanism; this is the subject of panels (b) and (c).

Panel (b) shows the reactive current $\mathbf{J}_{AB}(\mathbf{x})$, overlaid on the $A \rightarrow B$ reactive density in the background, i.e., the orange curves in the right column of Fig. 2. Gray patches indicate the metastable regions A and B , where reactive density is zero by definition (modulo the projected-out dimensions). Whereas the trajectories in Fig. 1(c) and the time series in Fig. 2 give us only a few noisy transition path samples, following the reactive current through state space allows to understand the transition dynamics at an ensemble level. To corroborate the faithful representation of transition pathways, and to compare with a more classical method, we have also plotted five realized transition paths (in both directions) from the reference simulation in blue, as well as the minimum-action pathways in cyan.

In Fig. 5(b), \mathbf{J}_{AB} emerges from set A with gradually increasing wave amplitude and decreasing zonal wind. Fol-

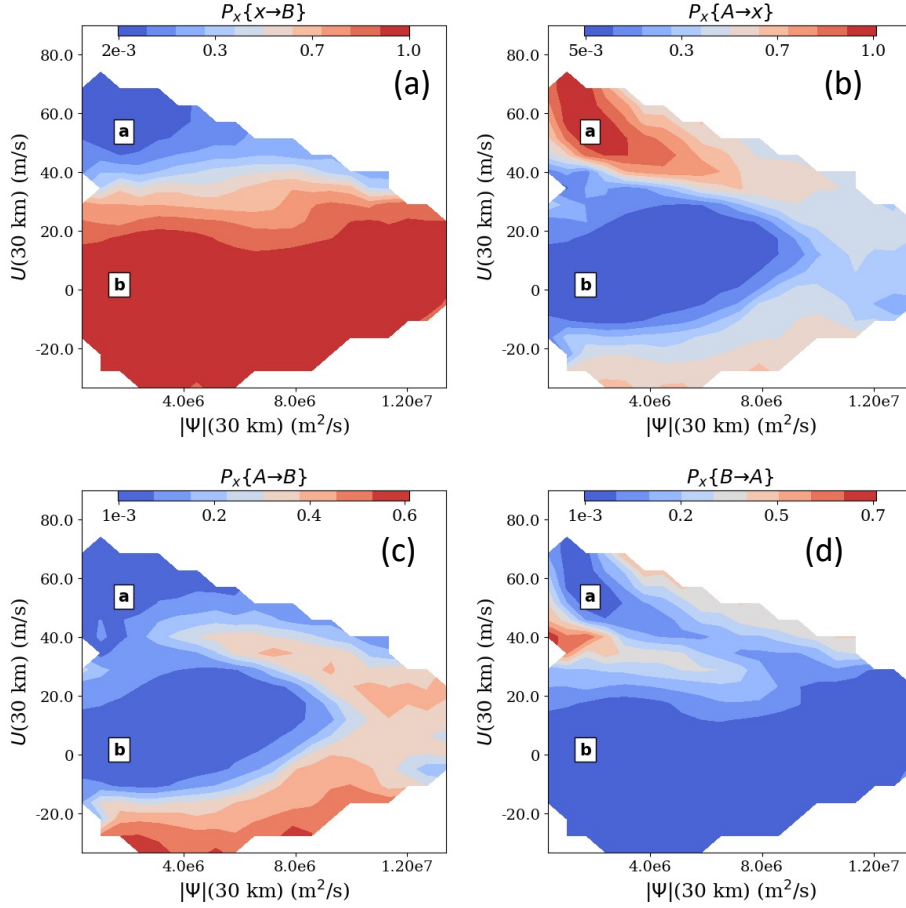


FIG. 4. **Committors and transition probabilities.** First row: committors $q^+(\mathbf{x}) = \mathbb{P}_{\mathbf{x}}\{\mathbf{x} \rightarrow B\}$ (a), the probability of hitting B next, and $q^-(\mathbf{x}) = \mathbb{P}_{\mathbf{x}}\{A \rightarrow \mathbf{x}\}$ (b), the probability of hitting A next. Second row: $q^+(\mathbf{x})q^-(\mathbf{x}) = \mathbb{P}_{\mathbf{x}}\{A \rightarrow B\}$ (c), the probability of being en route from A to B , and $(1 - q^+(\mathbf{x}))(1 - q^-(\mathbf{x})) = \mathbb{P}_{\mathbf{x}}\{B \rightarrow A\}$ (d), the probability of being en route from B to A . All fields are projected onto the $(|\Psi|, U)$ plane at 30 km.

lowing this current highway farther away from A , the field weakens and then spreads out, indicating that pathways tend to meander more widely through this stage. The current flows to large streamfunction amplitude, and eventually extremely negative zonal wind before approaching set B from below, generally tracking along the red crescent of Fig. 4(a). But the current does not completely stall out when hitting B ; instead, the current loops cyclically around B , indicating that some paths *orbit* the weak-zonal-wind fixed point, perhaps several times, before approaching closely enough to enter B . This behavior, of course, depends on the size and shape we assigned to set B : with a smaller radius, more cycling occurs, and with a larger radius, paths will generally terminate sooner and cycling will become less prevalent.

From a weather forecasting perspective, there may be little interest in charting the evolution of an SSW after the first dip into negative zonal wind territory, which is

why we used a different definition of A and B in Finkel et al. (2021). In that case, transition paths end by definition as soon as the wind reverses, and any further oscillations are considered part of the $B \rightarrow B$ phase, not the $A \rightarrow B$ phase. But from a dynamical systems perspective, these oscillations are robust features of the event's final stages, and distinct from the following oscillations in the $B \rightarrow B$ phase.

The cyan-colored minimum-action pathway 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). The pathway solves an optimization problem, deviating as minimally as possible from the deterministic dynamics while still bridging the gap all the way from A to B . These deviations from deterministic behavior represent the minimum necessary stochastic forcing required to actuate the transition. We have computed the least action path by minimizing the integrated magnitude of perturba-

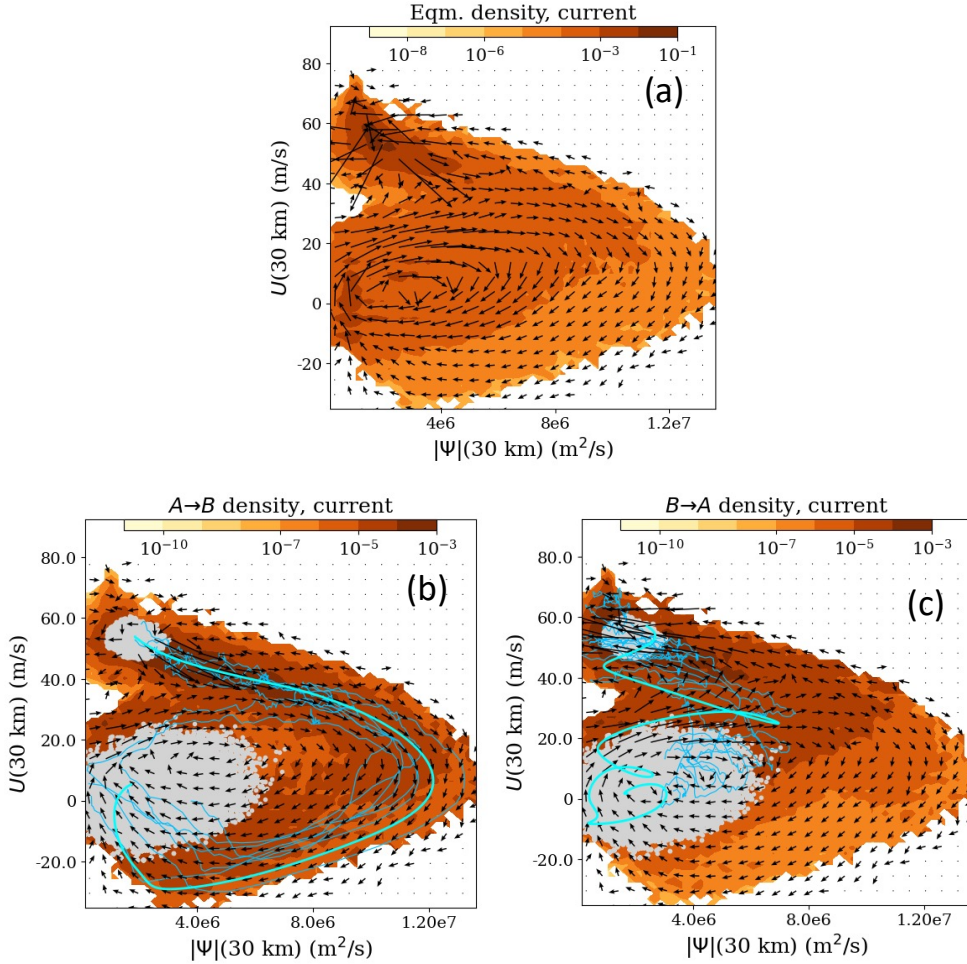


FIG. 5. **Densities and currents.** (a) shows the equilibrium density $\pi(\mathbf{x})$ and equilibrium current $\mathbf{J}(\mathbf{x})$. (b) and (c) show $\mathbf{J}_{AB}(\mathbf{x})$ overlaid on $\pi_{AB}(\mathbf{x})$ and $\mathbf{J}_{BA}(\mathbf{x})$ overlaid on $\pi_{BA}(\mathbf{x})$ respectively. Cyan curves mark the minimum-action pathways in each direction, while blue curves are sampled realized transition pathways. Gray dots are data points inside A and B , which are spheres in 75 dimensions, so some transition path segments appear to overlap with A and B in this projection.

tions while enforcing the end constraint with a quadratic penalty, a procedure that follows Plotkin et al. (2019) and is described further in the supplement. As the noise strength (i.e., stochastic forcing) shrinks to zero, we expect the reactive current to collapse into a single streamline following the minimum-action path. The finite-noise transition path ensemble, however, departs significantly from it. In the initial stages of transition in panels (c,d) of Fig. 5, the minimum-action path tracks right down the center of the reactive current, suggesting this feature is stable with noise. But in the latter part of the process, the minimum-action path seems to exaggerate the large-wave amplitude and low-wind extremes of transition paths. The general effect of nonzero noise is to mitigate the extreme spike in zonal wind.

We turn now to the $B \rightarrow A$ current in Fig. 5(c), which is very different from a reversed $A \rightarrow B$ current (another sign of irreversible dynamics). However, there is symmetry between the two directions. \mathbf{J}_{AB} dives beneath set B in an exceptional wind collapse, and subsequently merges smoothly from below with the equilibrium vacillation cycles. After an extended stay in B , \mathbf{J}_{BA} escapes out the *top* of the vacillation loop, breaking the cycle during the wave amplification phase by climbing upward in U space ever so slightly, to breach the “basin boundary” from below. At this point, the vortex has reached a sufficiently strong state to inhibit wave propagation, and radiative relaxation takes over.

The minimum-action path from B to A begins with some tortuous maneuvers to escape from the stable spiral **b**, but upon reaching the boundary of set B it follows a simi-

lar route as the reactive current. The path escapes from a vacillation cycle at the top of B , during the strong-wind phase, giving the radiative forcing a chance to pull the vortex back into shape. But as with the $A \rightarrow B$ direction, the minimum-action path strays significantly from the vector field streamlines, especially in the final stages with the negative IHF anomaly: while \mathbf{J}_{BA} enters A from the right (high $|\Psi|$), the minimum-action path enters A from below.

The minimum-action paths depend in part on choices in the optimization strategy, as detailed in the supplement. However, the systematic differences with reactive current are robust with respect to the optimization parameters. Moreover, any *single* pathway cannot possibly represent the variety of paths present that the reactive current finds.

e. Visualizing transition states

In the visualization of reactive current and minimum-action paths, we have chosen a two-dimensional space to provide some physical insight into the transition path mechanisms; details on this projection are provided in the supplement and in Strahan et al. (2021). However, any two-dimensional projection must sacrifice all other degrees of freedom, here 73, missing interaction between dynamical fields and altitudes which simply cannot be visualized in this kind of two-dimensional plot. In Finkel et al. (2021), we partially addressed this problem with sparse regression to approximate the committor through a small number of highly predictive physical proxies. This seems promising for prediction—it identifies the key observations required for a forecast—but it leaves much to be desired for detailed physical understanding. We offer further visualizations in this section to aggregate the higher-dimensional dependencies, and to compare the transition according to TPT and action minimization. For the remainder of the paper, we focus on the $A \rightarrow B$ transition, the prototype for a SSW.

The minimum-action method is easy to visualize; it is one single path, whereas TPT provides the full range of paths. However, TPT quantities can be used to reconstruct snapshots of “transition states” that are typical in the following sense. A defining property of reactive current is *constant flux across dividing surfaces*: if we draw a surface C in the $(|\Psi|, U)$ plane of Fig. 5 that encircles A without intersecting B , and compute the total outward flux $\int_C \mathbf{J}_{AB} \cdot \mathbf{n} d\sigma$, the result is a constant regardless of which surface C is chosen. That constant is the rate, or inverse return time, which is the average number of SSW events in a given long time period. (Here $d\sigma$ is an area element on C .) An expanding sequence of dividing surfaces can be constructed to bridge the gap all the way from A to B , each one supporting a different flux distribution with the same integral. Stringing together the peaks of the distributions, we can assemble a (discretized) typical path from A to B , as well as a spread around it. This is a simpler version of the transition tubes defined in Vanden-Eijnden (2006); we call

it the *high probability flux* path. It is not to be interpreted as the path of a single event, but rather as the flow of SSW “traffic” through a sequence of thresholds.

A natural choice of dividing surface is a level set of the forward committor shown in 4(a), i.e., all states where the likelihood of a SSW is equal. To represent both typical paths and their variability at different stages through the transition, we pick out three committor levels (0.1, 0.5, 0.9) and identify as candidates all data points (from our finite data set) whose estimated committor is within a small margin (0.05) of the selected level. For committor level sets, the surface normal vector is $\mathbf{n} = \nabla q^+ / |\nabla q^+|$. At each point on the level set, we can therefore calculate $\mathbf{J}_{AB} \cdot \mathbf{n}$, the local contribution to the total flux (the rate) across the surface. The points with the greatest local flux density are in that sense most representative of the pathway through this surface. For each committor level, we have identified all points with flux density with at least 10% of the maximum magnitude, giving a discrete weighted sampling of the flux distribution on each surface. Fig. 6 shows the mean and spread of this distribution in terms of the zonal wind and heat flux profiles. Blue and red dashed curves represent the fixed points **a** and **b**, the solid curves show the mean profiles at each committor level, and the shading shows the range between the 5th and 95th percentile (note they are not symmetric about the mean). The supplement provides details on how these statistics are calculated.

The progression from A to B starts off gradual and constrained, but in the latter stages explodes in both magnitude and variability. The $q^+ = 0.1$ and $q^+ = 0.5$ profiles form tight bundles, consistent with the narrow stream of \mathbf{J}_{AB} close to A , while the $q^+ = 0.9$ profiles spread out dramatically, especially in terms of heat flux. Another striking aspect is that the upper level winds remain quite close to the radiative solution even for high committor levels, a long way into the transition process; the collapse of upper level wind happens very fast and very late after the vortex has practically committed to breaking down. Comparing the wind profiles at a forecast probability of 0.1 vs. 0.5 suggests that the key evidence of an impending SSW is a subtle, but significant weakening of the midlevel vortex between 20 and 40 km. There is more spread in the lower-level winds around 10–20 km at the 50% point, but not a significant difference in the mean. The key role of the mid-level winds suggests the importance of preconditioning the vortex (McIntyre 1982; Albers and Birner 2014). Strong winds inhibit upward wave propagation (e.g., Charney and Drazin 1961) and an initial weakening of the mid-level winds allows for the abrupt burst of wave activity in the SSW event.

In terms of the heat flux, the key evidence of an impending warming event is the subtle increase in heat flux (i.e., upward wave propagation) at lower levels, 10 km and lower. The increase in wave propagation from the boundary is the key precursor. The explosion of upper levels (the 0.9

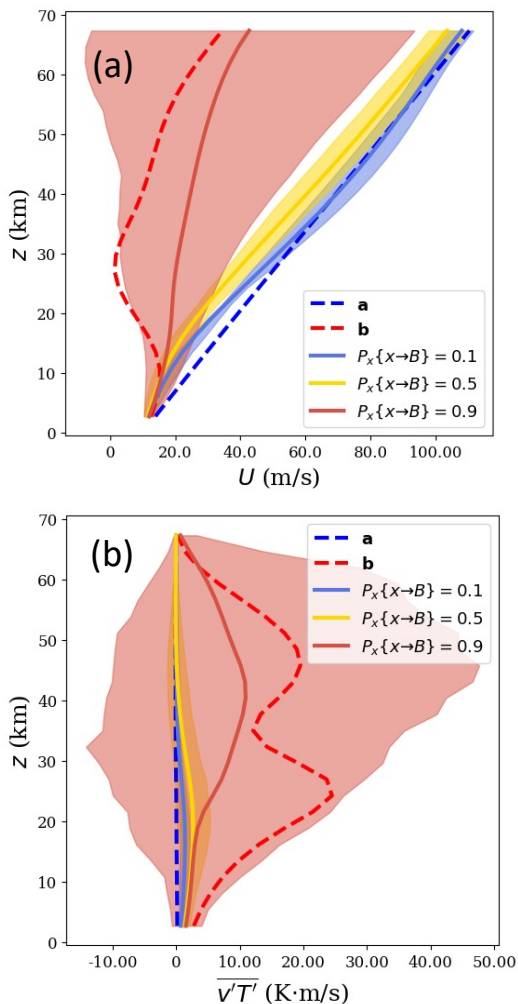


FIG. 6. Typical transition states and variability. For a sequence of three committor level surfaces, we plot (a) the zonal wind profile and (b) the meridional heat flux profile that is most typical of that surface in the sense of reactive current flux density. Shading represents the 5th and 95th percentiles of the flux distribution. Blue, gold, and red curves come from the 0.1, 0.5, and 0.9 level sets of the committor, respectively. Blue and red dashed curves represent the profiles for the fixed points **a** and **b**, respectively.

surface) comes when the event is already in progress. The middle altitudes may be critical for this model’s preconditioning, or receptivity to planetary waves propagating from below. Preconditioning is a key concept for predictability, and has been explored in Bancalá et al. (2012) and Albers and Birner (2014) in reanalysis data. The details of preconditioning are more complex in three-dimensional resolved models, and it may furthermore differ between split- and displacement-type warmings.

While Fig. 6 divides the transition into stages based on the committor, it does not indicate how long each stage takes. To compare TPT more directly with the minimum-

action pathway, which is parameterized by time, we choose another set of dividing surfaces defined as level sets of the *lead time* to the SSW, i.e., to reach set B :

$$\eta^+(\mathbf{x}) = \mathbb{E}_{\mathbf{x}}[\tau^+ | \mathbf{x} \rightarrow B]. \quad (25)$$

This measures the progress from A to B in terms of time, as presented in Finkel et al. (2021) as an important forecasting metric. We select 15 different level sets of η^+ from 100 days to 0 days, matching the time horizon of the minimum-action path, and compute the mean and standard deviation wind profile as described above. Fig. 7 displays both minimum-action path (left) and high probability flux path (right), showing the evolution of $U(30 \text{ km})$, $U(z)$, and $|\Psi(z)|$ profiles. The two paths are qualitatively similar, both zonal wind profiles slowly decaying and then rapidly plummeting down to **b**. But they differ, primarily in the magnitude of extremal behavior: the zonal wind dips lower and the wave magnitude rises higher in the minimum-action path as opposed to the high probability flux path.

A vertical dashed line marks the beginning of the conventional SSW event, when $U(30 \text{ km})$ first dips below the threshold for state **a**. While this event is clearly defined for the minimum-action path, in Fig. 7(b) the high probability flux path actually stagnates in its descent, not crossing the threshold until the transition is almost finished. This is not because real transition paths don’t plummet to negative wind—Fig. 5 clearly shows that they do, both from the current and from the observed path samples—but because the subsequent large oscillations around set B cancel out the negative spike, on average. The orange flux distribution spreads out rapidly even as the mean profile flatlines, confirming that mean behavior belies some important variability in the transition path ensemble. We thus put the vertical line in the right-hand column where the 5th percentile of zonal wind drops below **a**, as the minimum-action path corresponds most closely to the negative tail of the transition path ensemble. This is consistent with the comparison between the minimum-action path and reactive flux in Fig. 5: nonzero noise tends to (1) blunt the largest negative zonal wind spikes, and (2) lead to subsequent large positive oscillations before the vortex enters B .

4. Transition path statistics and numerical benchmarks

In this section we present various statistics over the transition path ensemble, including metrics of SSW severity. We also report numerical benchmarks to assess the accuracy of the DGA method.

a. Numerical method: DGA

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 to provide context for the statistics to follow. We generate a data set by

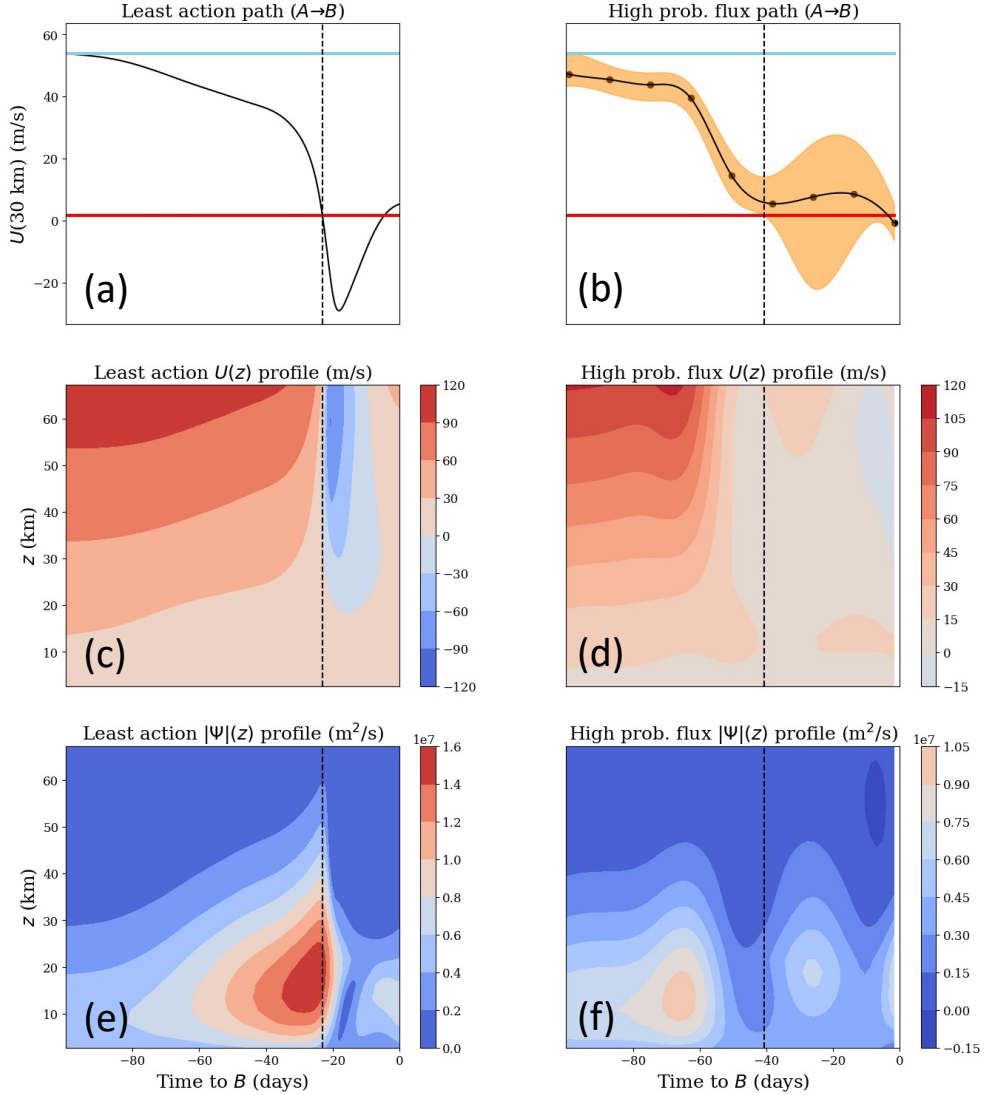


FIG. 7. **Minimum-action paths (left) and high probability flux paths (right).** The first row shows $U(30 \text{ km})$ as a function of time, while the second row displays the evolution of the whole zonal wind profile $U(z)$. The third row shows the evolution of the profile of wave amplitude $|\Psi(z)|$. The high probability flux path is defined as a sequence of maximizers of the flux distribution on level sets of the lead time to B (see text for details). The two paths are similar, but the minimum-action path goes through larger extremes in both wind and wave amplitude.

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 $\{\mathbf{X}_n(t) : 0 \leq t \leq \Delta t\}_{n=1}^N$. This sampling measure, the number $N = 3 \times 10^5$ trajectories, and the length $\Delta t = 20$ days, are key parameters of the method. The trajectories are significantly shorter than the typical ~ 100 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 control simulation to seed the initial points,

but here we focus on TPT and DGA as a proof of concept rather than optimizing the numerical procedure.

After generating the data, we expand several unknown “forecast functions” of interest— $q^+(\mathbf{x})$, $q^-(\mathbf{x})$, $\eta^+(\mathbf{x})$, $\pi(\mathbf{x})$ etc, see the supplement—in basis sets informed by the data, and then solve matrix equations for the expansion coefficients. The choice of basis is another key parameter of the method. The entries in the matrix equations 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. The forecast functions then lead directly to all the other quantities presented here,

including reactive densities, currents, and the path integral moments to be described next.

DGA is an attractive alternative to DNS because it is parallelizable and flexible with respect to the sampling distribution μ . DNS, by contrast, samples state space according to π , which is small in the transition regions between A and B . DNS generates transition events only rarely, and thus the transition path statistics of interest may take a long time to converge. As DGA is a new, rapidly developing method, rigorous error bounds are not yet available for general systems, and certainly beyond the scope of this paper. In practice, the performance of DGA depends strongly on the specific choice of sampling measure, basis functions, and lag time Δt . Because we have performed both DGA and DNS on the Holton-Mass model, we can compare them quantitatively, and furthermore put error bars on DNS, as a rough validation, albeit with significant room for improvement.

b. Forecast maps: distinguishing transition routes

With reactive densities, we describe the properties of instantaneous configurations along transition paths. With reactive currents, we describe the movement between successive configurations during a transition. Here, we go one step further and describe the properties of full, coherent pathways as discrete objects drawn from an ensemble. We focus on just two simple summary statistics that are physically relevant to our insights so far. The first statistic is the *transit time*: how long it takes for the vortex to break down completely. Formally, transit time is $\tau^+ - \tau^-$, where τ^- is the beginning of the path (when it leaves A) and τ^+ is the end of the path (when it hits B). The second statistic is the total heat flux, $\int_{\tau^-}^{\tau^+} \overline{v'T'}(30\text{km}) dt$: a quantification of the size of the spikes observed in Fig. 2(c) during the $A \rightarrow B$ SSW phase. Both quantities are random variables that differ from event to event, and we would like to distinguish different pathways based on their transit time and total heat flux. We can apply the same analysis to any *path integral* of the form

$$\int_{\tau^-}^{\tau^+} \Gamma(\mathbf{X}(t)) dt \quad (26)$$

where Γ is a physical observable to represent the magnitude, or severity, of an event.

Fig. 8(a) shows the expected transit time of all $(A \rightarrow B)$ -transition paths passing through \mathbf{x} ,

$$\mathbb{E}_{\mathbf{x}}[\tau^+ - \tau^- | A \rightarrow B]. \quad (27)$$

The blue regions indicate a small $\tau^+ - \tau^-$, i.e., the *fast* routes through state space between A and B . The red regions indicate traps, where stray transition pathways encounter delays. The fastest route generally follows the reactive current, with a large spike in wave amplitude preceeding the zonal wind dropoff.

Next consider the total heat flux at 30 km. Panel (b) shows the expected total heat flux of transition paths crossing through \mathbf{x} , while panel (c) shows the instantaneous heat flux at 30 km as a static function of state space, viz. Equation (9), highlighting its association with large wave magnitude and weak zonal wind. Counterintuitively, the low-heat flux path route passes through the high-heat flux region of state space. To resolve this paradox, we turn back to the reactive current \mathbf{J}_{AB} in Fig. 5 and note again the presence of extra loops around set B in the final stages of transition. A proportion of transition paths execute some extra loops, having missed the set the first time around, and accumulate more heat flux into the path integral. The red interior region of large total heat flux in Fig. 8(c) is a signature of these multi-loop pathways, one of which is already shown in the relatively brief simulation window of Figs. 1 and 2. Panel (d) draws attention to where total heat flux is large and total time is small: i.e., the most intense events. It appears greatest right where transition paths tend to enter set B after the first major deceleration of the vortex, similar to the minimum-action path shown in Fig. 5(b). This is the sweet spot where paths reach the weak vortex state without getting trapped in a number of cycles about B .

c. Path integral distributions

Whereas Fig. 8 shows expectations over all transition paths passing through a given \mathbf{x} , the same expectations—and higher moments—can be computed for all transition paths irrespective of their route, using DGA. While TPT applications in chemistry have mainly focused on the *mean* passage time, higher moments may be important for risk analysis. When transition path integrals are defined to represent the severity or damage of a certain event, knowledge of their full distribution is important for building *uncertainty* estimates into disaster response planning. While the Holton-Mass model is too idealized to interpret any path integral as a concrete damage indicator, natural disasters have clear quantifiable notions of damage that vary widely between events. Earthquakes and wildfires, for example, seem to follow power laws (Christensen et al. 2002; Malamud et al. 2005), as do failures in some engineered systems such as power grids (Dobson et al. 2007) and financial markets (Gabaix 2009). Recent analysis of birth-death processes in ecological models has revealed universal behavior in fixation time distributions (Ashcroft et al. 2015; Hathcock and Strogatz 2019, 2021).

The ability of DGA to access higher moments of distributions may help to estimate the upper echelons of risk without having to run long, expensive simulations. Chebyshev's inequality, in particular, allows us to bound tail probabilities using moments. With this motivation in mind, we have estimated moments of time and total heat flux in the Holton-Mass model using DGA. Of course, DNS provides

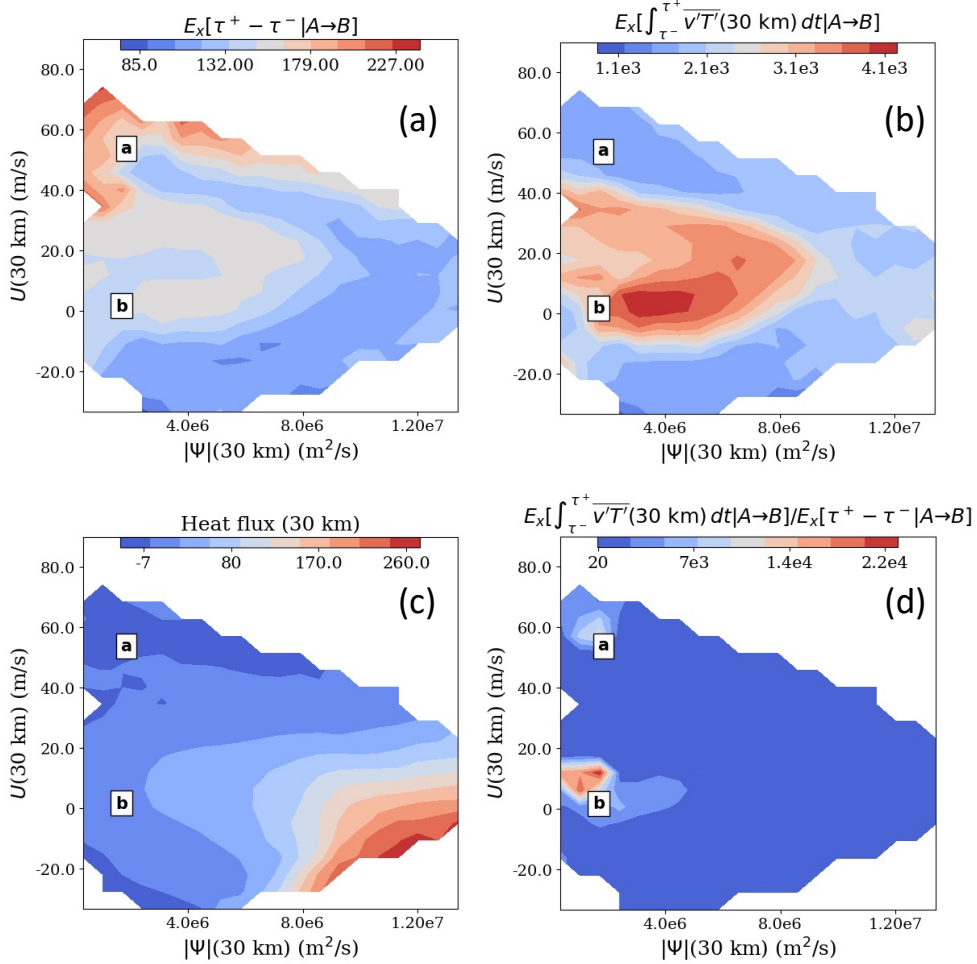


Fig. 8. **Transition path expectations.** Panel (a) shows a map of expected duration $\tau^+ - \tau^-$ of all transition paths crossing through \mathbf{x} , so red and blue regions represent slow and fast pathways. Panel (b) shows the corresponding expected total heat flux at 30 km. Panel (c) shows heat flux as a function of state space, and (d) shows the ratio of (b) to (a).

the same information given enough time, but in more expensive models we expect DGA to become more favorable. We will present both sets of results together, simultaneously bolstering our scientific claims and validating the DGA method.

The first two moments are visually assessed in Fig. 9(a), where we have plotted the transition path duration against the integrated heat flux for all observed transitions from the control simulation. Cyan and red crosses indicates the mean and ± 1 standard deviation envelope estimated from DNS and DGA respectively. Note that the crosses are not error bars on the mean, but estimates of the variances of the transition path integral distributions. DNS and DGA agree approximately in these first two moments, but the distribution is plainly skewed, indicating that higher moments are necessary to fully describe it. Panels (b) and (c) show the first three moments, normalized to the same units

by taking a k -norm,

$$\left\{ \mathbb{E} \left[\left(\int_{\tau^-}^{\tau^+} \Gamma(\mathbf{X}(t)) dt \right)^k \right] \right\}^{1/k}, \quad (28)$$

which have the same dimensions of $[\Gamma] \cdot [T]$ for $k = 1, 2, 3, \dots$

The two methods are consistent given the 95% confidence interval on DNS and produce similar moment patterns as a function of moment number k . The errors suffered by DNS are due to poorly sampled tails in the distribution, which are rare events among rare events. The fundamental data scarcity problem applies to the transition path ensemble itself, and we have found the moment error bars to decrease slowly as the DNS data set increases.

DGA is not perfect either, displaying a consistent bias in the time moments relative to DNS, which may depend sensitively on the definition of B . In principle, one can use

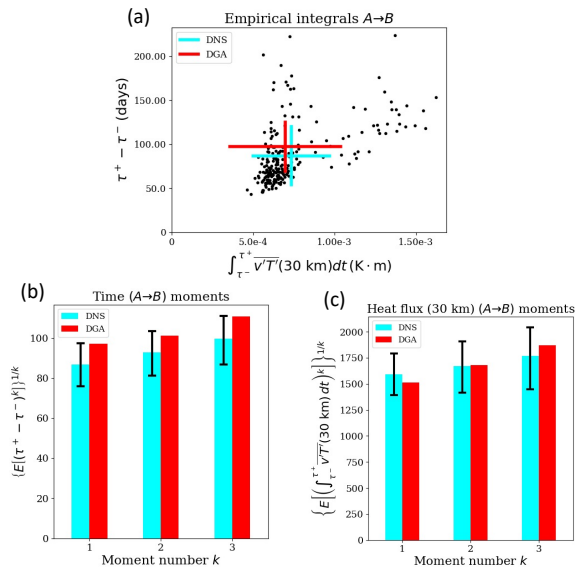


FIG. 9. **Transition path integral distributions.** In (a), each black dot is a different transition path realized from the long control simulation, and the cyan crosses mark the mean and standard deviation of the empirical distribution. Red crosses mark the mean and standard deviation computed from DGA. The lower two panels show expected transition path integral moments 1-3, both of duration and total heat flux, and both from DGA and DNS.

these moments to describe the distribution’s shape in more detail, e.g., with skew and kurtosis. At this stage of our method, these higher-order statistics are not yet reliable to draw confident conclusions. We present these results to fully document the methodology and to highlight areas in need of further improvements. While DNS accuracy is fundamentally limited by the number of sampled paths, DGA has the potential for improvement through the basis sets and the sampling strategy. As these avenues are explored in the future, DGA may become increasingly advantageous relative to DNS as one becomes more interested in the tails of the path integral distribution.

Fig. 9(a) also makes the important point that the two path integrals—total time and total heat flux—are not perfectly correlated; that is, transitions with large heat flux aren’t necessarily so just because they have longer duration, and likewise very long-lasting transition paths can maintain a fairly low total heat flux. The DNS distribution of $A \rightarrow B$ path integrals appears to have two branches, one with no correlation between time and heat flux and the other with high correlation. Hence the two path integrals are truly independent pieces of information. Here, we can only see this because of DNS, but in principle DGA can also estimate the joint moments of time and heat flux.

5. Conclusion

Extreme weather events are central to the modern challenge of climate modeling. As the impacts of flooding, heat waves, cold spells, and other natural disasters become increasingly stark, the scientific community is endeavoring to better capture these events, as well as average behavior, in numerical models. While many exciting techniques are being developed to simulate and diagnose rare events, there is an overall lack of standard language and benchmarks to do so, due to the sheer variety of extreme-weather phenomena under study. Closely coupled to this conceptual problem is the computational problem that rare events take a long time to appear and hence a very long time to produce a significant statistical distribution, whether in an observational record or in a simulation.

We have advocated two major ideas to advance an organized approach to extreme weather modeling. First, as a conceptual demonstration, we have presented a detailed transition path theory (TPT) analysis of a prototypical extreme event, sudden stratospheric warmings in the Holton-Mass model. TPT provides a set of summary statistics that encapsulate important features of rare events, including frequency, precursors, and various severity metrics. Reactive densities and reactive currents tell us how the system evolves through state space to a SSW event, as well as momentum and heat transfers along the way. The minimum-action method provides a useful but limited point of comparison, as it provides no information about the *variability* of transitions. We have furthermore extended the TPT framework to calculate statistics of the total time elapsed and integrated heat flux during a vortex breakdown event, both for the entire path ensemble and for different kinds of paths through state space. Comparing reactive current with maps of various forecasts and aftcasts, we can infer many qualitative properties of the path ensemble, especially what makes transition paths unique relative to “everyday” dynamics.

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 demonstrated in Finkel et al. (2021) for the narrow goal of forecasting sudden stratospheric warming events in the Holton-Mass model. In the current paper we have used the same computational method to ask more intricate statistical questions about the evolution of SSW from start to finish.

Our work is an early application of TPT to atmospheric science, where 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 becoming recognized as informative for 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). We stress that TPT and DGA are specialized to describe certain aspects of rare events, but not all aspects. Our method targets specific integral quantities of the form (26), but nonlinear functionals may also be worth computing. Furthermore, successive rare events may carry long-time correlations, which we have not quantified. For example, a large earthquake might release enough tectonic stress to make the next one less severe. Our approach will require further extensions to address such issues.

Significant challenges 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 communicating the 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 dataset 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. (2006b); 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); Webber et al. (2019); Bouchet et al. (2019a,b); Plotkin et al. (2019); Simonnet et al. (2020); 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. In the meantime, our calculations here serve as a conceptual and numerical foundation for holistic rare event description in atmospheric dynamics.

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. R.J.W. is 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. The computations in the paper were done

on the high-performance computing clusters 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.
- Ashcroft, P., A. Traulsen, and T. Galla, 2015: When the mean is not enough: Calculating fixation time distributions in birth-death processes. *Phys. Rev. E*, **92**, 042 154, doi:10.1103/PhysRevE.92.042154, URL <https://link.aps.org/doi/10.1103/PhysRevE.92.042154>.
- Bancalá, S., K. Krüger, and M. Giorgetta, 2012: The preconditioning of major sudden stratospheric warmings. *Journal of Geophysical Research: Atmospheres*, **117** (D4), doi:10.1029/2011JD016769.
- Berner, J., and Coauthors, 2017: Stochastic parameterization: Toward a new view of weather and climate models. *Bulletin of the American Meteorological Society*, **98** (3), 565 – 588, doi: 10.1175/BAMS-D-15-00268.1, URL <https://journals.ametsoc.org/view/journals/bams/98/3/bams-d-15-00268.1.xml>.
- 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.
- Chan, P.-W., P. Hassanzadeh, and Z. Kuang, 2019: Evaluating indices of blocking anticyclones in terms of their linear relations with surface hot extremes. *Geophysical Research Letters*, **46** (9), 4904–4912, doi:<https://doi.org/10.1029/2019GL083307>, URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019GL083307>, <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2019GL083307>.

- 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>.
- Chattopadhyay, A., E. Nabizadeh, and P. Hassanzadeh, 2020: Analog forecasting of extreme-causing weather patterns using deep learning. *Journal of Advances in Modeling Earth Systems*, **12** (2), e2019MS001958, doi:https://doi.org/10.1029/2019MS001958, URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001958>, <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2019MS001958>.
- 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>.
- Chen, N., F. Gilani, and J. Harlim, 2021: A bayesian machine learning algorithm for predicting enso using short observational time series. 2104.01435.
- 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.
- Christensen, K., L. Danon, T. Scanlon, and P. Bak, 2002: Unified scaling law for earthquakes. *Proceedings of the National Academy of Sciences*, **99** (suppl 1), 2509–2513, doi:10.1073/pnas.012581099, URL https://www.pnas.org/content/99/suppl_1/2509, https://www.pnas.org/content/99/suppl_1/2509.full.pdf.
- 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.
- Dobson, I., B. A. Carreras, V. E. Lynch, and D. E. Newman, 2007: Complex systems analysis of series of blackouts: Cascading failure, critical points, and self-organization. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **17** (2), 026 103, doi:10.1063/1.2737822, URL <https://doi.org/10.1063/1.2737822>, <https://doi.org/10.1063/1.2737822>.
- 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/jas190278.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.
- Gabaix, X., 2009: Power laws in economics and finance. *Annual Review of Economics*, **1** (1), 255–294, doi:10.1146/annurev.economics.050708.142940, URL <https://doi.org/10.1146/annurev.economics.050708.142940>, <https://doi.org/10.1146/annurev.economics.050708.142940>.
- Gálfi, V. M., V. Lucarini, and J. Wouters, 2019: A large deviation theory-based analysis of heat waves and cold spells in a simplified model of the general circulation of the atmosphere. *Journal of Statistical Mechanics: Theory and Experiment*, **2019** (3), 033 404.
- Hathcock, D., and S. H. Strogatz, 2019: Fitness dependence of the fixation-time distribution for evolutionary dynamics on graphs. *Phys. Rev. E*, **100**, 012 408, doi:10.1103/PhysRevE.100.012408, URL <https://link.aps.org/doi/10.1103/PhysRevE.100.012408>.
- Hathcock, D., and S. H. Strogatz, 2021: Universal absorption-time distributions in birth-death markov processes. 2103.08522.
- 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., J. Henderson, S. Leidner, C. Grassotti, and T. Nehr Korn, 2006a: Using 4d-var to move a simulated tropical cyclone

- in a mesoscale model. *Computers & Mathematics with Applications*, **52** (8), 1193–1204, doi:<https://doi.org/10.1016/j.camwa.2006.11.013>, URL <https://www.sciencedirect.com/science/article/pii/S0898122106003257>, variational Data Assimilation and Optimal Control.
- Hoffman, R. N., J. M. Henderson, S. M. Leidner, C. Grassotti, and T. Nehr Korn, 2006b: 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), 083102, 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), 164902, doi:10.1063/1.2186317, URL <https://doi.org/10.1063/1.2186317>, <https://doi.org/10.1063/1.2186317>.
- Katz, R. W., 2010: Statistics of extremes in climate change. *Climatic Change*, **100**, 71–76, doi:10.1007/s10584-010-9834-5.
- 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>.
- Li, J., C. Wasko, F. Johnson, J. P. Evans, and A. Sharma, 2018: Can regional climate modeling capture the observed changes in spatial organization of extreme storms at higher temperatures? *Geophysical Research Letters*, **45** (9), 4475–4484, doi:<https://doi.org/10.1029/2018GL077716>, URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018GL077716>, <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2018GL077716>.
- Liu, Y., D. P. Hickey, S. D. Minter, A. Dickson, and S. Calabrese Barton, 2019: Markov-State Transition Path Analysis of Electrostatic Channeling. *The Journal of Physical Chemistry C*, **123** (24), 15284–15292, doi:10.1021/acs.jpcc.9b02844, URL <https://doi.org/10.1021/acs.jpcc.9b02844>, publisher: American Chemical Society.
- Lucarini, V., and Coauthors, 2016: Extremes and recurrence in dynamical systems. 1605.07006.
- Lucente, D., S. Duffner, C. Herbert, J. Rolland, and F. Bouchet, 2019: Machine learning of committor functions for predicting high impact climate events. *Climate Informatics*, Paris, France, URL <https://hal.archives-ouvertes.fr/hal-02322370>.
- 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.
- Malamud, B. D., J. D. A. Millington, and G. L. W. Perry, 2005: Characterizing wildfire regimes in the united states. *Proceedings of the National Academy of Sciences*, **102** (13), 4694–4699, doi:10.1073/pnas.0500880102, URL <https://www.pnas.org/content/102/13/4694>, <https://www.pnas.org/content/102/13/4694.full.pdf>.
- 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), 45242.
- Matthewman, N. J., and J. G. Esler, 2011: Stratospheric sudden warmings as self-tuning resonances. part i: Vortex splitting events. *Journal of the Atmospheric Sciences*, **68** (11), 2481 – 2504, doi:10.1175/JAS-D-11-07.1, URL <https://journals.ametsoc.org/view/journals/atsc/68/11/jas-d-11-07.1.xml>.
- McIntyre, M. E., 1982: How well do we understand the dynamics of stratospheric warmings? *Journal of the Meteorological Society of Japan. Ser. II*, **60** (1), 37–65, doi:10.2151/jmsj1965.60.1_37.
- 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>.
- 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.
- 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), 11138–11143, doi:10.1073/pnas.1813263115, URL <https://www.pnas.org/content/115/44/11138>, <https://www.pnas.org/content/115/44/11138.full.pdf>.
- Muratov, C. B., E. Vanden-Eijnden, and W. E, 2005: Self-induced stochastic resonance in excitable systems. *Physica D: Nonlinear Phenomena*, **210** (3), 227–240, doi:<https://doi.org/10.1016/j.physd.2005.07.014>, URL <https://www.sciencedirect.com/science/article/pii/S0167278905003155>.
- 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), 19011–19016, 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), 19011–19016, doi:10.1073/pnas.0905466106, URL <https://www.pnas.org/content/106/45/19011>, <https://www.pnas.org/content/106/45/19011.full.pdf>.
- O’Gorman, P. A., and J. G. Dwyer, 2018: Using machine learning to parameterize moist convection: Potential for modeling of climate, climate change, and extreme events. *Journal of Advances in Modeling Earth Systems*, **10** (10), 2548–2563.

- 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:10.1029/2018MS001419.
- Qi, D., and A. J. Majda, 2020: Using machine learning to predict extreme events in complex systems. *Proceedings of the National Academy of Sciences*, **117** (1), 52–59, doi:10.1073/pnas.1917285117, URL <https://www.pnas.org/content/117/1/52>, <https://www.pnas.org/content/117/1/52.full.pdf>.
- 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>.
- 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>.
- Scott, R. K., and L. M. Polvani, 2006: Internal variability of the winter stratosphere. part i: Time-independent forcing. *Journal of the Atmospheric Sciences*, **63** (11), 2758–2776, doi:10.1175/JAS3797.1, URL <https://doi.org/10.1175/JAS3797.1>.
- Scott, R. K., L. M. Polvani, and D. W. Waugh, 2008: Internal variability of the winter stratosphere. part ii: Time-dependent forcing. *Journal of the Atmospheric Sciences*, **65** (7), 2375 – 2388, doi:10.1175/2007JAS2619.1, URL <https://journals.ametsoc.org/view/journals/atsc/65/7/2007jas2619.1.xml>.
- Simonnet, E., J. Rolland, and F. Bouchet, 2020: Multistability and rare spontaneous transitions between climate and jet configurations in a barotropic model of the jovian mid-latitude troposphere. 2009.09913.
- Sjoberg, J. P., and T. Birner, 2012: Transient tropospheric forcing of sudden stratospheric warmings. *Journal of the Atmospheric Sciences*, **69** (11), 3420–3432, doi:10.1175/JAS-D-11-0195.1.
- 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>.
- 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.
- 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.
- Wang, X., J. Slawinska, and D. Giannakis, 2020: Extended-range statistical ENSO prediction through operator-theoretic techniques for nonlinear dynamics. *Scientific Reports*, **10** (1), 2636, doi:10.1038/s41598-020-59128-7, URL <https://doi.org/10.1038/s41598-020-59128-7>.
- 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.
- Weiss, J. B., B. Fox-Kemper, D. Mandal, A. D. Nelson, and R. K. Zia, 2020: Nonequilibrium oscillations, probability angular momentum, and the climate system. *Journal of Statistical Physics*, **179** (5), 1010–1027.
- Woollings, T., and Coauthors, 2018: Blocking and its response to climate change. *Current Climate Change Reports*.
- 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](https://doi.org/10.1175/1520-0469(1987)044<3683:DAOSVI>2.0.CO;2).