Path Properties of Atmospheric Transitions: Illustration with a Low-Order Sudden Stratospheric Warming Model🖉

JUSTIN FINKEL

Committee on Computational and Applied Mathematics, University of Chicago, Chicago, Illinois

DORIAN S. ABBOT

Department of the Geophysical Sciences, University of Chicago, Chicago, Illinois

JONATHAN WEARE

Courant Institute of Mathematical Sciences, New York University, New York, New York

ABSTRACT

Many rare weather events, including hurricanes, droughts, and floods, dramatically impact human life. To accurately forecast these events and characterize their climatology requires specialized mathematical techniques to fully leverage the limited data that are available. Here we describe *transition path theory* (TPT), a framework originally developed for molecular simulation, and argue that it is a useful paradigm for developing mechanistic understanding of rare climate events. TPT provides a method to calculate statistical properties of the paths into the event. As an initial demonstration of the utility of TPT, we analyze a low-order model of sudden stratospheric warming (SSW), a dramatic disturbance to the polar vortex that can induce extreme cold spells at the surface in the midlatitudes. SSW events pose a major challenge for seasonal weather prediction because of their rapid, complex onset and development. Climate models struggle to capture the long-term statistics of SSW, owing to their diversity and intermittent nature. We use a stochastically forced Holton–Mass-type model with two stable states, corresponding to radiative equilibrium and a vacillating SSW-like regime. In this stochastic bistable setting, from certain probabilistic forecasts TPT facilitates estimation of dominant transition pathways and return times of transitions. These "dynamical statistics" are obtained by solving partial differential equations in the model's phase space. With future application to more complex models, TPT and its constituent quantities promise to improve the predictability of extreme weather events through both generation and principled evaluation of forecasts.

1. Introduction and background

The polar winter stratosphere typically supports a strong, cyclonic polar vortex, maintained by the thermal wind relation and meridional temperature gradient. A sudden stratospheric warming (SSW) event is a large excursion from this normal state, which can take many different forms. In split-type SSWs, the vortex splits completely in two. In displacement-type SSWs the vortex displaces far away from the pole (these can be considered wavenumber-2 and wavenumber-1 disturbances, respectively). Both types are "major" warmings, as the mean zonal wind reverses. In a "minor" warming, the zonal wind slows down significantly without completely reversing (Butler et al. 2015).

SSW is a rare event occurring about twice every 3 years, depending on the definition used (Butler et al. 2015). Its effects can propagate downward into the troposphere, altering the tropospheric jet stream and inducing extreme midlatitude surface weather events, including cold spells and precipitation (Baldwin and Dunkerton 2001; Thompson et al. 2002). Abrupt cold spells severely stress infrastructures, economies and human lives, and every bit of extra prediction lead time is helpful for adaptation. Unfortunately, numerical weather prediction struggles to forecast SSW at any

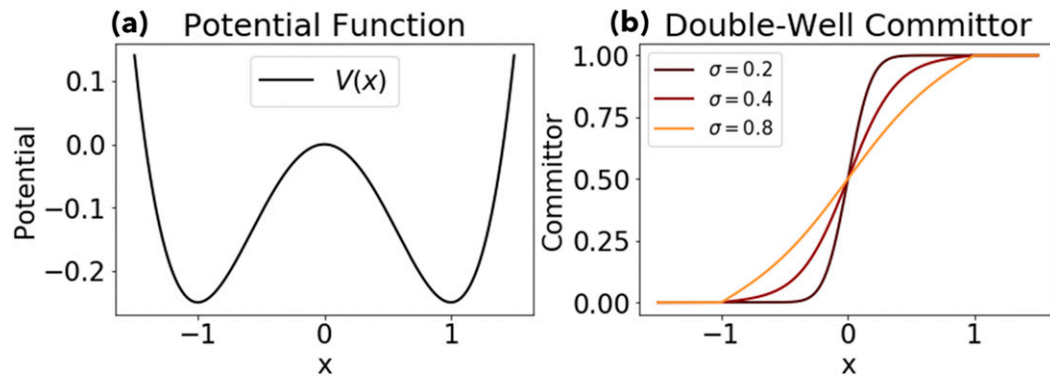*Corresponding author*: Justin Finkel, jfinkel@uchicago.edu

FIG. 1. The committor function for a double-well potential under the dynamics $\dot{x} = -V'(x) + \sigma\dot{w}$. (a) The potential function $V(x)$ and (b) the committor function. The committor has value zero on the left minimum, one on the right minimum, and one-half at the top of the barrier. The stronger the stochastic forcing, the less the actual potential shape matters and the more gradual the committor's slope. For small noise, the dynamics become more deterministic and the committor approaches a step function, since $x(t)$ will directly approach whichever minimum is closer.

lead time longer than about 2 weeks (Tripathi et al. 2016). Understanding SSW is therefore important for practical forecasting as well as science, but this task remains difficult. Several different geophysical fields are often used as indices of SSW onset. One simple indicator is zonal-mean zonal wind at 60°N, which defines thresholds for minor and major warming (Charlton and Polvani 2007; Butler et al. 2015). Another common indicator is the 10 hPa geopotential height field, which was used by Inatsu et al. (2015) to estimate a fluctuation–dissipation relation in its leading empirical orthogonal functions (EOFs). Many studies have examined SSW precursors and dominant pathways through simulation and observation. Limpasuvan et al. (2004), for instance, catalogued the various wavenumber forcings, heat fluxes and zonal wind anomalies that accompanied each stage of SSW events from reanalysis data. While planetary wave forcing from the troposphere is an accepted proximal cause of SSW, the polar vortex's susceptibility to such forcing, or "preconditioning," is a nontrivial and debated function of its geometry (Albers and Birner 2014; Bancalá et al. 2012). Tropospheric blocking is also thought to be linked to SSW; Martius et al. (2009) and Bao et al. (2017) found blocking to precede many major SSW events of the past half century. The diversity and complex life cycle of SSWs makes it difficult to build a unified picture of their onset.

In this article, we are interested in developing a detailed understanding of transition events between two states, at least one of which is typically long lived. Consider, for example, a particle with position $x(t)$ moving in the double-well potential energy landscape $V(x) = x^4/4 - x^2/2$ (illustrated in Fig. 1) and forced by stochastic white noise $\dot{W}$: $\dot{x} = -V'(x) + \sigma\dot{W}$. If the

system starts in the left well, it will tend to remain there a while, but occasionally the stochastic forcing will push it over the barrier into the right well. The natural predictor for this event is the *committor*: the probability of reaching the right well before the left well.

We denote this function by $q(x)$, which solves the Kolmogorov backward equation (to be introduced later). For this simple system the equation takes a form that can be solved exactly:

$$
\begin{cases}
-V'(x)q'(x) + \dfrac{\sigma^2}{2}q''(x) = 0 & x \in (-1, 1) \\[2mm]
q(-1) = 0,\ q(1) = 1
\end{cases}
$$

$$
\Rightarrow q(x) = \frac{\displaystyle\int_{-1}^{x} \exp\left[\frac{2}{\sigma^2}V(x')\right] dx'}{\displaystyle\int_{-1}^{1} \exp\left[\frac{2}{\sigma^2}V(x')\right] dx'}. \tag{1}
$$

Note that the boundary conditions are implied by the probabilistic interpretation.

The committor is plotted in the right panel of Fig. 1 for various noise levels. N.b., the potential landscape picture is not fully general, but is a useful mental model. The equations that determine $q(x)$ will be presented in section 2.

In the case of SSW, the long-lived states are the steady and disturbed circulation regimes of the stratospheric polar vortex. Recent work by Yasuda et al. (2017) has studied SSW in an equilibrium statistical mechanics framework, with these two stable states as saddle points of energy functionals. Transition path theory (TPT) takes a complementary nonequilibrium view, describing the long-time (steady state) statistics of trajectories

between the two states. For example, TPT introduces a probability density of reactive trajectories (or "reactive density") indicating the regions where trajectories tend to spend their time en route from A to B. The system is said to be *reactive* at a point in time if it has most recently visited A and will next visit B. The associated probability current of reactive trajectories (or "reactive current") indicates the preferred direction and speed of transition paths. These detailed descriptors of the mechanism underlying a rare event can be expressed in terms of probabilistic forecasts like the committor $q(x)$, the probability of entering state A before reaching state B from a given initial condition x (not in either A or B). The committor is the ideal probabilistic forecast in the usual variance-minimizing sense of conditional expectations (Durrett 2013). Any other predictor of a transition derived through experiments and observations, such as vortex preconditioning and forcing at different wavenumbers (Albers and Birner 2014; Bancalá et al. 2012; Martius et al. 2009; Bao et al. 2017) necessarily corresponds to an approximation of the committor.

Ensemble simulation is a commonly used method to estimate the committor at a single given initial condition by measuring the fraction of ensemble trajectories that achieve the rare event. This is challenging because many simulations are needed to generate enough rare events for significant statistical power. Recent and ongoing work aims to channel this computing power more efficiently in weather simulation using importance sampling and large deviation theory (Hoffman et al. 2006; Weare 2009; Vanden-Eijnden and Weare 2013; Ragone et al. 2018; Dematteis et al. 2018; Plotkin et al. 2019; Webber et al. 2019). The committor, and other quantities of interest, are fundamentally averages over sample trajectories. We can also express them as solutions to a concrete set of partial differential equations (PDEs) using basic stochastic calculus. TPT provides a framework to exploit these quantities and enhance our understanding of rare events, both from simulation data and from the fundamental equations of motion.

TPT has been applied primarily to molecular dynamics simulation to determine reaction rates and pathways of complex conformational transitions (E and Vanden-Eijnden 2010; Metzner et al. 2006; E and Vanden-Eijnden 2006); however, the framework does not depend on the details of the underlying dynamical system. TPT deals particularly with stochastically forced systems, such as Brownian dynamics of particles. Stochastic forcing applies quite generally; while the climate system is deterministic in principle, nonlinear interactions between resolved and unresolved scales inevitably leads to resolution-dependent model errors that can be approximated as stochastic. Hasselmann (1976)

originally formulated stochastic climate models to capture the influence of quickly evolving "weather" variables on the slowly evolving "climate" variables. Stochastic parameterization remains an active area of research. For example, Franzke and Majda (2006) had success in capturing energy fluxes of a three-layer quasigeostrophic model by projecting onto ten EOF modes and treating the remainder as stochastic forcing. Kitsios and Frederiksen (2019) addressed the challenge of designing consistent numerical schemes for subgrid-scale parameterization. Deep convection in the atmosphere and turbulence in the ocean boundary layer are two examples of multiscale processes that are especially challenging to resolve.

The aim of this article is to introduce the key quantities and relations describing the path properties of rare atmospheric events. Computing those quantities for more complicated systems than the low-order model studied here is a significant and, we argue, worthwhile challenge. We do not address that computational challenge here. Instead we note that development of approximation techniques for TPT and related quantities in high dimensional settings is an active area of research Thiede et al. (2019), Bowman et al. (2009), and Chodera and Noe (2014). These methods incorporate both simulated and observational data and we outline them briefly in the conclusions.

The paper is organized as follows. Section 2 describes the dynamical model we use, building on work by Ruzmaikin et al. (2003) and Birner and Williams (2008). Section 3 describes the mathematical framework of TPT, with detailed, but informal, derivations mainly put in the online supplement. Section 4 explains the methodology and results particular to this model.

2. Dynamical model

Holton and Mass (1976) studied "stratospheric vacillation cycles," a certain kind of minor warming in which zonal wind oscillates on a roughly seasonal time scale. They posited a mechanism of wave-mean flow interaction, which continues to be an important modeling paradigm. The quasigeostrophic equations are confined to a β-plane channel from 60° to the North Pole, and the streamfunction is perturbed from below by orographically induced planetary waves, specified through the lower boundary condition. The Holton–Mass model combines the zonal-mean flow equations,

$$\frac{\partial \overline{u}}{\partial t} - f_0 \overline{v} = 0, \tag{2}$$

$$f_0 \overline{u} = -\frac{\partial \overline{\Phi}}{\partial y}, \tag{3}$$

$$\frac{\partial}{\partial t}\left(\frac{\partial \overline{\Phi}}{\partial z}\right) + N^2 \overline{w} = -\alpha\left(\frac{\partial \overline{\Phi}}{\partial z} - \frac{R\overline{T}^*}{H}\right) - \frac{\partial}{\partial y}(\overline{v'\Phi'_z}), \quad (4)$$

$$\frac{\partial \overline{v}}{\partial y} + \frac{1}{\rho_s}\frac{\partial}{\partial z}(\rho_s \overline{w}) = 0, \quad (5)$$

and the linearized quasigeostrophic potential vorticity equation,

$$\left(\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right)q' + \beta_e \frac{\partial \psi'}{\partial x} + \frac{f_0^2}{\rho_s}\frac{\partial}{\partial z}\left(\frac{\alpha \rho_s}{N^2}\frac{\partial \psi'}{\partial z}\right) = 0, \quad (6)$$

where

$$q' := \nabla^2 \psi' + \frac{1}{\rho_s}\frac{\partial}{\partial z}\left(\frac{f_0^2}{N^2}\rho_s \frac{\partial \psi'}{\partial z}\right), \quad (7)$$

and

$$\beta_e = \beta - \frac{\partial^2 \overline{u}}{\partial y^2} - \frac{1}{\rho_s}\frac{\partial}{\partial z}\left(\rho_s \frac{f_0^2}{N^2}\frac{\partial \overline{u}}{\partial z}\right). \quad (8)$$

These equations represent zonal [(2)] and meridional [(3)] momentum balance, conservation of energy [(4)], conservation of mass [(5)], and a combination of all these for the perturbations [(6)]. Overbars and primes represent zonal averages and perturbations. Φ is the geopotential height; $H = 7\,\mathrm{km}$ is the atmospheric scale height; $z = -H\ln(p/p_0)$ is log pressure; $\rho_s = \rho_0 e^{-z/H}$ is a standard density profile; $\alpha = \alpha(z)$ is an altitude-dependent damping coefficient; and $\overline{T}^* = \overline{T}^*(y, z)$ is the radiative equilibrium temperature. Holton and Mass (1976) studied the solution consisting of single Fourier modes in the zonal and meridional directions:

$$\overline{u}(y, z, t) = U(z, t)\sin\ell y, \quad (9)$$

$$\psi'(x, y, t) = \mathrm{Re}\{\Psi(z, t)e^{ikx}\}e^{z/2H}\sin\ell y, \quad (10)$$

where ψ' is the zonal perturbation of $\psi = (g/f_0)\Phi$. $k = 2/(a\cos 60°)$ and $\ell = 3/a$ are the zonal and meridional wavenumbers, where a is Earth's radius. These wavenumbers are commonly observed in real SSWs and used in theoretical studies (Birner and Williams 2008; Ruzmaikin et al. 2003; Yoden 1987; Holton and Mass 1976); a split-type SSW is an extreme wavenumber-two perturbation. The lower boundary condition at $z = z_B$ (the tropopause) is

$$\Psi(z_B, t) = \frac{g}{f_0}h(t), \quad (11)$$

where h is a topographically induced perturbation to geopotential height at the tropopause. Holton and Mass

found that for a certain range of h, this system has qualitatively different regimes: a steady eastward zonal flow close to radiative equilibrium, and a weaker zonal flow with quasi-periodic "vacillations" from eastward to westward, even under constant forcing. Each vacillation cycle consists of a sudden warming and cooling over the time scale of weeks. Although these individual cycles are interesting weather events unto themselves, in this paper we think of the vacillations as occurring within a general *climate regime* that is conducive to sudden warming, as opposed to the steady flow state, which is not. Transitions between these two regimes, which we focus on here, are more accurately described as climatological shifts than weather events. The study by Ruzmaikin et al. (2003) varies h on an interannual time scale, with each single winter season occupying one of the two stable states and generating its daily weather accordingly. Hence, for this paper we will use the term "climate transitions."

The original Holton–Mass model discretizes the above PDE with finite differences across 27 vertical levels, which is assumed to be close to a continuum limit. Following several studies at this resolution (Holton and Mass 1976; Yoden 1987; Christiansen 2000), Ruzmaikin et al. (2003) did the most severe truncation possible, resolving only three vertical levels (including fixed boundaries) for easy analysis and exploration of parameter space. This reduces phase space to only three degrees of freedom: $U(t)$, which modulates $\overline{u}$ as a sine jet; $X(t) = \mathrm{Re}\{\Psi(t)\}$; and $Y(t) = \mathrm{Im}\{\Psi(t)\}$, where X and Y modulate the amplitude and phase of the perturbation streamfunction:

$$\psi'(x, y, t) = (X\cos kx - Y\sin kx)e^{z/2H}\sin\ell y. \quad (12)$$

Carrying the ansatz through the quasigeostrophic equations, Ruzmaikin et al. (2003) derived the following system:

$$\dot{X} = -\frac{1}{\tau_1}X - rY + sUY - \xi h + \delta_w \dot{h}, \quad (13)$$

$$\dot{Y} = -\frac{1}{\tau_1}Y + rX - sUX + \zeta hU, \quad (14)$$

$$\dot{U} = -\frac{1}{\tau_2}(U - U_R) - \eta hY - \delta_\Lambda \dot{\Lambda}. \quad (15)$$

The primary control parameter h represents topographic forcing and other sources of planetary waves, such as land–sea ice contrast. While Ruzmaikin et al. (2003) also varies Λ, representing vertical wind shear, we will only vary h and set Λ constant. Time derivatives $\dot{h}$ and $\dot{\Lambda}$ are zeroed, removing transient forcings such as seasonality
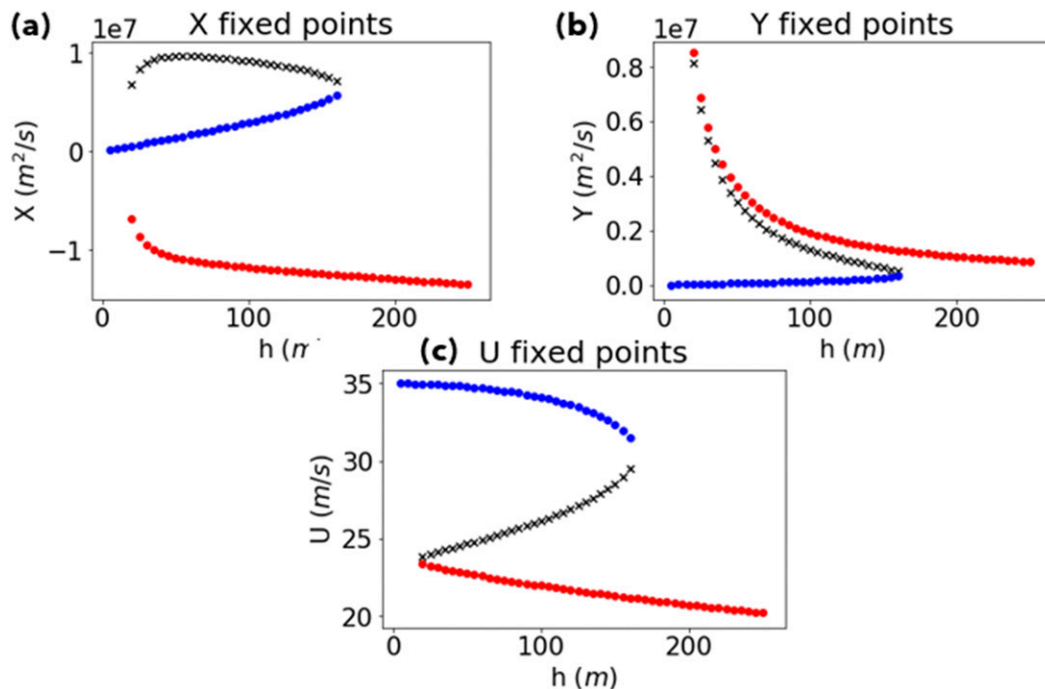
FIG. 2. Fixed points of (13)–(15) in the state space (X, Y, U), where X and Y represent the real and imaginary parts of the streamfunction and U the mean zonal wind amplitude. Fixed points vary as a function of the topographic forcing parameter h. (a)–(c) Fixed points of X, Y, and U, respectively, on the vertical axis, while h varies across the horizontal axis. Circles and crosses denote linearly stable and unstable fixed points, respectively. The range of h between $\sim$20 and $\sim$160 m supports three fixed points, two stable and one unstable. In this range, the blue points correspond to the radiative solution, while the red points represent the vacillating regime. This corresponds to a winter climatology that is conducive to sudden stratospheric warming events.

effects. The appendix and Ruzmaikin et al. (2003) have a detailed list of parameters.

Remarkably, this hugely simplified model retains the qualitative structure of the Holton–Mass model as a bistable system for a certain range of h between the critical values $h_1 \approx 20$ m and $h_2 \approx 160$ m, as shown in the bifurcation diagram of Fig. 2. Blue points represent the normal state of the vortex, in approximate thermal wind balance with the radiative equilibrium temperature field (henceforth called the "radiative solution"). Red points represent a disturbed vortex, with weaker zonal wind and vacillations. This climatological regime supports more SSW events, and is henceforth called the "vacillating solution." We use the same blue-red color scheme consistently here to represent these two states. Transitions between them happen on interannual time scales, affecting each year's likelihood of SSW events. The structure of transitions is illustrated in Fig. 3: as h increases slowly past the bifurcation threshold h_2, the system enters a series of rapid, large-amplitude oscillations that spiral into the weaker-circulation state.

In Figs. 2 and 3, transitions require crossing the bifurcation threshold h_2, where the radiative solution ceases to exist. Birner and Williams (2008) introduced additive white-noise forcing in the U variable to model unresolved gravity waves and found that these perturbations were sufficient to excite the system out of its normal state and into a vacillating regime. In Fig. 4 we illustrate stochastic trajectories of the system for three different (fixed) values of h. (For numerical reasons we also add a small amount of independent white noise to X and Y variables). Even when h is far below h_2, transitions still occur, and in fact the preference for the vacillating solution branch increases quickly with h.

Birner and Williams (2008) used direct numerical simulation and the Fokker–Planck equation to calculate long-term occupation statistics, that is, how much time on average was spent in each regime and the mean first passage time before a transition to the vacillating regime, all for a range of forcing and noise levels. Our approach differs in both target and methodology. We aim to characterize the transition process between the two states, to monitor its progress in real time, as well as to describe statistics of the transition over many realizations. Methodologically, transition path theory phrases these questions in terms of the *generator* of the stochastic process, a differential operator that encodes all information about the behavior of the process.
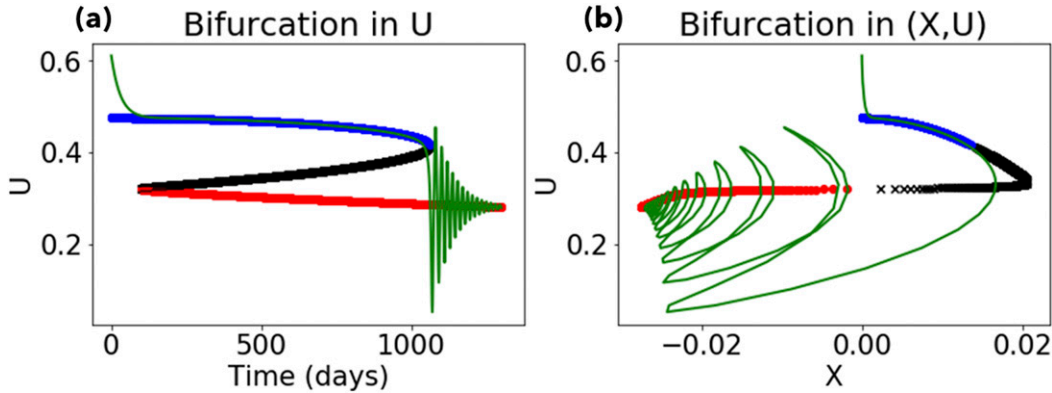
FIG. 3. Trajectories in (X, Y, U) space, where X and Y represent the real and imaginary parts of the streamfunction and U the mean zonal wind amplitude. In this simulation, the topographic forcing h increases linearly from 0 to 200 m in 1300 days. (a) The fixed points, with blue, red, and black for the radiative solution A, the vacillating solution B, and the unstable fixed point between them respectively. The trajectory of U over time is superimposed in gray. (b) As in (a), but plotted parametrically, in X–U space. Before the bifurcation, the trajectories follow the existing fixed point; after the bifurcation, they spiral into the new fixed point through a series of "vacillations."

3. Path properties

TPT characterizes the statistics of transitions between states. In this section, we introduce the key quantities needed for TPT as applied to the Ruzmaikin model to obtain a more complete picture than we get from the sample paths shown above. For mathematical details, see the online supplement and background literature (E and Vanden-Eijnden 2006; Metzner et al. 2006; E and Vanden-Eijnden 2010).

a. Infinitesimal generator

The noisy Ruzmaikin model can be expressed compactly as a stochastic differential equation (SDE)—specifically a *diffusion* process—in the variable $\mathbf{Z} = (X, Y, U) \in \mathbb{R}^3$ (or a more general phase space Ω) with a deterministic drift vector $\mathbf{b}(z) = [b_1(z), b_2(z), b_3(z)]$ and a 3×3 diffusion matrix $\boldsymbol{\sigma}(z)$:

$$d\mathbf{Z}_t = \mathbf{b}(\mathbf{Z}_t)\, dt + \boldsymbol{\sigma}(\mathbf{Z}_t)d\mathbf{W}_t. \qquad (16)$$

Here, $\mathbf{W}_t$ is a three vector of independent Brownian motions. We use the Ito convention for stochastic integration. While $\boldsymbol{\sigma}$ can in principle be any z-dependent matrix, we make $\boldsymbol{\sigma}$ diagonal and constant: $\boldsymbol{\sigma}(z) = \mathrm{diag}(\sigma_1, \sigma_2, \sigma_3)$, creating independent additive noise in the X, Y, and U variables; σ_1 and σ_2 have units of $\mathrm{m^2\,s^{-1}\,day^{-1/2}}$, while σ_3 has units of $\mathrm{m\,s^{-1}\,day^{-1/2}}$. Associated with this equation is the infinitesimal generator $\mathcal{L}$, an operator describing the evolution of observable functions forward in time following a trajectory. Explicitly, if $f(\cdot)$ is a smooth function of phase-space variables, then

$$\mathcal{L}f(z) := \frac{d}{dt}\mathbb{E}[f(\mathbf{Z}_t)|\mathbf{Z}_0 = z]|_{t=0}, \qquad (17)$$

where $\mathbb{E}$ is an expectation over sample paths.

Ito's lemma (the chain rule for diffusion SDEs) gives the Kolmogorov backward equation, which represents $\mathcal{L}$ as a partial differential operator (for details, see Pavliotis 2014):

$$\mathcal{L}f(z) = \sum_i b_i(z)\frac{\partial f(z)}{\partial z_i} + \frac{1}{2}\sum_{i,j}(\boldsymbol{\sigma}\boldsymbol{\sigma}^{\mathrm{T}})_{ij}\frac{\partial^2 f(z)}{\partial z_i \partial z_j}, \qquad (18)$$

$$= \mathbf{b}(z) \cdot \nabla f(z) + \mathrm{Tr}\left[\frac{1}{2}\boldsymbol{\sigma}\boldsymbol{\sigma}^{\mathrm{T}}\mathbf{Hf}(z)\right]. \qquad (19)$$

The diffusion matrix $(1/2)\sigma\sigma^T$ is also called $\mathbf{D}$ for convenience, which we will use interchangeably. $\mathbf{Hf}$ denotes the Hessian matrix: $[\mathbf{Hf}]_{ij} = \partial^2 f/\partial z_i \partial z_j$. The generator provides path statistics as the solution to PDEs, as illustrated in the following subsections.

b. Equilibrium probability density

This stochastic process admits a time-dependent probability density $\rho(z, t)$, which can be derived from the generator. For example, if the system starts in a known position $Z_0 = z$, then $\rho(z', 0) = \delta(z - z')$. The density spreads out from this initial point over time according to the Fokker–Planck equation, which can be written in terms of the adjoint of the generator:

$$\frac{\partial \rho(z,t)}{\partial t} = \mathcal{L}^*\rho(z,t), \qquad (20)$$

$$= \sum_i \frac{\partial}{\partial z_i}\left\{-b_i(z)\rho(z,t) + \sum_j \frac{\partial}{\partial z_j}[\rho(z,t)D_{ij}(z)]\right\}, \qquad (21)$$

$$= \nabla \cdot \{-\mathbf{b}(z)\rho(z,t) + \nabla \cdot [\rho(z,t)\mathbf{D}(z)]\}. \qquad (22)$$
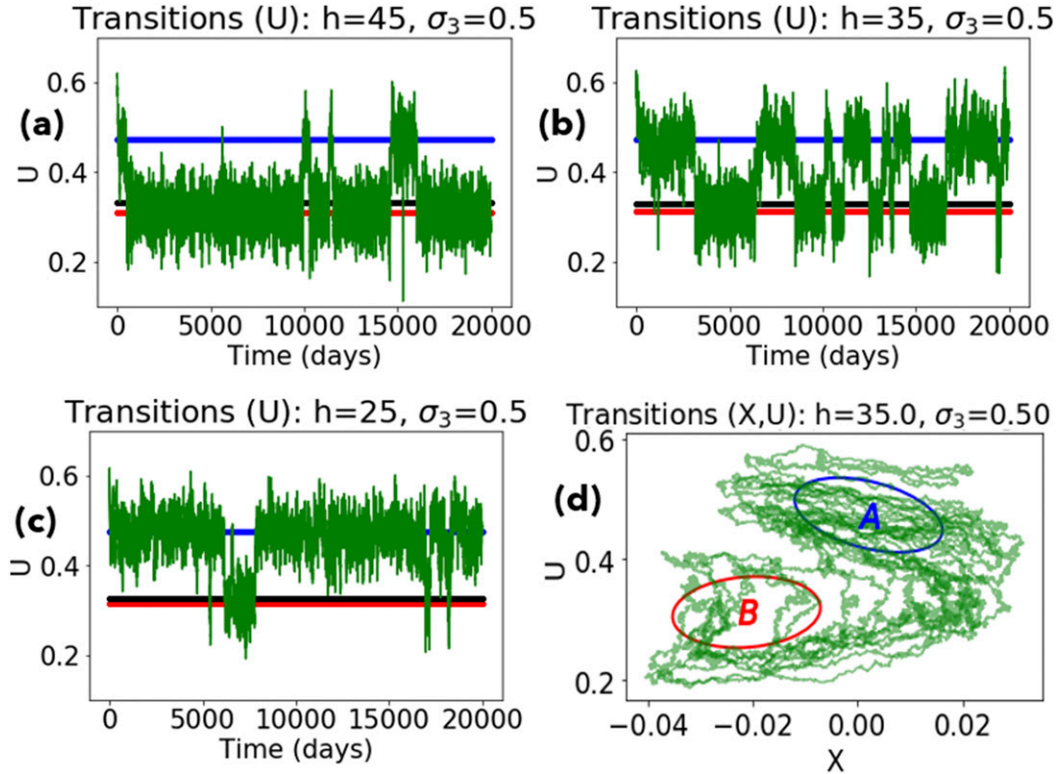
FIG. 4. Stochastic trajectories of the system for various fixed values of the parameters h (topographic forcing) and σ_3 (amplitude of stochastic forcing). (a)–(c) $U(t)$ for three different forcing levels: $h = 45$, 35, and 25 m, respectively, with $\sigma_3 = 0.5 \, \mathrm{m \, s^{-1} \, day^{-1/2}}$. In keeping with the bifurcation diagrams, the blue, black, and red lines mark the radiative, unstable, and vacillating solutions, respectively. Note that their relative positions vary slightly with h, as fixed points depend on parameters. As h increases from left to right, the system spends increasingly more of its time in the vacillating state. (d) A parametric plot of the transitions through (X, U) space, for $h = 35$ m [another view of (b)]. The transition happens seven times, and hence (d) shows seven different transition paths superimposed on each other. Most of the transitions follow a similar characteristic curve through X–U space, with a rapid decrease in U followed by a decrease in X.

When **D** is constant and diagonal, the last term simplifies to $\nabla \cdot [\nabla \cdot (\rho\mathbf{D})] = \sum_i D_{ii}\partial^2\rho/\partial z_i^2$. In the case of pure Brownian motion, $d\mathbf{Z}_t = d\mathbf{W}_t$, then $\mathbf{b} = 0$ and $\mathbf{D} = (1/2)I$, giving the heat equation $\partial_t\rho = (1/2)\nabla^2\rho \; (\nabla^2 = \sum_i \partial^2/\partial z_i^2)$. Assuming that the process is ergodic, the density eventually forgets the initial condition and stabilizes into a long-term (or equilibrium, or stationary) probability density $\pi(z)$, which solves $\mathcal{L}^*\pi = 0$. This can be approximated by either simulating the SDE for a very long time and binning data points, or directly solving the stationary PDE $\mathcal{L}^*\pi = 0$, subject to the normalization and positivity constraints $\int_\Omega \pi(z) \, dz = 1$ and $\pi(z) \geq 0$.

c. Committor probability

The stationary density is an equilibrium quantity characterizing the long-term occupation statistics. But it is insufficient to describe the events of interest to us, which are *transition paths*: trajectory segments beginning inside the radiative state and ending inside the vacillating state. Specifically, we define the

sets A and B as ellipsoids around these two fixed points, respectively. Their size is determined by contours of a local approximation to the stationary density π; see supplement for details. We say that a snapshot $\mathbf{Z}_t$ of the system is undergoing a *transition* (or reaction) at time t if it is on the way from set A to set B. This involves information about both its future and its past, for which we introduce the forward and backward committor probabilities in this section.

The forward committor q^+ (denoted q when context is clear) describes the progress of a stochastic trajectory traveling from set A to set B, as follows:

$$q^+(z) = \mathbb{P}\{\mathbf{Z}_t \text{ next hits } B \text{ before } A | \mathbf{Z}_0 = z\}, \quad (23)$$

$$q^+(z \in A) = 0, \quad q^+(z \in B) = 1. \quad (24)$$

The boundary conditions on A and B follow naturally from the probabilistic definition. If the system begins in set A, by path continuity it will certainly next find itself

in A, with zero chance of hitting B first. Starting in set B the opposite is true. The committor therefore obeys the boundary value problem (see online supplement for derivation)

$$\begin{cases} \mathcal{L}q^+(z) = 0 & z \in (A \cup B)^c \\ q^+(z) = 0 & z \in A \\ q^+(z) = 1 & z \in B \end{cases}. \qquad (25)$$

This equivalence of a conditional expectation with respect to a Markov process like the committor and the solution to PDE involving the generator of the process is generally referred to as a Feynman–Kac relation (Karatzas and Shreve 1998) and is well studied. The PDE in (25) is most naturally posed on an infinite domain, but as a numerical approximation we solve it in a large rectangular domain and impose homogeneous Neumann conditions at the domain boundary. A limiting example is the noise-dominated case, where $\mathbf{b}(z)$ is negligible and $\mathbf{D} = \mathbf{I}$. The Kolmogorov backward equation then becomes Laplace's equation:

$$\mathcal{L}q^+(z) = \nabla^2 q^+(z) = 0. \qquad (26)$$

If posed on the interval $[0, 1]$, with $A = \{0\}$ and $B = \{1\}$, the solution is $q^+(z) = z$. The linear increase from set A to set B reflects the greater likelihood of entering B when beginning closer to it. This limit is reflected in Fig. 1, which shows the committor of the double-well potential approaching a straight line for large σ values.

Prediction is naturally much harder in high-dimensional systems such as stratospheric models. A number of physically interpretable fields, such as zonal wind and geopotential height anomalies, seem to have some predictive power for SSW, but prediction by any single such diagnostic is suboptimal. Insofar as they are successful, these variables approximate certain aspects of the committor. For example, the committor might increase monotonically with the quasi-biennial oscillation index. Furthermore, statistical correlations potentially obscure the conditional relationships needed. For example, Martius et al. (2009) and Bao et al. (2017) examined tropospheric precursors to SSW events in reanalysis records, finding that blocking events preceded most major SSWs, potentially by enhancing upward-propagating planetary waves. (We use "precursor" only to mean an event that sometimes happens before SSW.) Blocking influences SSW through height perturbations at the tropopause, which would enter the Ruzmaikin model as low-frequency variations in lower boundary forcing h. Since we fix h constant, the blocking precursor is outside our scope here. However, farther down the dynamic chain are other measurable precursors such vertical wave activity flux and meridional heat flux, which are also found to have predictive power (Sjoberg and Birner 2012). However comprehensive the model, we would naturally expect the true committor probability to exhibit similar patterns to canonical precursors of that model such as blocking (for a troposphere-coupled model) and heat flux (for a stratosphere-only model). However, there is an important difference: while a precursor P may appear with high probability *given* that a SSW is imminent, the committor specifies the probability of a SSW given an observed pattern. As acknowledged in Martius et al. (2009), many blocking events did not lead to SSW events, meaning that $\mathbb{P}\{\text{SSW}|\text{blocking}\} \neq \mathbb{P}\{\text{blocking}|\text{SSW}\}$. Such distinctions highlight the need for a precise mathematical formulation that provides and distinguishes both kinds of information.

While q^+ describes the future of a transition, the backward committor q^- describes its past. It is defined as

$$q^-(z) \coloneqq \mathbb{P}\{\mathbf{Z}_t \text{ last visited } A \text{ rather than } B|\mathbf{Z}_0 = z\}, \qquad (27)$$

$$q^-(A) = 1, \quad q^-(B) = 0. \qquad (28)$$

Here, q^- solves the time-reversed Kolmogorov backward equation $\tilde{\mathcal{L}}q^- = 0$, where $\tilde{\mathcal{L}}$ is the time-reversed generator, which evolves observables backward in time. See supplement for a detailed description of $\tilde{\mathcal{L}}$ and its relationship the forward generator $\mathcal{L}$.

We now describe the fundamental statistics characterizing transition events as identified by TPT and explain how they can be expressed in terms of quantities such as q^+, q^-, and π. The probability density of reactive trajectories $\rho_R(z)$, the probability of observing the system $\mathbf{Z}_t$ at the location z during a transition, is proportional (up to a normalization constant) to the product $\pi(z)q^-(z)q^+(z)$. This density is large in regions of phase space that are highly trafficked by reactive trajectories. This is how TPT gives information about precursors, indicating regions of phase space that are usually visited by the system over the course of a transition path.

The direction and intensity of this traffic is specified by the *reactive current*. To develop this concept, we start by introducing the probability current $\mathbf{J}$, a vector field that satisfies a continuity equation with the time-dependent density ρ:

$$\frac{\partial \rho}{\partial t} = \mathcal{L}^*\rho = -\nabla \cdot \mathbf{J}.$$

If ρ were the density and υ the velocity field of a fluid, $\mathbf{J}$ would be $\rho\upsilon$. One can think of $\mathbf{J}$ as an instantaneous (in time and position) average over all possible system trajectories, though a precise mathematical description requires some care. In equilibrium, when $\rho = \pi$ is no

longer changing, $\nabla \cdot \mathbf{J} = 0$, or equivalently $\oint_C \mathbf{J} \cdot n \, dS = 0$ where C is any closed surface.

The reactive current $\mathbf{J}_{AB}$ is also an "average velocity," but restricted to reactive paths. Unlike $\mathbf{J}$, $\mathbf{J}_{AB}$ is not divergence-free, with a source in A and a sink in B (where transition paths start and end). We define $\mathbf{J}_{AB}$ implicitly via surface integrals. If C is any surface enclosing set A but not set B, with outward normal n, then the flux $\oint_C \mathbf{J}_{AB} \cdot n \, dS$ is the number of forward transitions per unit time, also called the *transition rate R_{AB}*; see E and Vanden-Eijnden (2010) for details. The supplement describes another expression in terms of the generator. The result is (Metzner et al. 2006)

$$\mathbf{J} = \pi \mathbf{b} - \nabla \cdot (\pi \mathbf{D}), \qquad (29)$$

$$\mathbf{J}_{AB} = q^+ q^- \mathbf{J} + \pi \mathbf{D}(q^- \nabla q^+ - q^+ \nabla q^-), \qquad (30)$$

where again π is the stationary density. This expression has intuitive ingredients. Multiplying $\mathbf{J}$ by $q^+ q^-$ conditions the equilibrium probability current on the trajectory being reactive, meaning en route from A to B. The $q^- \nabla q^+ - q^+ \nabla q^-$ reflects the fact that trajectories from A to B must ascend a gradient of q^+, going from $q^+ = 0$ to $q^+ = 1$, while descending a gradient of q^-.

Just as $\mathbf{J}_{AB}(z)$ describes the average reactive velocity, a streamline z_t of $\mathbf{J}_{AB}(z)$ (solving $\dot{z}_t = \mathbf{J}_{AB}(z_t)$, with $z_0 = \mathbf{a} \in A$ and $z_T = \mathbf{b} \in B$ for some $T > 0$) is a kind of "average" transition path. Although the streamline will not be realized by any particular transition path, it will have common geometric features in phase space with many actual path samples. At low noise the reactive trajectories will cluster in a thin corridor about the streamline. The streamline is a more dynamical description of precursors: whereas regions of high reactive density are commonly observed states along reactive trajectories, streamlines of reactive current are commonly observed *sequences* of states along reactive trajectories. The study by Limpasuvan et al. (2004), for example, described a sequence of events in a prototypical SSW life cycle based on reanalysis involving vortex preconditioning, wave forcing, and anomalous heat fluxes at various levels in the troposphere and stratosphere. The sequence described there likely corresponds to a streamline of the reactive trajectory.

The committor also quantifies the relative balance of time spent on the way to each set. If more probability mass lies in the region where $q^+ > 1/2$, set B is globally more imminent, whereas more mass where $q^+ < 1/2$ indicates set A is. A single summary statistic of imminence is the average committor during a long trajectory, $\mathbb{E}[q^+(\mathbf{Z}_t)]$, computed as a weighted average against the equilibrium density:

$$\mathbb{E}[q^+(\mathbf{Z}_t)] = \mathbb{E}_\pi[q^+(\mathbf{Z})] = \int_\Omega q^+(z) \pi(z) \, dz =: \langle q^+ \rangle_\pi. \qquad (31)$$

An average below (above) 1/2 *would indicate more time spent on the way* to A (B).

Another statistic, the forward transition rate, captures the frequency of transitions between A and B rather than the overall time spent in each. We earlier defined R_{AB} as the number of $A \rightarrow B$ transitions per unit time. Since a $B \rightarrow A$ transition must occur between every two $A \rightarrow B$ transitions, $R_{AB} = R_{BA} =: R$. The inverse of the transition rate is the return time, a widely used metric for changing frequency of extreme events under climate change scenarios (Easterling et al. 2000). However, the forward and backward transitions may differ in important characteristics like speed. To capture this asymmetry, we need a *dynamical* analog to the equilibrium statistic $\langle q^+ \rangle_\pi$. The typical quantity of choice is the rate *constant k_{AB}*, which is larger if $A \rightarrow B$ transitions happen faster than $B \rightarrow A$ transitions. We therefore normalize by the overall time spent having come from A, which is $\langle q^- \rangle_\pi$:

$$k_{AB} = \frac{R}{\langle q^- \rangle_\pi} = \frac{R}{\int q^-(z) \pi(z) \, dz}. \qquad (32)$$

This rate constant, defined in Vanden-Eijnden (2014), parallels the chemistry definition. If X_A and X_B are two chemical species, with $[\cdot]$ denoting concentration, the forward and backward-rate constants k_{AB} and k_{BA} are defined so that

$$R = [X_A] k_{AB} = [X_B] k_{BA}. \qquad (33)$$

In the language of transition path theory, $[X_A]$ is the long-term probability of the system existing most recently in state A, which is $\langle q^- \rangle_\pi$. Rates are also expressible in terms of expected passage times. Thinking of $[X_A]$ as the total probability of having last visited set A, $1/k_{AB} = [X_A]/R$ estimates the total transition time between entering A (having last visited B) and next reentering B. It is these inverse quantities we display in the results section.

These quantities together make an informative description of the typical transition process from A to B. We now proceed to analyze the transition path properties of the Ruzmaikin stratospheric model.

4. Methodology

Spatial discretization

The quantities of interest described above (π, q^+, q^-, and $\mathbf{J}_{AB}$) emerge as solutions to PDEs involving the

generator $\mathcal{L}$, which must be approximated by spatial discretization. In the supplement we describe a finite volume scheme to directly discretize the adjoint $\mathcal{L}^*$ as a matrix, which we name $\mathbf{L}^*$, on a regular grid in d dimensions. Here we use the same domain and noise levels as Birner and Williams (2008): $-0.06 \le X \le 0.04$, $-0.05 \le Y \le 0.05$, $0 \le U \le 0.8$ in units nondimensionalized in terms of the radius of Earth and the length of a day. We tile this with a grid of $40 \times 40 \times 80$ grid cells. We choose a noise constant σ_3 in the U variable in the range 0.4–1.5. This is a similar range to observed atmospheric gravity wave momentum forcing (Birner and Williams 2008). For numerical reasons, we also add small noise to the streamfunction variables X and Y, in proportion to the domain size. Specifically, as U spans a range of 0.8 and X, Y span a smaller range of 0.1, we choose σ_1 and σ_2 to be $\sigma_3 \times (0.1/0.8)$. This adjustment does change our results with respect to Birner and Williams (2008), causing more transitions in both directions at lower h than if only the U variable were perturbed. While gravity wave drag forces the zonal wind, eddy interactions and other sources of internal variability can perturb the streamfunction as well, and it is not uncommon to represent these effects stochastically (DelSole and Farrell 1995). There are surely more accurate representations of noise, but this important issue is not our focus. We retain these perturbations for numerical convenience, but stress that the general principles of the TPT framework are independent of any specific form of stochasticity. In the forthcoming experiments, we will refer only to σ_3 with the understanding that σ_1 and σ_2 are adjusted proportionally. The discretization we use has strengths and limitations. Given the matrix $\mathbf{L}^*$ on this grid, the discretized generator $\mathbf{L}$ is just the transpose. To ensure certain properties of solutions, such as positivity of probabilities, $\mathbf{L}$ should ideally retain characteristics of the infinitesimal generator of a discrete-space, continuous-time Markov process: rows that sum to zero, and nonnegative off-diagonal entries. Such a discretization is called "realizable" (Bou-Rabee and Vanden-Eijnden 2015). One can check that our discretization always satisfies the former property, and so is realizable provided small enough grid spacing (δX, δY, δU). In our current example, the spacing is not nearly small enough to guarantee this (matrix entries were just as often negative as positive), but results are still accurate, as verified by stochastic simulations to be described in the results section. While we could have used one-sided finite differences to enforce positivity, this would have degraded the overall numerical accuracy of the solutions. We opted instead to zero out negatives, which were always negligible in magnitude.

The discretized Kolmogorov backward equation is $\mathbf{L}q^+ = 0$, augmented with appropriate boundary conditions. The definition of A and B is a design choice that should satisfy three conditions: 1) they are disjoint, 2) A contains the radiative fixed point and B the fixed point of the vacillating regime, and 3) both sets are relatively stable in the chosen noise range. We choose A and B to be ellipses with orientations determined by the covariance of the equilibrium density of the linearized stochastic dynamics about their respective fixed points, as described in the online supplement. The choice of the sizes of A and B is a subjective decision that alters the very definition of a reactive trajectory; hence, different sizes emphasize different features of the transition path ensemble, especially in oscillatory systems like this one. We made A and B large enough to enclose the many loops that often accompany the escape from A and the descent into B, so that we can focus on the relatively rare crossing of phase space. More sophisticated techniques exist for shrinking the two sets while erasing resulting loops (Lu and Vanden-Eijnden 2014; Banisch and Vanden-Eijnden 2016); for simplicity, we forgo these techniques for the current study.

Careful discretization is important for constructing the dominant pathways discussed above, that is, the streamlines z_t satisfying $\dot{z}_t = \mathbf{J}_{AB}$. Standard integration techniques such as Euler or Runge–Kutta will accumulate errors, not only from Taylor expansion but also from the discretized solution of q^+, q^-, and π. These can be severe enough to prevent z_t from reaching set B. To guarantee that full transitions are extracted, we instead solve shortest-path algorithms on the graph induced by the discretization, as described in Metzner et al. (2009). The supplement contains more details on this computation.

5. Results

We begin this section by describing the kinematic path characteristics of the process in its three-dimensional phase space, according to the quantities described above. Following this purely geometrical description, we will suggest some dynamical interpretations and compare with previous studies. Finally, we will map statistical features as functions of background parameters.

The Ruzmaikin model is attractive for demonstrating use of the tools introduced in section 3 due to its low-dimensional state space, in which PDEs can be solved numerically using standard methods such as our finite volume scheme. We tested the committor's accuracy empirically by randomly selecting 50 cells in our grid (this is 0.04% of the grid) and evolving $n = 60$ stochastic trajectories forward in time from each, stopping when
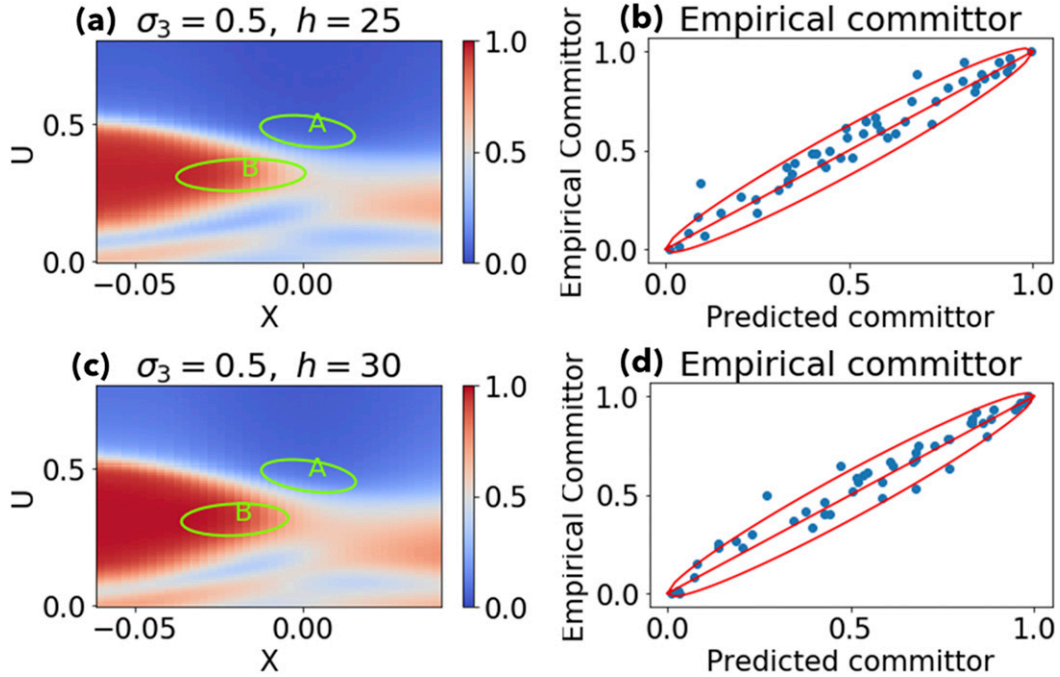
FIG. 5. The committor gives the probabilistic forecast of the system, and this plot is an empirical demonstration of its predictive capacity. Here, noise is fixed at $\sigma_3 = 0.5$ m s^{-1} day$^{-1/2}$, while topographic forcing (a),(b) $h = 25$ m and (c),(d)$h = 30$ m. (left) The forward committor q^+ solved by the finite-volume method, averaged in the Y direction. The ellipses labeled A and B are projections of the actual sets onto the X–U plane, where X and U are the real part of the streamfunction and the mean zonal wind amplitude. Committor values range from 0 (blue) to 1 (red), with the white contour showing the surface $q^+ = 1/2$. (right) Comparison of the PDE solution of the committor with a Monte Carlo solution from running many trajectories. For 50 randomly chosen grid points [sampled uniformly across the committor range (0, 1)], we launched 60 independent stochastic trajectories and counted the fraction that reached set B first. We call this the empirical committor, $\tilde{q}^+$. (b),(d) Plots of $\tilde{q}^+$ vs q^+ for these 50 random grid cells. The middle red line is the curve $\tilde{q}^+ = q^+$, and the envelope around it is the 95% confidence interval for sampling errors, based on a Gaussian approximation to the binomial distribution.

they reach either set A or set B. The fraction of trajectories starting from z that first reach B is taken as the empirical committor at point z and is denoted $\tilde{q}^+$, which we compare with the predicted committor q^+ from the finite volume scheme. Figure 5 clearly demonstrates the usefulness of the committor for probabilistic forecasting. The left column displays the committor calculated from finite volumes, averaged in the Y direction, for two different forcing levels h. The right column shows a scatterplot of $\tilde{q}^+$ versus q^+ at the 50 randomly selected grid cells. We expect the points to fall along the line $\tilde{q}^+ = q^+$ with some spread proportional to the standard deviation of the binomial distribution, $\sqrt{q^+(1 - q^+)/n}$. Approximating this as a Gaussian, we have plotted red curves enclosing the 95% confidence interval, which indeed contains approximately 95% of the data points. Statistical sampling error can explain the observed level of deviation, although discretization error (from solving the PDE and from time-stepping) may also contribute.

Figure 5 also shows how q^+ responds to increasing h, even far below the bifurcation threshold: the committor values throughout state space become rapidly skewed toward unity (meaning redder in the picture). This means that even slight perturbations can kick the system out of state A toward state B. Another indicator is the "isocommittor surface," the set of points z such that $q^+(z) = 1/2$; that is, the system has equal probability of next entering set A or B. In the left-hand column this is the set of gray points (averaging out the variable Y). For low forcing, this surface tightly encloses set B, meaning the system must wander very close before a transition is imminent. For high forcing values, the isocommittor hugs set A more closely, meaning that small perturbations from this normal state can easily push the system into dangerous territory. In Fig. 6, the isocommittor is shown as a set of gray points in a 3D plot viewed from various vantage points. In the low-U region, the isocommittor resembles a spiral staircase, reflecting the spiral-shaped stable manifold of the fixed point in set
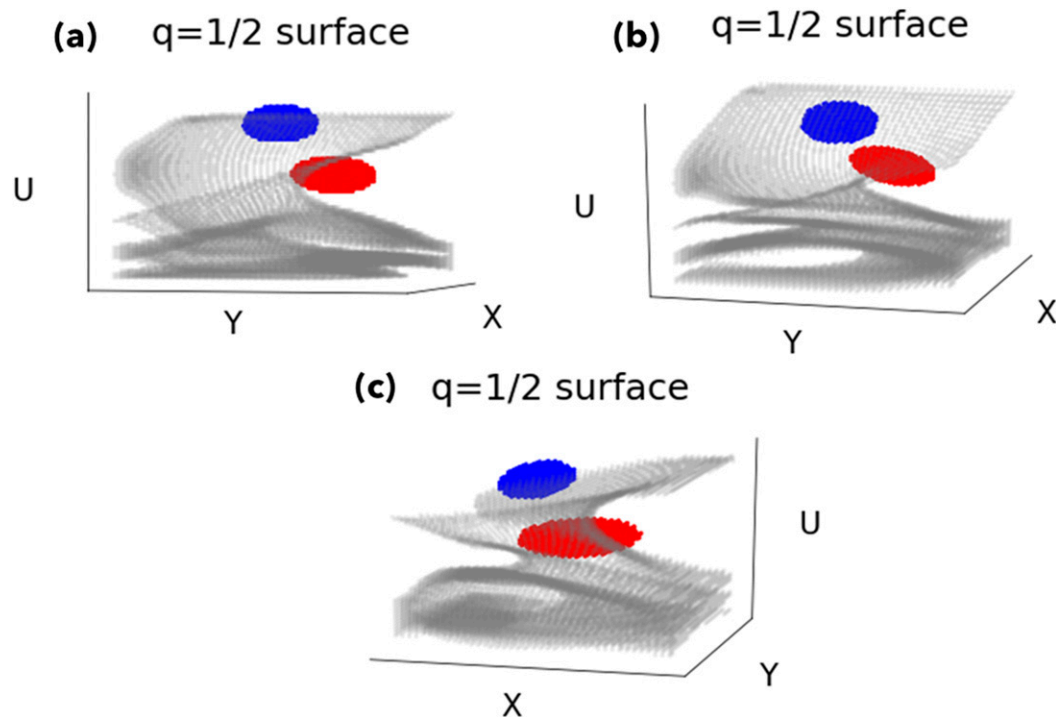
FIG. 6. The committor 1/2 surface is the set of all points in state space where the $q^+ = 1/2$, and sets A and B have equal probabilities of being visited next. Here the surface is rendered as a set of points and viewed from various vantage points in state space (the online supplement shows a video with rotation). The topographic forcing is fixed to $h = 25$ m and the noise level to $\sigma_3 = 0.5$ m s^{-1} day$^{-1/2}$. The blue and red clusters mark sets A and B respectively, centered around the two stable fixed points. The gray points show the location of the surface $q^+ = 1/2$. The most striking feature is the "spiral staircase" structure in the low-U region of phase space. For any given streamfunction phase, the likelihood of heading toward state A or B depends sensitively on U, in an oscillatory manner. Even at very low values of U, there are narrow channels that are likely to lead back to set A rather than set B. This accounts for the blue regions in the lower part of Fig. 5. These disappear, however, at higher noise, when set B overtakes the lower half of the picture.

B. Different initial positions with the same stream-function phase, differing only slightly in the U direction, can have drastically different final destinations. These spiral surfaces are responsible for the blue lobes in the lower part of Fig. 5, but they disappear at higher noise.

Figures 7 and 8 display numerical solutions of the equilibrium density π and reactive density $\rho_R \propto \pi q^+ q^-$ for two forcing levels. While π indicates where $\mathbf{Z}_t$ tends to reside, ρ_R indicates where $\mathbf{Z}_t$ resides *given* a transition from A to B is underway. As h increases, even far below the bifurcation threshold, π responds strongly, shifting weight toward state B. On the other hand, the reactive density displays similar characteristics for all h values. In the X–U plane, the two lobes of high reactive density surrounding A indicate that zonal wind tends to remain strong for a while before dipping into the weaker regime. Viewing the same field in the X–Y plane (Fig. 8) reveals a halo of intermediate density about set A. While many different motions would be consistent with this pattern, the coming figures verify that the early stages of transition have circular loops in the X–Y plane, meaning zonal movement of the streamfunction's peaks

and troughs. The exact streamfunction phase corresponding to the (X, Y) position is calculated as follows. Recall the streamfunction is $\psi' = \mathrm{Re}\{\Psi(t)e^{ikx}\cos(\ell y)\}$, where $\Psi = X + iY$. In polar coordinates, $\Psi = \sqrt{X^2 + Y^2}e^{i\phi}$, where $\phi = \tan^{-1}(Y/X)$. The full streamfunction is

$$\psi' \propto \mathrm{Re}\{\Psi e^{ikx}\} = \mathrm{Re}\{\sqrt{X^2 + Y^2}\, e^{i(\phi+kx)}\}$$

$$= \sqrt{X^2 + Y^2}\cos(\phi + 2\lambda),$$

where λ is longitude.

The angle from the origin in the X–Y plane indicates the zonal streamfunction phase, and circular motion indicates zonal movement. (This "looping" motion is indeed shared by the transition path samples shown in Figs. 9 and 10, to be described later.)

The darkest (most-trafficked) region of this loop is the sector $(\pi/4) \lesssim \phi \lesssim (\pi/4)$. The relationship between (X, Y) and λ indicates ψ' is maximized at longitudes $\lambda = \{-\phi/2, \pi - \phi/2\}$. As the maximum reactive density occurs around $\phi = 3\pi/8$, the streamfunction peaks are at
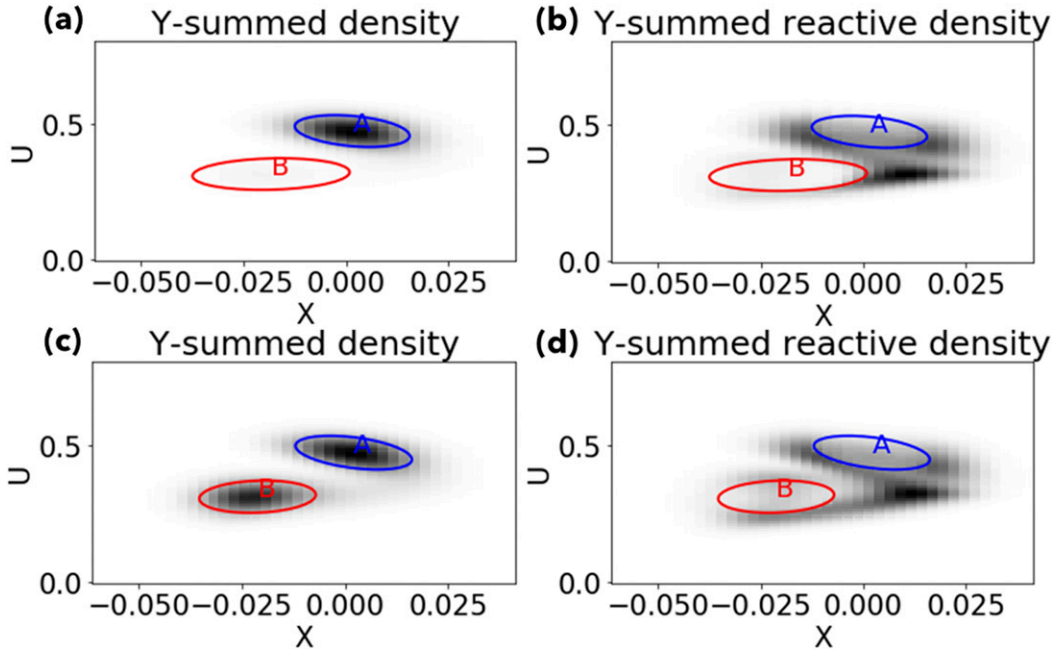
FIG. 7. (a),(c) Equilibrium densities $\pi(z)$ and (b),(d) reactive densities $\rho_R(z) = \pi(z)q^+(z)q^-(z)$ for two different forcing levels: (a),(b) $h = 25$ m and (c),(d) $h = 35$ m. $\pi(z)$ is the long-term probability density of finding the system at point z at any given time; $\rho_R(z)$ is the same probability, but *conditional* on also being reactive at that time, meaning having last visited A and next destined to visit B. Darker color indicates higher density. These densities are summed in the Y direction to give a marginal density as a function of X and U. The red and blue ellipses show the projected boundaries of sets A and B, respectively. These reactive densities capture the patterns of transition path samples shown in Fig. 4, but through continuous fields instead. At low forcing levels, most of the equilibrium density is concentrated around set A, whereas higher forcing shifts some of the mass to set B. Meanwhile, the characteristic curved shape of transition paths in phase space is borne out by the reactive densities, with a sickle-shaped high-density region bridging the gap between set A and B.

$\{-3\pi/16, 13\pi/16\} \approx \{326°, 146°\}$. What is the significance of this phase relative to the lower boundary forcing? Recalling the forcing form $\mathrm{Re}\{\Psi(z_B, t)e^{ikx}\} = h\mathrm{Re}\{e^{i2\lambda}\} \propto \cos(2\lambda)$, the bottom peaks are located at $\lambda = \{0, \pi\}$. Hence, the bulk of the transition process happens when the perturbation streamfunction at the midstratosphere lags the lower boundary condition by $3/16 \pm 1/16$ of a wavelength. Meanwhile, the X–U plane reveals what happens to the zonal wind speed during the SSW transition. The high-reactive density region discussed above coincides with the crescent-shaped bridge of high density between the sets in Fig. 7. This suggests that in an SSW, the zonal wind weakens while the streamfunction stays in that particular phase window.

The pictures of reactive density suggest that reactive trajectories tend to loop around set A, physically meaning the streamfunction tends to travel in one direction before slowing down, but they technically convey no *directional* information to explicitly support this claim. For this, we turn to the reactive current. We computed the discrete-space effective current matrix $\mathbf{f}^+_{ij}$, directly from the finite-volume discretization and numerical solutions of π, q^+ and q^-. Physically, this matrix

represents the flux of a vector field from grid cell i to cell j. From this we calculated the maximum-current paths as described in Metzner et al. (2009) and displayed the results in Fig. 9 for a forcing level of $h = 30$ m (other levels are qualitatively similar). Both the X–U and X–Y views are shown. Superimposed on these paths are seven actual reactive trajectories that occurred during a long stochastic simulation, to demonstrate features that are captured by the dominant pathways. The dominant path from A to B indeed contains a half loop in the X–Y plane in the clockwise direction, which means an eastward phase velocity. With a smaller set A, this dominant path would contain more of these loops. However, during the next transition stage, the streamfunction slows to a halt at the phase angle $\phi = \pi/2$, doubles back and travels westward as zonal wind loses strength. The smear of high density in the neighborhood $\phi \sim 3\pi/16$ therefore comprises not only a precipitous drop in zonal wind (which happens at the edge of that region) but also a backtrack, this time with weaker background zonal wind. This behavior is borne out by the trajectory samples, which vacillate in the upper-middle section of the X–Y plot. These paths are displayed as space–time
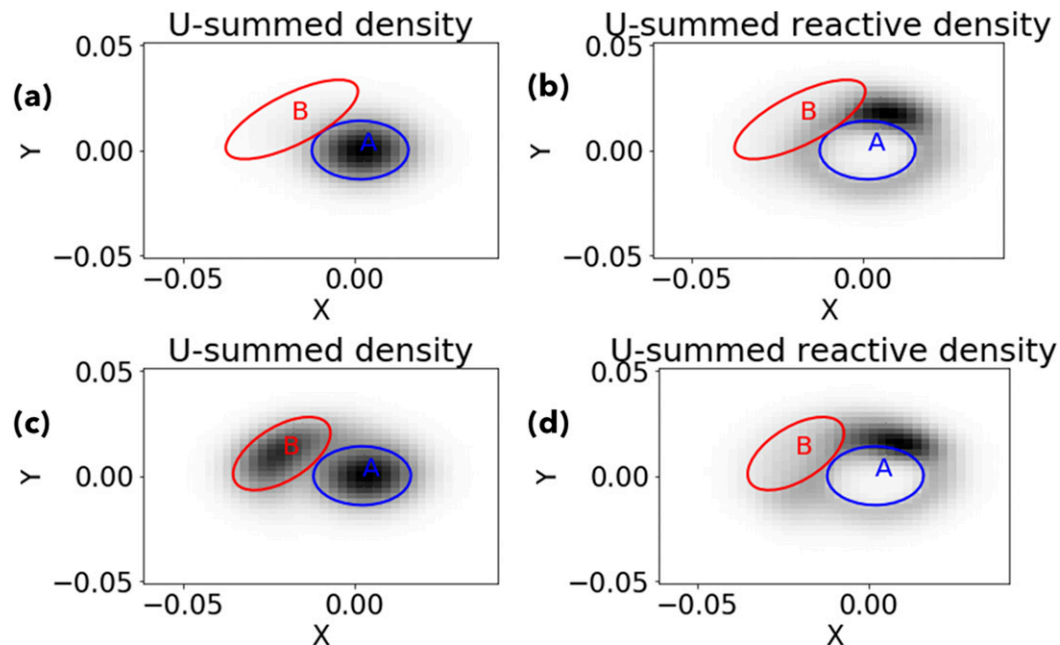
FIG. 8. As in Fig. 7, but projected onto the X–Y plane, which is the complex plane that characterizes the perturbation streamfunction. The center of set A, approximately at $X = Y = 0$, corresponds to a zonally symmetric streamfunction with no perturbation. Counterclockwise motion of trajectories around A represents an eastward phase velocity of the streamfunction, which is the dominant modality in the radiative regime. The region of high density above set A shows the phase in the streamfunction at which zonal wind is most likely to begin to weaken.

diagrams of the streamfunction in Fig. 10. In Fig. 10a, the dominant path's two loops correspond to two troughs moving east past a fixed longitude before the slowdown. The random streamfunction trajectories shown in Figs. 10b–g do not follow this representative history exactly, but they do combine elements of it: steady eastward wave propagation followed by slowdown and reversal. Each stage can have multiple false starts. Notably, the slowdown consistently happens at the same phase, with peaks at ~120° and 300°E, at roughly the same phase as found from the density plots. In fact, the figures show a brief slowdown every time the streamfunction passes this phase. This can be thought of as a representation of blocking events that often accompany sudden stratospheric warmings. The third transition path shown is an exception to the general pattern, making a final turn toward the east instead of to the west. This outlier of a reactive trajectory can also be seen in Fig. 9, as the single green trajectory that decreases in X before decreasing in U instead of the other way around.

This kinematic sequence of events has a dynamical interpretation with precedent in prior literature. A critical ingredient of SSW is meridional eddy heat flux, which in this model takes the form $\overline{v'\Phi'_z} \propto hY$ (see the online supplement for a detailed derivation). This term

shows up explicitly as a negative forcing on U in (15), showing that a reduced equator-to-pole temperature gradient in turn weakens the vortex via the thermal wind relation. The association of heat flux with SSW has been demonstrated in reanalysis (Sjoberg and Birner 2012) and in detailed numerical simulations of internal stratospheric dynamics, even with time-independent lower boundary forcing (Scott and Polvani 2006). This relationship favors the phase $\phi = \pi/2$ as the most susceptible state for SSW onset, which is exactly picked out by the dominant transition pathway in Fig. 9.

However, immediately after the wind starts weakening at $\phi = \pi/2$, where the streamfunction lines up with its lower boundary condition, the phase velocity reverses, giving rise to the westward lag of $3\pi/16$ we observed in the reactive density. A similar phase lag has also been observed in more detailed numerical studies. For example, Scott and Polvani (2006) observed a lag of $\pi/2$ across the whole stratosphere ($\pi/4$ at the midlevel), quite similar to our result. They found that vortex breakup was preceded by a long, slow build-up phase in which the vortex became increasingly vertically coherent, only to be ripped apart by an upward- and west-propagating wave. In an experiment with slowly increasing lower boundary forcing, Dunkerton et al. (1981) saw a phase lag across the whole stratosphere that increased from ~100° to ~180°
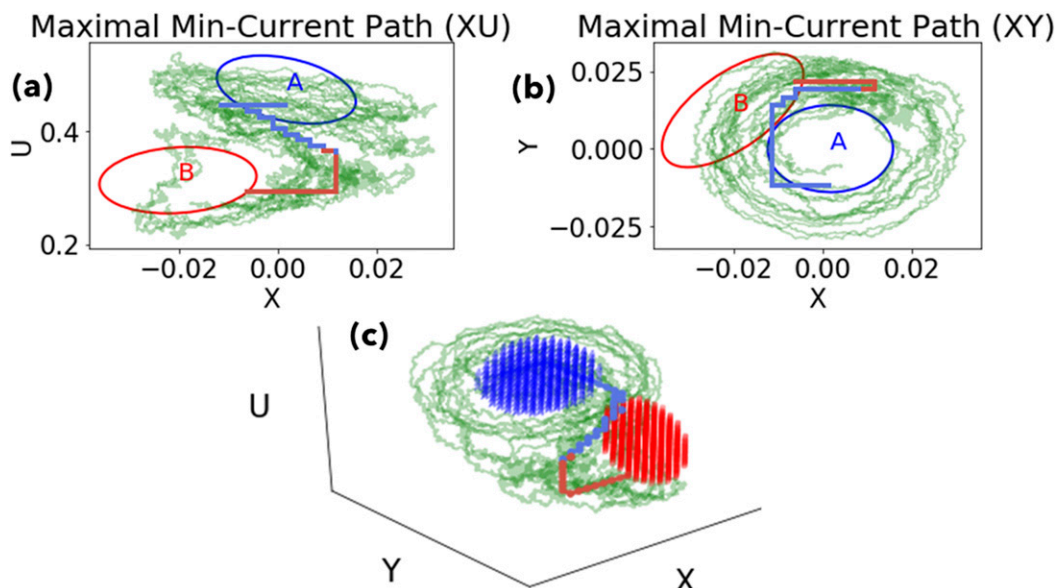
FIG. 9. Paths of maximal current superimposed on transition path samples. All four figures shown the same path, but from different vantage points. While the reactive probability density (Figs. 7 and 8) says where transition paths spend their time, the reactive current is a vector field of the transition paths' local average directionality. The path shown in a color gradient from blue to red is a streamline of this vector field, representing an "average" transition path. The path is colored blue where the local committor is less than 0.5, and red otherwise. Note that the path can cross back and forth. The transition from red to blue, where the path first crosses the threshold $q^+ = 1/2$ and enters the probabilistic B basin, is marked by a sudden drop in the U variable—a deceleration in zonal wind. At the same time, the path's rotations about A reverse direction, from clockwise to counterclockwise, corresponding to a reversal in phase velocity of the streamfunction. This path accurately captures geometric tendencies of actual transition paths; five random samples of reactive trajectories are superimposed in green, the bulk of which cluster around the maximum current path. (a),(b) Cross sections in X–U and X–Y space, respectively. (c) A three-dimensional view. The parameters are $h = 30$ m and $\sigma_3 = 0.5$ m s^{-1} day$^{-1/2}$. Figure 10 shows the corresponding space–time diagrams of the streamfunction.

(50° to 90° between the lower boundary and the mid-stratosphere) over the course of the warming event. They attribute this phase tilt to the zonal wind rapidly reversing and carrying the streamfunction along. The weakening zonal wind simply removes the Doppler shift from the Rossby wave dispersion relation, $\omega = Uk - \beta k/(k^2 + \ell^2)$, allowing the waves to revert to their preferred westward phase velocity. This balance is also clear in (13) and (14): ignoring the damping and forcing terms, the dynamics read $[\dot{X}, \dot{Y}] = [(sU - r)Y, -(sU - r)X]$. For weak U, rotation in the X–Y plane is counterclockwise and phase speed is westward.

Let us reemphasize the probabilistic interpretation of reactive density. We have found that transitions from A to B are accompanied by anomalous increases in meridional heat flux. In other words, $\mathbb{P}\{\text{High heat flux}|\text{SSW}\}$ is high. As noted earlier, this does not imply that $\mathbb{P}\{\text{SSW}|\text{high heat flux}\}$ is also high. The identified streamfunction phase is not a "danger zone" in the sense that a trajectory entering this region is at higher risk of falling into state B; the committor alone conveys that information. Rather, a trajectory is highly likely to pass through that

region *given* that it is reactive. Notably, reactive trajectories are unlikely to take a straight-line path from A to B with U, X, and Y changing linearly. This unrealistic path would represent a zonally stationary streamfunction growing steadily in magnitude, while zonal wind falls off gradually. At higher noise levels, however, the system would be increasingly dominated by pure Brownian motion, and such a path would become more plausible.

We now turn to a quantitative comparison of committors and transition rates for different forcing and noise levels. These trends illustrate the effects of modeling choices and global change on the climatology of SSW. Planetary wave forcing h varies across days and seasons as well as different planets. The strength of additive noise σ_3 (which determines the full diffusion matrix $\boldsymbol{\sigma}$ by proportionality) is a modeling choice intended to represent gravity wave drag. Different stochastic parameterizations will vary in their effective σ_3 value, and it is important to understand the sensitivity of SSW to model choices (Sigmond and Scinocca 2010). Furthermore, long-term climate change may cause both parameters to drift, altering the occurrence of SSW-induced severe weather events.
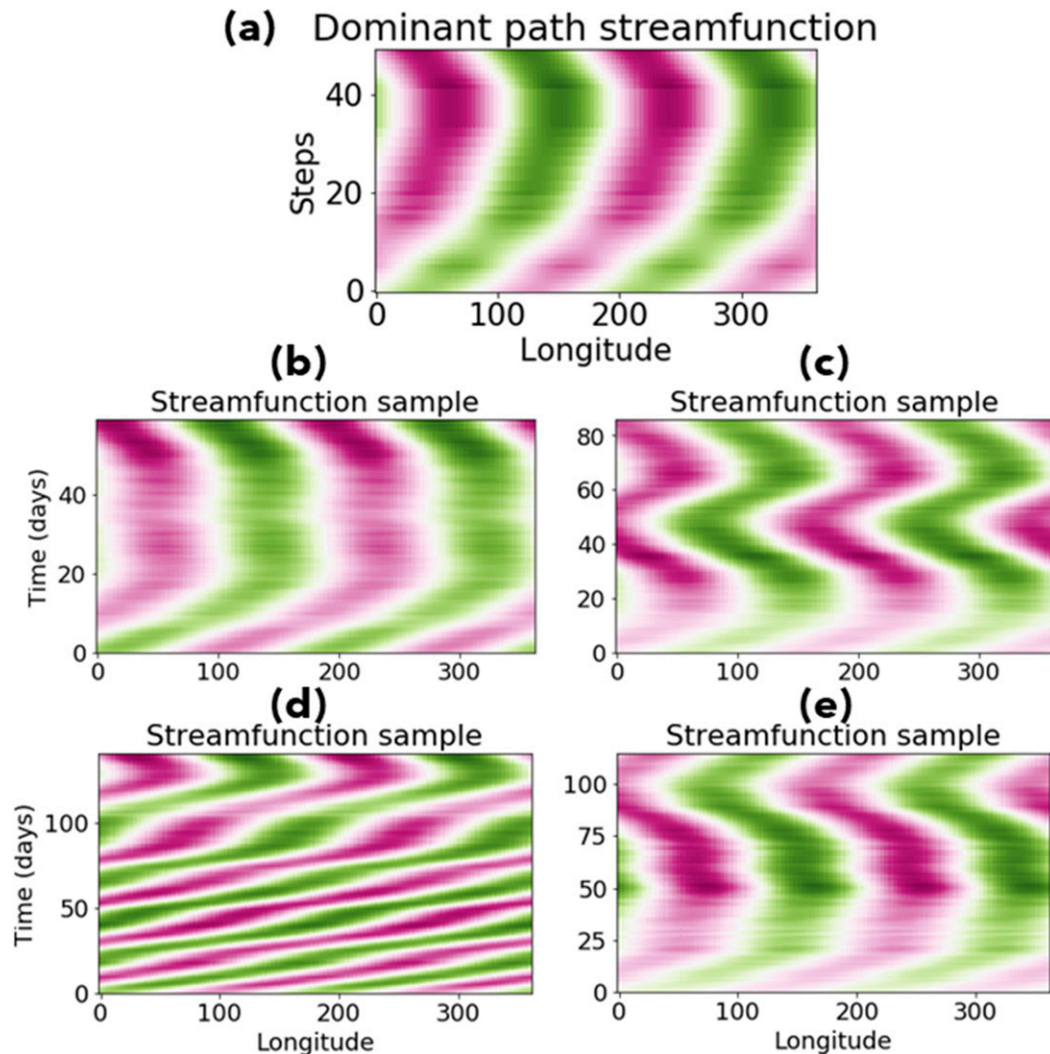
FIG. 10. Streamfunctions over time corresponding to the trajectories shown in Fig. 9. As $\psi' \propto (X \cos kx - Y \sin kx)$, the X and Y variables represent the phase of the streamfunction, whose movement we plot over time as a space–time diagram. (a) The dominant transition path. The phase velocity is initially eastward, matching with the clockwise rotations in the X–Y plane, as shown in Fig. 9. The waves then slow down and reverse direction, matching with the counterclockwise turn and zonal wind drop in Fig. 9. The vertical axis plays the role of time, but the dominant path technically conveys only geometrical information. Hence, we measure it in discrete steps. (b)–(d) The streamfunctions over time corresponding to four of the green transition path samples in Fig. 9, chosen randomly. Most exhibit the same slowdown and reversal behavior exemplified by the dominant path. The exception is (e), which turns to the east at the end of its path. This corresponds to the stray green trajectory visible in Fig. 9a, which enters set B from above and in the clockwise direction. Sample (d) undergoes some winding before the slowdown but does slow down every time it reaches the same phase.

The measure of the relative "imminence" of a vacillating solution versus a radiative solution, as described in the background section on committors, is the equilibrium density-weighted average committor, denoted $\langle q^+ \rangle_\pi$. Figure 11a shows this quantity for $25 \leq h \leq 45$ m and $0.4 \leq \sigma_3 \leq 2.0$ m s^{-1} day$^{-1/2}$. Two trends are clearly expected from the basic physics of the model. First, as seen in Figs. 7 and 8, $\langle q^+ \rangle_\pi$ should increase with h. Second, in the limit of large noise and infinite domain

size, the dominance of Brownian motion will smooth out the committor function and make $\langle q^+ \rangle_\pi$ tend to an intermediate value between zero and one. On the other hand, as noise approaches zero, the dynamical system becomes increasingly deterministic, and the ultimate destination of a trajectory will depend entirely on which basin of attraction it starts in. The boundary, or *separatrix*, between these two basins is the stable manifold of the third (unstable) fixed point. In the case of a
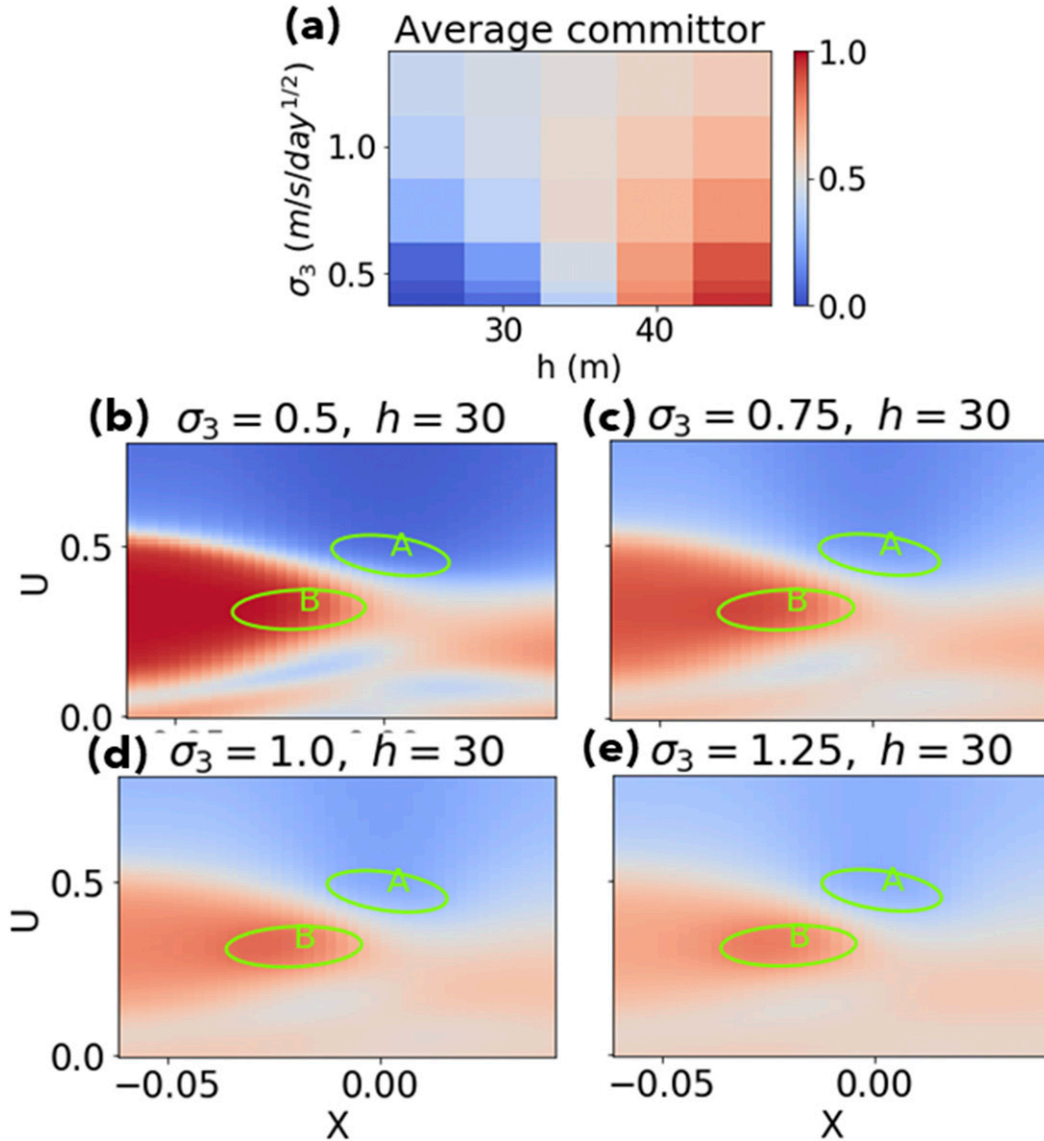
FIG. 11. Behavior of the committor as a function of forcing h and noise σ_3. (a) The average committor (weighted by equilibrium density): $\int q^+(x)\pi(x)\,dx$ evaluated for a range of h and σ_3 values. (b)–(e) Images of the committor for $h = 30$ m and $\sigma_3 = 0.5, 0.75, 1.0$, and 1.25, respectively. The blue lobes in the lower part of the images, a shadow of the spiral structure from Fig. 6, thin out and disappear with increasing forcing h.

potential system, of the form $\dot{x} = -\nabla V(x)$, this boundary would be a literal ridge of the function V. Our system admits no such potential function, but this is a useful visual analogy. The committor function becomes a step function in the deterministic limit, with the discontinuity located exactly on this boundary. The addition of low noise moves the committor 1/2 surface away from the separatrix, possibly asymmetrically: one basin will shrink, becoming more precarious with respect to random perturbations, while the other will expand, becoming a stronger global attractor. Which basin will shrink is not evident a priori, so we compute the averaged committor,

$\langle q^+ \rangle_\pi$, as a summary statistic that will increase when the basin of B expands.

Figure 11a plots the trends in $\langle q^+ \rangle_\pi$ as a function of h (along the horizontal axis) and σ_3 (along the vertical axis). The two basic hypotheses are verified: $\langle q^+ \rangle_\pi$ increases monotonically as h increases, and $\langle q^+ \rangle_\pi \sim 1/2$ as σ_3 increases, no matter the value of h. The less predictable behavior is in the range $h = 35$ m, $0.75 \leq \sigma_3 \leq 1.0\,\mathrm{m\,s^{-1}\,day^{-1/2}}$, where $\langle q^+ \rangle_\pi$ displays non-monotonicity with respect to noise, at low noise levels. As σ_3 increases, the average committor increases from $\sim$0.4 to $\sim$0.6, and then decreases again. The four
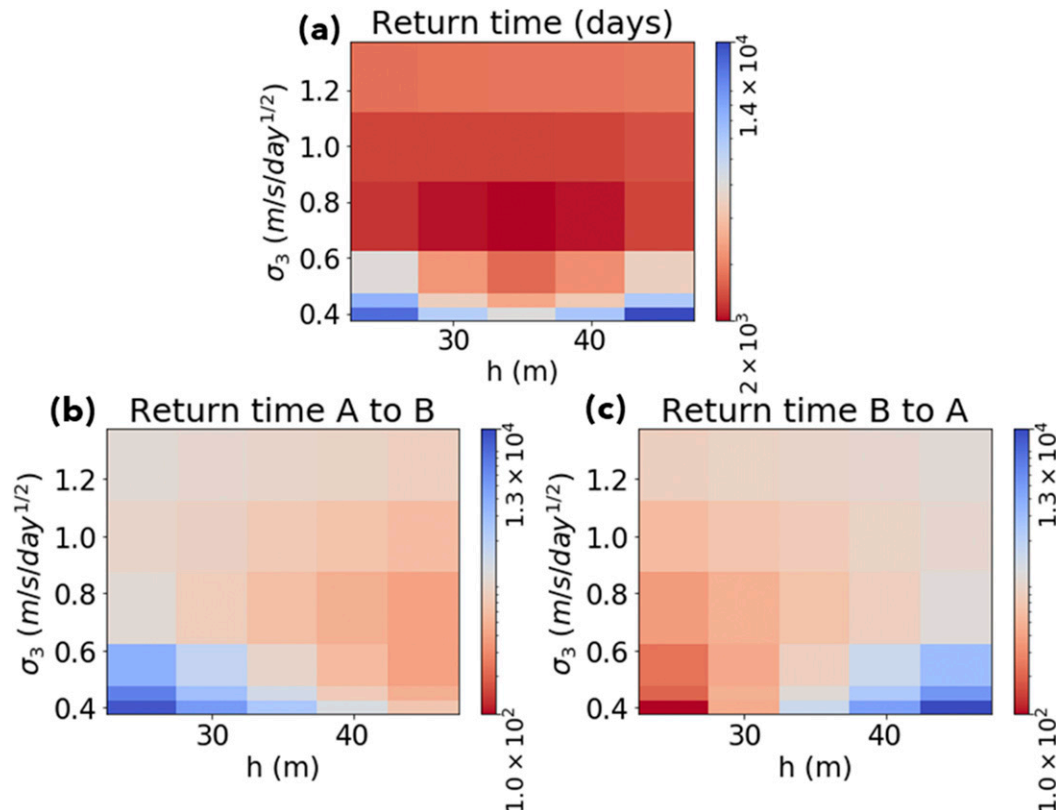
FIG. 12. Behavior of return times as a function of forcing h and noise σ_3. (a) The average period between the start of one transition event and the start of the next. Red here means many transitions per unit time, for both $A \rightarrow B$ and $B \rightarrow A$. We call this the return time, and calculate it as the reciprocal of R, the number of forward (or backward) transitions per unit time. There is clearly a parameter set: $(h, \sigma_3) \approx (35\,\text{m}, 0.75\,\text{m s}^{-1}\,\text{day}^{-1/2})$ that optimizes the number of transitions per unit time. Below this noise level, internal variability is scarcely enough to jump between regions. Above this noise level, sets A and B are no longer metastable, and excursions are so wide and frequent that passing from set A to set B is a very spatially restricted event. (b),(c) Forward and backward transition times, respectively. More specifically, (b) shows the expected passage time T_{AB}, the interval between the end of a $B \rightarrow A$ transition and the end of the next $A \rightarrow B$ transition, and (c) shows the analogous backward passage time T_{BA}. Note the scales are logarithmic, and here red simply means faster transitions, regardless of which direction is being considered.

committor plots at the bottom of 11 illustrate the trend graphically. At low noise, the A basin includes winding passageways leading from the small-U region back to A. Small additive noise closes them off, effectively expanding the B basin. As noise increases and Brownian motion dominates the dynamics, committor values everywhere relax back to less-extreme values, reflecting the unbiased nature of Brownian motion.

Despite the coarse grid resolution, the first-order effect of noise is clear. At the low and high margins of h, where falling into state A and B respectively is virtually certain, an increase in noise decreases this virtual certainty, and the trend continues at larger noise to attenuate $\langle q^+ \rangle_\pi$ to its limit of 1/2. The middle h range, however, behaves differently. Whereas $h = 35\,\text{m}$ appears to balance out the basin sizes at low noise, a slight noise increase tends to kick the system out of the A basin

and toward B, more so than the other way around. At higher noise, the committor relaxes back to 1/2. Examining the committor fields, it is clear that the isocommittor surface location does not move back and forth; rather, it moves toward A, and then the rest of the field flattens out.

Figure 12 shows trends in the return times of SSW with varying h and σ_3. There are several different return times of interest. The first, shown in Fig. 12a, is the total expected time between one transition event and the next, whose reciprocal is the *rate* R_{AB}, the number of transitions per unit time. The return time is a symmetric quantity between A and B, since every forward transition is accompanied by a backward one. Among the parameter combinations, $(h, \sigma_3) \approx (35\,\text{m}, 0.75\,\text{m s}^{-1}\,\text{day}^{-1/2})$ is the one that minimizes return time, or equivalently maximizes the transitions per unit time. $h = 35\,\text{m}$ is a

forcing level that approximately balances out the time spent between the two sets, making transitions relatively common. At lower noise, transitions are exceedingly rare, and at higher noise the two states cease to be long lived. However, this symmetric quantity does not capture information about the relative speed of transition from A to B versus from B back to A. Figure 12b shows a different passage time, which is the average time between the end of a backward $(B \to A)$ transition and the end of the next forward $(A \to B)$ transition, which we call T_{AB}. This is computed as the reciprocal of the rate constant k_{AB}, as described in the previous section. In other words, the stopwatch begins when the system returns to A after having last visited B, and ends when the system next hits B. This metric is asymmetric: a smaller $A \to B$ return time indicates that the forward transition is faster than the backward transition. Figure 12b shows the complementary $B \to A$ return time T_{BA}. Unsurprisingly, an increase in h causes a decrease in T_{AB} and an increase in T_{BA} regardless of the noise level. The noise level has a less obvious effect. Whereas T_{BA} decreases monotonically with increasing noise, regardless of h, the forward time T_{AB} is minimized by a midrange noise level of $\sigma_3 \approx 0.75\,\mathrm{m\,s^{-1}\,day^{-1/2}}$. This is another reflection of the bias toward state B that is effected by adding noise to a very low baseline.

6. Conclusions

Transition path theory is a framework for describing rare transitions between states. We have described TPT along with a number of its key ingredients like the forward and backward committor functions. While it has been applied primarily to molecular systems, we believe it offers valuable insight into climate and weather phenomena such as sudden stratospheric warming, primarily through committors, reactive densities, and reactive currents. Of interest apart from its role in TPT, the committor defines an optimal probabilistic forecast, borne out by direct numerical simulation experiments. The reactive densities and currents describe the geometric properties of dominant transition mechanisms at low noise. In applying TPT to a noisy, truncated Holton–Mass model, we find that transitions tend to begin with a drop in mean zonal wind and a reversal of the streamfunction's phase velocity at a particular streamfunction phase. This is consistent with the significance of blocking precursors to SSW as found in Martius et al. (2009), insofar as this idealized model can represent them. We also find that noise has a nonmonotonic effect on the overall preference for a vacillating state, measured by the average committor. At a forcing of $h = 35\,\mathrm{m}$, where the iso-committor surface (essentially the basin boundary of the deterministic dynamics) divides the space approximately

in half, we find that raising the noise tilts the balance decisively toward the vacillating solution. Still larger noise evens the whole field out. The transition rate constant shows a similar dependence on h and σ_3.

In future work, we plan to scale these methods to more realistic and complex models as well as observational data, where predicting SSW remains an active area of research. In this high-dimensional setting, the generator may be unknown or computationally intractable. Any state space with more than ~5 degrees of freedom is beyond the reach of a finite-volume discretization, because the number of grid cells increases exponentially with dimension. However, there is a generic insight that physical models evolve on very low-dimensional manifolds within the full available state space. A growing body of research in molecular dynamics (Thiede et al. 2019), fluid dynamics (Giannakis et al. 2018; Froyland and Junge 2018), climate dynamics (Giannakis and Majda 2012; Sabeerali et al. 2017), and general multiscale systems (Harlim and Yang 2018; Berry et al. 2015; Giannakis 2015, 2019) exploits the intrinsic low-dimensionality to represent the infinitesimal generator more efficiently. While here we represented the generator as a finite-volume or finite-difference operator on a grid, one can also write it in a basis of globally coherent functions, such as Fourier modes, or more generally harmonic functions on a manifold. Given only data, without an explicit form of the dynamics, this manifold and the basis functions can be estimated from (for example) the diffusion maps algorithm, and the generator's action on this basis can be approximated from short trajectories. These ideas can be applied to computing the dynamical statistics that have been the focus of this paper (Thiede et al. 2019). We hope that these techniques will enable more efficient observation strategies for targeted data assimilation procedures with the goal of tracking the progression of specific extreme events, including hurricanes and heat waves as well as sudden stratospheric warming.

APPENDIX

Numerical Constants

Below are the numerical coefficients used in the reduced-order Ruzmaikin model, with very similar values to Ruzmaikin et al. (2003) and Birner and Williams (2008). The relationship with physical parameters is described in the appendix of Ruzmaikin et al. (2003). Note that our notation differs slightly: following Birner and Williams (2008), we write the topographic forcing in terms of h rather than $\Psi_0 = gh/f_0$, a difference that results in numerical factors of ~ 1000 depending on the convention used:

$$\tau_1 = 122.6, \tag{A1}$$

$$\tau_2 = 30.4, \tag{A2}$$

$$r = 0.63, \tag{A3}$$

$$s = 1.96, \tag{A4}$$

$$\xi = 1.75, \tag{A5}$$

$$\delta_w = 70.84, \tag{A6}$$

$$\zeta = 240.54, \tag{A7}$$

$$U_R = 0.47, \tag{A8}$$

$$\eta = 9.13 \times 10^4, \tag{A9}$$

$$\delta_\Lambda = 4.91 \times 10^{-3}, \tag{A10}$$

$$\dot{\Lambda} = 0. \tag{A11}$$

REFERENCES

Albers, J. R., and T. Birner, 2014: Vortex preconditioning due to planetary and gravity waves prior to sudden stratospheric warmings. *J. Atmos. Sci.*, **71**, 4028–4054, https://doi.org/10.1175/JAS-D-14-0026.1.

Baldwin, M. P., and T. J. Dunkerton, 2001: Stratospheric harbingers of anomalous weather regimes. *Science*, **294**, 581–584, https://doi.org/10.1126/science.1063315.

Bancalá, S., K. Krüger, and M. Giorgetta, 2012: The preconditioning of major sudden stratospheric warmings. *J. Geophys. Res.*, **117**, D04101, https://doi.org/10.1029/2011JD016769.

Banisch, R., and E. Vanden-Eijnden, 2016: Direct generation of loop-erased transition paths in non-equilibrium reactions. *Faraday Discuss.*, **195**, 443–468, https://doi.org/10.1039/C6FD00149A.

Bao, M., X. Tan, D. L. Hartmann, and P. Ceppi, 2017: Classifying the tropospheric precursor patterns of sudden stratospheric warmings. *Geophys. Res. Lett.*, **44**, 8011–8016, https://doi.org/10.1002/2017GL074611.

Berry, T., D. Giannakis, and J. Harlim, 2015: Nonparametric forecasting of low-dimensional dynamical systems. *Phys. Rev.*, **91E**, 032915, https://doi.org/10.1103/PhysRevE.91.032915.

Birner, T., and P. D. Williams, 2008: Sudden stratospheric warmings as noise-induced transitions. *J. Atmos. Sci.*, **65**, 3337–3343, https://doi.org/10.1175/2008JAS2770.1.

Bou-Rabee, N., and E. Vanden-Eijnden, 2015: Continuous-time random walks for the numerical solution of stochastic differential equations. American Mathematical Society Paper, 124 pp.

Bowman, G. R., K. A. Beauchamp, G. Boxer, and V. S. Pande, 2009: Progress and challenges in the automated construction of Markov state models for full protein systems. *J. Chem. Phys.*, **131**, 124101, https://doi.org/10.1063/1.3216567.

Butler, A. H., D. J. Seidel, S. C. Hardiman, N. Butchart, T. Birner, and A. Match, 2015: Defining sudden stratospheric warmings. *Bull. Amer. Meteor. Soc.*, **96**, 1913–1928, https://doi.org/10.1175/BAMS-D-13-00173.1.

Charlton, A. J., and L. M. Polvani, 2007: A new look at stratospheric sudden warmings. Part I: Climatology and modeling benchmarks. *J. Climate*, **20**, 449–469, https://doi.org/10.1175/JCLI3996.1.

Chodera, J. D., and F. Noe, 2014: Markov state models of biomolecular conformational dynamics. *Curr. Opin. Struct. Biol.*, **25**, 135–144, https://doi.org/10.1016/j.sbi.2014.04.002.

Christiansen, B., 2000: Chaos, quasiperiodicity, and interannual variability: Studies of a stratospheric vacillation model. *J. Atmos. Sci.*, **57**, 3161–3173, https://doi.org/10.1175/1520-0469(2000)057<3161:CQAIVS>2.0.CO;2.

DelSole, T., and B. F. Farrell, 1995: A stochastically excited linear system as a model for quasigeostrophic turbulence: Analytic results for one- and two-layer fluids. *J. Atmos. Sci.*, **52**, 2531–2547, https://doi.org/10.1175/1520-0469(1995)052<2531:ASELSA>2.0.CO;2.

Dematteis, G., T. Grafke, and E. Vanden-Eijnden, 2018: Rogue waves and large deviations in deep sea. *Proc. Natl. Acad. Sci. USA*, **115**, 855–860, https://doi.org/10.1073/pnas.1710670115.

Dunkerton, T., C.-P. F. Hsu, and M. E. McIntyre, 1981: Some Eulerian and Lagrangian diagnostics for a model stratospheric warming. *J. Atmos. Sci.*, **38**, 819–844, https://doi.org/10.1175/1520-0469(1981)038<0819:SEALDF>2.0.CO;2.

Durrett, R., 2013: *Probability: Theory and Examples*. Cambridge University Press, 430 pp.

E, W., and E. Vanden-Eijnden, 2006: Towards a theory of transition paths. *J. Stat. Phys.*, **123**, 503–523, https://doi.org/10.1007/s10955-005-9003-9.

——, and ——, 2010: Transition-path theory and path-finding algorithms for the study of rare events. *Annu. Rev. Phys. Chem.*, **61**, 391–420, https://doi.org/10.1146/annurev.physchem.040808.090412.

Easterling, D. R., G. A. Meehl, C. Parmesan, S. A. Changnon, T. R. Karl, and L. O. Mearns, 2000: Climate extremes: Observations, modeling, and impacts. *Science*, **289**, 2068–2074, https://doi.org/10.1126/science.289.5487.2068.

Franzke, C., and A. J. Majda, 2006: Low-order stochastic mode reduction for a prototype atmospheric GCM. *J. Atmos. Sci.*, **63**, 457–479, https://doi.org/10.1175/JAS3633.1.

Froyland, G., and O. Junge, 2018: Robust fem-based extraction of finite-time coherent sets using scattered, sparse, and incomplete trajectories. *SIAM J. Appl. Dyn. Syst.*, **17**, 1891–1924, https://doi.org/10.1137/17M1129738.

Giannakis, D., 2015: Dynamics-adapted cone kernels. *SIAM J. Appl. Dyn. Syst.*, **14**, 556–608, https://doi.org/10.1137/140954544.

——, 2019: Data-driven spectral decomposition and forecasting of ergodic dynamical systems. *Appl. Comput. Harmon. Anal.*, **47**, 338–396, https://doi.org/10.1016/j.acha.2017.09.001.

——, and A. J. Majda, 2012: Nonlinear Laplacian spectral analysis for time series with intermittency and low-frequency variability. *Proc. Natl. Acad. Sci. USA*, **109**, 2222–2227, https://doi.org/10.1073/pnas.1118984109.

——, A. Kolchinskaya, D. Krasnov, and J. Schumacher, 2018: Koopman analysis of the long-term evolution in a turbulent convection cell. *J. Fluid Mech.*, **847**, 735–767, https://doi.org/10.1017/jfm.2018.297.

Harlim, J., and H. Yang, 2018: Diffusion forecasting model with basis functions from QR-decomposition. *J. Nonlinear Sci.*, **28**, 847–872, https://doi.org/10.1007/s00332-017-9430-1.

Hasselmann, K., 1976: Stochastic climate models: Part I. Theory. *Tellus*, **28**, 473–485, https://doi.org/10.3402/tellusa.v28i6.11316.

Hoffman, R. N., J. M. Henderson, S. M. Leidner, C. Grassotti, and T. Nehrkorn, 2006: The response of damaging winds of a simulated tropical cyclone to finite-amplitude perturbations of different variables. *J. Atmos. Sci.*, **63**, 1924–1937, https://doi.org/10.1175/JAS3720.1.

Holton, J. R., and C. Mass, 1976: Stratospheric vacillation cycles. *J. Atmos. Sci.*, **33**, 2218–2225, https://doi.org/10.1175/1520-0469(1976)033<2218:SVC>2.0.CO;2.

Inatsu, M., N. Nakano, S. Kusuoka, and H. Mukougawa, 2015: Predictability of wintertime stratospheric circulation examined using a nonstationary fluctuation–dissipation relation. *J. Atmos. Sci.*, **72**, 774–786, https://doi.org/10.1175/JAS-D-14-0088.1.

Karatzas, I., and S. E. Shreve, 1998: *Brownian Motion and Stochastic Calculus.* Springer, 470 pp.

Kitsios, V., and J. S. Frederiksen, 2019: Subgrid parameterizations of the eddy–eddy, eddy–mean field, eddy–topographic, mean field–mean field, and mean field–topographic interactions in atmospheric models. *J. Atmos. Sci.*, **76**, 457–477, https://doi.org/10.1175/JAS-D-18-0255.1.

Limpasuvan, V., D. W. J. Thompson, and D. L. Hartmann, 2004: The life cycle of the Northern Hemisphere sudden stratospheric warmings. *J. Climate*, **17**, 2584–2596, https://doi.org/10.1175/1520-0442(2004)017<2584:TLCOTN>2.0.CO;2.

Lu, J., and E. Vanden-Eijnden, 2014: Exact dynamical coarse-graining without time-scale separation. *J. Chem. Phys.*, **141**, 044109, https://doi.org/10.1063/1.4890367.

Martius, O., L. M. Polvani, and H. C. Davies, 2009: Blocking precursors to stratospheric sudden warming events. *Geophys. Res. Lett.*, L14806, https://doi.org/10.1029/2009GL038776.

Metzner, P., C. Schutte, and E. Vanden-Eijnden, 2006: Illustration of transition path theory on a collection of simple examples. *J. Chem. Phys.*, **125**, 084110, https://doi.org/10.1063/1.2335447.

——, ——, and ——, 2009: Transition path theory for Markov jump processes. *Multiscale Model. Simul.*, **7**, 1192–1219, https://doi.org/10.1137/070699500.

Pavliotis, G. A., 2014: *Stochastic Processes and Applications.* Springer, 339 pp.

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. *J. Adv. Model. Earth Syst.*, **11**, 863–891, https://doi.org/10.1029/2018MS001419.

Ragone, F., J. Wouters, and F. Bouchet, 2018: Computation of extreme heat waves in climate models using a large deviation algorithm. *Proc. Natl. Acad. Sci. USA*, **115**, 24–29, https://doi.org/10.1073/pnas.1712645115.

Ruzmaikin, A., J. Lawrence, and C. Cadavid, 2003: A simple model of stratospheric dynamics including solar variability. *J. Climate*, **16**, 1593–1600, https://doi.org/10.1175/1520-0442-16.10.1593.

Sabeerali, C. T., R. S. Ajayamohan, D. Giannakis, and A. J. Majda, 2017: Extraction and prediction of indices for monsoon intraseasonal oscillations: An approach based on nonlinear Laplacian spectral analysis. *Climate Dyn.*, **49**, 3031–3050, https://doi.org/10.1007/s00382-016-3491-y.

Scott, R. K., and L. M. Polvani, 2006: Internal variability of the winter stratosphere. Part I: Time-independent forcing. *J. Atmos. Sci.*, **63**, 2758–2776, https://doi.org/10.1175/JAS3797.1.

Sigmond, M., and J. F. Scinocca, 2010: The influence of the basic state on the Northern Hemisphere circulation response to climate change. *J. Climate*, **23**, 1434–1446, https://doi.org/10.1175/2009JCLI3167.1.

Sjoberg, J. P., and T. Birner, 2012: Transient tropospheric forcing of sudden stratospheric warmings. *J. Atmos. Sci.*, **69**, 3420–3432, https://doi.org/10.1175/JAS-D-11-0195.1.

Thiede, E. H., D. Giannakis, A. R. Dinner, and J. Weare, 2019: Galerkin approximation of dynamical quantities using trajectory data. *J. Chem. Phys.*, **150**, 244111, https://doi.org/10.1063/1.5063730.

Thompson, D. W. J., M. P. Baldwin, and J. M. Wallace, 2002: Stratospheric connection to Northern Hemisphere wintertime weather: Implications for prediction. *J. Climate*, **15**, 1421–1428, https://doi.org/10.1175/1520-0442(2002)015<1421:SCTNHW>2.0.CO;2.

Tripathi, O. P., and Coauthors, 2016: Examining the predictability of the stratospheric sudden warming of January 2013 using multiple NWP systems. *Mon. Wea. Rev.*, **144**, 1935–1960, https://doi.org/10.1175/MWR-D-15-0010.1.

Vanden-Eijnden, E., 2014: Transition path theory. *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*, Springer, 91–100.

——, and J. Weare, 2013: Data assimilation in the low noise regime with application to the Kuroshio. *Mon. Wea. Rev.*, **141**, 1822–1841, https://doi.org/10.1175/MWR-D-12-00060.1.

Weare, J., 2009: Particle filtering with path sampling and an application to a bimodal ocean current model. *J. Comput. Phys.*, **228**, 4312–4331, 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**, 053109, https://doi.org/10.1063/1.5081461.

Yasuda, Y., F. Bouchet, and A. Venaille, 2017: A new interpretation of vortex-split sudden stratospheric warmings in terms of equilibrium statistical mechanics. *J. Atmos. Sci.*, **74**, 3915–3936, https://doi.org/10.1175/JAS-D-17-0045.1.

Yoden, S., 1987: Bifurcation properties of a stratospheric vacillation model. *J. Atmos. Sci.*, **44**, 1723–1733, https://doi.org/10.1175/1520-0469(1987)044<1723:BPOASV>2.0.CO;2.