Learning forecasts of rare stratospheric transitions from short simulations

Justin Finkel*

Committee on Computational and Applied Mathematics, University of Chicago

Robert J. Webber

Courant Institute of Mathematical Sciences, New York University

Edwin P. Gerber

Courant Institute of Mathematical Sciences, New York University

Dorian S. Abbot

Department of the Geophysical Sciences, University of Chicago

Jonathan Weare

Courant Institute of Mathematical Sciences, New York University

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

ABSTRACT

Nonlinear atmospheric dynamics produce rare events that are hard to predict and attribute due to many interacting degrees of freedom. Sudden stratospheric warming event is a model example. Approximately once every other year, the winter polar vortex in the boreal stratosphere rapidly breaks down, inducing a shift in midlatitude surface weather patterns persisting for up to 2-3 months. In principle, lengthy numerical simulations can be used to predict and understand these rare transitions. For complex models, however, the cost of the direct numerical simulation approach is often prohibitive. We describe an alternative approach which only requires relatively short-duration computer simulations of the system. The methodology is illustrated by applying it to a prototype model of an SSW event developed by Holton and Mass (1976) and driven with stochastic forcing. While highly idealized, the model captures the essential nonlinear dynamics of SSWs and exhibits the key forecasting challenge: the dramatic separation in timescales between the dynamics of a single event and the return time between successive events. We compute optimal forecasts of sudden warming events and quantify the limits of predictability. Statistical analysis relates these optimal forecasts to a small number of interpretable physical variables. Remarkably, we are able to estimate these quantities using a data set of simulations much shorter than the timescale of the warming event. This methodology is designed to take full advantage of the high-dimensional data from models and observations, and can be employed to find detailed predictors of many complex rare events arising in climate dynamics.

1. Introduction

As computing power increases and weather models grow more intricate and capable of generating a vast wealth of realistic data, the once-distant goal of extreme weather event prediction is starting to become plausible (Vitart and Robertson 2018). To take full advantage of the increased computing power, we must develop new approaches to efficiently manage and parse the data we generate (or observe) to derive physically interpretable, actionable insights. Extreme weather events are worthy targets for simulation owing to their destructive potential to life and property. Rare events have attracted significant simulation efforts recently, including hurricanes (Zhang and Sippel 2009; Webber et al. 2019; Plotkin et al. 2019), heat waves (Ragone et al. 2018), rogue waves (Dematteis et al. 2018), and space weather events, e.g., coronal mass ejections (Ngwira et al. 2013). These are very difficult to characterize and predict, being exceptionally rare and pathological outliers in the spectrum of weather events.

Large ensemble simulations are the most detailed source of data to assess the frequency, intensity, and correlates of extreme weather events (e.g., Schaller et al. 2018). A single simulation must span decades to incorporate the possible impacts of climate change and decadal-scale variability and the full state space of a climate model may be billions of dimensions large, depending on grid resolution. Unfortunately, the data-richness of a long simulation comes at the cost of sample size: even the largest ensembles are limited to tens or hundreds of members as a matter of computational necessity. Under stationary background parameters and some ergodicity properties, a single simulation will eventually sample state space thoroughly and provide all relevant statistics. In practice, however, simulations are often not run long enough to reach steady state, and furthermore one may wish to change parameters over time. A much larger number of independent ensemble members would then be needed to quantify the effects of initial conditions, changing climatology, feedbacks, and unresolved high-frequency variability with statistical confidence (Sillmann et al. 2017; Webber et al. 2019).

While the last decade has seen exciting progress in the development of targeted rare event simulation in geophysical contexts (Hoffman et al. 2006; Weare 2009; Bouchet et al. 2011, 2014; Vanden-Eijnden and Weare 2013; Chen et al. 2014; Yasuda et al. 2017; Farazmand and Sapsis 2017; Dematteis et al. 2018; Mohamad and Sapsis 2018; Dematteis et al. 2019; Webber et al. 2019; Bouchet et al. 2019a,b; Plotkin et al. 2019; Simonnet et al. 2020; Ragone and Bouchet 2020; Sapsis 2021), predicting long time-scale behavior of complex dynamical systems remains a difficult task. A traditional approach to addressing this issue is through dimensional reduction techniques which seek to replace an expensive, high-fidelity model with a lower-dimensional and less costly model. Physics-based reduced-order models have a long and very successful history in atmospheric science, especially as prototypes of chaos and multistability (Lorenz 1963; Charney and DeVore 1979; Legras and Ghil 1985; Crommelin 2003; Timmermann et al. 2003; Ruzmaikin et al. 2003). Observationally, regime behavior has been diagnosed by projecting the empirical steady-state distributions onto leading EOFs (e.g., Crommelin 2003). More recently, significant attention has been paid to data-based dimensional reduction techniques that use data generated by the high-fidelity model to specify a more quantitatively accurate reduced-order model (e.g., Giannakis et al. 2018; Berry et al. 2015; Sabeerali et al. 2017; Majda and Qi 2018; Wan et al. 2018; Bolton and Zanna 2019; Chattopadhyay et al. 2020; Chen and Majda 2020). However the reduced-order model is derived, it can subsequently be thoroughly interrogated by direct computer simulation.

We advance an alternative computational approach to predicting and understanding rare events without sacrificing model fidelity. Like data-informed reduced order modeling, our method relies on data generated by a high-fidelity model. However, unlike dimensional reduction techniques, our approach focuses on computing specific quantities of interest rather than on capturing all aspects of a very complicated, high dimensional dynamical system. In particular we will compute estimators of statistically optimal forecasts using a data set of many short forward simulations. To accomplish this we represent these forecasts as solutions to Feynman-Kac equations. In the continuous time limit, these become partial differential equations (PDE) with a number of independent variables equal to the dimension of the model state space. It is therefore hopeless to solve the equations using any standard spatial discretization. As we demonstrate nonetheless, the equations can be solved with remarkable accuracy via an expansion in a basis of functions informed by the data set. Importantly, our approach to solving these equations is independent of the model used to generate the data, avoiding unrealistic simplifications or structural assumptions.

As typical examples of the forecasts computatable within our framework, we focus on the probability that a warming event occurs before a return to a "typical" state, as well as the expected time that it takes for that event to occur. Both quantities depend on the initial condition, and are therefore functions over all of state space. We will follow the convention in computational statistical mechanics and refer to these as the committor function and mean first passage time (MFPT) respectively. The committor has been computed previously for low dimensional atmospheric models in Tantet et al. (2015); Lucente et al. (2019); Finkel et al. (2020). Forecasts like these quantify the risk associated with an event given the current state of the system. They also encode important information regarding the rare event itself.

Even putting aside the difficulty of computing the committor and MFPT, they still must be 'decoded'; knowledge of these functions does not automatically reveal insights into the fundamental causes or precursers of a rare event. Nor are they easily applied to observations of a limited collection of variables. They are, after all, complicated functions of a high dimensional model state space. In Section 5 we will demonstrate a detailed statistical analysis of our computed committor function aimed at identifying a relatively small subset of the original variables capable of describing the committor (in the sense defined below).

We illustrate our approach on the highly simplified Holton-Mass model (Holton and Mass 1976; Christiansen 2000) with stochastic velocity perturbations in the spirit of Birner and Williams (2008). The Holton-Mass model is well-understood dynamically in light of decades of analysis and experiments, yet complex enough to present the essential computational difficulties of probabilistic forecasting and test our methods for addressing them. Despite the challenges posed by its 75-dimensional state space, our computational framework can indeed accurately characterize of extreme events with unprecedented detail using only a data set of short model simulations. In the future, the same methodology could be applied to query the properties of more complex models where less theoretical understanding is available.

Section 2 reviews the dynamical model, and section 3 describes a general class of methods which we then apply to our problem specifically in section 4. We present the results in section 5, including a discussion of optimal forecasting and physical insights gleaned from our approach. We then lay out future prospects and conclude in section 6.

2. Holton-Mass model

Holton and Mass (1976) devised a simple model of the stratosphere aimed at reproducing observed intra-seasonal oscillations of the polar vortex, which they termed “stratospheric vacillation cycles." Earlier SSW models, originating with that of Matsuno (1971), proposed upward-propagating planetary waves as the major source of disturbance to the vortex. This was a significant step outside the bounds of the nonacceleration theorem of Charney and Drazin (1961), which stated that the vortex would be robust to disturbances under a variety of conditions. While Matsuno (1971) used impulsive forcing from the troposphere as the source of planetary waves, Holton and Mass

(1976) suggested that stationary tropospheric forcing, if large enough, could lead to an oscillatory response, purely through dynamics internal to the stratosphere.

Radiative cooling through the stratosphere and wave perturbations at the tropopause are the two competing forces that drive the vortex in the Holton-Mass model. Altitude-dependent cooling relaxes the zonal wind toward a strong vortex in thermal wind balance with a radiative equilibrium temperature field. Gradients in potential vorticity along the vortex, however, can allow the propagation of Rossby waves. When conditions are just right, a Rossby wave emerges from the tropopause and rapidly propagates upward, sweeping heat poleward and stalling the vortex by depositing a burst of negative momentum. The vortex is destroyed and begins anew the rebuilding process.

Yoden (1987a) found that for a certain range of parameter settings, these two effects balance each other to create two distinct stable regimes: a strong vortex with zonal wind close to the radiative equilibrium profile, and a weak vortex with a possibly oscillatory wind profile. We focus our study on this bistable setting as a prototypical model of atmospheric regime behavior. The transition from strong to weak vortex state captures the essential dynamics of an SSW. The methodology presented here, using only observed short trajectories, can be applied equally to any of these models as well as observational data, which the reader should keep in mind as we present the specifics of the present application.

The Holton-Mass model takes the linearized quasigeostrophic potential vorticity (QGPV) equation for a perturbation streamfunction $\psi'(x,y,z,t)$ on top of a zonal mean flow $\overline{u}(y,z,t)$, and projects these two fields onto a single zonal wavenumber $k = 2/(a\cos 60°)$ and a single meridional wavenumber $\ell = 3/a$, where a is the Earth's radius. The resulting ansatz is

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

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

which is fully determined by the reduced state space $U(z,t)$, and $\Psi(z,t)$, the latter being complex. Inserting this into the linearized QGPV equations yields the coupled PDE system

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

$$\begin{aligned}
&\left(-\mathcal{G}^2\ell^2-\frac{\partial}{\partial z}+\frac{\partial^2}{\partial z^2}\right)\frac{\partial U}{\partial t}=\left[(\alpha_z-\alpha)U_z^R-\alpha U_{zz}^R\right] \\
&\quad-\left[(\alpha_z-\alpha)\frac{\partial}{\partial z}+\alpha\frac{\partial^2}{\partial z^2}\right]U+\frac{\varepsilon k\ell^2}{2}e^z\mathrm{Im}\left\{\Psi\frac{\partial^2\Psi^*}{\partial z^2}\right\}
\end{aligned} \tag{3}$$

where we have nondimensionalized the equations with the parameter $\mathcal{G}^2=H^2N^2/(f_0^2L^2)$ in order to create a homogeneously shaped dataset more suited to our analysis. Boundary conditions are prescribed at the bottom of the stratosphere, which in this model corresponds to $z=0$, and the top of the stratosphere $z_{top}=70\,km$.

$$\begin{aligned}
\Psi(0,t)&=\frac{gh}{f_0} &\qquad \Psi(z_{top},t)&=0 \\
U(0,t)&=U^R(0) &\qquad \partial_z U(z_{top},t)&=\partial_z U^R(z_{top})
\end{aligned} \tag{4}$$

The vortex-stabilizing influence is represented by $\alpha(z)$, the altitude-dependent cooling coefficient, and the linear relaxation profile $U^R(z)=U^R(0)+\frac{\gamma}{1000}z$, which forces the vortex toward radiative equilibrium. Here $\gamma=\mathcal{O}(1)$ is the vertical wind shear in m/s/km. The competing force of wave perturbation is encoded through the lower boundary condition $\Psi(0,t)=gh/f_0$.

Detailed bifurcation analysis of the model by both Yoden (1987a) and Christiansen (2000) in (γ,h) space revealed the bifurcations that lead to bistability, vacillations, and ultimately quasiperiodicity and chaos. Here we will focus on an intermediate parameter setting of $\gamma=1.5$ m/s/km and $h=38.5$ m, where two stable states coexist: a strong vortex with U closely following U^R and an almost barotropic streamfunction, as well as a weak vortex with U dipping close to zero at an intermediate altitude and a disturbed streamfunction with strong westward phase tilt. The two stable equilibria, which we call **a** and **b**, are represented in the first row of Figure 1 by their z-dependent zonal wind and streamfunction profiles.

The two equilibria can be interpreted as two different winter climatologies, one with a strong vortex and one with a weak vortex susceptible to vacillation cycles. To explore transitions between these two states, we follow Birner and Williams (2008) and modify the Holton-Mass equations with small additive noise in the U variable to mimic momentum perturbations by smaller scale Rossby waves, gravity waves, and other unresolved sources. While the details of the additive noise are ad hoc, this approach can be more rigorously justified through the Mori-Zwanzig formalism (Zwanzig 2001). Because many hidden degrees of freedom are being projected onto the low-dimensional space of the Holton-Mass model, the dynamics on small observable subspaces can be considered stochastic. This is the perspective taken in stochastic parameterization of turbulence and other high-dimensional chaotic systems (Hasselmann 1976; DelSole and Farrell 1995; Franzke and Majda 2006; Majda et al. 2001; Gottwald et al. 2016). More sophisticated parameterizations would surely influence the results (Hu et al. 2019), but would not present a fundamental problem for our purely data-driven numerical method.

We follow Holton and Mass (1976) and discretize the equations using a finite-difference method in z, with 27 vertical levels (including boundaries). After constraining the boundaries, there are $d = 3 \times (27 - 2) = 75$ degrees of freedom in the model. Christiansen (2000) investigated higher resolution and found negligible differences. The full discretized state is represented by a long vector

$$
\begin{aligned}
\mathbf{X}(t) = \Big[&\mathrm{Re}\Psi(\Delta z, t), \ldots, \mathrm{Re}\Psi(z_{top} - \Delta z, t), \\
&\mathrm{Im}\Psi(\Delta z, t), \ldots, \mathrm{Im}\Psi(z_{top} - \Delta z, t), \\
&U(\Delta z, t), \ldots, U(z_{top} - \Delta z, t\Big] \in \mathbb{R}^d = \mathbb{R}^{75}
\end{aligned}
\tag{5}
$$

Boundary terms are not included because they are constrained by the boundary conditions. The deterministic system can be written $d\mathbf{X}(t)/dt = m(\mathbf{X}(t))$ for a vector field $m : \mathbb{R}^d \to \mathbb{R}^d$ specified by discretizing (2) and (3). Under deterministic dynamics, $\mathbf{X}(t) \to \mathbf{a}$ or $\mathbf{X}(t) \to \mathbf{b}$ as $t \to \infty$ depending on initial conditions. The addition of white noise changes the system into an Itô diffusion

$$
d\mathbf{X}(t) = m(\mathbf{X}(t))\,dt + \sigma(\mathbf{X}(t))\,d\boldsymbol{W}(t) \tag{6}
$$

where $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ imparts a correlation structure to the vector $\boldsymbol{W}(t) \in \mathbb{R}^d$ of independent standard white noise processes. We design σ to be a low-rank, constant matrix that adds spatially smooth stirring to only the zonal wind U (not the streamfunction Ψ) and respects boundary conditions at the bottom and top of the stratosphere. We simulate the model using the Euler-Maruyama method: in a timesetep δt, after a deterministic forward Euler step we add the stochastic perturbation to zonal wind on large vertical scales

$$
\delta U(z) = \sigma_U \sum_{k=0}^{2} \eta_k \sin\left[\left(k + \frac{1}{2}\right)\pi \frac{z}{z_{top}}\right]\sqrt{\delta t} \tag{7}
$$

where η_k $(k = 0, 1, 2)$ are independent unit normal samples. We set the magnitude of σ by

$$
\sigma_U^2 = \frac{\mathbb{E}[(\delta U)^2]}{\delta t} \approx (1\,\mathrm{m/s})^2/\mathrm{day} \tag{8}
$$

σ_U technically has units of $(L/T)/T^{1/2}$, where the square-root of time comes from the quadratic variation of the Wiener process. It is best interpreted in terms of the daily root-mean-square velocity perturbation of 1.0 m/s.

A long stochastic simulation of the model reveals metastability, with the system tending to remain close to one fixed point for a long time before switching quickly to the other, as shown by the timeseries of $U(30\,km)$ in panel (c) of Figure 1. We display the zonal wind U at 30 km following Christiansen (2000), because this is about where zonal wind strength is minimized in the weak vortex. While the two regimes are clearly associated with the two fixed points, they are better characterized by extended *regions* of state space with strong and weak vortices. We thus define the two metastable subsets of $\mathbb{R}^d$

$$A = \{\mathbf{X} : U(30\,km)(\mathbf{X}) \geq U(30\,km)(\mathbf{a}) = 53.8\,m/s\}$$

$$B = \{\mathbf{X} : U(30\,km)(\mathbf{X}) \leq U(30\,km)(\mathbf{b}) = 1.75\,m/s\}$$

This straightforward definition roughly follows the convention of Charlton and Polvani (2007), which defines an SSW as a reversal of zonal winds at 10 hPa. In the Holton-Mass model, where $z = 0$ at the tropopause, this translates to $z = -7\,km\ln(10/1000) - 10\,km = 22.2\,km$, but we have adjusted the specific altitude here to 30 km where the zonal wind reduction is most drastic. There is lively debate around the definition of SSW events (e.g., Butler et al. 2015), with different thresholds leading to different statistics. The details are affected by the definition in our analysis, but the results are qualitatively similar over a wide range. Our method is equally applicable to any definition, and so to illustrate we choose one that enjoys broad acceptance. Incidentally, the analysis tools we present may be helpful in distinguishing predictability properties between different definitions.

The green highlights in Figure 1 (c) begin precisely when the system exits the A region bound for B, and end when the system enters B. The orange highlights start when the system leaves B bound for A, and end when A is reached. Note that $A \to B$ transitions are much shorter in duration than $B \to A$ transitions. Figure 1 (d) shows the same paths, but viewed in the space $(|\Psi|, U)$ at 30 km. The $A \to B$ and $B \to A$ transitions are again highlighted in green and orange respectively, showing geometrical differences between the two directions. We will refer to the $A \to B$ transition as an SSW event, even though it is more accurately a transition between climatologies according to the Holton-Mass interpretation. The $B \to A$ transition is a vortex restoration event. Our focus in this paper is on predicting transition events and monitoring their progress in a principled way. In the next section we explain the formalism for doing so.

3. Theory and computation

a. Definitions

We will introduce the quantities of interest by way of several simple, important examples. Suppose the stratosphere is observed in an initial state $\mathbf{X}(0) = \mathbf{x}$ that is neither in A nor B, so $U(\mathbf{b})(30\,km) < U(\mathbf{x})(30\,km) < U(\mathbf{a})(30\,km)$ and the vortex is somewhat weakened, but not completely broken down. We call this intermediate zone $D = (A \cup B)^{\mathrm{c}}$ (the complement of the two metastable sets). Because A and B are attractive, the system will soon find its way to one or the other at the *first-exit time* from D, denoted

$$\tau_{D^{\mathrm{c}}} = \min\{t \geq 0 : \mathbf{X}(t) \in D^{\mathrm{c}}\} \tag{9}$$

Because of stochastic forcing (which in practice arises from unobserved variables), the first-exit time is a random variable, formally called a "stopping time" (Oksendal 2003; Durrett 2013), meaning measurable from the history of $\mathbf{X}(t)$. The first-exit location $\mathbf{X}(\tau_{D^{\mathrm{c}}})$ is itself a random variable which importantly determines how the system exits D: either $\mathbf{X}(\tau_{D^{\mathrm{c}}}) \in A$, meaning the vortex restores to radiative equilibrium, or $\mathbf{X}(\tau_{D^{\mathrm{c}}}) \in B$, meaning the vortex breaks down into vacillation cycles. A fundamental goal of probabilistic forecasting is to determine the probabilities of these two events, which naturally leads to the definition of the (forward) committor function

$$q^+(\mathbf{x}) = \begin{cases} \mathbb{P}_{\mathbf{x}}\{\mathbf{X}(\tau_{D^c}) \in B\} & \mathbf{x} \in D = (A \cup B)^c \\ 0 & \mathbf{x} \in A \\ 1 & \mathbf{x} \in B \end{cases} \tag{10}$$

where the subscript $\mathbf{x}$ indicates that the probability is conditional on a fixed initial condition $\mathbf{X}(0) = \mathbf{x}$, i.e., $\mathbb{P}_{\mathbf{x}}\{\cdot\} = \mathbb{P}\{\cdot|\mathbf{X}(0) = \mathbf{x}\}$. (The superscript "+" distinguishes the forward committor from the *backward committor*, an analogous quantity for the time-reversed process which we do not use in this paper.) Throughout, we will use capital $\mathbf{X}(t)$ to denote a stochastic process, and lower-case $\mathbf{x}$ to represent a specific point in state space, typically an initial condition, i.e., $\mathbf{X}(0) = \mathbf{x}$. Both are $d = 75$-dimensional vectors. The boundary conditions in Equation (10) naturally extend the definition of q^+ in D: if the vortex starts out very weak, $\mathbf{x}$ is close to set B and the system will probably land in B next, making $q^+(\mathbf{x}) \approx 1$. If it starts out strong and close to A, it will most likely restore to A next, making $q^+(\mathbf{x}) \approx 0$. The committor is clearly a function of initial condition $\mathbf{x}$, but assuming the process is Markovian, it does not depend on the history of the system that led it to $\mathbf{x}$ in the first place.

Another important forecasting quantity is the lead time to the event of interest. While the forward committor reveals the probability of experiencing vortex breakdown *before* returning to a strong vortex, it does not say how long either event will take in absolute terms. Furthermore, even if the vortex is restored first, how long will it be until the next SSW does occur? The time until the next SSW event is denoted τ_B, again a random variable, whose distribution depends on the initial condition $\mathbf{x}$. We call $\mathbb{E}_{\mathbf{x}}[\tau_B]$ the *mean first passage time* (MFPT) to B. Conversely, we may ask how long a vortex disturbance will persist before normal conditions return; the answer (on average) is $\mathbb{E}_{\mathbf{x}}[\tau_A]$, the mean first passage time to A. Dissecting the expectations further, we may condition

τ_B on the event that an SSW is coming before the strong vortex returns, leading to the conditional first passage time $\mathbb{E}_{\mathbf{x}}[\tau_B|\tau_B < \tau_A]$, which in some sense quantifies the suddenness of SSW.

All of these quantities can, in principle, be estimated by collecting averages over very long simulations. For example, to estimate the committor at a given $\mathbf{x}$, one can shoot N trajectories starting from $\mathbf{x}$ and count the numbers N_A and N_B hitting A and B first. Then N_A/N will be an estimate for the committor at $\mathbf{x}$. The mean first passage time can be estimated using these same sampled trajectories. But this direct method can be prohibitively expensive, especially if applied to many points all over state space. By definition, transitions between A and B are infrequent. Therefore, if starting from $\mathbf{x}$ far from B, then a huge number of sampled trajectories (N) will be required to observe even a small number ending in B (N_B). Likewise, transition path statistics such as return times can, in principle, be computed from an extremely long model run, but in most cases of interest this direct simulation approach will not be feasible.

In Subsection 3(b) we will write these forecasts in a single general form, and describe a computational approach to compute them using only a data set of short forward model integrations. The method is called the Dynamical Galerkin Approximation (DGA), introduced in Thiede et al. (2019). It takes advantage of the Feynman-Kac formula (Oksendal 2003), recasting conditional expectations as PDE problems over state space. These equations are *local* and thus approximable by short trajectories. We perform these calculations on the Holton-Mass model and present the results in Section 4. In Section 5(a) we describe a statistical analysis to aid interpretation of the estimated committor.

b. Dynamical Galerkin Approximation

In this section we describe the methodology, which involves some technical results from stochastic processes and measure theory. The forecast functions described above—committors and MFPTs—

can all be written as conditional expectations of the form

$$F(\mathbf{x}) = \mathbb{E}_{\mathbf{x}}\bigg[G(\mathbf{X}(\tau))e^{-\int_0^\tau V(\mathbf{X}(s))\,ds} + \int_0^\tau H(\mathbf{X}(s))e^{-\int_0^s V(\mathbf{X}(r))\,dr}\,ds\bigg] \tag{11}$$

where again the subscript $\mathbf{x}$ denotes conditioning on $\mathbf{X}(0) = \mathbf{x}$; G, H and V are arbitrary known functions over $\mathbb{R}^d$; and τ is a stopping time, specifically a first-exit time like Equation (9) but possibly with D replaced by another set. To see that the forward committor takes on this form, set $G(\mathbf{x}) = \mathbb{1}_B(\mathbf{x})$ (one on set B and zero everywhere else), $V = 0$, $H = 0$, and $\tau = \tau_{D^c}$. Then $F(\mathbf{x}) = \mathbb{E}_{\mathbf{x}}\big[\mathbb{1}_B(\mathbf{X}(\tau))\big] = \mathbb{P}_{\mathbf{x}}\{\mathbf{X}(\tau_{D^c}) \in B\} = q^+(\mathbf{x})$. For the mean first passage time to B, set $\tau = \tau_B$, $G = 0$, $V = 0$, and $H = 1$. Then $F(\mathbf{x}) = \mathbb{E}_{\mathbf{x}}\big[\int_0^{\tau_B} dt\big] = \mathbb{E}_{\mathbf{x}}[\tau_B]$. For the conditional first passage time to B, set $G = \mathbb{1}_B$, $\tau = \tau_{D^c}$, $V = -\lambda$ (a constant) and $H = 0$. Then the expectation can be computed in two steps, by computing and differentiating a moment-generating function:

$$\begin{aligned} F(\mathbf{x};\lambda) &= \mathbb{E}_{\mathbf{x}}\big[\mathbb{1}_B(\mathbf{X}(\tau_{D^c}))e^{\lambda\tau_{D^c}}\big] \\ \frac{1}{q^+(\mathbf{x})}\frac{\partial}{\partial\lambda}F(\mathbf{x};0) &= \frac{\mathbb{E}_{\mathbf{x}}[\tau_{D^c}\mathbb{1}_B(\mathbf{X}(\tau_{D^c}))}{\mathbb{E}_{\mathbf{x}}[\mathbb{1}_B(\mathbf{X}(\tau_{D^c}))]} \\ &= \mathbb{E}_{\mathbf{x}}[\tau_B|\tau_B < \tau_A] \end{aligned} \tag{12}$$

Note that the event $\tau_B < \tau_A$ is equivalent to $\mathbf{X}(\tau_{D^c}) \in B$, where again $D^c = A \cup B$. We approximate $\partial F/\partial\lambda$ with a centered finite difference, after computing $F(\mathbf{x};\lambda)$ for several λ in the neighborhood of zero. In principle we could compute higher moments in the same way and get a more detailed understanding of the conditional passage time distribution. Alternatively we could estimate $F(\mathbf{x};\lambda)$ for a large range of λ and recover the distribution with a Laplace transform.

More generally, the function G is chosen by the user to quantify risk at the terminal time τ; in the case of the forward committor, that risk is binary, with an SSW representing a positive risk and a radiative vortex no risk at all. The function H is chosen to quantify the risk accumulated up until time τ, which might be simply an event's duration, but other integrated risks may be of more interest for the application. For example, one could express the total thermal energy absorbed by the polar vortex by setting $H = \overline{v'T'}$, or the momentum lost by the vortex by setting $H(\mathbf{x}) = U(\mathbf{a}) - U(\mathbf{x})$, or a vertically integrated version. Using the moment generating function in (12), one can compute not only means but higher moments of such integrals by expressing the risk with V.

Let us now describe how to numerically compute $F(\mathbf{x})$ of the form (11) with short trajectories, starting with the special case of the forward committor and then generalizing. Consider starting a random trajectory at $\mathbf{x} = \mathbf{X}(0) \in D = (A \cup B)^{\mathrm{c}}$ and evolving it for a short time Δt. Its probability of reaching B first, $q^+(\mathbf{x})$, is simply the probability that it reaches B first starting from $q^+(\mathbf{X}(\Delta t))$ instead, averaged over all possible $\mathbf{X}(\Delta t)$ (ignoring momentarily the small probability that $A \cup B$ is reached before time Δt). That is, $q^+(\mathbf{x}) \approx \mathbb{E}_{\mathbf{x}}[q^+(\mathbf{X}(\Delta t))] =: \mathcal{T}^{\Delta t} q^+(\mathbf{x})$. The operator $\mathcal{T}^{\Delta t}$ is known as the (stochastic) transition operator, which maps a function on state space to the expectation of that function at a future time. We could furthermore divide by Δt and take the limit $\Delta t \to 0$, eliminating the event $\tau_{D^{\mathrm{c}}} < \Delta t$, and obtain the Kolmogorov Backward PDE (e.g., Oksendal 2003; Weinan et al. 2019). Instead, to represent our numerical method more directly, we implement a purely finite-time approach from Strahan et al. (2020): artifically halt the dynamics upon first arrival in $D^{\mathrm{c}} = A \cup B$ and modify the equation to

$$\begin{aligned} &\mathbb{E}_{\mathbf{x}}[q^+(\mathbf{X}(\Delta t \wedge \tau_{D^{\mathrm{c}}}))] - q^+(\mathbf{x}) \\ &\qquad = (\mathcal{T}_{D^{\mathrm{c}}}^{\Delta t} - 1) q^+(\mathbf{x}) = 0 \end{aligned} \tag{13}$$

where $\Delta t \wedge \tau_{D^{\mathrm{c}}} := \min(\Delta t, \tau_{D^{\mathrm{c}}})$ and $\mathcal{T}_{D^{\mathrm{c}}}^{\Delta t}$ is a "stopped" transition operator. This equation holds for $\mathbf{x} \in D$, and comes with the boundary condition $q^+(\mathbf{x}) = \mathbb{1}_B(\mathbf{x})$ for $\mathbf{x} \in A \cup B$. Applying similar logic to the mean first passage time to B, let $\mathbf{x} \in B^{\mathrm{c}}$, denote $m_B(\mathbf{x}) = \mathbb{E}_{\mathbf{x}}[\tau_B]$ and observe that $\tau_{D^{\mathrm{c}}}$

decreases by $\Delta t \wedge \tau_{B^c}$ during the short timespan. So for all $\mathbf{x} \in B^c$,

$$\begin{aligned}\mathbb{E}_{\mathbf{x}}[m_B(\mathbf{X}(\Delta t \wedge \tau_B^+))] - m_B(\mathbf{x}) \\ = (\mathcal{T}_B^{\Delta t} - 1)m_B(\mathbf{x}) = -\mathbb{E}_{\mathbf{x}}[\Delta t \wedge \tau_B]\end{aligned} \tag{14}$$

with the boundary condition $m_B(\mathbf{x}) = 0$ for $\mathbf{x} \in B$. Now in the general case, let D stand in for the relevant region of state space and let G, H and V be arbitrary. The corresponding operator equation is

$$\begin{aligned}(\mathcal{T}_{D^c}^{\Delta t} - 1)F(\mathbf{x}) - \mathbb{E}_{\mathbf{x}}\left[\int_0^{\Delta t \wedge \tau_{D^c}} V(\mathbf{X}(t))F(\mathbf{X}(t))\,dt\right] \\ = -\mathbb{E}_{\mathbf{x}}\left[\int_0^{\Delta t \wedge \tau_{D^c}} H(\mathbf{X}(t))\,dt\right]\end{aligned} \tag{15}$$

for $\mathbf{x} \in D$, with boundary condition $F(\mathbf{x}) = G(\mathbf{x})$ for $\mathbf{x} \in D^c$. This linear equation comes from Dynkin's formula, an integrated version of the Feynman-Kac; see Oksendal (2003); Karatzas and Shreve (1998); Weinan et al. (2019) theoretical background. The remarkable aspect of this formula is that while F is an expectation over paths going all the way to the boundary D^c (a strong or weak vortex), it obeys a *local* equation with expectations over short trajectories of length Δt. By collecting many short-trajectory samples, we can compute statistical properties of the event without ever actually observing one happen in simulation. Note that (15) reduces to (13) with $V = H = 0$ and (14) with $V = 0, H = 1$.

Like a PDE with a high dimensional independent variable space, Equation (15) cannot be solved using any classical discretization of the possible values of $\mathbf{x}$. Successful approaches will involve a representation of the solution, F, suitable for the high dimensional setting, i.e. representations of the type commonly employed for machine learning tasks. The DGA method, in particular, consists of expanding the unknown function F in a "data-informed" basis (to be specified later). The expectations in Equation (15) are estimated by launching short trajectories from all over state space. Finally, a finite system of equations is solved for the unknown coefficients in the basis expansion of F, in effect stitching together information from all trajectories at once.

We can express the essential idea using the example of Equation (13) for $q^+(\mathbf{x})$, while the supplement contains a more general version. We first homogenize the boundary conditions with a guess function $\hat{q}^+(x)$ that obeys the boundary conditions $\hat{q}^+|_A = 0$, $\hat{q}^+|_B = 1$, and let $r(x) = q^+(x) - \hat{q}^+(x)$, so that r obeys homogeneous Dirichlet conditions and satisfies

$$(\mathcal{T}_{D^c}^{\Delta t} - 1)r(x) = -(\mathcal{T}_{D^c}^{\Delta t} - 1)\hat{q}^+(x) \tag{16}$$

We next expand r in a finite-dimensional basis of functions $\{\phi_1, \ldots, \phi_M\}$ with unknown coefficients c_j: $r(x) = \sum_{j=1}^{M} c_j(r)\phi_j(x)$. Each ϕ_j obeys the homogeneous boundary conditions. Finally, we take the inner product of both sides with ϕ_i, with respect to some measure μ, to produce a system of M linear equations

$$\sum_{j=1}^{M} \langle \phi_i, (\mathcal{T}_{D^c}^{\Delta t} - 1)\phi_j \rangle_\mu c_j(r) = -\langle \phi_i, (\mathcal{T}_{D^c}^{\Delta t} - 1)\hat{q}^+ \rangle_\mu$$

$$i = 1, \ldots, M \tag{17}$$

These inner products are intractable integrals over high-dimensional state space, but can be approximated using Monte Carlo integration. If $\mathbf{X}$ is an $\mathbb{R}^d$-valued random variable distributed according to μ, and we have access to random samples $\{\mathbf{X}_1, \ldots, \mathbf{X}_N\}$, the law of large numbers gives

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(\mathbf{X}_n) = \int_{\mathbb{R}^d} f(\mathbf{x})\mu(d\mathbf{x}) \tag{18}$$

This is where the short trajectory data enters the picture. We generate a dataset of length-Δt trajectories $\{\mathbf{X}_n(t) : 0 \le t \le \Delta t, n = 1, \ldots, N\}$. These short trajectories might enter A or B before time Δt, and to account for this we also store the stopping times $\Delta t \wedge \tau_{n,D^c}$. The $\mathbf{X}_n(0)$'s are sampled from an arbitrary measure μ, called the sampling measure, which is determined by the sampling procedure for initial points. For example, if points are selected randomly from a long trajectory, $\mu \approx \pi$ (the steady-state probability density) by ergodicity. However, we may choose μ so that many samples appear in regions of particular interest, such as transition regions far away from A and B and to which π assigns very little probability. Once the dataset is generated, we use μ as the reference measure for the inner products in (17), allowing us to approximate them with Monte Carlo integration. For example,

$$\begin{aligned}&\langle \phi_i, (\mathcal{T}_{D^c}^{\Delta t} - 1)\phi_j\rangle_\mu \\ &\approx \frac{1}{N}\sum_{n=1}^{N} \phi_i(\mathbf{X}_n(0))\left[\phi_j(\mathbf{X}_n(\Delta t \wedge \tau_{n,D^c}) - \phi_j(\mathbf{X}_n(0))\right]\end{aligned} \tag{19}$$

We can similarly estimate any expectation of the form (15) using different basis functions adapted to the specific region of interest.

The formulation above works for any class of basis functions that becomes increasingly expressive as the library grows, capable of estimating any function of interest. However, with a finite truncation, choosing the basis functions is a crucial ingredient of DGA, greatly impacting the efficiency and accuracy of the results. In our current study, we restrict to the simplest kind of basis, which consists of indicator functions $\phi_i(x) = \mathbb{1}_{S_i}(x)$, where $\{S_1, \dots, S_M\}$ is a disjoint partition of state space. In practice we will construct these sets by clustering data. This basis set construction is borrowed from common practice in the computational statistical mechanics community for building a Markov State Model (MSM) (Frank and Fischer 2008; Pande et al. 2010; Bowman et al. 2013; Chodera and Noé 2014). MSMs are a dimensional reduction technique that has also been used in conjuction with analysis of metastable transitions, primarily in protein folding dynamics (Noé et al. 2009) and were recently used to study ocean circulation in Miron et al. (2021). DGA can be viewed as an extension of MSMs, though, rather than producing any reduced complexity model, the explicit goal in DGA is the estimation of specific functions as in Equation (11). The supplement spells out DGA in considerably greater detail, which may be more helpful to view after seeing the forthcoming results.

4. Methods

In this section we explain our specific application of the Dynamical Galerkin Approximation (DGA) method (Thiede et al. 2019; Strahan et al. 2020) to the Holton-Mass model and validate our results empirically from simulation.

a. Data generation

There are many possible ways to choose starting points for the short trajectories. Whatever procedure we use will induce a *sampling measure* μ on state space. $\mu(\mathbf{x})$ is a probability density that specifies the expected number of starting points per unit volume in the region of state space near $\mathbf{x}$. This is a natural reference measure for the Monte Carlo inner products described in Subsection 3(b). Because μ has minimal requirements, the user is afforded great flexibility in sampling the data. How to efficiently generate maximally informative data is an active and nontrivial research question, but a few heuristics are obvious. In a metastable system, setting $\mu = \pi$ would be a poor choice, because the data would be strongly concentrated in the immediate neighborhoods of A and B, whereas the regions of primary interest are the transition regions somewhere in between A and B. Different physical observables, such as the Eliassen-Palm flux, may be important prior candidates for their predictive power, and we might like to seed data samples uniformly over a certain range of that variable. On the other hand, the samples should fall within a physically realistic region of state space, not just any point in $\mathbb{R}^{75}$. To see why, recall that the last 25 entries of the state vector $\mathbf{X}$ represent the velocity field at discretized vertical levels from $z = 0\,km$ to $z = 70\,km$. Because velocity is a continuous function of altitude, adjacent entries should be close together, which is not at all guaranteed for a randomly chosen 75-dimensional vector.

Because a goal of this article is to demonstrate interpretable results of rare event analysis on a climate model, we choose an easy, probably suboptimal sampling strategy. We defer optimization to later work, perhaps for a more expensive model that demands it. We define our sampling distribution as the equilibrium distribution, re-weighted to be uniform over the space $(U(30\,km), |\Psi|(30\,km))$ within the bounds realized by the control simulation, which are approximately $-30\,m/s \leq U(30\,km) \leq 70\,m/s$ and $0\,m^2/s \leq |\Psi|(30\,km) \leq 2\times 10^7\,m^2/s$. Without direct access to the equilibrium distribution, we approximate it by running a very long trajectory of 500,000 days, producing many transitions like those shown in Figure 1, with an Euler-Maruyama timestep of 0.005 days (for comparison, a single transition event takes on the order of 100 days). We acknowledge this is cheating on our claim to only use short trajectories; however, we use the long simulation only to seed the initial conditions for those short trajectories, as well as to empirically validate the results of DGA later on. This way we can emphasize the power of DGA itself, which will motivate more efficient upstream data generation methods. Alternatives exist for sampling state space thoroughly without a long simulation, for example trajectory-splitting (e.g., L'Ecuyer et al. 2007). One could initialize trajectories in one of the metastable sets, say in A directly on the fixed point $\mathbf{a}$, and integrate the trajectories for a short time to explore the surrounding region. These new data points can be used as initial conditions for the next round of simulation, at each stage exploring a wider region of state space until the bulk of the attractor is covered. This initialization procedure may require a long total simulation time, but is parallelizable. We will explore and optimize such methods in future work with more sophisticated models, where efficient initialization is more critical. For now we settle for initial data points from a long simulation.

After downsampling the long simulation to a resolution of 0.5 days, we sample snapshots from the trajectory, reweighted to induce a uniform distribution on the space $(U(30\,km), |\Psi(30\,km)|)$. Specifically, we compute a discrete histogram over the two-dimensional space and weight each sample by the inverse of its density on that histogram. We collect $N = 1\times 10^6$ snapshots $\{\mathbf{X}_n(0)\}$ directly from the long simulation, and then launch independent (hence completely parallelizable) short 10-day trajectories from each, to obtain the short trajectory database $\{\mathbf{X}_n(t) : 0 \leq t \leq \Delta t, n = 1\ldots,N\}$. Afterward we identify the first-entry times to D^c for each trajectory, called τ_{n,D^c}. This strategy is straightforward and guarantees that μ gives substantial probability to candidate transition regions and that only physically reasonable points are sampled.

b. Computation and validation

The partition $\{S_1,\ldots,S_M\}$ to build the basis function library $\{\mathbb{1}_{S_j}(\mathbf{x})\}_{n=1}^N$ should be chosen with a number of considerations in mind. The partition elements should be small enough to accurately represent the functions they are used to approximate, but large enough to contain sufficient data to robustly estimate transition probabilities. We form these sets by a hierarchical modification of K-means clustering on $\{\mathbf{X}_n(0)\}_{n=1}^N$. K-means is a robust method that can incorporate new samples by simply identifying the closest centroid, and is commonly used in molecular dynamics (Pande et al. 2010). However, straightforward application of K-means, as implemented in the `scikit-learn` software (Pedregosa et al. 2011), can produce a very imbalanced cluster size distribution, even with empty clusters. This leads to unwanted singularities in the constructed Markov matrix. To avoid this problem we cluster hierarchically, starting with a coarse clustering of all points and iteratively refining the larger clusters, at every stage enforcing a minimum cluster size, until we have the desired number of clusters (M). After clustering on the initial points $\{\mathbf{X}_n(0)\}$, the other points $\{\mathbf{X}_n(t), 0 < t \leq \Delta t\}$ are placed into clusters using an address tree produced by the K-means cluster hierarchy. To guarantee that D and D^c consist exactly of a union of subsets, we cluster points in D and D^c separately, with a number of clusters proportional to the number of points therein. (We remind the reader that the domain and boundary depend on which quantity of interest is being computed. For the forward and backward committor, D^c consists of A and B, which are defined *a priori* by thresholds of $U(30\,km)$.) The total number of clusters is fixed to $M = 1500$. When doing out-of-sample extension on a point z, we first identify whether $z \in D$ or D^c, and assign it to a cluster accordingly.

Figure 2 demonstrates the accuracy of the calculated forward committor and mean first passage time to B by taking advantage of the long trajectory from which we sampled the short trajectories. We divide the interval $(0,1)$ into 20 bins, and identify for each interval (ζ_1, ζ_2) which data points $\{\mathbf{X}_n(0) : \zeta_1 < q^+(\mathbf{X}_n(0)) < \zeta_2\}$ were en route to the vacillating regime at the instant they were selected from the long simulation. If the committor is computed accurately, the proportion of data points headed to B should fall in the interval (ζ_1, ζ_2). For example, about 20-25% of data points $\mathbf{X}_n(0)$ whose committor is calculated to be within $(0.2, 0.25)$ should be headed to set B. Analogously, we expect rain 20% of the time the National Weather Service forecasts a 20% chance of rain. This is a very coarse measure of accuracy, and only a necessary condition, but the strong empirical match shown in the scatter plots of Figure 2 gives us confidence in our numerical results. The mean first passage time calculation is evaluated similarly: for all data points $\mathbf{X}_n(0)$ with the estimated $m_B(\mathbf{X}_n(0))$ in a certain range (t_1, t_2), we average the true first-passage time observed from the long trajectory. The match is quite good up until very long lag times, where DGA underestimates the long tail. The accuracy of committors and passage times improve as the dataset grows and clusters are refined. More sophisticated basis sets and sampling methods may significantly improve the convergence rate.

The committor and first passage time relate to the weather forecasting problem of predicting the next rare event given the current initial condition. However, they can also characterize the polar vortex climatology, meaning its average behavior over very long time periods as pertains to A and B. To wit, how much does the system "prefer" to be in a weak or strong state, as measured by the fraction of time it spends in either? This can be quantified by the steady state distribution (also called the invariant or stationary measure) $\pi(\mathbf{x})$, the probability distribution function produced by binning data points from a very long simulation. Figure 3 illustrates that the metastable Holton-Mass model has a starkly bimodal distribution, with the system tending to spend a long time in state A or B before occasionally switching quickly to the other state. We have estimated π here using a variation on the DGA recipe described above. The details of the calculation can be found in the supplement. We have projected π onto the two-dimensional subspace $(|\Psi(30\,km)|, U(30\,km))$ on a log scale, along with a one-dimensional projection onto the latter coordinate $U(30\,km)$ on a linear scale. The preferences for A and B can be quantitatively compared by the fraction of time spent inside each set, as well as the fraction of time spent between the two sets but destined for either one. These ergodic averages can be found by averaging the forward committor over different regions of state space. For example, the fraction of time spent inside A is $\int_A \pi(d\mathbf{x}) = \int_{\mathbb{R}^d} \mathbb{1}_A(\mathbf{x})\pi(d\mathbf{x}) = \langle \mathbb{1}_A \rangle_\pi$. Similarly, the fraction of time spent inside B is $\langle \mathbb{1}_B \rangle_\pi$; the fraction spent outside A and B but destined for B is $\langle \mathbb{1}_{(A\cup B)^c} q^+ \rangle_\pi$; and the fraction spent outside A and B but destined for B is $\langle \mathbb{1}_{(A\cup B)^c}(1-q^+) \rangle_\pi$. Table 1 displays these fractions calculated from DGA and empirically from the long trajectory. The time spent either in A or destined for A (the first two rows) is about equal to the time spent in or destined for B (last two rows). However, more time is spent destined for B than strictly inside B, as vacillation cycles often increase the zonal wind above 1.75 m/s before it dips back down. Furthermore, Figure 3 shows a higher and narrower peak in the A regime. We interpret that a strong vortex is much less variable than a weak vortex, which is consistent with the vacillation cycles that characterize the latter.

5. Results and Discussion

Our analysis can be roughly divided into two parts. First, from a forecasting perspective, we demonstrate that the committor is more robust than naïve proxies from the model as a leading indicator of an oncoming SSW. We also find a low-rank representation of the committor in terms of the system's basic observables using a sparsity-promoting LASSO regression (Tibshirani 1996). Second, we quantitatively relate the *risk* of an oncoming event with the *lead time* to the event, an important consideration in extreme weather prediction.

a. The committor as an early warning

Operational forecasting requires continuous updating of probabilities from incoming observations, which provide only partial information on the state of the atmosphere. The choice of which observables to monitor is constrained by measurement capabilities, but is also informed by prediction efficacy; we desire warning signs that are highly correlated with the event and occur as early as possible to give some buffer time to brace for impacts. Figure 4 visually demonstrates the advantage of considering the committor as a forecasting metric compared to two other observables: zonal wind U and meridional eddy heat flux $\overline{v'T'}$, both measured at the same altitude of 30 km. We have extracted a typical complete SSW event ($A \to B$ transition path) from the long simulation and plotted a timeseries of the observables on a common time axis. The time $t = 0$ corresponds to the central date of a warming event, the moment when the system first enters set B, with zonal wind at 30 km dropping below the threshold of 1.75 m/s. The committor timeseries (Figure 4a) is estimated by nearest neighbor interpolation from the dataset $\{\mathbf{X}_n(0)\}_{n=1}^{N}$.

The committor curve timeseries first exceeds the threshold of 0.5 around 27 days before the event while rising sharply in a roughly S-shaped curve. A perfect committor-measuring instrument would be sending a strong signal of increasing risk at that time. Compare this with $U(t)$, which is plateauing, or very gradually decreasing, around $40\,m/s$ when the threshold $q^+ = 0.5$ is crossed. The apparently mild behavior belies the rapid increase in SSW risk shown by the committor timeseries. The dramatic drop in zonal wind occurs well after the committor exceeds 0.5, and so a reading of U directly would not give a strong warning sign until late in the progress of transition. One could write the committor as an approximate function of $U(30\,km)$, which is plotted in Figure 5(a) as explained below, and would find that the U-level corresponding to $q^+ = 0.5$ is around $37\,m/s$. Unfortunately, $U(30\,km)$ does not drop below $37\,m/s$ until the SSW is 12 days away, providing much less lead time than if the full committor were known. The considerable gap in prediction date is shown by a blue strip. Meanwhile, the heat flux over time plotted in panel (c) suffers the same deficiency as a predictor, having an analogous threshold of $1.2 \times 10^{-6}\,K \cdot m/s$. The $\overline{v'T'}$ level hardly budges while critical preconditions are falling into place, and only after the die is already cast in favor of a SSW does the heat flux rise sharply. A monitoring system based on heat flux alone would be very ill-informed about the risk of impending SSW event. While heat flux is a dynamically consequential quantity for describing the evolution of an SSW, this does not directly translate into good predictive properties. These prediction gaps are typical: over many simulated transitions, the average delay between $q^+(\mathbf{X}(t))$ clearing 0.5 for the last time the other observables clearing their thresholds for the last time are 9.1 days for $U(30\,km)$ and 9.8 days for $\overline{v'T'}(30\,km)$.

We use the caveat "directly" because the possibility remains that the vertical scale in Figures 4 (b-c) unfairly downplay the predictive power of zonal wind and heat flux. Perhaps they could be very robust predictors, if examined on the right scale and with appropriate (possibly nonlinear) transformations. Calculating such a transformation on theoretical grounds alone would be a daunting task, especially in light of stochastic perturbations. But even if this were possible, at best this calculation would approximate nothing other than the forward committor itself. Furthermore, we incorporate all state variables at once into the committor calculation, which is at least as flexible as considering heat flux or zonal wind alone. Nonetheless, for the sake of dynamical transparency and practical observational constraints, it would be helpful to have a parsimonious representation of the committor in terms of a small number of state variables, if possible. We pursue this prospect in the following subsection.

b. Sparse representation of the committor

The committor's superiority as a probabilistic forecast is not surprising, because it is built into the definition. The committor combines information from every degree of freedom in just the right way to give the probability of next hitting B rather than A. However, these degrees of freedom may not all be "observable" in a practical sense, given the sparsity and resolution limits of weather sensors. It is therefore important to ask: what is the best possible estimate of the committor given an observed subset of state variables? A related question arises in the design of observational systems: which variables should be measured to optimally estimate the committor, under cost and engineering constraints? In this section we will propose a systematic method to address these questions in the context of the Holton-Mass model.

Consider a single-variable observable like U (30 km). If constrained to observe only U (30 km) and forced to approximate $q^+(x)$ as a function of this one variable, we would average $q^+(\mathbf{x})$ across the remaining 74 model dimensions, weighted by the invariant measure. We would assess the quality of this observable by the variance across those projected-out dimensions: a large projected variance would imply strong dependence on unobserved variables. Figure 5 applies this projection to the committor (first row) and mean first passage time to B (second row), using three different single-variable observables: U (first column), $\overline{v'T'}$ (second column), both at 30 km, and the LASSO regression (third column). The solid curves show the projected means, and the dotted curves indicate the one-standard-deviation envelope. $q^+(U(30\,km))$ is a smooth, mostly monotonic curve with a consistently small projection error never exceeding ~ 0.2, which occurs near $q^+ = 0.5$. Compare this to $q^+(\overline{v'T'}(30\,km))$, which is essentially discontinuous as heat flux increases from zero, and which has a large standard deviation approaching 0.3 when heat flux is small. This is consistent with its prediction properties as shown in Figure 4: while the heat flux reading hardly changes at all from zero, crucial processes are acively destabilizing the vortex, with the committor increasing significantly without any response from $\overline{v'T'}$. The projected mean first passage times tell a similar story, being strongly negatively correlated with the committor. Weaker zonal wind generally signals less lead time before entering state B, but while heat flux stays small, an observer is in the dark about how soon, as well as how certain, a transition is.

Let us briefly formalize the projection idea before exploring other variables. We want to approximate a function $F : \mathbb{R}^d \to \mathbb{R}$, such as the committor or mean first passage time, as a function of some reduced coordinates $\boldsymbol{\theta} : \mathbb{R}^d \to \mathbb{R}^k$, called "collective variables" (CVs) in chemistry literature. That is, we wish to find $f : \mathbb{R}^k \to \mathbb{R}$ such that $F(\mathbf{x}) \approx f(\boldsymbol{\theta}(\mathbf{x}))$. For instance, $\boldsymbol{\theta}(\mathbf{x}) = (\theta_1(\mathbf{x}), \theta_2(\mathbf{x}))$ where $\theta_1(\mathbf{x})$ is the mean zonal wind at 30 km and $\theta_2(\mathbf{x})$ is the perturbation streamfunction magnitude $|\Psi|$ at 30 km. Typically the projected dimension $k \ll d$, for instance $k = 1$ or 2 for visualization purposes. The "best" function f is chosen by minimizing some function-space metric between $f \circ \boldsymbol{\theta}$ and F. The simplest choice would be the mean-squared error, so the projection problem is to minimize over functions $f : \mathbb{R}^k \to \mathbb{R}$ the penalty

$$\begin{aligned} S[f;\boldsymbol{\theta}] :&= \|f \circ \boldsymbol{\theta} - F\|^2_{L^2(\pi)} \\ &= \int_{\mathbb{R}^d} \Big[f(\boldsymbol{\theta}(\mathbf{x})) - F(\mathbf{x}) \Big]^2 \pi(d\mathbf{x}) \end{aligned} \tag{20}$$

The optimal f for this purpose is the conditional expectation $f(\mathbf{y}) = \mathbb{E}_{\mathbf{X}\sim\pi}[F(\mathbf{X})|\boldsymbol{\theta}(\mathbf{X}) = \mathbf{y}] = \int f(\mathbf{x})\delta(\boldsymbol{\theta}(\mathbf{x}) - \mathbf{y})\pi(d\mathbf{x})$. We derive a discretized version of this formula in the supplement, and this is how we display all the low-dimensional projections. We call the square root of $S[f;\boldsymbol{\theta}]$ the projected standard deviation, or projection error, which determines the dotted envelope in Figure 5.

A much harder problem than optimizing over f given $\boldsymbol{\theta}$ is the problem of optimizing over sets of coordinates $\boldsymbol{\theta}$. CVs can be arbitrarily complex nonlinear functions of the basic state variables $\mathbf{x}$. Modern machine learning algorithms such as artificial neural networks are designed exactly for that purpose: to represent functions nonparametrically from observed input-output pairs. However, we wish to maintain some interpretability in the committor representation. For this reason, in searching for optimal projections, we begin with more constrained and physics-informed feature spaces before allowing for more complex relationships. We focus on observables coming from the Eliassen-Palm (EP) relation, which relates wave activity, PV fluxes and gradients, and heating source terms in a conservation equation. From Yoden (1987b), the EP relation for the Holton-Mass model takes the form

$$\begin{aligned}
&\partial_t\left(\frac{\overline{q'^2}}{2}\right) + (\partial_y\overline{q})\rho_s^{-1}\nabla\cdot\boldsymbol{F} \\
&\qquad = -\frac{f_0^2}{N^2}\rho_s^{-1}\overline{q'\partial_z(\alpha\rho_s\partial_z\psi')} \qquad (21)\\
&\text{where } \boldsymbol{F} = (-\rho_s\overline{u'v'})\mathbf{j} + (\rho_s\overline{v'\partial_z\psi'})\mathbf{k}
\end{aligned}$$

In the highly idealized Holton-Mass model, the EP flux divergence has two alternative expressions: $\rho_s^{-1}\nabla\cdot\boldsymbol{F} = \overline{v'q'} = \frac{R}{Hf_0}\rho_s^{-1}\overline{v'T'}$. If there were no dissipation ($\alpha = 0$) and the background zonal state were time-independent ($\partial_t\overline{q} = 0$), dividing both sides by $\partial_y\overline{q}$ would express local conservation of wave activity $\mathcal{A} = \rho_s\overline{q'^2}/(2\partial_y\overline{q})$. Neither of these is true in the stochastic Holton-Mass model, so we use the quantities in Equation (21) as diagnostics: enstrophy $\overline{q'^2}$, PV gradient $\partial_y\overline{q}$, PV flux $\overline{v'q'}$, and heat flux $\overline{v'T'}$. Each field is a function of (y,z) and takes on very different profiles in A and B, as found by Yoden (1987b). A transition from A to B, where the vortex weakens dramatically, must entail a reduction in $\partial_y \overline{q}$ and a burst in positive $\overline{v'T'}$ and negative $\overline{v'q'}$ as a Rossby wave propagates from the tropopause vertically up through the stratosphere. This is the general physical narrative of a sudden warming event, and these same fields might be expected to be useful observables to track for qualitative understanding and prediction, along with the basic state variables U and $|\Psi|$.

One option is to take vertical averages of any of these fields, but there may be particularly salient altitude levels that clarify the role of vertical interactions. The first three rows of Figure 6 display, for three of these fields (U, $|\Psi|$ and $\overline{v'T'}$) and for a range of altitude levels, the mean and standard deviation of the committor projected onto that field at that altitude. Each altitude has a different range of the CV; for example, because U has a Dirichlet condition at the bottom and a Neumann condition at the top, the lower levels have a much smaller range of variability than the high levels. We also plot the integrated variance, or L^2 projection error, at each level in the right-hand column. A low projected committor variance over U at altitude z_0 means that the committor is mostly determined by the single observable $U(z_0)$, while a high projected variance indicates significant dependence of q^+ on variables other than $U(z_0)$. In order to compare different altitudes and fields as directly as possible, the L^2 projection error at each altitude is an average over discrete bins of the observable, not a proper integral.

In selecting good CV's, we generally look for a simple, hopefully monotonic, and sensitive relationship with the committor. Of all the candidate fields, U and $\partial_y \overline{q}$ stand out the most in this respect, being clearly negatively correlated with the forward committor at all altitudes. The associated projection error tends to be greatest in the region $q^+ \approx 0.5$, as observed before, but interestingly there is a small altitude band around $20-25$ km where its magnitude is minimized. This suggests an optimal altitude for monitoring the committor through zonal wind, giving the most reliable estimate possible for a single state variable. In contrast, the projection of q^+ onto $|\Psi|$, displays a large variance across all altitudes. The eddy heat flux is also rather unhelpful as an early warning sign, despite its central role in SSW evolution, which is consistent with Figure 5. For example, the large, positive spikes in heat flux across all altitudes generally occur after the committor ≈ 0.5 threshold has already been crossed. Furthermore, the relationship of $\overline{v'T'}$ with the committor is not smooth. The $q^+ < 0.5$ region at each altitude is a thin band near zero. Even so, the optimal altitude for observing the committor through heat flux is also 20 km.

The exhaustive observable search in Figure 6 is visually compelling, but not completely numerically satisfactory as a comparison between fields. Differences between units and ranges make it difficult to objectively compare the L^2 projection error, despite the normalization mentioned above. Furthermore, restricting to one variable at a time is limiting. Accordingly, in a second, more automated approach to identify salient variables, we perform a sparsity-promoting LASSO regression for the forward committor (Tibshirani 1996; Pedregosa et al. 2011), using as input features all state variables $U, \mathrm{Re}\Psi, \mathrm{Im}\Psi$ and their vertical derivatives. We leave out eddy fluxes, which seem to have poor prediction properties. The advantage of a sparsity-promoting regression is to isolate a small number of observables that can decently approximate the committor in linear combination. Considering that regions close to A and B have low committor uncertainty, we regress only on data points with $q^+ \in (0.2, 0.8)$, and of those only a subset weighted by the reactive probability density $q^+q^-\pi$, since we wish to isolate the dominant transition pathways. To enforce predicted committors being between zero and one, we regress on the probit-transformed committor $\ln(q^+/(1-q^+))$. First we do this at each altitude separately, and in Figure 7 (a) we plot the coefficients of each component as a function of altitude.

Each component is salient for some altitude range. In general, U and U_z dominate as causal variables at low altitudes, while Ψ and Ψ_z dominate at high altitudes. The overall prediction quality, as measured by R^2 and plotted in Figure 7 (b), is greatest around 21.5 km, consistent with our qualitative observations of Figure 6. Note that not all single-altitude slices are sufficient for approximating the committor, even with LASSO regression; in the altitude band $50-60\,km$, the LASSO predictor is not monotonic and has a large projected variance. The specific altitude can matter a great deal. But by using all altitudes at once, the committor approximation may be improved further. We thus repeat the LASSO with all altitudes simultaneously and find the sparse coefficient structure shown in 7 (c), with a few variables contributing the most: U (21.5 km), U (29.6 km), ReΨ_z (13.5 km), and ImΨ (21.5 km). The results of LASSO regression are also displayed in the bottom row of Figure 4, the right column of Figure 5, and the bottom row of Figure 6 for direct comparison with the other candidate fields. With multiple lines of evidence indicating 21.5 km as an altitude with high predictive value for the forward committor, we can make a strong recommendation for targeting observations there. This conclusion applies only to the Holton-Mass model under these parameters, but the methodology explained above can be applied similarly to models of arbitrary complexity.

c. Relationship to lead time

A skillful forecast is only useful if it comes early and leaves some buffer time before impact. Having identified the committor as optimally skillful among all observables, we can now assess the limits of early prediction by relating certainty levels and lead times. Such a relationship would answer two dual questions: during the transition to an SSW winter phase, (1) how far in advance will we be aware of it with some prescribed confidence, say 80%? (2) given some prescribed lead time, say 42 days, how aware or in the dark could we be of it?

These questions clearly involve some kind of first-passage time, like the curves in the bottom row of Figure 5. The same quantity has been calculated previously in other simplified models, e.g.

Birner and Williams (2008) and Esler and Mester (2019). But $\mathbb{E}[\tau_B]$ has an obvious shortcoming. From Figure 5, we see that when $q^+ \approx 0.5$, $\mathbb{E}[\tau_B^+] \approx 600$, an average which includes half the paths going straight into B and the other half returning to A and lingering there before eventually crossing into B. The conditional passage time $\mathbb{E}[\tau_B|\tau_B < \tau_A]$ is designed to highlight only the contribution of the latter half and measure the mean time of paths going directly to B, which can be computed by DGA using a Laplace transform as described in Subsection 3(b). Figure 8 shows all three quantities—the forward committor, mean passage time to B, and conditional passage time to B—this time projected on a two-dimensional observable space $(\mathrm{Im}\Psi(21.5km), U(21.5km))$ identified as salient by sparse regression. Physically, these levels operate as a valve regulating wave propagation into the stratosphere.

The committor has a clear negative relationship with both conditional and unconditional first passage time: as the risk of imminent SSW grows, the time until impact shrinks. Figure 9 shows this relationship more quantitatively, for both the $A \rightarrow B$ process (panel (a)) and the $B \rightarrow A$ process (panel (b)). The relationship is roughly quantified by a least-squares regressions, weighted by the change of measure, between the SSW probability q^+ (resp. the restoration probability $1-q^+$) and the conditional lead time to the SSW event $\mathbb{E}[\tau_B|\tau_B < \tau_A]$ (resp. the conditional lead time to vortex restoration, $\mathbb{E}[\tau_A|\tau_A < \tau_B]$). While the relationships are nonlinear and the spreads significant, the linear fits offer two meaningful numerical insight. The vertical intercept says how long the next excursion to a given state will take when the system starts trapped in the other state. The negative slope says how fast the remaining time shrinks as the risk grows. The vertical intercepts of 79 days and 107 days offer further evidence that the vortex breaks down faster than it restores.

These metrics can inform preparation for extreme weather. For example, a threatened community might decide in advance on an “alarm threshold" of, say, 50%, meaning they plan to prepare for an SSW event only once it is 50% certain to occur. According to the linear fit in panel (a), they must be ready to do so in ~ 48 days time. The nonlinear deviations are, however, significant. The spread around the linear fit increases suddenly towards the lower-right corner of the plot, meaning that the uncertainty in timing, viewed as a function of the committor, increases as the SSW certainty increases. Lead time must therefore depend strongly on more than just the forward committor, and must be estimated by taking more details of the current state into account. We emphasize that the choice of A, B and alarm thresholds are more of a community and policy decision than a scientific one. The strength of our approach is that it provides a flexible numerical framework to quantify and optimize the consequences of those decisions.

6. Conclusion

Forecasting rare events is, by the very nature of rare events, an extremely difficult computational task. Given the dangers posed by climate change, it also one of science's most pressing challenges. We suggest a computational framework that uses relatively short model simulations to make predictions on much longer time scales. Our numerical results point to its promise for forecasting.

Within the context of a stochastically forced Holton-Mass model with 75 degrees of freedom, we have computed fundamental quantities of the SSW transition process, including committor probabilities and expected lead times, for both the vortex destruction and vortex restoration processes. The system is irreversible, making these two directions very statistically distinct from each other. By systematically evaluating many model variables for their utility in predicting the fate of the vortex, we have identified some salient physical descriptions of early warning signs. We have furthermore quantified the relationship between probability and lead time for a given rare event, a potentially useful paradigm for assessing predictability and preparing for extreme weather. Our results suggest that the slow evolution of vortex preconditioning is an important source of predictability. In particular, the zonal wind and streamfunction at 20 km seems to be optimal among a large class of dynamically motivated observables.

The committor and mean first passage time have obvious utility for forecasting, but they are also ingredients in a larger framework called Transition Path Theory (TPT) for describing rare steady state transition events. In principle, interrogating the ensemble of transition paths requires direct simulation of the system long enough to observe many transition events. However, using TPT, quantities computable by our framework can be combined to yield key statistics describing the ensemble of *transition paths* connecting regions in state space, (Finkel et al. 2020; Metzner et al. 2006, 2009; Vanden-Eijnden and E 2010; E. and Vanden-Eijnden 2006). In a following paper we will apply the same short-trajectory forecasting approach together with TPT to compute transition path statistics such as return times and extract insight about physical mechanisms of the transition process.

Our numerical pipeline is promising and robust, but leaves plenty of room for improvement. Our sampling method, while advantageous for validation of results, wastes a great deal of data. Targeted sampling from the transition region has the potential to achieve the same precision for the quantities of interest with much less data. Also, moving beyond a basis expansion of the forecast functions, in upcoming work we will explore more flexible representations using kernel methods and neural networks. The solution of high-dimensional PDEs is an active research area that is making innovative use of machine learning, particularly in the fields of computational chemistry and quantum mechanics (e.g., Chen and Majda 2017; Carleo and Troyer 2017; Han et al. 2018; Khoo et al. 2018; Li et al. 2020; Mardt et al. 2018; Li et al. 2019; Lorpaiboon et al. 2020). Similar approaches may hold great potential for understanding predictability in atmospheric science.

Acknowledgments. J.F. is supported by the U.S. Department of Energy, 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 National Science Foundation via grant RTG/DMS-1646339. E.P.G. acknowledges support from the U. S. National Science Foundation 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 and the Research Computing Center at the University of Chicago. We thank John Strahan, Aaron Dinner, and Chatipat Lorpaiboon for helpful methodological advice. Mary Silber, Noboru Nakamura, and Richard Kleeman offered invaluable scientific insight. J.F. benefitted from many helpful discussions with Anya Katsevich.

Data availability statement. The stochastic Holton-Mass model and analysis techniques are fully described in the text and references, and can be integrated quickly at low computational costs. J.F. is happy to share code and data upon request.

References

Berry, T., D. Giannakis, and J. Harlim, 2015: Nonparametric forecasting of low-dimensional dynamical systems. *Phys. Rev. E*, **91**, 032 915, doi:10.1103/PhysRevE.91.032915.

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.

Bolton, T., and L. Zanna, 2019: Applications of deep learning to ocean data inference and subgrid parameterization. *Journal of Advances in Modeling Earth Systems*, **11 (1)**, 376–399, doi:https://doi.org/10.1029/2018MS001472, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018MS001472, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2018MS001472.

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.

Bowman, G. R., V. S. Pande, and F. Noé, 2013: *An introduction to Markov state models and their application to long timescale molecular simulation*, Vol. 797. Springer Science & Business

Media.

Butler, A. H., D. J. Seidel, S. C. Hardiman, N. Butchart, T. Birner, and A. Match, 2015: Defining sudden stratospheric warmings. *Bulletin of the American Meteorological Society*, **96 (11)**, 1913–1928, doi:10.1175/BAMS-D-13-00173.1.

Carleo, G., and M. Troyer, 2017: Solving the quantum many-body problem with artificial neural networks. *Science*, **355 (6325)**, 602–606, doi:10.1126/science.aag2302, URL https://science.sciencemag.org/content/355/6325/602, https://science.sciencemag.org/content/355/6325/602.full.pdf.

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

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

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

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)**, e2019MS001 958, doi:https://doi.org/10.1029/2019MS001958, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001958, e2019MS001958 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., and A. J. Majda, 2017: Beating the curse of dimension with accurate statistics for the fokker–planck equation in complex turbulent systems. *Proceedings of the National Academy of Sciences*, **114 (49)**, 12 864–12 869, doi:10.1073/pnas.1717017114, URL https://www.pnas.org/content/114/49/12864, https://www.pnas.org/content/114/49/12864.full.pdf.

Chen, N., and A. J. Majda, 2020: Predicting observed and hidden extreme events in complex nonlinear dynamical systems with partial observations and short training time series. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **30 (3)**, 033 101, doi:10.1063/1.5122199, URL https://doi.org/10.1063/1.5122199, https://doi.org/10.1063/1.5122199.

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

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

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

DelSole, T., and B. F. Farrell, 1995: A stochastically excited linear system as a model for quasi-geostrophic turbulence: Analytic results for one- and two-layer fluids. *Journal of the Atmospheric Sciences*, **52 (14)**, 2531–2547, doi:10.1175/1520-0469(1995)052<2531:ASELSA>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.

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

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

Esler, J. G., and M. Mester, 2019: Noise-induced vortex-splitting stratospheric sudden warmings. *Quarterly Journal of the Royal Meteorological Society*, **145 (719)**, 476–494, doi:https://doi.org/

10.1002/qj.3443, URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3443, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3443.

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

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

Frank, N., and S. Fischer, 2008: Transition networks for modeling the kinetics of conformational change in macromolecules. *Current Opininion in Structural Biology*, **18**, 154–163, doi:10.1016/j.sbi.2008.01.008.

Franzke, C., and A. J. Majda, 2006: Low-order stochastic mode reduction for a prototype atmospheric gcm. *Journal of the Atmospheric Sciences*, **63 (2)**, 457–479, doi:10.1175/JAS3633.1.

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

Gottwald, G. A., D. T. Crommelin, and C. L. E. Franzke, 2016: Stochastic climate theory. 1612.07474.

Han, J., A. Jentzen, and W. E, 2018: Solving high-dimensional partial differential equations using deep learning. *Proceedings of the National Academy of Sciences*, **115 (34)**, 8505–8510, doi:10.1073/pnas.1718942115, URL https://www.pnas.org/content/115/34/8505, https://www.pnas.org/content/115/34/8505.full.pdf.

Hasselmann, K., 1976: Stochastic climate models part i. theory. *Tellus*, **28 (6)**, 473–485, doi: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. *Journal of the Atmospheric Sciences*, **63 (7)**, 1924 – 1937, doi:10.1175/JAS3720.1, URL https://journals.ametsoc.org/view/journals/atsc/63/7/jas3720.1.xml.

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

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

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

Khoo, Y., J. Lu, and L. Ying, 2018: Solving for high-dimensional committor functions using artificial neural networks. *Research in the Mathematical Sciences*, **6**, doi:10.1007/s40687-018-0160-2, URL https://doi.org/10.1007/s40687-018-0160-2.

L'Ecuyer, P., V. Demers, and B. Tuffin, 2007: Rare events, splitting, and quasi-monte carlo. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, **17 (2)**, 9–es.

Legras, B., and M. Ghil, 1985: Persistent anomalies, blocking and variations in atmospheric predictability. *Journal of Atmospheric Sciences*, **42 (5)**, 433 – 471, doi:10.1175/1520-0469(1985)042<0433:PABAVI>2.0.CO;2, URL https://journals.ametsoc.org/view/journals/atsc/42/5/1520-0469_1985_042_0433_pabavi_2_0_co_2.xml.

Li, H., Y. Khoo, Y. Ren, and L. Ying, 2020: Solving for high dimensional committor functions using neural network with online approximation to derivatives. 2012.06727.

Li, Q., B. Lin, and W. Ren, 2019: Computing committor functions for the study of rare events using deep learning. *The Journal of Chemical Physics*, **151 (5)**, 054 112, doi:10.1063/1.5110439, URL https://doi.org/10.1063/1.5110439.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *Journal of Atmospheric Sciences*, **20 (2)**, 130 – 141, doi:10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2, URL https://journals.ametsoc.org/view/journals/atsc/20/2/1520-0469_1963.

Lorpaiboon, C., E. H. Thiede, R. J. Webber, J. Weare, and A. R. Dinner, 2020: Integrated variational approach to conformational dynamics: A robust strategy for identifying eigenfunctions of dynamical operators. *The Journal of Physical Chemistry B*, **124 (42)**, 9354–9364, doi:10.1021/acs.jpcb.0c06477, URL https://doi.org/10.1021/acs.jpcb.0c06477.

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.

Majda, A. J., and D. Qi, 2018: Strategies for reduced-order models for predicting the statistical responses and uncertainty quantification in complex turbulent dynamical systems. *SIAM Review*,

60 (3), 491–549, doi:10.1137/16M1104664, URL https://doi.org/10.1137/16M1104664, https://doi.org/10.1137/16M1104664.

Majda, A. J., I. Timofeyev, and E. Vanden Eijnden, 2001: A mathematical framework for stochastic climate models. *Communications on Pure and Applied Mathematics*, **54 (8)**, 891–974, doi:https://doi.org/10.1002/cpa.1014, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.1014, https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpa.1014.

Mardt, A., L. Pasuali, H. Wu, and F. Noé, 2018: Vampnets for deep learning of molecular kinetics. *Nature Communications*, **9**, doi:10.1038/s41467-017-02388-1, URL https://doi.org/10.1038/s41467-017-02388-1.

Matsuno, T., 1971: A Dynamical Model of the Stratospheric Sudden Warming. *Journal of the Atmospheric Sciences*, **28 (8)**, 1479–1494, doi:10.1175/1520-0469(1971)028<1479:ADMOTS>2.0.CO;2, URL https://doi.org/10.1175/1520-0469(1971)028<1479:ADMOTS>2.0.CO;2, https://journals.ametsoc.org/jas/article-pdf/28/8/1479/3417422/1520-0469(1971)028_1479_admots_2_0_co_2.pdf.

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

Metzner, P., C. Schutte, and E. Vanden-Eijnden, 2009: Transition path theory for markov jump processes. *Multiscale Modeling and Simulation*, **7 (3)**, 1192–1219, doi:10.1137/070699500.

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)**, 11 138–11 143, doi:10.1073/pnas.1813263115, URL https://www.pnas.org/content/115/44/11138, https://www.pnas.org/content/115/44/11138.full.pdf.

Ngwira, C. M., and Coauthors, 2013: Simulation of the 23 july 2012 extreme space weather event: What if this extremely rare cme was earth directed? *Space Weather*, **11 (12)**, 671–679, doi:https://doi.org/10.1002/2013SW000990, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2013SW000990, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2013SW000990.

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

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

Pande, V. S., K. Beauchamp, and G. R. Bowman, 2010: Everything you wanted to know about markov state models but were afraid to ask. *Methods*, **52 (1)**, 99–105, URL https://doi.org/10.1016/j.ymeth.2010.06.002.

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

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

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

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

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

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 Dynamics*, **49 (9)**, 3031–3050, doi:10.1007/s00382-016-3491-y.

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.

Schaller, N., J. Sillmann, J. Anstey, E. M. Fischer, C. M. Grams, and S. Russo, 2018: Influence of blocking on northern european and western russian heatwaves in large climate model ensembles. *Environmental Research Letters*, **13 (5)**, 054 015, doi:10.1088/1748-9326/aaba55, URL https://doi.org/10.1088%2F1748-9326%2Faaba55.

Sillmann, J., and Coauthors, 2017: Understanding, modeling and predicting weather and climate extremes: Challenges and opportunities. *Weather and Climate Extremes*, **18**, 65 – 74, doi: https://doi.org/10.1016/j.wace.2017.10.003.

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.

Strahan, J., A. Antoszewski, C. Lorpaiboon, B. P. Vani, J. Weare, and A. R. Dinner, 2020: Long-timescale predictions from short-trajectory data: A benchmark analysis of the trp-cage miniprotein. 2009.04034.

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.

Tibshirani, R., 1996: Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, **58 (1)**, 267–288, doi:https://doi.org/10.1111/j.2517-6161.1996.tb02080.x, URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1996.tb02080.x, https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.2517-6161.1996.tb02080.x.

Timmermann, A., F.-F. Jin, and J. Abshagen, 2003: A nonlinear theory for el niño bursting. *Journal of the Atmospheric Sciences*, **60 (1)**, 152 – 165, doi:10.1175/1520-0469(2003)060<0152:

ANTFEN>2.0.CO;2, URL https://journals.ametsoc.org/view/journals/atsc/60/1/1520-0469_2003.

Vanden-Eijnden, E., and W. E, 2010: Transition-path theory and path-finding algorithms for the study of rare events. *Annual Review of Physical Chemistry*, **61 (1)**, 391–420, doi:10.1146/annurev.physchem.040808.090412.

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**, URL https://doi.org/10.1038/s41612-018-0013-0.

Wan, Z. Y., P. Vlachas, P. Koumoutsakos, and T. Sapsis, 2018: Data-assisted reduced-order modeling of extreme events in complex dynamical systems. *PLOS ONE*, **13 (5)**, 1–22, doi:10.1371/journal.pone.0197704, URL https://doi.org/10.1371/journal.pone.0197704.

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.

Weinan, E., T. Li, and E. Vanden-Eijnden, 2019: *Applied stochastic analysis*, Vol. 199. American Mathematical Soc.

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

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

Zhang, F., and J. A. Sippel, 2009: Effects of moist convection on hurricane predictability. *Journal of the Atmospheric Sciences*, **66 (7)**, 1944 – 1961, doi:10.1175/2009JAS2824.1, URL https://journals.ametsoc.org/view/journals/atsc/66/7/2009jas2824.1.xml.

Zwanzig, R., 2001: *Nonequilibrium statistical mechanics*. Oxford University Press.

LIST OF TABLES

Region	Fraction (DGA)	Fraction (empirical)
Inside A	0.24	0.27
$(A \cup B)^c \to A$	0.27	0.28
Inside B	0.14	0.13
$(A \cup B)^c \to B$	0.36	0.32

TABLE 1. Fraction of time that the system spends (i) inside A, (ii) outside A but destined for A, (iii) inside B, (iv) outside B but destined for B. Each fraction is computed directly from DGA and empirically verified from the long simulation.

LIST OF FIGURES

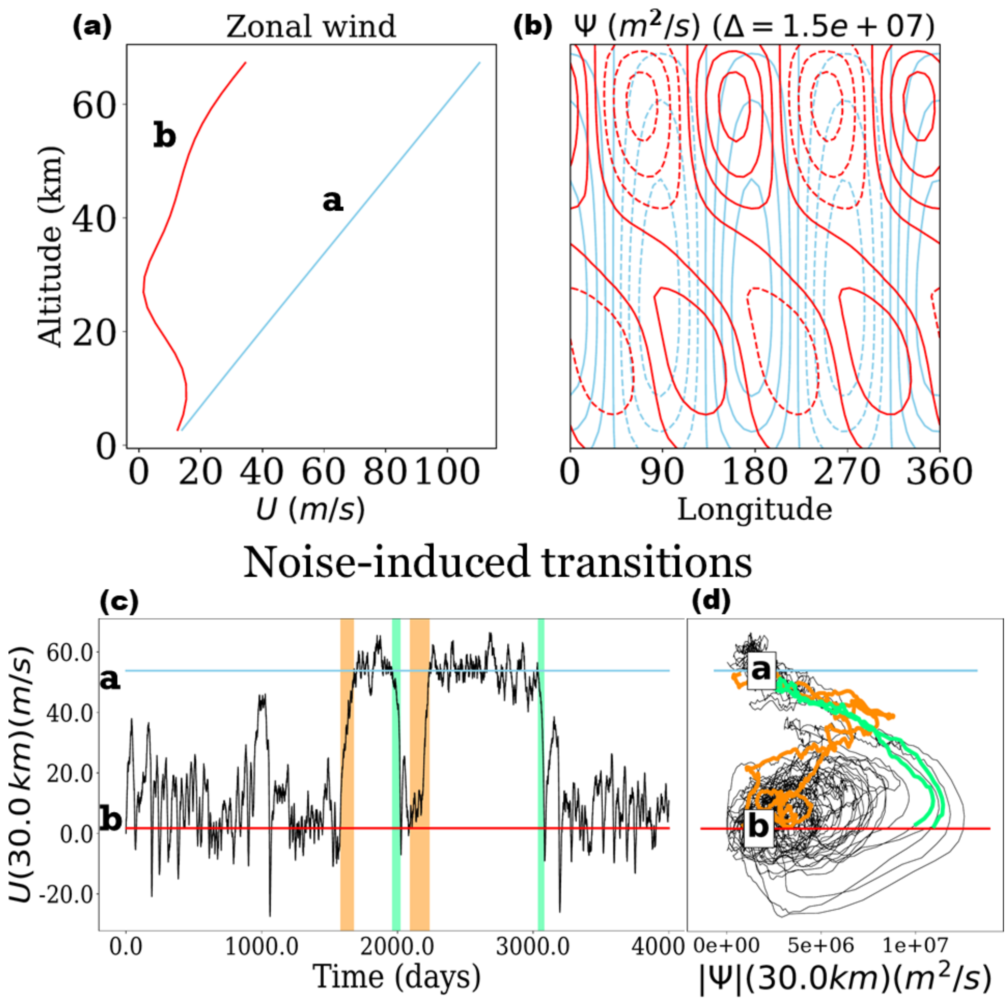


FIG. 1. **Illustration of the two stable states of the Holton-Mass model and transitions between them.** (a) Zonal wind profiles of the radiatively maintained strong vortex (the fixed point **a**, blue) which increases linearly with altitude, and the weak vortex (the fixed point **b**, red) which dips close to zero in the mid-stratosphere. (b) Streamfunction contours are overlaid for the two equilibria **a** and **b**, the weak vortex exhibiting strong westward phase tilt with altitude. (c) Timeseries of $U(30\,km)$ from a long stochastic simulation, including several noise-induced transitions from A to B (green) and from B to A (orange). Although both states **a** and **b** are equilibria in this parameter regime ($h = 38.5\,m$), the stochastic perturbations uncover the vacillation cycles that would appear beyond the Hopf bifurcation if h were increased. (d) A parametric curve of the same trajectory segment as in (c) with the same color highlights for transition paths, but in the space $(|\Psi|, U)$ at 30 km. The two equilibria are indicated with horizontal blue and red lines.

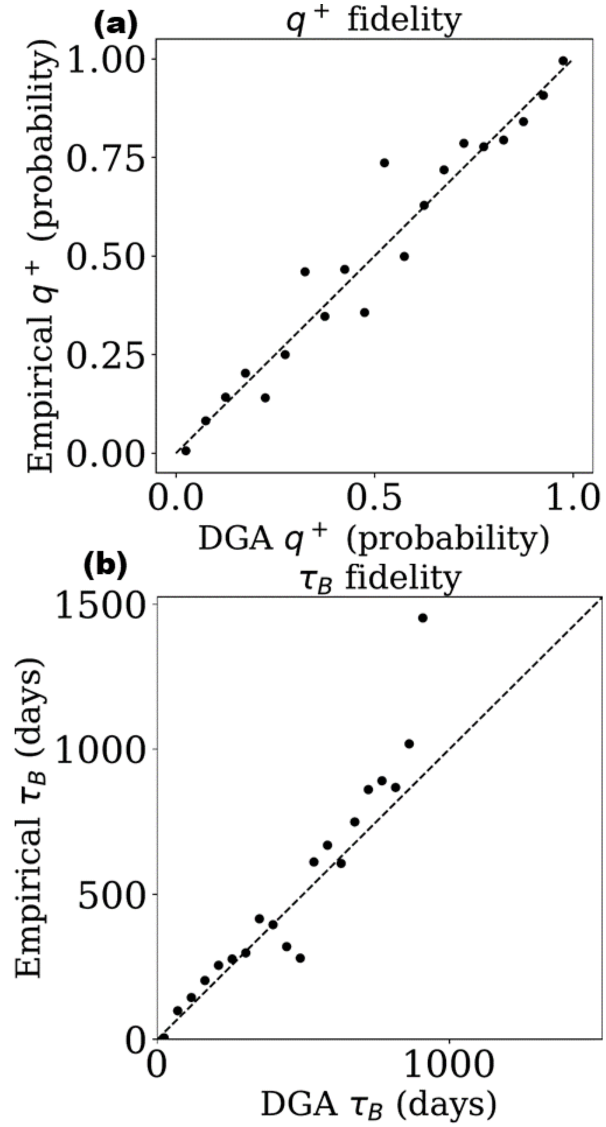
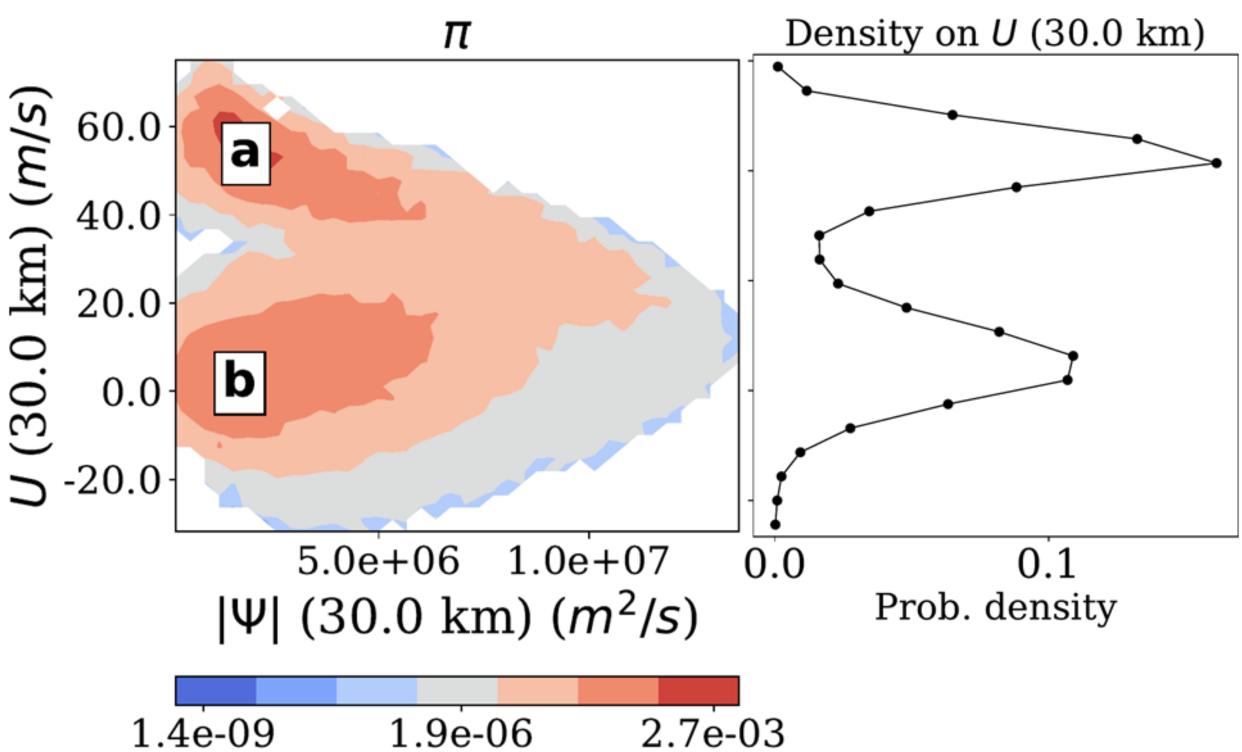


FIG. 2. **Accuracy of the committor and mean first passage time calculations verified with long trajectory data.** The DGA calculations assign approximate committor and mean first passage time values $q^+(\mathbf{X}_n(0))$ and $m_B(\mathbf{X}_n(0))$ to each data point. Because each snapshot x_n was collected from a long trajectory, its destination and the time to get there are known and can provide an empirical validation of the committor and lead time. For 20 equal partitions (ζ_1, ζ_2) of the interval $(0,1)$, we assemble all trajectory starts $\mathbf{X}_n(0)$ with $q^+(\mathbf{X}_n(0)) \in (\zeta_1, \zeta_2)$ and count the fraction heading toward B. These are the empirical committors for the interval (ζ_1, ζ_2), and are plotted on the vertical axis against $(\zeta_1 + \zeta_2)/2$. Similarly, we bin the space of calculated first passage times and for each bin average the empirical first passage time to B. Both quantities line up well between DGA computations and empirical values, with the exception of the longest passage times, which are underestimated.



1139 FIG. 3. **Steady-state distribution.** The density $\pi(\mathbf{x})$ is projected onto the two-dimensional space $(|\Psi|, U)$ at
1140 30 km, on a log scale. The density is peaked in the neighborhoods of the two fixed points. On the right is a
1141 projection of π onto the single variable U(30 km), on a linear scale, confirming strong bimodality.

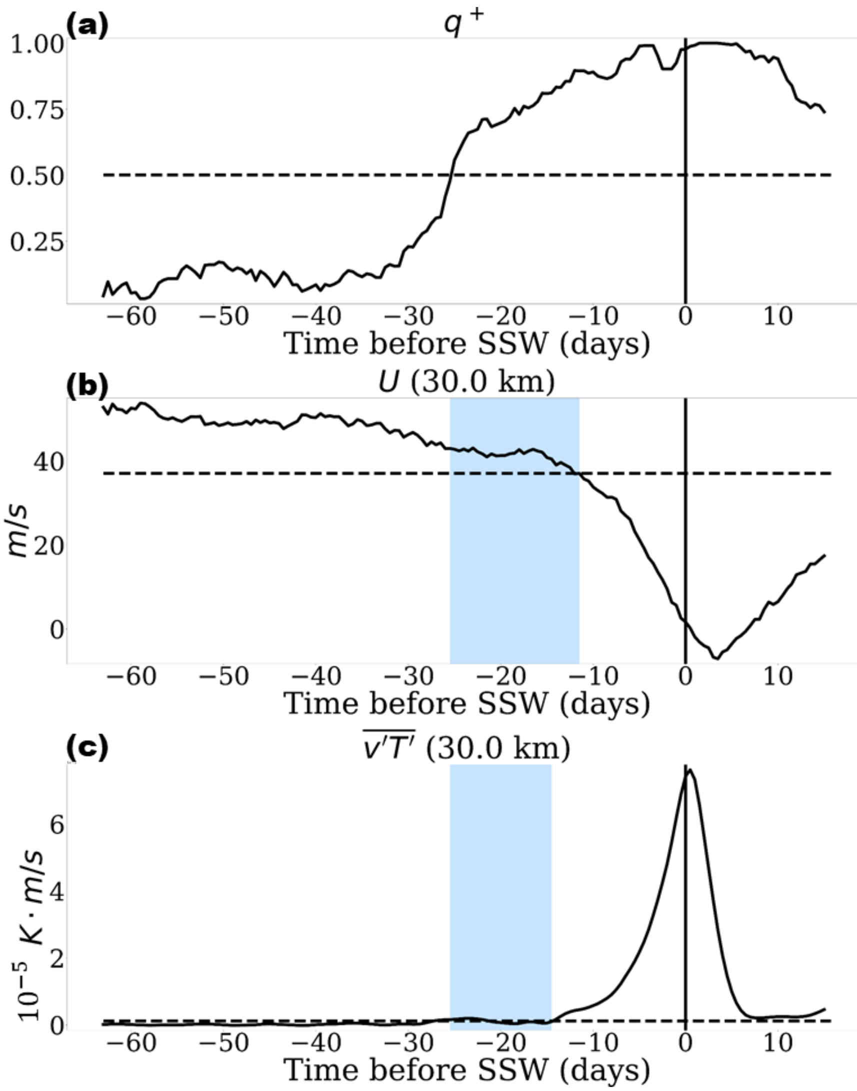


Fig. 4. **The committor vs. other observables as a forecasting tool.** A representative simulated SSW event from the long simulation is plotted over time, starting 65 days in advance of the official event when $U(30\,km)$ first drops below 1.75 m/s, which is marked by a vertical solid line. Panel (a) shows the committor over time following the trajectory, panel (b) shows the zonal wind $U(30\,km)$, and panel (c) shows the eddy heat flux $\overline{v'T'}(30\,km)$. Horizontal dashed lines mark the natural forecasting threshold of $q^+ = 0.5$ (panel (a)) or the value of the observable most closely associated with $q^+ = 0.5$: $37\,m/s$ (panel (b)) and $1.2 \times 10^{-6}\,K \cdot m/s$ (panel (c)). The sharp increase in q^+ as it crosses the threshold provides a clear and early warning sign of oncoming SSW, about 26 days in advance. U and $\overline{v'T'}$ are moving slowly at that time, and don't clear their respective thresholds for the last time until the event is much closer at hand. The gap in lead time is marked by blue strips.

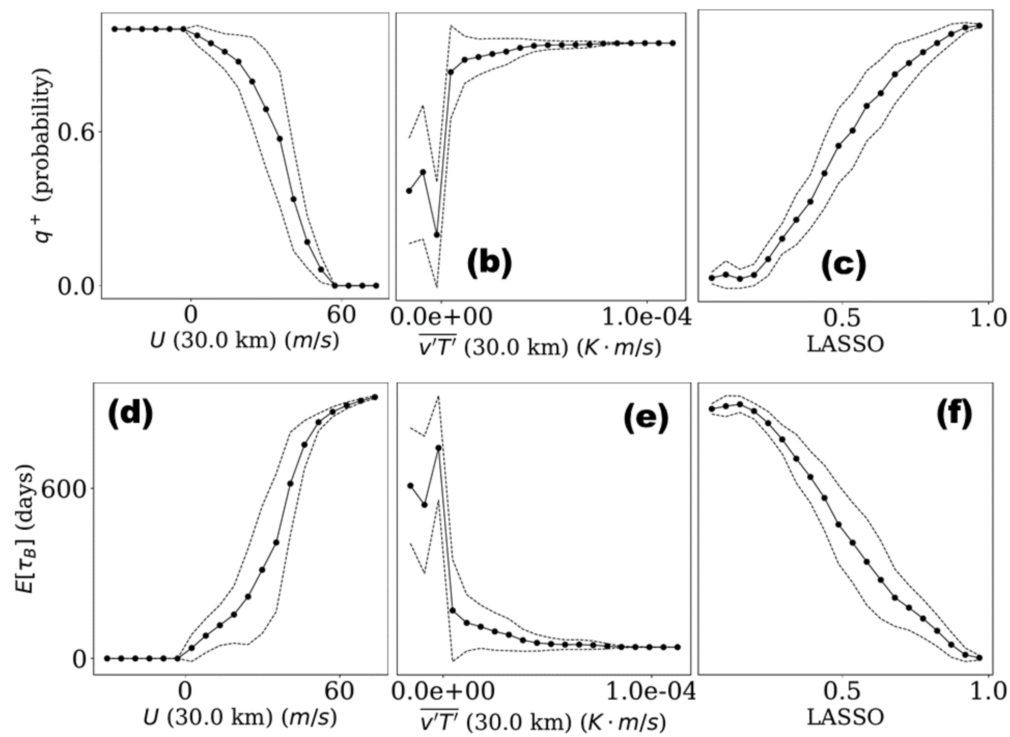


FIG. 5. **One-dimensional projections of the forward committor and mean first passage time to B, computed with DGA**. These functions depend on all $d = 75$ degrees of freedom in the model, but we have averaged across $d - 1 = 74$ dimensions to visualize the committor (first row) and mean first passage time to B (second row) as rough functions of three single degrees of freedom: $U(30\,km)$ (first column), $\overline{v'T'}(30\,km)$ (second column), and the LASSO-regressed committor (third column). The forward committor measures proximity to B in probability, while mean passage time to B measures proximity in time, hence the negative correlation between the two quantities. The general trends reveal fairly obvious relationships: stronger wind is associated with tendency towards the strong-vortex state A, and larger poleward eddy heat flux is associated with tendency toward the weak vortex state B. In addition, curves like this assess the quality of single-variable observables as proxies for an oncoming transition event. The committor and passage time vary smoothly and (mostly) monotonically with U, but discontinuously with $\overline{v'T'}$: the heat flux burst that accompanies a SSW gives no advance warning for the event, while a small negative change in U indicates incrementally higher transition probability and shorter lead time.

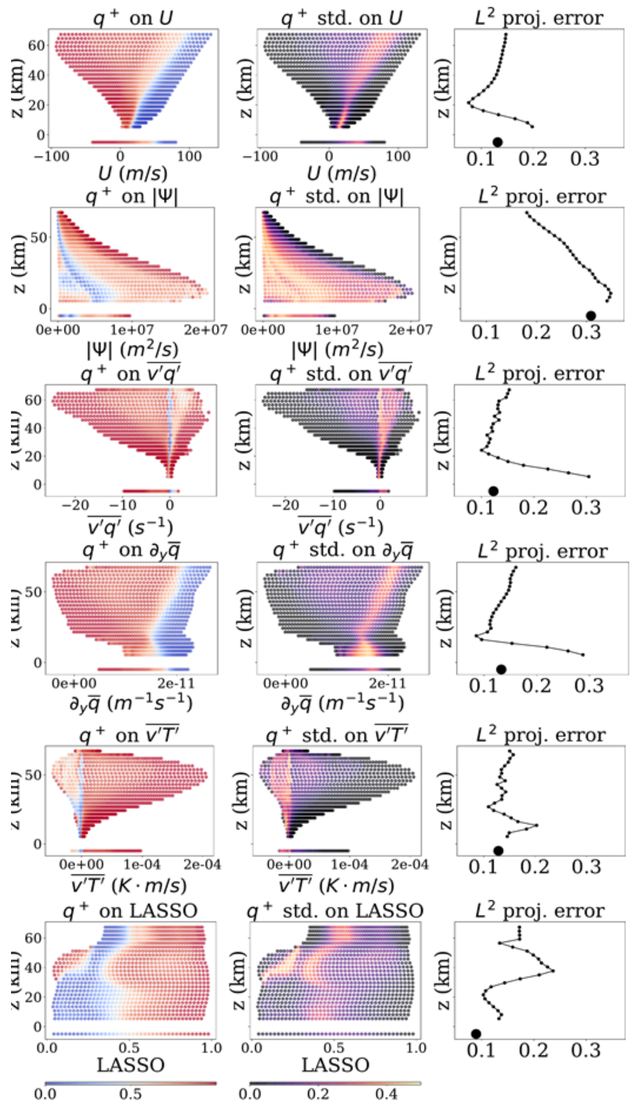
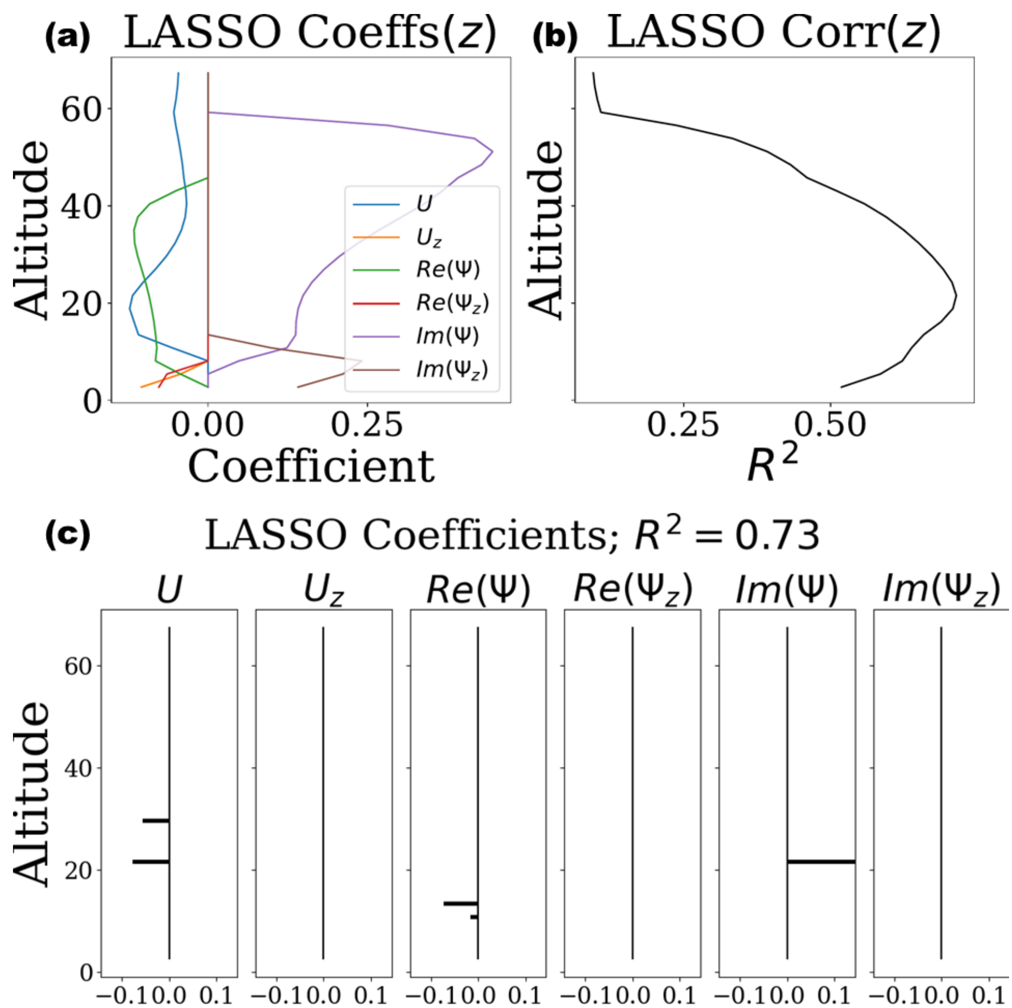


FIG. 6. **Projection of the forward committor onto a large collection of one-dimensional CVs**, along with the associated standard deviation, or projection error, of the committor along the remaining 74 model dimensions. Consider the first two panels. The left-hand panel shows, for each discretized altitude z, a heatmap of the committtor as $U(z)$ ranges from its minimum to its maximum realized strength at that altitude. At the bottom is an additional heatmap of the committor as a function of $\langle U\rangle_z$, the vertical average. These are conditional expectations, with the corresponding conditional standard deviations displayed in the right-hand panels. The following rows display analogous plots for wave magnitude, eddy PV flux, background PV gradient, eddy heat flux, and the LASSO predictor specified in Figure 7. The bottom line on the last plot is not a vertical average, but the results of regression on all altitudes at once.



1173 Fig. 7. **Results of LASSO regression of the forward committor with $U, \mathrm{Re}\Psi, \mathrm{Im}\Psi$ as input features**. Panel (a) shows the coefficients when q^+ is regressed as a function of only the variables at a given altitude, and panel (b) shows the corresponding correlation score. 21.5 km seems the most predictive (where $z \equiv 0$ at the tropopause, not the surface). Panel (c) shows the coefficient structure when all altitudes are considered simultaneously. By design, most of the coefficients are zero, but most of the nonzero coefficients appear at 21.5 km, once again distinguishing that level as highly relevant for prediction.

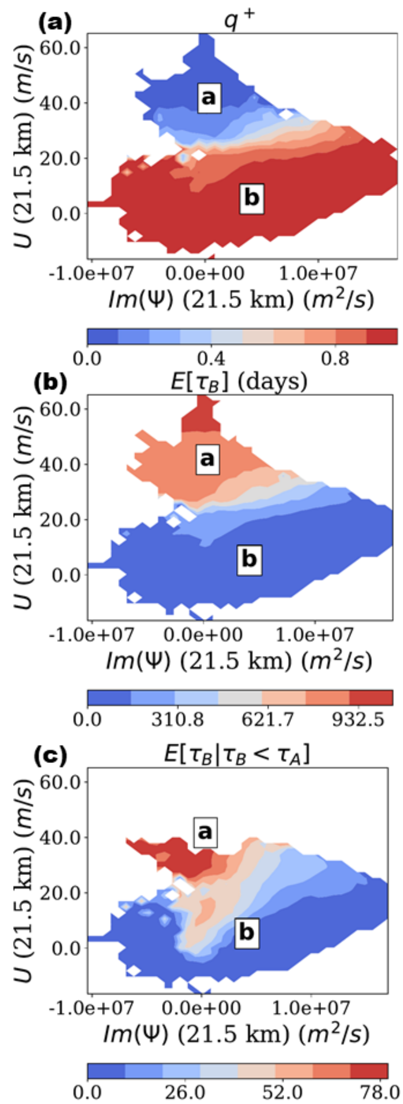


FIG. 8. **Two-dimensional projections of the committor and mean first passage times.** We have projected three quantities onto the observable subspace of zonal wind and imaginary part of streamfunction at 21.5 km. (a) forward committor $q^+(x) = \mathbb{P}_x\{\tau_B < \tau_A\}$, (b) first passage time to B $\mathbb{E}[\tau_B]$, and (c) conditional mean first passage time to B $\mathbb{E}[\tau_B|\tau_B < \tau_A]$. The condition $\tau_B < \tau_A$ decreases the passage time by an order of magnitude, because it excludes the possibility of getting trapped in A first. Figure 9 quantifies the relationship between the committor and conditional passage time, and its forecasting implications.

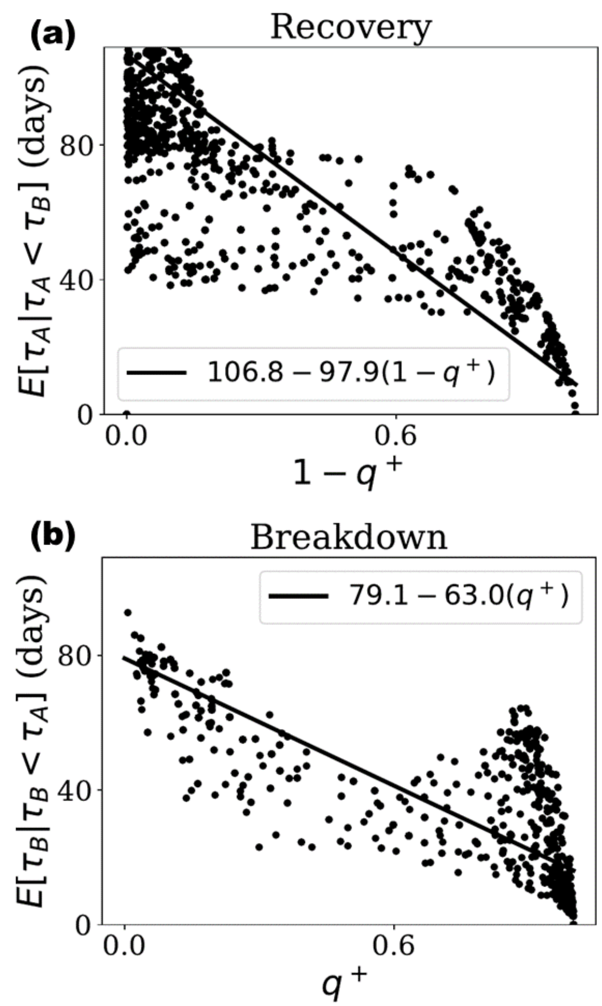


FIG. 9. **Relationship between committor and mean first passage time.** Panel (a) shows the relationship between q^+ (probability of next hitting B) and $\mathbb{E}[\tau_B|\tau_B < \tau_A]$, the time until hitting B conditional on avoiding A. These quantities correspond to panels (a) and (c) of Fiure 8. Panel (b) shows the same relationship but in the $B \to A$ direction. In both cases, we performed a least squares regression weighted by the change of measure. A +0.1 increase in the probability q^+ of next hitting B comes with a 6.3-day decrease in the expected time to get there, whereas a +0.1 increase in the opposite probability $1-q^+$ comes with a 9.8-day reduction in the time to reach A. Meanwhile, the vertical intercepts indicate the mean time of a full transition from $A \to B$ (79 days) and $B \to A$ (170 days).