THE UNIVERSITY OF CHICAGO

ATMOSPHERIC EXTREMES THROUGH THE LENS OF TRANSITION PATH THEORY

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

COMMITTEE ON COMPUTATIONAL AND APPLIED MATHEMATICS

BY
JUSTIN FINKEL

CHICAGO, ILLINOIS
AUGUST 2022

To my grandparents, whose intellect and integrity reverberate through generations.

I took month-long vacations in the stratosphere,

and you know it's really hard to hold your breath.

—Bruce Springsteen, "Growin' Up"

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

It takes a village to produce a Ph.D., and I am no exception. In advising my projects directly, I owe tremendous thanks to my gaggle of advisors, who span multiple institutions: Jonathan Weare, Dorian Abbot, Mary Silber, Edwin Gerber, and Aaron Dinner. They sacrificed many hours of brainstorming, patiently listening, proofreading, and advocating for me to make my experience positive and productive. Their contributions are incomparable, but I must especially thank Ed for joining a project with a student who drifted into NYU from another institution and pitched some very technical and esoteric ideas. He went above and beyond to study and understand my work, and was instrumental in helping me bridge the gap between shiny mathematical tools and concrete science.

I wish to also acknowledge Noboru Nakamura and Andrew Charlton-Perez, for the fortuitous conversations with both of them that sparked entire threads of research within my Ph.D.

My colleagues and collaborators helped me through many research hurdles, from technical to "soft." First and foremost, Robert Webber was an endlessly generous, insightful, and supportive academic older brother. Always willing to proofread and collaborate, he provided an important practical source of feedback and ideas. I've learned a lot about mathematics and research from him. Erik Thiede, as well, was a fundamental contributor, having invented the very algorithm which I spent my Ph.D. applying to a new system. During our relatively brief overlap at UChicago, he kickstarted my technical development. Continuing on, John Strahan and Chatipat Lorpaiboon continued to push forward Erik's ideas, and I benefitted substantially from the best practices they developed.

The Department of Energy Computational Science Graduate Fellowship (DOE CSGF) played a key role in my graduate career as well, providing the funding and flexibility to move between departments and institutions as I did. I gained some valuable perspective and professional contacts in the CSGF community and the practicum.

More importantly, I must thank the CSGF for connecting me with Anya Katsevich. Our relationship, and my move to New York City, marked a decisive positive turning point in my Ph.D. and

indeed my life. I am unbelievably fortunate to have her companionship, in mathematics as well as other areas of life.

Together with Anya, my parents and sister are my greatest supporters. My accomplishments are their accomplishments. They inspire me to make the world better in big and small ways. I cannot thank them enough for believing in me even when I doubt myself. It will take a lifetime to repay this gift like one solves the Kolmogorov equations: both backward and forward.

ABSTRACT

Extreme weather events have large consequences, dominating the impact of climate on society, but are very difficult to characterize and predict, being exceptionally rare and pathological outliers in the spectrum of weather events. A rare event with a 100-year return period takes, on average, 100 years of simulation time to appear just once, let alone a statistically significant number of times. One can collect more statistics by running models at reduced resolution, but this comes at the cost of bias. High-fidelity models are needed to resolve the relevant dynamics.

Furthermore, even if we had abundant data on extreme events, they make up a complex and diverse ensemble that is difficult to describe. Extremes come in different shapes, sizes, and magnitudes. Precursors and first causes are highly sought after for forecasting, but untangling these from background weather variability can raise thorny statistical issues.

This thesis addresses both questions, by advancing two ideas: (1) Transition path theory, or TPT, as a mathematical framework to describe the statistical ensemble of rare events of a certain type; and (2) dynamical Galerkin approximation, or DGA, as a computational method to compute those important quantities. Both ideas emerged from the molecular dynamics community, and, I believe, have considerable potential for use by the climate modeling community. I demonstrate these ideas by way of example, on a hierarchy of models of one particular atmospheric phenomenon: sudden stratospheric warming (SSW), a rapid, large-scale disturbance in the stratosphere. SSW is an archetype of a complicated, extreme event that develops suddenly, often defying long-term prediction, and with disputed mechanisms. The TPT lens reveals some interesting features of SSW as a statistical ensemble, including its precursors, rate, and seasonal distribution. The hope is that this new tool will inspire fruitful investigation of many other kinds of atmospheric extremes.

1 INTRODUCTION

The atmosphere's extreme, irregular behavior is, in some ways, more important to characterize than its typical climatology. Our society is optimized for historical weather patterns, and therefore highly exposed to damage from extreme heat and cold, flooding, and other natural hazards. From a human perspective, weather is inconsequential when it follows mean behavior; it is the anomalies that challenge society [Lesk et al., 2016, Kron et al., 2019].

Extremes may respond more sensitively to climate change than does mean behavior, an argument supported by elementary statistics [Wigley, 2009], empirical climate observations [Coumou and Rahmstorf, 2012, AghaKouchak et al., 2014, O'Gorman, 2012, Huntingford et al., 2014, Naveau et al., 2020] and climate simulations [Pfahl et al., 2017, Myhre et al., 2019]. The overall "climate sensitivity" [Hansen et al., 1984], summarized by a change in global-mean temperature, does not do justice to these extreme-weather consequences. Unprecedented extreme weather events in the past decade hint at the range of possibilities [Mishra and Shah, 2018, Van Oldenborgh et al., 2017, Goss et al., 2020, Fischer et al., 2021]. Extreme weather is taking an increasing toll on ecosystems, economies, and human life, due to both a changing climate and increasing reliance on weather-susceptible infrastructure [e.g., Mann et al., 2017, Frame et al., 2020]. Rare events have attracted significant simulation efforts recently, especially hurricanes [e.g., Zhang and Sippel, 2009, Webber et al., 2019, Plotkin et al., 2019], heat waves [e.g., Ragone et al., 2018], rogue waves [e.g., Dematteis et al., 2018], and space weather events [e.g., coronal mass ejections; Ngwira et al., 2013].

The intermittency of extreme events makes precise risk assessment exceedingly difficult. 100 flips of a loaded coin with $\mathbb{P}\{\text{Heads}\} = 0.01$ is almost as likely to yield zero heads (probability 0.366) as one head (probability 0.370), and half as likely to yield two heads (probability 0.185). Similarly, in a 100-year climate simulation or historical record, a once-per-century event may appear either non-existent or twice as likely as it really is, with more than 50% probability. The difficulty exists even in a stationary climate, but worsens in the presence of time-dependent forc-

ing (such as CO_2). The limited historical record forces us to use numerical models as approximations, introducing a dilemma: we can run cheaper, coarse-resolution models for long integrations, providing reliable statistics of a biased system, or expensive, high-resolution models for short integrations, sacrificing our ability to properly sample the system for a reduction in bias. Coarse models produce larger sample sizes more efficiently, which is why long-term climate simulations are usually performed with a low resolution of $O(50-100)$ km per grid cell [Haarsma et al., 2016]. A coarse model might suffice to estimate global-mean temperature and other aggregated statistics, but cannot resolve convective systems, e.g., tropical cyclones and precipitation over complex topography, that deliver localized but heavy damage [O'Brien et al., 2016, He et al., 2019]. Even extremes that manifest at a large scale, such as a sudden stratospheric warming (SSW, the topic of this thesis) might arise out of multi-scale interactions that are poorly represented in coarse model grids.

To take full advantage of the computing power available, we must develop new approaches to efficiently manage and parse the data we generate (or observe) to derive physically interpretable, actionable insights. Ensemble forecasting, the traditional operating procedure of numerical weather prediction, is best suited to estimate statistics of the average or most likely scenarios, and specialized methods are needed to examine the more extreme outlier scenarios.

One candidate framework to address rare events is transition path theory (TPT). This thesis documents my attempts to harness TPT as a tool to provide insights into a dramatic atmospheric phenomenon, sudden stratospheric warming (SSW), which is responsible for major re-organization of wintertime circulation and extreme weather regimes. This goal was open-ended, with terms such as "harness" and "insights" open for interpretation. TPT is an elegant set of ideas and relationships whose stated mission is to describe rare events in dynamical systems, and SSW fits the bill. However, some significant translation is needed to forge this new link between disciplines. TPT is couched in the language of statistical mechanics and was largely developed in the context of molecular dynamics, a domain radically different from atmospheric science in spatiotemporal

scale, computational practices, and most importantly, quantities of interest. Chemical reactions are driven by electrostatic forces ranging from nanometers to micrometers, while atmospheric storms are driven by waves, currents, and phase changes that synchronize across tens to thousands of kilometers. However, the two disciplines share a common computational challenge of *timescale separation*: in computer simulations of both molecular and atmospheric dynamics, the timestep is exceptionally short compared to the timescale of the important events. Molecular simulation proceeds femtoseconds at a time, while full reactions take milliseconds. Atmospheric simulations proceed for minutes or hours at a time, while it might take weeks or years to see a high-impact storm. A once-per-century event, for example, would appear only once (on average) in a 100-year climate simulation. Yet one realization is not enough: a complex, turbulent system like the Earth's climate is best conceptualized as a *random* process, which demands a statistically significant sample for description. What TPT brings to the table is a bridge between the disparate timescales, in the form of relations between long-term probabilities (rates) and the short-term evolution of the system. In other words, TPT links the *initial-value problem*—how a system evolves from a given initial condition—to the *boundary-value problem* of the system's steady-state: how it explores and fills the phase space available to it over long timescales. In atmospheric dynamics, these two problems go by the names "weather" and "climate", a perspective advanced by Bryson [1997].

TPT is a relative newcomer to the world of climate dynamics, competing with existing frameworks such as extreme value theory [e.g., Lucarini et al., 2016] and large deviation theory [Gálfi et al., 2019, Gálfi et al., 2021] as the tool of choice to describe rare events. Extreme value theory describes the long-term probability of some observable, like surface temperature, exceeding a range of thresholds, with a universal family of probability densities. These probabilities are also called *rates*, which are inversely related to *return times* (which we define precisely in the following background section). Large deviation theory, on the other hand, describes the specific mechanism of an extreme event as a path winding through phase space. The two approaches are complementary, but fail to address two important issues. First, they are asymptotic theories: EVT only holds

at the very upper tail of the probability distribution, and LDT only holds in the low-noise limit, with vanishingly small random perturbations and hence very low rates. Second, they only grasp small pieces of the full statistical ensemble of extreme events: LDT recovers a single most-likely path, whereas EVT recovers a distribution over endpoints of that path. TPT fills in these gaps, providing probability distributions over pathways at finite noise and for very flexible extreme event definitions. Of course, this more ambitious goal has higher demands in terms of data. A major goal in this thesis is to demonstrate ways to meet that demand by allocating computational resources efficiently.

This work explores the utility, and some drawbacks, of TPT in climate applications by way of example. Chapters 2, 3, and 4 progress upward on a hierarchy of SSW models, from a low-order conceptual equation to a state-of-the-art weather forecasting system. The emphasis and language evolve along the way, reflecting my changing understanding of the needs and interests of the climate science community. Before jumping into the research projects, the following two subsections cover the essential background of (i) the mathematical framework, transition path theory, and (ii) the scientific application, sudden stratospheric warming.

There remain many chapters yet to write in the co-evolution of computational non-equilibrium statistical mechanics with climate science. I hope this work serves as a useful starting point.

2 BACKGROUND: TRANSITION PATH THEORY

Here we give an overview of transition path theory (TPT). Consider some stochastic process $\mathbf{X}(t)$ evolving through some state space Ω as a function of time t. In the forthcoming examples, Ω will generally be a Euclidean space $\mathbb{R}^d$ or a finite, discrete space $\{1,\ldots,M\}$, although more general state spaces are possible such as manifolds and product spaces. Similarly, time t might be continuous or discrete.

TPT conceives of a rare event as a trajectory of the system from some "normal" set $A \subset \Omega$ to some "abnormal" set $B \subset \Omega$, without touching either set in between: $\{(t,\mathbf{X}(t)) : \tau_A^- < t < \tau_B^+\}$, where τ_A^- is the time of departure from A and τ_B^+ is the time of arrival to B. A and B are user-defined, which makes TPT flexible. In chemistry, A and B usually refer to reactants and products, whereas in the applications to SSW to follow, we define A and B as strong and weak circulation regimes of the stratospheric polar vortex.

Given a process $\mathbf{X}(t)$ and a pair of states (A,B), TPT describes aspects of the system's statistical behavior, such as:

- How often do transitions happen? In other words, what is the rate?
- Given an instantaneous snapshot of the system, how "close" is it to the abnormal state B vs. the normal state A? As I explain below, TPT formulates "closeness" in terms of either time or probability.
- How much time does $\mathbf{X}$ spend in or near A, in or near B, and en route from A to B?
- During the $A \to B$ transition process, what route does the system take? Is there one preferred route, a few different preferred routes, or a continuum?

All of these questions could, in principle, be answered by integrating computer models for as long as needed to collect statistically significant samples. However, the insight of TPT is to express the answers in terms of a few essential functions over state space, including the following.

1. The **probability density**, π, measures the probability of the system being located near $\mathbf{x}$ at a given time.

$$\pi(\mathbf{x})\, d\mathbf{x} = \mathbb{P}\{\mathbf{X}(t) \in d\mathbf{x}\} \tag{2.1}$$

where $d\mathbf{x}$ is an infinitesimal region in space near $\mathbf{x}$, and also stands for the volume of that region.

2. The **first hitting time** τ_B^+, a random variable, is the time until the system reaches state B after some starting time t_0:

$$\tau_B^+(t_0) = \inf\{t \geq t_0 : \mathbf{X}(t) \in B\} \tag{2.2}$$

Similarly, the first-hitting time to A is τ_A^+, and the first-hitting time to their union is $\tau_{A\cup B}^+ = \min(\tau_A^+, \tau_B^+)$, etc. We will be interested in *expectations* of these times, in various combinations and with various initial conditions.

3. The **forward committor**, q_B^+, measures the progress of the transition path. It is defined as the probability of next reaching the abnormal state B rather than the normal state A, from a given initial condition:

$$q_B^+(\mathbf{x}) = \mathbb{P}_{\mathbf{x}}\{\mathbf{X} \text{ next reaches } B \text{ rather than } A\} = \mathbb{P}_{\mathbf{x}}\{\tau_B^+ < \tau_A^+\} \tag{2.3}$$

where the subscript $\mathbf{x}$ means "conditional on starting the system at $\mathbf{x}$".

The superscripts $(+)$ emphasize that the expectations look forward into the future. Below, in section 2.3, we label backward-in-time expectations with a superscript $(-)$. These definitions are deliberately vague, glossing over details that depend on the application at hand. The key takeaway is that global statistics, such as the rate, can be assembled from local quantities over state space such as committors and probability densities.

With this essential preamble, the application-oriented reader can skip to the SSW background or to chapter 4. Below I provide a more complete account of TPT in several steps that roughly parallel the sequence of applications in chapters 4-6. Section 2.1 defines the transition path ensemble. Section 2.2 presents necessary background on diffusion processes, which are the subject of most existing TPT analyses, including chapters 4-6. Section 2.3 defines the specific quantities of interest for TPT analysis, such as committors, and writes down equations (the Kolmogorov equations) for them. These equations are solved with a classical finite-volume method in chapter 4 to analyze a one-layer stratospheric model. Section 2.4 presents an alternative data-driven numerical method, dynamical Galerkin approximation (DGA), to deal with high-dimensional systems. DGA was invented in Thiede et al. [2019] and further developed in Strahan et al. [2021], and is used in chapters 5 and 6 to analyze a vertically stratified (but still idealized) stratospheric model. The messier application with an operational forecasting model in chapter 7 uses a different version of TPT that is fully discrete and time-dependent, but those methodological details are postponed to that chapter. In my opinion, this background chapter should be used as a reference for the application chapters, which are friendlier. There is some redundancy between the material in this introductory chapter and the application chapters, in order to serve the reader who wants to see applications without getting overwhelmed in background material first.

2.1 Transition path ensemble

The following theoretical development parallels Vanden-Eijnden [2006], but expands on it in several ways. Consider a stochastic ergodic dynamical system $\mathbf{X}(t) \in \Omega$, evolving through a very long time interval $(-T, T)$. As $T \to \infty$, $\mathbf{X}(t)$ will explore the full dynamical state space available to it, eventually filling out a *steady-state probability density* $\pi(\mathbf{x})$. It will also cross from A to B and back a number M_T of times. As $T \to \infty$, ergodicity guarantees that $M_T \to \infty$ as well. Let $\tau^-_{A,m}$ and $\tau^+_{B,m}$ denote the beginning and ending time of the mth transition path (so $\mathbf{X}(\tau^-_{A,m}) \in A$ and $\mathbf{X}(\tau^+_{B,m}) \in B$). Technically, we assume $\mathbf{X}(t)$ is right-continuous with left limits, meaning

$\mathbf{X}(\tau^-_{A,m}) \notin A$ but $\lim_{t\uparrow\tau^-_{A,m}} \mathbf{X}(t) \in A$. We won't concern ourselves with such details.

We will describe the transition path ensemble at two levels of granularity. At the first level, we consider the set of *reactive snapshots*, which are the instantaneous model states $\mathbf{X}(t)$ realized in the course of a transition without regard to their ordering in time or their grouping into separate transition events:

$$\text{Reactive snapshots} = \bigcup_{m=-\infty}^{\infty} \bigcup_{t=\tau^-_{A,m}}^{\tau^+_{B,m}} \{\mathbf{X}(t)\}. \tag{2.4}$$

We use "reactive" for consistency with the chemistry literature. At the second level, we distinguish each transition path as a unique, coherent object, containing a sequence of snapshots ordered in time. We formally define the $(A \to B)$ transition path ensemble as

$$\text{Transition path ensemble} = \left\{ \{(t, \mathbf{X}(t)), \tau^-_m < t < \tau^+_m\}, m = \ldots, -2, -1, 0, 1, 2, \ldots \right\} \tag{2.5}$$

The inner set is the collection of snapshots along the mth transition path, where the index m is assigned arbitrarily so that $\tau_{A,-1} \leq 0$ and $\tau_{A,0} > 0$. The outer set is the collection of paths, which becomes infinite as $T \to \infty$. There is no fixed duration of transition paths; each one has a different duration $\tau^+_{m,B} - \tau^-_{m,A}$. For this reason, the space of paths has infinite dimension, and there is no probability density to describe the path ensemble. However, *functionals* of transition paths do have well-defined distributions. Using the abbreviation $\mathbf{X}^{(m)} := \{(t, \mathbf{X}(t)) : \tau^-_{m,A} \leq t < \tau^+_{m,B}\}$ for the mth transition path, we can define arbitrary functionals $\mathscr{G}$ such as

$$\mathscr{G}_1[\mathbf{X}^{(m)}] = \tau^+_{m,B} - \tau^-_{m,A}, \tag{2.6}$$

$$\mathscr{G}_2[\mathbf{X}^{(m)}, \mathbf{X}^{(m+1)}] = \tau^+_{m+1,B} - \tau^+_{m,B} \tag{2.7}$$

$$\mathscr{G}_3[\mathbf{X}^{(m)}] = \max \left\{ \left| \frac{U(\mathbf{X}(t_2)) - U(\mathbf{X}(t_1))}{t_2 - t_1} \right| : \tau^-_{m,A} < t_1 < t_2 < \tau^+_{m,B} \right\}, \tag{2.8}$$

where $U(\mathbf{X}(t))$ is shorthand for the eastward wind velocity in the stratosphere. $\mathscr{G}_1$ quantifies the

elapsed time over the course of a transition path; $\mathscr{G}_2$ is the return time between one extreme event and the next; and $\mathscr{G}_3$ is the fastest average rate of change of wind speed recorded over the whole transition path. The quantities of interest $\mathscr{G}$ will, of course, depend on the application.

In principle, we could collect statistics over the transition path ensemble by "direct numerical simulation" (DNS): integrate the system for a long time, collect many $A \to B$ transition paths $\mathbf{X}^{(m)}$, calculate any quantities of interest $\mathscr{G}_1, \mathscr{G}_2, \mathscr{G}_3, \ldots$, and estimate summary statistics. The number of samples needed for a given level of confidence may vary greatly, depending on the process and the $\mathscr{G}$'s chosen. Although DNS is simple and general, it is expensive for high-dimensional models, particularly for rare events with a very long return interval ($\mathscr{G}_2$ above) relative to the simulation timestep.

The following subsection will specify, and chapters 5-7 will demonstrate, an alternative approach known as dynamical Galerkin approximation [DGA; Thiede et al., 2019], which circumvents DNS and uses only short numerical simulations instead—a calculation made possible by TPT. For example, I will show how to compute statistical averages of $\mathscr{G}_1$ and $\mathscr{G}_2$ above without the prohibitive costs of DNS. Yet the approach is not a panacea: nonlinear functionals like $\mathscr{G}_3$ do not fall so neatly within the purview of the method. The mathematical development below will make clear what is possible.

2.2 Itô diffusions

2.2.1 Feynman-Kac formulae

To get more specific with TPT, it will help to get more specific with the process. Following standard references such as Oksendal [2003] and E et al. [2019], I will present a brief overview of the most common and natural setting for TPT analysis, in which $\mathbf{X}(t)$ solves a stochastic differential

equation known as an Itô diffusion:

$$d\mathbf{X}(t) = v(\mathbf{X}(t))\,dt + \sigma(\mathbf{X}(t))\,d\mathbf{W}(t). \tag{2.9}$$

Here, $v : \mathbb{R}^d \to \mathbb{R}^d$ is the deterministic part of the dynamics known as the "drift." $\mathbf{W}(t) \in \mathbb{R}^k$ ($k \leq d$) is a vector of independent Brownian motions that injects randomness, and $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times k}$ is a diffusion matrix which distributes that randomness among components of the physical system. Randomness may represent unresolved processes, such as forcing from fast internal oscillations (e.g., gravity waves), uncertain initial conditions, or model error. Chapters 4-6 work in the setting of Itô diffusion models. Other forms of noise and non-autonomous dynamics, which can and do appear in real applications, can still be analyzed approximately in a TPT framework with appropriate discretization. We will address these issues in chapter 7 when working with real data.

Associated to a diffusion process is the *infinitesimal generator*, $\mathscr{L}$, which acts on suitable functions of state space (also called "observable" functions) by evolving their expectation forward in time:

$$\mathscr{L}H(\mathbf{x}) = \lim_{\Delta t\to 0} \frac{\mathbb{E}_{\mathbf{x}}[H(\mathbf{X}(\Delta t))] - H(\mathbf{x})}{\Delta t} \tag{2.10}$$

where $\mathbb{E}_{\mathbf{x}}[\cdot] := \mathbb{E}[\cdot|\mathbf{X}(0) = \mathbf{x}]$. The Itô chain rule gives an evolution equation for H, following $\mathbf{X}(t)$:

$$dH(\mathbf{X}(t)) = \mathscr{L}H(\mathbf{X}(t))\,dt + dM(t) \tag{2.11}$$

where $M(t)$ is a martingale. For an Itô diffusion, the infinitesimal generator and martingale terms

are partial differential operators:

$$\mathscr{L}H(\mathbf{x}) = \sum_{i=1}^{d} \nu_i(\mathbf{x}) \frac{\partial H(\mathbf{x})}{\partial x_i} + \sum_{i=1}^{d} \sum_{j=1}^{d} \frac{1}{2} \left[\sigma(\mathbf{x})\sigma(\mathbf{x})^\top\right]_{ij} \frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j} \tag{2.12}$$

$$dM(t) = \sum_{i=1}^{d} \sum_{j=1}^{k} \frac{\partial H(\mathbf{x})}{\partial x_i} \sigma_{ij}(\mathbf{x}) dW_j(t) \tag{2.13}$$

It turns out that the committors $q^{\pm}$, probability density π, and other key ingredients in TPT can be expressed as solutions to linear equations, known as *Feynman-Kac formulae*, involving the generator. With a diffusion process, these equations take the form of partial differential equations (PDEs) over $\mathbb{R}^d$. Feynman-Kac formulae are an engine of the DGA method, and I will informally derive them here.

Let D be a sub-domain of Ω, usually the region $(A \cup B)^c$ between A and B where the system has not decided between the normal and abnormal state; depending on the stochastic forcing it could go into either one from time $t = 0$. Consistent with the above notation for first-hitting times, let $\tau_{D^c} = \min\{t \geq 0 : \mathbf{X}(t) \notin D\}$ be the first exit time from this domain starting at time zero. This is a random variable which depends on the starting condition $\mathbf{x} \in D$. Let $G : \partial D \to \mathbb{R}$ be a boundary condition, $\lambda \in \mathbb{R}$ a real number, and $\Gamma : D \to \mathbb{R}$ a term to represent accumulated risk, all three of which can be chosen by the user. We seek a PDE for the *forecast function*

$$F(\mathbf{x}) = \mathbb{E}_{\mathbf{x}}\left[G(\mathbf{X}(\tau_{D^c}^+)) \exp\left(\lambda \int_0^{\tau_{D^c}^+} \Gamma(\mathbf{X}(s))\, ds\right)\right], \tag{2.14}$$

which can express various notions of risk depending on the choice of G, Γ, and λ. To derive the PDE, for F, consider the following stochastic process:

$$Z(t) = F(\mathbf{X}(t))Y(t) \tag{2.15}$$

where $Y(t) := \exp\left(\lambda \int_0^t \Gamma(\mathbf{X}(s))\,ds\right)$. Itô's lemma gives us that $dY(t) = \lambda\Gamma(\mathbf{X}(t))Y(t)\,dt$. Hence, applying the product rule to $Z(t)$,

$$\begin{aligned} dZ(t) &= dF(\mathbf{X}(t))Y(t) + F(\mathbf{X}(t))\,dY(t) \qquad (2.16)\\ &= \mathscr{L}F(\mathbf{X}(t))Y(t)\,dt + dM(t)Y(t) + \lambda F(\mathbf{X}(t))\Gamma(\mathbf{X}(t))Y(t)\,dt \\ &= \left[\mathscr{L}F + \lambda\Gamma F\right](\mathbf{X}(t))Y(t)\,dt + Y(t)dM(t) \end{aligned}$$

where in (2.16) we have left out the quadratic cross-variation of $F(\mathbf{X}(t))$ and $Y(t)$ because Y has finite variation. If the bracketed term $(\mathscr{L} + \lambda\Gamma(\mathbf{x}))F(\mathbf{x}) = 0$ for all $\mathbf{x}$, then $Z(t)$ is a martingale and it follows that

$$Z(0) = \mathbb{E}_{\mathbf{x}}[Z(t)] \qquad (2.17)$$

$$F(\mathbf{x}) = \mathbb{E}_{\mathbf{x}}\left[F(\mathbf{X}(t))\exp\left(\lambda\int_0^t \Gamma(\mathbf{X}(s))\,ds\right)\right] \qquad (2.18)$$

Finally, the formula still holds if we substitute a stopping time for t. Here, a *stopping time* technically means a random variable time τ that is measurable with respect to the filtration up until τ [Oksendal, 2003]. By choosing τ_{D^c}, the first exit time from D, the $F(\mathbf{X}(t))$ inside the brackets becomes its boundary value $G(\mathbf{X}(\tau_{D^c}))$. Thus $F(\mathbf{x})$ as defined in (2.14) also solves the boundary value problem

$$\begin{cases} (\mathscr{L} + \lambda\Gamma(\mathbf{x}))F(\mathbf{x}) = 0 & \mathbf{x} \in D \\ F(\mathbf{x}) = G(\mathbf{x}) & \mathbf{x} \in D^c \end{cases} \qquad (2.19)$$

In a diffusion process, this boundary value problem takes the form of a PDE according to the expression for the generator.

It will also prove useful to treat F as a moment-generating function for the Γ-integral, by differentiating in λ and setting $\lambda = 0$. This is related to the Kac moment method [Fitzsimmons and

Pitman, 1999]. The corresponding expectations are

$$\partial_\lambda^k F(\mathbf{x};\lambda=0) = \mathbb{E}_{\mathbf{x}}\left[G(\mathbf{X}(\tau_{D^c}^+))\left(\int_0^{\tau_{D^c}^+}\Gamma(\mathbf{X}(t))\,dt\right)^k\right] \tag{2.20}$$

In other words, λ-derivatives of F give us moments of the probability distribution of Γ-integrals, which may be exploited to ask about various statistical properties of the passage to B. Simple manipulations of Eq. (2.19) gives a sequence of linear equations for $\partial_\lambda^k F(\mathbf{x};\lambda)$ at $\lambda = 0$:

$$(\mathscr{L}+\lambda\Gamma(\mathbf{x}))\partial_\lambda F(\mathbf{x};\lambda)+\Gamma(\mathbf{x})F(\mathbf{x};\lambda)=0 \tag{2.21}$$

$$\mathscr{L}\left[\partial_\lambda F\right](\mathbf{x};0) = -\Gamma(\mathbf{x})F(\mathbf{x};0) \tag{2.22}$$

$$(\mathscr{L}+\lambda\Gamma(\mathbf{x}))\partial_\lambda^2 F(\mathbf{x};\lambda)+2\Gamma(\mathbf{x})\partial_\lambda F(\mathbf{x};\lambda)=0 \tag{2.23}$$

$$\mathscr{L}\left[\partial_\lambda^2 F\right](\mathbf{x};0) = -2\Gamma(\mathbf{x})\partial_\lambda F(\mathbf{x};0) \tag{2.24}$$

$$(\mathscr{L}+\lambda\Gamma(\mathbf{x}))\partial_\lambda^3 F(\mathbf{x};\lambda)+3\Gamma(\mathbf{x})\partial_\lambda^2 F(\mathbf{x};\lambda)=0 \tag{2.25}$$

$$\mathscr{L}\left[\partial_\lambda^3 F\right](\mathbf{x};0) = -3\Gamma(\mathbf{x})\partial_\lambda^2 F(\mathbf{x};0) \tag{2.26}$$

$$\vdots \tag{2.27}$$

$$(\mathscr{L}+\lambda\Gamma(\mathbf{x}))\partial_\lambda^k F(\mathbf{x};\lambda)+k\Gamma(\mathbf{x})\partial_\lambda^{k-1} F(\mathbf{x};\lambda)=0 \tag{2.28}$$

$$\mathscr{L}\left[\partial_\lambda^k F\right](\mathbf{x};0) = -k\Gamma(\mathbf{x})\partial_\lambda^{k-1} F(\mathbf{x};0) \tag{2.29}$$

This sequence is recursive: the $(k-1)$th derivative of F provides a ready-made source term for the equation for the kth derivative.

What about the boundary condition? For $\mathbf{x}\in D^c$, $\tau_{D^c}^+ = 0$, so all the Γ-integrals are zero, which makes homogeneous boundary conditions for all $k \geq 1$.

$$\partial_\lambda^k F(\mathbf{x};0) = \mathbb{E}_{\mathbf{x}}\left[G(\mathbf{X}(0))\left(\int_0^0\Gamma(\mathbf{X}(t))\,dt\right)^k\right] = 0 \tag{2.30}$$

2.2.2 *Dynkin's formula and finite lag time*

In practice, when we later discretize Eq. (2.19), we achieve better numerical stability integrating the generator to a finite lag time Δt, following Strahan et al. [2021], rather than estimating a numerical limit as $\Delta t \to 0$ in Eq. (2.10). The theorem that allows this is called Dynkin's formula, which states that for any suitable function $H : \mathbb{R}^d \to \mathbb{R}$ and a stopping time θ,

$$\mathbb{E}_{\mathbf{x}}[H(\mathbf{X}(\theta))] = H(\mathbf{x}) + \mathbb{E}_{\mathbf{x}}\left[\int_0^\theta \mathscr{L}H(\mathbf{X}(t))\,dt\right]. \tag{2.31}$$

The left-hand side, $\mathbb{E}_{\mathbf{x}}[H(\mathbf{X}(\theta))]$, is known as the *transition operator* $\mathscr{T}^\theta H(\mathbf{x})$, a finite-time version of the generator. Note that this is a deterministic operator despite θ being a random variable, because by definition $\mathscr{T}^\theta$ only has θ inside of expectations.

Applying Dynkin's formula to $F(\mathbf{x})$, with $\theta = \min(\Delta t, \tau_{D^c})$, we find

$$\begin{aligned}
\mathbb{E}_{\mathbf{x}}[F(\mathbf{X}(\theta))] &= F(\mathbf{x}) + \mathbb{E}_{\mathbf{x}}\left[\int_0^\theta \mathscr{L}F(\mathbf{X}(t))\,dt\right] \\
&= F(\mathbf{x}) - \lambda\,\mathbb{E}_{\mathbf{x}}\left[\int_0^\theta \Gamma(\mathbf{X}(t))F(\mathbf{X}(t))\,dt\right] \\
\mathscr{T}^\theta F(\mathbf{x}) &= F(\mathbf{x}) - \lambda\,\mathscr{K}_\Gamma^\theta F(\mathbf{x})
\end{aligned} \tag{2.32}$$

where $\mathscr{K}_\Gamma^\theta$ is shorthand notation for the integral operator on the right.

2.2.3 *Steady-state distribution*

We will generally assume that the process $\mathbf{X}(t)$ has a steady-state distribution π, which we will typically represent as $\pi(\mathbf{x})$: a density with respect to Lebesgue measure. π is defined by several equivalent properties, the simplest being Eq. (2.1). Another property, known as ergodicity, is that for any suitable observable function H, the long-term average of H following $\mathbf{X}(t)$ equates to a

π-weighted average of f over state space:

$$\lim_{T\to\infty}\int_{-T}^{T} H(\mathbf{X}(t))\,dt = \int_{\Omega} H(\mathbf{x})\pi(\mathbf{x})\,d\mathbf{x} =: \langle f\rangle_{\pi} \tag{2.33}$$

The third property will provide an equation for $\pi(\mathbf{x})$ in terms of the generator. If $\mathbf{X}(0)$ is a random variable drawn from π, then its time-integrated image $\mathbf{X}(\Delta t)$ is also drawn from π. This means that an observable $H(\mathbf{x})$ can equally well be averaged over $\mathbf{X}(0)$ or $\mathbf{X}(\Delta t)$: they have the same distribution.

$$\mathbb{E}_{\mathbf{X}(0)\sim\pi}\big[H(\mathbf{X}(\Delta t))\big] = \mathbb{E}_{\mathbf{X}(\Delta t)\sim\pi}\big[H(\mathbf{X}(\Delta t))\big] \tag{2.34}$$

$$\int \mathscr{T}^{\Delta t} H(\mathbf{x})\pi(\mathbf{x})\,d\mathbf{x} = \int H(\mathbf{y})\pi(\mathbf{y})\,d\mathbf{y} \tag{2.35}$$

$$\int (\mathscr{T}^{\Delta t} - 1)H(\mathbf{x})\pi(\mathbf{x})\,d\mathbf{x} = 0 \tag{2.36}$$

$$\int H(\mathbf{x})(\mathscr{T}^{\Delta t} - 1)^{*}\pi(\mathbf{x})\,d\mathbf{x} = 0 \tag{2.37}$$

where * denotes the adjoint operator with respect to the Lebesgue measure. Because this holds for arbitrary H, we can write

$$(\mathscr{T}^{\Delta t} - 1)^{*}\boldsymbol{\pi} = 0 \tag{2.38}$$

Or, dividing by Δt and taking the limit $\Delta t \to 0$,

$$\mathscr{L}^{*}\boldsymbol{\pi}(\mathbf{x}) = 0 \tag{2.39}$$

$$\int \pi(\mathbf{x})\,d\mathbf{x} = 1 \tag{2.40}$$

where the second equation ensures that π is a properly normalized probability density. More generally, for any "reference measure" ζ that is absolutely continuous with respect to π, we can write

an equation for the change of measure $d\pi/d\zeta$:

$$\int \mathscr{T}^{\Delta t} H(\mathbf{x}) \frac{d\pi}{d\zeta}(\mathbf{x}) \zeta(\mathbf{x})\, d\mathbf{x} = \int H(\mathbf{y}) \frac{d\pi}{d\zeta}(\mathbf{y}) \zeta(\mathbf{y})\, d\mathbf{y} \tag{2.41}$$

$$\int (\mathscr{T}^{\Delta t} - 1) H(\mathbf{x}) \frac{d\pi}{d\zeta}(\mathbf{x}) \zeta(\mathbf{x})\, d\mathbf{x} = 0 \tag{2.42}$$

$$\int H(\mathbf{x}) (\mathscr{T}^{\Delta t} - 1)^*_\zeta \left[\frac{d\pi}{d\zeta}\right](\mathbf{x})\, d\mathbf{x} = 0 \tag{2.43}$$

$$\therefore (\mathscr{T}^{\Delta t} - 1)^*_\zeta \left[\frac{d\pi}{d\zeta}\right](\mathbf{x}) = 0 \tag{2.44}$$

This will become important when we use data to estimate π, in which the reference measure is not Lebesgue but a sampling measure.

2.2.4 Time reversal

We refer the conditional expectation (2.14) as a *forecast*, simply because it considers the future evolution of $\mathbf{X}(t)$. But transition paths are defined by their past as well as their future: if a snapshot $\mathbf{X}(t)$ is part of a transition path, the next hitting time $\tau^+_{D^c}(t)$ will find the system in B, while the most recent hitting time $\tau^-_{D^c}(t)$ saw the system in A. This calls for a time-reversed forecast, or *aftcast* (the term "hindcast" is already taken; see chapter 7), and therefore a time-reversed generator $\widetilde{\mathscr{L}}$ that pushes expectations backward in time:

$$\widetilde{\mathscr{L}} H(\mathbf{x}) = \lim_{\Delta t \to 0} \frac{\widetilde{\mathbb{E}}[H(\mathbf{X}(-\Delta t))] - H(\mathbf{x})}{\Delta t} =: \lim_{\Delta t \to 0} \frac{(\mathscr{T}^{-\Delta t} - 1) H(\mathbf{x})}{\Delta t} \tag{2.45}$$

where $\widetilde{\mathbb{E}}$ denotes backward-in-time expectation. We specifically consider backward dynamics *at stationarity*, i.e., the dynamics such that the correlation between a function f at time 0 and g at

time Δt are equivalent, whether we evolve forward from $0 \to \Delta t$ or backward from $\Delta t \to 0$:

$$\mathbb{E}_{\mathbf{X}(0)\sim\pi}\big[f(\mathbf{X}(0))g(\mathbf{X}(\Delta t))\big] = \mathbb{E}_{\mathbf{X}(\Delta t)\sim\pi}\big[f(\mathbf{X}(0))g(\mathbf{X}(\Delta t))\big] \tag{2.46}$$

$$\int f(\mathbf{x})\mathscr{T}^{\Delta t}g(\mathbf{x})\pi(\mathbf{x})\,d\mathbf{x} = \int \mathscr{T}^{-\Delta t}f(\mathbf{y})g(\mathbf{y})\pi(\mathbf{y})\,d\mathbf{y} \tag{2.47}$$

$$\int (\mathscr{T}^{\Delta t})^*[f\pi](\mathbf{x})g(\mathbf{x})\,d\mathbf{x} = \int \pi(\mathbf{y})\mathscr{T}^{-\Delta t}f(\mathbf{y})g(\mathbf{y})\,d\mathbf{y} \tag{2.48}$$

$$\therefore \frac{1}{\pi(\mathbf{x})}(\mathscr{T}^{\Delta t})^*[\pi f](\mathbf{x}) = \mathscr{T}^{-\Delta t}f(\mathbf{x}) \tag{2.49}$$

since g is arbitrary. The reversed infinitesimal generator $\widetilde{\mathscr{L}}$ inherits the same property: $\frac{1}{\pi}\mathscr{L}^*[\pi f] = \widetilde{\mathscr{L}}f$. For Itô diffusions, we have an explicit formula for $\mathscr{L}^*$ through integration by parts. For brevity, we abbreviate the diffusion matrix $\frac{1}{2}\sigma\sigma^\top$ by $\mathbf{a} = [a_{ij}]$, $\partial_i = \partial/\partial x_i$, and implicitly sum over repeated indices. For any two twice-differentiable functions f and g which vanish at the boundary of our domain,

$$\langle \mathscr{L}f, g\rangle = \int \mathscr{L}f(\mathbf{x})g(\mathbf{x})\,d\mathbf{x} \tag{2.50}$$

$$= \int \big[v_i\partial_i f + a_{ij}\partial_i\partial_j f\big]g\,d\mathbf{x} \tag{2.51}$$

$$= \int \partial_i\big[v_i f g + a_{ij}g\partial_j f\big]\,d\mathbf{x} - \int \big[\partial_i(v_i g)f + \partial_i(a_{ij}g)\partial_j f\big]\,d\mathbf{x} \tag{2.52}$$

$$= -\int \partial_i(v_i g)f\,d\mathbf{x} - \int \partial_j\big[\partial_i(a_{ij}g)f\big]\,d\mathbf{x} + \int \partial_i\partial_j(a_{ij}g)f\,d\mathbf{x} \tag{2.53}$$

$$= \int \big[-\partial_i(v_i g) + \partial_i\partial_j(a_{ij}g)\big]f\,d\mathbf{x} \tag{2.54}$$

$$\therefore \mathscr{L}^*g(\mathbf{x}) = -\sum_{j=1}^{d}\frac{\partial}{\partial x_i}\big[v_i(\mathbf{x})g(\mathbf{x})\big] + \sum_{i=1}^{d}\sum_{j=1}^{d}\frac{\partial^2}{\partial x_i\partial x_j}\left[\frac{1}{2}\Big(\sigma(\mathbf{x})\sigma(\mathbf{x})^\top\Big)_{ij}g(\mathbf{x})\right]. \tag{2.55}$$

All integrals of total derivatives have vanished, thanks to the homogeneous boundary conditions.

2.3 TPT quantities of interest

Now that we have equations for the general forecast function (2.14), we can start assembling the building blocks of a full TPT analysis with different choices of G and Γ.

2.3.1 Forecasts

The first building block is the forward committor, for which we set $D = (A \cup B)^{\mathrm{c}}$, $G(\mathbf{x}) = \mathbb{1}_B(\mathbf{x})$, and $\Gamma \equiv 0$:

$$F^+(\mathbf{x}) = \mathbb{E}_{\mathbf{x}}\left[\mathbb{1}_B(\mathbf{X}(\tau^+_{A\cup B}))\right] \tag{2.56}$$

$$= \mathbb{P}_{\mathbf{x}}\{\mathbf{X}(\tau^+_{A\cup B}) \in B\} =: q^+_B(\mathbf{x}). \tag{2.57}$$

We can therefore immediately write down a boundary value problem for q^+_B,

$$\begin{cases} \mathscr{L} q^+_B(\mathbf{x}) = 0 & \mathbf{x} \in (A \cup B)^{\mathrm{c}} \\ q^+_B(\mathbf{x}) = 0 & \mathbf{x} \in A \\ q^+_B(\mathbf{x}) = 1 & \mathbf{x} \in B \end{cases} \tag{2.58}$$

How long does the system take to get from its starting point $\mathbf{x}$ to the abnormal state B? We obtain an equation for the mean first passage time to B (MFPT_B) by defining $D = B^{\mathrm{c}}$ (including A), $G(\mathbf{x}) = 1$, $\Gamma = 1$. With respect to λ, we retain it as an auxiliary variable and differentiate:

$$F^+(\mathbf{x};\lambda) = \mathbb{E}_{\mathbf{x}}\left[e^{\lambda \tau^+_B}\right] \tag{2.59}$$

$$\partial_\lambda F^+(\mathbf{x};0) = \mathbb{E}_{\mathbf{x}}\left[\tau^+_B\right] = \mathrm{MFPT}_B(\mathbf{x}). \tag{2.60}$$

Reading off the first recursive equation for moments above, we have

$$\mathscr{L}\left[\partial_\lambda F^+\right](\mathbf{x};0) = -F^+(\mathbf{x};0) = -1. \tag{2.61}$$

The last equality follows from plugging in $\lambda = 0$ to $F^+ = \mathbb{E}_{\mathbf{x}} e^{\lambda \tau_B^+}$. The boundary condition for MFPT_B is simply zero, because starting at B, the process needs no additional time to arrive at B. Thus the full boundary value problem for MFPT_B is

$$\begin{cases} \mathscr{L}\left[\text{MFPT}_B\right](\mathbf{x}) = -1 & \mathbf{x} \in B^{\text{c}} \\ \text{MFPT}_B(\mathbf{x}) = 0 & \mathbf{x} \in B. \end{cases} \tag{2.62}$$

Any other set S can be substituted for B here, to get the mean first passage time to S.

MFPT_B gives a notion of the system's overall propensity to reach B. However, this expectation includes possible paths that visit A first, perhaps for a long time. In certain weather forecasting contexts, the real-time speed of the event itself might be more immediately important (see chapter 5). To quantify the time to B *after the transition path has already started*, we condition the MFPT on going directly to B by choosing $D = (A \cup B)^{\text{c}}$, $G(\mathbf{x}) = \mathbb{1}_B(\mathbf{x})$, and $\Gamma = 1$. Manipulating the Feynman-Kac formula (2.14) results in an expression for the *lead time*:

$$F^+(\mathbf{x};\lambda) = \mathbb{E}_{\mathbf{x}}\left[\mathbb{1}_B(\mathbf{X}(\tau_{A\cup B}^+)) e^{\lambda \tau_{A\cup B}^+}\right] \tag{2.63}$$

$$\partial_\lambda F^+(\mathbf{x};0) = \mathbb{E}_{\mathbf{x}}\left[\mathbb{1}_B(\mathbf{X}(\tau_{A\cup B}^+))\tau_{A\cup B}^+\right] \tag{2.64}$$

$$\frac{\partial_\lambda F^+(\mathbf{x};0)}{q^+(\mathbf{x})} = \frac{\mathbb{E}_{\mathbf{x}}[\mathbb{1}_B(\mathbf{X}(\tau_{A\cup B}^+))\tau_{A\cup B}^+]}{\mathbb{E}_{\mathbf{x}}[\mathbb{1}_B(\mathbf{X}(\tau_{A\cup B}^+))]} \tag{2.65}$$

$$= \mathbb{E}\left[\tau_{A\cup B}^+ | \mathbf{X}(\tau_{A\cup B}^+) \in B\right] =: \eta_B^+(\mathbf{x}) = \text{Lead time to } B \tag{2.66}$$

The denominator and numerator correspond to the zeroth and first λ-derivatives of the forecast function (5.14), and an be solved with the corresponding recursive Feynman-Kac formulae. The lead time is an important object in the SSW application in chapter 5.

2.3.2 *Aftcasts*

Every forecast function has a corresponding aftcast function for the time-reversed process. The *backward committor* is defined as the probability of having emerged from A more recently than B:

$$q_A^-(\mathbf{x}) = \mathbb{E}_{\mathbf{x}}[\mathbb{1}_A(\mathbf{X}(\tau_{A\cup B}^-))] = \mathbb{P}_{\mathbf{x}}\{\mathbf{X}(\tau_{A\cup B}^-) \in A\}, \tag{2.67}$$

where the backward hitting time is the random variable defined as

$$\tau_{A\cup B}^-(t_0) = \max\{t \le t_0 : \mathbf{X}(t) \in A\cup B\}. \tag{2.68}$$

Similarly, the backward lead time is the elapsed time since leaving A, conditional on having left A more recently:

$$\eta_A^-(\mathbf{x}) = \mathbb{E}_{\mathbf{x}}\left[\tau_{A\cup B}^- | \mathbf{X}(\tau_{A\cup B}^-) \in A\right] = \frac{\mathbb{E}_{\mathbf{x}}[\mathbb{1}_A(\mathbf{X}(\tau_{A\cup B}^-))]}{q_A^-(\mathbf{x})} \tag{2.69}$$

More generally, we introduce a backward version of the general forecast F above, and decorate both versions to indicate their time orientation as well as their dependence on λ and Γ:

$$F_\Gamma^+(\mathbf{x};\lambda) = \mathbb{E}_{\mathbf{x}}\left[\mathbb{1}_B(\mathbf{X}(\tau_{A\cup B}^+))\exp\left(\lambda\int_0^{\tau_{A\cup B}^+}\Gamma(\mathbf{X}(r))\,dr\right)\right] \tag{2.70}$$

$$F_\Gamma^-(\mathbf{x};\lambda) = \mathbb{E}_{\mathbf{x}}\left[\mathbb{1}_A(\mathbf{X}(\tau_{A\cup B}^-))\exp\left(\lambda\int_{\tau_{A\cup B}^-}^{0}\Gamma(\mathbf{X}(r))\,dr\right)\right] \tag{2.71}$$

Using the fact that the past and future are independent *conditional on the present*, we multiply the aftcast and hindcast together to get an expectation over transition paths crossing through $\mathbf{x}$:

$$F_\Gamma^-(\mathbf{x},\lambda)F_\Gamma^+(\mathbf{x},\lambda) = \tag{2.72}$$
$$\mathbb{E}_{\mathbf{x}}\left[\mathbb{1}_A(\mathbf{X}(\tau_{A\cup B}^-))\mathbb{1}_B(\mathbf{X}(\tau_{A\cup B}^+))\exp\left(\lambda\int_{\tau_{A\cup B}^-}^{\tau_{A\cup B}^+}\Gamma(\mathbf{X}(r))\,dr\right)\right]$$

Differentiating repeatedly in λ at $\lambda = 0$ provides us with all moments of the probability distribution of *transition path integrals*:

$$\partial_\lambda^k \left[F^-(\mathbf{x},\lambda)F_\Gamma^+(\mathbf{x},\lambda)\right]_{\lambda=0} = \mathbb{E}_{\mathbf{x}}\left[\mathbb{1}_A(\mathbf{X}(\tau_{A\cup B}^-))\mathbb{1}_B(\mathbf{X}(\tau_{A\cup B}^+))\left(\int_{\tau_{A\cup B}^-}^{\tau_{A\cup B}^+}\Gamma(\mathbf{X}(r))\,dr\right)^k\right] \tag{2.73}$$

The expectation is restricted to paths crossing through $\mathbf{x}$. Setting $k = 0$, this is simply $q_B^+(\mathbf{x})q^-(\mathbf{x})$, the probability of an observed snapshot $\mathbf{x}$ being part of a transition path. With $k \geq 1$, it is natural to condition on snapshots being reactive by dividing by $q_A^-(\mathbf{x})q_B^+(\mathbf{x})$.

$$\begin{aligned}
&\frac{\partial_\lambda^k[F_\Gamma^+(\mathbf{x};\lambda)F_\Gamma^-(\mathbf{x};\lambda)]_{\lambda=0}}{q_A^-(\mathbf{x})q_B^+(\mathbf{x})} \\
&= \mathbb{E}_{\mathbf{x}}\left[\left(\int_{\tau_{A\cup B}^-}^{\tau_{A\cup B}^+}\Gamma(\mathbf{X}(r))\,dr\right)^k\middle|\mathbf{X}(\tau_{A\cup B}^-)\in A,\mathbf{X}(\tau_{A\cup B}^+)\in B\right] \\
&= \mathbb{E}_{\mathbf{x}}\left[\left(\int_{\tau_{A\cup B}^-}^{\tau_{A\cup B}^+}\Gamma(\mathbf{X}(r))\,dr\right)^k\middle|A\to B\right] \text{ (abbreviation)}
\end{aligned} \tag{2.74}$$

Everything we say about transition paths stems originally from the functions F_Γ^+ and F_Γ^- for various Γ, as well as the steady-state distribution π.

Having introduced TPT's building blocks, we now present some common statistical averages that distinguish transition path behavior from ordinary behavior of $\mathbf{X}(t)$. Depending on the application, these averages might provide direct insight into the dynamics of interest. There are two broad categories of statistics, the first pertaining to the reactive snapshots from Eq. (2.4) and the second pertaining to the transition paths from Eq. (2.5).

2.3.3 Reactive snapshot averages

Just as the process $\mathbf{X}(t)$ is has a limiting distribution of $\pi(\mathbf{x})$, the reactive trajectories have their own limiting *reactive density* $\pi_{AB}(\mathbf{x})$. The ergodic property says that time averages *restricted to*

reactive times are equivalent to spatial averages weighted by π_{AB}: for any observable f,

$$\int_\Omega f(\mathbf{x})\pi_{AB}(\mathbf{x}) = \lim_{T\to\infty} \frac{\int_{-T}^{T} f(\mathbf{X}(t))\mathbb{1}_A(\mathbf{X}(\tau^-_{A\cup B}(t)))\mathbb{1}_B(\mathbf{X}(\tau^+_{A\cup B}(t)))\,dt}{\int_{-T}^{T} \mathbb{1}_A(\mathbf{X}(\tau^-_{A\cup B}(t)))\mathbb{1}_B(\mathbf{X}(\tau^+_{A\cup B}(t)))\,dt} \tag{2.75}$$

$$= \frac{\langle f q_A^- q_B^+\rangle\pi}{\langle q_A^- q_B^-\rangle\pi} = \int_\Omega f(\mathbf{x})\frac{q_A^-(\mathbf{x})q_B^-(\mathbf{x})\pi(\mathbf{x})}{\langle q_A^- q_B^-\rangle\pi}\,d\mathbf{x} \tag{2.76}$$

where we used ergodicity in the second line. Hence, $\pi_{AB}(\mathbf{x})$ is the product of $\pi(\mathbf{x})$ with the forward and backward committors—the probability of being reactive at any given point—and normalized so that π_{AB} integrates to one. The normalizing factor, $\langle q_A^- q_B^-\rangle$, is the time fraction spent en route from A to B. Various choices of f give various cross-sections of the system's transitory behavior. With $f(\mathbf{x}) = \mathbb{1}\{U(\mathbf{x}) > U_0\}$, $\langle f\rangle_{\pi_{AB}}$ is the fraction of transition time with the observable $U(\mathbf{x})$ (such as wind speed) exceeding a threshold U_0. A systematic difference from $\langle f\rangle_\pi$ would indicate that transition paths are systematically different from everyday behavior, at least with respect to f. One can also swap different combinations of symbols for AB. For instance, π_{AA} is the probability density of the system on its way from A, back to A. In chapter 6, we will compare the probability densities on all four phases of the SSW "life cycle" AA, AB, BB, BA.

2.3.4 Transition path averages and currents

Thinking about transition paths from start to finish as discrete, coherent objects unlocks a richer description of the ensemble, including their average dynamical tendency, variability, and most importantly their rate. Here we generalize the concept of rate to not only count transitions, but to characterize certain functionals of distributions. The *generalized rate* is defined as

$$R_\Gamma(\lambda) := \lim_{T\to\infty} \frac{1}{T} \sum_{m=m_0}^{m_0+M_T-1} \exp\left(\lambda \int_{\tau^-_{A,m}}^{\tau^+_{B,m}} \Gamma(\mathbf{X}(r))\,dr\right) \tag{2.77}$$

where $m_0 \le 0$ is index of the first transition path beginning during the interval $(-T, T)$ and M_T is the number of transition paths fully contained within that interval. The notation emphasizes

dependence on the observable Γ and the real parameter λ. To unpack this formula, first set $\lambda = 0$ and observe that $R_\Gamma(0) = \frac{M_T}{T}$ is the number of transitions per unit time—the ordinary rate—whose inverse is the average period of the full "life cycle" $A \to B \to A$. This is not to be confused with the asymmetric forward and backward rates,

$$k_{AB} = \frac{R_\Gamma(0)}{\langle q_A^- \rangle \pi}, \qquad\qquad k_{BA} = \frac{R_\Gamma(0)}{\langle q_B^- \rangle \pi} \tag{2.78}$$

which distinguish the $A \to B$ and $B \to A$ directions by how fast they occur. The factor $\langle q_A^- \rangle \pi$ is the time fraction spent *having last been in A* rather than B, and $\langle q_B^- \rangle$ is its complement. For example, if A were very stable and B very unstable, the system would spend most of its time in the basin of attraction of A, making $\langle q_A^- \rangle \pi$ large and $k_{AB} \ll k_{BA}$. Asymmetric rates (or "rate constants") are very important for chemistry applications, and we do present some in 4, but the symmetric rate turns out more useful overall.

Returning to (2.77), we divide through by $R_\Gamma(0)$:

$$\begin{aligned} \frac{R_\Gamma(\lambda)}{R_\Gamma(0)} &= \lim_{T\to\infty} \frac{1}{M_T} \sum_{m=m_0}^{m_0+M_T-1} \exp\left(\lambda \int_{\tau_m^-}^{\tau_m^+} \Gamma(\mathbf{X}(r))\, dr\right) \\ &= \mathbb{E}_{\text{paths}}\left[\exp\left(\lambda \int_{\tau_A^-}^{\tau_B^+} \Gamma(\mathbf{X}(r))\, dr\right)\right] \end{aligned} \tag{2.79}$$

where the subscript "paths" distinguishes the expectation as over *all* transition paths, not just those crossing through a fixed $\mathbf{x}$ as in (2.74). The right side of (2.79) is a moment-generating function for the *transition path integral* $\int \Gamma\, dt$. Differentiating in λ yields the moments of that distribution, including its variance, skew, and kurtosis:

$$\frac{\partial_\lambda^k R_\Gamma(0)}{R_\Gamma(0)} = \mathbb{E}_{\text{paths}}\left[\left(\int_{\tau_A^-}^{\tau_B^+} \Gamma(\mathbf{X}(r))\, dr\right)^k\right], \tag{2.80}$$

Thus, $R_\Gamma(\lambda)$ contains much information about the transition ensemble as measured by path integrals.

How can $R_\Gamma(\lambda)$ be computed? Consistent with the promise of the last section, it is fully expressible in terms of the forecast F_Γ^+ and the corresponding aftcast F_Γ^-. We must convert Equation (2.77), a sum over transition paths $\sum_{m=1}^{M_T}(\cdot)$, into an integral over time $\int_{-T}^{T}(\cdot)\,dt$ and then (by ergodicity) into an integral over space $\int_{\mathbb{R}^d}(\cdot)\pi(\mathbf{x})\,d\mathbf{x}$. This approach extends the rate derivation in Vanden-Eijnden [2006] and Strahan et al. [2021] to generalized rates.

To write the rate as a time integral, we introduce a set S which fully contains A and does not intersect B. Its surface $C = \partial S$ is called a *dividing surface*. We make the simple observation that every transition path must cross from the inside of C to the outside of C at least once. It may cross back again, but then it must exit again, and so on. All told, each transition path has to cross C an odd number of times, with #(outward crossings) $-$ #(inward crossings) $= 1$. We then implement the counting operation by applying a mask under a time integral to select only the time segments when a reactive trajectory segment is crossing this surface ($+1$ for positive crossings and -1 for negative crossings), resulting in unit weight for each transition path. To be explicit,

$$R_\Gamma(\lambda) = \lim_{\Delta t\to 0}\frac{1}{\Delta t}\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} \tag{2.81}$$

$$\exp\left(\lambda\int_{\tau^-_{A\cup B}(t)}^{\tau^+_{A\cup B}(t+\Delta t)}\Gamma(\mathbf{X}(r))\,dr\right)\times \tag{2.82}$$

$$\mathbb{1}_A(\mathbf{X}(\tau^-_{A\cup B}(t)))\mathbb{1}_B(\mathbf{X}(\tau^+_{A\cup B}(t+\Delta t)))\times \tag{2.83}$$

$$\Big[\mathbb{1}_S(\mathbf{X}(t))\mathbb{1}_{S^c}(\mathbf{X}(t+\Delta t)) \tag{2.84}$$

$$-\mathbb{1}_{S^c}(\mathbf{X}(t))\mathbb{1}_S(\mathbf{X}(t+\Delta t))\Big]\,dt \tag{2.85}$$

The idea is to restrict the interval $(0,T)$ to the collection of time intervals $(t,t+\Delta t)$ during which the path crosses the surface ∂S. Line (2.83) applies a mask picking out transition path segments, which are those that come from A and next go to B. Line (2.84) applies a further mask picking out the narrow time intervals when $\mathbf{X}(t)$ exits the region from S to S^c, while line (2.85) subtracts the backward crossings from S^c to S. Using ergodicity, we can replace the time integral with a space

integral and insert conditional expectations inside. For example, the part of the integrand

$$\exp\left(\lambda \int_{t+\Delta t}^{\tau^+_{A\cup B}(t+\Delta t)} \Gamma(\mathbf{X}(r))\,dr\right)\times \qquad (2.86)$$
$$\mathbb{1}_B(\mathbf{X}(\tau^+_{A\cup B}(t+\Delta t)))\mathbb{1}_{S^c}(\mathbf{X}(t+\Delta t))$$

becomes, after taking conditional expectations,

$$\mathbb{E}\left[\mathbb{1}_{S^c}(\mathbf{X}(t+\Delta t))F^+_\Gamma(\mathbf{X}(t+\Delta t))\middle|\mathbf{X}(t)=\mathbf{x}\right] \qquad (2.87)$$
$$=: \mathscr{T}^{\Delta t}\left[\mathbb{1}_{S^c}F^+_\Gamma\right](\mathbf{x})$$

Where the the transition operator, as above, is defined by $\mathscr{T}^{\Delta t}f(\mathbf{x}) = \mathbb{E}_{\mathbf{x}}[f(\mathbf{X}(\Delta t))]$. Applying similar logic to all terms in the integrand, we have the following generalized rate formula:

$$R_\Gamma(\lambda) = \lim_{\Delta t\to 0}\frac{1}{\Delta t}\int_{\mathbb{R}^d} F^-_\Gamma(\mathbf{x};\lambda)\times \qquad (2.88)$$
$$\left\{\mathbb{1}_S\mathscr{T}^{\Delta t}\left[\mathbb{1}_{S^c}F^+_\Gamma\right] - \mathbb{1}_{S^c}\mathscr{T}^{\Delta t}\left[\mathbb{1}_S F^+_\Gamma\right]\right\}(\mathbf{x})\pi(\mathbf{x})\,d\mathbf{x}$$

which holds for any S enclosing A and disjoint from B. A more compact, although less symmetric, version is found by replacing $\mathbb{1}_S$ with $1-\mathbb{1}_{S^c}$ and canceling two terms:

$$R_\Gamma(\lambda) = \lim_{\Delta t\to 0}\frac{1}{\Delta t}\int \pi F^-_\Gamma\left\{(1-\mathbb{1}_{S^c})\mathscr{T}^{\Delta t}\left[\mathbb{1}_{S^c}F^+_\Gamma\right] - \mathbb{1}_{S^c}\mathscr{T}^{\Delta t}\left[(1-\mathbb{1}_{S^c})F^+_\Gamma\right]\right\} \qquad (2.89)$$
$$= \lim_{\Delta t\to 0}\frac{1}{\Delta t}\int \pi F^-_\Gamma\left\{\mathscr{T}^{\Delta t}\left[\mathbb{1}_{S^c}F^+_\Gamma\right] - \mathbb{1}_{S^c}\mathscr{T}^{\Delta t}F^+_\Gamma\right\}d\mathbf{x} \qquad (2.90)$$

This is a form estimable from short simulation data, which the next section will explain. However, in the case of diffusion processes, we can go a step further using the differential form of the generator. The trick will be to transform the $\mathscr{T}^{\Delta t}$s inside the integral into $\mathscr{L}$s, whose form we know explicitly. By subtracting $\mathbb{1}_{S^c}F^+_\Gamma$ from the first term, adding the same thing to the second

term, and bringing $1/\Delta t$ inside the integral, we get

$$R_\Gamma(\lambda) = \lim_{\Delta t \to 0} \int \pi F_\Gamma^- \left(\frac{\mathscr{T}^{\Delta t} - 1}{\Delta t} \left[\mathbb{1}_{S^c} F_\Gamma^+ \right] - \mathbb{1}_{S^c} \frac{\mathscr{T}^{\Delta t} - 1}{\Delta t} F_\Gamma^+ \right) d\mathbf{x} \tag{2.91}$$

It is tempting to take the limit $\Delta t \to 0$ inside the integral, taking the formal limit $(\mathscr{T}^{\Delta t} - 1)/\Delta t \to \mathscr{L}$ as $\Delta t \to 0$, but $\mathscr{L}$ cannot act on discontinuous functions like indicators. Instead, we introduce a smooth mollifier ϕ_δ with parameter $\delta > 0$, such that ϕ_δ is zero deep in S and one deep in S^c, but varies smoothly from zero to one through a transition region around ∂S of thickness δ. This means $\phi_\delta \to \mathbb{1}_{S^c}$ as $\delta \to 0$. Then we switch the order of limits and take the $\Delta t \to 0$ first, giving

$$R_\Gamma(\lambda) = \lim_{\delta \to 0} \int \pi F_\Gamma^- \left\{ \mathscr{L} \left[\phi_\delta F_\Gamma^+ \right] - \phi_\delta \mathscr{L} F_\Gamma^+ \right\} d\mathbf{x} \tag{2.92}$$

Now let us insert the specific form of the generator. The generator acts on on a product of functions fg as follows:

$$\begin{aligned} \mathscr{L}\left[fg\right] &= v_i \partial_i (fg) + a_{ij} \partial_i \partial_j (fg) && (2.93) \\ &= f \partial_i g + g \partial_i f + a_{ij} \left[f \partial_i \partial_j g + \partial_i f \partial_j g + \partial_j f \partial_i g + g \partial_i \partial_j f \right] && (2.94) \\ &= f \mathscr{L} g + g \mathscr{L} f + 2 a_{ij} \partial_i f \partial_j g && (2.95) \end{aligned}$$

where we've used the fact that $\mathbf{a}$ is symmetric to combine two terms. Applying this to the integrand,

$$R_\Gamma(\lambda) = \lim_{\delta \to 0} \int \pi F_\Gamma^- \left\{ F_\Gamma^+ \mathscr{L} \phi_\delta + 2 a_{ij} \partial_i \phi_\delta \partial_j F_\Gamma^+ \right\} d\mathbf{x} \tag{2.96}$$

Our goal is now to isolate $\partial_i \phi_\delta$ as a global coefficient for the rest of the integrand. We do this by

expanding $\mathscr{L}\phi_\delta$ and integrating by parts.

$$R_\Gamma(\lambda) = \lim_{\delta\to 0}\int \pi F_\Gamma^- \left\{ F_\Gamma^+ \nu_i \partial_i \phi_\delta + F_\Gamma^+ a_{ij}\partial_i\partial_j\phi_\delta + 2a_{ij}\partial_i\phi_\delta \partial_j F_\Gamma^+ \right\} d\mathbf{x} \tag{2.97}$$

$$= \lim_{\delta\to 0}\left(\int \partial_i\phi_\delta \left[\pi F_\Gamma^- F_\Gamma^+ \nu_i + 2\pi a_{ij} F_\Gamma^- \partial_j F_\Gamma^+ - \partial_j(\pi a_{ij} F_\Gamma^- F_\Gamma^+) \right] d\mathbf{x} \right. \tag{2.98}$$

$$\left. + \int \partial_j \left[\pi F_\Gamma^- F_\Gamma^+ a_{ij}\partial_i\phi_\delta \right] d\mathbf{x} \right) \tag{2.99}$$

We are assuming π vanishes fast enough at the boundaries so that the final term, an integral of a total derivative, is zero. The first integrand, meanwhile, can be manipulated into a more symmetric expression:

$$R_\Gamma(\lambda) = \lim_{\delta\to 0}\int \partial_i\phi_\delta \left[\pi F_\Gamma^- F_\Gamma^+ \nu_i + 2\pi a_{ij} F_\Gamma^- \partial_j F_\Gamma^+ - \pi a_{ij} F_\Gamma^- \partial_j F_\Gamma^+ - F_\Gamma^+ \partial_j(\pi a_{ij} F_\Gamma^-) \right] d\mathbf{x} \tag{2.100}$$

$$= \lim_{\delta\to 0}\int \partial_i\phi_\delta \left\{ F_\Gamma^- F_\Gamma^+ \left[\pi\nu_i - \partial_i(\pi a_{ij}) \right] + \pi a_{ij}\left(F_\Gamma^- \partial_j F_\Gamma^+ - F_\Gamma^+ \partial_j F_\Gamma^- \right) \right\} d\mathbf{x} \tag{2.101}$$

$$= \lim_{\delta\to 0}\int \nabla\phi_\delta \cdot \left\{ F_\Gamma^- F_\Gamma^+ \mathbf{J} + \pi\mathbf{a}(F_\Gamma^- \nabla F_\Gamma^+ - F_\Gamma^+ \nabla F_\Gamma^-) \right\} d\mathbf{x} \tag{2.102}$$

We have introduced the *equilibrium current* $\mathbf{J} = \pi\nu - \nabla\cdot(\pi\mathbf{a})$, which is divergence-free at steady-state and zero in systems with detailed balance (which we are not assuming). Let us then define the full term in braces $\mathbf{J}_{AB}$, and finally dispense with ϕ_δ by the following argument. As $\delta \to 0$, $\nabla\phi_\delta = |\nabla\phi_\delta|\mathbf{n}$, where $\mathbf{n}$ is the outward unit normal from S, pointing away from A and towards B. The volume element $d\mathbf{x}$ can be locally decomposed as $d\mathbf{x} = dx_\mathbf{n}\, d\sigma$, where $dx_\mathbf{n}$ is the element of length along the direction of $\mathbf{n}$ and $d\sigma$ is the element of surface area on S. We accordingly decompose the integral, and treat $\mathbf{J}_{AB}$ as constant over that vanishingly small region of size δ where ϕ_δ varies. By construction, the inner integral is $\int |\nabla\phi_\delta|\, dx_\mathbf{n} = 1$. All that remains is a $(d-1)$-

dimensional integral over a manifold, the surface ∂S:

$$R_\Gamma(\lambda) = \int_{\partial S} \mathbf{J}_{AB} \cdot \mathbf{n} \, d\sigma \tag{2.103}$$

We have derived an explicit form for $\mathbf{J}_{AB}$ for diffusion processes, but a more general definition of $\mathbf{J}_{AB}$ is the implicit definition in Eq.(2.103). For $\lambda = 0$, $\mathbf{J}_{AB}$ is the vector field whose surface integral on ∂S gives the rate. We could also take derivatives to find a generalized current contributing to the moments of transition path integrals. For example, using the product rule,

$$\partial_\lambda R_\Gamma(\lambda) = \int_{\partial S} \Big\{ \Big(\partial_\lambda F_\Gamma^- F_\Gamma^+ + F_\Gamma^- \partial_\lambda F_\Gamma^+ \Big) \mathbf{J} \tag{2.104}$$

$$+ \pi \mathbf{a} \Big(\partial_\lambda F_\Gamma^- \nabla F_\Gamma^+ + F_\Gamma^- \nabla \partial_\lambda F_\Gamma^+ - \partial_\lambda F_\Gamma^+ \nabla F_\Gamma^- - F_\Gamma^+ \nabla \partial_\lambda F_\Gamma^- \Big) \Big\} \cdot \mathbf{n} \, d\mathbf{x} \tag{2.105}$$

Note that the integral does *not* depend on the specific dividing surface we choose, which implies that $\mathbf{J}_{AB}$ is divergence-free outside of $A \cup B$, but has a source of field lines at A and a sink at B. Since every dividing surface supports the same total flux, large local current magnitude means a constrained reaction mechanism.

We have now completely described the mathematics of TPT, and our extensions to it. All the above quantities of interest can be computed from $\pi(\mathbf{x})$, $F_\Gamma^+(\mathbf{x})$, $F_\Gamma^-(\mathbf{x})$, and their λ-derivatives. In chapter 4, we solve the Feynman-Kac PDEs numerically with a finite volume method, which is possible because the state space is only three-dimensional. In chapters 5 and 6, however, I study a 75-dimensional model that cannot be handled the same way due to the curse of dimensionality. The next section explains both how to compute them from data instead, making use of the probabilistic definitions of the generator and the transfer operator.

2.4 Numerical method: dynamical Galerkin approximation (DGA)

The DGA method [Thiede et al., 2019] is an approach to solving the Feynman-Kac formulae, and subsequently estimating the TPT quantities of interest, using a dataset of short trajectories:

$$\{\mathbf{X}_n(t) : 0 \leq t \leq \Delta t\}_{n=1}^{N}, \tag{2.106}$$

where the initial conditions $\mathbf{X}_n(0)$ have been sampled from all over state space. One can use many different procedures to do so, including stratification and splitting methods, but this thesis uses only very straightforward schemes. We define a *sampling measure* $\mu(\mathbf{x})$ to encapsulate the procedure of generating $\mathbf{X}_n(0)$, and take it as given in the development below. μ is very flexible; it just has to be absolutely continuous with respect to π.

2.4.1 Discretization of Feynman-Kac formulae

The forecast $F_\Gamma^+(\mathbf{x})$, the steady-state density $\pi(\mathbf{x})$, and the aftcast $F_\Gamma^-(\mathbf{x})$ call for three similar but distinct algorithms, which we address in turn. Each step requires expanding an unknown function in a pre-defined basis, which can be quite general. However, I ended up using only locally supported indicator functions as basis elements, which allow us to interpret DGA as an application of Markov state modeling (MSM). I will spell out this special case in parallel with the more general case.

DGA for forecasts

First we address the forecast, which solve the boundary value problem from Eq. (2.32):

$$\begin{cases} (\mathscr{T}^{\theta} - 1 + \lambda \mathscr{K}_{\Gamma}^{\theta}) F_{\Gamma}^{+}(\mathbf{x}) = 0 & \mathbf{x} \in D \\ F_{\Gamma}^{+}(\mathbf{x}) = G(\mathbf{x}) & \mathbf{x} \in D^{\mathrm{c}} \end{cases} \tag{2.107}$$

where for any suitable function ϕ, (2.108)

$$\theta := \min(\Delta t, \tau_{D^{\mathrm{c}}}^{+}) \tag{2.109}$$

$$\mathscr{T}^{\theta} \phi(\mathbf{x}) := \mathbb{E}_{\mathbf{x}}\big[\phi(\mathbf{X}(\theta))\big] \tag{2.110}$$

$$\mathscr{K}_{\Gamma}^{\theta} \phi(\mathbf{x}) := \mathbb{E}_{\mathbf{x}}\left[\int_0^{\theta} \Gamma(\mathbf{X}(t))\phi(\mathbf{X}(t))\, dt\right] \tag{2.111}$$

Note that the operator acting on F_{Γ}^{+} is linear. To discretize this equation and impose regularity on the solution, we approximate F_{Γ}^{+} as a finite linear combination with coefficients $w_j(F_{\Gamma}^{+}(\mathbf{x};\lambda))$, abbreviated $w_j(\lambda)$:

$$F_{\Gamma}^{+}(\mathbf{x};\lambda) \approx \hat{F}_{\Gamma}^{+}(\mathbf{x};\lambda) + \sum_{j=1}^{M} w_j(\lambda)\phi_j(\mathbf{x};\lambda) \tag{2.112}$$

where $\hat{F}_{\Gamma}^{+}$ is a guess function obeying the appropriate boundary conditions ($\hat{F}_{\Gamma}^{+}|_{D^{\mathrm{c}}} = G$) and $\{\phi_j\}_{j=1}^{M}$ is a collection of basis functions that are zero on D^{c}. The basis functions can be defined in many ways, including with nonlinear dimensionality reduction and machine-learned features [Thiede et al., 2019, Strahan et al., 2021]. In this thesis, we restrict to simple indicator functions, rendering the method very similar to a Markov state model. The performance of DGA hinges on the expressive capacity of the basis functions.

The task is now to solve for the coefficients $w_j(\lambda)$. Equation (2.107) becomes a system of

linear equations in $w_j(\lambda)$:

$$\sum_{j=1}^{M} w_j(\lambda)(\mathscr{T}^\theta - 1 + \lambda\mathscr{K}_\Gamma^\theta)\phi_j(\mathbf{x};\lambda) = -(\mathscr{T}^\theta - 1 + \lambda\mathscr{K}_\Gamma^\theta)\hat{F}_\Gamma^+(\mathbf{x};\lambda) \tag{2.113}$$

Since the operators $\mathscr{T}^\theta, \mathscr{K}_\Gamma^\theta$ produce expectations over the future state of the system beginning at $\mathbf{x}$, we can estimate their action at $\mathbf{x} = \mathbf{X}_n(0)$ (a short-trajectory starting point) as

$$\begin{aligned}(\mathscr{T}^\theta - 1 + \lambda\mathscr{K}_\Gamma^\theta)\phi_j(\mathbf{X}_n(0);\lambda) \approx \phi_j(\mathbf{X}_n(\theta_n);\lambda) - \phi_j(\mathbf{X}_n(0);\lambda) \\ + \lambda\int_0^{\theta_n} \Gamma(\mathbf{X}_n(t))\phi_j(\mathbf{X}_n(t);\lambda)\,dt\end{aligned} \tag{2.114}$$

or, if multiple independent trajectories are launched from $\mathbf{x}$, we can average over them. Here, $\theta_n = \min(\Delta t, \tau^+_{D^c,n})$ is the nth sample realization of the stopping time θ. In other words, if a trajectory hits the boundary D^c before its prescribed duration Δt, we consider it stopped there. On the right-hand side, the integral can be approximated by any quadrature method, which will become more accurate with sampling frequency.

Applying this to every short trajectory and plugging into Eq. (2.113), we obtain a system of N equations in M unknowns. In practice, $N \gg M$, meaning we have many more trajectories than basis functions, and the system is overdetermined. A unique, and regularized, solution is obtained by casting it into weak form: we multiply both sides by $\phi_i(\mathbf{x})$ and integrate over state space:

$$\sum_{j=1}^{M} w_j(\lambda)\left\langle \phi_i, (\mathscr{T}^\theta - 1 + \lambda\mathscr{K}_\Gamma^\theta)\phi_j \right\rangle_\zeta = -\left\langle \phi_i, (\mathscr{T}^\theta - 1 + \lambda\mathscr{K}_\Gamma^\theta)\hat{F}_\Gamma^+ \right\rangle_\zeta \tag{2.115}$$

where the inner products are defined with respect to a measure ζ:

$$\langle f, g\rangle_\zeta = \int f(\mathbf{x})g(\mathbf{x})\zeta(\mathbf{x})\,d\mathbf{x} \tag{2.116}$$

With our finite data set, we approximate the inner product by a sum over pairs of points. Given that

$\mathbf{X}_n(0) \sim \mu$, the law of large numbers ensures that for any bounded function $H(\mathbf{x})$,

$$\frac{1}{N}\sum_{n=1}^{N} H(\mathbf{X}_n(0)) \approx \int H(\mathbf{x})\mu(\mathbf{x})\,d\mathbf{x} \tag{2.117}$$

becomes more accurate as $N \to \infty$. Thus we set $H(\mathbf{x}) = \phi_i(\mathbf{x})(\mathscr{T}^\theta - 1 + \lambda\mathscr{K}_\Gamma^\theta)\phi_j(\mathbf{x})$ as estimated by (2.114), approximate the inner products with $\zeta = \mu$, plug them into (2.115), and solve the $M \times M$ system of linear equations for $w_j(\lambda)$.

With the inner products in hand, we now have (2.115) as a family of matrix equations with λ a continuous parameter:

$$(P + \lambda Q)\xi(\lambda) = \mathbf{v} + \lambda\mathbf{r}, \tag{2.118}$$

where $P_{ij} = \langle\phi_i, (\mathscr{T}^\theta - 1)\phi_j\rangle_\zeta$, $Q_{ij} = \langle\phi_i, \mathscr{K}_\Gamma^\theta\phi_j\rangle_\zeta$, $v_i = \langle\phi_i, (\mathscr{T}^\theta - 1)\hat{F}_\Gamma^+\rangle_\zeta$, and $r_i = \langle\phi_i, \mathscr{K}_\Gamma^\theta\hat{F}_\Gamma^+\rangle_\zeta$. We can then differentiate in λ and evaluate at $\lambda = 0$ to obtain a ready-to-solve discretization of the recursion (2.29):

$$P\xi(0) = \mathbf{v} \tag{2.119}$$

$$P\xi'(0) = \mathbf{r} - Q\xi(0) \tag{2.120}$$

$$P\xi^{(k)}(0) = -kQ\xi^{(k-1)}(0) \text{ for } k \geq 2 \tag{2.121}$$

where the k'th derivative $\xi^{(k)}(0)$ is the coefficient expansion in the basis $\{\phi_j\}$ of the k'th moment from (5.18):

$$\partial_\lambda^k F_\Gamma^+(\mathbf{x};0) = \mathbb{E}_\mathbf{x}\left[G(\mathbf{X}(\tau_{D^c}^+))\left(\lambda\int_0^{\tau_{D^c}^+}\Gamma(\mathbf{X}(s))\,ds\right)^k\right] \tag{2.122}$$

It is helpful to write the equations more explicitly in the special case of the forward committor (where $D = (A \cup B)^c$, $\Gamma = 0$, and $G = \hat{F}_\Gamma^+ = \mathbb{1}_B$), and where the basis functions are indicators. To

construct this basis, we partition D into a disjoint collection of sets $\{S_j\}_{j=1}^M$, which could come from a regular grid in low dimensions or a clustering algorithm in high dimensions. We define $\phi_j(\mathbf{x}) = \mathbb{1}_{S_j}(\mathbf{x})$, i.e., one if $\mathbf{x} \in S_j$ and 0 otherwise. Note that $S_1, \ldots, S_M$ are all subsets of D and so disjoint with $A \cup B$. The matrix elements on the left-hand side above then become

$$\langle \phi_i, (\mathscr{T}^\theta - 1)\phi_j \rangle_\mu \approx \frac{1}{N} \sum_{n=1}^{N} \phi_i(\mathbf{X}_n(0)) \big[\phi_j(\mathbf{X}_n(\theta_n)) - \phi_j(\mathbf{X}_n(0)) \big] \tag{2.123}$$

$$= \frac{\#\{n : \mathbf{X}_n(0) \in S_i, \mathbf{X}_n(\theta_n) \in S_j\} - \#\{n : \mathbf{X}_n(0) \in S_i, \mathbf{X}_n(0) \in S_j\}}{N} \tag{2.124}$$

$$= \frac{N_{ij} - N_i \delta_{ij}}{N} \tag{2.125}$$

where N_{ij} is defined as the number of transitions from S_i at the beginning of the trajectory to S_j at the end of the trajectory, and N_i is the number of trajectories starting in S_i. The right-hand side then becomes

$$-\langle \phi_i, (\mathscr{T}^\theta - 1)\mathbb{1}_B \rangle_\mu \approx -\frac{1}{N} \sum_{n=1}^{N} \phi_i(\mathbf{X}_n(0)) \big[\mathbb{1}_B(\mathbf{X}_n(\theta_n)) - \mathbb{1}_B(\mathbf{X}_n(0)) \big] \tag{2.126}$$

$$= -\frac{\#\{n : \mathbf{X}_n(0) \in S_i, \mathbf{X}_n(\theta_n) \in B\} - \#\{\mathbf{X}_n(0) \in S_i, \mathbf{X}_n(0) \in B\}}{N} \tag{2.127}$$

$$= -\frac{N_{iB}}{N} \tag{2.128}$$

where in a slight abuse of notation, N_{iB} counts the transitions from S_i at the beginning of the trajectory to B at the end. The second term in the numerator is zero, as the sets are all disjoint from B and hence no $\mathbf{X}_n(0)$ can be simultaneously in S_i and B. Equating the two sides of the equation, multiplying by N/N_i, and rearranging, we have

$$\frac{1}{N} \sum_{j=1}^{M} w_j(q_B^+)(N_{ij} - N_i \delta_{ij}) = -\frac{N_{iB}}{N} \tag{2.129}$$

$$\sum_{j=1}^{M} w_j(q_B^+) \left(\frac{N_{ij}}{N_i} - \delta_{ij} \right) = -\frac{N_{iB}}{N_i} \tag{2.130}$$

We can now see a connection with Markov state modeling: N_{ij}/N_i is the maximum-likelihoood estimate of the Markov transition matrix entry from S_i to S_j, i.e., $\mathbb{P}\{\mathbf{X}(\theta) \in S_j | \mathbf{X}(0) \in S_i\}$. Likewise, N_{iB}/N_i is the estimate of the transition probability from S_i to B. Therefore we label these ratios P^{θ}_{ij} and P^{θ}_{iB} respectively, and write

$$\sum_{j=1}^{M} w_j(q_B^+)(P^{\theta} - I)_{ij} = -P^{\theta}_{iB} \tag{2.131}$$

Rearranging gives a very intuitive equation for the committor:

$$w_i(q_B^+) = P^{\theta}_{iB} + \sum_{j=1}^{M} P^{\theta}_{ij} w_j(q_B^+) \tag{2.132}$$

In words, going from set i to B before A could happen either in a single trajectory's lifetime (with probability P^{θ}_{iB}) or through some intermediate state first. The probability can be decomposed into all possible steps from (i, right now) to (j, in one timestep) and subsequently from j to B. Hence the recursive nature of the equation.

We should think of P^{θ} as a $(M+2) \times (M+2)$ Markov matrix, where the first M states correspond to the M basis functions, and A and B the last two states:

$$P^{\theta} = \begin{bmatrix} P^{\theta}_{11} & \cdots & P^{\theta}_{1M} & P^{\theta}_{1A} & P^{\theta}_{1B} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ P^{\theta}_{M1} & \cdots & P^{\theta}_{MM} & P^{\theta}_{MA} & P^{\theta}_{MB} \\ 0 & \ldots & 0 & 1 & 0 \\ 0 & \ldots & 0 & 0 & 1 \end{bmatrix} \tag{2.133}$$

Thus, Eq. (2.132) is really an eigenvalue equation of P^{θ}. Since P^{θ} is a properly normalized stochastic matrix, with all entries nonnegative and all rows summing to 1, each $w_i(q_B^+)$ is guaranteed to be a *bona fide* probability between 0 and 1.

DGA for steady-state density

To solve for the steady-state density, we will return to the weak form of the stationary Fokker-Planck equation (2.43), and set $\zeta = \mu$:

$$\int (\mathscr{T}^{\Delta t} - 1) H(\mathbf{x}) \frac{d\pi}{d\mu}(\mathbf{x}) \mu(\mathbf{x})\, d\mathbf{x} = 0 \text{ for all } H \tag{2.134}$$

Once again, we expand the unknown function $\frac{d\pi}{d\mu}(\mathbf{x})$ in the basis $\{\phi_j\}_{j=1}^M$ (although it could be a different basis from the one used for F_Γ^+) and enforce the above for each $H = \phi_i$:

$$\frac{d\pi}{d\mu}(\mathbf{x}) \approx \sum_{j=1}^M w_j \left(\frac{d\pi}{d\mu}\right) \phi_j(\mathbf{x}) \tag{2.135}$$

$$\int (\mathscr{T}^{\Delta t} - 1) \phi_i(\mathbf{x}) \frac{d\pi}{d\mu}(\mathbf{x}) \mu(\mathbf{x})\, d\mathbf{x} \approx \sum_{j=1}^M w_j \left(\frac{d\pi}{d\mu}\right) \int (\mathscr{T}^{\Delta t} - 1) \phi_i(\mathbf{x}) \phi_j(\mathbf{x}) \mu(\mathbf{x})\, d\mathbf{x} \tag{2.136}$$

$$0 = \sum_{j=1}^M w_j \left(\frac{d\pi}{d\mu}\right) \left\langle (\mathscr{T}^{\Delta t} - 1) \phi_i, \phi_j \right\rangle_\mu \tag{2.137}$$

This is a homogeneous linear system, with matrix entries on the right-hand side estimated by Monte Carlo:

$$\langle (\mathscr{T}^{\Delta t} - 1) \phi_i, \phi_j \rangle_\mu \approx \frac{1}{N} \sum_{n=1}^N \Big[\phi_i(\mathbf{X}_n(\Delta t)) - \phi_i(\mathbf{X}_n(0)) \Big] \phi_j(\mathbf{X}_n(0)) \tag{2.138}$$

We solve this homogeneous system by *QR* decomposition. Note that there are no boundary conditions, and the trajectories need not be stopped early. Instead there is a normalization condition, which we enforce as $\sum_{n=1}^N \frac{d\pi}{d\mu}(\mathbf{X}_n(0)) = 1$. To ensure that the matrix has a nontrivial null vector, one can add a vector of constants. However, the basis of indicators that we use guarantees a null space automatically. Again, let's work out the equation in the case of indicators. The matrix

element is

$$\langle(\mathscr{T}^{\Delta t}-1)\phi_i,\phi_j\rangle_\mu \approx \frac{\#\{n:\mathbf{X}_n(\Delta t)\in S_i,\mathbf{X}_n(0)\in S_j\}-\#\{n:\mathbf{X}_n(0)\in S_i,\mathbf{X}_n(0)\in S_j\}}{N} \tag{2.139}$$

$$=\frac{N_{ji}}{N}-\frac{N_j\delta_{ji}}{N}=\frac{N_j}{N}(P^{\Delta t}-I)_{ji} \tag{2.140}$$

$$\therefore 0=\sum_{j=1}^{M}(P^{\Delta t}-I)_{ji}N_j w_j\left(\frac{d\pi}{d\mu}\right) \tag{2.141}$$

Apparently, $N_j w_j(d\pi/d\mu)$ is the jth entry of the invariant measure of a finite Markov chain, i.e., the left null eigenvector of $P^{\Delta t}$. This gives concrete meaning to the words *change of measure*:

$$w_j\left(\frac{d\pi}{d\mu}\right)=\frac{\text{Steady-state probability of } S_j}{\text{Number of initial points sampled in } S_j} \tag{2.142}$$

The change of measure is a very important tool for estimating TPT quantities. Given the weights $\frac{d\pi}{d\mu}(\mathbf{X}_n(0))$, we can take any ergodic average $\langle H\rangle_\pi$ by inserting the change of measure:

$$\langle H\rangle_\pi=\int_{\mathbb{R}^d}H(\mathbf{x})\pi(\mathbf{x})\,d\mathbf{x}=\int_{\mathbb{R}^d}H(\mathbf{x})\frac{d\pi}{d\mu}(\mathbf{x})\mu(\mathbf{x})\,d\mathbf{x}\approx\sum_{n=1}^{N}H(\mathbf{X}_n(0))\frac{d\pi}{d\mu}(\mathbf{X}_n(0)) \tag{2.143}$$

For the specific case of a Markov state model, we can decompose the sum into clusters and reduce to an intuitive formula.

$$\langle H\rangle_\pi\approx\sum_{n=1}^{N}H(\mathbf{X}_n(0))\frac{d\pi}{d\mu}(\mathbf{X}_n(0))\sum_{j=1}^{M}\mathbb{1}_{S_j}(\mathbf{X}_n(0)) \tag{2.144}$$

$$=\sum_{j=1}^{M}w_j\left(\frac{d\pi}{d\mu}\right)\sum_{n=1}^{N}H(\mathbf{X}_n(0))\mathbb{1}_{S_j}(\mathbf{X}_n(0)) \tag{2.145}$$

$$=\sum_{j=1}^{M}N_j w_j\left(\frac{d\pi}{d\mu}\right)\left(\frac{1}{N_j}\sum_{n:\mathbf{X}_n(0)\in S_j}H(\mathbf{X}_n(0))\right) \tag{2.146}$$

$$=\sum_{j=1}^{M}\Big(\text{Steady-state probability of } S_j\Big)\times\Big(\text{Average of } H \text{ over } S_j\Big) \tag{2.147}$$

DGA for aftcasts

Finally, we address the backward-in-time expectations, which means approximating the action of $\mathscr{T}^{-\theta}$ rather than $\mathscr{T}^{\theta}$. It is almost enough to simply reverse time on all trajectories, with $\mathbf{X}_n(\Delta t)$ becoming the trajectory's beginning, $\mathbf{X}_n(0)$ its end, and

$$\mathscr{T}^{-\theta}H(\mathbf{x}) = \mathbb{E}\big[H\big(\max(0,\tau^{-}_{D^{\mathrm{c}}}(\Delta t))\big)|\mathbf{X}(\Delta t)=\mathbf{x}\big]. \tag{2.148}$$

But there is one problem: $\mathbf{X}_n(\Delta t)$ is not distributed according to μ, and so we cannot use the same Monte Carlo inner product with a reference measure of $\zeta = \mu$. However, we can solve the problem by reweighting with the change of measure, leading to $\zeta = \pi$ instead. Because π is the stationary measure, reweighting so $\mathbf{X}(0) \sim \pi$ is equivalent to reweighting such that $\mathbf{X}(\Delta t) \sim \pi$. To derive this simply, let the trajectory be discrete in time, i.e.,

$$\mathbf{X} = \left[\mathbf{X}(0), \mathbf{X}\left(\frac{\Delta t}{K}\right), \mathbf{X}\left(\frac{2\Delta t}{K}\right), \ldots, \mathbf{X}(\Delta t)\right] \tag{2.149}$$

and consider functionals $\mathscr{H}[\mathbf{X}]$ of the whole trajectory. Defining the transition density $p(\mathbf{x},\mathbf{y})$ for each step of size Δt, the expectation of $\mathscr{H}$ with $\mathbf{X}(0) \sim \pi$ is given by

$$\mathbb{E}_{\mathbf{X}(0)\sim\pi}\mathscr{H}[\mathbf{X}] = \int d\mathbf{x}_0 \pi(\mathbf{x}_0) \int d\mathbf{x}_1 p(\mathbf{x}_0,\mathbf{x}_1) \int \ldots \int d\mathbf{x}_K p(\mathbf{x}_{K-1},\mathbf{x}_K)\mathscr{H}[\mathbf{x}_0,\ldots,\mathbf{x}_K] \tag{2.150}$$

The time reversal step explicitly assumes the *equilibrium* backward process, leading to a backward transition kernel $\widetilde{p}(\mathbf{y},\mathbf{x}) = \frac{\pi(\mathbf{x})}{\pi(\mathbf{y})}p(\mathbf{x},\mathbf{y})$. Inserting this throughout converts the expectation over $\mathbf{X}(0)$

into an expectation over $\mathbf{X}(\Delta t)$:

$$\mathbb{E}_{\mathbf{X}(0)\sim\pi}\mathscr{H}[\mathbf{X}] = \int d\mathbf{x}_0 \pi(\mathbf{x}_0) \int d\mathbf{x}_1 \frac{\pi(\mathbf{x}_1)}{\pi(\mathbf{x}_0)} \widetilde{p}(\mathbf{x}_1, \mathbf{x}_0) \int \tag{2.151}$$

$$\ldots \int d\mathbf{x}_K \frac{\pi(\mathbf{x}_K)}{\pi(\mathbf{x}_{K-1})} \widetilde{p}(\mathbf{x}_K, \mathbf{x}_{K-1}) \mathscr{H}[\mathbf{x}_0, \ldots, \mathbf{x}_K] \tag{2.152}$$

$$= \int d\mathbf{x}_K \pi(\mathbf{x}_K) \int d\mathbf{x}_{K-1} \widetilde{p}(\mathbf{x}_K, \mathbf{x}_{K-1}) \int \ldots \int d\mathbf{x}_0 \widetilde{p}(\mathbf{x}_1, \mathbf{x}_0) \mathscr{H}[\mathbf{x}_0, \ldots, \mathbf{x}_K] \tag{2.153}$$

$$= \widetilde{\mathbb{E}}_{\mathbf{X}(\Delta t)\sim\pi}\mathscr{H}[\mathbf{X}] \tag{2.154}$$

where $\widetilde{\mathbb{E}}$ denotes backward-in-time expectation. This is precisely what we need to apply (2.115) to the time-reversed process, namely, define $\mathscr{H}$ as

$$\mathscr{H}[\mathbf{X}] := \phi_i(\mathbf{X}(\Delta t))(\mathscr{T}^{-\theta} - 1 + \lambda \mathscr{K}_\Gamma^{-\theta})\phi_j(\mathbf{X}(\Delta t)) \tag{2.155}$$

and then integrate over state space weighted by π, turning the right-hand side into an inner product:

$$\langle \phi_i, (\mathscr{T}^{-\theta} - 1 + \lambda \widetilde{\mathscr{K}_\Gamma^{-\theta}})\phi_j \rangle_\pi = \widetilde{\mathbb{E}}_{\mathbf{X}(\Delta t)\sim\pi}\mathscr{H}[\mathbf{X}] \tag{2.156}$$

$$= \mathbb{E}_{\mathbf{X}(0)\sim\pi}\mathscr{H}[\mathbf{X}] \approx \sum_{n=1}^{N} \mathscr{H}[\mathbf{X}_n] \frac{d\pi}{d\mu}(\mathbf{X}_n(0)) \tag{2.157}$$

The right-hand side of Eq. (2.115) can be estimated similarly, also with $\zeta = \pi$.

In both forward- and backward-time estimates, we never solve for $F_\Gamma^+(\mathbf{x};\lambda)$ or $F_\Gamma^-(\mathbf{x};\lambda)$ with nonzero λ; rather, we repeat the recursion process with Eq. (2.29). This is equivalent to implicitly differentiating the discretized system Eq. (2.115).

Once again, to connect with the simpler theory of Markov chains, let's specialize to the problem of the backward committor q_A^- with indicators as basis functions. The matrix elements are

$$\langle \phi_i, (\mathscr{T}^{-\theta} - 1)\phi_j \rangle_\pi \approx \sum_{n=1}^{N} \frac{d\pi}{d\mu}(\mathbf{X}_n(0)) \mathbb{1}_{S_i}(\mathbf{X}_n(\Delta t)) \left[\mathbb{1}_{S_j}\left(\mathbf{X}_n(\widetilde{\theta}_n)\right) - \mathbb{1}_{S_j}(\mathbf{X}_n(\Delta t)) \right] \tag{2.158}$$

where $\widetilde{\theta}_n = \max(0, \tau^-_{D^c,n}(\Delta t))$. The source term is

$$-\langle \phi_i, (\mathscr{T}^{-\theta} - 1)\mathbb{1}_A \rangle_\pi \approx -\sum_{n=1}^{N} \frac{d\pi}{d\mu}(\mathbf{X}_n(0))\mathbb{1}_{S_i}(\mathbf{X}_n(\Delta t))\left[\mathbb{1}_A(\mathbf{X}_n(\widetilde{\theta}_n)) - \mathbb{1}_A(\mathbf{X}_n(\Delta t))\right] \tag{2.159}$$

The second term is zero, since all $\mathbb{1}_{S_i}$s are supported strictly outside $A \cup B$. The matrix equation is then

$$\sum_{j=1}^{M} w_j(q_A^-) \sum_{n=1}^{N} \frac{d\pi}{d\mu}(\mathbf{X}_n(0))\mathbb{1}_{S_i}(\mathbf{X}_n(\Delta t))\left[\mathbb{1}_{S_j}(\mathbf{X}_n(\widetilde{\theta}_n)) - \mathbb{1}_{S_j}(\mathbf{X}_n(\Delta t))\right] \tag{2.160}$$

$$= -\sum_{n=1}^{N} \frac{d\pi}{d\mu}(\mathbf{X}_n(0))\mathbb{1}_{S_i}(\mathbf{X}_n(\Delta t))\mathbb{1}_A(\mathbf{X}_n(\widetilde{\theta}_n)) \tag{2.161}$$

Or, isolating the δ_{ij} term on one side,

$$w_i(q_A^-) \sum_{n=1}^{N} \frac{d\pi}{d\mu}(\mathbf{X}_n(0))\mathbb{1}_{S_i}(\mathbf{X}_n(\Delta t)) = \tag{2.162}$$

$$\sum_{j=1}^{M} w_j(q_A^-) \sum_{n=1}^{N} \frac{d\pi}{d\mu}(\mathbf{X}_n(0))\mathbb{1}_{S_i}(\mathbf{X}_n(\Delta t))\mathbb{1}_{S_j}(\mathbf{X}_n(\widetilde{\theta}_n)) \tag{2.163}$$

$$+\sum_{n=1}^{N} \frac{d\pi}{d\mu}(\mathbf{X}_n(0))\mathbb{1}_{S_i}(\mathbf{X}_n(\Delta t))\mathbb{1}_A(\mathbf{X}_n(\widetilde{\theta}_n)) \tag{2.164}$$

Dividing through by the coefficient of $w_i(q_A^-)$ on the left,

$$w_i(q_A^-) = \sum_{j=1}^{M} w_j(q_A^-) \frac{\sum_{n=1}^{N} \frac{d\pi}{d\mu}(\mathbf{X}_n(0))\mathbb{1}_{S_i}(\mathbf{X}_n(\Delta t))\mathbb{1}_{S_j}(\mathbf{X}_n(\widetilde{\theta}_n))}{\sum_{n=1}^{N} \frac{d\pi}{d\mu}(\mathbf{X}_n(0))\mathbb{1}_{S_i}(\mathbf{X}_n(\Delta t))} \tag{2.165}$$

$$+\frac{\sum_{n=1}^{N} \frac{d\pi}{d\mu}(\mathbf{X}_n(0))\mathbb{1}_{S_i}(\mathbf{X}_n(\Delta t))\mathbb{1}_A(\mathbf{X}_n(\widetilde{\theta}_n))}{\sum_{n=1}^{N} \frac{d\pi}{d\mu}(\mathbf{X}_n(0))\mathbb{1}_{S_i}(\mathbf{X}_n(\Delta t))} \tag{2.166}$$

Finally, we have on the right-hand side the entries of a $(M+2)\times(M+2)$ Markov transition matrix. It is properly normalized: summing the right-hand coefficients over the possibilities $j = 1, \ldots, M$ as well as A and B, the rows sum to 1. Hence q_A^-, like q_B^+, is guaranteed to be a probability for the

Markov state model.

2.4.2 Rate estimate and numerical benchmarking

To estimate generalized rates (in particular, the ordinary rate), we reproduce here the rate estimate from Strahan et al. [2021], which is an almost-direct implementation of the formula (2.90), repeated here:

$$R_\Gamma(\lambda) = \lim_{\Delta t\to 0}\frac{1}{\Delta t}\int_{\mathbb{R}^d} F_\Gamma^-(\mathbf{x};\lambda)\Big\{\mathscr{T}^{\Delta t}\big[\mathbb{1}_{S^c}F_\Gamma^+\big] - \mathbb{1}_{S^c}\mathscr{T}^{\Delta t}F_\Gamma^+\Big\}(\mathbf{x};\lambda)\pi(\mathbf{x})\,d\mathbf{x} \tag{2.167}$$

In principle, the integral could be estimated directly with any choice of dividing surface S, but the sum would only use the very small fraction of data either exiting S or entering S. We can use all the data at once and improve numerical stability by averaging over multiple such surfaces. We first replace $\mathbb{1}_{S^c}$ with a smooth function (on D), as follows.

Let $K:\mathbb{R}^d\to[0,1]$ be a function that increases from 0 on set A to 1 on set B (for instance, the committor). Let $S_\zeta = \{\mathbf{x} : K(\mathbf{x})\le\zeta\}$ for $\zeta\in(0,1)$, and integrate both sides over ζ, noting that $\int_0^1 \mathbb{1}_{S_\zeta^c}(\mathbf{x})\,d\zeta = \int_0^1 \mathbb{1}\{K(\mathbf{x})>\zeta\}\,d\zeta = K(\mathbf{x})$.

$$\int_0^1 R_\Gamma(\lambda)\,d\zeta = \lim_{\Delta t\to 0}\int_{\mathbb{R}^d} F_\Gamma^-(\mathbf{x};\lambda)\left\{\frac{\mathscr{T}^{\Delta t}-1}{\Delta t}\big[KF_\Gamma^+\big] - K\mathscr{L}F_\Gamma^+\right\}(\mathbf{x})\pi(\mathbf{x})\,d\mathbf{x} \tag{2.168}$$

Now we can move the limit inside and use the PDE to find

$$R_\Gamma(\lambda) = \int_{\mathbb{R}^d} F_\Gamma^-(\mathbf{x};\lambda)\big\{\mathscr{L}[KF_\Gamma^+](\mathbf{x}) + \lambda K(\mathbf{x})\Gamma(\mathbf{x})F_\Gamma^+(\mathbf{x})\big]\big\}\pi(\mathbf{x})\,d\mathbf{x} \tag{2.169}$$

This formula can be estimated directly from knowledge of F_Γ^-,F_Γ^+, and π, using the ergodic assumption and with a discrete finite difference in time to estimate $\mathscr{L}[KF_\Gamma^+]$, i.e.,

$$\mathscr{L}[KF_\Gamma^+](\mathbf{X}_n(0)) \approx \frac{K(\mathbf{X}_n(\Delta t))F_\Gamma^+(\mathbf{X}_n(\Delta t)) - K(\mathbf{X}_n(0))F_\Gamma^+(\mathbf{X}_n(0))}{\Delta t} \tag{2.170}$$

Generalized rates are obtained by differentiating $R_\Gamma(\lambda)$ with respect to λ. Assuming we already know derivatives of F_Γ^+ and F_Γ^-, we can simply iterate the product rule.

2.4.3 Visualization method

Visualization is an essential step in interpreting and diagnosing the results of DGA. Chapters Chapters 5-6 contain several one- and two-dimensional projections. Because we can only plot in one or two dimensions at a time, it is critical to average out the remaining dimensions in a statistically consistent way. We do so with the following procedure.

Let $\mathbf{y} = \mathbf{Y}(\mathbf{x})$ be an observable subspace, typically with dimension much less than that of $\mathbf{x}$ (usually two). Any scalar field $F(\mathbf{x})$, such as the committor, has a projection $F^{\mathbf{Y}}(\mathbf{y})$ onto this subspace by

$$F^{\mathbf{Y}}(\mathbf{y}) = \int F(\mathbf{x})\pi(\mathbf{x})\delta(\mathbf{Y}(\mathbf{x}) - \mathbf{y})\, d\mathbf{x} \tag{2.171}$$

In practice, the $\mathbf{y}$ space is partitioned into grid boxes $d\mathbf{y}$, and the integral is estimated from the dataset, yielding

$$F^{\mathbf{Y}}(\mathbf{y}) = \frac{1}{N}\sum_{n=1}^{N} F(\mathbf{X}_n(0))\frac{d\pi}{d\mu}(\mathbf{X}_n(0))\mathbb{1}_{d\mathbf{y}}(\mathbf{Y}(\mathbf{X}_n(0))) \tag{2.172}$$

where $\mathbb{1}_{d\mathbf{y}}(\mathbf{Y}(\mathbf{x})) = 1$ if $\mathbf{Y}(\mathbf{x}) \in d\mathbf{y}$ and zero otherwise. In words, we simply take a weighted average over all data points $\mathbf{X}_n(0)$ that project onto the grid box $d\mathbf{y}$, with weights given by the change of measure. In Fig. 7.6, we use q_B^+ and q_A^- for F. In Fig. 6.4, we use $F = \pi$ (a), $F = q_A^- q_A^+$ (b), $F = q_A^- q_B^+$ (c), $F = q_B^- q_A^+$ (d), and $F = q_B^- q_B^+$ (e) to generate the background colors.

To display overlaid vector fields such as the reactive current requires a more involved formula. We use the exact same reactive current formula as in the supplement of Strahan et al. [2021], but

repeat it here for reference. The projected current is defined as

$$\mathbf{J}_{AB}^{\mathbf{Y}}(\mathbf{y}) = \int \mathbf{J}_{AB}(\mathbf{x}) \cdot \nabla \mathbf{Y}(\mathbf{x}) \delta\big(\mathbf{Y}(\mathbf{x}) - \mathbf{y}\big)\, d\mathbf{x} \tag{2.173}$$

In the discretized $\mathbf{y}$ space, this leads to the discretized projected current:

$$\mathbf{J}_{AB}^{\mathbf{Y}}(\mathbf{y}) \approx \frac{1}{2\Delta t} \sum_{n=1}^{N} \frac{d\pi}{d\mu}(\mathbf{X}_n(0)) \Bigg[\mathbb{1}_{d\mathbf{y}}(\mathbf{X}_n(0)) q_A^-(\mathbf{X}_n(0)) q_B^+(\mathbf{X}_n(\theta_n)) \frac{\mathbf{Y}(\mathbf{X}_n(\theta_n)) - \mathbf{Y}(\mathbf{X}_n(0))}{\theta_n} \tag{2.174}$$

$$+ \mathbb{1}_{d\mathbf{y}}(\mathbf{X}_n(\Delta t)) q_A^-(\mathbf{X}_n(\widetilde{\theta}_n)) q_B^+(\mathbf{X}_n(\Delta t)) \frac{\mathbf{Y}(\mathbf{X}_n(\Delta t)) - \mathbf{Y}(\mathbf{X}_n(\widetilde{\theta}_n))}{\Delta t - \widetilde{\theta}_n} \Bigg] \tag{2.175}$$

where θ_n and $\widetilde{\theta}_n$ are the "first-entry times" to $D = (A \cup B)^{\mathrm{c}}$ in the nth trajectory with time running forward and backward, respectively. To visualize $\mathbf{J}_{AA}$, $\mathbf{J}_{BA}$, and $\mathbf{J}_{BB}$, we swap symbols accordingly on the committor subscripts. For the steady-state current $\mathbf{J}$, we replace all committors with 1.

3 BACKGROUND: SUDDEN STRATOSPHERIC WARMING

I have selected sudden stratospheric warming (SSW) as the subject of TPT analysis in this thesis. Most TPT applications thus far have been confined to molecular dynamics applications, e.g., to determine reaction rates and pathways of complex conformational transitions [Thiede et al., 2019, Strahan et al., 2021]. TPT has also inspired many innovations in applied mathematics, with idealized potential landscapes as testbeds [e.g., Metzner et al., 2006, Banisch and Vanden-Eijnden, 2016, Khoo et al., 2018]. But only recently has TPT started to proliferate in climate science. The handful of climate applications so far include atmospheric blocking [Tantet et al., 2015], ocean eddy organization [Miron et al., 2021], and the El Niño Southern Oscillation [Lucente et al., 2019, 2021a]. However, we lack a general recipe for extracting meaningful physical insight with TPT in climate phenomena. One thing is clear: TPT should be applied across the *hierarchy* of models, widely recognized as fundamental to the practice of climate science [Held, 2005]. Low-order models aim to capture only the essential physics and discard the extraneous complex details, while fully coupled general circulation models (GCMs) aim to capture these details for high-fidelity forecasting. Each level strikes a different balance between interpretability and precision.

SSW is a quintessential example of why hierarchical modeling is so important. It was first detected in the 1950s as a mysterious temperature anomaly high aloft, and in following decades the community gradually discerned more of the massive scale and far-reaching consequences of SSW events. For a historical summary, see Baldwin et al. [2021]. A hierarchy of models developed, each advancing slightly different but related physical mechanisms [e.g., Matsuno, 1970, Holton and Mass, 1976, Scott and Polvani, 2006, Scott et al., 2008, Matthewman and Esler, 2011]. GCMs also produce plausible SSW events, thanks to the push to better resolve the stratosphere due to its surface impacts [Baldwin et al., 2021]. Yet despite the decades of physical insight and a mature model hierarchy, a *statistical* theory of SSW behavior is lacking. Rates, seasonality distributions, and dynamical characteristics common to SSW events depend sensitively on choices of definition and modeling choices. As an important driver of winter surface weather, SSW statistics are of great

practical as well as academic interest.

This situation is ripe for analysis with tools like TPT that characterize statistics and dynamics simultaneously. The following chapters ascend a hierarchy of SSW models, computing various quantities of interest for TPT at each level. We do not claim to "solve" the SSW problem, but only to introduce a tool that has potential for exploring the statistical consequences of various modeling choices. Before diving into these applications, this background chapter will present the essentials of SSW. Section 3.1 summarizes the observed characteristics and impacts of SSW, and section 3.2 describes the Holton-Mass model that supplies the part of the model hierarchy covered in chapters 4-6.

3.1 SSW observed characteristics

The polar winter stratosphere typically supports a strong, cyclonic polar vortex over the north pole, maintained by the thermal wind relation and meridional temperature gradient. A sudden stratospheric warming (SSW) event is a large excursion from this normal state, which can take many different forms. In split-type SSWs, the vortex splits completely in two. In displacement-type SSWs the vortex displaces far away from the pole. These can be considered wavenumber-2 and wavenumber-1 disturbances, respectively [Butler et al., 2015]. The subsidence and adiabatic warming associated with vortex breakdown can cause lower-stratospheric temperatures to rise by more than 40 K over several days [Baldwin et al., 2021]. With the reversal of stratospheric winds, upward-propagating planetary waves break at lower and lower levels, exerting a "downward influence" on tropospheric circulation and inducing equatorward shifts of the midlatitude jet and associated storm tracks [Baldwin and Dunkerton, 2001, Thompson et al., 2002, Baldwin et al., 2003, Hitchcock and Simpson, 2014, Kidston et al., 2015].

Some geographical regions near the storm track are then subject to anomalous weather regimes, such as extreme cold spells [Kolstad et al., 2010, Kretschmer et al., 2018a]. King et al. [2019] documents the impact of an SSW on extreme winter weather over the British Isles, the so-called "Beast

from the East" in February 2018. SSWs are a demonstrated source of predictability for surface weather on the subseasonal-to-seasonal (S2S) timescale, a frontier of weather forecasting with many implications for helping humanity deal with meteorological extremes [Sigmond et al., 2013, Scaife et al., 2016, White et al., 2017, Vitart and Robertson, 2018, Butler et al., 2019, Bloomfield et al., 2021, Scaife et al., 2022]. Abrupt cold spells severely stress infrastructures, economies and human lives, and every bit of extra prediction lead time is helpful for adaptation. Unfortunately, numerical weather prediction struggles to forecast SSW at any lead time longer than about two weeks [Tripathi et al., 2016]. For these reasons, there is keen interest in improving (i) the prediction of SSW itself beyond the horizon of $\sim$10 days that marks the current state-of-the-art [Tripathi et al., 2016, Domeisen et al., 2020], and (ii) understanding of the long-term frequency, seasonal distribution, and other climatological statistics of SSW.

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

3.2 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 [1970], proposed upward-propagating planetary waves as the major source of disturbance to the vortex. While Matsuno [1970] used impulsive forcing from the troposphere as the source of planetary waves, Holton and Mass [1976] suggested that even stationary tropospheric forcing could lead to an oscillatory response, suggesting that the stratosphere can self-sustain its own oscillations. While the Holton-Mass model is meant to represent internal stratospheric dynamics, Sjoberg and Birner [2014] point out that the stationary boundary condition does not lead to stationary wave activity flux, meaning that even the Holton-Mass model involves some dynamic interaction between the troposphere and stratosphere. Isolating internal from external dynamics is a subtle modeling question, but in chapters 4-6 we adhere to the original Holton-Mass framework for simplicity.

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, inducing a poleward flow of heat and stalling the vortex by depositing a burst of negative momentum. The vortex is destroyed and begins anew the rebuilding process.

Holton and Mass [1976] found that different values of the orographic "height" parameter h led to two distinct equilibrium regimes: a strong vortex with zonal wind close to the radiative equilibrium profile, and a weak vortex with a possibly oscillatory wind profile. A more detailed bifurcation analysis by Yoden [1987a] found that these two regimes coexist for a certain range of h. 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 Holton-Mass model is a wave-mean flow (or eddy-mean flow) interaction model on a β-plane channel between 60°N and 90°N, with zero-flux boundary conditions in y (position in the North-South or meridional direction), periodic boundary conditions in x (distance in the East-West or zonal direction), and z boundary conditions to be specified later. The model is derived from two basic ingredients. First, we have a set of prediction equations for the zonal mean flow $\overline{u}$:

$$\frac{\partial \overline{u}}{\partial t} - f_0 \overline{v} = 0 \tag{3.1}$$

$$f_0 \overline{u} = -\frac{\partial \overline{\Phi}}{\partial y} \tag{3.2}$$

$$\frac{\partial}{\partial t}\left(\frac{\partial \overline{\Phi}}{\partial z}\right) + N^2 \overline{w} = -\alpha\left(\frac{\partial \overline{\Phi}}{\partial z} - \frac{R\overline{T}^*}{H}\right) - \frac{\partial}{\partial y}(\overline{v'\Phi'_z}) \tag{3.3}$$

$$\frac{\partial \overline{v}}{\partial y} + \frac{1}{\rho_s}\frac{\partial}{\partial z}(\rho_s \overline{w}) = 0 \tag{3.4}$$

These equations represent (3.1) zonal and (3.2) meridional momentum balance, (3.3) conservation of energy, and (3.4) conservation of mass. Overbars and primes represent zonal averages and perturbations. Φ is the geopotential height; $H = 7$ km is a characteristic atmospheric scale height; $z = -H\ln(\frac{p}{p_0})$ is a vertical log-pressure coordinate; $\rho_s = \rho_0 e^{-z/H}$ is a standard density profile; $\overline{T}^* = \overline{T}^*(y,z)$ is the radiative equilibrium temperature field; $N^2 = 4 \times 10^{-4}\,\mathrm{s}^{-2}$ is a constant stratification (Brunt-Väisälä frequency); and

$$\alpha(z) = \left[1.5 + \tanh\left(\frac{z - 35\,\mathrm{km}}{7\ \mathrm{km}}\right)\right] \times 10^{-6}\mathrm{s}^{-1} \tag{3.5}$$

is the altitude-dependent cooling coefficient. z here represents log-pressure pseudo-height *above the surface*; in following papers z is referenced from the tropopause, and the 35 becomes 25 in the equation for $\alpha(z)$. The four equations (3.1)-(3.4) can be combined by eliminating $\overline{v}$, $\overline{\Phi}$, and $\overline{w}$ in

favor of $\overline{u}$:

$$\frac{\partial}{\partial t}\left[\frac{f_0^2}{N^2}\frac{1}{\rho_s}\frac{\partial}{\partial z}\left(\rho_s\frac{\partial\overline{u}}{\partial z}\right)+\frac{\partial^2\overline{u}}{\partial y^2}\right]=-\frac{f_0^2}{N^2}\frac{1}{\rho_s}\frac{\partial}{\partial z}\left[\alpha\rho_s\left(\frac{\partial\overline{u}}{\partial z}-\frac{\partial u_R}{\partial z}\right)\right]+\frac{\partial^2}{\partial y^2}\left[\frac{f_0^2}{N^2}\frac{1}{\rho_s}\frac{\partial}{\partial z}\left(\rho_s\overline{v'\frac{\partial\psi'}{\partial z}}\right)\right] \tag{3.6}$$

The left-hand side is the tendency of the (negative) meridional gradient of the background quasi-geostrophic potential vorticity (QGPV), $\partial_y\overline{q}$. There are two competing forces on the right-hand side. The first term is Newtonian cooling, which relaxes the zonal wind profile $\overline{u}$ towards a "radiative zonal wind profile" $u_R(z)$, defined to be in thermal wind balance with the radiative equilibrium temperature field: $f_0\partial_z u_R=-(R/H)\partial_y\overline{T}^*$. The second term is an eddy flux of potential vorticity, which is a destabilizing force. The parameter h will be specified as a lower boundary condition for ψ' below, and here it enters through the third term.

Second, we have the linearized QGPV equation, for the perturbation QGPV q' associated with the perturbation streamfunction ψ':

$$\left(\frac{\partial}{\partial t}+\overline{u}\frac{\partial}{\partial x}\right)q'+\beta_e\frac{\partial\psi'}{\partial x}+\frac{f_0^2}{\rho_s}\frac{\partial}{\partial z}\left(\frac{\alpha\rho_s}{N^2}\frac{\partial\psi'}{\partial z}\right)=0 \tag{3.7}$$

$$\text{where}\quad q':=\nabla^2\psi'+\frac{1}{\rho_s}\frac{\partial}{\partial z}\left(\frac{f_0^2}{N^2}\rho_s\frac{\partial\psi'}{\partial z}\right) \tag{3.8}$$

$$=\text{Zonal-perturbation QGPV field} \tag{3.9}$$

$$\text{and}\quad \beta_e=\beta-\frac{\partial^2\overline{u}}{\partial y^2}-\frac{1}{\rho_s}\frac{\partial}{\partial z}\left(\rho_s\frac{f_0^2}{N^2}\frac{\partial\overline{u}}{\partial z}\right) \tag{3.10}$$

$$=\text{Background meridional QGPV gradient} \tag{3.11}$$

At this point, we have two coupled PDEs—for u and ψ'—in three spatial dimensions and one temporal dimension. To simplify further and focus on the specific dynamics of the "polar night" (winter over the north pole), Holton and Mass [1976] projected these two fields onto a single zonal wavenumber $k=2/(a\cos 60^\circ)$ and a single meridional wavenumber $\ell=3/a$, where a is the Earth's radius. This notation is consistent with Holton and Mass [1976] and Christiansen [2000], and we

refer the reader to these earlier papers for complete description of the equations and parameters. The resulting ansatz is

$$\begin{aligned} \overline{u}(y,z,t) &= U(z,t)\sin(\ell y) \\ \psi'(x,y,z,t) &= \mathrm{Re}\{\Psi(z,t)e^{ikx}\}e^{z/2H}\sin(\ell y) \end{aligned} \tag{3.12}$$

which is fully determined by the reduced state space $U(z,t)$, and $\Psi(z,t)$, the latter being complex-valued. Inserting this into the linearized QGPV equations yields the coupled PDE system in one spatial dimension (z) as well as time:

$$\begin{aligned} &\left[-\left(\mathscr{G}^2(k^2+\ell^2)+\frac{1}{4}\right)+\frac{\partial^2}{\partial z^2}\right]\frac{\partial\Psi}{\partial t} \\ &= \left[\left(\frac{\alpha}{4}-\frac{\alpha_z}{2}-i\mathscr{G}^2k\beta\right)-\alpha_z\frac{\partial}{\partial z}-\alpha\frac{\partial^2}{\partial z^2}\right]\Psi \\ &\quad+\left\{ik\varepsilon\left[\left(k^2\mathscr{G}^2+\frac{1}{4}\right)-\frac{\partial}{\partial z}+\frac{\partial^2}{\partial z^2}\right]U\right\}\Psi-ik\varepsilon\frac{\partial^2\Psi}{\partial z^2}U \end{aligned} \tag{3.13}$$

for $\Psi(z,t)$, and

$$\begin{aligned} &\left(-\mathscr{G}^2\ell^2-\frac{\partial}{\partial z}+\frac{\partial^2}{\partial z^2}\right)\frac{\partial U}{\partial t} = \left[(\alpha_z-\alpha)U_z^R+\alpha U_{zz}^R\right] \\ &\quad-\left[(\alpha_z-\alpha)\frac{\partial}{\partial z}+\alpha\frac{\partial^2}{\partial z^2}\right]U+\frac{\varepsilon k\ell^2}{2}e^z\mathrm{Im}\left\{\Psi\frac{\partial^2\Psi^*}{\partial z^2}\right\} \end{aligned} \tag{3.14}$$

for $U(z,t)$. Here, $\varepsilon = 8/(3\pi)$ is a coefficient for projecting $\sin^2(\ell y)$ onto $\sin(\ell y)$. We have nondimensionalized the equations with the parameter $\mathscr{G}^2 = H^2N^2/(f_0^2L^2)$, where f_0 is the Coriolis parameter, and $L = 2.5\times10^5$ m is a horizontal length scale, selected in order to create a homogeneously shaped data set more suited to our analysis. Boundary conditions are prescribed at the bottom of the stratosphere, which in this model corresponds to $z = 0$ km, and the top of the strato-

sphere $z_{top} = 70$ km.

$$\Psi(0,t) = \frac{gh}{f_0}, \qquad \Psi(z_{top},t) = 0, \tag{3.15}$$
$$U(0,t) = U^R(0), \qquad \partial_z U(z_{top},t) = \partial_z U^R(z_{top}).$$

The vortex-stabilizing influence is represented by $\alpha(z)$, the altitude-dependent cooling coefficient, and the radiative wind profile $U^R(z) = U^R(0) + (\gamma/1000)z$ (with z in m), which relaxes the vortex toward radiative equilibrium. Here $\gamma = \mathscr{O}(1)$ is the vertical wind shear in m/s/km. The competing force of wave perturbation is encoded through the lower boundary condition $\Psi(z = 0,t) = gh/f_0$, which is a wavenumber-2 perturbation because k corresponds to wavenumber 2.

We now have specified the Holton-Mass model as a 2-dimensional PDE in z and t, with several free parameters. The following three chapters instantiate the model in two different ways for TPT analysis. In chapter 4, we follow Ruzmaikin et al. [2003] and discretize z to a single layer, resulting in a system of three ordinary differential equations (ODEs). We further impose stochastic noise following Birner and Williams [2008] to excite transitions between the two metastable states. In chapter 5-6, we discretize z to 27 layers, as in the original numerical studies of Holton and Mass [1976], which reduces the PDE to a system of 75 ODEs. Again, we find two metastable states, and impose spatiotemporal noise to excite transitions between them. Because TPT analysis requires solving PDEs over state space, the two levels of discretization require fundamentally different numerical and visualization techniques.

4 PATH PROPERTIES OF ATMOSPHERIC TRANSITIONS: ILLUSTRATION WITH A LOW-ORDER SUDDEN STRATOSPHERIC WARMING MODEL

As an initial demonstration of the utility of TPT, we analyze a stochastically forced Holton-Mass-type model with two stable states, corresponding to radiative equilibrium and a vacillating SSW-like regime. In this stochastic bistable setting, from certain probabilistic forecasts TPT facilitates estimation of dominant transition pathways and return times of transitions. These "dynamical statistics" are obtained by solving partial differential equations in the model's phase space. This chapter is adapted from the publication Finkel et al. [2020].

4.1 Introduction and background

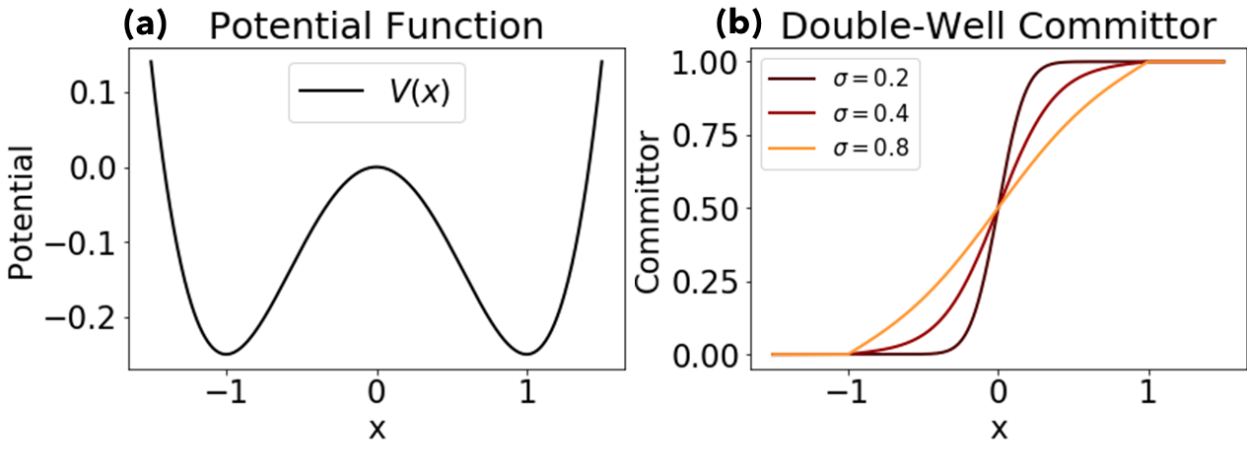
The general goal of this thesis is to develop a detailed understanding of transition events between two states, at least one of which is typically long-lived. Consider, for example, a particle with position $x(t)$ moving in the double-well potential energy landscape $V(x) = \frac{x^4}{4} - \frac{x^2}{2}$ (illustrated in Figure 4.1) and forced by stochastic white noise $\dot{W}$: $\dot{x} = -V'(x) + \sigma\dot{W}$. If the system starts in the left well, it will tend to remain there a while, but occasionally the stochastic forcing will push it over the barrier into the right well. The natural predictor for this event is the *committor*: the probability of reaching the right well before the left well.

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

$$\begin{cases} V'(x)q'(x) + \frac{\sigma^2}{2}q''(x) = 0 \quad x \in (-1,1) \\ q(-1) = 0, \;\; q(1) = 1 \end{cases} \implies q(x) = \frac{\int_{-1}^{x} \exp\left(\frac{2}{\sigma^2}V(x')\right) dx'}{\int_{-1}^{1} \exp\left(\frac{2}{\sigma^2}V(x')\right) dx'} \tag{4.1}$$

Note that the boundary conditions are implied by the probabilistic interpratation. The committor for this system is plotted in the right panel of Figure 4.1 for various noise levels. N.B., the potential

Figure 4.1: **The committor function for a double-well potential under the dynamics** $\dot{x} = -V'(x) + \sigma \dot{w}$**.** Panel (a) shows the potential function $V(x)$, and panel (b) shows the committor function. The committor has value zero on the left minimum, one on the right minimum, and one half at the top of the barrier. The stronger the stochastic forcing, the less the actual potential shape matters and the more gradual the committor's slope. For small noise, the dynamics become more deterministic and the committor approaches a step function, since $x(t)$ will directly approach whichever minimum is closer.



landscape picture is not fully general: many dynamical systems in > 1 dimension, including the Holton-Mass model, are not gradient systems, but this picture is the easiest way to understand the committor probability. The equations that determine $q(x)$ will be presented in section 2.

In the case of SSW, the long-lived states are the steady and disturbed circulation regimes of the stratospheric polar vortex. Recent work by Yasuda et al. [2017] has studied SSW in an equilibrium statistical mechanics framework, with these two stable states as saddle points of energy functionals. TPT takes a complementary non-equilibrium view, describing the long-time (steady-state) statistics of trajectories between the two states. For example, TPT introduces a probability density of reactive trajectories (or "reactive density") indicating the regions where trajectories tend to spend their time en route from A to B. The system is said to be *reactive* at a point in time if it has most recently visited A and will next visit B. The associated probability current of reactive trajectories (or "reactive current") indicates the preferred direction and speed of transition paths. These detailed descriptors of the mechanism underlying a rare event can be expressed in terms of probabilistic forecasts like the committor $q(x)$, the probability of entering state A before reaching state B from a given initial condition x (not in either A or B). The committor is the ideal probabilistic forecast in the usual variance-minimizing sense of conditional expectations [Durrett, 2013]. Any other predictor of a transition derived through experiments and observations, such as vortex preconditioning and forcing at different wavenumbers [Albers and Birner, 2014, Bancalá et al., 2012, Martius et al., 2009, Bao et al., 2017] necessarily corresponds to an approximation of the committor.

In many simplified climate models as well as in the double-well potential, stochastic forcing is needed to excite transitions between the metastable states. Stochastic forcing applies quite generally; while the climate system is deterministic in principle, nonlinear interactions between resolved and unresolved scales inevitably leads to resolution-dependent model errors that can be approximated as stochastic. Hasselmann [1976] originally formulated stochastic climate models to capture the influence of quickly evolving "weather" variables on the slowly evolving "climate" variables.

Stochastic parameterization remains an active area of research. For example, Franzke and Majda [2006] had success in capturing energy fluxes of a 3-layer quasigeostrophic model by projecting onto ten EOF modes and treating the remainder as stochastic forcing. Kitsios and Frederiksen [2019] addressed the challenge of designing consistent numerical schemes for subgrid-scale parametrization. Deep convection in the atmosphere and turbulence in the ocean boundary layer are two examples of multiscale processes that are especially challenging to resolve.

The aim of this chapter is to introduce the key quantities and relations of TPT in a conceptually simple climate model. TPT analysis on more complicated systems is a significant and worthwhile challenge, which is addressed in subsequent chapters. The chapter is organized as follows. Section 4.2 describes the dynamical model we use, building on work by Ruzmaikin et al. [2003] and Birner and Williams [2008]. Section 4.3 describes the mathematical framework of TPT, with detailed, but informal, derivations mainly put in the supplement. Section 4.4 explains the methodology, and section 4.5 presents the results particular to this model. Section 4.6 concludes the chapter.

4.2 Dynamical model

In the Holton-Mass model presented in chapter 3, a certain range of values for orographic forcing h allows the coexistence of two qualitatively different stable regimes: a steady eastward zonal flow close to radiative equilibrium, and a weaker zonal flow with quasi-periodic "vacillations" from eastward to westward, even under constant forcing. Each vacillation cycle consists of a sudden warming and cooling over the timescale of weeks. Although these individual cycles are interesting weather events unto themselves, in this chapter we think of the vacillations as occurring within a general *climate regime* that is conducive to sudden warming, as opposed to the steady flow state, which is not. Transitions between these two regimes, which we focus on here, are more accurately described as climatological shifts than weather events. The study by Ruzmaikin et al. [2003] varies h on an interannual timescale, with each single winter season occupying one of the two stable states and generating its daily weather accordingly. Hence, we will use the term "climate transitions."

The original Holton-Mass model discretizes the coupled PDE system (3.13), (3.14) with finite differences across 27 vertical levels, which is assumed to be close to a continuum limit. But to simplify even further, Ruzmaikin et al. [2003] did the most severe truncation possible, resolving only three vertical levels (including fixed boundaries) for easy analysis and exploration of parameter space. This reduces phase space to only three degrees of freedom: $U(t)$, which modulates $\overline{u}$ as a sine jet; $X(t) = \mathrm{Re}\{\Psi(t)\}$; and $Y(t) = \mathrm{Im}\{\Psi(t)\}$. (To avoid notational conflict, this chapter will use $\mathbf{Z} = (X, Y, U)$ instead of $\mathbf{X}$ to represent the state vector, while staying consistent with Ruzmaikin et al. [2003].) X and Y modulate the amplitude and phase of the perturbation streamfunction:

$$\psi'(x,y,t) = (X\cos kx - Y\sin kx)e^{z/2H}\sin \ell y \tag{4.2}$$

Carrying the Ansatz through the QG equations, Ruzmaikin et al. [2003] derived the following system:

$$\dot{X} = -\frac{1}{\tau_1}X - rY + sUY - \xi h + \delta_w \dot{h} \tag{4.3}$$

$$\dot{Y} = -\frac{1}{\tau_1}Y + rX - sUX + \zeta hU \tag{4.4}$$

$$\dot{U} = -\frac{1}{\tau_2}(U - U_R) - \eta hY - \delta_\Lambda \dot{\Lambda} \tag{4.5}$$

The primary control parameters are Λ (vertical wind shear) and h (topographic forcing and other sources of planetary waves, such as land-sea ice contrast). In Ruzmaikin et al. [2003], both are systematically varied between experiments and also subjected to seasonal and astronomical cycles. We simplify the parameter space by fixing Λ and only varying h between experiments, while also removing the seasonal and astronomical cycles Hence the the final terms in Eqs. (4.3) and (4.5) are zero. The full parameter list is specified in Table 4.1.

Remarkably, this hugely simplified model retains the qualitative structure of the Holton-Mass model as a bistable system for a certain range of h between the critical values $h_1 \approx 20\,m$ and

τ_1	122.6
τ_2	30.4
r	0.63
s	1.96
ξ	1.75
δ_w	70.84
ζ	240.54
U_R	0.47
η	9.13×10^4
δ_Λ	4.91×10^{-3}
$\dot{\Lambda}$	0

Table 4.1: **Numerical coefficients used in the reduced-order Ruzmaikin model**. The values are very similar to Ruzmaikin et al. [2003] and Birner and Williams [2008]. The relationship with physical parameters is described in the appendix of Ruzmaikin et al. [2003]. Note that our notation differs slightly: following Birner and Williams [2008], we write the topographic forcing in terms of h rather than $\Psi_0 = \frac{gh}{f_0}$, a difference that results in numerical factors of ~ 1000 depending on the convention used.

$h_2 \approx 160\,m$, as shown in the bifurcation diagram of Figure 4.2. Blue points represent the strong-vortex state, where zonal wind is in approximate thermal wind balance with the radiative equilibrium temperature field (henceforth called the "radiative solution"). Red points represent a disturbed vortex, with weaker zonal wind and vacillations. This climatological regime supports more SSW events, and is henceforth called the "vacillating solution." We use the same blue-red color scheme consistently here to represent these two states. Transitions between them happen on interannual time scales, affecting each year's likelihood of SSW events. The structure of transitions is illustrated in Figure 4.3: as h increases slowly past the bifurcation threshold h_2, the system enters a series of rapid, large-amplitude oscillations that spiral into the weaker-circulation state.

In Figures 4.2 and 4.3, transitions require crossing the bifurcation threshold h_2, where the radiative solution ceases to exist. Birner and Williams [2008] introduced additive white-noise forcing in the U variable to model unresolved gravity waves and found that these perturbations were sufficient to excite the system out of its normal state and into a vacillating regime. In Figure 4.4 we illustrate stochastic trajectories of the system for three different (fixed) values of h. (For numerical

Figure 4.2: **Fixed points of Equations (4.3)-(4.5) in the state space** (X,Y,U). Here X and Y represent the real and imaginary parts of the streamfunction and U the mean zonal wind amplitude. Fixed points vary as a function of the topographic forcing parameter, h. Panels (a), (b) and (c) show fixed points of X, Y and U respectively on the vertical axis, while h varies across the horizontal axis. Circles and crosses denote linearly stable and unstable fixed points, respectively. The range of h between $\sim 20\,m$ and $\sim 160\,m$ supports three fixed points, two stable and one unstable. In this range, the blue points correspond to the radiative solution, while the red points represent the vacillating regime. (In fact this is a stable fixed point with one real and two complex eigenvalues; vacillations refer to the oscillatory motion *near* the fixed point, which is excited by the stochastic noise pecified below.) This corresponds to a winter climatology that is conducive to sudden stratospheric warming events.

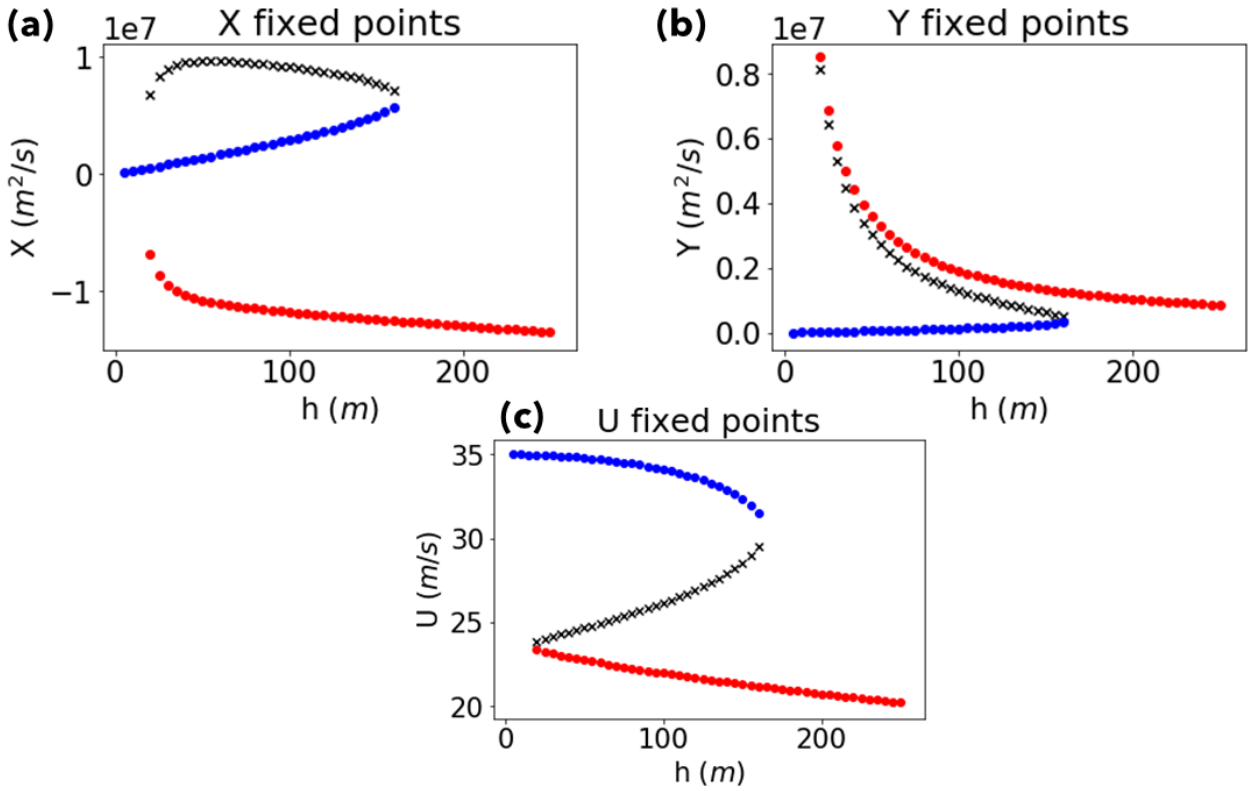
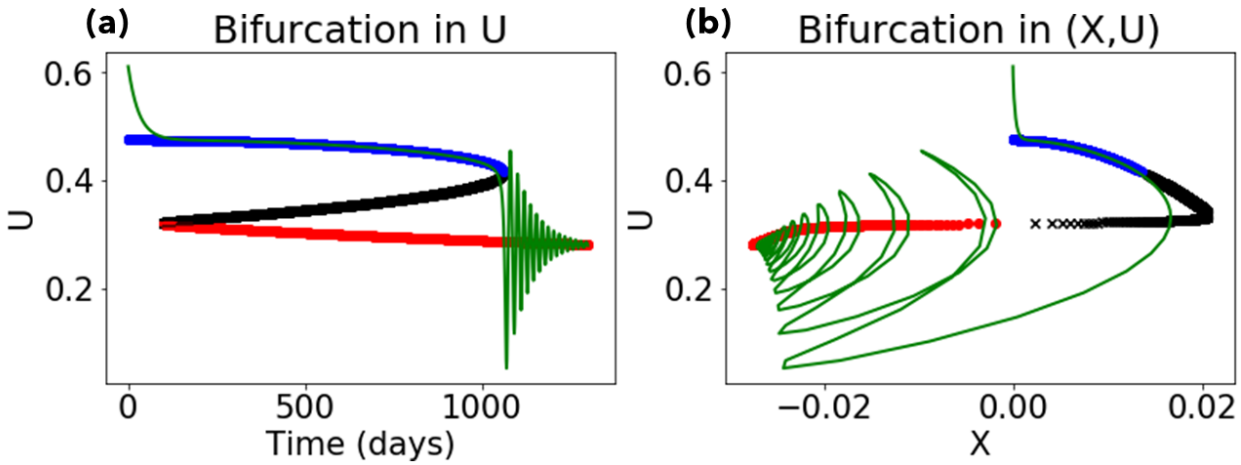


Figure 4.3: **Trajectories in (X,Y,U) space**. Here X and Y represent the real and imaginary parts of the streamfunction and U the mean zonal wind amplitude. In this simulation, the topographic forcing h increases linearly from $0\,m$ to $200\,m$ in 1300 days. (a) shows the fixed points, with colors blue, red and black for the radiative solution (A), the vacillating solution (B) and the unstable fixed point between them respectively. The trajectory of U over time is superimposed in gray. (b) plots this same curve parametrically, in XU space. Before the bifurcation, the trajectories follow the existing fixed point; after the bifurcation, they spiral into the new fixed point through a series of "vacillations."



reasons we also add a small amount of independent white noise to X and Y variables). Even when h is far below h_2, transitions still occur, and in fact the preference for the vacillating solution branch increases quickly with h.

Birner and Williams [2008] used direct numerical simulation and the Fokker-Planck equation to calculate long-term occupation statistics, i.e., how much time on average was spent in each regime and the mean first passage time before a transition to the vacillating regime, all for a range of forcing and noise levels. Our approach differs in both target and methodology. With direct access to the system's infinitesimal generator, as defined in chapter 2, we solve Feynman-Kac PDEs numerically to compute several relevant quantities of interest drawn from TPT: the committor, reactive densities, and reactive current. We then go on to validate these quantities using direct simulation.

4.3 Path properties

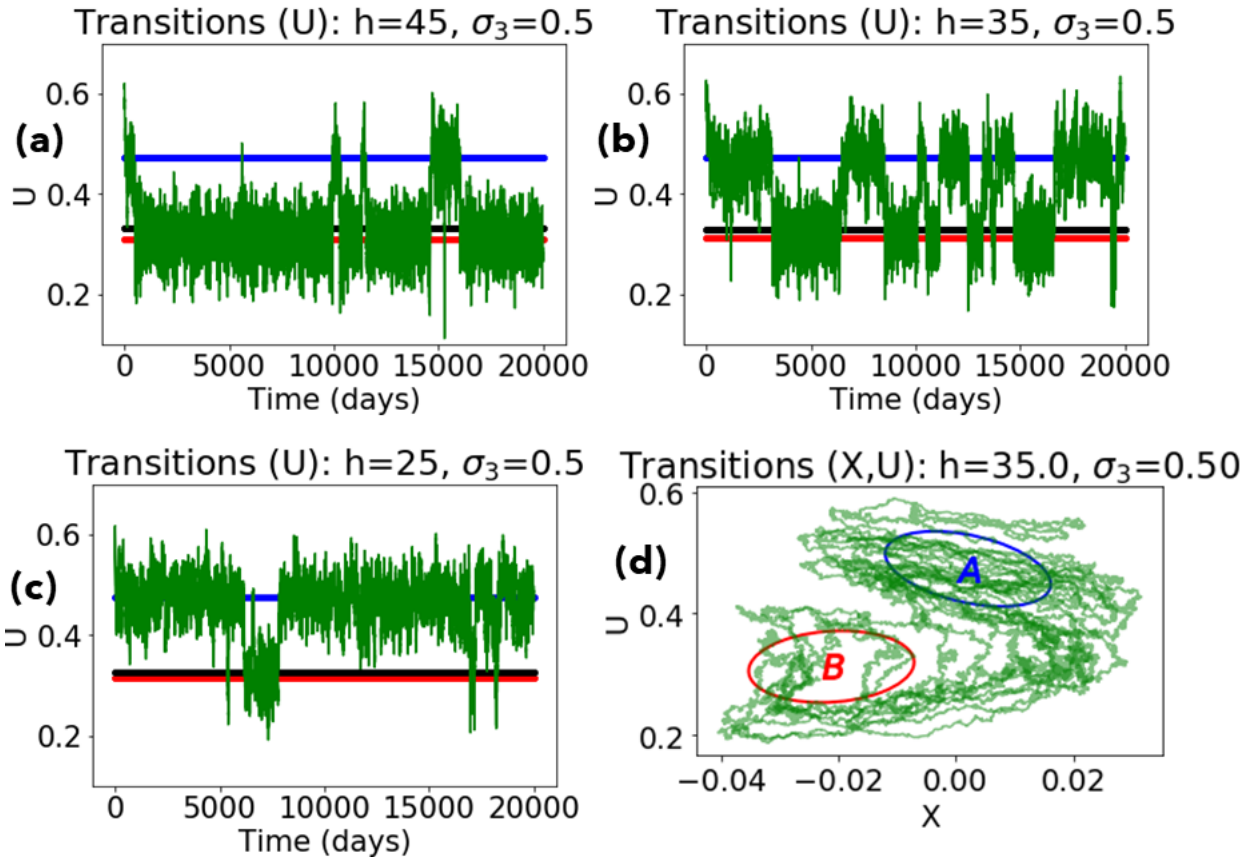
TPT characterizes the steady state statistics of transitions between states. In this section, we introduce key quantities needed to introduce TPT as applied to the Ruzmaikin model to obtain a more complete picture than we get from individual sample paths.

The noisy Ruzmaikin model can be expressed compactly as a stochastic differential equation (SDE)—specifically an Itô diffusion process of the type in Eq. (2.9)—in the variable $\mathbf{Z} = (X,Y,U) \in \mathbb{R}^3$ with a deterministic drift vector $\nu(z) = \big(\nu_1(z), \nu_2(z), \nu_3(z)\big)$ and a 3×3 diffusion matrix $\sigma(z)$.

$$d\mathbf{Z}(t) = \nu(\mathbf{Z}(t))\,dt + \sigma(\mathbf{Z}(t))\,d\mathbf{W}(t) \tag{4.6}$$

Here, $\mathbf{W}(t)$ is a 3-dimensional vector of independent Brownian motions. While σ can in principle be any state space-dependent matrix, we make σ diagonal and constant: $\sigma(\mathbf{z}) = \mathrm{diag}(\sigma_1, \sigma_2, \sigma_3)$, creating independent additive noise in the X, Y and U variables. σ_1 and σ_2 have units of $m^2/s/day^{1/2}$,

Figure 4.4: **Stochastic trajectories of the system**. We show trajectories with various fixed values of the parameters h (topographic forcing) and σ_3 (amplitude of stochastic forcing). Panels (a), (b) and (c) show $U(t)$ for three different forcing levels: $h = 25, 35, 45\,m$ with $\sigma_3 = 0.5$ m/s/day$^{1/2}$ (see text for specification of σ_1 and σ_2). All three h levels are within the zone of bistability in the bifurcation diagram in Fig. 4.2. In keeping with the bifurcation diagrams, the blue, black and red lines mark the radiative, unstable and vacillating solutions respectively. Note that their relative positions vary slightly with h, as fixed points depend on parameters. As h increases from left to right, the systems spends increasingly more of its time in the vacillating state. Panel (d) shows a parametric plot of the transitions through (X,U) space, for $h = 35\,m$ (another view of panel (b)). The $A \to B$ transition happens seven times, and hence panel (d) shows seven different transition paths superimposed on each other. Most of the transitions follow a similar characteristic path through XU space, with a rapid decrease in U followed by a decrease in X.



while σ_3 has units of $m/s/day^{1/2}$. Associated with this equation is the infinitesimal generator $\mathscr{L}$, an operator describing the evolution of observable functions forward in time following a trajectory (see chapter 2). Explicitly, if $f(\cdot)$ is a smooth function of phase space variables, then

$$\mathscr{L}f(\mathbf{z}) := \frac{d}{dt}\mathbb{E}[f(\mathbf{Z}(t))|\mathbf{Z}(0)=\mathbf{z}]\Big|_{t=0} \tag{4.7}$$

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

Itô's lemma (the chain rule for diffusion SDEs) gives the Kolmogorov backward equation, which represents $\mathscr{L}$ as a partial differential operator [e.g., Pavliotis, 2014]:

$$\mathscr{L}f(z) = \sum_i v_i(z)\frac{\partial f(z)}{\partial z_i} + \frac{1}{2}\sum_{i,j}(\sigma\sigma^\top)_{ij}\frac{\partial^2 f(z)}{\partial z_i \partial z_j} \tag{4.8}$$

$$= v(z)\cdot\nabla f(z) + \mathrm{Tr}\Big(\frac{1}{2}\sigma\sigma^\top Hf(z)\Big) \tag{4.9}$$

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

b. Equilibrium probability density

This stochastic process admits a time-dependent probability density, $\rho(\mathbf{z},t)$. If the system starts in a known position $\mathbf{Z}(0)=\mathbf{z}$, then $\rho(\mathbf{z}',0)=\delta(\mathbf{z}-\mathbf{z}')$. The density spreads out from this initial point over time according to the Fokker-Planck equation, which can be written in terms of the adjoint of

the generator:

$$\frac{\partial \rho(\mathbf{z},t)}{\partial t} = \mathscr{L}^* \rho(\mathbf{z},t) \tag{4.10}$$

$$= \sum_i \frac{\partial}{\partial z_i} \Big[-\nu_i(\mathbf{z})\rho(\mathbf{z},t) + \sum_j \frac{\partial}{\partial z_j} \big(\rho(\mathbf{z},t) D_{ij}(\mathbf{z})\big) \Big] \tag{4.11}$$

$$= \nabla \cdot \big[-\nu(\mathbf{z})\rho(\mathbf{z},t) + \nabla \cdot \big(\rho(\mathbf{z},t) D(\mathbf{z})\big) \big] \tag{4.12}$$

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

c. Metastable sets

The stationary density is an equilibrium quantity characterizing the long-term occupation statistics. But it is insufficient to describe the events of interest to us, which are *transition paths*: trajectory segments beginning inside the radiative state and ending inside the vacillating state. Specifically, we define the sets A and B as ellipsoids around these two fixed points, respectively. Their size is determined by contours of a local approximation to the stationary density π, using the *linearized* dynamical system about these stable fixed points. Let $\mathbf{a}$ be the upper stable fixed point of the

dynamics $\dot{\mathbf{z}} = \nu(\mathbf{z})$, and linearize

$$d\mathbf{Z} = \nu(\mathbf{Z})\,dt + \sigma\,d\mathbf{W} \tag{4.13}$$

$$\approx \left[\nu(\mathbf{a}) + \sum_j \frac{\partial \nu_i}{\partial z_j}(\mathbf{a})(\mathbf{Z}-\mathbf{a})_j\right] dt + \sigma\,d\mathbf{W} \tag{4.14}$$

$$=: G(\mathbf{Z}-\mathbf{a})\,dt + \sigma\,d\mathbf{W} \tag{4.15}$$

where $\nu(\mathbf{a}) = 0$ since $\mathbf{a}$ is a fixed point, and $G_{ij} = \frac{\partial \nu_i}{\partial x_j}(\mathbf{a})$ is the Jacobian matrix of the drift. This linear system, being stable, has its own equilibrium density $\pi_{\mathbf{a}}$ with covariance $C = \mathbb{E}_{\nu}[(\mathbf{Z}-\mathbf{a})(\mathbf{Z}-\mathbf{a})^{\top}]$. In fact, it is a Gaussian process, with an invariant density determined completely by its mean ($\mathbf{a}$) and its covariance matrix C, to be determined. Letting $f_{ij}(\mathbf{z}) = (z_i - a_i)(z_j - a_j)$, we can solve for the covariance directly using the generator on the (i,j)th entry of C. (For an alternative derivation, see Zwanzig [2001], ch. 1.) We use the basic fact that in steady-state, C is unchanging in time.

$$C_{ij} = \mathbb{E}_{\mathbf{Z}(t)\sim\pi_{\mathbf{a}}}[(\mathbf{Z}(t)-\mathbf{a})_i(\mathbf{Z}(t)-\mathbf{a})_j] \tag{4.16}$$

$$= \mathbb{E}_{\mathbf{Z}(0)\sim\pi_{\mathbf{a}}}[(\mathbf{Z}(t)-\mathbf{a})_i(\mathbf{Z}(t)-\mathbf{a})_j] \text{ for any } t \geq 0 \tag{4.17}$$

$$\frac{dC_{ij}}{dt} = 0 = \int \pi_{\mathbf{a}}(\mathbf{z})\frac{d}{dt}\mathbb{E}[(\mathbf{Z}(t)-\mathbf{a})_i(\mathbf{Z}(t)-\mathbf{a})_j|\mathbf{Z}(0)=\mathbf{z}]\,d\mathbf{z} \tag{4.18}$$

$$= \mathbb{E}_{\mathbf{Z}\sim\pi_{\mathbf{a}}}\{\mathscr{L} f_{ij}(\mathbf{Z})\} \text{ by setting } t = 0 \tag{4.19}$$

The subscript means that the initial position, $\mathbf{Z}$, is drawn from the density $\pi_{\mathbf{a}}$. Calculating $\mathscr{L} f_{ij}$ in detail, with implicit summation over repeated indices,

$$f_{ij}(\mathbf{z}) = (z_i - a_i)(z_j - a_j) \tag{4.20}$$

$$(\nabla f_{ij})_k = (z_i - a_i)\delta_{kj} + (z_j - a_j)\delta_{ki} \tag{4.21}$$

$$(Hf_{ij})_{k\ell} = \delta_{i\ell}\delta_{kj} + \delta_{j\ell}\delta_{ki} \tag{4.22}$$

$$\mathscr{L} f_{ij}(\mathbf{z}) = \nu(\mathbf{z}) \cdot \nabla f_{ij}(\mathbf{z}) + \mathrm{Tr}\big(D^\top H f_{ij}(z)\big) \tag{4.23}$$

$$= [G(\mathbf{z}-\mathbf{a})]_k [\nabla f_{ij}(z)]_k + D_{k\ell}[Hf_{ij}(z)]_{k\ell} \tag{4.24}$$

$$= G_{k\ell}(\mathbf{z}-\mathbf{a})_\ell [(\mathbf{z}-\mathbf{a})_i \delta_{kj} + (\mathbf{z}-\mathbf{a})_j \delta_{ki}] + D_{k\ell}(\delta_{i\ell}\delta_{kj} + \delta_{j\ell}\delta_{ki}) \tag{4.25}$$

$$= G_{j\ell}(\mathbf{z}-\mathbf{a})_\ell (\mathbf{z}-\mathbf{a})_i + G_{i\ell}(\mathbf{z}-\mathbf{a})_\ell (\mathbf{z}-\mathbf{a})_j + D_{ji} + D_{ij} \tag{4.26}$$

$$= [G(\mathbf{z}-\mathbf{a})(\mathbf{z}-\mathbf{a})^\top]_{ji} + [G(\mathbf{z}-\mathbf{a})(\mathbf{z}-\mathbf{a})]_{ij} + D_{ji} + D_{ij} \tag{4.27}$$

$$= [G(\mathbf{z}-\mathbf{a})(\mathbf{z}-\mathbf{a})^\top + (\mathbf{z}-\mathbf{a})(\mathbf{z}-\mathbf{a})^\top A^\top + 2D]_{ij} \tag{4.28}$$

We finally set the expectation of this expression to zero, obtaining a matrix equation for C.

$$\mathbb{E}_{\mathbf{x}}\{\mathscr{L} f_{ij}(\mathbf{Z})\} = \big[G\mathbb{E}_{\pi_{\mathbf{a}}}\{(\mathbf{Z}-\mathbf{a})(\mathbf{Z}-\mathbf{a})^\top\} + \mathbb{E}_{\pi_{\mathbf{a}}}\{(\mathbf{Z}-\mathbf{a})(\mathbf{Z}-\mathbf{a})^\top\}G^\top + 2D\big]_{ij} \tag{4.29}$$

$$0 = CG^\top + GC + 2D \tag{4.30}$$

This is known as a Sylvester equation which can be solved for C. The steady-state density of the linear system is therefore $\pi_{\mathbf{a}}(\mathbf{z}) \propto \exp[-(\mathbf{z}-\mathbf{a})^\top C^{-1}(\mathbf{z}-\mathbf{a})]$. The set A can then be defined as the set of all points z such that $(\mathbf{z}-\mathbf{a})^\top C^{-1}(\mathbf{z}-\mathbf{a}) < r^2$ for some hand-picked radius r, and analogously for B. For our experiments, we solved for C using the diffusion matrix D associated with $\sigma_3 = 0.5$ m/s/day$^{1/2}$ and $r = 1.48$ (a non-dimensional ratio). This choice was empirical and based on the characteristic oscillations about fixed point A that accompany the typical transition. These oscillations constitute the basic state of the system in the A basin, and therefore are dynamically distinct from the rare transition paths that are our focus. We chose an r parameter large enough that

set A encloses most of this looping motion, but not so large that the sets A and B start to dominate the phase space. Fig. 4.8 shows that typical transition paths still do contain several wide loops, but the maximum-current path traced in black does not. This picture is robust to variations in r in the range 1-2. More sophisticated methods for erasing loops while maintaining small sets are described in Lu and Vanden-Eijnden [2014] and Banisch and Vanden-Eijnden [2016].

We say that a snapshot $\mathbf{Z}(t)$ of the system is undergoing a *transition* (or reaction) at time t if it is on the way from set A to set B. This involves information about both its future and its past, for which we introduce the forward and backward committor probabilities.

d. Committor probabilities

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

$$q^+(\mathbf{z}) = \mathbb{P}\{\mathbf{Z}(t) \text{ next hits } B \text{ before } A | \mathbf{Z}(0) = \mathbf{z}\} \tag{4.31}$$

$$q^+(\mathbf{z} \in A) = 0, \quad q^+(\mathbf{z} \in B) = 1 \tag{4.32}$$

The boundary conditions on A and B follow naturally from the probabilistic definition. If the system begins in set A, by path continuity it will certainly next find itself in A, with zero chance of hitting B first. Starting in set B the opposite is true. The committor therefore obeys the boundary value problem (see chapter 2 for a derivation)

$$\begin{cases} \mathscr{L} q^+(\mathbf{z}) = 0 & \mathbf{z} \in (A \cup B)^{\mathrm{c}} \\ q^+(\mathbf{z}) = 0 & \mathbf{z} \in A \\ q^+(\mathbf{z}) = 1 & \mathbf{z} \in B \end{cases} \tag{4.33}$$

The equivalence of conditional expectations with respect to a Markov process like the committor and solutions to PDE involving the generator of the process are generally referred to as Feynman-

Kac relations [e.g., Karatzas and Shreve, 1998] and are well studied. The PDE in (25) is most naturally posed on an infinite domain, but as a numerical approximation we solve it in a large rectangular domain and impose homogeneous Neumann conditions at the domain boundary. A limiting example is the noise-dominated case, where $\nu(z)$ is negligible and $D = I$. The Kolmogorov backward equation then becomes Laplace's equation:

$$\mathscr{L}q^+(\mathbf{z}) = \nabla^2 q^+(\mathbf{z}) = 0 \tag{4.34}$$

If posed on the interval $[0,1]$, with $A = \{0\}$ and $B = \{1\}$, the solution is $q^+(\mathbf{z}) = \mathbf{z}$. The linear increase from set A to set B reflects the greater likelihood of entering B when beginning closer to it. This limit is reflected in Figure 4.1, which shows the committor of the double-well potential approaching a straight line for large σ values.

Prediction is naturally much harder in high-dimensional systems such as stratospheric models. A number of physically interpretable fields, such as zonal wind and geopotential height anomalies, seem to have some predictive power for SSW, but prediction by any single such diagnostic is suboptimal. Insofar as they are successful, these variables approximate certain aspects of the committor. For example, the committor might increase monotonically with the quasi-biennial oscillation (QBO) index. Furthermore, statistical correlations potentially obscure the conditional relationships needed. For example, Martius et al. [2009] and Bao et al. [2017] examined tropospheric precursors to SSW events in reanalysis records, finding that blocking events preceded most major SSWs, potentially by enhancing upward-propagating planetary waves. (We use "precursor" only to mean an event that sometimes happens before SSW.) Blocking influences SSW through height perturbations at the tropopause, which would enter the Ruzmaikin model as low-frequency variations in lower boundary forcing h. Since we fix h to be constant, the blocking precursor is outside our scope here. However, farther down the dynamic chain are other measurable precursors such vertical wave activity flux and meridional heat flux, which are also found to have predictive power [Sjoberg and Birner, 2012]. However comprehensive the model, we would naturally expect the

true committor probability to exhibit similar patterns to canonical precursors of that model such as blocking (for a troposphere-coupled model) and heat flux (for a stratosphere-only model). Yet, there is an important difference: while a precursor P may appear with high probability *given* that a SSW is imminent, the committor specifies the probability of a SSW given an observed pattern. As acknowledged in Martius et al. [2009], many blocking events did not lead to SSW events, meaning that $\mathbb{P}\{\text{SSW}|\text{blocking}\} \neq \mathbb{P}\{\text{blocking}|\text{SSW}\}$. Such distinctions highlight the need for a precise mathematical formulation that provides and distinguishes both kinds of information.

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

$$q^-(z) := \mathbb{P}\{\mathbf{Z}(t) \text{ last visited } A \text{ rather than } B|\mathbf{Z}(0) = \mathbf{z}\} \tag{4.35}$$

$$q^-(\mathbf{z} \in A) = 1, \quad q^-(\mathbf{z} \in B) = 0 \tag{4.36}$$

q^- solves the time-reversed Kolmogorov backward equation $\widetilde{\mathscr{L}}q^- = 0$, where $\widetilde{\mathscr{L}}$ is the time-reversed generator, which evolves observables backward in time (see chapter 2 for a detailed description of $\widetilde{\mathscr{L}}$ and its relationship the forward generator $\mathscr{L}$).

e. Densities and currents

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

The direction and intensity of this traffic is specified by the *reactive current*. To develop this

concept, we start by introducing the probability current $\mathbf{J}$, a vector field that satisfies a continuity equation with the time-dependent density ρ:

$$\frac{\partial \rho(\mathbf{z},t)}{\partial t} = \mathscr{L}^* \rho(\mathbf{z},t) = -\nabla \cdot \mathbf{J}(\mathbf{z},t) \tag{4.37}$$

If ρ were the density and v the velocity field of a fluid, $\mathbf{J}$ would be ρv. One can think of $\mathbf{J}$ as an instantaneous (in time and position) average over all possible system trajectories, though a precise mathematical description requires some care. In equilibrium, when $\rho = \pi$ is no longer changing, $\nabla \cdot \mathbf{J} = 0$, or equivalently $\oint_C \mathbf{J} \cdot \mathbf{n}\, dS = 0$ where C is any closed surface with outward unit normal vector $\mathbf{n}$.

The reactive current $\mathbf{J}_{AB}$ is also an "average velocity", but restricted to reactive paths. Unlike $\mathbf{J}$, $\mathbf{J}_{AB}$ is not divergence-free, with a source in A and a sink in B (where transition paths start and end). $\mathbf{J}_{AB}$ is defined implicitly via surface integrals. If C is any surface enclosing set A but not set B, with outward normal $\mathbf{n}$, then the flux $\oint_C \mathbf{J}_{AB} \cdot \mathbf{n}\, dS$ is the number of forward transitions per unit time, also called the *transition rate* R_{AB}. The result is (see chapter 2 or Metzner et al. [2006] for a derivation)

$$\mathbf{J} = \pi v - \nabla \cdot (\pi D) \tag{4.38}$$

$$\mathbf{J}_{AB} = q^+ q^- \mathbf{J} + \pi D (q^- \nabla q^+ - q^+ \nabla q^-) \tag{4.39}$$

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

Just as $\mathbf{J}_{AB}(z)$ describes the average reactive velocity, a streamline $\mathbf{Z}(t)$ of $\mathbf{J}_{AB}(\mathbf{z})$ (solving $\dot{\mathbf{Z}}(t) = \mathbf{J}_{AB}(\mathbf{Z}(t))$, with $\mathbf{Z}(0) = \mathbf{a} \in A$ and $\mathbf{Z}(T) = \mathbf{b} \in B$ for some $T > 0$ depending on $\mathbf{Z}(0)$) is a kind of "average" transition path. Although the streamline will not be realized by any particular

transition path, it will have common geometric features in phase space with many actual path samples. At low noise the reactive trajectories will cluster in a thin corridor about the streamline. The streamline is a more dynamical description of precursors: whereas regions of high reactive density are commonly observed states along reactive trajectories, streamlines of reactive current are commonly observed *sequences* of states along reactive trajectories. The study by Limpasuvan et al. [2004], for example, described a sequence of events in a prototypical SSW life cycle based on reanalysis including vortex preconditioning, wave forcing, and anomalous heat fluxes at various levels in the troposphere and stratosphere. The sequence described there likely corresponds to a streamline of the reactive trajectory.

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

$$\mathbb{E}[q^+(\mathbf{Z}(t))] = \mathbb{E}_\pi[q^+(\mathbf{Z}(t))] = \int_\Omega q^+(\mathbf{z})\pi(\mathbf{z})\,d\mathbf{z} =: \langle q^+ \rangle_\pi \tag{4.40}$$

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

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

A, which is $\langle q^- \rangle_\pi$.

$$k_{AB} = \frac{R}{\langle q^- \rangle_\pi} = \frac{R}{\int q^-(z)\pi(z)\,dz} \tag{4.41}$$

This rate constant, defined in Bowman et al. [2013], parallels the chemistry definition. If X_A and X_B are two chemical species, with $[\cdot]$ denoting concentration, the forward and backward rate constants k_{AB} and k_{BA} are defined so that

$$R = [X_A]k_{AB} = [X_B]k_{BA} \tag{4.42}$$

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

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

4.4 Methodology

a. Spatial discretization

The quantities of interest described above (π, q^+, q^-, and $\mathbf{J}_{AB}$) emerge as solutions to PDEs involving the generator $\mathscr{L}$, which must be approximated by spatial discretization. We use a finite volume scheme to directly discretize the adjoint $\mathscr{L}^*$ as a matrix, which we name L^*, on a regular grid in $d = 3$ dimensions. The grid consists of boxes $\{R_k\}_{k=1}^N$ with edge lengths $h_1 \times \ldots \times h_d$, with $\mathbf{z}_k$ as the centers. Denoting the unit vectors as $\{\mathbf{e}_i\}_{i=1}^d$, the nearest neighbors of $\mathbf{z}_k$ are $\{\mathbf{z}_k \pm h_i\mathbf{e}_i\}_{i=1}^d$.

Now, writing the Fokker-Planck equation in divergence form and exploiting the divergence theorem, and assuming D is diagonal,

$$\mathscr{L}^*\rho = \frac{\partial \rho}{\partial t} = -\nabla \cdot \left[\rho v - D\nabla\rho\right] \tag{4.43}$$

$$\frac{d}{dt}\int_{R_k} \rho \, d\mathbf{z} = -\int_{\partial R_k} \left(\rho v - D\nabla\rho\right) \cdot \mathbf{n} \, dS \tag{4.44}$$

$$= \sum_{\ell \neq k} \int_{\partial R_k \cap \partial R_\ell} \left(D\nabla\rho - \rho v\right) \cdot \mathbf{n} \, dS \tag{4.45}$$

$$\approx \sum_{i=1}^{d} \left\{ \left[D_{ii} \frac{\rho(\mathbf{z}_k + h_i\mathbf{e}_i) - \rho(\mathbf{z}_k)}{h_i} - \frac{\rho(\mathbf{z}_k) + \rho(\mathbf{z}_k + h_i\mathbf{e}_i)}{2} v\left(\mathbf{z}_k + \frac{1}{2}h_i\mathbf{e}_i\right) \right] \right. \tag{4.46}$$

$$\left. - \left[D_{ii} \frac{\rho(\mathbf{z}_k) - \rho(\mathbf{z}_k - h_i\mathbf{e}_i)}{h_i} - \frac{\rho(\mathbf{z}_k) + \rho(\mathbf{z}_k - h_i\mathbf{e}_i)}{2} v\left(\mathbf{z}_k - \frac{1}{2}h_i\mathbf{e}_i\right) \right] \right\} \tag{4.47}$$

$$= \sum_{i=1}^{d} \left\{ - \left[\frac{2D_{ii}}{h_i} + \frac{v(\mathbf{z}_k + \frac{1}{2}h_i\mathbf{e}_i) - v(\mathbf{z}_k - \frac{1}{2}h_i\mathbf{e}_i)}{2} \right] \rho(\mathbf{z}_k) \right. \tag{4.48}$$

$$+ \left[\frac{D_{ii}}{h_i} - \frac{v(\mathbf{z}_k + \frac{1}{2}h_i\mathbf{e}_i)}{2} \right] \rho(\mathbf{z}_k + h_i\mathbf{e}_i) \tag{4.49}$$

$$\left. + \left[\frac{D_{ii}}{h_i} + \frac{v(\mathbf{z}_k - \frac{1}{2}h_i\mathbf{e}_i)}{2} \right] \rho(\mathbf{z}_k - h_i\mathbf{e}_i) \right\} \tag{4.50}$$

The coefficients of each density term gives the corresponding matrix entry in L^*. At the domain boundaries, we consider the flux through outward facing edges to be zero. For example, if $\mathbf{z}_k$ is the upper boundary of the domain in the i'th direction, the terms involving $\rho(\mathbf{z}_k + h_i\mathbf{e}_i) - \rho(\mathbf{z}_k)$ will vanish. In terms of the PDE, this imposes homogeneous Neumann boundary conditions. In terms of the stochastic process, these are reflecting boundary conditions. This is somewhat unnatural for the dynamical system we consider, but provided the domain is big enough relative to the noise, the system will visit the boundary sufficiently rarely for this to have negligible impact.

Here we use the same domain and noise levels as Birner and Williams [2008]: $-0.06 \leq X \leq 0.04$, $-0.05 \leq Y \leq 0.05$, $0 \leq U \leq 0.8$ in units non-dimensionalized in terms of the radius of Earth and the length of a day. We tile this with a grid of $40 \times 40 \times 80$ grid cells. We choose a noise constant σ_3 in the U variable in the range $0.4 - 1.5$. This is a similar range to observed

atmospheric gravity wave momentum forcing [Birner and Williams, 2008]. For numerical reasons, we also add small noise to the streamfunction variables X and Y, in proportion to the domain size. Specifically, as U spans a range of 0.8 and X,Y span a smaller range of 0.1, we choose σ_1 and σ_2 to be $\sigma_3 \times (0.1/0.8)$. This adjustment does change our results with respect to Birner and Williams [2008], causing more transitions in both directions at lower h than if only the U variable were perturbed. While gravity wave drag forces the zonal wind, eddy interactions and other sources of internal variability can perturb the streamfunction as well, and it is not uncommon to represent these effects stochastically [DelSole and Farrell, 1995]. There are surely more accurate representations of noise, but this important issue is not our focus. We retain these perturbations for numerical convenience, but stress that the general principles of the TPT framework are independent of any specific form of stochasticity. In the forthcoming experiments, we will refer only to σ_3 with the understanding that σ_1 and σ_2 are adjusted proportionally.

The discretization has strengths and limitations. Given the matrix L^* on this grid, the discretized generator L is just the transpose. To ensure certain properties of solutions, such as positivity of probabilities, L should ideally retain characteristics of the infinitesimal generator of a discrete-space, continuous-time Markov process: rows that sum to zero, and nonnegative off-diagonal entries. Such a discretization is called "realizable" [Bou-Rabee and Vanden-Eijnden, 2015]. One can check that our discretization always satisfies the former property, and so is realizable provided small enough grid spacing $(\delta X, \delta Y, \delta U)$. In our current example, the spacing is not nearly small enough to guarantee this (matrix entries were just as often negative as positive), but results are still accurate, as verified by stochastic simulations to be described in section 4.5. While we could have used one-sided finite differences to enforce positivity, this would have degraded the overall numerical accuracy of the solutions. We opted instead to zero out negatives, which were always negligible in magnitude.

The discretized Kolmogorov backward equation is $Lq^+ = 0$, augmented with appropriate boundary conditions. The definition of A and B is a design choice that should satisfy three conditions:

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

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

Specifically, we start with the continuous-space relationship (2.92) and convert it into a discretized form consistent with the numerical grid. Recall that $\phi_\delta \to \mathbb{1}_{S^c}$ as $\delta \to 0$.

$$R_{AB} = \lim_{\delta\to 0} \int \pi q^- \left\{ \mathscr{L}\left[\phi_\delta q^+\right] - \phi_\delta \mathscr{L} q^+ \right\} d\mathbf{z} \tag{4.51}$$

$$= \lim_{\delta\to 0} \langle q^-, \mathscr{L}\left[\phi_\delta q^+\right]\rangle_\pi - \langle q^- \phi_\delta, \mathscr{L} q^+\rangle_\pi \tag{4.52}$$

$$= \lim_{\delta\to 0} \left(\langle q^-, \mathscr{L}\left[\phi_\delta q^+\right]\rangle_\pi - \langle \widetilde{\mathscr{L}}[q^- \phi_\delta], q^+\rangle_\pi \right) \tag{4.53}$$

At this point, we transition from continuous space to discrete space, and label the average value of

q^+ over the i'th cell by q_i^+.

$$R_{AB} \approx \sum_{i\in S} \pi_i \left(q_i^- \sum_{j\in\Omega} L_{ij} q_j^+ \mathbb{1}_{S^c}(j) - q_i^+ \sum_{j\in\Omega} \widetilde{L}_{ij} q_j^- \mathbb{1}_{S^c}(j) \right) \tag{4.54}$$

$$= \sum_{i\in S} \pi_i \left(q_i^- \sum_{j\in S^c} L_{ij} q_j^+ - q_i^+ \sum_{j\in S^c} \frac{\pi_j}{\pi_i} L_{ji} q_j^- \right) \tag{4.55}$$

$$= \sum_{i\in S, j\in S^c} (\pi_i q_i^- L_{ij} q_j^+ - \pi_j q_j^- L_{ji} q_i^+) \tag{4.56}$$

$$=: \sum_{i\in S, j\in S^c} (f_{ij}^{AB} - f_{ji}^{AB}) \tag{4.57}$$

Intuitively, f_{ij}^{AB} describes the number of reactive trajectories per unit time making the transition from cell i to cell j. Both f_{ij}^{AB} and f_{ji}^{AB} are positive with small enough grid spacing, but our discretization can reliably compute only the difference between them. In practice, we take S as the set A and compute the rate as

$$R = \sum_{i\in A, j\in A^c} f_{ij}^{AB} \tag{4.58}$$

since for $i \in A$, $q_i^+ = 0$ and therefore $f_{ji}^{AB} = \pi_j q_j^- L_{ji} q_i^+ = 0$. As detailed in Metzner et al. [2009], we compute the dominant transition pathway $(i_1, \ldots, i_n)$ as the path whose bottleneck,

$$\min_k \max(0, f_{i_k,i_{k+1}}^{AB} - f_{i_{k+1},i_k}^{AB}), \tag{4.59}$$

is maximized.

4.5 Results

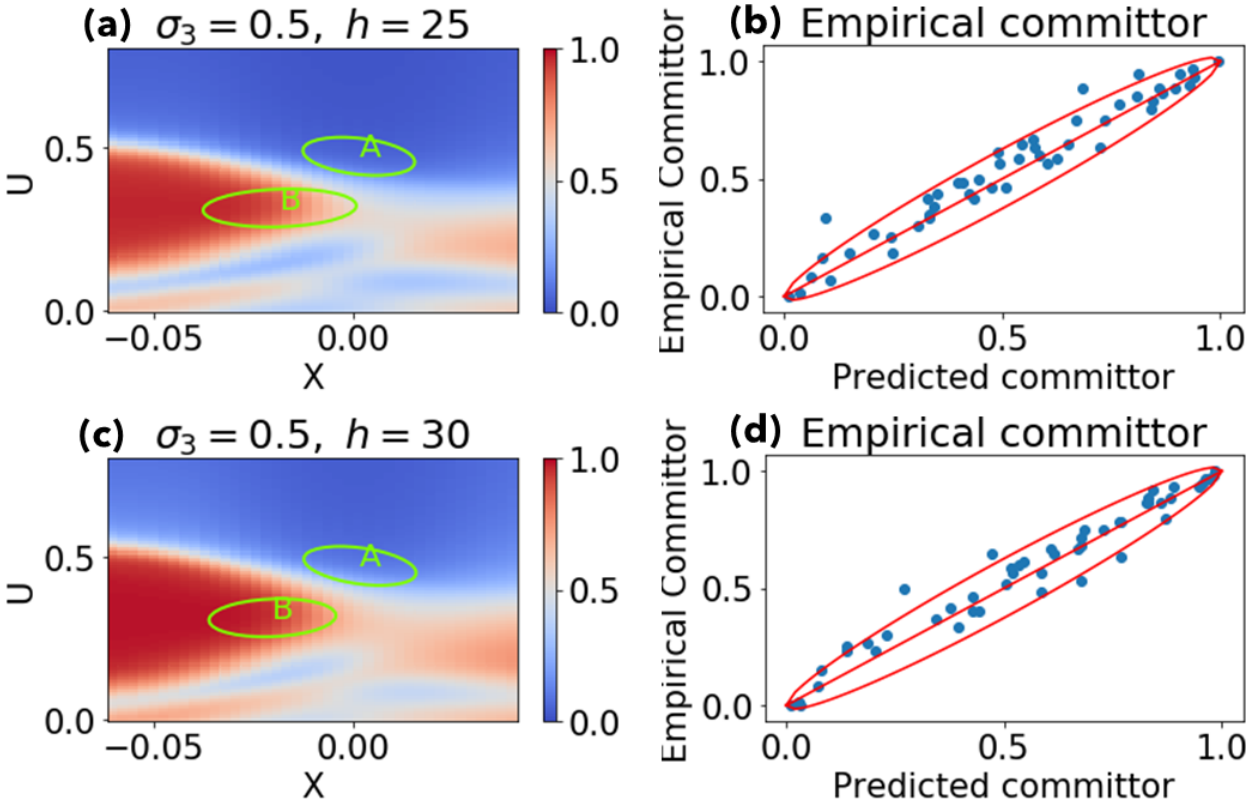
We begin this section by describing the kinematic path characteristics of the process in its three-dimensional phase space, according to the quantities described above. Following this purely geometrical description, we will suggest some dynamical interpretations and compare with previous

studies. Finally, we will map statistical features as functions of background parameters.

The Ruzmaikin model is attractive for demonstrating use of the tools introduced in Section 3 due to its low-dimensional state space, in which PDEs can be solved numerically using standard methods such as our finite volume scheme. We tested the committor's accuracy empirically by randomly selecting 50 cells in our grid (this is 0.04% of the grid) and evolving $n = 60$ stochastic trajectories forward in time from each, stopping when they reach either set A or set B. The fraction of trajectories starting from $\mathbf{z}$ that first reach B is taken as the empirical committor at point $\mathbf{z}$ and is denoted $\widetilde{q^+}$, which we compare with the predicted committor q^+ from the finite volume scheme. Figure 4.5 clearly demonstrates the usefulness of the committor for probabilistic forecasting. The left column displays the committor calculated from finite volumes, averaged in the Y direction, for two different forcing levels h. The right column shows a scatterplot of $\widetilde{q^+}$ vs. q^+ at the 50 randomly selected grid cells. We expect the points to fall along the line $\widetilde{q^+} = q^+$ with some spread proportional to the standard deviation of the binomial distribution, $\sqrt{q^+(1-q^+)/n}$. Approximating this as a Gaussian, we have plotted red curves enclosing the 95% confidence interval, which indeed contains approximately 95% of the data points. Statistical sampling error can explain the observed level of deviation, although discretization error (from solving the PDE and from time-stepping) may also contribute.

Figure 4.5 also shows how q^+ responds to increasing h, even far below the bifurcation threshold: the committor values throughout state space become rapidly skewed toward unity (meaning more red in the picture). This means that even slight perturbations can kick the system out of state A toward state B. Another indicator is the "isocommittor surface", the set of points z such that $q^+(z) = 1/2$; that is, the system has equal probability of next entering set A or B. In the left-hand column this is the set of gray points (averaging out the variable Y). For low forcing, this surface tightly encloses set B, meaning the system must wander very close before a transition is imminent. For high forcing values, the isocommittor hugs set A more closely, meaning that small perturbations from this normal state can easily push the system into dangerous territory. In Figure 4.6, the

Figure 4.5: **Empirical demonstration of the committor**. Here, noise is fixed at $\sigma_3 = 0.5\,m/s/day^{1/2}$, while topographic forcing $h = 25\,m$ in (a), (b) and $30\,m$ in (c), (d). The left column shows the forward committor q^+ solved by the finite volume method, averaged in the Y direction. The ellipses labeled A and B are projections of the actual sets onto the XU plane, where X and U are the real part of the streamfunction and the mean zonal wind amplitude. Committor values range from 0 (blue) to 1 (red), with the white contour showing the surface $q^+ = \frac{1}{2}$. The right column compares the PDE solution of the committor with a Monte Carlo solution from running many trajectories. For 50 randomly chosen grid points (sampled uniformly across the committor range $(0,1)$), we launched 60 independent stochastic trajectories and counted the fraction that reached set B first. We call this the empirical committor, $\widetilde{q^+}$. Plots (b) and (d) show $\widetilde{q^+}$ vs. q^+ for these 50 random grid cells. The middle red line is the curve $\widetilde{q^+} = q^+$, and the envelope around it is the 95% confidence interval for sampling errors, based on a Gaussian approximation to the binomial distribution.



isocommittor is shown as a set of gray points in a 3D plot viewed from various vantage points. In the low-U region, the isocommittor resembles a spiral staircase, reflecting the spiral-shaped stable manifold of the fixed point in set B. Different initial positions with the same streamfunction phase, differing only slightly in the U direction, can have drastically different final destinations. These spiral surfaces are responsible for the blue lobes in the lower part of Figure 4.5, but they disappear at higher noise.

Figures 4.7 and 4.8 display numerical solutions of the equilibrium density π and reactive density $\rho_R \propto \pi q^- q^+$ for two forcing levels. While π indicates where $\mathbf{Z}(t)$ tends to reside, ρ_R indicates where $\mathbf{Z}(t)$ resides *given* a transition from A to B is underway. As h increases, even far below the bifurcation threshold, π responds strongly, shifting weight toward state B. On the other hand, the reactive density displays similar characteristics for all h values. In the XU plane, the two lobes of high reactive density surrounding A indicate that zonal wind tends to remain strong for a while before dipping into the weaker regime. Viewing the same field in the XY plane (Figure 4.8) reveals a halo of intermediate density about set A. While many different motions would be consistent with this pattern, the coming figures verify that the early stages of transition have circular loops in the XY plane, meaning zonal movement of the streamfunction's peaks and troughs. The exact streamfunction phase corresponding to the (X,Y) position is calculated as follows. Recall the streamfunction is $\psi' = \mathrm{Re}\{\Psi(t)e^{ikx}\cos(\ell y)\}$, where $\Psi = X + iY$. In polar coordinates, $\Psi = \sqrt{X^2+Y^2}e^{i\phi}$, where $\phi = \tan^{-1}(Y/X)$. The full streamfunction is

$$\begin{aligned}\psi' = \mathrm{Re}\{\Psi e^{ikx}\}\cos(\ell y) &= \mathrm{Re}\{\sqrt{X^2+Y^2}e^{i(\phi+kx)}\}\cos(\ell y)\\ &= \sqrt{X^2+Y^2}\cos(\phi+2\lambda)\cos(\ell y)\end{aligned}$$

where λ is longitude.

The angle from the origin in the XY plane indicates the zonal streamfunction phase, and circular motion indicates zonal movement. (This looping motion is indeed shared by the transition path samples shown in Figures 4.9 and 4.10, to be described later.)

Figure 4.6: **The committor-1/2 surface**. This is the set of all points in state space where the $q^+ = \frac{1}{2}$, and sets A and B have equal probabilities of being visited next. Here the surface is rendered as a set of points and viewed from various vantage points in state space (the supplement shows a video with rotation). The topographic forcing is fixed to $h = 25\,m$ and the noise level to $\sigma_3 = 0.5\,m/s/day^{1/2}$. The blue and red clusters mark sets A and B respectively, centered around the two stable fixed points. The gray points show the location of the surface $q^+ = \frac{1}{2}$. The most striking feature is the "spiral staircase" structure in the low-U region of phase space. For any given streamfunction phase, the likelihood of heading toward state A or B depends sensitively on U, in an oscillatory manner. Even at very low values of U, there are narrow channels which are likely to lead back to set A rather than set B. This accounts for the blue regions in the lower part of Figure 4.5. These disappear, however, at higher noise, when set B overtakes the lower half of the picture.

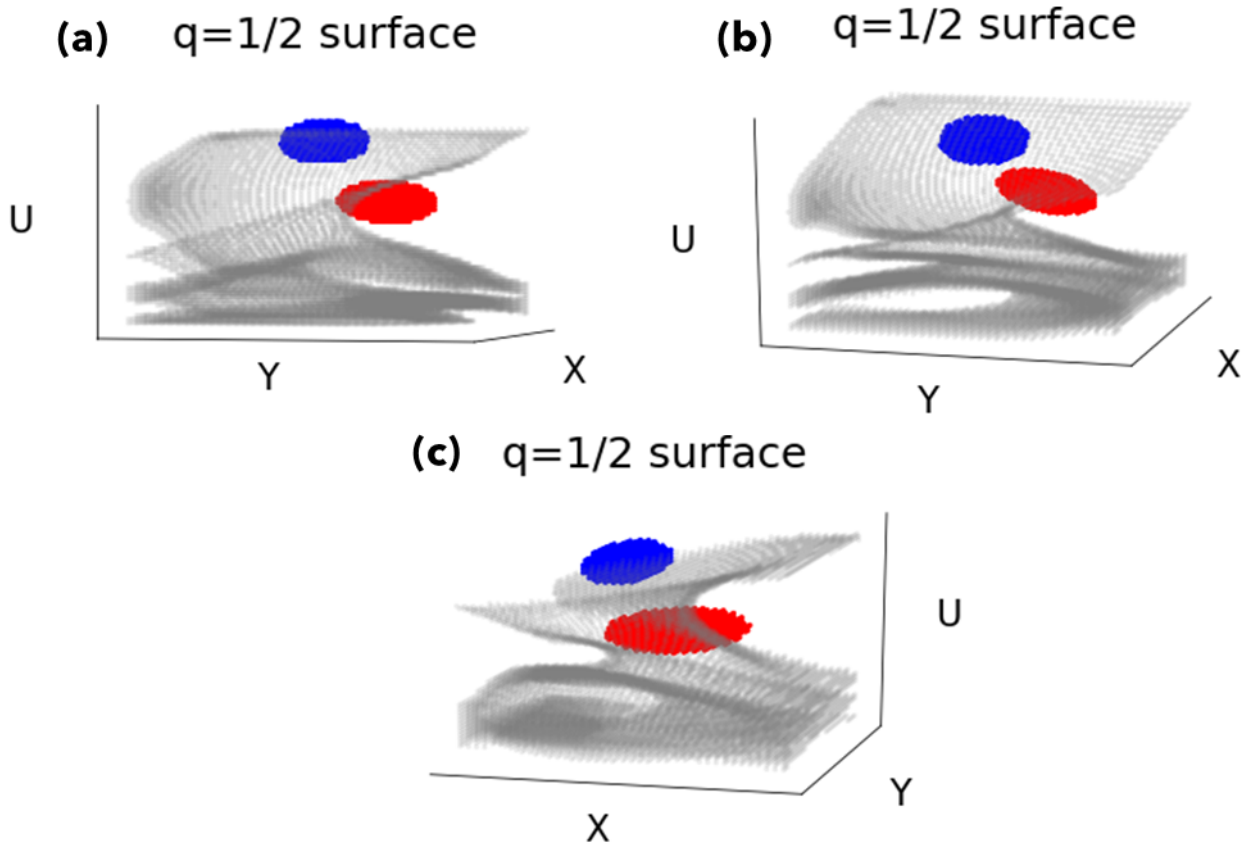


Figure 4.7: **Equilibrium and reactive densities in (X,U) space**. The equilibrium density $\pi(z)$ (left column) and reactive density $\rho_R(z) = \pi(z)q^+(z)q^-(z)$ (right column) are displayed for two different forcing levels, $h = 25\,m$ (top row) and $h = 35\,m$ (bottom row). $\pi(z)$ is the long-term probability density of finding the system at point z at any given time; $\rho_R(z)$ is the same probability, but *conditional* on also being reactive at that time, meaning having last visited A and next destined to visit B. Dark color indicates higher density. These densities are summed in the Y direction to give a marginal density as a function of X and U. The red and blue ellipses show the projected boundaries of sets A and B, respectively. These reactive densities capture the patterns of transition path samples shown in Figure 4.4, but through continuous fields instead. At low forcing levels, most of the equilibrium density is concentrated around set A, whereas higher forcing shifts some of the mass to set B. Meanwhile, the characteristic curved shape of transition paths in phase space is borne out by the reactive densities, with a sickle-shaped high-density region bridging the gap between set A and B.

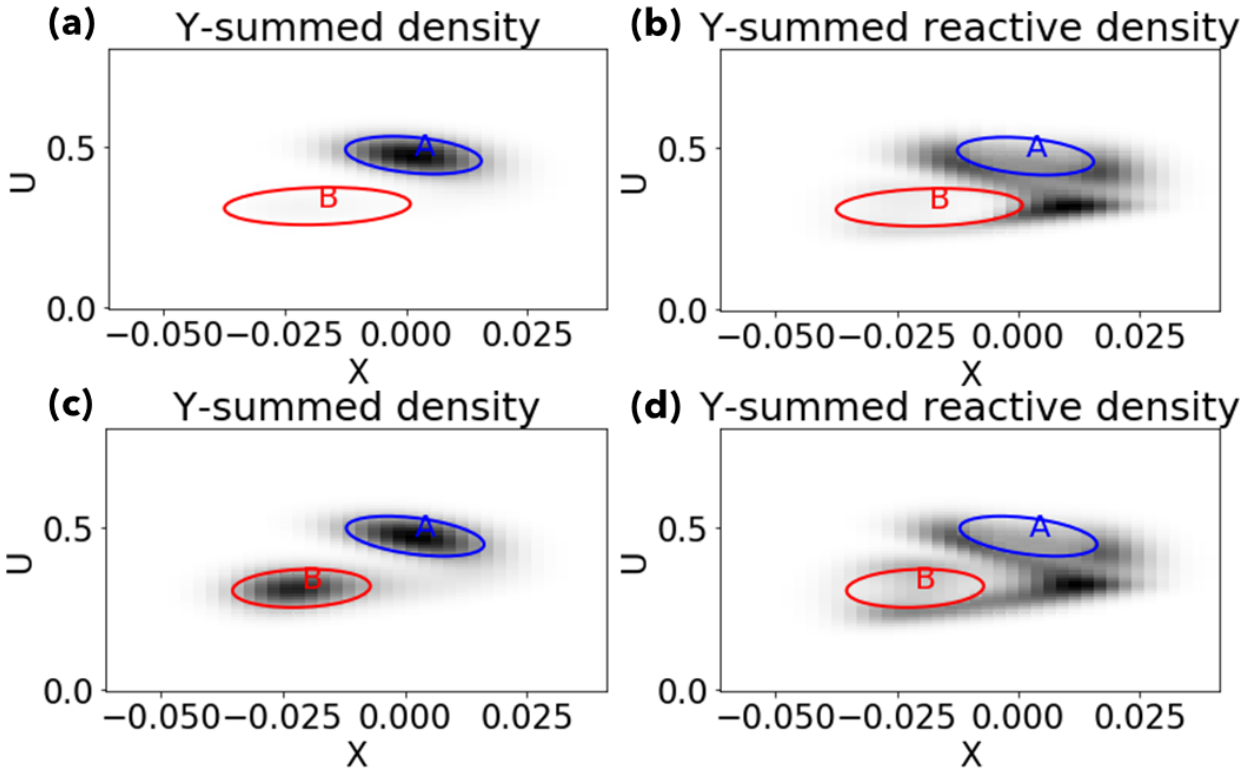
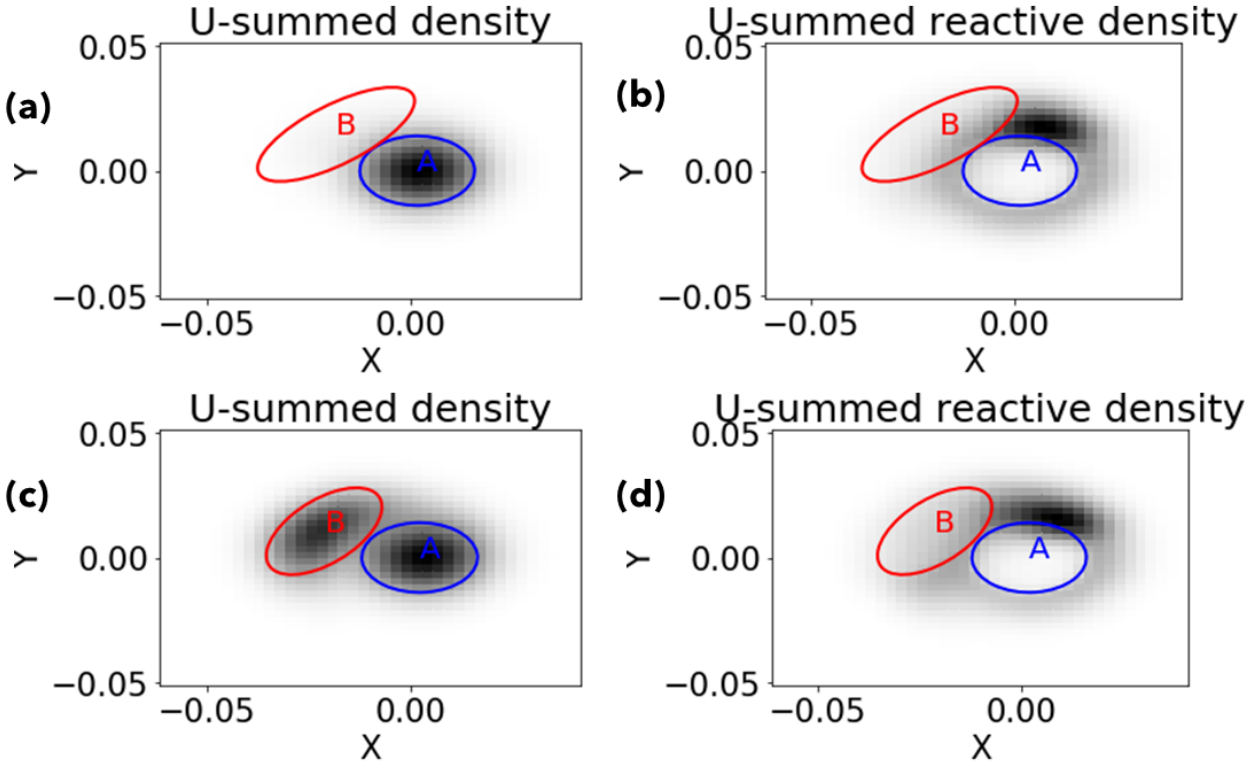


Figure 4.8: **Equilibrium and reactive densities in (X,Y) space.** The XY plane is the complex plane that characterizes the perturbation streamfunction. The center of set A, approximately at $X = Y = 0$, corresponds to a zonally symmetric streamfunction with no perturbation. Counterclockwise motion of trajectories around A represents an eastward phase velocity of the streamfunction, which is the dominant modality in the radiative regime. The region of high density above set A shows the phase in the streamfunction at which zonal wind is most likely to begin to weaken.



The darkest (most-trafficked) region of this loop is the sector $\frac{\pi}{4} \lesssim \phi \lesssim \frac{\pi}{2}$. The relationship between (X,Y) and λ indicates ψ' is maximized at longitudes $\lambda = \{-\frac{\phi}{2}, \pi - \frac{\phi}{2}\}$. As the maximum reactive density occurs around $\phi = \frac{3\pi}{8}$, the streamfunction peaks are at $\{-\frac{3\pi}{16}, \frac{13\pi}{16}\} \approx \{326^\circ, 146^\circ\}$. What is the significance of this phase relative to the lower boundary forcing? Recalling the forcing form $\mathrm{Re}\{\Psi(z_B,t)e^{ikx}\} = h\mathrm{Re}\{e^{i2\lambda}\} \propto \cos(2\lambda)$, the bottom peaks are located at $\lambda = \{0, \pi\}$. Hence, the bulk of the transition process happens when the perturbation streamfunction at the mid-stratosphere lags the lower boundary condition by $\frac{3}{16} \pm \frac{1}{16}$ of a wavelength. Meanwhile, the XU plane reveals what happens to the zonal wind speed during the SSW transition. The high-reactive density region discussed above coincides with the crescent-shaped bridge of high density between the sets in Figure 4.7. This suggests that in an SSW, the zonal wind weakens while the streamfunction stays in that particular phase window.

The pictures of reactive density suggest that reactive trajectories tend to loop around set A, physically meaning the streamfunction tends to travel in one direction before slowing down, but they technically convey no *directional* information to explicitly support this claim. For this, we turn to the reactive current. We computed the discrete-space effective current matrix f_{ij}^+, directly from the finite volume discretization and numerical solutions of π, q^+ and q^-. Physically, this matrix represents the flux of a vector field from grid cell i to cell j. From this we calculated the maximum-current paths as described in Metzner et al. [2009] and displayed the results in Figure 4.9 for a forcing level of $h = 30m$ (other levels are qualitatively similar). Both the XU and XY views are shown. Superimposed on these paths are seven actual reactive trajectories that occurred during a long stochastic simulation, to demonstrate features that are captured by the dominant pathways. The dominant path from A to B indeed contains a half-loop in the XY plane in the clockwise direction, which means an eastward phase velocity. With a smaller set A, this dominant path would contain more of these loops. However, during the next transition stage, the streamfunction slows to a halt at the phase angle $\phi = \frac{\pi}{2}$, doubles back and travels westward as zonal wind loses strength. The smear of high density in the neighborhood $\phi \sim \frac{3\pi}{16}$ therefore comprises not only a precipitous

drop in zonal wind (which happens at the edge of that region) but also a backtrack, this time with weaker background zonal wind. This behavior is borne out by the trajectory samples, which vacillate in the upper-middle section of the XY plot. These paths are displayed as space-time diagrams of the streamfunction in Figure 4.10. In panel (a), the dominant path's two loops correspond to two troughs moving east past a fixed longitude before the slowdown. The random streamfunction trajectories shown in (b)-(g) do not follow this representative history exactly, but they do combine elements of it: steady eastward wave propagation followed by slowdown and reversal. Each stage can have multiple false starts. Notably, the slowdown consistently happens at the same phase, with peaks at $\sim 120^\circ E$ and $300^\circ E$, at roughly the same phase as found from the density plots. In fact, the figures show a brief slowdown every time the streamfunction passes this phase. This can be thought of as a representation of blocking events that often accompany sudden stratospheric warmings. The third transition path shown is an exception to the general pattern, making a final turn toward the east instead of to the west. This outlier of a reactive trajectory can also be seen in Figure 4.9, as the single green trajecory that decreases in X before decreasing in U instead of the other way around.

This kinematic sequence of events has a dynamical interpretation with precedent in prior literature. A critical ingredient of SSW is meridional eddy heat flux, which in this model takes the form $\overline{v'\Phi'_z} \propto hY$. This follows from the relationship between velocity, temperature, and the streamfunction:

$$\overline{v'\Phi'_z} \propto \overline{\psi'_x\psi'_z} \propto \overline{\mathrm{Re}\{ik\Psi e^{ikx}\}\mathrm{Re}\{(\Psi_z + (2H)^{-1}\Psi)e^{ikx}\}\sin^2(\ell y)e^{z/2H}} \tag{4.60}$$

$$\propto \overline{k(-X\sin kx - Y\cos kx)\mathrm{Re}\left\{\left(\frac{\Psi(z_T)-\Psi(z_B)}{z_T-z_B} + \frac{\Psi}{2H}\right)e^{ikx}\right\}} \tag{4.61}$$

$$\propto \overline{(-X\sin kx - Y\cos kx)\left[-\frac{\Psi_0}{z_T-z_B}\cos kx + \frac{X\cos kx - Y\sin kx}{2H}\right]} \tag{4.62}$$

$$\propto \frac{\Psi_0 Y}{2(z_T-z_B)} + \frac{XY-YX}{4H} \propto hY \tag{4.63}$$

where we have used the vertical finite-difference discretization in Eq. (4.61).

Figure 4.9: **Paths of maximal current superimposed on transition path samples.** All four Figures shown the same path, but from different vantage points. While the reactive probability density (Figures 4.7 and 4.8) says where transition paths spend their time, the reactive current is a vector field of the transition paths' local average directionality. The path shown in a color gradient from blue to red is a streamline of this vector field, representing an "average" transition path. The path is colored blue where the local committor is less than 0.5, and red otherwise. Note that the path can cross back and forth. The transition from red to blue, where the path first crosses the threshold $q^+ = \frac{1}{2}$ and enters the probabilistic B basin, is marked by a sudden drop in the U variable – a deceleration in zonal wind. At the same time, the path's rotations about A reverse direction, from clockwise to anti-clockwise, corresponding to a reversal in phase velocity of the streamfunction. This path accurately captures geometric tendencies of actual transition paths; five random samples of reactive trajectories are superimposed in green, the bulk of which cluster around the maximum current path. Panels (a) and (b) show cross-sections in XU and XY space respectively, while (c) shows a three-dimensional view. The parameters are $h = 30m$ and $\sigma_3 = 0.5 m/s/day^{1/2}$. Figure 4.10 shows the corresponding spacetime diagrams of the streamfunction.

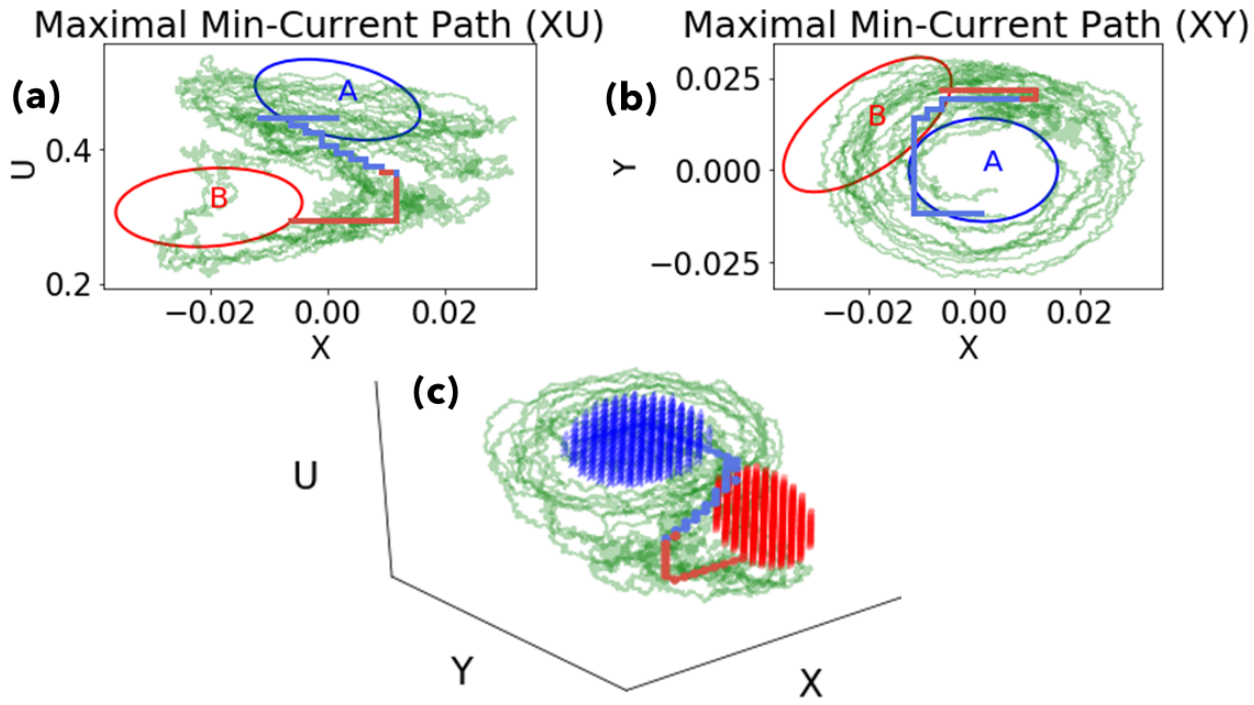
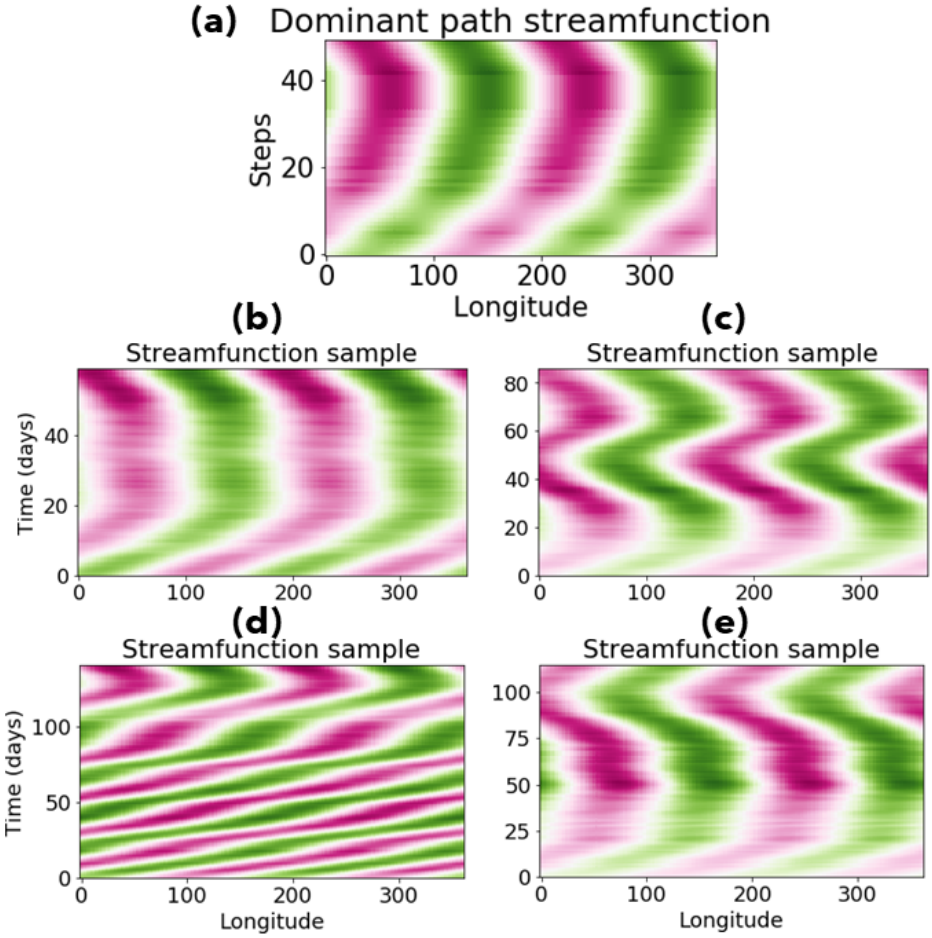


Figure 4.10: **Streamfunctions over time corresponding to the trajectories shown in Figure 4.9.** As $\psi' \propto (X \cos kx - Y \sin kx)$, the X and Y variables represent the phase of the streamfunction, whose movement we plot over time as a space-time diagram. Panel (a) shows the dominant transition path. The phase velocity is initially eastward, matching with the clockwise rotations in the XY plane as shown in Figure 4.9. The waves then slow down and reverse direction, matching with the anti-clockwise turn and zonal wind drop in Figure 4.9. The vertical axis plays the role of time, but the dominant path technically conveys only geometrical information. Hence, we measure it in discrete steps. Panels (b)-(d) show the streamfunctions over time corresponding to four of the green transition path samples in Figure 4.9, chosen randomly. Most exhibit the same slow-down and reversal behavior exemplified by the dominant path. The exception is sample (b), which turns to the east at the end of its path. This corresponds to the stray green trajectory visible in Figure 4.9(a), which enters set B from above and in the clockwise direction. Samples 1, 2 and 4 undergo some winding before the slowdown, but do slow down every time they reach the same phase.



This term shows up explicitly as a negative forcing on U in equation (15), showing that a reduced equator-to-pole temperature gradient in turn weakens the vortex via the thermal wind relation. The association of heat flux with SSW has been demonstrated in reanalysis [Sjoberg and Birner, 2012] and in detailed numerical simulations of internal stratospheric dynamics, even with time-independent lower boundary forcing [Scott and Polvani, 2006]. This relationship favors the phase $\phi = \frac{\pi}{2}$ as the most susceptible state for SSW onset, which is exactly picked out by the dominant transition pathway in Figure 4.9.

However, immediately after the wind starts weakening at $\phi = \frac{\pi}{2}$, where the streamfunction lines up with its lower boundary condition, the phase velocity reverses, giving rise to the westward lag of $\frac{3\pi}{16}$ we observed in the reactive density. A similar phase lag has also been observed in more detailed numerical studies. For example, Scott and Polvani [2006] observed a lag of $\frac{\pi}{2}$ across the whole stratosphere ($\frac{\pi}{4}$ at the mid-level), quite similar to our result. They found that vortex breakup was preceded by a long, slow build-up phase in which the vortex became increasingly vertically coherent, only to be ripped apart by an upward- and west-propagating wave. In an experiment with slowly-increasing lower boundary forcing, Dunkerton et al. [1981] saw a phase lag across the whole stratosphere that increased from $\sim 100^\circ$ to $\sim 180^\circ$ (50° to 90° between the lower boundary and the mid-stratosphere) over the course of the warming event. They attribute this phase tilt to the zonal wind rapidly reversing and carrying the streamfunction along. The weakening zonal wind simply removes the Doppler shift from the Rossby wave dispersion relation, $\omega = Uk - \beta k/(k^2 + \ell^2)$, allowing the waves to revert to their preferred westward phase velocity. This balance is also clear in equations (13-14): ignoring the damping and forcing terms, the dynamics read $[\dot{X}, \dot{Y}] = [(sU - r)Y, -(sU - r)X]$. For weak U, rotation in the XY plane is anti-clockwise and phase speed is westward.

Let us re-emphasize the probabilistic interpretation of reactive density. We have found that transitions from A to B are accompanied by anomalous increases in meridional heat flux. In other words, $\mathbb{P}\{\text{High heat flux}|\text{SSW}\}$ is high. As noted earlier, this does not imply that $\mathbb{P}\{\text{SSW}|\text{high heat flux}\}$

is also high. The identified streamfunction phase is not a "danger zone" in the sense that a trajectory entering this region is at higher risk of falling into state B; the committor alone conveys that information. Rather, a trajectory is highly likely to pass through that region *given* that it is reactive. Notably, reactive trajectories are unlikely to take a straight-line path from A to B with U, X and Y changing linearly. This unrealistic path would represent a zonally stationary streamfunction growing steadily in magnitude, while zonal wind falls off gradually. At higher noise levels, however, the system would be increasingly dominated by pure Brownian motion, and such a path would become more plausible.

We now turn to a quantitative comparison of committors and transition rates for different forcing and noise levels. These trends illustrate the effects of modeling choices and global change on the climatology of SSW. Planetary wave forcing, h, varies across days and seasons as well as different planets. The strength of additive noise, σ, is a modeling choice intended to represent gravity wave drag. Different stochastic parametrizations will vary in their effective σ value, and it is important to understand the sensitivity of SSW to model choices [Sigmond and Scinocca, 2010]. Furthermore, long-term climate change may cause both parameters to drift, altering the occurrence of SSW-induced severe weather events.

The measure of the relative "imminence" of a vacillating solution vs. a radiative solution, as described in the background section on committors, is the equilibrium density-weighted average committor, denoted $\langle q^+ \rangle_\pi$. Figure 11(a) shows this quantity for $25 \leq h \leq 45$ (m) and $0.4 \leq \sigma_3 \leq 2.0$ $(m/s/day^{1/2})$. Two trends are clearly expected from the basic physics of the model. First, as seen in Figures 4.7 and 4.8, $\langle q^+ \rangle_\pi$ should increase with h. Second, in the limit of large noise and infinite domain size, the dominance of Brownian motion will smooth out the committor function and make $\langle q^+ \rangle_\pi$ tend to an intermediate value between zero and one. On the other hand, as noise approaches zero, the dynamical system becomes increasingly deterministic, and the ultimate destination of a trajectory will depend entirely on which basin of attraction it starts in. The boundary, or *separatrix*, between these two basins is the stable manifold of the third (unstable) fixed point. In the case

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

Figure 4.11(a) plots the trends in $\langle q^+ \rangle_\pi$ as a function of h (along the horizontal axis) and σ_3 (along the vertical axis). The two basic hypotheses are verified: $\langle q^+ \rangle_\pi$ increases monotonically as h increases, and $\langle q^+ \rangle_\pi \sim \frac{1}{2}$ as σ increases, no matter the value of h. The less predictable behavior is in the range $h = 35\,m$, $0.75 \leq \sigma_3 \leq 1.0$, where $\langle q^+ \rangle_\pi$ displays non-monotonicity with respect to noise, at low noise levels. As σ_3 increases, the average committor increases from ~ 0.4 to ~ 0.6, and then decreases again. The four committor plots at the bottom of 4.11 illustrate the trend graphically. At low noise, the A basin includes winding passageways leading from the small-U region back to A. Small additive noise closes them off, effectively expanding the B basin. As noise increases and Brownian motion dominates the dynamics, committor values everywhere relax back to less-extreme values, reflecting the unbiased nature of Brownian motion.

Despite the coarse grid resolution, the first-order effect of noise is clear. At the low and high margins of h, where falling into state A and B respectively is virtually certain, an increase in noise decreases this virtual certainty, and the trend continues at larger noise to attenuate $\langle q^+ \rangle_\pi$ to its limit of 1/2. The middle h range, however, behaves differently. Whereas $h = 35\,m$ appears to balance out the basin sizes at low noise, a slight noise increase tends to kick the system out of the A basin and toward B, more so than the other way around. At higher noise, the committor relaxes back to 1/2. Examining the committor fields, it's clear that the isocommittor surface location does not move back and forth; rather, it moves toward A, and then the rest of the field flattens out.

Figure 4.11: **Behavior of the committor as a function of forcing h and noise σ_3.** (a) shows the average committor (weighted by equilibrium density): $\int q^+(x)\pi(x)\,dx$ evaluated for a range of h and σ values. Panels (b)-(e) show images of the committor for $h = 30$ and $\sigma = 0.5,\ 0.75,\ 1.0,\ 1.25$. The blue lobes in the lower part of the images, a shadow of the spiral structure from Figure 4.6, thin out and disappear with increasing forcing h.

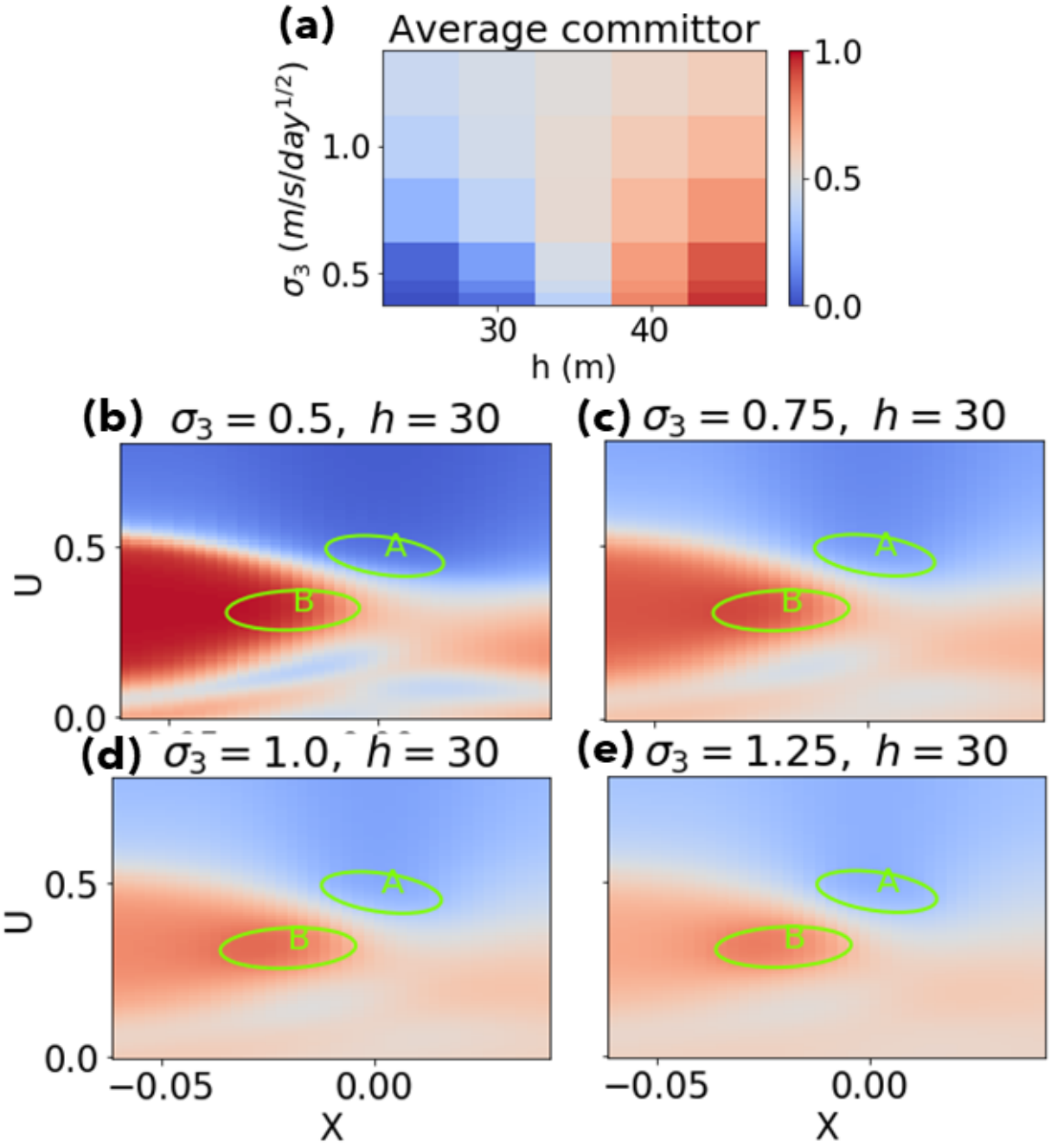
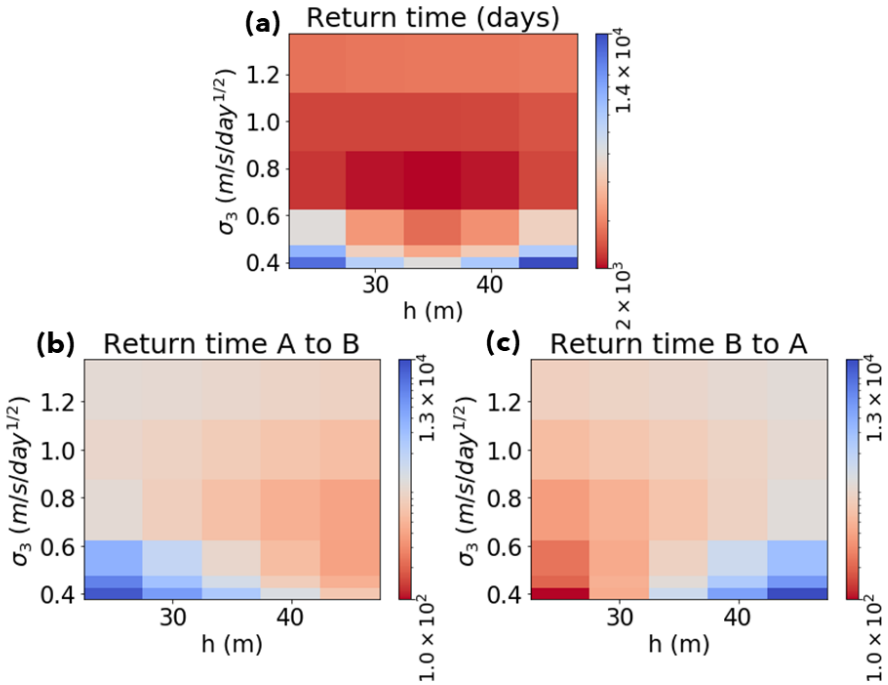


Figure 4.12 shows trends in the return times of SSW with varying h and σ_3. There are several different return times of interest. The first, shown in panel (a), is the total expected time between one transition event and the next, whose reciprocal is the *rate* R_{AB}, the number of transitions per unit time. The return time is a symmetric quantity between A and B, since every forward transition is accompanied by a backward one. Among the parameter combinations, $(h, \sigma_3) \approx (35m, 0.75 m/s/day^{1/2})$ is the one which minimizes return time, or equivalently maximizes the transitions per unit time. $h = 35m$ is a forcing level which approximately balances out the time spent between the two sets, making transitions relatively common. At lower noise, transitions are exceedingly rare, and at higher noise the two states cease to be long-lived. However, this symmetric quantity does not capture information about the relative speed of transition from A to B vs. from B back to A. Panel (b) shows a different passage time, which is the average time between the end of a backward $(B \to A)$ transition and the end of the next forward $(A \to B)$ transition, which we call T_{AB}. This is computed as the reciprocal of the rate constant k_{AB}, as described in the previous section. In other words, the stopwatch begins when the system returns to A after having last visited B, and ends when the system next hits B. This metric is asymmetric: a smaller $A \to B$ return time indicates that the forward transition is faster than the backward transition. Panel (b) shows the complementary $B \to A$ return time, T_{BA}. Unsurprisingly, an increase in h causes a decrease in T_{AB} and an increase in T_{BA} regardless of the noise level. The noise level has a less obvious effect. Whereas T_{BA} decreases monotonically with increasing noise, regardless of h, the forward time T_{AB} is minimized by a mid-range noise level of $\sigma_3 \approx 0.75 m/s/day^{1/2}$. This is another reflection of the bias toward state B that is effected by adding noise to a very low baseline.

4.6 Conclusion

Transition path theory is a framework for describing rare transitions between states. We have described TPT along with a number its key ingredients like the forward and backward committor functions. While TPT has been applied primarily to molecular systems, we believe it offers valu-

Figure 4.12: **Behavior of return times as a function of forcing h and noise σ_3.** Panel (a) shows the average period between the start of one transition event and the start of the next. Red here means many transitions per unit time, both $A \to B$ and $B \to A$. We call this the return time, and calculate it as the reciprocal of R, the number of forward (or backward) transitions per unit time. There is clearly a parameter set: $(h, \sigma_3) \approx (35\,m, 0.75\,m/s/day^{1/2})$, which optimizes the number of transitions per unit time. Below this noise level, internal variability is scarcely enough to jump between regions. Above this noise level, sets A and B are no longer metastable, and excursions are so wide and frequent that passing from set A to set B is a very spatially restricted event. Panels (b) and (c), below, distinguish forward and backward transition times. Panel (b) shows the expected passage time T_{AB}, the interval between the end of a $B \to A$ transition and the end of the next $A \to B$ transition. Panel (c) shows the analogous backward passage time, T_{BA}. Note the scales are logarithmic, and here red simply means faster transitions, regardless of which direction is being considered.



able insight into climate and weather phenomena such as sudden stratospheric warming, primarily through committors, reactive densities, and reactive currents. Of interest apart from its role in TPT, the forward committor defines an optimal probabilistic forecast, borne out by direct numerical simulation experiments. The reactive densities and currents describe the geometric properties of dominant transition mechanisms at low noise. In applying TPT to a noisy, truncated Holton-Mass model, we find that transitions tend to begin with a drop in mean zonal wind and a reversal of the streamfunction's phase velocity at a particular streamfunction phase. This is consistent with the significance of blocking precursors to SSW as found in Martius et al. [2009], insofar as this idealized model can represent them. We also find that noise has a non-monotonic effect on the overall preference for a vacillating state, measured by the average committor. At a forcing of $h \approx 35\,m$, where the isocommittor surface (essentially the basin boundary of the deterministic dynamics) divides the space approximately in half, we find that raising the noise tilts the balance decisively toward the vacillating solution. Still larger noise evens the whole field out. The transition rate constant shows a similar dependence on h and σ_3.

This study is a foundation for the following chapters and future studies on more realistic and complex models. In this high-dimensional setting, the generator is unknown or computationally intractable. Any state space with more than ~ 5 degrees of freedom is beyond the reach of a finite-volume discretization, because the number of grid cells increases exponentially with dimension. While here we represented the generator as a finite-volume or finite-difference operator on a grid, one can also write it in a basis of globally coherent functions, such as Fourier modes, or more generally harmonic functions on a manifold. Given only data, without an explicit form of the dynamics, this manifold and the basis functions can be estimated from (for example) the diffusion maps algorithm, and the generator's action on this basis can be approximated from short trajectories.

5 LEARNING FORECASTS OF RARE STRATOSPHERIC TRANSITIONS FROM SHORT SIMULATIONS

The previous chapter introduced TPT and its associated quantities in a low-order model. This chapter and the next take a step up the model hierarchy to work with a more complex, though still idealized, SSW model. While a full TPT analysis is postponed to chapter 6, this chapter focuses only on two key quantities of interest: the probability of an SSW occurring, and the expected lead time if it does occur, as functions of initial condition. These *statistically optimal forecasts* concretely measure the event's progress. Direct numerical simulation can estimate them in principle, but is prohibitively expensive in practice: each rare event requires a long integration to observe, and the cost of each integration grows with model complexity. We describe and implement an alternative approach which uses integrations that are *short* compared to the timescale of the warming event. We compute the probability and lead time efficiently by solving equations involving the transition operator, which encodes all information about the dynamics. We relate these optimal forecasts to a small number of interpretable physical variables, suggesting optimal measurements for forecasting. We illustrate the methodology on a prototype SSW model developed by Holton and Mass (1976) and modified by stochastic forcing. This model captures the essential nonlinear dynamics of SSWs and exhibits the key forecasting challenge: the dramatic separation in timescales between a single event and the return time between successive events. Our methodology is designed to fully exploit high-dimensional data from models and observations, and has the potential to identify detailed predictors of many complex rare events in meteorology.

This chapter is adapted from the publication Finkel et al. [2021b].

5.1 Introduction

We focus on two forecasts in particular to quantify risk. The *committor* is the probability that a given initial condition evolves directly into an "extreme" configuration B rather than a "typical"

configuration A, where extreme and typical are user-defined. Given that it does reach B first, the *conditional mean first passage time*, or *lead time*, is the expected time that it takes to get there. The committor appears prominently in the molecular dynamics literature, with some recent applications in geoscience including Tantet et al. [2015], Lucente et al. [2019], and Finkel et al. [2020], which compute the committor for low-dimensional atmospheric models.

Both quantities depend on the initial condition, defining functions over d-dimensional state space that encode important information regarding the fundamental causes and precursors of the rare event. However, "decoding" the physical insights is not automatic. With real-time measurement constraints, the risk metrics must be estimated from low-dimensional proxies. Even visualizing them requires projecting down to one or two dimensions. This calls for a principled selection of low-dimensional coordinates which are both physically meaningful and statistically informative for our chosen risk metrics. We address this problem using sparse regression, a simple but easily extensible solution with the potential to inform optimal measurement strategies to estimate risk as precisely as possible under constraints.

To overcome the curse of dimensionality—the main challenge in estimating the committor and lead time—we employ the dynamical Galerkin approximation (DGA) method presented in 2.4. In brief, the method takes a large data set of short-time independent simulations and approximately solves Feynman-Kac formulae [E et al., 2019], which relate long-time forecasts to instantaneous tendencies. These equations are elegant and general, but computationally daunting: in the continuous time and space limit, they become partial differential equations (PDE) with d independent variables—the same as the model state space dimension. It is therefore hopeless to solve the equations using any standard spatial discretization. But, as we demonstrate, the equations can be solved with remarkable accuracy by expanding in a basis of functions informed by the data set.

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. In particular, this system captures the key difficulty in sampling rare events. The vast majority of the time, the system sits in one of two metastable states, characterizing a strong or weak vortex respectively. Extreme events are the infrequent jumps from one state to another. Our computational framework can accurately characterize these rare transitions using only a data set of short model simulations, short not only compared to the long periods the system sits in one state or the other, but also relative to the timescale of the transition events themselves. In the future, the same methodology could be applied to query the properties of more complex models, such as GCMs, where less theoretical understanding is available.

In section 5.2, we specify the dynamical model and define the specific rare event of interest. In section 5.3, we formally define the risk metrics introduced above and visualize the results for the Holton-Mass model, including a discussion of physical and practical insights gleaned from our approach. In section 5.4 we identify an optimal set of reduced coordinates for estimating risk using sparse regression. These results will provide motivation for the computational method, which we present afterward in section 5.5 along with accuracy tests. We then lay out future prospects and conclude in section 5.6.

5.2 Holton-Mass model

We return to the original Holton-Mass model, the wave-mean flow interaction represented by Eqs. (3.13) and (3.14). Whereas chapter 4 worked with a one-level version of the model due to Ruzmaikin et al. [2003], here we work with the original discretization by Holton and Mass [1976]. The two parameters controlling the background forcing are γ, the vertical shear of the radiative zonal wind profile $U^R(z)$, and h, the amplitude of orographic forcing. 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

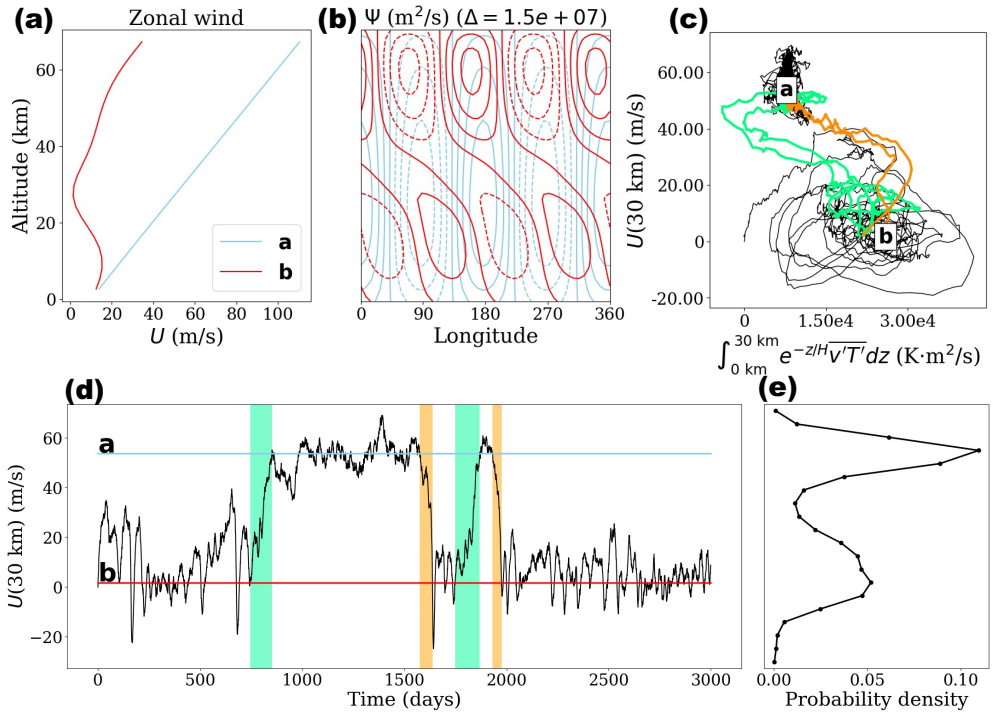


Figure 5.1: **Illustration of the two stable states of the Holton-Mass model and transitions between them.** (a) Zonal wind profiles of the radiatively maintained strong vortex (the fixed point **a**, blue) which increases linearly with altitude, and the weak vortex (the fixed point **b**, red) which dips close to zero in the mid-stratosphere. (b) Streamfunction contours are overlaid for the two equilibria **a** and **b**. (c) Parametric plot of a control simulation in a 2-dimensional state space projection, including two transitions from A to B (orange) and B to A (green). (d) Time series of $U(30\text{ km})$ from the same simulation. (e) The steady state density projected onto $U(30\text{km})$.

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 stationary wave, as well as a weak vortex with U dipping close to zero at an intermediate altitude and a stationary wave with strong westward phase tilt. The two stable equilibria, which we call **a** and **b**, are illustrated in Fig. 5.1(a,b) by their z-dependent zonal wind and perturbation streamfunction profiles.

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. The

form of noise will be specified in Eq. (5.3). While the details of the additive noise are ad hoc, the general 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]. In general, unobserved deterministic dynamics can make the system non-Markovian, which technically violates the assumptions of our methodology. However, with sufficient separation of timescales the Markovian assumption is not unreasonable. Furthermore, memory terms can be ameliorated by lifting data back to higher-dimensional state space with time-delay embedding [Berry et al., 2013, Thiede et al., 2019, Lin and Lu, 2021].

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

$$\begin{aligned}\mathbf{X}(t) = \Big[&\mathrm{Re}\{\Psi\}(\Delta z,t),\ldots,\mathrm{Re}\{\Psi\}(z_{top}-\Delta z,t),\\ &\mathrm{Im}\{\Psi\}(\Delta z,t),\ldots,\mathrm{Im}\{\Psi\}(z_{top}-\Delta z,t),\\ &U(\Delta z,t),\ldots,U(z_{top}-\Delta z,t)\Big] \in \mathbb{R}^d = \mathbb{R}^{75}\end{aligned} \tag{5.1}$$

The deterministic system can be written $d\mathbf{X}(t)/dt = v(\mathbf{X}(t))$ for a vector field $v : \mathbb{R}^d \to \mathbb{R}^d$ specified by discretizing (3.13) and (3.14). 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 from an

ordinary differential equation into a stochastic differential equation, spefically an Itô diffusion:

$$d\mathbf{X}(t) = v(\mathbf{X}(t))\,dt + \sigma(\mathbf{X}(t))\,d\mathbf{W}(t), \tag{5.2}$$

where $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ imparts a correlation structure to the vector $\mathbf{W}(t) \in \mathbb{R}^m$ of independent standard white noise processes. As discussed above, 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 which respects boundary conditions at the bottom and top of the stratosphere. Its structure is defined by the following Euler-Maruyama scheme: in a timesetep $\delta t = 0.005$ days, after a deterministic forward Euler step we add the stochastic perturbation to zonal wind on large vertical scales

$$\delta U(z) = \sigma_U \sum_{k=0}^{m} \eta_k \sin\left[\left(k+\frac{1}{2}\right)\pi\frac{z}{z_{top}}\right]\sqrt{\delta t} \tag{5.3}$$

where η_k $(k = 0,1,2)$ are independent unit normal samples, $m = 2$, and σ_U is a scalar that sets the magnitudes of entries in σ. In terms of physical units,

$$\sigma_U^2 = \frac{\mathbb{E}[(\delta U)^2]}{\delta t} \approx (1\,\text{m/s})^2/\text{day} \tag{5.4}$$

σ_U 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. We have experimented with this value, and found that reducing the noise level below 0.8 dramatically reduces the frequency of transitions, while increasing it past 1.5 washes out metastability. We keep σ_U constant going forward as a favorable numerical regime to demonstrate our approach, while acknowledging that the specifics of stochastic parameterization are important in general to obtain accurate forecasts. The resulting matrix σ is 75×3, with nonzero entries only in the last 25 rows as forcing only applies to $U(z)$.

A long 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 time series of $U(30\,\mathrm{km})$ in panel (d) of Fig. 5.1. Panel (e) shows a projection of the steady state distribution, also known as the equilibrium/invariant distribution, of U as a function of z. We call this density $\pi(\mathbf{x})$, which is a function over the full d-dimensional state space. We focus on the zonal wind U at 30 km following Christiansen [2000], because this is where its 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 $\mathbb{R}^d$

$$A = \{\mathbf{x} : U(30\,\mathrm{km})(\mathbf{x}) \geq U(30\,\mathrm{km})(\mathbf{a}) = 53.8\,\mathrm{m/s}\},$$
$$B = \{\mathbf{x} : U(30\,\mathrm{km})(\mathbf{x}) \leq U(30\,\mathrm{km})(\mathbf{b}) = 1.75\,\mathrm{m/s}\}.$$

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. We use 30 km for consistency with Christiansen (2000); this is technically higher than 10 hPa because $z = 0$ in the Holton-Mass model represents the tropopause. Our method is equally applicable to any definition, and the results are not qualitatively dependent on this choice. Incidentally, the analysis tools we present may be helpful in distinguishing predictability properties between different definitions. In fact, we will show that the height neighborhood of 20 km is actually more salient for predicting the event than wind at the 30-km level, even when the event is defined by wind at 30 km. This emerges from statistical analysis alone, and gives us confidence that essential SSW properties are stable with respect to reasonable changes in definition.

The orange highlights in Fig. 5.1 (d) begin when the system exits the A region bound for B, and end when the system enters B. The green highlights start when the system leaves B bound for A, and end when A is reached. Note that $A \to B$ transitions, SSWs, are much shorter in duration than $B \to A$ transitions. Fig. 5.1 (c) shows the same paths, but viewed parametrically in a two-dimensional state space consisting of integrated heat flux or IHF $\int_{0\,\mathrm{km}}^{30\,\mathrm{km}} e^{-z/H}\overline{v'T'}\,dz$, and zonal

wind $U(30\text{ km})$. IHF is an informative number because it captures both magnitude and phase information of the streamfunction in the Holton-Mass model:

$$\text{IHF} = \int_{0\text{ km}}^{30\text{ km}} e^{-z/H}\overline{v'T'}\,dz \propto \int_{0\text{ km}}^{30\text{ km}} |\Psi|^2 \frac{\partial \varphi}{\partial z}\,dz \tag{5.5}$$

where φ is the phase of Ψ. The $A \to B$ and $B \to A$ transitions are again highlighted in orange and green 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.

5.3 Forecast functions: the committor and lead time statistics

5.3.1 Defining risk and lead time

We will introduce the quantities of interest by way of example. First, suppose the stratosphere is observed in an initial state $\mathbf{X}(0) = \mathbf{x}$ that is neither in A nor B, so $U(30\text{km})(\mathbf{b}) < U(30\text{km})(\mathbf{x}) < U(30\text{km})(\mathbf{a})$ and the vortex is somewhat weakened, but not completely broken down. We call this intermediate zone $D = (A \cup B)^{\text{c}}$ (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

$$\tau_{D^{\text{c}}} = \min\{t \geq 0 : \mathbf{X}(t) \in D^{\text{c}}\} \tag{5.6}$$

Here, D^{c} emphasizes that the process has left D, i.e., gone to A or B. The first-exit location $\mathbf{X}(\tau_{D^{\text{c}}})$ is itself a random variable which importantly determines how the system exits D: either $\mathbf{X}(\tau_{D^{\text{c}}}) \in A$, meaning the vortex restores to radiative equilibrium, or $\mathbf{X}(\tau_{D^{\text{c}}}) \in B$, meaning the vortex breaks down into vacillation cycles. A fundamental goal of forecasting is to determine the probabilities of

these two events, which naturally leads to the definition of the (forward) committor function

$$q^+(\mathbf{x}) = \begin{cases} \mathbb{P}_{\mathbf{x}}\{\mathbf{X}(\tau_{D^c}) \in B\} & \mathbf{x} \in D = (A \cup B)^c \\ 0 & \mathbf{x} \in A \\ 1 & \mathbf{x} \in B \end{cases} \tag{5.7}$$

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 chapter (but which will be important in chapter 6). 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.

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. 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. These same quantities have been calculated previously in other simplified models, e.g. Birner and Williams [2008] and Esler and Mester [2019].

$\mathbb{E}_{\mathbf{x}}[\tau_B]$ has an obvious shortcoming: it is an average over all paths starting from $\mathbf{x}$, including those which go straight into B (i.e., an orange trajectory in Fig. 5.1c,d) and the rest which return to A i.e., a green trajectory) and linger there, potentially for a very long time, before eventually re-crossing back into B. It is more relevant for near-term forecasting to condition τ_B on the event that an SSW is coming before the strong vortex returns. For this purpose, we introduce the *conditional*

mean first passage time, or lead time, to B:

$$\eta^+(\mathbf{x}) := \mathbb{E}_{\mathbf{x}}[\tau_B | \tau_B < \tau_A] \tag{5.8}$$

which quantifies the suddenness of SSW.

All of these quantities can, in principle, be estimated by direct numerical simulation (DNS). (We use "DNS" to mean a single-threaded integration of a model, as opposed to a parallel integration from many initial conditions. This departs from the computational fluid dynamics usage, where it means "without subgrid closure".) For example, suppose we observe an initial condition $\mathbf{X}(0) = \mathbf{x}$ in an operational forecasting setting, and wish to estimate the probability and lead time for the event of next hitting B. We would initialize an ensemble $\{\mathbf{X}_n(0) = \mathbf{x}, n = 1, \ldots, N\}$ and evolve each member forward in time until it hits A or B at the random time τ_n. In an explicitly stochastic model, random forcing would drive each member to a different fate, while in a deterministic model their initial conditions would be perturbed slightly. To estimate the committor to B, we could calculate the fraction of members that hit B first. Averaging the arrival times (τ_n), over only those members gives an estimate of the lead time to B. For a single initial condition $\mathbf{x}$ reasonably close to B, DNS may be the most economical. But how do we systematically compute $q^+(\mathbf{x})$ over all of state space (here 75 variables, but potentially billions of variables in a GCM or other state-of-the-art forecast system)?

For this more ambitious goal, DNS is prohibitively expensive. By definition, transitions between A and B are infrequent. Therefore, if starting from $\mathbf{x}$ far from B, a huge number of sampled trajectories (N) will be required to observe even a small number ending in B, and they may take a long time to get there. If instead we could precompute these functions offline over all of state space, the online forecasting problem would reduce to "reading off" the committor and lead time with every new observation. Achieving this goal is the key point of our paper, and we achieve this using the dynamical Galerkin approximation, or DGA, recipe described by Thiede et al. [2019].

A brute force way to estimate these functions is to integrate the model for a long time until it

reaches statistical steady state, meaning it has explored its attractor thoroughly according to the steady state distribution. After long enough, it will have wandered close to every point $\mathbf{x}$ sufficiently often to estimate $q^+(\mathbf{x})$ and η^+ robustly as in DNS. We have performed such a "control simulation" of 5×10^5 days for validation purposes, but our main contribution in this chapter is to compute the forecast functions using only *short* trajectories with DGA, allowing for massive parallelization. However, we will defer the methodological details to Section 5, and first justify the effort with some results. We visualize the committor and lead time computed from short trajectories and elaborate on their interpretation, utility, and relationship to ensemble forecasting methods.

5.3.2 Steady state distribution

Before visualizing the committor and lead time, it will be helpful to have a precise notion of the steady state distribution, denoted $\pi(\mathbf{x})$, a probability density that describes the long-term behavior of a stochastic process $\mathbf{X}(t)$. Assuming the system is ergodic, averages over time are equivalent to averages over state space with respect to π. That is, for any well-behaved function $g : \mathbb{R}^d \to \mathbb{R}$,

$$\langle g \rangle_\pi := \lim_{T\to\infty} \frac{1}{T} \int_0^T g(\mathbf{X}(t))\, dt = \int_{\mathbb{R}^d} g(\mathbf{x})\pi(\mathbf{x})\, d\mathbf{x} \tag{5.9}$$

For example, if $g(\mathbf{x}) = \mathbb{1}_S(\mathbf{x})$ (an indicator function, which is 1 for $\mathbf{x} \in S \subset \mathbb{R}^d$ and 0 for $\mathbf{x} \notin S$), Eq. (5.9) says that the fraction of time spent in S can be found by integrating the density over S. The density peaks in Fig. 5.1(d) indicates clearly that the neighborhoods of **a** and **b** are two such regions with especially large probability under π. Note that both sides of (5.9) are independent of the initial condition, which is forgotten eventually. Short-term forecasts are by definition out-of-equilibrium processes, depending critically on initial conditions; however, $\pi(\mathbf{x})$ is important to us here as a "default" distribution for missing information. If the initial condition is only partially observed, e.g. in only one coordinate, we have no information about the other $d-1$ dimensions, and in many cases the most principled tactic is to assume those other dimensions are distributed

according to π, conditional on the observation.

5.3.3 Visualizing committor and lead times

The forecasts $q^+(\mathbf{x})$ and $\eta^+(\mathbf{x})$ are functions of a high-dimensional space $\mathbb{R}^d$. However, these degrees of freedom may not all be "observable" in a practical sense, given the sparsity and resolution limits of weather sensors, and visualizing them requires projecting onto reduced-coordinate spaces of dimension 1 or 2. We call these "collective variables" (CVs) following chemistry literature [e.g., Noé and Clementi, 2017], and denote them as vector-valued functions from the full state space to a reduced space, $\theta : \mathbb{R}^d \to \mathbb{R}^k$, where $k = 1$ or 2. For instance, Fig. 5.1 (c) plots trajectories in the CV space consisting of integrated heat flux and zonal wind at 30 km: $\theta(\mathbf{x}) = \left(\int_{0\text{ km}}^{30\text{km}} e^{-z/H} \overline{v'T'}\, dz, U(30\text{km}) \right)$. The first component is a nonlinear function involving products of $\mathrm{Re}\{\Psi\}$ and $\mathrm{Im}\{\Psi\}$, while the second component is a linear function involving only U at a certain altitude. For visualization in general, we have to approximate a function $F : \mathbb{R}^d \to \mathbb{R}$, such as the committor or lead time, as a function of reduced coordinates. That is, we wish to find $f : \mathbb{R}^k \to \mathbb{R}$ such that $F(\mathbf{x}) \approx f(\theta(\mathbf{x}))$. Given a fixed CV space θ, an "optimal" f is chosen by minimizing some function-space metric between $f \circ \theta$ and F.

A natural choice is the mean-squared error weighted by the steady state distribution π, so the projection problem is to minimize over functions $f : \mathbb{R}^k \to \mathbb{R}$ the penalty

$$\begin{aligned} S[f;\theta] &:= \|f \circ \theta - F\|^2_{L^2(\pi)} \\ &= \int_{\mathbb{R}^d} \Big[f(\theta(\mathbf{x})) - F(\mathbf{x}) \Big]^2 \pi(\mathbf{x})\, d\mathbf{x}. \end{aligned} \tag{5.10}$$

The optimal f for this purpose is the conditional expectation

$$\begin{aligned} f(\mathbf{y}) &= \mathbb{E}_{\mathbf{X}\sim\pi}[F(\mathbf{X}) | \theta(\mathbf{X}) = \mathbf{y}] \\ &= \lim_{|d\mathbf{y}|\to 0} \frac{\int f(\mathbf{x}) \mathbb{1}_{d\mathbf{y}}(\theta(\mathbf{x})) \pi(\mathbf{x})\, d\mathbf{x}}{\int \mathbb{1}_{d\mathbf{y}}(\theta(\mathbf{x})) \pi(\mathbf{x})\, d\mathbf{x}} \end{aligned} \tag{5.11}$$

where $d\mathbf{y}$ is a small neighborhood about $\mathbf{y}$ in CV space R^k. The subscript $\mathbf{X} \sim \pi$ means that the expectation is with respect to a random variable $\mathbf{X}$ distributed according to $\pi(\mathbf{x})$, i.e., at steady state. Fig. 5.2 uses this formula to display one-dimensional projections of the committor (first row) and lead time (second row), as well as the one-standard deviation envelope incurred by projecting out the other 74 degrees of freedoom. This "projection error" is defined as the square root of the conditional variance

$$V_F(\mathbf{y}) = \mathbb{E}_{\mathbf{X}\sim\pi}\Big[\Big(F(\mathbf{X}) - f(\mathbf{y})\Big)^2 \Big| \theta(\mathbf{X}) = \mathbf{y}\Big]. \tag{5.12}$$

Each quantity is projected onto two different one-dimensional CVs: $U(30\text{ km})$ (first column) and IHF (second column). In panel (a), for example, we see the committor is a decreasing function of U: the weaker the wind, the more likely a vortex breakdown. Moreover, the curve provides a conversion factor between risk (as measured by probability) and a physical variable, zonal wind. An observation of $U(30\text{ km}) = 38$ m/s implies a 50% chance of vortex breakdown. The variation in slope also tells us that a wind reduction from 40 m/s to 30 m/s represents a far greater increase in risk than a reduction from 30 m/s to 20 m/s. Meanwhile, panel (b) shows the committor to be an increasing function of IHF, since SSW is associated with large wave amplitude and phase lag. However, IHF seems inferior to zonal wind as a committor proxy, as a small change in IHF from ~ 15000 to ~ 20000 K·m^2/s corresponds to a sharp increase in committor from nearly zero to nearly one. In other words, knowing only IHF doesn't provide much useful information about the threat of SSW until it is already virtually certain. The dotted envelope is also wider in panel (b) than (a), indicating that projecting the committor onto IHF removes more information than projecting onto U. While the underlying noise makes it impossible to divine the outcome with certainty from *any* observation, the projection error clearly privileges some observables over others for their predictive power.

In panels (c) and (d), the lead time is seen to have the opposite overall trend as the committor: the weaker the wind, or the greater the heat flux, the closer you are on average to a vortex

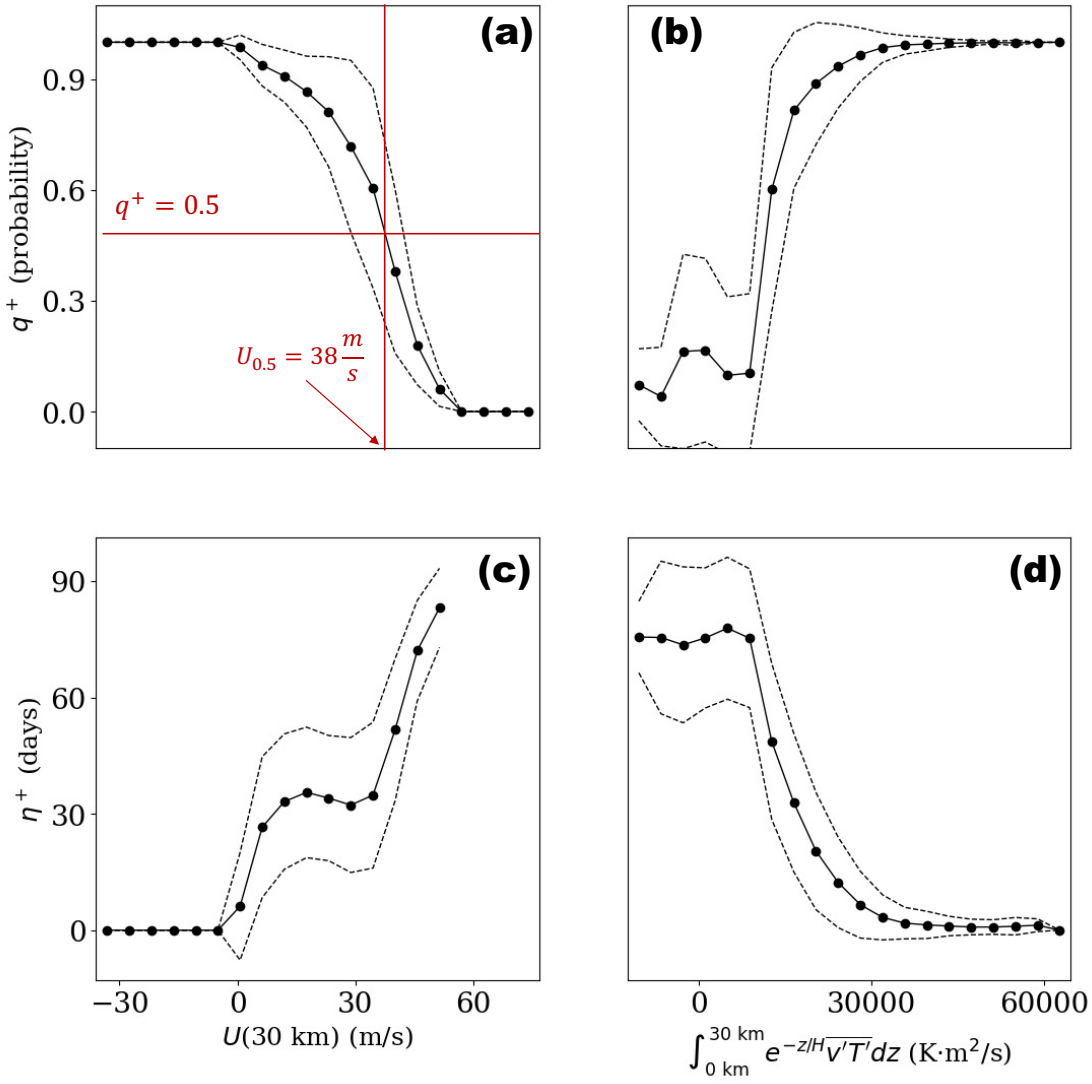


Figure 5.2: **One-dimensional projections of the forward committor (first row) and lead time to B (second row).** These functions depend on all $d = 75$ degrees of freedom in the model, but we have averaged across $d - 1 = 74$ dimensions to visualize them as rough functions of two single degrees of freedom: $U(30\text{ km})$ (first column) and integrated heat flux up to 30 km, IHF (second column). Panel (a) additionally marks the $q^+ = \frac{1}{2}$ threshold and the corresponding value of zonal wind.

breakdown. $\eta^+(\mathbf{x})$ is not defined when wind is strongest, as $\mathbf{x} \in A$ and so $q^+(\mathbf{x}) = 0$. However, an interesting exception to the trend occurs in the range $10\ \mathrm{m/s} \leq U \leq 40\ \mathrm{m/s}$: the expected lead time stays constant or slightly *decreases* as zonal wind increases, and the projection error remains large. This means that while the probability of vortex breakdown increases rapidly from 50% to 90%, the time until vortex breakdown remains highly uncertain. To resolve this seeming paradox, we will have to visualize the joint variation of q^+ and η^+.

It is of course better to consider multiple observables at once. Fig. 5.3 shows the information gained beyond observing $U(30\ \mathrm{km})$ by incorporating IHF as a second observable. In the top row we project π, q^+, and η^+ onto the two-dimensional subspace, revealing structure hidden from view in the one-dimensional projections. Panel (a) is a 2-dimensional extension of Fig. 5.1(d), with density peaks visible in the neighborhoods of **a** and **b**. The white space surrounding the gray represents physically insignificant regions of state space that was not sampled by the long simulation. The same convention holds for the following two-dimensional figures. The committor is displayed in panel (b) over the same space. It changes from blue at the top (an SSW is unlikely) to red at the bottom (an SSW is likely), bearing out the negative association between U and q^+. However, there are non-negligible horizontal gradients that show that IHF plays a role, too. Likewise, the lead time in panel (c) decreases from ~ 90 days near **a** to 0 days near **b**, when the transition is complete. Here, IHF appears even more critically important for forecasting how the event plays out, as gradients in η^+ are often completely horizontal.

A horizontal dotted line in Fig. 5.3(a-c) marks the 50% risk level $U(30\ \mathrm{km}) = 38$ m/s, but the committor varies along it from low risk at the left to high risk at the right: we show this concretely by selecting two points θ_0 and θ_1 along the line. According to U alone, i.e., the curve in Fig. 5.2(a), both would have the same committor of 0.5. According to both U and IHF together, i.e., the two-dimensional heat map in Fig. 5.3(b), they have very different probabilities of $q^+(\theta_0) = 0.31$ and $q^+(\theta_1) = 0.73$: an SSW is more than twice as likely to occur from starting point θ_1 as θ_0.

While those committor values come from the DGA method to be described in Section 5, we

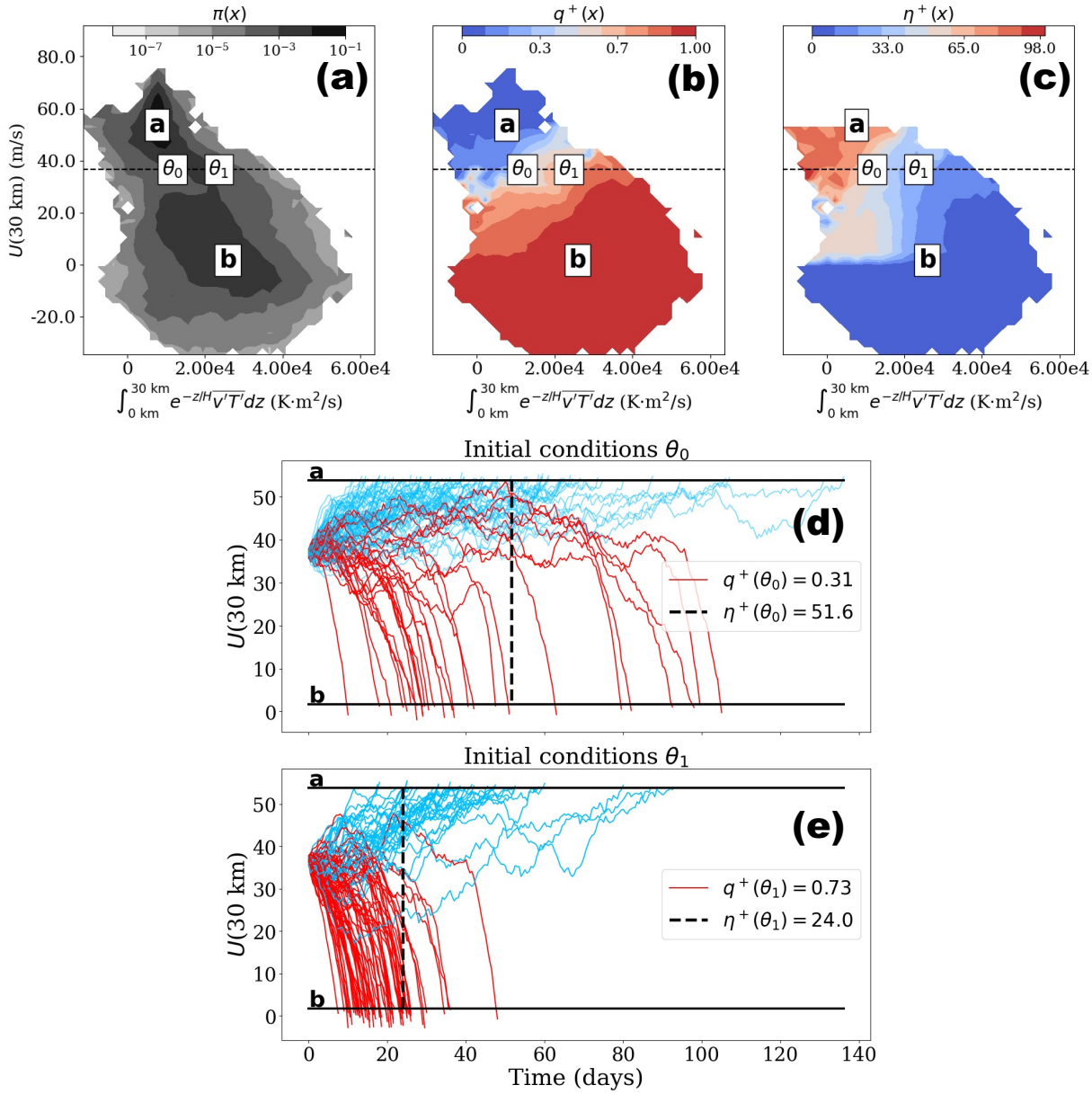


Figure 5.3: **The density, committor, and lead time as functions of zonal wind and integrated heat flux.** Panel (a) projects the steady state distribution $\pi(\mathbf{x})$ onto the two-dimensional subspace (U,IHF) at 30 km. The white regions surrounding the gray are unphysical states with negligible probability. Panels (b) and (c) display the committor and lead time in the same space. A horizontal transect marks the level $U(30\text{ km}) = 38.5$ m/s, where q^+ according to U only is 0.5. Panels (d) and (e) show ensembles initialized from two points θ_0 and θ_1 along the transect, verifying that their committor and lead time values differ from their values according to U, in a way that is predictable due to considering IHF in addition to U.

confirm them empirically by plotting an ensemble of 100 trajectories originating from each of the two initial conditions in panels (d) and (e) below, coloring A-bound trajectories blue and B-bound trajectories red. Only 28% of the sampled trajectories through θ_0 exhibit an SSW, next going to state B, while 68% of the integrations from θ_1 end at B. In both cases, the heatmaps and ensemble sample means roughly match. The small differences between the projected committor and the empirical "success" rate of trajectories arise both from errors in the DGA calculation (which we analyze in section 5) and the finite size of the ensemble.

The lead time prediction is improved similarly by incorporating the second observable. According to U alone, Fig. 5.2 predicts a lead time of 40 days for both θ_0 and θ_1. Considering IHF additionally, the two-dimensional heat map in Fig. 5.3 predicts a lead time of 52 days and 24 days for θ_0 and θ_1, respectively. Referring to the ensemble from θ_1 in panels (d) and (e), the arrival times of red trajectories to B provide a discrete sampling of the lead time distributions of $\tau_B|\tau_B < \tau_A$. The sample means are 50 and 32 days respectively from θ_0 and θ_1, again roughly matching with our predictions.

These two-dimensional projections still leave out 73 remaining dimensions, which we could incorporate to make the forecasts even better. After accounting for all 75 dimensions, we would obtain the full committor function $q^+ : \mathbb{R}^d \to \mathbb{R}$. This is still a probability, i.e., an expectation over the unresolved turbulent processes and uncertain initial condition. Low-dimensional committor projections simply treat the projected-out dimensions as random variables sampled according to π. Whether projected to a space of 1 or 75 dimensions, the committor is the function of that space that is closest, in the mean-square sense, to the binary indicator $\mathbb{1}_B(\mathbf{X}(\tau))$; this is the defining characteristic of conditional expectation [Durrett, 2013]. In the case that the system does hit B next, the lead time is closest in the mean-square sense to τ_B.

While high-dimensional systems offer many coordinates to choose from, we argue that the committor and lead time are the most important nonlinear coordinates to monitor for forecasting purposes. We will explore their relationship in the next subsection. Although both encode some

version of proximity to SSW, they are independent variables which deserve separate consideration.

5.3.4 Relationship between risk and lead time

A forecast is most useful if it comes sufficiently early (to leave some buffer time before impact) and is sufficiently precise to time your response. For example, in June we can say with certainty it will snow next winter in Minnesota. To be useful, we want to know the date of the first snow as early as possible. By relating levels of risk (quantified by q^+) and lead time (quantified by η^+), we can now assess the limits of early prediction. Such a relationship would answer two questions: for an SSW transition, (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 ignorant could we be of it?

The committor and lead time have an overall negative relationship, but they do not completely determine each other, as the contours in Fig. 5.3(a,b) do not perfectly line up. We treat them as independent variables in Fig. 5.4, which maps zonal wind and IHF as functions of the coordinates q^+ and η^+ in an inversion of Fig. 5.3. The density $\pi(\mathbf{x})$ projected on this space in 4(a) shows again a bimodal structure around **a** and **b**, which occupy opposite corners of this space by construction. Meanwhile, zonal wind and IHF are indicated by the shading in panels (b) and (c). The bridge between **a** and **b** is not a narrow band, but rather includes a curious high-committor, high-lead time branch which seems paradoxical: points at $q^+ = 0.9$ have a greater spread in η^+ than points at $q^+ = 0.5$, contrary to the intuition that closeness to B in probability means closeness in time. The color shading shows that q^+ is strongly associated with U(30 km), while η^+ is more strongly associated with IHF(30 km). In particular the horizontal contours in panel (c) show that the large spread in lead time near B is due almost completely to variation in IHF. In other words, the system can be highly committed to B with a low zonal wind, but if IHF is low, it may take a long time to get there. We can also see this from the lower-left region of Fig. 5.3(a) and (b), where committor is high and lead time is high.

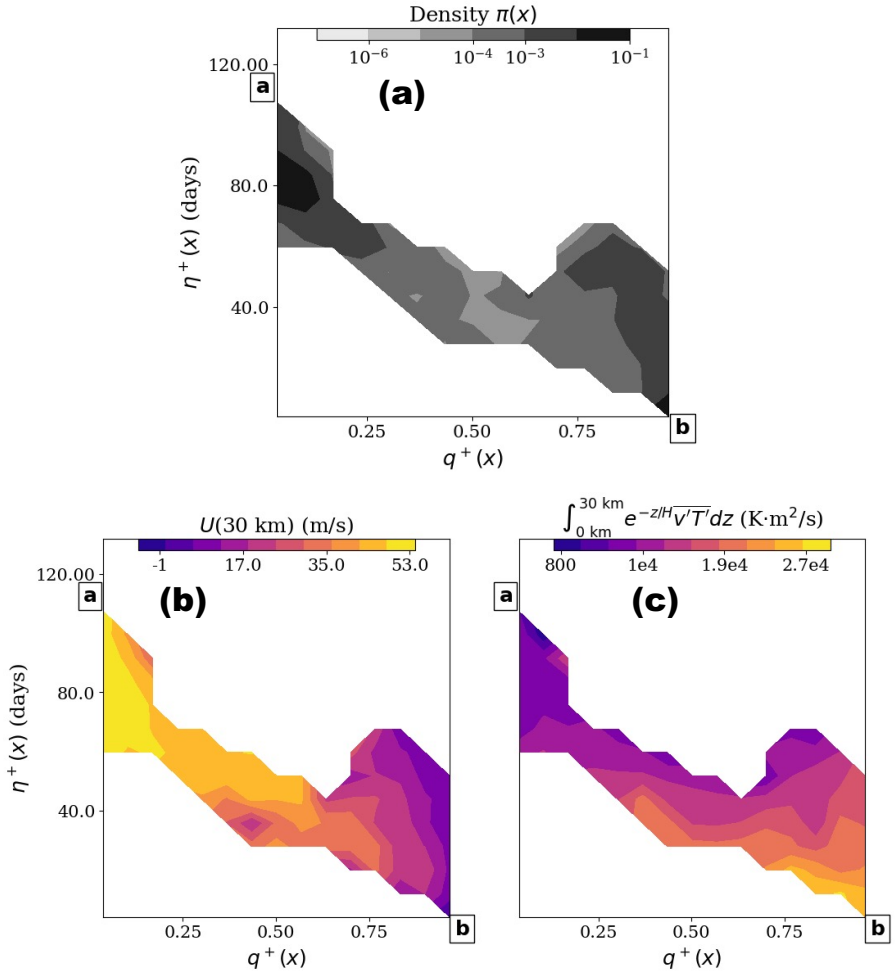


Figure 5.4: **Committor and lead time as independent coordinates.** This figure inverts the functions in Fig. 5.3, considering the zonal wind and integrated heat flux as functions of committor and lead time. The two-dimensional space they span is the essential goal of forecasting. Panel (a) shows the steady state distribution on this subspace, which is peaked near **a** and **b** (darker shading), weaker in the "bridge" region between them, and completely negligible the white regions unexplored by data. Panels (b) and (c) display zonal wind and heat flux in color as functions of the committor and lead time.

There are two complementary explanations for this phenomenon. First, the low-U, low-IHF region of state space corresponds to a temporary restoration phase in a vacillation cycle, which delays the inevitable collapse of zonal wind below the threshold defining B. In fact, the ensemble of pathways starting from θ_0 in Fig. 5.3(c) has several members whose zonal wind either stagnates at medium strength, or dips low and partially restores before finally plunging all the way down. The second explanation is that many of these partial restoration events are not part of an $A \to B$ transition, but rather a $B \to B$ transition. In a highly irreversible system such as the Holton-Mass model, these two situations are quite dynamically distinct. To distinguish them using DGA, we would have to account for the *past* as well as the future, calculating backward-in-time forecasts such as the backward committor $q^-(\mathbf{x}) = \mathbb{P}_{\mathbf{x}}\{\mathbf{X}(\tau^-) \in A\}$, where $\tau^- < 0$ is the most-recent hitting time. Backward forecasts will be analyzed thoroughly in a forthcoming paper, but they are beyond the scope of the present one.

In summary, q^+ and η^+ are principled metrics to inform preparation for extreme weather. For example, a threatened community might decide in advance to start taking action when an event is very likely, $q^+ \geq 0.8$, and somewhat imminent, $\eta^+ \leq 10$ days, or rather, when an event is somewhat likely, $q^+ \geq 0.5$, and very imminent, $\eta^+ \leq 3$ days. Because of partial restoration events, the committor does not determine the lead time or vice versa, and so a good real-time disaster response strategy should take both of them into account, defining an "alarm threshold" that is not a single number, but some function of both the committor and lead time. This idea is similar in spirit to that of the Torino scale, which assigns a single risk metric to an asteroid or comet impacts based on both probability and severity [Binzel, 2000]. Of course, after many near-SSW events, a lot of material damage may have already occurred, which may be a reason to define a higher threshold for the definition of B, or even a continuum for different severity levels of SSW. 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.

5.4 Sparse representation of the committor

The committor projections showed give only an impression of its high-dimensional structure. While Eq. (5.11) says how to optimally represent the committor over a given CV subspace, optimizing $S[f;\theta]$ over f, it does not say which subspace θ is optimal. If the committor does admit a sparse representation, we could specifically target observations on these high-impact signals. In this section we address this much harder problem of optimizing $S[f;\theta]$ over subspaces θ.

The set of CV spaces is infinite, as observables θ can be arbitrarily complex nonlinear functions of the basic state variables $\mathbf{x}$. Machine learning algorithms such as artificial neural networks are designed exactly for that purpose: to represent functions nonparametrically from observed input-output pairs. However, to keep the representation interpretable, we will restrict ourselves to physics-informed input features based on 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

$$\begin{aligned} &\partial_t\left(\frac{\overline{q'^2}}{2}\right) + (\partial_y\overline{q})\rho_s^{-1}\nabla\cdot\mathbf{F} \\ &\qquad = -\frac{f_0^2}{N^2}\rho_s^{-1}\overline{q'\partial_z(\alpha\rho_s\partial_z\psi')} \qquad (5.13) \\ &\text{where } \mathbf{F} = (-\rho_s\overline{u'v'})\mathbf{j} + (\rho_s\overline{v'\partial_z\psi'})\mathbf{k} \end{aligned}$$

The EP flux divergence has two alternative expressions: $\rho_s^{-1}\nabla\cdot\mathbf{F} = \overline{v'q'} = \rho_s^{-1}\frac{R}{Hf_0}\partial_z\left[\rho_s\overline{v'T'}\right]$. 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 $\mathscr{A} = \rho_s\overline{q'^2}/(2\partial_y\overline{q})$. Neither of these is exact in the stochastic Holton-Mass model, so we use the quantities in Eq. (5.13) 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 for the states **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 $\overline{v'T'}$ (negative $\overline{v'q'}$) as a Rossby wave propagates from the tropopause vertically up through the stratosphere and breaks. 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. For visualization, we have found $U(30km)$ and IHF(30 km) $= \int_{0\,\mathrm{km}}^{30\,\mathrm{km}} e^{-z/H}\overline{v'T'}\,dz$ to be particularly helpful. However, this doesn't necessarily imply they are optimal predictors of q^+, and regression is a more principled way to find them.

We start by projecting the committor onto each observable at each altitude separately, in hopes of finding particularly salient altitude levels that clarify the role of vertical interactions. The first five rows of Fig. 5.5 display, for five fields (U, $|\Psi|$, $\overline{q'^2}$, $\partial_y \overline{q}$, and $\overline{v'q'}$) 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.

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 $15-25$ 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 enstrophy and potential vorticity flux are also

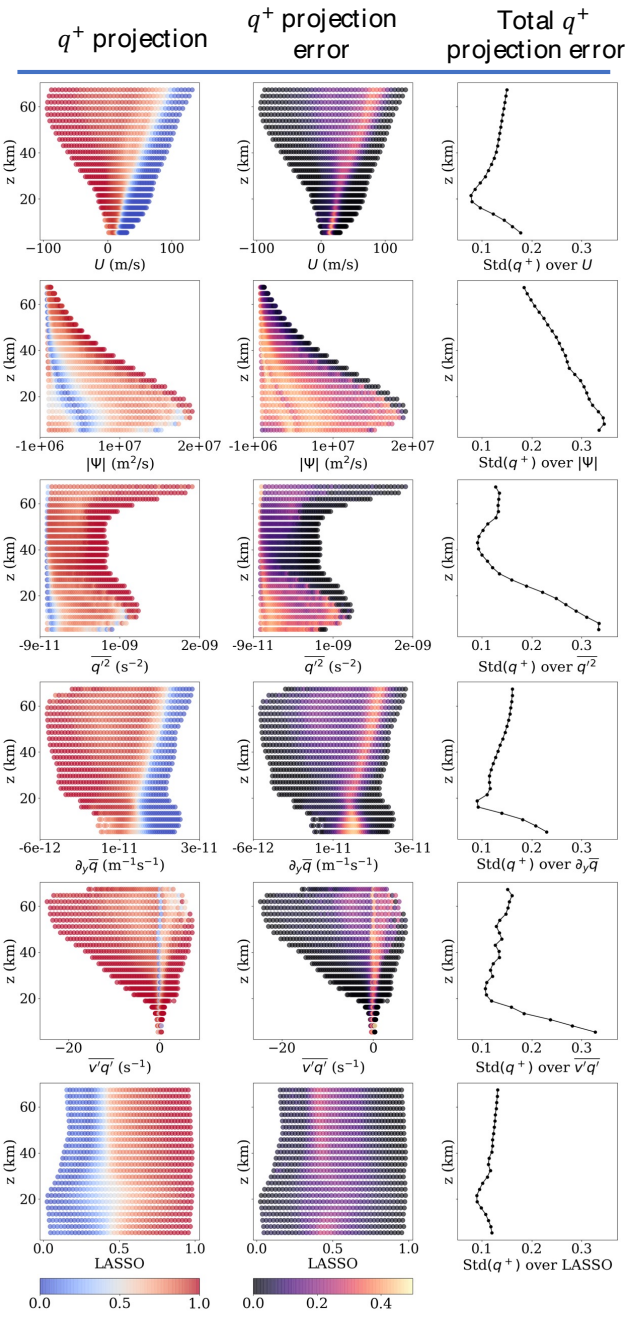


Figure 5.5: **Projection of the forward committor onto a large collection of altitude-dependent physical variables**. The top left panel shows heatmaps of q^+ as a function of U and z; white regions denote where $U(z)$ is negligibly observed. The top middle panel shows the standard deviation in q^+ as a function of U and z; this uncertainty stems from the remaining 74 model dimensions. The right-hand panel displays the total mean-squared error due to the projection for each altitude, i.e., $\sqrt{S[f;\theta]}$ from Eq. (5.10). A low value indicates that this level is ideal for prediction. The following rows show the same quantities for other physical variables: streamfunction magnitude, eddy enstrophy, background PV gradient, eddy PV flux, and LASSO.

rather unhelpful as early warning signs, despite their central role in SSW evolution. 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 $\overline{v'q'}$ with the committor is not smooth. The $q^+ < 0.5$ region at each altitude is a thin band near zero.

The exhaustive CV search in Fig. 5.5 is visually compelling in favor of some fields and some altitudes over others, but it is not satisfactory as a rigorous comparison. Differences between units and ranges make it difficult to objectively compare the L^2 projection error. Furthermore, restricting to one variable at a time is limiting. Accordingly, we also perform a more automated approach to identify salient variables in the form of a generalized linear model for the forward committor, using sparsity-promoting LASSO regression ("Least Absolute Shrinkage and Selection Operator") due to Tibshirani [1996], as implemented in the `scikit-learn` Python package [Pedregosa et al., 2011]. As input features, we use all state variables $\mathrm{Re}\{\Psi\}, \mathrm{Im}\{\Psi\}, U$, the integrated heat flux $\int_0^z e^{-z/H}\overline{v'T'}\,dz$, the eddy PV flux $\overline{v'q'}$, and the background PV gradient $\partial_y \overline{q}$, at all altitudes z simultaneously. The advantage of a sparsity-promoting regression is that it isolates a small number of observables that can accurately 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 $\pi(\mathbf{x})q^+(\mathbf{x})(1-q^+(\mathbf{x}))$ to further emphasize the transition region $q^+ \approx 0.5$. To constrain committor predictions to the range $(0,1)$, we regress on the committor after an inverse-sigmoid transformation, $\ln(q^+/(1-q^+))$. First we do this at each altitude separately, and in Fig. 5.6 (a) we plot the coefficients of each component as a function of altitude. The bottom row of Fig. 5.5 also displays the committor projected on the height-dependent LASSO predictor.

The height-dependent regression in 5.6(a) shows each component is salient for some altitude range. In general, U and $\mathrm{Im}\{\Psi\}$ dominate as causal variables at low altitudes, while $\mathrm{Re}\{\Psi\}$ dominates at high altitudes. The overall prediction quality, as measured by R^2 and plotted in Fig. 5.6 (b), is greatest around 21.5 km, consistent with our qualitative observations of Fig. 5.5. Note that not all

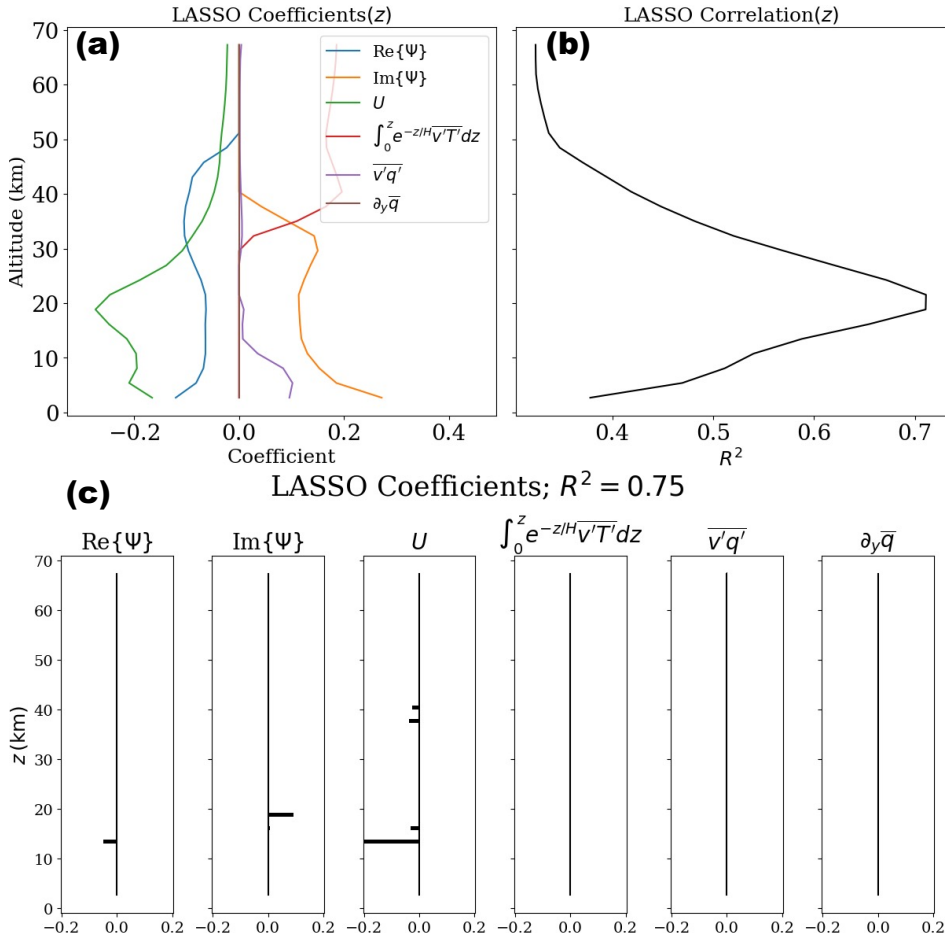


Figure 5.6: **Results of LASSO regression of the forward committor with linear and nonlinear input features**. Panel (a) shows the coefficients when q^+ is regressed as a function of only the variables at a given altitude, and panel (b) shows the corresponding correlation score. 21.5 km seems the most predictive (where $z \equiv 0$ at the tropopause, not the surface). Panel (c) shows the coefficient structure when all altitudes are considered simultaneously. Most of the nonzero coefficients appear between 15-22 km, distinguishing that range as highly relevant for prediction.

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, as seen in the bottom row of Fig. 5.5. 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 5.6 (c), with a few variables contributing the most, namely the state variables Ψ and U in the altitude range 15-22 km. The nonlinear CVs failed to make any nonzero contribution to LASSO, and this remained stubbornly true for other nonlinear combinations not shown, such as $\overline{v'T'}$. 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 here. 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.

We have presented the committor and lead time as "ideal" forecasts, especially the committor, which we have devoted considerable effort to approximating in this section. We want to emphasize that q^+ and η^+ are not competitors to ensemble forecasting; rather, they are two of its most important end results. So far, we have simply advocated including q^+ and η^+ as quantities of interest. Going forward, however, we do propose an alternative to ensemble forecasting aimed specifically at the committor, lead time, and a wider class of forecasting functions, as they are important enough in their own right to warrant dedicated computation methods. Our approach uses only short simulations, making it highly parallelizable, and shifts the numerical burden from online to offline. Figs. 5.2-5.6 were all generated using the short-simulation algorithm. While the method is not yet optimized and in some cases not competitive with ensemble forecasting, we anticipate such methods will be increasingly favorable with modern trends in computing.

5.5 The computational method

In this section we describe the methodology, which involves some technical results from stochastic processes and measure theory. Chapter 2 contains a more complete account, but here we include only the details essential to the forecasts computed in this chapter. After describing the theoretical motivation and the numerical pipeline in turn, we demonstrate the method's accuracy and discuss its efficiency compared to straightforward ensemble forecasting.

5.5.1 Feynman-Kac formulae

The forecast functions described above—committors and passage times—can all be derived from general conditional expectations of the form

$$F(\mathbf{x};\lambda) = \mathbb{E}_{\mathbf{x}}\left[G(\mathbf{X}(\tau))\exp\left(\lambda\int_0^\tau \Gamma(\mathbf{X}(s))\,ds\right)\right] \tag{5.14}$$

where again the subscript $\mathbf{x}$ denotes conditioning on $\mathbf{X}(0)=\mathbf{x}$; G,Γ are arbitrary known functions over $\mathbb{R}^d$; and τ is a stopping time, specifically a first-exit time like Eq. (5.6) but possibly with D replaced by another set. λ is a variable parameter that turns F into a moment-generating function. To see that the forward committor takes on this form, set $G(\mathbf{x})=\mathbb{1}_B(\mathbf{x})$, $\lambda=0$ (Γ can be anything), and $\tau=\tau_{A\cup B}$. 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 η^+, set $\tau=\tau_B$, $G=\mathbb{1}_B$, and $\Gamma=1$. Then

$$F(\mathbf{x};\lambda) = \mathbb{E}_{\mathbf{x}}\big[\mathbb{1}_B(\mathbf{X}(\tau))\exp(\lambda\tau)\big] \tag{5.15}$$

$$\frac{1}{q^+(\mathbf{x})}\frac{\partial}{\partial\lambda}F(\mathbf{x};0) = \frac{\mathbb{E}_{\mathbf{x}}[\tau\mathbb{1}_B(\mathbf{X}(\tau))]}{\mathbb{E}_{\mathbf{x}}[\mathbb{1}_B(\mathbf{X}(\tau))]} \tag{5.16}$$

$$= \eta^+(\mathbf{x}). \tag{5.17}$$

So we must also be able to differentiate F with respect to λ.

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 Γ 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 poleward heat flux by setting $\Gamma = \overline{v'T'}$, or the momentum lost by the vortex by setting $\Gamma(\mathbf{x}) = U(\mathbf{a}) - U(\mathbf{x})$. Extending (5.16), one can compute not only means but higher moments of such integrals by expressing the risk with Γ. Repeated differentiation of $F(\mathbf{x};\lambda)$ gives

$$\partial_\lambda^k F(\mathbf{x};0) = \mathbb{E}_\mathbf{x}\left[G(\mathbf{X}(\tau))\left(\int_0^\tau \Gamma(\mathbf{X}(s))\,ds\right)^k\right] \tag{5.18}$$

We choose to focus on expectations of the form (5.14) in order to take advantage of the Feynman-Kac formula, which represents $F(\mathbf{x};\lambda)$ as the solution to a PDE boundary value problem over state space. As PDEs involve local operators, this form is more amenable to solution with short trajectories which don't stray far from their source. The boundary value problem associated with (5.14) is

$$\begin{cases} (\mathscr{L} + \lambda\Gamma)F(\mathbf{x};\lambda) = 0 & \mathbf{x} \in D \\ F(\mathbf{x};\lambda) = G(\mathbf{x}) & \mathbf{x} \in D^\mathrm{c} \end{cases} \tag{5.19}$$

The domain D here is some combination of A^c and B^c. The operator $\mathscr{L}$ is known as the *infinitesimal generator* of the stochastic process, which acts on functions by pushing expectations forward in time along trajectories:

$$\mathscr{L}f(\mathbf{x}) := \lim_{\Delta t \to 0} \frac{\mathbb{E}_\mathbf{x}[f(\mathbf{X}(\Delta t))] - f(\mathbf{x})}{\Delta t} \tag{5.20}$$

In a diffusion process like the stochastic Holton-Mass model, $\mathscr{L}$ is an advection-diffusion partial differential operator which is analogous to a material derivative in fluid mechanics. The generator

encapsulates the properties of the stochastic process. In addition to solving boundary value problems (5.14), its adjoint $\mathscr{L}^*$ provides the Fokker-Planck equation for the stationary density $\pi(\mathbf{x})$:

$$\mathscr{L}^*\pi(\mathbf{x}) = 0 \tag{5.21}$$

We can also write equations for moments of F, as in (5.18), by differentiating (5.19) repeatedly and setting $\lambda = 0$:

$$\mathscr{L}\left[\partial_\lambda^k F\right](\mathbf{x};0) = -k\Gamma\partial_\lambda^{k-1}F \tag{5.22}$$

This is an application of the Kac Moment Method [Fitzsimmons and Pitman, 1999]. Note that we never actually have to solve (5.19) with nonzero λ. Instead we implement the recursion above. Note that the base case, $k = 0$, with $G = \mathbb{1}_B$ gives $F^+ = q^+$, no matter what the risk function Γ. In this chapter we compute only up to the first moment, $k = 1$. Further background regarding stochastic processes and Feynman-Kac formulae can be found in Karatzas and Shreve [1998], Oksendal [2003], E et al. [2019].

5.5.2 Dynamical Galerkin Approximation

To solve the boundary value problem (5.19) with $\lambda = 0$, we start by following the standard finite element recipe, converting to a variational form and projecting onto a finite basis. First, we homogenize boundary conditions by writing $F(\mathbf{x}) = \hat{F}(\mathbf{x}) + f(\mathbf{x})$, where $\hat{F}$ is a guess function that obeys the boundary condition $\hat{F}|_{D^c} = G$, and $f|_{D^c} = 0$. Next, we integrate the equation against any test function ϕ, weighting the integrand by a density μ (which is arbitrary for now, but will be specified

later):

$$\int_{\mathbb{R}^d} \phi(\mathbf{x}) \mathscr{L} f(\mathbf{x}) \mu(\mathbf{x})\, d\mathbf{x} = \int \phi(\mathbf{x}) (G - \mathscr{L}\hat{F})(\mathbf{x}) \mu(\mathbf{x}) d\mathbf{x}$$

$$\langle \phi, \mathscr{L} f \rangle_\mu = \langle \phi, G - \mathscr{L}\hat{F} \rangle_\mu \tag{5.23}$$

The test function ϕ should live in the same space as f, i.e., with homogeneous boundary conditions $\phi(\mathbf{x}) = 0$ for $\mathbf{x} \in A \cup B$. We refer to the inner products in (5.23) as being "with respect to" the measure (with density) μ. We approximate f by expanding in a finite basis $f(\mathbf{x}) = \sum_{j=1}^{M} \xi_j \phi_j(\mathbf{x})$ with unknown coefficients ξ_j, and enforce that (5.23) hold for each ϕ_i. This reduces the problem to a system of linear equations,

$$\sum_{j=1}^{M} \langle \phi_i, \mathscr{L}\phi_j \rangle_\mu \xi_j = \langle \phi_i, G - \mathscr{L}\hat{F} \rangle_\mu \qquad\qquad i = 1, \ldots, M \tag{5.24}$$

which can be solved with standard numerical linear algebra packages.

This procedure consists of three crucial subroutines. First, we must construct a set of basis functions ϕ_j. Second, we have to evaluate the generator's action on them, $\mathscr{L}\phi_j$. Third, we have to compute inner products. With standard PDE methods, the basis size would grow exponentially with dimension, quickly rendering the first and third steps intractable. 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. DGA is one such method, whose special twist is to construct a "data-informed" basis of reasonable size, evaluate the generator by implementing Eq. (5.20) with the same data set, and finally evaluate the inner products (5.23) with a Monte Carlo integral. The data consist of short trajectories launched from all over state space, which the system of linear equations stitches together into a global function estimate. We sketch the procedure here, but for the implementation details we refer to chapter 2 and to Thiede et al. [2019] and Strahan et al. [2021], where DGA has already been developed for molecular dynamics.

Step 1: Generate the data, in the format of N initial conditions $\{\mathbf{X}_n : 1 \leq n \leq N\}$. Evolve each initial condition forward for a "lag time" Δt to obtain a set of short trajectories $\{\mathbf{X}_n(t) : 0 \leq t \leq \Delta t, n = 1, \ldots, N\} \subset \mathbb{R}^d$. (Lag time is an algorithmic parameter for DGA. It is not to be confused with the forecast time horizon between the prediction and the event of interest in meteorology.) Here and going forward, $\mathbf{X}_n$ will mean $\mathbf{X}_n(0)$. The choice of starting points is flexible, but crucial for the efficiency and accuracy of DGA. Because our goal here is to demonstrate interpretable results, we prioritize simplicity and accuracy over efficiency, and defer optimization to later work. We simply draw initial conditions at random from the long control simulation of 5×10^5 days, and then generate new short trajectories from those points. We do not sample the points with equal probability, but instead re-weight to get a uniform distribution over the space $(U(30\,\text{km}), |\Psi|(30\,\text{km}))$, within the bounds realized by the control simulation, which are approximately $-30\,\text{m/s} \leq U(30\,\text{km}) \leq 70$ m/s and $0\,\text{m}^2/\text{s} \leq |\Psi|(30\,\text{km}) \leq 2 \times 10^7\,\text{m}^2/\text{s}$. This sampling procedure, and any other version, implicitly defines a *sampling measure* μ on state space, where $\mu(\mathbf{x})\,d\mathbf{x}$ is the expected fraction of starting points in the neighborhood $d\mathbf{x}$ about $\mathbf{x}$. Sampling points with equal weight from the control run would induce $\mu = \pi$, a very inefficient choice because probability concentrates around the metastable states **a** and **b**. The re-weighting procedure ensures data coverage of intermediate-wind regions between A and B, as well as the large bursts of wave amplitude that characterize the transition pathways. Our main results use $N = 5 \times 10^5$ short trajectories with a lag time of $\Delta t = 20$ days, sampled at a frequency of twice per day. This data set is more than needed to get a reasonable committor estimate, but we have sampled generously in order to visualize the functions in high detail. The final section will show the method is robust, capable of reasonably approximating the committor even with an order-of-magnitude reduction in data.

Step 2: Define the basis. The Galerkin method works for any class of basis functions that becomes increasingly expressive as the library grows and becomes capable of estimating any function of interest. However, with a finite truncation, choosing 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 $\{S_1, \ldots, S_M\}$ is a disjoint partition of state space.

This partition should be chosen with a number of considerations in mind. The 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 the initial points $\{\mathbf{X}_n\}_{n=1}^N$. 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 of five points, until we have the desired number of clusters (M). After clustering on the initial points $\{\mathbf{X}_n\}$, 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. For boundary value problems with a domain D and boundary D^c, we need only cluster points in D, since the basis should be homogeneous. The total number of clusters should scale with data set. In our main results with $N = 5 \times 10^5$, we found $M = 1500$ to be enough basis functions to resolve some of the finer details in the structure of the forecast functions, but not so many as to require an unmanageably deep address tree, which manifests in dramatic slowdown past a certain threshold. At this point, the cluster number is still a manually tuned hyperparameter.

Because the committor and lead time obey Dirichlet boundary conditions on $A \cup B$, the basis functions used to construct them should be zero on $A \cup B$, meaning only data points $\mathbf{X}_n \notin A \cup B$ should be used to produce the clusters. On the other hand, the steady state distribution has no boundary condition to satisfy, only a global normalization condition. Hence, the basis for the change of measure w must be different from the basis for q^+ and η^+, with its clusters including

all data points in $A \cup B$. Furthermore, the basis must be chosen so that the matrix $\langle (\mathscr{T}^{\Delta t} - 1)\phi_i, \phi_j \rangle$ has a nontrivial null space; this is guaranteed by the indicator basis set we use, but can otherwise be guaranteed by including a constant function in the basis.

The use of an indicator basis follows the Markov State Modeling (MSM) literature [Chodera et al., 2006, Noé and Fischer, 2008, Pande et al., 2010, Bowman et al., 2013, Chodera and Noé, 2014, e.g.,]. MSMs are a dimensionality reduction technique that has also been used in conjuction with analysis of metastable transitions, primarily in protein folding dynamics [Jayachandran et al., 2006, Noé et al., 2009]. MSMs have also been used recently to study garbage patch dynamics in the ocean [Miron et al., 2021] as well as complex social dynamics [Helfmann et al., 2021]. In Maiocchi et al. [2022], the authors take an interesting approach to MSMs by clustering points based on proximity to unstable periodic orbits, a potentially useful paradigm for general chaotic weather phenomena [Lucarini and Gritsun, 2020]. DGA can be viewed as an extension of MSMs, though, rather than producing any reduced complexity model, the explicit goal in DGA is estimating specific functions as in Eq. (5.14).

MSMs have the advantage of simplicity and robustness. In particular, the discretization of $\mathscr{T}^{\theta} - 1$ is a properly normalized stochastic matrix (with nonnegative entries and rows summing to 1), which guarantees the maximum principle $0 \leq q^{+}(\mathbf{x}) \leq 1$ and $0 \leq w(\mathbf{x})$ for all data points $\mathbf{x}$. However, alternative basis sets have been shown to be promising, perhaps with much less data. Thiede et al. [2019] used diffusion maps, while Strahan et al. [2021] used a PCA-like procedure to construct the basis. More generally, there is no requirement to use a linear Galerkin method to solve the Feynman-Kac formulae. More flexible functional forms may have an important role to play as well. In the low-data regime, some preliminary experiments have suggested that Gaussian process regression (GPR) is a useful way to constrain the committor estimate with a prior, following the framework in Bilionis [2016] to solve PDEs with Gaussian processes. As mentioned in the conclusion, there is rapidly growing interest in the use of artificial neural networks to solve PDEs. As with many novel methods, however, DGA is likely to work best on new applications when its

simplest form is applied first. This will be our approach in coming experiments on more complex models.

Step 3: Apply the generator. The forward difference formula

$$\widehat{\mathscr{L}\phi}(\mathbf{X}_n) = \frac{\phi(\mathbf{X}_n(\Delta t)) - \phi(\mathbf{X}_n)}{\Delta t} \tag{5.25}$$

suggested by the definition of the generator (5.20), results in a systematic bias when Δt is finite. On the other hand, small values of Δt lead to large variances in our Monte Carlo estimates of the inner products in (5.24). To resolve these issues we use an integrated form of the Feynman–Kac equations that involves stopping trajectories when they enter A or B. Details are provided in Appendix A.

Step 4: Compute the inner products. The inner products in Eq. (5.24) are integrals over high-dimensional state space that are intractable with standard quadrature, 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 $\{\mathbf{X}_1, \ldots, \mathbf{X}_N\}$ (which we do), the law of large numbers gives, for any function g with finite expectation,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} g(\mathbf{X}_n) = \int_{\mathbb{R}^d} g(\mathbf{x})\mu(\mathbf{x})d\mathbf{x} \tag{5.26}$$

Setting $g(\mathbf{x}) = \phi_i(\mathbf{x})\mathscr{L}\phi_j(\mathbf{x})$, the sample average on the left-hand side of (5.26) therefore provides an estimator of $\langle \phi_i, \mathscr{L}\phi_j \rangle_\mu$. Of course, our approximation uses finite N and nonzero Δt. A similar sample average approximation can be used to estimate the inner product on the right-hand side of (5.24).

These same steps apply to both q^+ and $\mathbb{E}[\tau_B]$, as well as the recursion in (5.22) for η^+. For the Fokker-Planck equation (5.21), one extra step is needed to convert an equation with $\mathscr{L}^*$ into an equation with $\mathscr{L}$. Our procedure for estimating π is described in chapter 2.

Step 5: Solve the equation (5.24). With a reasonable basis size $M \lesssim 1000$, an LU solver such as in LAPACK via Numpy can handle Eq. (5.24). In the case of the homogeneous system for $w(\mathbf{x})$, a QR decomposition can identify the null vector.

5.5.3 DGA fidelity and sensitivity analysis

To illustrate the effect of parameter choices on performance, we present here a simple sensitivity analysis. Fig. 5.7 verifies the numerical accuracy and convergence of DGA by plotting the committor as a function of $U(30\text{ km})$, estimated both with DNS and DGA, for various DGA parameters. The red curves $q^+_{\text{DGA}}(U(30\text{km}))$ are calculated by projecting the committor as in Fig. 5.2(a), while the black curve $q^+_{\text{DNS}}(U(30\text{km}))$ is an empirical committor estimate equal to the fraction of control simulation points seen at a particular value of $U(30\text{km})$ that next hit B.

In panels (a), (b), and (d), the lag time Δt increases from 5 to 10 to 20 days while the number of short trajectories stays fixed at $N = 5 \times 10^5$. Panel (c) has a long lag of 20 days, but a small data set of $N = 5 \times 10^4$, allowing us to see the tradeoff between N and Δt. The basis size M is chosen heuristically as large as possible within reason for the clustering algorithm (see Appendix A). While DGA tends to systematically overestimate q^+ relative to q^+_{DNS} in the mid-range of U, it seems to approach the empirical estimate as the data size and lag time increase. Each plot also displays the root-mean-square deviation between the two estimators over this subspace, $\varepsilon = \sqrt{\langle (q^+_{\text{DGA}} - q^+_{\text{DNS}})^2 \rangle_\pi}$. Within this regime, it seems that increasing the lag time has a greater impact on the deviation than increasing the number of data points. Panels (b) and (c) have approximately the same deviation ε, but (c) uses only one fifth the data, measured by total simulation time. On the other hand, more short trajectories can be parallelized more readily than fewer long trajectories, and the optimal choice will depend on computing resources.

It is natural to ask whether our short trajectory based approach is more efficient than DNS in which many independent "long" trajectories are launched from a single initial condition $\mathbf{x}$ and the committor probability $q^+(\mathbf{x})$ (or another forecast) is estimated directly. For a single value

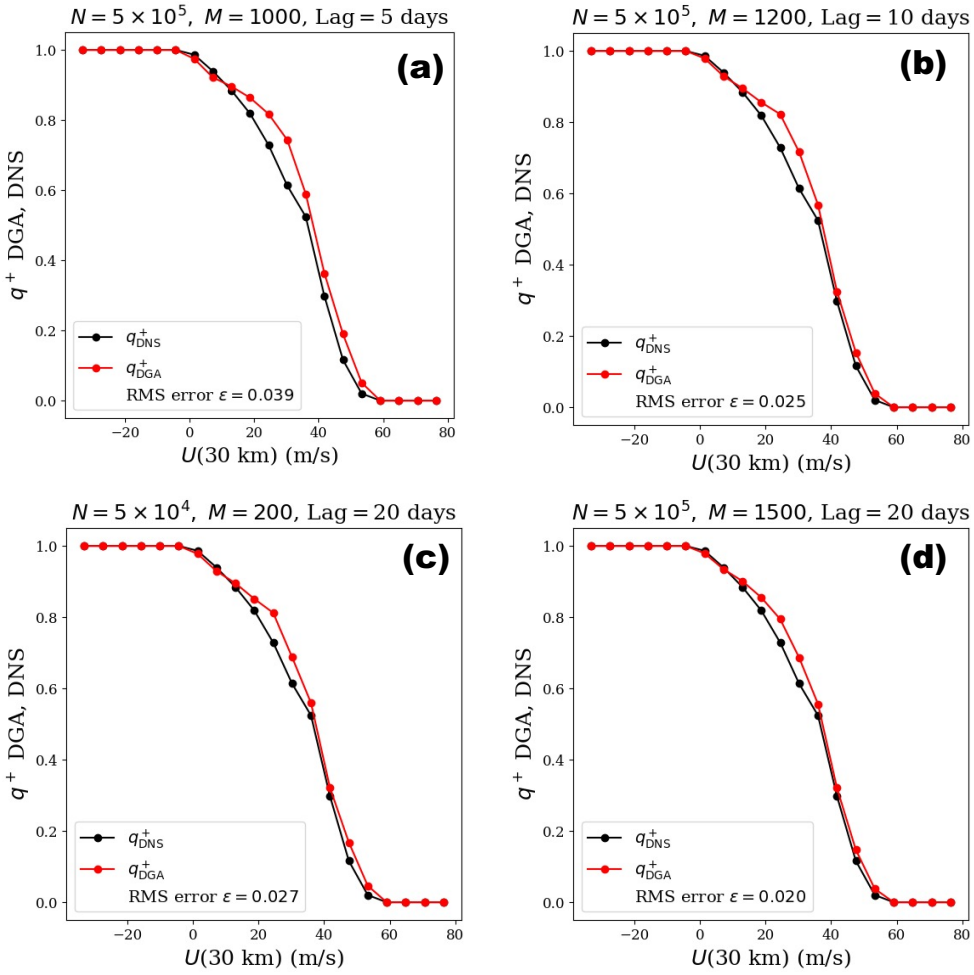


Figure 5.7: **Fidelity of DGA.** For several DGA parameter values of N (the number of data points), M (the number of basis functions) and lag time, we plot the committor calculated from DGA and DNS (from the long control simulation), both as a function of $U(30\text{ km})$. The mean-square difference ε in the legend is used as a global error estimate for DGA.

of $\mathbf{x}$ for which $q^+(\mathbf{x})$ is not very small (so that a non-negligible fraction of trajectories reach B before A) and for which the lead time $\eta^+(\mathbf{x})$ is not too large (so that trajectories reaching B do so without requiring long integration times), DNS will undoubtedly be more efficient. This is often the situation in real-time weather forecasting. However, a key feature of our approach is that it simultaneously estimates forecasts at all values of $\mathbf{x}$, allowing the subsequent analysis of those functions that has been the focus of much of this article. Global knowledge of the committor and lead time is more pertinent for oft-repeated forecasts, for long-term risk assessment of extreme event climatology, and for targeting observations optimally. Building accurate estimators in all of state space by DNS would be extremely costly even for the reduced complexity model studied here.

5.6 Conclusion

We have shown numerical results in the context of a stochastically forced Holton-Mass model with 75 degrees of freedom, which points to the method's promise for forecasting. 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 examined the relationship between probability and lead time for a given rare event, a powerful pairing 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 in the range of 10-20 km above the tropopause seems to be optimal among a large class of dynamically motivated observables.

The next chapter further analyzes the Holton-Mass model from a TPT perspective, relating the forecast functions to aftcast functions and path statistics such as the rate and reactive current.

6 EXPLORING SUDDEN STRATOSPHERIC WARMINGS WITH TRANSITION PATH THEORY

Transition Path Theory (TPT) is used for a comprehensive analysis of sudden stratospheric warming (SSW) events in a highly idealized wave-mean flow interaction system due to Holton and Mass [1976], augmented with stochastic forcing. TPT is a statistical mechanics framework that explicitly considers rare events as an ensemble, and provides relationships between short-term forecasting and long-term climatology. We use the probability current, a central TPT quantity, to build a picture of critical altitude-dependent interactions between waves and the mean flow that fuel SSW events, both average behavior and variability across the SSW ensemble. We find that the rapid deceleration of zonal wind tends to be preceded by a gradual, halting decay in wind strength and a steady increase in meridional heat flux, which conspire to precondition the vortex for collapse. The ensemble-level description allows us to identify the signal of an oncoming SSW emerging from background variability during preconditioning, well before the sudden collapse. To circumvent the costly approach of extensive direct simulation of the full rare event ensemble, we implement a highly parallel computational method that launches a large collection of short simulations from many initial conditions, estimating long-timescale rare event statistics from short-term tendencies.

This chapter is adapted from the preprint Finkel et al. [2021a].

6.1 Introduction

In the previous chapter, we computed key forecasting functions—the *forward committor* and *lead time*—that give the probability of SSW and its expected arrival time, as a function of initial conditions. We worked in the context of the Holton-Mass model, which we examine further in this chapter. The TPT analysis we undertake here is related to the forecasting problem, but furthermore addresses the event's mechanism all the way from start to finish, not just forward in time from a fixed initial condition. Crucially, TPT distinguishes between the *onset* of an atmospheric distur-

bance (in this case, a breakdown of the polar vortex from strong to weak) and the *persistence* of that disturbance (the "vacillation cycles" of an already weakened jet; Holton and Mass [1976]). In this paper, we use TPT to connect short-term weather forecast statistics, encoded by the committor and lead time, to the long-term climatology of SSW events, including their frequency, duration, and the distribution of pathways encoded by the *probability current*: the average tendency of the system conditioned on the occurrence of an SSW. By visualizing the probability current, we quantitatively assess the interaction between wave disturbances and zonal wind anomalies, and the extent to which they are uniquely associated with an SSW. TPT gives information about the *variability* of these processes, not just their mean behavior. In particular, we will show differences in the variability between successive stages of an SSW event. The preconditioning of the polar vortex manifests as a steady, predictable weakening of the lower-level zonal wind. The latter stage is an abrupt burst of heat flux and collapse of zonal wind that is much more variable in its timing and intensity. These are only a few deliverables of TPT, which can be adapted to probe many other weather phenomena.

As in chapter 5, we use the DGA method of aggregating together many short, parallel simulations to capture rare event statistics without ever observing a complete event. This chapter further applies the method to *backward-in-time* forecasts, which are needed to recover steady-state statistics from short-trajectories, as described in chapter 2.

This chapter is organized as follows. In section 6.2, we visualize the evolution of SSW events through the probability current, and compare to the minimum action method. Section 6.4 assesses the numerical accuracy of the DGA method on quantities of interest in the Holton-Mass model. We assess future possibilities and conclude in section 6.5.

6.2 Transition path ensemble

Every SSW event, or transition path, is a sample from a high-dimensional distribution called the *transition path ensemble*, which refers to the infinite collection of paths one would obtain by running the model forever. We will first give an account of the transition path ensemble based on

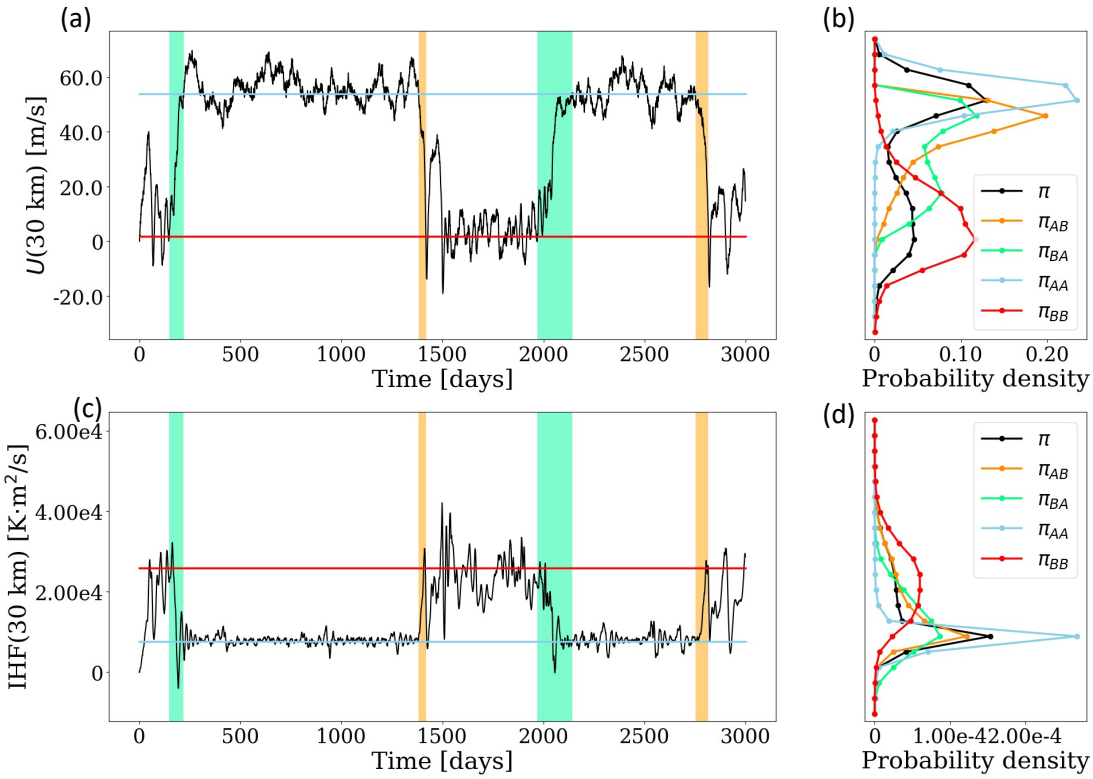


Figure 6.1: **Bistable time series**. (a) Zonal wind at 30 km over time, with $A \to B$ transitions (SSWs) highlighted in orange and $B \to A$ transitions highlighted in green. (b) Conditional probability distributions of each of the four phases. (c-d) Same as a-b but with integrated heat flux up to 30 km plotted instead of zonal wind at 30 km. Blue and red lines show the position of the two fixed points, **a** and **b**, along these two observables.

storylines of the few individual events shown Fig. 6.1. Subsequently, we will present the TPT analysis, which describes the distribution as a whole using a specific collection of functions including probability densities, committors, and currents.

6.2.1 SSW storylines

Fig. 5.1c shows a 3000-day model integration in a two-dimensional subspace consisting of zonal wind $U(30\text{ km})$ and vertically integrated eddy meridional heat flux, which is abbreviated IHF (integrated heat flux) and defined as

$$\text{IHF}(30\text{ km}) = \int_0^{30\text{ km}} e^{-z/H}\overline{v'T'}(z)\,dz \tag{6.1}$$

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

$$e^{-z/H}\overline{v'T'}(z) = e^{-z/H}\frac{Hf_0}{R}\overline{\frac{\partial\psi'}{\partial y}\frac{\partial\psi'}{\partial z}} \propto |\Psi(z)|^2\frac{\partial\varphi}{\partial z}, \tag{6.2}$$

where R is the ideal gas constant for dry air, and φ is the phase of Ψ. Hence the heat flux is related to the amplitude and phase tilt of the waves, both of which rise significantly during an SSW event. In Fig. 5.1c, the fixed point **b** has more than twice the IHF of **a**, and the $A \to B$ transitions (orange segments) begin with a simultaneous decrease in U and increase in IHF. The $B \to A$ transitions (green segments) do not retrace the same route backward, but rather linger in the vicinity of B before gaining zonal wind strength and decreasing in IHF, which even dips slightly negative in the late stages of vortex recovery.

The same two variables, U and IHF, are plotted over time in Fig. 6.1(a,c), with transition paths highlighted in the same colors. The neighborhoods A and B are clearly metastable: the system tends to linger in one of the regions for an extended period before quickly switching to the other. We can also see bistability by looking at the steady-state probability density, denoted $\pi(\mathbf{x})$, which is plotted

as black curves in Fig. 6.1(b,d). The curve is bimodal over $U(30\text{ km})$. Over IHF(30 km), $\pi(\mathbf{x})$ is sharply peaked over A but low and flat over B, reflecting persistent fluctuations, the "vacillation cycles" of Holton and Mass [1976], in the weak-vortex regime.

We can decompose the distribution more explicitly into four separate "phases" induced by the presence of sets A and B. (i) In the $A \to B$ phase, marked by orange, the vortex is breaking down, en route from A to B. (ii) In the $B \to A$ phase, marked by green, the vortex is recovering from the vacillating regime back to the radiatively driven regime. (iii) In the $A \to A$ phase, the vortex is strong and remaining strong for the time being, either inside set A or taking a brief excursion before returning back to A. (iv) In the $B \to B$ phase, the vortex is weak, caught in ongoing vacillation cycles in the vicinity of B. We denote the corresponding probability densities as π_{AB}, π_{BA}, π_{AA}, and π_{BB}, and plot them in Fig. 6.1(b,d) along with the overall density π. Concretely, π_{AB} can be obtained from DNS by running a long simulation, extracting only the $A \to B$ transition paths, and plotting a histogram of those states. The other phases can be obtained analogously, although for all of them we strictly use DGA as described in chapter 2. The two peaks in $\pi(\mathbf{x})$, over both observables U(30 km) and IHF(30 km), are seen to come from two unimodal distributions, π_{AA} and π_{BB}. In both panels (b) and (d) the peak over A is narrow and tall compared to the low, wide peak over B, indicating a higher degree of variability associated with vacillation cycles.

When the system is en route from A to B, we say it is (AB)-*reactive*, using a term from chemistry literature where the passage from A (reactant) to B (product) models a chemical reaction. Therefore we refer to π_{AB} and π_{BA} as (AB and BA)-*reactive densities*, which reveal structure hidden from view within the sparsely-populated region between A and B. Along $U(30\text{ km})$, π_{AB} is peaked near A and falls off rapidly toward B, suggesting that transition paths spend much of their time slowly crawling away from A before speeding up later on. π_{BA} has two peaks in the transition region, suggesting that the system takes a long time to escape from B, and also a long time to re-enter A. This asymmetry is not so clear over the observable IHF(30 km), in which π_{AB} and π_{BA} look quite similar, underscoring the need to examine multiple subspaces to distinguish the phases.

The two events in Figs. 5.1c and 6.1(a,b) are only samples from the full transition path ensemble. Any small sample of events cannot fully represent the whole ensemble of transition paths (for example, in the real world, SSWs have two distinct types: split and displacement). How should we describe this complicated ensemble faithfully? The distributions π_{AB} and π_{BA} tell us where transition paths tend to linger, on average, but not much about their detailed movement through state space. A standard approach is to average together multiple events to obtain a composite evolution, which can reveal important features of SSW climatology [e.g., Charlton and Polvani, 2007, Albers and Birner, 2014, Mitchell et al., 2011]. However, lining up multiple time series with different durations requires some arbitrary choices. Conventionally, the "central date" of the warming—when zonal wind first reverses—is used as a reference point, but this may obscure the initial seeds of SSW that happen at different times in advance.

The issue is illustrated in Fig. 6.2. Panel (a) shows zonal wind over time for 300 observed transition events leading up the warming. Three of these paths are colored, only in between the last-exit time from A (denoted τ_A^-) and the first-entrance time to B (denoted τ_B^+), to illustrate some of the variability between transition paths. The red curve sinks steadily downward until accelerating into an SSW, while the black curve spends a long time trapped in a partially weakened vortex state before its ultimate decline. The cyan pathway does something in between. The remaining gray trajectories include several deep dives and partial recoveries of zonal wind before ultimately descending into B. Panel (b) shows the composite evolution of these 300 trajectories: at every point in time, the black curve shows the median, while the three red envelopes show the middle 20th, 50th, and 90th percentile ranges. (We include in this average the timeseries that have not yet left set A, although the definitions to follow will exclude these early segments from the analysis). The composite evolution successfully captures the sharp nosedive in zonal wind at the end of the transition pathway, but misses the large meanders that some paths, including the black path, go through before the precipitous decline. A comprehensive account of the transition path ensemble should include the stagnations as well. In order to capture these initial stages, we have defined

SSW in such a way that the full process takes ~ 80 days, much longer than the ~ 10 days time horizon that traditionally comprises an SSW event. This model, like the true atmosphere, sees the most dramatic zonal wind collapse only in the last few days; however, we will show that most of the probabilistic progress occurs during the longer preceding "preconditioning" stage.

The TPT approach averages trajectories together in a different way, aligning them by their position in state space rather than by the time until SSW (which is itself a random variable). This new kind of composite evolution is the essence of the probability current, which highlights the sequence of events that must happen between A and B regardless of the time horizon. In the rest of this section, we define and visualize probability currents, starting with their basic ingredients: committor functions. This background material is also contained in chapter 2, but the essential ingredients are repeated here.

6.2.2 Committors, densities, and currents

Let us fix an initial condition $\mathbf{X}(t_0) = \mathbf{x}$ with a vortex that is neither strong nor fully broken down, so $\mathbf{x} \notin A \cup B$. $\mathbf{X}(t)$ will soon evolve into either A or B, since both are attractive. The probability of hitting B first is called the *forward committor* (to B):

$$q_B^+(\mathbf{x}) = \mathbb{P}_{\mathbf{x}}\{\mathbf{X}(\tau_{A\cup B}^+(t_0)) \in B\} \tag{6.3}$$

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

$$\tau_S^+(t_0) = \min\{t > t_0 : \mathbf{X}(t) \in S\}. \tag{6.4}$$

Here, S is the union of A and B, i.e., the trajectory has returned to a metastable state. The probability of hitting A first instead—the "forward committor to A"—is $q_A^+(\mathbf{x}) = 1 - q_B^+(\mathbf{x})$. Unless specified otherwise, we call q_B^+ the forward committor, as the SSW event is our main interest. Committors

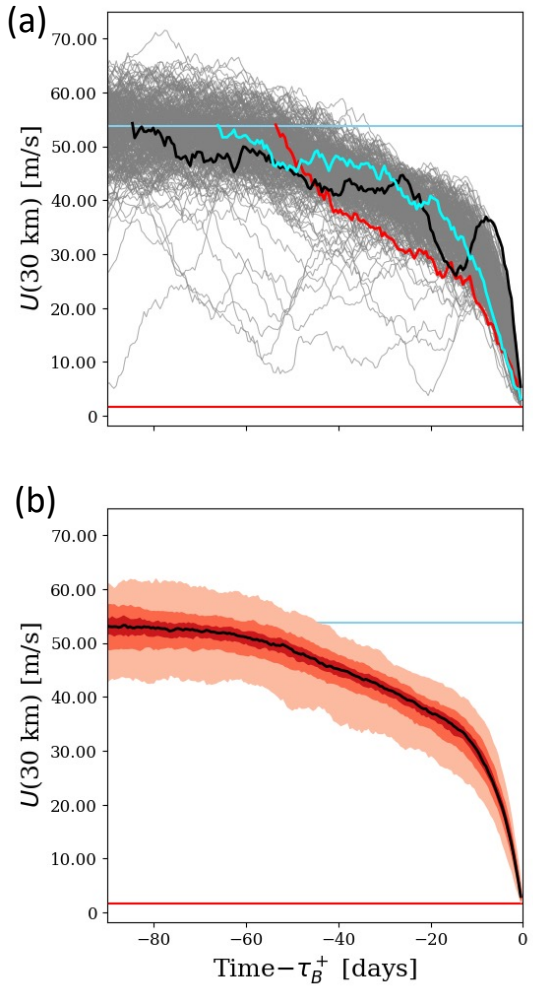


Figure 6.2: **SSW ensemble and composites.** (a) 100 SSW realizations in gray in terms of $U(30$ km), aligned by the central date of the warming when zonal wind dips below 1.75 m/s. Three of the realizations are colored in between their last-exit time from A (τ_A^-) and their next-hitting time to B (τ_B^+). (b) Composite evolution of $U(30$ km). The black curve shows the pointwise median, and the three red-orange envelopes show the middle 20, 50, and 90 percentile ranges.

are deterministic functions of state space involving ensemble averages of $\mathbf{X}(t)$, whereas hitting times are random variables depending on the realization of $\mathbf{X}(t)$. Our system is autonomous, with no external time-dependent forcing, so we can set $t_0 = 0$ and drop the argument from $\tau^+_{A\cup B}$ without loss of generality. The autonomous assumption can be relaxed, either by augmenting $\mathbf{x}$ with a periodic variable for time (e.g., to include the seasonal cycle) or by augmenting A and B to include initial and terminal times (e.g., to examine climate change effects). Periodic- and finite-time TPT has been presented in Helfmann et al. [2020], and we plan to utilize this framework in a forthcoming paper using state-of-the-art ensemble forecasts. As a conceptual demonstration, the autonomous Holton-Mass model makes for a clearer exposition.

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

$$q^-_A(\mathbf{x}) = \mathbb{P}_{\mathbf{x}}\{\mathbf{X}(\tau^-_{A\cup B}(t_0)) \in A\} \tag{6.5}$$

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

$$\tau^-_S(t_0) = \max\{t < t_0 : \mathbf{X}(t) \in S\} \tag{6.6}$$

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

The forward and backward committors are shown in Fig. 6.3(a,b). In this and later figures, the white regions of state space have insignificant probability. Note that q^+_B and q^-_A have very different contour structures, a sign of irreversible behavior (in a stochastic system with detailed balance, i.e., a reversible system, $q^-_A = 1 - q^+_B$). Both q^+_B and q^-_A are large in the upper-right flank of state space, meaning that whenever medium-strength zonal wind and large IHF are observed together,

chances are high that the system both came from A and will next hit B. In other words, an SSW is underway. Compare to the middle-left flank of state space, where q_B^+ is large but q_A^- is small: there, the system is likely headed toward B, *from* B, which does not count as an SSW event.

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

$$\pi_{AB}(\mathbf{x}) = \frac{1}{Z_{AB}} \pi(\mathbf{x}) q_A^-(\mathbf{x}) q_B^+(\mathbf{x}) \tag{6.7}$$

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

Each probability density (π, π_{AB}, π_{BB}, etc.) is associated with a *probability current* ($\mathbf{J}$, $\mathbf{J}_{AB}$, $\mathbf{J}_{BB}$, etc.). The steady-state current $\mathbf{J}(\mathbf{x})$ is a vector field that describes the probability mass flux through $\mathbf{x}$. It is related to the deterministic flow $\dot{\mathbf{X}}(t) = v(\mathbf{X}(t))$, but differs by a factor of $\pi(\mathbf{x})$ to account for density variations and a diffusion term to account for the stochastic perturbations. For a diffusion process of the form (2.9), these currents have the explicit form

$$\mathbf{J}(\mathbf{x}) = \pi v - \nabla \cdot (\mathbf{D}\pi), \tag{6.8}$$

$$\mathbf{J}_{AB}(\mathbf{x}) = q_A^- q_B^+ \mathbf{J} + \pi \mathbf{D} \left[q_A^- \nabla q_B^+ - q_B^+ \nabla q_A^- \right], \tag{6.9}$$

where the diffusion matrix $\mathbf{D}(\mathbf{x}) = \frac{1}{2}\sigma(\mathbf{x})\sigma(\mathbf{x})^\top$, and ∇ represents the gradient operator over state space. One can substitute A and B for other symbols to single out the phase of interest. Dependence on $\mathbf{x}$ has been suppressed throughout. Unlike the deterministic flow field $v(\mathbf{x})$, $\mathbf{J}(\mathbf{x})$ is divergence-free, reflecting the steady-state property that every region of state space has a constant probability mass. (See Vanden-Eijnden [2006] and Metzner et al. [2006] for a thorough mathematical treatment, or Finkel et al. [2020] for an application to a simpler SSW model.) Fig. 6.4a overlays $\mathbf{J}(\mathbf{x})$ (black arrows) atop $\pi(\mathbf{x})$ (orange logarithmic color scale). The vector field lives in $\mathbb{R}^{75}$, but we have projected it into two dimensions using a visualization procedure due to Strahan et al. [2021]

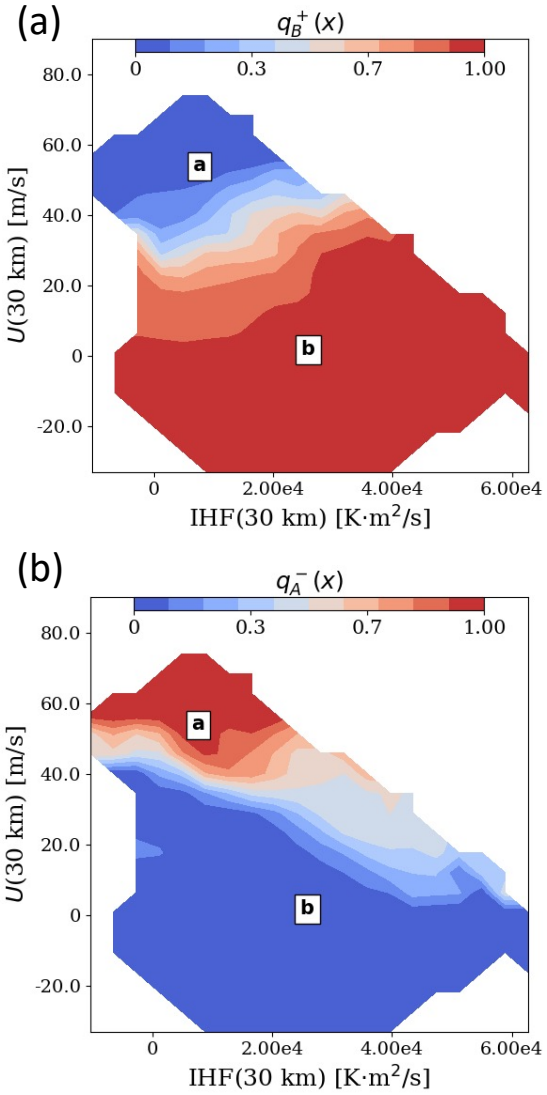


Figure 6.3: **Committors.** (a) Forward committor $q_B^+(\mathbf{x})$, the probability to hit B next starting from $\mathbf{x}$, and (b) backward committor $q_A^-(\mathbf{x})$, the probability to have come from A last given the current state $\mathbf{x}$. The committors are projected on a two-dimensional space (IHF(30 km),U(30 km)).

and described in chapter 2. The two black curves in Fig. 6.1 are the two marginals of the orange density in Fig. 6.4. The two probability peaks around A and B are seen as dark blobs, each of which is surrounded by strong probability currents and separated by a region of weaker current.

To understand this vector field, we make a fluid-dynamical analogy. If A and B are two coherent eddies in a body of water, a tracer particle spends most of its time trapped in one of the two, but is occasionally ejected from one eddy and entrained in the other. The equilibrium current is thus dominated by the velocity fields of the two eddies, but the smaller filaments that connect them are responsible for occasional transition events, which of course are our primary interest. To single out the dynamics of each phase, we decompose $\mathbf{J}(\mathbf{x})$ just as we decomposed $\pi(\mathbf{x})$, conditioning on the past and future of $\mathbf{X}(t)$ as it passes through $\mathbf{x}$. $\mathbf{J}_{AB}(\mathbf{x})$, shown in Fig. 6.4b, is the average flow of trajectories moving from A to B through $\mathbf{x}$; $\mathbf{J}_{AA}(\mathbf{x})$, shown in Fig. 6.4c, is the flow from A back to A through $\mathbf{x}$, etc. The background colors are the probability densities for the corresponding phase. For example, panel (c) shows $\pi_{AB}(\mathbf{x})$, the probability of finding a trajectory at $\mathbf{x}$ given that it is en route from A to B.

By visualizing transition pathways as static vector fields in state space, we switch from a Lagrangian to an Eulerian reference frame and fulfill our promise to "align transition paths by their position in state space." The averaging choices in Fig. 6.2 were challenging because each "particle" (ensemble member) approaches B through a different pathway. The probability currents protray the global behavior of transitions, as opposed to "case studies" provided by individual trajectories.

Let us examine the characteristics of each phase. The current $\mathbf{J}_{AA}$ is disorderly and suggests that typical fluctuations around A are usually extinguished swiftly by the restoring force of radiative equilibrium. On the other hand, $\mathbf{J}_{BB}$ is a highly organized "eddy" around $\mathbf{b}$. This reflects the vacillation cycles seen in the time series of Fig. 6.1, and offers a dynamic perspective not available from the stationary distribution $\pi_{BB}(\mathbf{x})$. Each cycle consists of a slow buildup of zonal wind driven by radiative cooling, wave enhancement allowed by the growing PV gradient, and subsequent collapse of zonal wind. Mathematically, the linearized system near $\mathbf{b}$ is stable with complex

eigenvalues; **b** is an attracting fixed point, and without noise the oscillations would die out eventually. Stochastic forcing injects enough energy to excite the system off of the fixed point, and a nearby limit cycle beyond a Hopf bifurcation directs this energy into sustained oscillations [Yoden, 1987b].

Comparing Fig. 6.4(a,b,e), we see that the steady-state current is approximately the sum of $\mathbf{J}_{AA}$ and $\mathbf{J}_{BB}$, two coherent eddies separated by a barrier at $U(30\text{ km}) \approx 35$ m/s. The occasional $A \to B$ transition breaches this barrier in a way described by $\mathbf{J}_{AB}$ in Fig. 6.4c. $\mathbf{J}_{AB}$ emerges from set A with gradually increasing IHF and decreasing zonal wind. At first $\mathbf{J}_{AB}$ matches approximately with $\mathbf{J}_{AA}$, extending out of the lower-right corner of A, but at $U(30\text{ km}) \approx 40$ m/s, $\mathbf{J}_{AB}$ separates decisively into its own unique stream. Down to $U(30\text{ km}) \approx 30$ m/s, $\mathbf{J}_{AB}$ remains strong and localized in a narrow tube going downward and rightward. Subsequently, $\mathbf{J}_{AB}$ weakens and spreads out as it turns downward for its final descent into B, indicating that pathways tend to meander more widely through this late stage of an SSW in the Holton-Mass system.

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

As a second point of comparison, we have also plotted the minimum-action pathways (both from $A \to B$ and $B \to A$) with thick cyan lines, representing the most likely transition path in the low-noise limit [e.g., Freidlin and Wentzell, 1970, E et al., 2004, Forgoston and Moore, 2018]. The pathway solves an optimization problem, deviating as minimally as possible from the deterministic dynamics while still bridging the gap all the way from A to B. We use sequential quadratic programming to approximate the minimum-action path following Plotkin et al. [2019]. We use a completely discrete approach for simplicity and to accommodate the low-rank nature of the stochastic forcing. Heuristically, the minimum-action path is the maximum-probability path con-

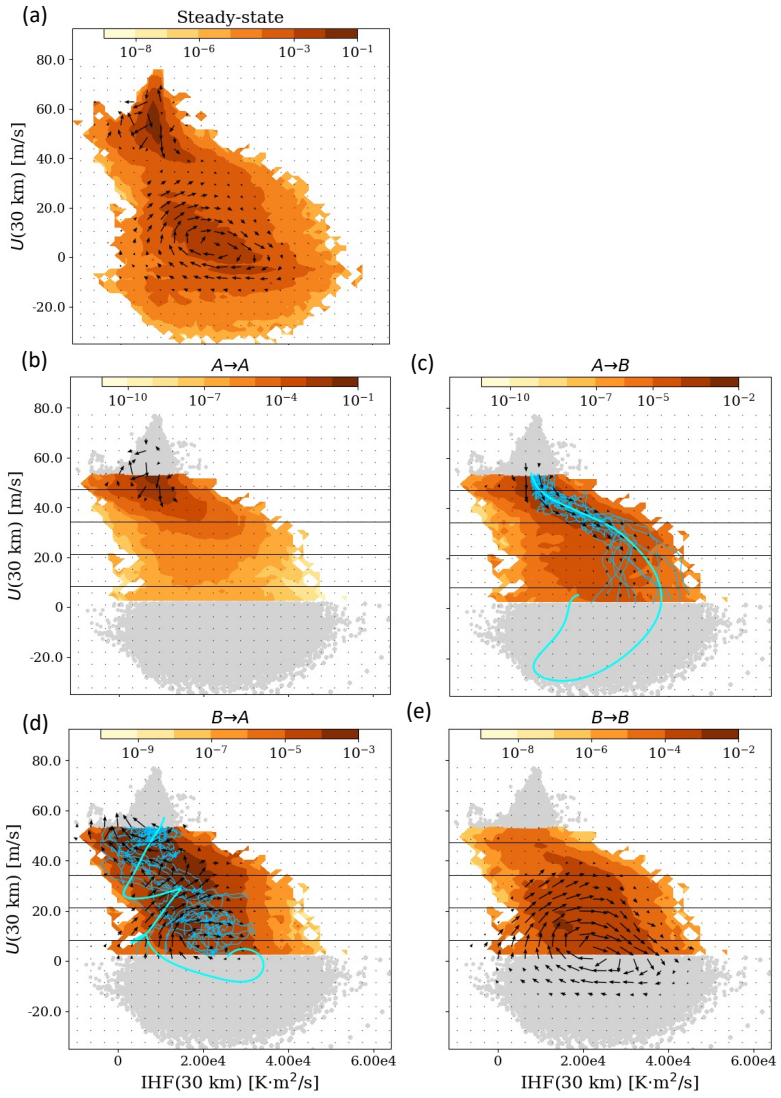


Figure 6.4: **Densities and currents.** (a) shows the equilibrium density $\pi(\mathbf{x})$ and equilibrium current $\mathbf{J}(\mathbf{x})$. (b-e) show the reactive densities and currents for $A \to A$, $A \to B$, $B \to A$, and $B \to B$ transitions, respectively. For example, (c) shows the reactive current $\mathbf{J}_{AB}(\mathbf{x})$ overlaid on the reactive $\pi_{AB}(\mathbf{x})$, illustrating the most common pathways of SSW trajectories from the strong to weak vortex state. Thick cyan curves in (c) and (d) mark the minimum-action pathways from $A \to B$ and $B \to A$, respectively, while thin blue curves show a few sampled realized transition pathways. Gray dots are data points inside states A and B.

necting A and B, which we take as the mode of the (discretized) path density over the distribution of paths from A to B. For concreteness, fix $\mathbf{x}(0) = \mathbf{x}_0 \in A$ and a time horizon T discretized into K intervals, with a timestep $\delta t = T/K = 0.005$ days. The discretized dynamics evolve according to the Euler-Maruyama method as

$$\mathbf{x}(k\delta t) = \mathbf{x}((k-1)\delta t) + v\big(\mathbf{x}((k-1)\Delta t)\big)\Delta t + \sigma\eta_k\sqrt{\delta t} \tag{6.10}$$

where η_k is a vector of i.i.d. unit normal samples. We perform optimization in the space of perturbations rather than paths, which results in a simple convex objective function at the expense of a more complicated constraint. The probability density of a particular forcing sequence $(\eta_1, \ldots, \eta_K)$ is given by

$$\prod_{k=1}^{K} \frac{1}{(2\pi)^{m/2}} \exp\left(-\frac{1}{2}\eta_k^\top \eta_k\right) = \frac{1}{(2\pi)^{mK/2}} \exp\left(-\frac{1}{2}\sum_{k=1}^{K} \eta_k^\top \eta_k\right) \tag{6.11}$$

The objective inside the exponential is now a simple quadratic in perturbation space which can be easily differentiated with respect to those perturbations. The constraint, meanwhile, takes the form of a complicated iterated function. Define the flow map $F(\mathbf{x}) = \mathbf{x} + v(\mathbf{x})\delta t$ as the deterministic part of the timestep, so $\mathbf{x}_k = F(\mathbf{x}_{k-1}) + \sigma\eta_k\sqrt{\delta t}$. In terms of F, the endpoint has to be written as a recursive function

$$\mathbf{x}_K = F(\mathbf{x}_{K-1}) + \sigma\eta_K\sqrt{\delta t} \tag{6.12}$$

$$\mathbf{x}_{K-1} = F(\mathbf{x}_{K-2}) + \sigma\eta_{K-1}\sqrt{\delta t} \tag{6.13}$$

$$\vdots \tag{6.14}$$

$$\mathbf{x}_1 = F(\mathbf{x}_0) + \sigma\eta_1\sqrt{\delta t} \tag{6.15}$$

We impose the end constraint by adding to the action a penalty $\Phi(\mathbf{x}_K) = \mathrm{dist}(\mathbf{x}_K, B)$, a function

which linearly increases with distance to B. The full optimization problem is

$$\min_{\eta}\left\{\frac{1}{2K}\sum_{k=1}^{K}\eta_k^{\top}\eta_k+\alpha\Phi(\mathbf{x}_K)\right\} \tag{6.16}$$

$$\mathbf{x}_0 \in A \text{ is fixed} \tag{6.17}$$

$$\mathbf{x}_k = F(\mathbf{x}_{k-1})+\sigma\eta_k\sqrt{\delta t} \text{ for } k=1,\ldots,K \tag{6.18}$$

Here α is a weight which can be increased to harden the end constraint. We divide by K so that the path action does not overwhelm the endpoint penalty as $K\to\infty$. (This makes the sum converge to an integral.) We set $\mathbf{x}_0$ to be the fixed point $\mathbf{a}\in A$ when finding the least-action path from A to B and the fixed point $\mathbf{b}\in B$ when finding the least-action path from B to A. We used the L-BFGS method as implemented in `scipy`, with a maximum of 10 iterations. We differentiate $\Phi(x_K)$ with respect to η_k using knowledge of the adjoint model, with a backward pass through the path to compute each gradient. At each descent step, we refine the stepsize with backtracking line search. One way to guarantee the end constraint is ultimately satisfied is to gradually increase α and lengthen T; however, we found it sufficient to fix $\alpha = 1.0$ and $T = 100$, in keeping with the typical observed transit time. We have kept the algorithm simple, not devoting too much effort to finding the global optimimum over all time horizons, as we only care for a qualitative assessment to compare with results of TPT.

As the stochastic forcing shrinks to zero, we expect $\mathbf{J}_{AB}$ to collapse into a single streamline following the minimum-action path (but becoming increasingly unlikely as we approach this limit). The finite-noise transition path ensemble, however, departs significantly from it. In the initial stages of transition in Fig. 6.4c, the minimum-action path tracks right down the center of $\mathbf{J}_{AB}$, suggesting this feature is stable with noise. At the end of the process, widening of current streamlines makes it impossible for the minimum-action path to represent the full ensemble meaningfully.

After an SSW event and ensuing vacillation cycles, the vortex eventually recovers, returning from B back to A, which is encoded by the current $\mathbf{J}_{BA}$ in Fig. 6.4d. The $B\to A$ current is very

different from a reversed $A \to B$ current. After many loops around B, $\mathbf{J}_{BA}$ emerges upward out of B just as in any other vacillation cycle, with a partial restoration of wind. The current then bifurcates: one branch continues its upward creep in zonal wind while reversing course in the IHF direction, eventually rebuilding a strong enough polar vortex to inhibit wave propagation and allowing radiative relaxation to take over, drawing it back into A. Meanwhile, the other branch of current continues to track with $\mathbf{J}_{BB}$ halfway through the wave amplification phase, as if about to execute another loop around $\mathbf{b}$. But $\mathbf{J}_{BA}$ stalls in the middle of the wave amplification phase, near IHF(30 km) $= 3 \times 10^4$ K·m^2/s. Where does the current go from there? Fig. 6.4(d,e) indicates that the eddy is centered slightly above the top of B, allowing some room for small vacillation cycles to proceed without technically re-entering B. This is the likely fate of some trajectories along the second branch of $\mathbf{J}_{BA}$, which finally achieve "escape velocity" the second time around.

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

6.2.3 Stages of an SSW from probability current

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

$$\oint_C \mathbf{J}_{AB} \cdot \mathbf{n} \, d\sigma = \text{Transition rate} \tag{6.19}$$

where $\mathbf{n}$ is an outward unit normal from C, σ is a surface element, and the transition rate is the average number of $A \to B$ transitions (SSW events) per unit time, or equivalently the inverse return period. The stochastic Holton-Mass model has a rate of $\sim (1700 \text{ days})^{-1}$, which changes with

parameters such as noise strength. (This is not a realistic rate for the real atmosphere, where there is a seasonal cycle and SSWs occur at least several times per decade.) The integral relationship (6.19) holds for any dividing surface, implying that the current is divergence-free outside of A and B, but has a source in A and a sink in B (vice versa for $\mathbf{J}_{BA}$). The integrand $\mathbf{J}_{AB} \cdot \mathbf{n}$, which we will henceforth call the $\mathbf{J}_{AB}$-*flux density* (not to be confused with heat flux or IHF) can be interpreted as a quasi-probability density, which is normalized to integrate to a constant (the transition rate) but may take on negative values for some choices of dividing surfaces. Because the number of $A \to B$ transitions per unit time must equal the number of $B \to A$ transitions per unit time, Eq. (6.19) must also hold when $\mathbf{J}_{AB}$ is replaced by $\mathbf{J}_{BA}$ and $\mathbf{n}$ is replaced by $-\mathbf{n}$. The reactive current essentially decomposes the rate among a continuum of possible pathways, which is much more dynamically insightful than the numerical value of the rate itself.

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

Surfaces of constant zonal wind

The simplest choice of dividing surfaces is a series of hyperplanes with constant $U(30\text{ km})$, represented as horizontal black lines in Fig. 6.4(b-e). To get from A to B, a trajectory must pass downward once through each threshold. It may also cross down, then up, then down; or three times down and two times up, etc., as long as the *net* number of downward crossings is one for each surface. The $\mathbf{J}_{AB}$-flux density element $\mathbf{J}_{AB}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x})\, d\sigma(\mathbf{x})$ can be interpreted as the long-term average number of net crossings through the surface at $\mathbf{x}$. Note that in the $A \to B$ direction, $\mathbf{n}$ points in the direction of negative $U(30\text{ km})$, i.e., $\mathbf{n} = -\nabla U(30\text{ km})/\|\nabla U(30\text{ km})\|$.

Fig. 6.5 shows the $\mathbf{J}_{AB}$-flux densities (a) and $\mathbf{J}_{BA}$-flux densities (b) across each surface. The

horizontal axis is IHF(30 km), as in Fig. 6.4, which instantiates the $\mathbf{J}_{AB}$-flux density element as

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

Here, the differential $d[\text{IHF}(30 \text{ km})]$ is shorthand for $\int dx_1 \ldots \int dx_{73}\, d\big[\text{IHF}(30 \text{ km})\big]$, where $x_1, \ldots, x_{73}$ are the 73 dimensions of state space orthogonal to both the IHF(30 km) and the $U(30 \text{ km})$ axes. Accordingly, the vertical axis of Fig. 6.5 has the units needed to normalize the integrals to a transition rate in days^{-1}. At the first $A \to B$ threshold $U(30 \text{ km}) = 47.3$ m/s, the flux distribution has a tall, narrow, negative spike, where $\mathbf{J}_{AB}$ points downward across the surface. There is also a small positive spike to the left due to a small amount of backflow where transition paths temporarily regain a bit of the lost zonal wind—not enough to re-enter A—before weakening again. This backflow corresponds to the small wiggles early in the black and cyan time series in Fig. 6.2. Moving from blue to red curves, as zonal wind drops further, we see the negative spike widen and slightly flatten, while the positive spike shrinks and disappears. By the last threshold $U(30 \text{ km}) = 8.3$ m/s, the $\mathbf{J}_{AB}$-flux density appears entirely negative, consistent with the sharp downturn into B seen in both Figs. 6.2 and 6.4c. It also covers a wider range of integrated heat flux, consistent with the weaker current magnitude pointing into B in Fig. 6.2c. The $\mathbf{J}_{BA}$-flux density somewhat mirrors the $\mathbf{J}_{AB}$-flux density, but with a larger backflow spike relative to the forward flow: in the early stages of vortex recovery (red and orange curves in panel (b)), a strengthening zonal wind at low values of IHF is accompanied by weakening zonal wind at higher value of IHF. This is consistent with the winding, branching character of $\mathbf{J}_{AB}$ in Fig. 6.4d, which inherits some clockwise circulation from $\mathbf{J}_{BB}$. In other words, the early $B \to A$ transition stages experience residual vacillation cycles, which ultimately dampen and die by the time zonal wind has reached 47.3 m/s (there is no noticeable negative dip in the dark blue curve in Fig. 6.5b).

These flux densities trace out a simpler version of the "transition tubes" defined in Vanden-Eijnden [2006]. The distributions cannot be interpreted as the path of a single event, but rather as the flow of SSW "traffic" through a sequence of thresholds, indicating the most frequently traveled

paths. Another important caveat is that a single-signed $\mathbf{J}_{AB}$-flux density (such as the red curve in Fig. 6.5a) does not imply strictly monotonic changes in zonal wind across that surface: it only means that the backflow, if present, is not systematically displaced from the forward flow along the IHF axis, as it is in the red curve in panel (b). However, a different choice of horizontal axis might reveal more coherent cyclical behavior. In general, reactive currents generally contain much more information that can be queried by slicing it along in different dimensions, which should be chosen with some physical intuition.

Surfaces of constant committor

Zonal wind, the defining coordinate for A and B, is an obvious measure of progress which we have used in Fig. 6.5. However, in some ways it is not the most natural. First, the presence of backflow, while it does reveal some interesting dynamics of transition paths, suggests that a particular zonal wind level might be associated with forward or backward progress depending on other variables. Second, by the time a typical transition path reaches the halfway point of $U(30\text{ km}) \approx 25$ m/s, its committor probability has risen to nearly 100% (cf. Figs. 6.3 and 6.4c; "typical" means along the main channel of $\mathbf{J}_{AB}$). The subsequent collapse of zonal wind is locked in by that point. The committor itself is a more balanced metric of progress toward B, and can be used the same way to find transition routes. A committor level set $\{\mathbf{x} : q_B^+(\mathbf{x}) = q_0\}$, i.e., all states with equal likelihood q_0 of SSW, is a dividing surface just like a level set of U, and thus supports a $\mathbf{J}_{AB}$-flux density similar to those in Fig. 6.5. We will see that this flux density is almost uniformly positive.

In Fig. 6.6, we plot a larger collection of $\mathbf{J}_{AB}$-flux densities, represented by gray histograms, across 15 level sets of the committor. The $\mathbf{J}_{AB}$-flux density elements for panels (a) and (b), are,

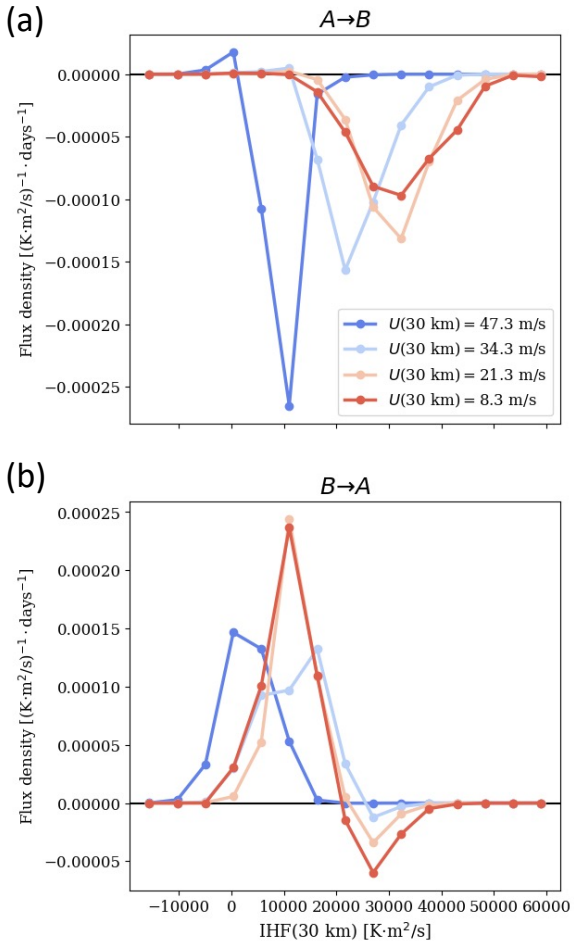


Figure 6.5: **J_{AB}-flux density (a) and J_{BA}-flux density (b) as a function of IHF(30 km), over four different level sets of U(30 km).** These cross sections of the reactive current from A to B and B to A illustrate the mean direction of trajectories crossing different zonal wind thresholds as a function the IHF. For an SSW (a), the progression marches from high winds (blue curves) to low winds (red) with increasing mean and variabilty of the IHF, while for the recovery of the vortex (b), the main progresssion is up toward higher wind, albeit with more substantial cycling down at higher values of IHF. Each density should have the same integral (in absolute value), equal to the rate. Due to numerical error, the integrals can vary and the rate is calculated by an averaging procedure (see chapter 2). For visual clarity, we have normalized each curve to have the same integral. To integrate to a rate, in days^{-1}, the vertical axis must have units of [K·m^2/s]$^{-1}$days^{-1}. This unit depends on the orientation of the dividing surface in state space, as well as the coordinates along that surface chosen for projection.

respectively,

$$\mathbf{J}_{AB} \cdot \left(\frac{\nabla q_B^+}{\|\nabla q_B^+\|} \right) d\big[U(30 \text{ km})\big] \tag{6.20}$$

$$\mathbf{J}_{AB} \cdot \left(\frac{\nabla q_B^+}{\|\nabla q_B^+\|} \right) d\big[\text{IHF}(30 \text{ km})\big] \tag{6.21}$$

We also display the minimum-action path with a dashed black curve for comparison. Panel (a) confirms that the zonal wind-committor relationship is nonlinear: approximately half of the total decline in zonal wind happens after the committor has surpassed 80% probability. The $\mathbf{J}_{AB}$-flux density widens across U in the late transition stages, past $q_B^+ \sim 0.7$. This indicates, somewhat puzzlingly, that the system may "commit" to B at a range of zonal wind strengths, even though B itself is defined by a fixed threshold $U(30 \text{ km}) \leq 1.75$ m/s. Fig. 6.3a offers some explanation: as the committor increases towards 1, the level sets become increasingly tilted across this two-dimensional state space. The last visible level set (the boundary between dark orange and red) spans the approximate range 5 m/s $\lesssim U(30 \text{ km}) \lesssim$ 30 m/s, depending on the value of IHF(30 km) along the horizontal. A large heat flux carries the promise of imminent SSW by sending waves into the stratosphere that will deposit enough negative momentum to surely destroy the vortex, even if the vortex is still persisting for the time being. If heat flux is weak, on the other hand, zonal wind must also be very weak to ensure the same degree of SSW certainty. Thus, the spread in zonal wind is closely tied with the spread in heat flux. This is consistent with Fig. 6.6b which shows the integrated heat flux distribution across each level set of q_B^+. Indeed, the distribution widens progressively from $q_B^+ \approx 0.5$ until the end of the path, consistent with the diffusing $\mathbf{J}_{AB}$ vector field and the diverging sample paths in 6.4c, as well as the broadening flux distributions in Fig. 6.5a. An interesting difference between the flux distributions and minimum-action path is that the latter decisively chooses the high-heat flux route, far outstripping the bulk of the flux distribution in Fig. 6.6b and hugging the right end of state space in Fig. 6.4c. We speculate that because stochastic forcing only acts on zonal wind, rather than the streamfunction (which determines heat flux), the

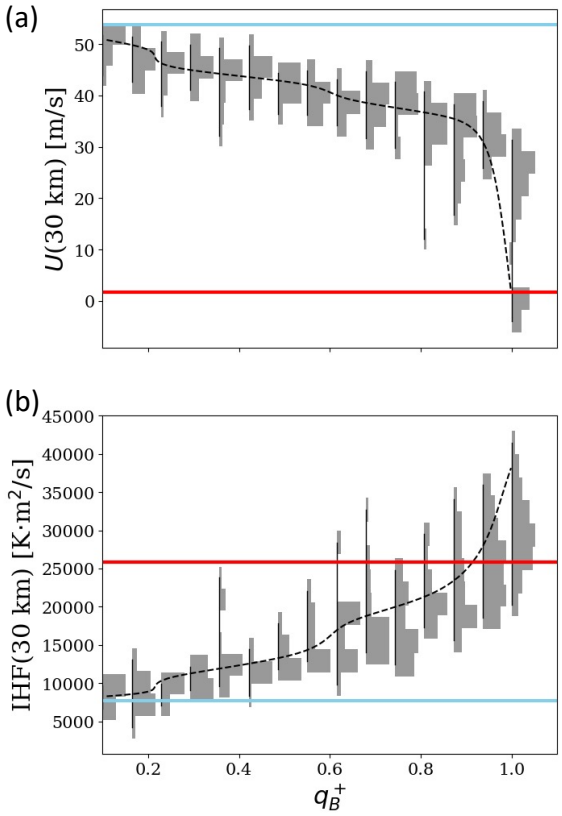


Figure 6.6: **Minimum-action paths and path distributions**. At a series of level sets in the committor q_B^+, gray histograms indicate the $\mathbf{J}_{AB}$-flux density of (a) zonal wind $U(30\text{ km})$ and (b) integrated heat flux IHF(30 km). Dashed curves show the minimum-action pathway in the same space. The minimum-action path tracks the mean of the full ensemble except very near SSW (q_B^+ near 1), where the jet breaks down more rapidly, accompanied by an extreme heat flux. The more extreme nature of the minimum-action path was also observed in Figure 6.4c, where it tracks along the rightmost envelope of more typical trajectories.

minimum-action path recruits the heat flux mechanism to do the "heavy lifting" of decelerating the zonal wind, thereby achieving SSW with a lower cost. An interesting future experiment would be to vary the form of stochasticity (5.3) and explore the consequences for flux distributions and minimum-action paths. TPT may thus offer an important rare event-oriented calibration tool for stochastic parameterization of climate models.

We have so far focused on observables at a fixed altitude of $z = 30$ km (or integrated up to 30 km), but the vertical structure of zonal wind and heat flux is essential to understand the physical processes of SSW onset. Every altitude z has a separate observable $U(z)$, with its own $\mathbf{J}_{AB}$-flux

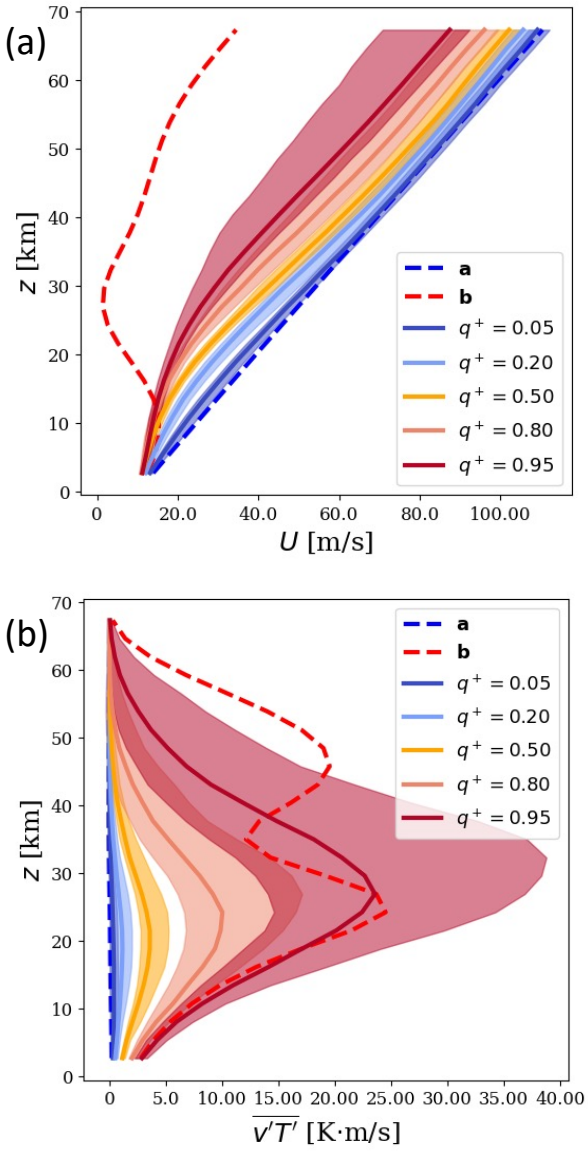


Figure 6.7: **Typical transition states and variability**. For a sequence of five committor ranges, we plot (a) the zonal wind profile and (b) the meridional heat flux profile that is most typical of that committor range in the sense of reactive current flux density. Shading represents the 25th-75th percentile range of the flux distribution. Blue and red dashed curves represent the profiles for the fixed points **a** and **b**, respectively. The widening of the distribution of both winds and IHF at high committor values (close to the SSW) highlights the diversity in late stage events which is lost in a composite approach (as in Figure 6.2) that pins all events together by the point of the vortex reversal. Even at a committor value of 0.95, the vortex is still largely intact above 15 km, emphasizing the importance of preconditioning the low level winds.)

density $\mathbf{J}_{AB} \cdot \nabla U(z)/\|\nabla U(z)\|$ of the same kind as Figs. 6.5 and 6.6. We visualize this z-indexed family of distributions in Fig. 6.7a by plotting their medians (solid curves) as functions of z, for five different committor level sets from 0 to 1. The background shading covers the interquartile range (25th-75th percentiles) of the $\mathbf{J}_{AB}$-flux density. There is essentially zero "backflow" across these surfaces, so the $\mathbf{J}_{AB}$-flux densities are ordinary nonnegative probability densities. Blue and red dashed curves represent the fixed points **a** and **b**. Fig. 6.7b shows the same construction, but with z-dependent meridional heat flux $\overline{v'T'}(z)$ as the independent variable. Together, these profiles give an idea of the joint evolution of propagating waves and weakening mean flow during the course of SSW.

As the committor increases from 0 to 0.6 (blue to yellow), the zonal wind profile slackens most noticeably at a low altitude range of 10-20 km, and the interquartile range remains narrow, suggesting that transitions are constrained to play out along a range of pathways with low variability. At the same time, meridional heat flux develops a positive bulge at the same low altitude range, indicating some upward flux of wave activity emanating from the troposphere. Later, as the committor increases to 1.0 (yellow to red), the wind profile stagnates at altitudes below 20 km, and above that continues weakening gradually. Most noticeably, the *variability*, both in zonal wind and heat flux, increases at higher altitudes of 30-50 km. At $q_B^+ = 0.95$, the distribution of zonal wind at high altitudes begins to skew sharply toward weak winds. Meanwhile, the distribution of heat flux profiles grows and widens, and the bulge moves slightly upward toward 30 km. This is consistent with the broadening of $\mathbf{J}_{AB}$ in IHF space in the final transition stages (Fig. 6.4c), and indicates a continued upward flow of wave activity. A slight change in zonal wind belies a substantial increase in SSW probability, which will eventually bring about an abrupt breakdown and explosion of variability in zonal wind.

Evolution over time

We have now measured SSW progress by two different coordinates, $U(30\text{ km})$ and $q_B^+(\mathbf{x})$, and visualized its composite evolution in both spaces. What neither of them captures directly is time: how long does SSW take to complete, and how long is each stage? We wish to produce a TPT version of the composite evolution shown in Fig. 6.2. To do this, we replace the hitting time τ_B^+ (a random variable) with its conditional expectation, the *lead time*,

$$\eta_B^+(\mathbf{x}) = \mathbb{E}_{\mathbf{x}}\big[\tau_{A\cup B}^+ | \mathbf{X}(\tau_{A\cup B}^+) \in B\big], \tag{6.22}$$

in other words, the average time from $\mathbf{x}$ to B conditional on hitting B before A. The composites in Fig. 6.2b parameterize the SSW process by τ_B^+ itself, which varies randomly from path to path, whereas $\eta_B^+(\mathbf{x})$ is the average value of τ_B^+ over all possible paths and hence a deterministic function of state space. We used η_B^+ as a forecast function in Finkel et al. [2021b], and we display it here in Fig. 6.8 over the same two-dimensional subspace, along with several committor level sets for comparison. $\eta_B^+(\mathbf{x})$ is uniformly zero on set B, increases farther away from B, and becomes undefined on set A. Fig. 6.8b gives an idea of how the certainty of SSW is related to the time until it happens. It turns out that along transition paths, the committor increases at an approximately linear rate with respect to time. Both the flux distributions and the minimum-action path indicate that the lead time drops by $\sim$8 days for every additional $\sim$10% in the likelihood of SSW. In particular, this means that the ultimate collapse of zonal wind in Fig. 6.6 is not only "sudden" with respect to the committor, but also with respect to the lead time. The final 20 days of the transition path (as measured by η_B^+) corresponds to the final $\sim 5\%$ of probability needed to achieve SSW, from 95% to 100%, and yet this same interval sees approximately 30 m/s reduction in zonal wind—the entire second half of its journey from A to B. This is the sense in which the pre-sudden part of SSW constitutes most of the probabilistic progress. Dynamically, it seems that this half-weakened polar vortex has been accompanied by "irreversible" changes in the flow field, the Holton-Mass version

of the threshold behavior found in Nakamura et al. [2020].

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

$$\mathbf{J}_{AB} \cdot \left(\frac{\nabla q_B^+}{\|\nabla q_B^+\|} \right) d\left[\eta_B^+\right] d\left[U(30\text{ km})\right] \tag{6.23}$$

whose two-dimensional integral is, again, the transition rate. For a sequence of 30 committor level surfaces, Fig. 6.9a shows quantiles of $U(30\text{ km})$ (a) and IHF(30 km) (b) vs. the median lead time η_B^+ in the horizontal. These "TPT composites" resemble the traditional composite of Fig. 6.2b. but differ in several important ways. The traditional composite narrows toward the end, by construction, since the entrance to B is defined by a single value of $U(30\text{ km})$. In contrast, the TPT composite widens toward the end before the final narrowing: as Fig. 6.8a demonstrates, the level sets of q_B^+ and η_B^+ closest to B both cover a range of $U(30\text{ km})$ values. The final collapse of zonal wind, which typically happens in the lower-right corner of state space, is so sudden that the lead time hardly changes, and so inevitable that the committor hardly changes. Of course, formally $\eta_B^+ = 0$ and $q_B^+ = 1$ if and only if $U(30\text{ km}) \leq 0$, a boundary condition we have enforced in Fig. 6.9a. From the TPT perspective, however, the process is essentially complete.

The TPT composite also has a wavy character not captured by the traditional composite. The individual samples in Fig. 6.2a do seem to proceed in pulses of steady downward progress punctuated by brief, partial recoveries. Because these partial recoveries are staggered in time between paths, the traditional composite in Fig. 6.2b cannot capture them. However, these wiggles may correspond robustly to various level sets of committor or lead time, which would suggest the waviness of the TPT composite is indeed capturing this same phenomenon. Some of the gray transition paths in Fig. 6.2 go through even larger oscillations after approaching close to B, which may correspond to the rapid expansion of the outer envelope (middle 90 percentile) in Fig. 6.9a at $\eta_B^+ = 30$ days.

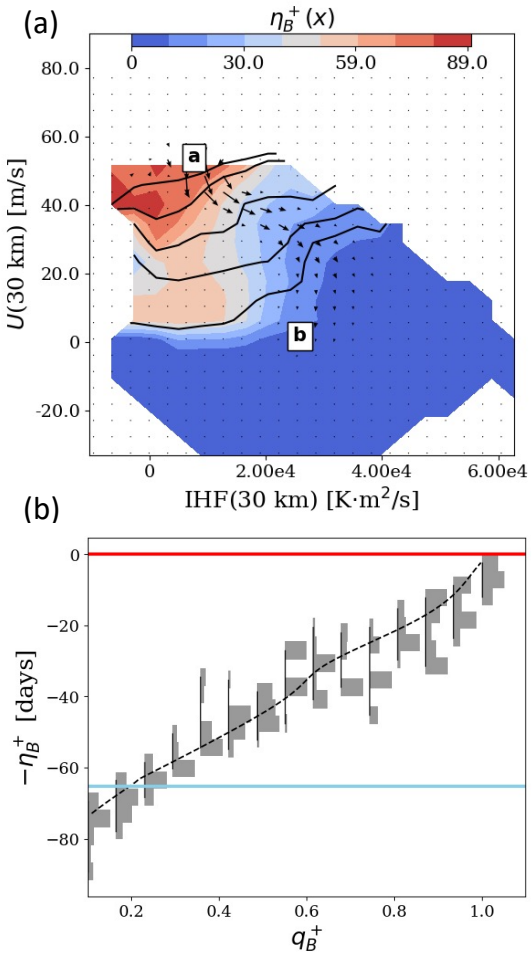


Figure 6.8: **Lead time-committor relationship.** (a) Background color shows η_B^+, the expected time to reach B from initial condition $\mathbf{x}$, conditional on hitting B next. Note that the contour structure is very different from that of the forward committor, whose level sets $q_B^+ = 0.1$, 0.2, 0.5, 0.8, and 0.9 are shown in solid black lines (cf. Fig. 6.3). Notable differences are in the light red region where the wind is approximately 20 m/s and IHF near 10^4 K· m/s: SSW events rarely occur from these initial conditions, and are associated with long trajectories (lead time of about 60 days) that often cycle back towards state A before swinging down to state B. Probability current $\mathbf{J}_{AB}$ is overlaid, the same as in Fig. 6.4c. (b) The distribution of lead time across a series of level sets of the committor, the same level sets as in Fig. 6.6.

The probability currents in the lower left corners of Fig. 6.4(a,d,e) indicate, indeed, that this region is associated with *increasing* zonal wind strength, which of course is only temporary if the trajectory is bound for B. These partial recoveries may be interpreted as minor warmings preceding the major warming. Nevertheless, the individual pathways are only case studies, and their detailed correspondence with the TPT composite is speculative. A more refined DGA discretization, or a large-scale time series statistical analysis, would confirm or deny the robustness of these oscillatory features, but such analysis is beyond the scope of this paper.

6.3 Numerical method

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

DGA is detailed in chapter 2, but we briefly sketch the procedure here. The key observation underpinning DGA is that unknown "forecast functions" of interest — $q_B^+(\mathbf{x}), q_A^-(\mathbf{x}), \eta_B^+(\mathbf{x}), \pi(\mathbf{x})$, etc — can be expressed as solutions to equations involving only short-time evolution of $\mathbf{X}$. For example, the committor, q_B^+, solves the equation

$$\begin{aligned} q_B^+(\mathbf{x}) &= \mathbb{E}_{\mathbf{x}}\left[q_B^+(\mathbf{X}(\Delta t))|\mathbf{X}(0)=\mathbf{x}\right], \quad \mathbf{x}\notin A\cup B \\ q_B^+(\mathbf{x}) &= 1,\ \mathbf{x}\in B \quad \text{and} \quad q_B^+(\mathbf{x}) = 0,\ \mathbf{x}\in A \end{aligned} \tag{6.24}$$

for $\mathbf{x}$ outside of A and B. In this equation we interpret evolution of $\mathbf{X}(t)$ to stop upon entrance to A or B. The user-chosen parameter Δt limits the length of the simulated trajectories. Crucially, Eq. (6.24) identifies q_B^+ exactly for any choice of Δt.

To approximately solve Eq. (6.24) and similar equations for other quantities of interest, we first generate a data set by sampling many points $\mathbf{X}_n(0)$ from all over state space according to

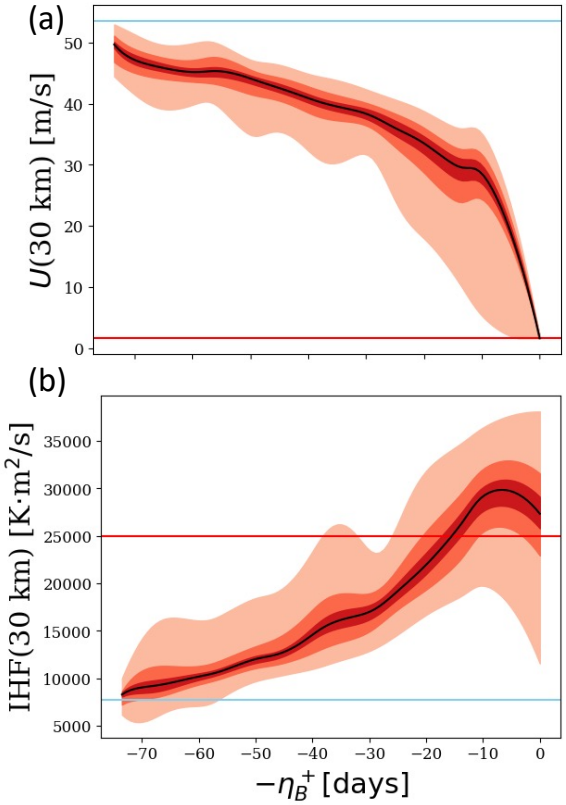


Figure 6.9: **TPT composite evolution vs. time.** For 15 committor level sets (the same as in Figs. 6.6 and 6.8b) we approximate the joint distribution of (a) lead time and zonal wind, and (b) lead time and integrated heat flux, according to the flux density of $\mathbf{J}_{AB} \cdot \mathbf{n}$ through the committor level surface. The three red-orange envelopes represent the middle 20%, 50%, and 90% percentile ranges. Black curves connect the medians. Unlike the traditional SSW composite shown in Figure 6.2, the variability in trajectories is more uniform in lead time, actually increasing near the event. This is due to use of the committor as the ordering coordinate, which aligns paths by the future predictability of an event. The widening at near -10 days reflects the diversity of model states when an SSW is approximately 95% likely to occur, as seen in Figure 6.7. All of these states are equally likely to move to an SSW with an expected lead time of 10 days, but there is a distribution of actual lead times which contributes to the spread in winds and heat flux.

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

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

$$q_B^+(\mathbf{x}) \approx \sum_{j=1}^{M} w_j(q_B^+)\phi_j(\mathbf{x}) \tag{6.25}$$

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

$$\phi_j(\mathbf{x}) = \mathbb{1}_{S_j}(\mathbf{x}) := \begin{cases} 1 & \mathbf{x} \in S_j \\ 0 & \text{otherwise.} \end{cases} \tag{6.26}$$

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

The Galerkin method proceeds by inserting the expansion in Eq. (6.25) into the short-trajectory equation solved by the quantity of interest (Eq. (6.24) in the case of q_B^+) and then integrating both sides against a test function ϕ_i, also from the basis. The result is in an $M \times M$ linear system. With an indicator basis as in Eq. (6.26), the matrix elements yield a Markov transition probability matrix

$$P_{ij} = \mathbb{P}_\mu\{\mathbf{X}(\Delta t) \in S_j \,|\, \mathbf{X}(0) \in S_i\}, \;\; i, j \in \{1, \ldots, M\}. \tag{6.27}$$

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

$$P_{ij} = \frac{\#\{n : \mathbf{X}_n(0) \in S_i, \mathbf{X}_n(\Delta t) \in S_j\}}{\#\{n : \mathbf{X}_n(0) \in S_i\}}, \tag{6.28}$$

Given the transition matrix P_{ij}, the committor coefficient vector obeys a discrete version of Eq. (6.24):

$$\begin{aligned} w_i(q_B^+) &= \sum_{j=1}^{M} P_{ij} w_j(q_B^+), \quad S_i \not\subseteq A \cup B \\ w_i(q_B^+) &= 1, \; S_i \subseteq B \quad \text{and} \quad w_i(q_B^+) = 0, \; S_i \subseteq A \end{aligned} \tag{6.29}$$

We have assumed that A, B, and $(A \cup B)^{\mathrm{c}}$ are partitioned separately, meaning each S_i is either completely inside A, completely inside B, or disjoint from both, which we ensure in the clustering step. As in Eq. (6.24), in Eq. (6.29) we interperet evolution of $\mathbf{X}_n(t)$ to be stopped upon entrance to A or B.

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

$$\begin{aligned} w_i(\pi) &= \sum_{j=1}^{M} w_j(\pi) P_{ji} \qquad\qquad i = 1, \ldots, M \\ \sum_{j=1}^{M} w_j(\pi) &= 1. \end{aligned} \tag{6.30}$$

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

The backward committor obeys a similar equation to (6.29), but with two differences. First, P_{ij} is replaced by $\widetilde{P}_{ij} = \frac{w_j(\pi)}{w_i(\pi)} P_{ji}$, which represents the process under time reversal at steady-state. Second, for q_A^-, the boundary conditions are flipped from those of q_B^+: $w_i(q_A^-) = 1$ for $S_i \subseteq A$ and $w_i(q_A^-) = 0$ for $S_i \subseteq B$. Because the time-reversed matrix depends on the steady-state density, q_A^- must be solved after π. The lead time η_B^+ solves a similar, but slightly more intricate, equation, which can be found in chapter 2.

With approximations to the committors and steady-state density provided by DGA (or any other means), TPT provides recipes to assemble approximations of the transition path statistics examined in this chapter. For example, the reactive density π_{AB} can be computed directly from its definition in (6.7). The transition rate and projections of the reactive current $\mathbf{J}_{AB}$ are estimated by more involved procedures presented in detail in the background chapter 2.

6.4 Numerical benchmarking of DGA

To validate DGA numerically, we can compare to the results of DNS. In Fig. 5.7, we saw convergence of the DGA committor to the DNS committor across state space as sample size and lag time were increased. Here, we turn our attention to summary statistics of interest for full transition paths, not just forecasting. This will benchmark our current DGA implementation for comparison with future algorithmic developments.

Fig. 6.10a displays the time fractions spent in each phase of the SSW lifecycle: $A \to B$, $B \to A$, $A \to A$, and $B \to B$, including estimates from DNS (cyan) and DGA (red) and their uncertainties. The DGA estimate of the $A \to B$ time fraction is a π-weighted average of $q_A^-(\mathbf{x})q_B^+(\mathbf{x})$ over state space,

$$\langle q_A^- q_B^+ \rangle_\pi = \int q_A^-(\mathbf{x}) q_B^+(\mathbf{x}) \pi(\mathbf{x})\, d\mathbf{x} \tag{6.31}$$

and similarly for the other phases. The DGA error bars are generated by repeating the entire pipeline three times with different short trajectory realizations. The bar height shows the mean, and the error bars show the minimum and maximum. The DNS error bars are generated by bootstrap resampling (with replacement) 500 times from the control simulation, treating an entire SSW lifecycle as a single unit (from the beginning of one $A \to B$ transition until the beginning of the next one). This assumes no memory between successive events, which we have found to be reasonable; there is insignificant autocorrelation between consecutive return periods. The bars extend two root-mean-squared errors in both directions, enclosing a 95% confidence interval. To first order, DGA agrees well with DNS on the fraction of time spent in each phase. A is the more stable of the two regimes, accounting for $\sim 50\%$ of the time compared to the $\sim 40\%$ of time spent in the orbit of B. The transition events are both an order of magnitude shorter, with $B \to A$ taking slightly longer on average. DGA ranks the $A \to B$ and $B \to A$ time fractions correctly, despite a bias in the absolute magnitudes.

The numbers in Fig. 6.10a are only relative durations; they do not tell us how long a full life cycle takes. That number is given by (one over) the rate. Fig. 6.10b shows three different rate estimates (that is, the generalized rate with $\Gamma = 0$) using the formulas above. The cyan bars come from DNS, counting the number of $A \to B$ transitions per unit time. Of course, this equals the number of $B \to A$ transitions per unit time, so the $A \to B$ and $B \to A$ cyan bars are identical. Error bars come from bootstrapping, as with the relative durations. The red bars come from DGA, and these estimates are not technically symmetric. The DGA estimate labeled $A \to B$ integrates $\mathbf{J}_{AB} \cdot \mathbf{n}$ over dividing surfaces with $\mathbf{n}$ pointing away from A toward B, while the estimate labeled $B \to A$ integrates $\mathbf{J}_{BA} \cdot \mathbf{n}$ over surfaces with $\mathbf{n}$ pointing away from B toward A. Numerical and sampling errors cause slight differences between them, but Fig. 6.10b shows them both to come within 20% of the DNS estimate.

DGA estimates should converge with increasing M (cluster number) and N (short-trajectory ensemble size). Larger M makes the approximation space $\{\phi_1, \ldots, \phi_M\}$ more expressive, making finer estimates possible. However, as M grows, we need more short trajectories N to robustly estimate the entries of the expanding matrix (6.28). Conversely, as M shrinks, P_{ij} will become closer to diagonal, because trajectories will escape from their starting cluster less frequently. Thus Δt would have to increase when M decreases. The optimal choice for a given model will depend on the relative costs of integrating the model, building basis sets, and solving large linear systems on different computer architectures. With our choice of $M = 1500$, increasing N from 5×10^4 to 3×10^6 does not change the DGA point estimates very much, but shrinks the error bars by a factor of $\sim$4. To further reduce the bias in Fig. 6.10, we would likely need more refined basis functions, perhaps using nonlinear features as input to K-means. For generalized rates such as transit time $\tau_B^+ - \tau_A^-$ and total heat flux $\int_{\tau_A^-}^{\tau_B^+} \overline{v'T'}(30\text{ km})\,dt$, a second-order calculation is required using Eq. (2.29), which causes errors to propagate further. The errors are not yet well-controlled enough to present the results of generalized rates. We do not yet have theoretical guarantees or optimal prescriptions for DGA parameters, but given the flexibility and parallelizability of the

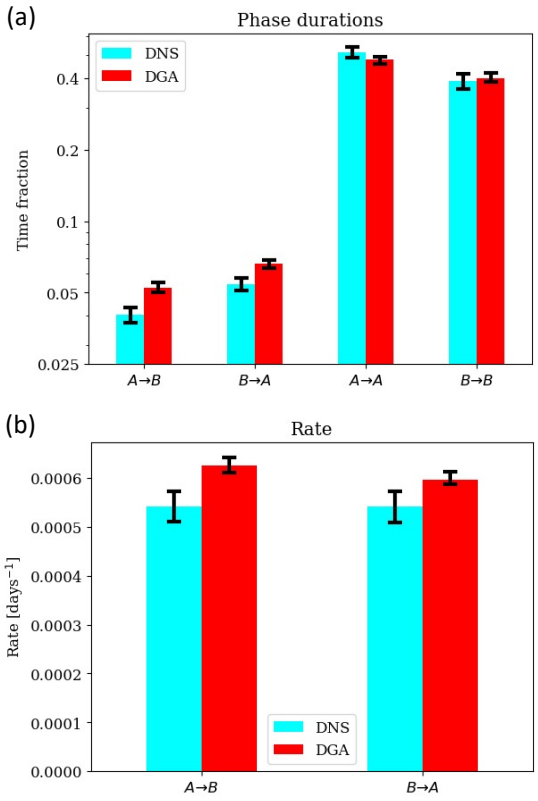


Figure 6.10: **DGA benchmarks and comparison to DNS.** (a) Time fractions spent in each phase. (b) Total SSW rate estimated using both $\mathbf{J}_{AB}$ and $\mathbf{J}_{BA}$; the two cyan columns, DNS estimates, are identical.

method, we believe it has much room for growth.

6.5 Conclusion

Using TPT analysis, we have shown that transition paths in the Holton-Mass model generally evolve through two distinct phases: (i) a gradual, halting decline in zonal wind strength in tandem with a slowly increasing meridional heat flux over a period of approximately 2 months, followed by (ii), a rapid burst of heat flux and deceleration of zonal wind in the last 10 days. The sudden breakdown of the vortex in the second stage encompasses the classic synoptic evolution of an SSW, but from a predictability standpoint, it is changes in the precondition phase that are most critical, allowing one to forecast a warming before the event is already in motion. Our key conclusion is

the SSW committor probability rises the most during the preconditioning phase. The committor signals an upcoming SSW before changes in the vortex (as quantified by just the zonal mean zonal wind) can be clearly identified above the noise in an individual trajectory.

A judicious choice of the "climatological state" A is essential to maximize predictive and dynamical understanding of the rare event's origin when using the TPT framework. In defining A relative to winds in the strong vortex meta-stable state, we were able to fully include stage (i). This lengthened the window over which we could tracked SSW trajectories to seasonal time scales. Extending this work to the atmosphere, where the climatological state is itself evolving on comparable time scales, remains a challenge. Set B, too, may be adjusted to compare between different kinds of rare events. The following chapter 7 does exactly that, varying B systematically to lower and lower thresholds to examine the dependence of the rate on the severity of extreme events.

7 REVEALING THE STATISTICS OF EXTREME EVENTS HIDDEN IN SHORT WEATHER FORECAST DATA

Extreme weather events have significant consequences, dominating the impact of climate on society, but occur with small probabilities that are inherently difficult to compute. A rare event with a 100-year return period takes, on average, 100 years of simulation time to appear just once. Computational constraints limit the resolution of models used for such long integrations, but high resolution is necessary to resolve extreme event dynamics. We demonstrate a method to exploit short-term forecasts from a high-fidelity weather model and lasting only weeks rather than centuries, to estimate the long-term climatological statistics of rare events. Using only two decades of forecast data, we are able to robustly estimate return times on the centennial scale. We use the mathematical framework of transition path theory to compute the rate and seasonal distribution of sudden stratospheric warming (SSW) events of varying intensity. We find SSW rates consistent with those derived from reanalysis data, but with greater precision. Our method performs well even with simple feature spaces of moderate dimension, and holds potential for assessing extreme events beyond SSW, including heat waves and floods.

Plain Language Summary

Weather extremes are a continually recurring threat to human life, infrastructure, and economies. Yet, we only have sparse datasets of extremes, both simulated and observed, because by definition they occur rarely. We introduce an approach to extract reliable extreme event statistics from a non-traditional data source: short, high-resolution weather simulations. With 21 years of 47-day weather forecasts, we estimate probabilities of once-in-500-year events.

Key points

1. Extreme weather risk, as measured by rate or return times, is inherently difficult to analyze because of data scarcity.

2. Transition path theory reveals climatological statistics of sudden stratospheric warming events from high-fidelity subseasonal forecasts.

3. Rates and seasonal distributions of 100-year stratospheric extremes are robustly computed from 47-day hindcast ensembles across 21 winters.

7.1 Introduction

The atmosphere's extreme, irregular behavior is, in some ways, more important to characterize than its typical climatology. A society optimized for historical weather patterns is highly exposed to damage from extreme heat and cold, flooding, and other natural hazards. Extremes may respond more sensitively than mean behavior to climate change, an argument supported by elementary statistics [Wigley, 2009], empirical observations [Coumou and Rahmstorf, 2012, AghaKouchak et al., 2014, O'Gorman, 2012, Huntingford et al., 2014, Naveau et al., 2020] and simulations [Pfahl et al., 2017, Myhre et al., 2019]. Recent unprecedented extreme weather events demonstrate the serious human impacts [Mishra and Shah, 2018, Van Oldenborgh et al., 2017, Goss et al., 2020, Fischer et al., 2021]. The overall "climate sensitivity" [Hansen et al., 1984], summarized by a change in global-mean temperature, does not do justice to these consequences, which has led the community to develop "event-based storylines" [Shepherd et al., 2018, Sillmann et al., 2021] as a more tangible expression of climate risk.

The intermittency of extreme events makes precise risk assessment exceedingly difficult. 100 flips of a biased coin with $\mathbb{P}\{\text{Heads}\} = 0.01$ is almost as likely to yield zero heads (probability 0.366) as one head (probability 0.370), and half as likely to yield two heads (probability 0.185). Similarly, in a 100-year climate simulation or historical record, a once-per-century event may easily

appear either non-existent or twice as likely as it really is. The difficulty exists even in a stationary climate, but worsens in the presence of time-dependent forcing, anthropogenic or otherwise. The limited historical record forces us to use numerical models as approximations, introducing a dilemma: we can run cheap, coarse-resolution models for long integrations, providing reliable statistics of a biased system, or expensive, high-resolution models for short integrations, which have lower bias but higher-variance due to under-sampling. Long-term climate simulations are usually performed with a low resolution of $O(50-100)$ km per grid cell [Haarsma et al., 2016]. A coarse model might suffice to estimate global-mean temperature and other aggregated statistics, but cannot resolve convective systems, e.g., tropical cyclones and precipitation over complex topography, that deliver localized but heavy damage [O'Brien et al., 2016, He et al., 2019]. Even large-scale events, such as a sudden stratospheric warming (SSW, the specific application of this paper) might arise from multi-scale interactions that are poorly represented in coarse model grids.

To obtain accurate dynamics and statistics, we must use the highest-fidelity models available, currently exemplified by the Integrated Forecast System (IFS) of the European Center for Medium-Range Weather Forecasts (ECMWF). Running at high resolutions of $\sim$16-32 km [ECMWF, 2016e], the IFS produces skillful ensemble forecasts spanning $\sim$1 week-1 month. Such a high-resolution model can generate a highly plausible "storyline", but cannot feasibly run long enough to estimate the climatological rate of an extreme event.

In this work, we help close this gap by assembling fragmented weather forecast ensembles together to cover the full dynamically relevant phase space. By re-weighting ensemble members in a principled way, we estimate probabilities of sudden stratospheric warming (SSW) events, in which the winter stratospheric polar vortex rapidly breaks down from its typical state, a strong cyclonic circulation over the winter-hemisphere pole. The associated subsidence and adiabatic warming can cause lower-stratospheric temperatures to rise by more than 40 K over several days [Baldwin et al., 2021]. The reversal of stratospheric winds forces upward-propagating planetary waves to break at lower and lower levels, exerting a "downward influence" on tropospheric circulation [Baldwin and

Dunkerton, 2001, Baldwin et al., 2003, Hitchcock and Simpson, 2014, Kidston et al., 2015]. The midlatitude jet and storm track shift equatorward, bringing extreme cold spells and other anomalous weather to nearby regions [Kolstad et al., 2010, Kretschmer et al., 2018a]. King et al. [2019] documents the impact of an SSW on extreme winter weather over the British Isles, the so-called "Beast from the East" in February 2018. SSWs are a demonstrated source of surface weather predictability on the subseasonal-to-seasonal (S2S) timescale, a frontier of weather forecasting with many implications for helping humanity deal with meteorological extremes [Sigmond et al., 2013, Scaife et al., 2016, White et al., 2017, Vitart and Robertson, 2018, Butler et al., 2019, Lang et al., 2020, Bloomfield et al., 2021, Scaife et al., 2022]. For these reasons, there is keen interest in improving (i) the prediction of SSW itself beyond the horizon of $\sim$10 days that marks the current state-of-the-art [Tripathi et al., 2016, Domeisen et al., 2020], and (ii) understanding of the long-term frequency, seasonal distribution, and other climatological statistics of SSW.

The ensemble forecasts archived in the S2S project at ECMWF [Vitart et al., 2017] have the potential to provide more precise statistics than the limited historical data. We describe our data sources in section 7.2. To realize this potential requires a method to stitch the short trajectories together, which we outline in section 7.3 and describe more fully in Supporting Information. Section 7.4 presents our main result: with data consisting of 47-day forecasts over a 21-year period, we estimate rates and seasonal distributions of SSW events which, depending on severity, occur as rarely as once in 500 years. We discuss the implications in section 7.5 and conclude in section 7.6.

7.2 Data and definitions

Fig. 7.1(a,b) show the evolution of zonal-mean zonal wind at 10 hPa and 60°N (which we abbreviate $U_{10,60}$), a standard index for the strength of the stratospheric polar vortex. Black timeseries show $U_{10,60}$ through two consecutive winters where SSW occurred, 2008-2009 (a) and 2009-2010 (b), superimposed on its 70-year climatology in gray from the ERA-5 reanalysis dataset [Hersbach et al., 2020]. $U_{10,60}$ is typically positive throughout the winter months, characterizing a strong cir-

cumpolar jet that forms in the stratosphere during the polar night. Occasionally, however, the vortex breaks down and $U_{10,60}$ reverses direction, becoming negative in the middle of winter. This is the standard definition of an SSW event [e.g., Butler et al., 2015], but it does not capture the range of intensities between events. Clearly, January 2009 achieved a much more negative $U_{10,60}$ level than February 2010. More intense SSW events have been linked to stronger tropospheric impacts [Karpechko et al., 2017, Baldwin et al., 2021], which motivates our efforts to distinguish between them. Historical data can provide reasonably robust estimates of moderately rare events such as February 2010, in which $U_{10,60}$ barely reversed sign; events of this magnitude occur on average every two years. On the other hand, extraordinary events like January 2009 are quite poorly constrained due to small sample size, while carrying an outsize risk in a nonstationary climate [Fischer et al., 2021].

To quantify SSW intensity, we vary the the $U_{10,60}$ threshold—henceforth called $U^{(\mathrm{th})}_{10,60}$—from 0 m/s to -35 m/s in 5 m/s increments and consider each case separately. Horan and Reichler [2017] and Butler and Gerber [2018] have suggested the utility of examining different thresholds, as SSW events form a continuum. Horizontal red lines in Fig. 7.1(a,b) mark each threshold. Vertical blue lines frame the winter period of November 1-February 28 in which we allow SSWs to occur, to exclude "final warmings" at winter's end when the vortex dissipates for the summer [Black et al., 2006]. We only count the first event of the season, to avoid counting the subsequent oscillations of $U_{10,60}$ about $U^{(\mathrm{th})}_{10,60}$ as separate SSW events. A minimum separation time can also be imposed, as in [Charlton and Polvani, 2007], to allow multiple SSWs in a season, but these are rare and for the purpose of demonstration, we keep the definition as simple as possible.

In addition to reanalysis, panels (a,b) also display a small sample of the S2S dataset in purple. These are not forecasts but *reforecasts*, or *hindcasts*, generated by initializing a present-day model version on past weather conditions. The S2S archive compiles forecasts and hindcasts from 11 forecasting centers around the world [Vitart et al., 2017], with a principle goal of tracking improvements in skill from one version to the next. In this study we restrict ourselves to the ECMWF

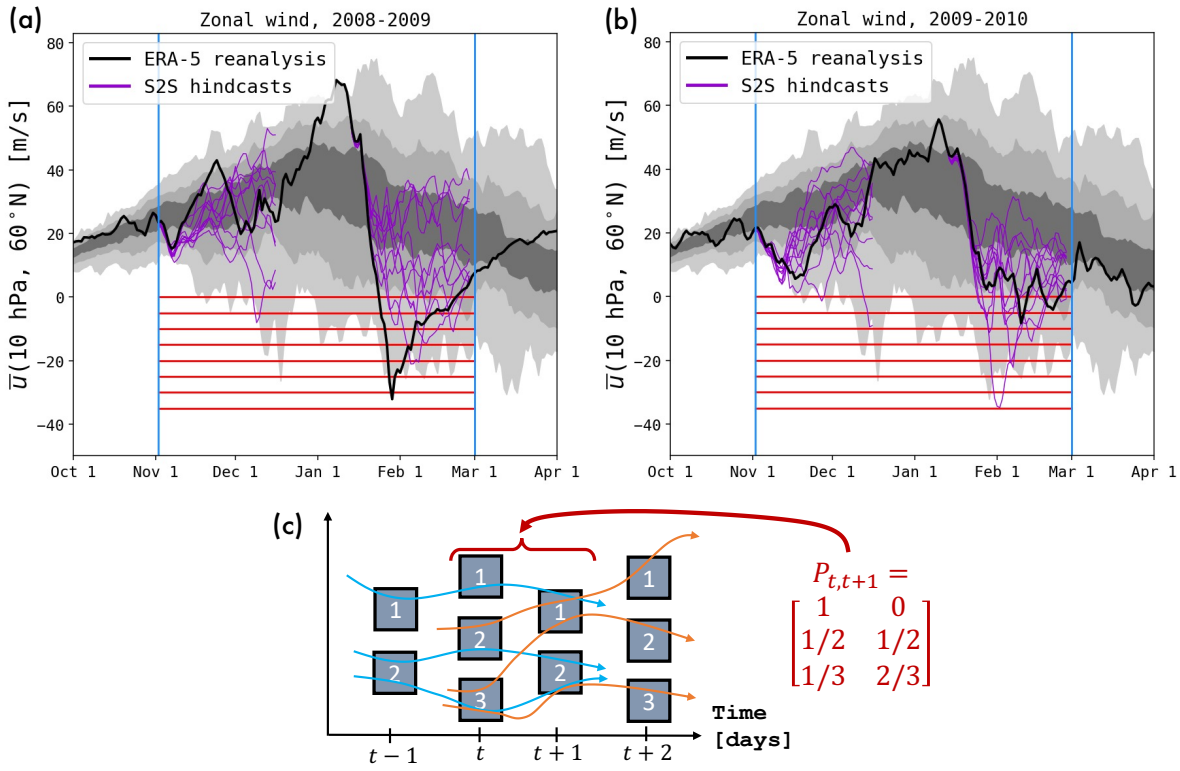


Figure 7.1: **Climatology of polar vortex and illustration of dataset**. (a,b): 70-year climatology of $U_{10,60}$ according to ERA-5, with the middle 40-, 80-, and 100-percentile envelopes in lightening gray envelopes. Two individual years are shown in black: 2008-2009 (a) and 2009-2010 (b). Two ensembles of S2S hindcasts (purple) are shown each winter, a small sample from the large S2S dataset of two ensembles *per week* from the ECMWF IFS. A range of SSW thresholds $U_{10,60}^{(\text{th})}$ from 0 m/s to -35 m/s are marked by horizontal red lines. When $U_{10,60}$ crosses this line from above, an SSW has occurred, provided it happens between the vertical blue lines marking November 1 and Feb. 28. (c) Schematic of the Markov state model approximation we use to estimate rates. Blue and orange curves represent the partial trajectories from S2S. At each time step the data are clustered into discrete boxes, and probability transition matrices estimated by counting transitions from one day to the next.

IFS, although our methodology can be repeated on other S2S datasets for intercomparison. We use data from the 2017 model version CY43R1, which produced 21 full winters of hindcasts between autumn of 1996 and spring of 2017. These are initialized using ERA-Interim (ERA-I) reanalysis [ECMWF, 2011], which is is almost identical to the more advanced ERA-5 from the standpoint of $U_{10,60}$. Two ensembles are launched every week, each with eleven members (one control and ten perturbed forecasts) that run for 47 days before terminating. We use only the ten perturbed members, which are initialized using a singular vector method and integrated with stochastic physics schemes [ECMWF, 2016e]. This introduces randomness into the ensemble, causing the members to drift apart over time after the initialization date, as shown in Fig. 7.1(c,d) for two sample ensembles. The specific strategy for perturbation of initial conditions and stochastic physics is informed by chaotic dynamical systems theory and has been refined by decades of numerical experiments [Mureau et al., 1993, Rabier et al., 1996, Palmer et al., 1998, Gelaro et al., 1998, Leutbecher, 2005, Lawrence et al., 2009, Buizza et al., 1999, Palmer et al., 2009] aimed at reducing forecast error due to under-dispersion, especially in the face of oncoming flow regime transitions [Trevisan et al., 2001]. In total, the S2S dataset contains over 900 years of simulation time. Many of them reach farther into the negative-$U_{10,60}$ tails than reanalysis, allowing us to calculate otherwise inaccessible probabilities.

7.3 Long-timescale dynamics from short trajectories

The advantage of sheer data volume comes with two attendant disadvantages. First, not all trajectories are independently sampled: on the contrary, all members of an ensemble are initialized close to reanalysis, and take several days to separate. Thus, the effective sample size is smaller than 900 years. Second, no individual ensemble can directly provide an SSW probability beyond the 47-day time horizon, which is well short of the 120 days between November 1 and February 28 when SSWs are allowed to happen. To make use of the "hanging" trajectory endpoints and infer what might have transpired were the simulation to continue, we construct a *Markov state model* (MSM)

[Deuflhard et al., 1999, Pande et al., 2010, Chodera and Noé, 2014] which is sketched in Fig. 7.1c. At every time sample $t = 1$ day, 2 days, ..., we partition state space into a disjoint collection of bins $S_{t,1}, S_{t,2}, \ldots, S_{t,M_t}$ and approximate the transition probability matrix for each time-step from t to $t+1$,

$$P_{t,t+1}(i,j) = \mathbb{P}\{\mathbf{X}(t+1) \in S_{t+1,j} | \mathbf{X}(t) \in S_{t,i}\}, \tag{7.1}$$

by counting the transitions between corresponding boxes. The matrices are row-normalized, which corrects for the redundancy and non-independence of ensemble members. Here, $\mathbf{X}(t)$ represents the full state vector of the ECMWF model. This sequence of matrices is the key ingredient that enables all downstream calculations, and it merits a brief note about the approximations involved. In a low-dimensional space, the partition could be created with a regular grid. However, every snapshot from the IFS has millions of degrees of freedom, including temperature and wind velocity in (latitude, longitude, pressure)-regular voxels. Any attempt to represent the dynamics of all these variables using a model such as (7.1) would suffer from large statistical error. On the other hand, if we only attempt to represent the dynamics of a small set of variables, our approximations may be very biased. To balance these concerns, we build the sets $S_{t,i}$ using k-means clustering of our data on a feature space Φ consisting of time-delays of $U_{10,60}$:

$$\Phi(\mathbf{X}(t)) = [U_{10,60}(\mathbf{X}(t)), U_{10,60}(\mathbf{X}(t-1)), \ldots, U_{10,60}(\mathbf{X}(t-\delta))] \tag{7.2}$$

where $\delta = 20$ days is the number of retained time-delays, which can range from 15 to 25 with only minor effects on the results. We have also experimented with richer feature spaces including EOFs of geopotential height, but found these unnecessary. A growing body of theoretical [Takens, 1981, Kamb et al., 2020] and empirical [Broomhead and King, 1986, Giannakis and Majda, 2012, Brunton et al., 2017, Thiede et al., 2019, Strahan et al., 2021] evidence supports the use of time-delay coordinates as reliable features for related methods. The k-means clustering is carried out

using `scikit-learn` [Pedregosa et al., 2011] with $k = M_t$ on the collection of hindcast trajectories that were running between days t and $t+1$. The number of clusters is set to $M_t = 170$ or the number of data points available on day t, whichever is smaller.

We use *transition path theory* (TPT) as a framework for combining several key forecast functions (both forward and backward-in-time) to compute the steady-state statistics of rare transition events [Vanden-Eijnden, 2014, Finkel et al., 2020, Miron et al., 2021, Finkel et al., 2021a]. TPT is most often applied in molecular dynamics applications [Noé et al., 2009, Meng et al., 2016, Strahan et al., 2021, **?**] and is typically formulated in a time-homogeneous setting. The different timescales of climate applications, in particular the seasonal cycle, demand incorporating time-dependence explicitly, which we do in a manner similar to [Helfmann et al., 2020]. Supporting Information provides more detail on TPT. All of the key forecast functions can be estimated directly using the transition matrix described above. In fact, the forecast functions each solve an infinite dimensional Feynman-Kac equation involving the transition operator of the process [Strahan et al., 2021], and our partitioning of space into clusters corresponds to a basis expansion approach to solving those equations. This more general perspective motivates the *dynamical Galerkin approximation* (DGA) method of which our MSM approach is a special case [Thiede et al., 2019, Strahan et al., 2021, Finkel et al., 2021b,a]. MSMs are similar in spirit to analogue forecasting [van den Dool, 1989], which is enjoying a renaissance with novel data-driven techniques, especially for characterizing extreme weather [Chattopadhyay et al., 2020, Lucente et al., 2021b]. Formally, the transition operator encoded by the matrix in (7.1) is related to linear inverse models [LIMs; Penland and Sardeshmukh, 1995], which have also been used to predict atmospheric rivers at the subseasonal timescale [Tseng et al., 2021]. Both MSMs and LIMs are finite-dimensional approximations of the Koopman operator [Mezić, 2013, Mezić, 2005, Klus et al., 2018]. For TPT analysis, however, an MSM is more convenient, which is explained in Supporting Information.

Detailed comparison in the following section reveals that the approach sketched here is statistically consistent with the direct method of sample-averaging over historical SSW events from

reanalysis. However, the MSM approach provides more precise estimates for the rarest of events like the SSW of January 2009.

7.4 Results

7.4.1 Rate estimates

Fig. 7.2 shows rate estimates computed from the S2S dataset using the MSM-based approach outlined in the previous section, as well as from several reanalysis datasets using the direct counting method. Each circle indicates a point estimate using all the data from a given source and timespan. In the case of S2S (red) the circle shows the mean rate from five independent trials with different seeds for k-means clustering. The thick and thin vertical lines represent the 50% and 90% confidence intervals respectively, estimated from the pivotal bootstrap procedure [Wasserman, 2004]. We treat a full winter as a single unit of data for resampling, and we resample 40 times with replacement to estimate error bars. Any error bar that reaches the bottom edge of the logarithmic plot is understood to include zero.

Different reanalysis datasets have different strengths for comparison with S2S. The most direct comes from ERA-5 (1996-2016)—meaning winter 1996/7-winter 2016/7, inclusive, the same time period as the S2S data—shown in orange. The S2S integrations from CY43R1 were initialized from ERA-I rather than ERA-5, but $U_{10,60}$ is virtually identical in both products (see Fig. S1). ERA-5 (1996-2016) is an appropriate baseline to compare with S2S, as both make use of the same observations. The key difference is that our MSM makes use of all the S2S hindcast integrations as well. Across the range of $U^{(\mathrm{th})}_{10,60}$, the S2S rate is less than or equal to the ERA-5 (1996-2016) rate. However, this does not mean the two results are statistically inconsistent: 21 flips of a fair coin can yield a range of outcomes, with 6-8 heads (combined probability 0.18) occurring slightly more often than either of the two most-likely outcomes of 10 or 11 heads (probability 0.17 each). The orange error bars in Fig. 7.2 show the 50% and 95% confidence intervals of $(K/21)$, where

K is a binomial random variable with $n = 21$ and $p =$(the corresponding S2S estimate). In other words, we treat the S2S estimate as a null hypothesis and consider the real world as a sequence of independent draws from a probability distribution. For $U^{(\mathrm{th})}_{10,60} = -15$ m/s and above, the 21-year ERA-5 (1996-2016) point estimates are well within the 50% S2S confidence intervals, i.e., the interquartile range of $K/21$. For the more extreme events, the two estimates remain consistent with 95%-level statistical significance, but ERA-5 (1996-2016) systematically indicates a higher frequency of extreme events in this 21-year timespan.

What climatology, then, is our MSM rate estimate inferring? Strictly speaking, it is a mixture between (i) the portion of phase space covered by 1996-2016 observations, and (ii) the *model climatology implied by the IFS*, including its stochastic parameterizations. Several recent studies have performed the same task of filling out a sparse climate distribution using models [Horan and Reichler, 2017, Kelder et al., 2020], but with uninterrupted long runs of a global climate model. Our technique is novel in using short runs of a weather model instead.

Does the IFS climatology then correspond to anything in the real world? We can answer this by comparing to longer reanalyses, such as the 70-year ERA-5 (1950-2019) shown in gray in Fig. 7.2. Results are encouraging: ERA-5 (1950-2019) agrees with S2S in estimating a rate systematically lower than ERA-5 (1996-2016), in other words suggesting this was an historically anomalous period. This tentative trend has been documented, and may explain some increasing cold-weather outbreaks despite an overall warming planet [Kretschmer et al., 2018b, Garfinkel et al., 2017]. Some studies indicate multi-decadal-scale variations in SSW frequency due to the quasi-biennial oscillation (QBO), El Niño southern oscillation (ENSO), Atlantic meridional overturning circulation, and other features of the coupled atmosphere-ocean system [Reichler et al., 2012, Dimdore-Miles et al., 2021]. Hence, the recent barrage of SSWs may represent a temporary internal fluctuation rather than a secular trend. The consistency of S2S with ERA-5 on more common events, and the improvement of consistency with record length, is an encouraging signal that the MSM estimate is extracting a meaningful statistic from the S2S dataset. This lends confidence in the S2S estimate

as we reach farther into the negative $U_{10,60}$ tail where reanalysis data are too sparse to give any rate estimate.

Longer reanalysis is helpful to generate better statistics. For this, we incorporate one more relevant product, ERA-20C, which spans the longer period 1900-2007, but assimilates only surface measurements as opposed to satellite data [Poli et al., 2016]. With these deliberate limitations, ERA-20C likely suffers higher bias than ERA-5 or ERA-I, but it enjoys lower variance due to its longer timespan. In their period of overlap (1950-2007, see Fig. S1), they roughly agree on the SSW rates with moderate thresholds of $U_{10,60}^{(\mathrm{th})} = 0$ and $U_{10,60}^{(\mathrm{th})} = -5$ m/s, but otherwise ERA-20C appears biased toward fewer SSW events. Nonetheless, ERA-20C is our best estimate for the SSW rate over the full 20th century.

In the upper range of thresholds from 0 m/s to -15 m/s, all datasets suggest a linear relationship between $U_{10,60}^{(\mathrm{th})}$ and rate. In the lower range from -20 m/s to -35 m/s, reanalysis becomes too noisy to discern clear trends, as these estimates rely on just a few exceptional events like January 2009 (Fig. 7.1). However, S2S clearly suggests an exponential trend with an e-folding scale of $\sim$4 m/s. Events become tenfold rarer as the threshold is lowered by 10 m/s. These results depend somewhat on parameter choices (see Supporting Information), but are robust to variations in the delay time δ from 15 to 25 days.

7.4.2 Probability current

To explain the rate calculation, we briefly expand on the TPT framework, whose real strength is to not only provide numerical rates, but to decompose them into a sum over possible pathways into the rare event. The spread of pathways is encoded by the *probability current*, a vector field $\mathbf{J}_{AB}(t, \mathbf{x})$ over state space that indicates the average tendency of the system $\mathbf{X}(t)$ as it passes through state $\mathbf{x}$, conditioned on an SSW occurring. The subscript AB refers to two distinguished sets A and B in

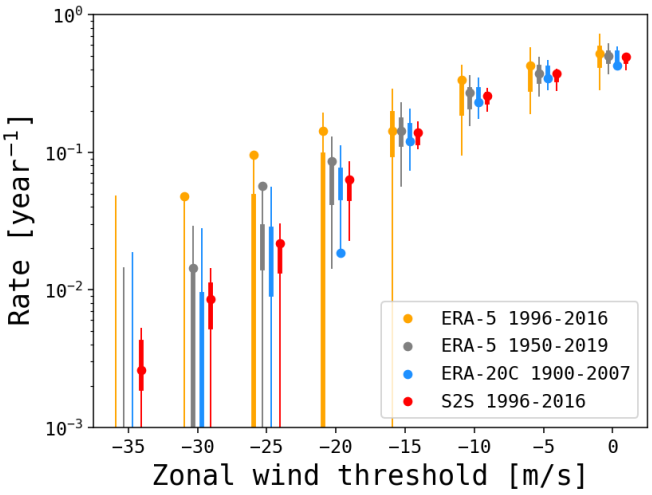


Figure 7.2: **Rate estimates derived from S2S and reanalysis**. Circles show point estimates of SSW rate according to each data source. S2S error bars show the 50% and 95% confidence intervals in thick and thin lines respectively, based on 40 bootstrap resamplings. Reanalysis error bars show the middle 50- and 95-percentile envelope of K/n, where K is a binomial random variable with p given by the corresponding S2S estimate, and n is the number of years in the reanalysis dataset. When an error bar overlaps with a reanalysis rate, the S2S rate is statistically consistent at the 95% confidence level.

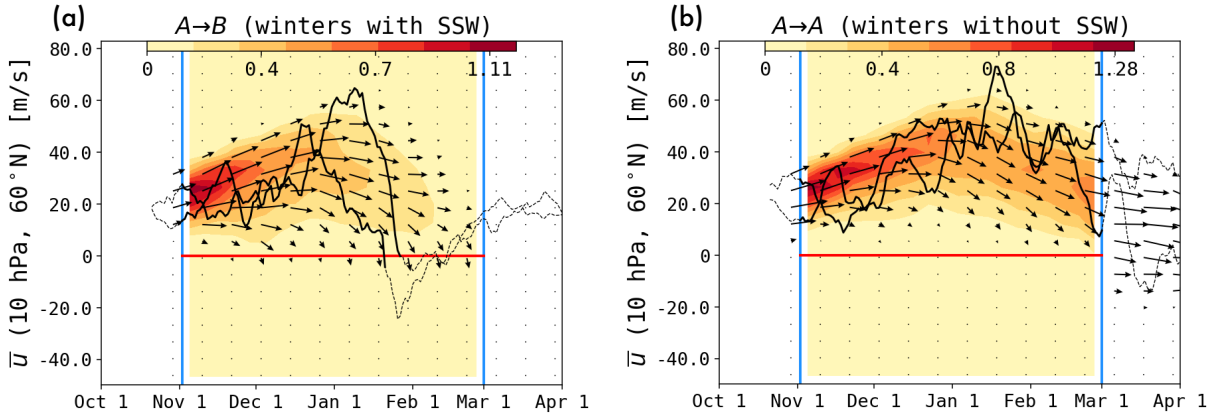


Figure 7.3: **Probability currents**. The probability currents $\mathbf{J}_{AB}$ (tendency of pre-SSW evolution) and $\mathbf{J}_{AA}$ (tendency of non-SSW evolution) overlaid on the corresponding time-dependent probability densities π_{AB} and π_{AA}. Horizontal red line shows the boundary of B. The flux density of $\mathbf{J}_{AB}$ across ∂B gives the seasonal distribution shown in Fig. 7.4.

space-time,

$$A = \{(t, \mathbf{x}) : t < \text{Nov. 1 or } t > \text{Feb. 28}\} \tag{7.3}$$

$$B = \{(t, \mathbf{x}) : \text{Nov. 1} \leq t \leq \text{Feb. 28, and } U_{10,60}(\mathbf{x}) < U_{10,60}^{(\text{th})}\}. \tag{7.4}$$

An SSW event can now be defined adhering to the TPT formalism [Vanden-Eijnden, 2014] as a passage of $\mathbf{X}(t)$ from A (the pre-winter part) to B, *before* returning to A (the post-winter part). Just as the symbol AB encodes an SSW, the symbol AA encodes a winter without SSW, in which the system departs A in the fall and re-enters A in the spring without ever hitting B. A second vector field, $\mathbf{J}_{AA}(t, \mathbf{x})$, indicates the average tendency of the system during non-SSW winters. Both currents, $\mathbf{J}_{AB}$ and $\mathbf{J}_{AA}$, are computable in discretized forms from the transition matrices $P_{t,t+1}(i, j)$ following Metzner et al. [2009]. Consistent projections of these reactive currents from the full delay-embedded space down to U_{10} can be defined following Strahan et al. [2021] and are shown in Fig. 7.3. Supporting Information details the visualization procedure. The streamlines of $\mathbf{J}_{AB}$ lead directly to the boundary ∂B of B, whereas the streamlines of $\mathbf{J}_{AA}$ avoid this boundary and lead instead to ∂A (the right edge of the plot). Background shading indicates the corresponding time-dependent probability densities $\pi_{AB}(t, \mathbf{x})$ (a) and $\pi_{AA}(t, \mathbf{x})$ (b), defined as the density of all system trajectories $\mathbf{X}(t)$ destined for an SSW event or a non-SSW winter, respectively. Two samples from each ensemble are superimposed: 1962-1963 and 2005-2006 as representative SSW winters, and 1966-1967 and 2004-2005 as representative non-SSW winters. The SSW trajectories drop out of the ensemble when they first enter B. The total probability $\int \pi_{AB}(t, \mathbf{x})\, d\mathbf{x}$ becomes steadily smaller as time progresses, because it is an average over fewer and fewer events. In fact, one can show (see Supporting Information) that the $\pi_{AB}(t, \mathbf{x})$ is identical to the t-component of $\mathbf{J}_{AB}(t, \mathbf{x})$, which roughly quantifies how many SSW-bound trajectories are temporarily maintaining steady—or even increasing—$U_{10,60}$ before the upcoming event. Note that the individual trajectories do not track along streamlines of the current: only their average evolution does. For example, the individual sample trajectories plummet toward B passing through *flat* $\mathbf{J}_{AB}$ arrows, which account for the

other SSW-bound trajectories that still persist at the same time of year.

These vector fields have concrete physical meaning: the field lines of $\mathbf{J}_{AB}$ poke through ∂B with a time-dependent flux density that integrates to the total rate, as seen in the equation

$$\int_{\text{Nov. 1}}^{\text{Feb. 28}} \mathbf{J}_{AB} \cdot \mathbf{n}\, dt = \frac{\text{\# SSW events}}{\text{Year}} \tag{7.5}$$

where $\mathbf{n}$ is the unit vector in state space pointing directly into B; in our case, $\mathbf{n} = -\nabla U_{10,60}(\mathbf{x}) / \|\nabla U_{10,60}(\mathbf{x})\|$. Moreover, SSW events can occur at different times during the winter, and the contribution from each time interval is equal to the corresponding partial flux integral. For example,

$$\int_{\text{Dec. 1}}^{\text{Dec. 31}} \mathbf{J}_{AB} \cdot \mathbf{n}\, dt = \frac{\text{\# Dec. SSW events}}{\text{Year}} \tag{7.6}$$

This relation allows us to examine more refined details of SSW climatology: the seasonal distribution of events.

7.4.3 Seasonal distribution

Past studies have found that seasonal differences are associated with dynamical differences in SSW events. For example, "Canadian warmings" shift the Aleutian high and occur earlier in the winter [Butler et al., 2015]. Categorizing SSWs by their seasonality may reveal preferred timings that indicate when and why the polar vortex is most vulnerable [Horan and Reichler, 2017]. Unfortunately, month-by-month rate estimates from reanalysis are noisier than full-winter rate estimates, as splitting data into finer categories makes the events even sparser. We can again use S2S data to enhance precision by recruiting the larger database of partial trajectories. Fig. 7.4 shows seasonal distributions at two thresholds, $U_{10,60}^{(\text{th})} = -15$ m/s (left) and $U_{10,60}^{(\text{th})} = 0$ m/s (right), according to the same four datasets used in Fig. 7.2. Each panel displays the distribution at two resolutions: monthly (hashed) and sub-monthly (solid, and rounded to the nearest day), both according to the same dataset and with the same total integrals equal to the rate estimate. To express the seasonal

cycle as a probability distribution, we normalize so that all histograms in Fig. 7.4 integrate to one, with units of probability per day. The two columns have different vertical scales to see features more readily.

Several features are noteworthy. For the conventional SSW, $U_{10,60}^{(\mathrm{th})} = 0$ m/s, the reanalysis histograms all exhibit a common seasonal trend of steadily rising SSW frequency from November to January and a small decline in February. The coarse S2S histogram disagrees, with a slight increase in February. Both trends are consistent with prior studies of seasonality at monthly resolution [e.g., Charlton and Polvani, 2007]. At a finer resolution of ~ 10 days, however, the S2S histogram reveals a frequency peak in late January/early February and declines thereafter. The January/February peak is documented in the literature, e.g., by [Horan and Reichler, 2017], who diagnosed the peak as a balance between two time-varying signals: the background strength of the polar vortex, and the vertical flux of wave activity capable of disturbing the vortex. Additionally, the 10-day resolved S2S histogram reveals a smaller December peak, which is absent from ERA-5 reanalysis and at best noisily present in ERA-20C. The bimodal structure seen in S2S has also been found tentatively in prior studies with both reanalysis and models [e.g., Horan and Reichler, 2017, Ayarzagüena et al., 2019]. We speculate that the early peak represents Canadian warmings [Meriwether and Gerrard, 2004], which our result suggests may deserve a more decisive classification.

All three reanalysis-based estimates of SSW distributions have a low signal-to-noise ratio, exemplified by the intermittent frequency spikes. The hint of a third peak at the end of February is clearer in reanalysis than S2S, and might be the beginning of the "final warmings", but its significance is questionable because of the histograms' general noisiness. This is even more of a problem at the more extreme threshold $U_{10,60}^{(\mathrm{th})} = -15$ m/s, where the ERA-5 (1996-2016) has degenerated to two isolated spikes while S2S retains a smoother shape, with little sign of bimodality. Early December still supports a nonzero rate of extreme SSW events, but is not a highly favorable time for them. This suggests that whatever distinct SSW type accounts for the December peak at $U_{10,60}^{(\mathrm{th})} = 0$ m/s is limited to weaker events. These results are sumathbfJect to all the caveats of our data-driven

procedure (see Supporting Information), but merit further investigation with numerical models.

7.5 Discussion

By comparing S2S results with reanalysis, we are measuring the composition of three separate error sources: (i) forecast model error, (ii) non-stationarity of the climate *with respect to SSW events* over the reanalysis period, and (iii) numerical errors in the MSM approach, both statistical (from the finite sample size) and systematic (from the projection of forecast functions onto a finite basis). We briefly address each error source in turn.

The S2S trajectories were realized only in simulation, not in the physical world. Accordingly, our S2S estimates apply strictly to the climatology of the 2017 IFS, a statistical ensemble that could be concretely realized by running the model uninterrupted for millennia, with external climatic parameters sampled from their variability in the short 21-year time window of 1996-2016. Such long, equilibrated simulations have been performed with coarser models by, e.g., Kelder et al. [2020] to assess UK flood risk (the so-called "UNSEEN" method), and by Horan and Reichler [2017] to assess SSW frequencies, but this is not practical given the constraints and mission of the ECMWF IFS. Given these constraints, we have assembled our best approximation using S2S trajectories. Indeed, the S2S dataset is an ensemble of opportunity for us. It was created to compare the skill of different forecast systems on S2S timescales, not at all for the purpose of establishing a climatology of SSWs.

The IFS model has proven outstanding in its medium-range forecast skill [Vitart, 2014, Kim et al., 2014, Vitart and Robertson, 2018]. However, there is a caveat that the IFS was designed for short forecasts, and it is not clear how it would behave if allowed to run for hundreds of years as a climate model, which requires careful attention to the boundary condition and conservation issues. Even if the climate were to remain stationary with its 1996-2016 parameters, numerical and model errors would inject some bias into the equilibrated simulation. Repeatedly initializing S2S forecasts with reanalysis ensures a realistic background climatology, and allows us to rely on the

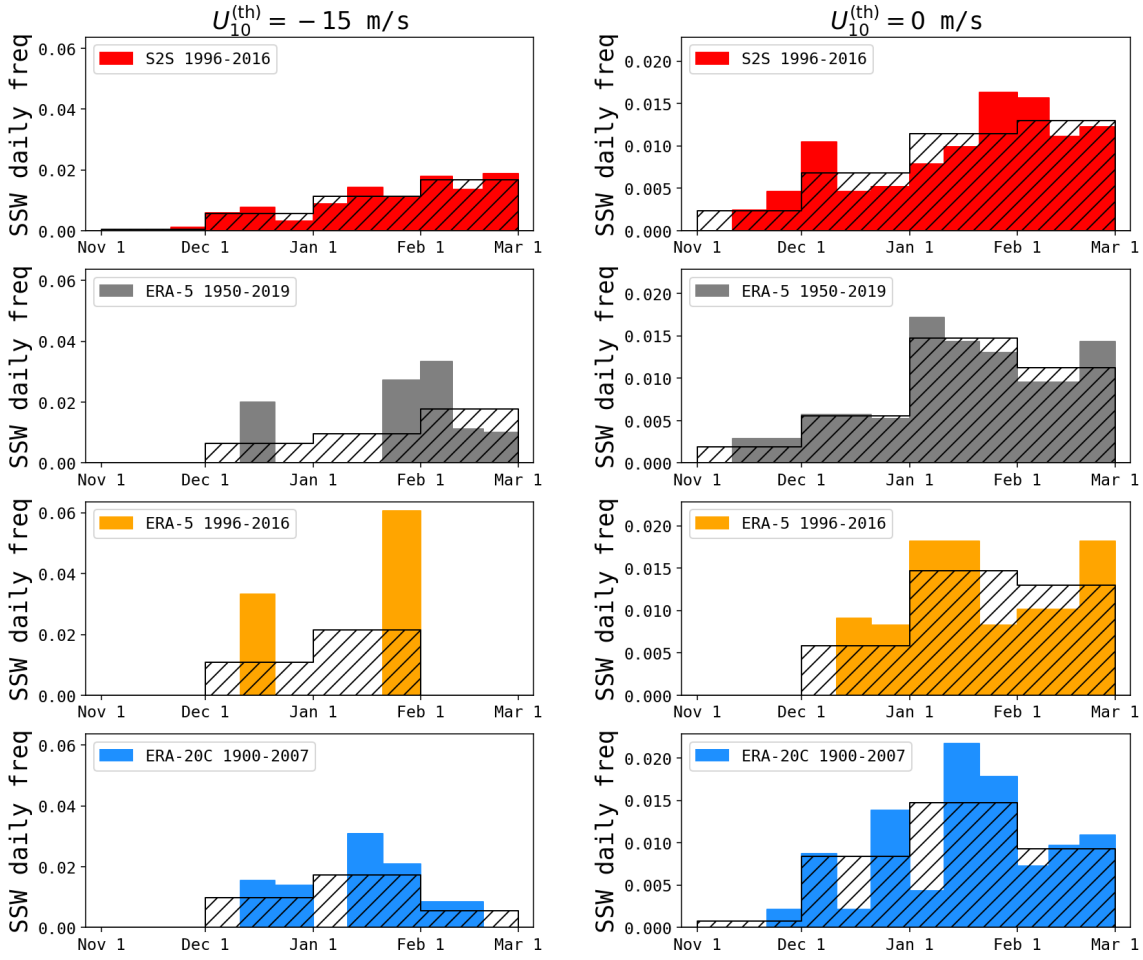


Figure 7.4: **Seasonal distributions of SSW events**. Left and right columns show statistics with threshold $U_{10,60}^{(\mathrm{th})} = -15$ m/s and $U_{10,60}^{(\mathrm{th})} = 0$ m/s, respectively, and each row uses a different data source. Each panel has a hashed histogram at monthly resolution, along with a solid-colored histogram at $\frac{1}{3}$-monthly resolution (rounded to the nearest day) with an equal area equal to unity. The vertical unit is SSW events per day. The vertical scales are shared within within each column, but different between columns in order to make the shape of the histogram at $U_{10,60}^{(\mathrm{th})} = -15$ m/s more easily visible.

IFS strictly for the short-term integrations that it was designed for. Our method may be used as a diagnostic tool to compare different models against each other, with specific attention paid to their rare event rates. A useful extension of this work would be to repeat the analysis on multiple data streams from all 11 forecasting centers worldwide that contribute to the S2S project, providing a new rare event-oriented intercomparison metric.

The rate we estimate with an MSM is the SSW rate of the climate system frozen in its 1996-2016 state. Comparing with a 70-year reanalysis dataset (ERA-5 1950-2019) measures the departure of the 21-year SSW climatology from the 70-year climatology, and likewise for the 108-year reanalyis ERA-20C (1900-2007). Of course, the 21-year SSW climatology itself may be estimated directly from reanalysis, but we have demonstrated in Fig. 7.2 that S2S gives more precise estimates that are different from the observations, but not at a statistically significant level. Our results indicate that according to the 2017 IFS, 1996-2021 was more similar to 1950-2019 than direct counting of SSW events would suggest, which could of course mean that the IFS was missing some key climatological variable during that period [Dimdore-Miles et al., 2021]. There is insufficient evidence on the anthropogenic influence on SSW to reject the hypothesis of stationarity [Ayarzagüena et al., 2020]. By running our method on different historical periods, we might discern a more decisive signal of secular changes than would be available from raw data.

Error source (iii) is the most open to scrutiny and improvement. In a sequence of preceding papers [Finkel et al., 2021b,a], we have benchmarked the performance of DGA (with a similar MSM basis set) on a highly idealized SSW model due to Holton and Mass [1976]. DGA was originally developed in molecular dynamics to study protein folding and has been benchmarked on a diverse set of low- and high-dimensional dynamical systems [Thiede et al., 2019, Strahan et al., 2021, Antoszewski et al., 2021]. Our parameter choices here, detailed further in Supporting Information, are informed by prior experience. Nevertheless, large-scale atmospheric models are a mostly-unexplored frontier for this class of methods. In this study, we have worked with static datasets produced by some of the most advanced models in the world; however, an even more

powerful procedure would be to generate data adaptively.

Our method exceeds what is possible directly from reanalysis, but we are not yet fully "liberated" from observations: every S2S trajectory is initialized near reanalysis, and it only has 47 days to explore state space before terminating. This fundamentally limits how far we can explore the tail of the SSW distribution. In other words, the real climate system sets the "sampling measure" which is a flexible but important component in the DGA pipeline [Thiede et al., 2019, Strahan et al., 2021, Finkel et al., 2021b]. On the other hand, with an executable model, we could initialize secondary and tertiary generations of short trajectories to push into more negative $U_{10,60}$ territory and maintain statistical power for increasingly extreme SSW events. This is the essence of many rare-event sampling algorithms, such as those reviewed in Bouchet et al. [2019b] and Sapsis [2021]. For example, a splitting large-deviation algorithm was used in Ragone et al. [2018] to sample extreme European heat waves and estimate their return times. Quantile diffusion Monte Carlo was used in Webber et al. [2019] to simulate intense hurricanes, and in [Abbot et al., 2021] to estimate the probability of extreme orbital variations of Mercury. Many other rare event sampling studies have been performed in fluid dynamics and other complex systems [Simonnet et al., 2021a, Hoffman et al., 2006, Weare, 2009, Vanden-Eijnden and Weare, 2013, Bouchet et al., 2014, Chen et al., 2014, Farazmand and Sapsis, 2017, Dematteis et al., 2018, Mohamad and Sapsis, 2018]. A natural extension of these various techniques would combine elements of active rare event sampling with the DGA method. Early developments of such a coupling procedure are presented in Lucente et al. [2021b].

7.6 Conclusion

Extreme weather events present a fundamental challenge to Earth system modeling. Many years of simulations are needed to generate sufficiently many extreme events to reduce statistical error, but high-fidelity models are needed to simulate those events accurately. Conventionally, no single model can provide both, simply because of computational costs. Here, we have demonstrated

an alternative approach that leverages *ensembles* of *short*, high-fidelity weather model forecasts to calculate extreme weather statistics, with specific application to sudden stratospheric warming (SSW). By exploiting the huge database of forecasts stored in the subseasonal-to-seasonal (S2S) database [Vitart et al., 2017], we have obtained plausible estimates of the rate and seasonal distribution of SSW events that are (i) more precise, and (ii) more robust in distribution tails, than reanalysis data.

Our method uses data to estimate the dynamics on a subspace relevant for SSW, namely the polar vortex strength as measured by zonal-mean zonal wind. This single observable, augmented by time-delay embedding, gives a simple set of coordinates sufficient to estimate rate and seasonal distributions. Our demonstration opens the door to address many other data-limited questions of basic physical interest. For example, how important are vortex preconditioning and upward wave activity as triggers of SSW? [Charlton and Polvani, 2007, Albers and Birner, 2014]. Do split-type and displacement-type events have fundamentally different mechanisms and/or different downstream effects? [Matthewman and Esler, 2011, Esler and Matthewman, 2011, O'Callaghan et al., 2014, Maycock and Hitchcock, 2015]. Will climate change affect the frequency of SSW, perhaps through arctic amplification? [Charlton-Perez et al., 2008, Garfinkel et al., 2017, Kretschmer et al., 2018b]. How do other slow climatic variables, such as ENSO, the QBO, and the Aleutian Low affect SSW propensity? [Dimdore-Miles et al., 2021]. These questions have been addressed in a number of coarse-resolution climate modeling studies, but high-resolution weather forecast data is an untapped source of potential for sharpening the answers. Our method offers a way forward, and is highly customizable to include physical features tailored for the problem at hand.

Another potential application of our methods is catastrophe modeling under climate change. Tropical cyclones pose a pressing problem for coastal communities, and have motivated several hybrid dynamical/statistical downscaling methods to project risk into the future under various climate change scenarios [Camargo et al., 2014, Lee et al., 2018, Jing and Lin, 2020, Sobel et al., 2021]. Extreme precipitation of many varieties threatens cities and agriculture and is expected to

change significantly with global warming [e.g., O'Gorman, 2012, Pfahl et al., 2017]. Model resolution, again, is the limiting factor [Laflamme et al., 2016, O'Brien et al., 2016, He et al., 2019]. Enlisting short weather forecasts, as we have done, may help identify precursors and drivers of changing frequency with unprecedented detail.

7.7 Acknowledgments

We extend special thanks to Andrew Charlton-Perez, who suggested the S2S dataset as a case study for the methodology, and Simon Lee, who helped familiarize us with the data. We thank Amy Butler for guidance on using the ERA-20C dataset. Our collaborators at the University of Chicago, including Aaron Dinner, John Strahan, and Chatipat Lorpaiboon, offered helpful methodological advice. Computations for this project were performed on the Greene cluster at New York University.

J.F. is supported by the U.S. DOE, Office of Science, Office of Advanced Scientific Computing Research, Department of Energy Computational Science Graduate Fellowship under Award Number DE-SC0019323. E.P.G. acknowledges support from the US National Science Foundation through award OAC-2004572. J.W. acknowledges support from the National Science Foundation through award DMS-2054306 and from the Advanced Scientific Computing Research Program within the DOE Office of Science through award DE-SC0020427.

7.8 Supporting information

Our work relies completely on publicly available datasets of reanalysis and hindcasts, which we describe in the subsequent section. We then lay out the numerical procedure to compute rates and seasonal distributions using transition path theory (TPT). We then present the formulas used to display results in the main text. Finally, we document the method used to select parameters.

Dataset description

We use four different datasets for this study.

- S2S: perturbed reforecast (hindcast) ensembles from the 2017 model version of the ECMWF IFS. We include all trajectories launched between October 1 and April 30 every year from 1996/97 through 2016/17. We downloaded geopotential height and zonal wind fields, sampled daily at time 00:00:00, at pressure levels 10, 50, 100, 200, 300, 500, 700, 800, 925, and 1000 hPa, and with horizontal resolution of $3^\circ \times 3^\circ$ latitude $\times$ longitude. We experimented with many feature spaces, and found that simply zonal-mean zonal wind at 10 hPa and 60°N (abbreviated $U_{10,60}$) was sufficient to capture robust rate and seasonality statistics.
- ERA-Interim: same fields and resolution as S2S, but between 1979/80 and 2017/18.
- ERA-20C: same fields and resolution as S2S, but between 1900/01 and 2007/08.
- ERA-5: only zonal wind at 10 hPa, in order to compare rates.

The first three datasets were downloaded from the ECMWF data portal `https://ecmwf.int`, and ERA-5 was downloaded from the Copernicus Data Store `https://cds.climate.copernicus.eu/`.

Each dataset spans a different period and gives somewhat different SSW rates, as shown in Fig. 7.2 of the main text. How much of that difference come from the non-overlapping timespans, and how much comes from the reconstruction methodology? Fig. 7.5 compares SSW rates in between

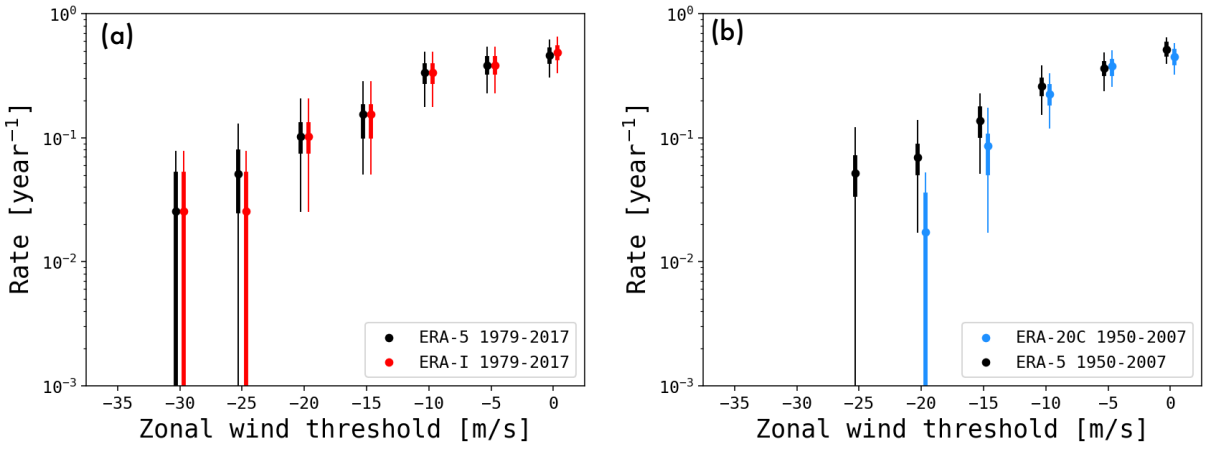


Figure 7.5: **Comparison of reanalyses on SSW frequency**.

pairs of reanalyses during their period of overlap. Circles are point estimates equal to the fraction of winters with SSW. Thick and thin vertical lines span the middle 50- and 90-percentile ranges according to the pivotal bootstrap procedure Wasserman [2004] with 40 resamplings. Panel (a) compares ERA-I to ERA-5 for 1996-2016, the same period as in S2S. The two are almost identical, save for slight differences at $U^{(\mathrm{th})}_{10,60} = 0$ m/s and $U^{(\mathrm{th})}_{10,60} = -25$ m/s. We therefore use ERA-5 (1996-2016) in place of ERA-I for the following comparisons in the main text. Panel (b) compares ERA-5 and ERA-20C on their period of overlap (1950-2007), revealing decent agreement for more common events but a low-SSW bias in ERA-20C at more extreme events. For this reason, ERA-20C should be interpreted cautiously, not as a most-likely estimate but as a lower bound. It is therefore a positive consistency check that in Fig. 7.2 of the main text, every threshold where ERA-20C does give a nonzero rate has a much higher S2S rate.

Numerical procedure

Here we present the computational procedure of Markov state modeling, and how we use it to calculate rates and seasonality distributions. As stated in the main text, an SSW event is a transition

of the atmospheric state vector, $\mathbf{X}(t) \in \mathbb{R}^d$, between two sets in space-time,

$$A = \{(t, \mathbf{x}) : t < t_1 := \text{Nov. 1 or } t > t_2 := \text{Feb. 28}\} \tag{7.7}$$

$$B = \{(t, \mathbf{x}) : t_1 \leq t \leq t_2 \text{ and } U_{10,60}(\mathbf{x}) < U_{10,60}^{(\text{th})}\}. \tag{7.8}$$

We could continue to set up the problem with $\mathbf{X}$ as a continuous variable, but for practical purposes we immediately discretize the process. Our dataset consists of a large collection of short trajectories

$$\left\{ \left(t_n(s), \mathbf{X}_n(t_n(s))\right) : s = 0, 1, \ldots, 46; \;\; n = 1, \ldots, N \right\} \tag{7.9}$$

where s represents the elapsed time since the initialization date of the nth trajectory, and $t_n(s) = t_n(0) + s$ is the calendar day of the nth trajectory after s days of integration. Each $\mathbf{X}_n(t_n(s))$ should be thought of as a partial realization of the stochastic process $\mathbf{X}(t)$ in the time interval $t_n(0) \leq t \leq t_n(46)$.

With the dataset in hand, we execute the following steps.

1. **Cluster the data**. For every calendar day t we apply k-means clustering to only the snapshots $\mathbf{X}_n(t_n(s))$ such that $t_n(s) = t$, i.e., the trajectories that are running on day t. We cluster using only the feature space of time-delays of $U_{10,60}$, after subtracting the seasonal mean and dividing by the seasonal standard deviation. (The seasonal statistics for a day t are found by aggregating data from days $t-4, \ldots, t+4$.) We set the number of clusters to $M_t = 170$ by default, but if fewer than 170 trajectories are live on that day we reduce M_t to that smaller number. The outcome of clustering is, for each calendar day t, a disjoint collection of sets $S_{t,1}, \ldots, S_{t,M_t}$ and a mapping from snapshots to clusters. Formally, for $s \in \{0, \ldots, 46\}$ and

$n \in \{1,\ldots,N\}$, we define the cluster assignment function

$$\mathbf{Z}_n(t_n(s)) = \left[\text{the cluster on calendar day } t_n(s) \text{ that contains } \mathbf{X}_n(t_n(s))\right] \in \{1,\ldots,M_{t_n(s)}\} \tag{7.10}$$

$$\implies \mathbf{X}_n(t_n(s)) \in S_{t_n(s),\mathbf{Z}_n(s)} \tag{7.11}$$

We can now consider each $\mathbf{Z}_n(t_n(s))$ as partial realizations of a fully discrete process $\mathbf{Z}(t)$ in the time interval $t_n(0) \le t \le t_n(46)$. Furthermore, A and B are transformed to index sets:

$$\overline{A} = \{(t,z) : t < t_1 \text{ or } t > t_2\} \tag{7.12}$$

$$\overline{B} = \{(t,z) : t_1 \le t \le t_2 \text{ and } U_{10,60}(z) \le U_{10,60}^{(\text{th})}\} \tag{7.13}$$

In the last line, $U_{10,60}(z)$ is understood to be the value of $U_{10,60}$ at the centroid of cluster z. Here we explicitly ignore "leakage", in which some data points $\mathbf{X}$ with $U_{10,60}(\mathbf{X}) > U_{10,60}^{(\text{th})}$ land in a cluster z with $U_{10,60}(z) \le U_{10,60}^{(\text{th})}$. This way, we need cluster the data only once at the outset and can subsequently calculate quantities of interest for every threshold using the same clustering. A more rigorous procedure is to separately cluster points inside A and B.

2. **Construct the Markov state model.** We then estimate $T-1$ probability transition matrices $P_{t,t+1}$ with shape $M_t \times M_{t+1}$ by counting trajectory transitions between the sets at time t and $t+1$. Explicitly, we compute a count matrix

$$C_{t,t+1}(i,j) = \sum_{n=1}^{N} \sum_{s=0}^{46-1} \mathbb{1}\{t_n(s) = t\}\mathbb{1}\{\mathbf{Z}_n(t_n(s)) = i\}\mathbb{1}\{\mathbf{Z}_n(t_n(s+1)) = j\} \tag{7.14}$$
$$\text{for} \quad i = 1,\ldots,M_t \text{ and } j = 1,\ldots,M_{t+1}$$

For the calculations to follow, every row and column of every $C_{t,t+1}$ has at least one entry. To enforce this condition, we artificially insert out-going transitions from any "dead-end" clus-

ter i (with $C_{t,t+1}(i,j) = 0$ for all i) to its four nearest neighbors, with uniform weights. We then do the same for columns. After this small correction, the transition matrix is estimated as

$$P_{t,t+1}(i,j) = \frac{C_{t,t+1}(i,j)}{\sum_{j'=1}^{M_{t+1}} C_{t,t+1}(i,j')} \tag{7.15}$$

3. **Estimate the three core ingredients of a rate calculation**. The TPT framework expresses rates using the following three functions of space-time. In the discretized state space, they will be finite-dimensional vectors, one entry for each cluster, and we will be able to compute them recursively.

   (a) The probability density π is the climatology of the system on day t, but estimated from S2S data rather than reanalysis (as in Fig. 7.1):

   $$\pi_t(\mathbf{z}) = \mathbb{P}\{\mathbf{Z}(t) = \mathbf{z}\} \tag{7.16}$$

   In our finite-time setting, π_t depends on some initial condition π_0, which we simply take as the empirical distribution of S2S trajectories which were live on the first day of available data. Explicitly,

   $$\pi_0(\mathbf{z}) = \frac{\sum_{n=1}^{N} \mathbb{1}\{t_n(0) = 0\}\mathbb{1}\{\mathbf{Z}_n(t_n(0)) = \mathbf{z}\}}{\sum_{n=1}^{N} \mathbb{1}\{t_n(0) = 0\}} \tag{7.17}$$

   To propagate π_t forward in time from $t = 0$, we use the following simple recursion relation. The probability of occupying a given cluster j at time $t+1$ can be found by summing transitions into j from time t:

   $$\pi_{t+1}(j) = \sum_{i=1}^{M_t} \pi_t(i) P_{t,t+1}(i,j) \tag{7.18}$$

This amounts to right-multiplying the vector $\pi_t \in \mathbb{R}^{M_t}$ by the matrix $P_{t,t+1} \in \mathbb{R}^{M_t \times M_{t+1}}$. Thus, in $T-1$ matrix multiplications, we obtain π_t for every timestep.

(b) The forward committor q^+ is the probability of an SSW before the end of winter, given some initial condition:

$$q_t^+(\mathbf{z}) = \begin{cases} \mathbb{P}\{\mathbf{Z} \text{ next reaches } B \text{ before } A | \mathbf{Z}(t) = \mathbf{z}\} & (t,\mathbf{z}) \notin \overline{A} \cup \overline{B} \\ 0 & (t,\mathbf{z}) \in \overline{A} \\ 1 & (t,\mathbf{z}) \in \overline{B} \end{cases} \tag{7.19}$$

We can find the forward committor at time t ("today") recursively by writing it as a sum over possibilities at time $t+1$ ("tomorrow"). In other words, we decompose the pathway $\mathbf{z}(t) \to \overline{B}$ into a sum of $\mathbf{z}(t) \to \mathbf{z}(t+1) \to \overline{B}$ over all possible $\mathbf{z}(t+1)$:

$$q_t^+(i) = \sum_{j=1}^{M_{t+1}} P_{t,t+1}(i,j) q_{t+1}^+(j) \tag{7.20}$$

Thus, $q_t^+(i)$ comes from left-multiplying q_{t+1}^+ by $P_{t,t+1}$. Because the recursion moves backward in time, we need a terminal condition. Because we have defined $\overline{A}$ to include all days beyond the end of winter, the terminal condition is simply $q_T^+(i) = 0$ for all $i \in \{1,\ldots,M_T\}$.

(c) The backward committor q_t^- is the probability that the winter *so far* is SSW-free; in other words, that $\mathbf{Z}(t)$ last came from A (pre-winter) rather than B (the SSW state):

$$q_t^-(\mathbf{z}) = \begin{cases} \mathbb{P}\{\mathbf{Z} \text{ most recently came from } \overline{A} \text{ rather than } \overline{B} | \mathbf{Z}_t = \mathbf{z}\} & (t,\mathbf{z}) \notin \overline{A} \cup \overline{B} \\ 1 & (t,\mathbf{z}) \in \overline{A} \\ 0 & (t,\mathbf{z}) \in \overline{B} \end{cases} \tag{7.21}$$

This definition requires some sensible definition of "backward-in-time" dynamics. For this we construct a time-reversed transition matrix $\widetilde{P}_{t+1,t} \in \mathbb{R}^{M_{t+1} \times M_t}$, which we compute using Bayes' rule:

$$\widetilde{P}_{t+1,t}(j,i) = \mathbb{P}\{\mathbf{Z}(t) = i | \mathbf{Z}(t+1) = j\} \tag{7.22}$$

$$= \frac{\mathbb{P}\{\mathbf{Z}(t) = i\}\mathbb{P}\{\mathbf{Z}(t+1) = j | \mathbf{Z}(t) = j\}}{\mathbb{P}\{\mathbf{Z}(t+1) = j\}} \tag{7.23}$$

$$= \frac{\pi_t(i) P_{t,t+1}(i,j)}{\pi_{t+1}(j)} \tag{7.24}$$

The requirement of $\widetilde{P}$ to be a properly normalized stochastic matrix is why we stipulated that each column, as well as each row, of the count matrix $C_{t,t+1}$ must also have some nonzero entries. We can now compute q^- with the same procedure as q^+ above, but now using $\widetilde{P}$ and sweeping forward in time:

$$q^-_{t+1}(j) = \sum_{i=1}^{M_t} \widetilde{P}_{t+1,t}(j,i) q^-_t(i) \text{ for } t = 1, \ldots, T \tag{7.25}$$

Because $\overline{A}$ includes all states at time $t = 0$, the initial condition for q^- is simply $q^-_0(i) = 1$ for all $i \in \{1, \ldots, M_0\}$.

The forward and backward committors are displayed as functions of $(t, U_{10,60})$ in Fig. 7.6. It is the above calculations that reveal the advantage of an MSM over a linear inverse model (LIM): with a discrete state space and properly normalized transition matrices, the committor probabilities are guaranteed to fall between zero and one, while the probability density π_t remains properly normalized at each timestep. No such guarantee exists for calculations with a LIM.

4. **Estimate the rate.** Given the three quantities above, the rate can be written as a weighted

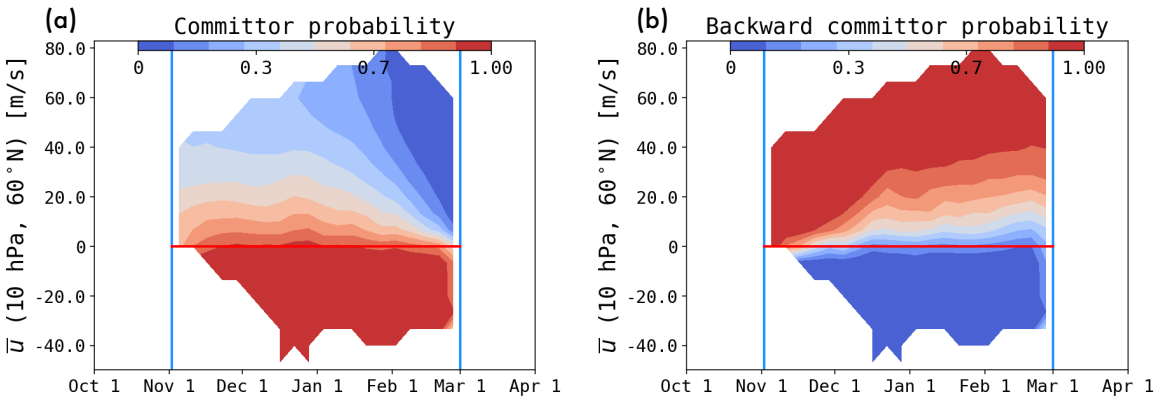


Figure 7.6: **Committor probabilities**. (Left) Forward committor q_t^+, the probability of reaching set B (the SSW state) before returning to A at the end of winter. (Right) Backward committor q_t^-, the probability that the winter so far has been SSW-free.

sum over trajectories leaving $\overline{A}$,

$$\text{Rate} = \sum_{i=1}^{M_{t_1}} \pi_{t_1}(i) q_{t_1}^+(i) \tag{7.26}$$

or alternatively as a weighted sum over trajectories entering $\overline{B}$,

$$\text{Rate} = \sum_{t=t_1}^{t_2} \sum_{i=1}^{M_t} \pi_t(i) q_t^-(i) \sum_{j:(t+1,j)\in B} P_{t,t+1}(i,j). \tag{7.27}$$

5. **Estimate the seasonal distribution.** A different decomposition of the rate formula can reveal the seasonal distribution. More generally than in the two formulas above, one can estimate the rate by partitioning space-time into two disjoint components, $C(\overline{A})$ containing $\overline{A}$ and $C(\overline{B})$ containing $\overline{B}$, with $C(\overline{A}) \cup C(\overline{B}) = [0,T] \times \mathbb{R}^d$, and write the rate as a weighted sum of transitions from one component to the other:

$$\text{Rate} = \sum_{t=0}^{T-1} \sum_{i=1}^{M_t} \mathbb{1}_{C(\overline{A})}\big((t,i)\big) \pi_t(i) q_t^-(i) \sum_{j=1}^{M_{t+1}} \mathbb{1}_{C(\overline{B})}\big((t+1,j)\big) P_{t,t+1}(i,j) q_{t+1}^+(j) \tag{7.28}$$

This formula follows Metzner et al. [2009] and is exact for a discrete Markov chain. We can

write it more compactly by collapsing the π, q^-, P, and π terms into a single *reactive flux* $F^{(AB)}$, such that

$$F^{(AB)}_{t,t+1}(i,j) = \pi_t(i) q^-_t(i) P_{t,t+1}(i,j) q^+_{t+1}(j) \tag{7.29}$$

$$\implies \text{Rate} = \sum_{t=0}^{T-1} \sum_{i=1}^{M_t} \sum_{j=1}^{M_{t+1}} \mathbb{1}_{C(\overline{A})}\big((t,i)\big) \mathbb{1}_{C(\overline{B})}\big((t+1,j)\big) F^{(AB)}_{t,t+1}(i,j) \tag{7.30}$$

$F^{(AB)}_{t,t+1}(i,j)$ is the flow of probability mass per unit time en route from $\overline{A}$ to $\overline{B}$ by way of $(t,i) \to (t+1,j)$. It encodes a discretized version of the continuous-space-time current $\mathbf{J}_{AB}(t,\mathbf{x})$ displayed in Fig. 7.3a. To make this connection explicit, we identify the boundary between $C(\overline{A})$ and $C(\overline{B})$ with a surface S, whose unit normal vector $\mathbf{n}$ points into the B side. $\mathbf{J}_{AB}$ is then defined implicitly as the vector field such that

$$\int_S \mathbf{J}_{AB} \cdot \mathbf{n} \, d\sigma = \text{ Rate} \tag{7.31}$$

where $d\sigma$ is a surface element on S. We have chosen to focus on one particular surface of interest: $S = \{(t,\mathbf{x}) : U_{10,60}(\mathbf{x}) = U^{(\text{th})}_{10,60}\}$, i.e., the surface of $\overline{B}$ itself, which is used in Eq. (7.27) above. This way, the crossing time t is identified with the central date of the SSW, and everything inside the outer sum of Eq. (7.30) can be considered the probability mass function at t of the seasonal distribution of SSW events.

Visualization

Figs. 7.3 in the main text and 7.6 involve two-dimensional projections of scalar fields and vector fields. We briefly describe the procedure for projecting scalar and vector fields, which closely follows Strahan et al. [2021] and Finkel et al. [2021b,a].

After building the Markov state model, and solving for the probability distribution $\pi_t(i)$ for all

times t and clusters i, we assign a weight to each snapshot known as the *change of measure*:

$$\gamma(n,t_n(s)) = \frac{\pi_{t_n(s)}(\mathbf{Z}_n(t_n(s)))}{\sum_{n'=1}^{N}\sum_{s'=0}^{46}\mathbb{1}\{t_{n'}(s') = t_n(s)\}\mathbb{1}\{\mathbf{Z}_{n'}(s') = \mathbf{Z}_n(s)\}} \tag{7.32}$$

The change of measure converts the *sampling distribution* μ—the distribution that $\mathbf{x}$ is drawn from—to the the climatological distribution π. The change of measure obeys the normalization condition $\sum_{n=1}^{N}\gamma(n,t) = 1$, which follows directly from the normalization condition $\sum_{i=1}^{M_t}\pi_t(i) = 1$.

Suppose we wish to visualize a function $G(t,\mathbf{x})$ in a low-dimensional space $\mathbf{y} = \mathbf{Y}(\mathbf{x})$, a vector-valued observable function with a dimension k much less than the dimension d of $\mathbf{x}$ (usually 1 or 2). Abbreviate $\mathbf{Y}(t_n(s),\mathbf{X}_n(t_n(s)))$ as $\mathbf{Y}_n(t_n(s))$. We discretize the projection space $\mathbb{R}^k$ into small pieces d$\mathbf{y}$, and define the projection

$$G^{\mathbf{Y}}(\mathbf{y}) = \frac{\sum_{n=1}^{N}\sum_{s=0}^{T}\mathbb{1}_{\mathrm{d}\mathbf{y}}\big(\mathbf{Y}_n(t_n(s))\big)G\big(t_n(s),\mathbf{X}_n(t_n(s))\big)\gamma\big(n,t_n(s)\big)}{\sum_{n'=1}^{N}\sum_{s'=0}^{T}\mathbb{1}_{\mathrm{d}\mathbf{y}}(\mathbf{Y}_{n'}(t_{n'}(s')))\gamma(n',t_{n'}(s'))} \tag{7.33}$$

In words, we take a weighted average of G evaluated at all snapshots $\mathbf{X}_n$ that map to $\mathbf{y}$ under the action of $\mathbf{Y}$. The weighting is the change of measure, γ.

This formula now positions us easily to project the vector field $\mathbf{J}_{AB}$. For every trajectory that transitions from $(t_n(s),\mathbf{X}_n(t_n(s)))$ to $(t_n(s+1),\mathbf{x}(t_n(s+1)))$, we define the projected current

$$\mathbf{J}^{\mathbf{Y}}_{AB}\big(t_n(s),\mathbf{Z}(t_n(s))\big) = q^{-}_{t_n(s)}(\mathbf{Z}_n(t_n(s)))q^{+}_{t_n(s+1)}(\mathbf{Z}_n(t_n(s+1)))\big[\mathbf{Y}_n(t_n(s+1)) - \mathbf{Y}_n(t_n(s))\big] \tag{7.34}$$

$\mathbf{J}^{\mathbf{Y}}_{AB}$ is a vector field with the same dimension as $\mathbf{Y}$. To project it, we simply treat each component as a scalar field like $G^{\mathbf{Y}}$ and apply the formula above. This gives us the arrows in Fig. 7.3a, where the first component of $\mathbf{Y}$ is t itself and the second component is $U_{10,60}$. Meanwhile, the background color of Fig. 7.3a in the main text is the probability density of $A \to B$ transition paths, a projection of the product $\gamma q^{-} q^{+}$, which happens to be identical to the t component of $\mathbf{J}^{\mathbf{Y}}_{AB}$. Fig. 7.3b shows

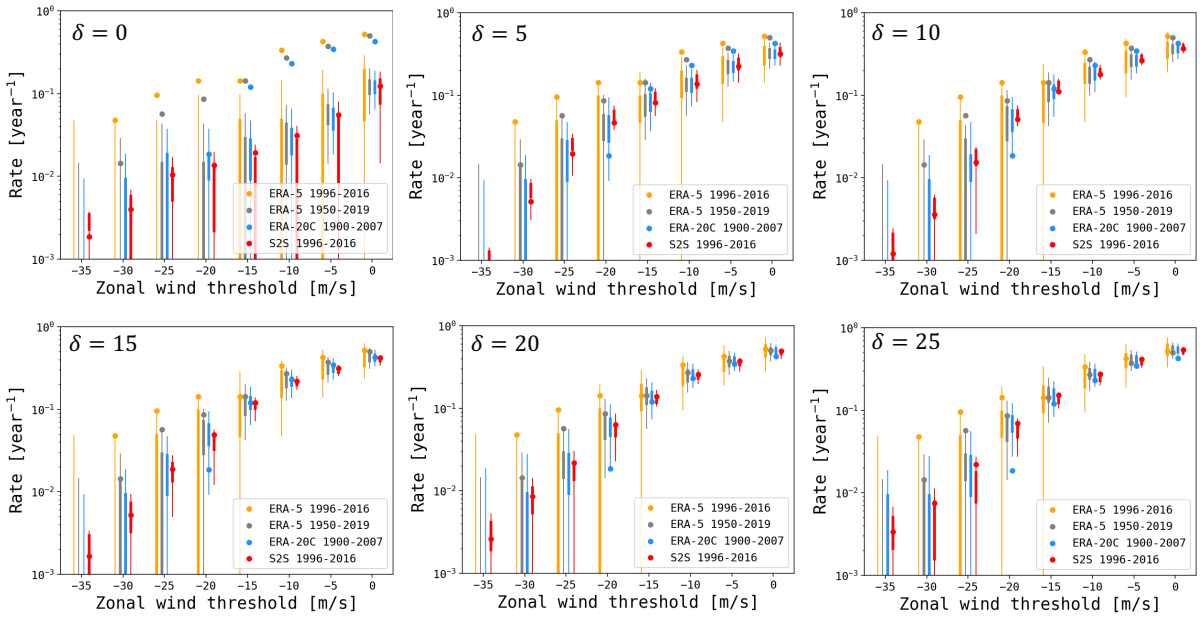


Figure 7.7: **Behavior of rate estimates as a function of time delay**. From upper left to bottom right, the number of time delays δ increases from 0 to 25 m/s. In every case, the feature space has dimension $\delta+1$ ($U_{10,60}$ at times $t, t-1, \ldots, t-\delta$).

the analogous current $\mathbf{J}_{AA}$ and density π_{AA} for $A \to A$ paths, which simply replaces q^+ with $1-q^+$ in the formulas above ($1-q^+$ is the probability of reaching A, the end of winter, with no SSW).

Feature selection and parameter tuning

We experimented with several feature spaces including empirical orthogonal functions (EOFs) and heat fluxes, but found simple time-delay embedding of $U_{10,60}$ to give the best tradeoff between simplicity and accuracy, as measured by agreement with ERA-5 (1950-2019) for less-extreme SSW events. It is unclear *a priori* how many time delays to include, however. We systematically varied the number δ of time-delays from 0 to 25 and show the results of each in Fig. 7.7, in the same layout as Fig. 7.2 of the main text.

The trends with δ are informative. To use $\delta = 0$ is to predict an SSW probability knowing only a snapshot $U_{10,60}$. The result is a systematic underestimate of rates. Increasing δ to 5 days

already provides vast improvement, which continues gradually upon increasing δ further. The choice of $\delta = 20$ days seems to approximately optimize three different notions of plausibility at once: (i) agreement with ERA-5 (1950-2019) on the more common events, (ii) narrowness of the bootstrapped S2S error bars, and (iii) symmetry of the S2S error bars about the point estimates. For time delays less than 15 days, S2S systematically underestimates rates relative to ERA-5 (1950-2019), and comes with negative error bar skew. This means that removing a year of data at random tends to pull the estimate systematically downward. As δ increases to 20 days, the S2S estimates climb steadily toward the ERA-5 rates. Increasing δ to 25 days increases the S2S estimates even slightly farther, but begins to produce negatively skewed error bars again, a possible sign of overfitting. These trends suggest an optimal tradeoff between the expressiveness of the feature space and the diminishing performance of k-means with increasing dimensionality. Our ultimate choice of δ partially uses the answer that we want to get, but only for more common SSW events on which reanalysis is reliable. The true strength of our method is to extrapolate, in a way informed by dynamics, to the more extreme rates.

8 CONCLUSION

Extreme weather events present a fundamental challenge to Earth system modeling. Many years of simulations are needed to generate many extreme events and reduce statistical error, but high-fidelity models are needed to simulate those event accurately. Conventionally, no single model can provide both, simply because of computational costs.

This thesis has described a computational framework, transition path theory (TPT), and a computational method, dynamical Galerkin approximation (DGA) to compute rare event statistics efficiently by combining the minimalistic philosophy of dimensionality reduction with the fidelity of high-resolution models. We have identified a small number of reduced coordinates, including the committor and lead time, that convey essential dynamics and statistics about the event. In its focus on directly estimating statistics of interest, DGA differs from previous reduced-order modeling methods that attempt to capture general qualities of the system, including both physics-based models [Lorenz, 1963, Charney and DeVore, 1979, Legras and Ghil, 1985, Crommelin, 2003, Timmermann et al., 2003, Ruzmaikin et al., 2003] and more recent data-driven models making use of machine learning [Giannakis and Majda, 2012, 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, Kashinath et al., 2021, Chattopadhyay et al., 2021].

We have demonstrated TPT analysis systematically across a hierarchy of SSW models, from a one-layer quasi-geostrophic model to a state-of-the-art ensemble forecasting system. However, there remain many challenges to broad, large-scale application. DGA would be most powerful if computed adaptively, in parallel with targeted sampling schemes with an executable climate model. To do this effectively, we will have to move beyond the simple sampling procedure of chapters 5-6, which generates a trajectory long enough to thoroughly sample transitions. This would not be practical for realistic models. One promising alternative would be to launch many trajectories in parallel and selectively replicate those that explore new regions of state space, especially transition regions. Such an approach could build on exciting progress over the last decade in targeted rare

event simulation schemes [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., 2021b, Ragone et al., 2018, Sapsis, 2021]. A potential challenge here is that GCMs may not be set up for short simulations that start and stop frequently. For this reason, it may be sensible to use longer lag times and a sliding window to define short trajectories. Furthermore, the communication overhead required for adaptive sampling with GCMs would impose additional costs. We have deferred the sampling problem to future work, acknowledging that this step is crucial to make DGA competitive. The utility of TPT quantities for scientific insight, however, is independent of the method for computing them.

Defining the source of stochasticity is also an important step that varies between models. Explicitly stochastic parameterization [e.g., Berner et al., 2009, Porta Mana and Zanna, 2014] will automatically lead to a spread in the short-trajectory ensemble, but in deterministic models, uncertainty will arise from perturbing the initial conditions. This may require special care depending on the model.

Another area of algorithmic improvement is selecting 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, quantum mechanics, and fluid dynamics [e.g., Carleo and Troyer, 2017, Han et al., 2018, Khoo et al., 2018, Li et al., 2020, Mardt et al., 2018, Li et al., 2019, Raissi et al., 2019, Lorpaiboon et al., 2020, Rotskoff and Vanden-Eijnden, 2020]. Similar approaches may hold great potential for understanding predictability in atmospheric science.

Despite these limitations, our demonstrations so far open the door to address many other data-limited questions of basic physical interest, for SSW and other phenomena. For example, how important are vortex preconditioning and upward wave activity as triggers of SSW? [Charlton

and Polvani, 2007, Albers and Birner, 2014]. Do split-type and displacement-type events have fundamentally different mechanisms and/or different downstream effects? [Matthewman and Esler, 2011, Esler and Matthewman, 2011, O'Callaghan et al., 2014, Maycock and Hitchcock, 2015]. Will climate change affect the frequency of SSW, perhaps through arctic amplification? [Charlton-Perez et al., 2008, Garfinkel et al., 2017, Kretschmer et al., 2018b]. How do other slow climatic variables, such as El Niño Southern Oscillation (ENSO), the quasi-biennial oscillation (QBO), and the Aleutian Low affect SSW propensity? [Dimdore-Miles et al., 2021]. These questions have been addressed in a number of coarse-resolution climate modeling studies, but high-resolution weather forecast data is an untapped source of potential for sharpening the answers. Our method offers a way forward, and is highly customizable to include physical features tailored for the problem at hand.

As a tool for simulation, TPT and DGA can help to constrain and diagnose various modeling choices when it comes to extremes. Machine-learning parameterizations are also gaining momentum in the climate modeling community, as a cheap and accurate alternative to extreme refinement of grids [e.g., O'Gorman and Dwyer, 2018, Kim et al., 2019, Bolton and Zanna, 2019, Frezat et al., 2021, Gentine et al., 2018, Chattopadhyay et al., 2021, Kashinath et al., 2021, Yuval et al., 2021]. Such data-driven models are usually trained to minimize some notion of mean-squared error relative to direct simulation. But it is the extreme behavior, not average behavior, where those simulations matter the most. TPT offers a set of diagnostics that could help constrain data-driven models to perform well on extreme events. At minimum, without knowledge of what "performing well" means due to lack of data, the TPT framework and DGA method can efficiently extract the consequences of a certain modeling choice for the occurrence of extremes.

Another potential application of DGA is catastrophe modeling under climate change. Tropical cyclones pose a pressing problem for coastal communities, and have motivated several hybrid dynamical/statistical downscaling methods to project risk into the future under various climate change scenarios [Camargo et al., 2014, Lee et al., 2018, Jing and Lin, 2020, Sobel et al., 2021]. Extreme

precipitation of many varieties threatens cities and agriculture and is expected to change significantly with global warming [e.g., O'Gorman, 2012, Pfahl et al., 2017]. Model resolution, again, is the limiting factor [Laflamme et al., 2016, O'Brien et al., 2016, He et al., 2019]. Enlisting short weather forecasts, as we have done, may help identify antecedent conditions and drivers of changing frequency with unprecedented detail. Especially in concert with other emerging techniques such as rare event sampling and machine learning, DGA will be an asset for quantifying extreme event probabilities and aiding efforts to understand their fundamental physical mechanisms.

REFERENCES

Dorian S. Abbot, Robert J. Webber, Sam Hadden, Darryl Seligman, and Jonathan Weare. Rare event sampling improves mercury instability statistics. *The Astrophysical Journal*, 923(2):236, dec 2021. doi: 10.3847/1538-4357/ac2fa8. URL https://doi.org/10.3847/1538-4357/ac2fa8.

Amir AghaKouchak, Linyin Cheng, Omid Mazdiyasni, and Alireza Farahmand. Global warming and changes in risk of concurrent climate extremes: Insights from the 2014 california drought. *Geophysical Research Letters*, 41(24):8847–8852, 2014. doi: https://doi.org/10.1002/2014GL062308. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014GL062308.

John R. Albers and Thomas Birner. Vortex preconditioning due to planetary and gravity waves prior to sudden stratospheric warmings. *Journal of the Atmospheric Sciences*, 71(11):4028 – 4054, 2014. doi: 10.1175/JAS-D-14-0026.1. URL https://journals.ametsoc.org/view/journals/atsc/71/11/jas-d-14-0026.1.xml.

Adam Antoszewski, Chatipat Lorpaiboon, John Strahan, and Aaron R. Dinner. Kinetics of phenol escape from the insulin r6 hexamer. *The Journal of Physical Chemistry B*, 125(42):11637–11649, 2021. doi: 10.1021/acs.jpcb.1c06544. URL https://doi.org/10.1021/acs.jpcb.1c06544. PMID: 34648712.

B. Ayarzagüena, F. M. Palmeiro, D. Barriopedro, N. Calvo, U. Langematz, and K. Shibata. On the representation of major stratospheric warmings in reanalyses. *Atmospheric Chemistry and Physics*, 19(14):9469–9484, 2019. doi: 10.5194/acp-19-9469-2019. URL https://acp.copernicus.org/articles/19/9469/2019/.

B. Ayarzagüena, A. J. Charlton-Perez, A. H. Butler, P. Hitchcock, I. R. Simpson, L. M. Polvani, N. Butchart, E. P. Gerber, L. Gray, B. Hassler, P. Lin, F. Lott, E. Manzini, R. Mizuta, C. Orbe, S. Osprey, D. Saint-Martin, M. Sigmond, M. Taguchi, E. M. Volodin, and S. Watanabe. Uncertainty in the response of sudden stratospheric warmings and stratosphere-troposphere coupling to quadrupled co2 concentrations in cmip6 models. *Journal of Geophysical Research: Atmospheres*, 125(6):e2019JD032345, 2020. doi: https://doi.org/10.1029/2019JD032345. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019JD032345. e2019JD032345 2019JD032345.

Mark P. Baldwin and Timothy J. Dunkerton. Stratospheric harbingers of anomalous weather regimes. *Science*, 294(5542):581–584, 2001. doi: 10.1126/science.1063315. URL https://www.science.org/doi/abs/10.1126/science.1063315.

Mark P. Baldwin, David B. Stephenson, David W. J. Thompson, Timothy J. Dunkerton, Andrew J. Charlton, and Alan O'Neill. Stratospheric memory and skill of extended-range weather forecasts. *Science*, 301(5633):636–640, 2003. doi: 10.1126/science.1087143. URL https://www.science.org/doi/abs/10.1126/science.1087143.

Mark P. Baldwin, Blanca Ayarzagüena, Thomas Birner, Neal Butchart, Amy H. Butler, Andrew J. Charlton-Perez, Daniela I. V. Domeisen, Chaim I. Garfinkel, Hella Garny, Edwin P. Gerber, Michaela I. Hegglin, Ulrike Langematz, and Nicholas M. Pedatella. Sudden stratospheric warmings. *Reviews of Geophysics*, 59(1):e2020RG000708, 2021. doi: https://doi.org/10.1029/2020RG000708. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020RG000708. e2020RG000708 10.1029/2020RG000708.

S. Bancalá, K. Krüger, and M. Giorgetta. The preconditioning of major sudden stratospheric warmings. *Journal of Geophysical Research: Atmospheres*, 117(D4), 2012. doi: https://doi.org/10.1029/2011JD016769. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011JD016769.

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

Ming Bao, Xin Tan, Dennis L. Hartmann, and Paulo Ceppi. Classifying the tropospheric precursor patterns of sudden stratospheric warmings. *Geophysical Research Letters*, 44(15):8011–8016, 2017. doi: https://doi.org/10.1002/2017GL074611. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017GL074611.

J. Berner, G. J. Shutts, M. Leutbecher, and T. N. Palmer. A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ecmwf ensemble prediction system. *Journal of the Atmospheric Sciences*, 66(3):603 – 626, 2009. doi: 10.1175/2008JAS2677.1. URL https://journals.ametsoc.org/view/journals/atsc/66/3/2008jas2677.1.xml.

P. Berrisford, D.P. Dee, P. Poli, R. Brugge, Mark Fielding, Manuel Fuentes, P.W. Kållberg, S. Kobayashi, S. Uppala, and Adrian Simmons. The era-interim archive version 2.0. (1):23, 11 2011. URL https://www.ecmwf.int/node/8174.

T. Berry, J. R. Cressman, Z. Gregurić-Ferenček, and T. Sauer. Time-scale separation from diffusion-mapped delay coordinates. *SIAM Journal on Applied Dynamical Systems*, 12(2):618–649, 2013. doi: 10.1137/12088183X. URL https://doi.org/10.1137/12088183X.

Tyrus Berry, Dimitrios Giannakis, and John Harlim. Nonparametric forecasting of low-dimensional dynamical systems. *Phys. Rev. E*, 91:032915, Mar 2015. doi: 10.1103/PhysRevE.91.032915. URL https://link.aps.org/doi/10.1103/PhysRevE.91.032915.

Ilias Bilionis. Probabilistic solvers for partial differential equations, 2016. URL https://arxiv.org/abs/1607.03526.

Richard P Binzel. The torino impact hazard scale. *Planetary and Space Science*, 48(4):297–303, 2000. ISSN 0032-0633. doi: https://doi.org/10.1016/S0032-0633(00)00006-4. URL https://www.sciencedirect.com/science/article/pii/S0032063300000064.

Thomas Birner and Paul D. Williams. Sudden stratospheric warmings as noise-induced transitions. *Journal of the Atmospheric Sciences*, 65(10):3337 – 3343, 2008. doi: 10.1175/2008JAS2770.1. URL https://journals.ametsoc.org/view/journals/atsc/65/10/2008jas2770.1.xml.

Robert X. Black, Brent A. McDaniel, and Walter A. Robinson. Stratosphere–troposphere coupling during spring onset. *Journal of Climate*, 19(19):4891 – 4901, 2006. doi: 10.1175/JCLI3907.1. URL https://journals.ametsoc.org/view/journals/clim/19/19/jcli3907.1.xml.

H. C. Bloomfield, D. J. Brayshaw, P. L. M. Gonzalez, and A. Charlton-Perez. Sub-seasonal forecasts of demand and wind power and solar power generation for 28 european countries. *Earth System Science Data*, 13(5):2259–2274, 2021. doi: 10.5194/essd-13-2259-2021. URL https://essd.copernicus.org/articles/13/2259/2021/.

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

Nawaf Bou-Rabee and Eric Vanden-Eijnden. Continuous-time random walks for the numerical solution of stochastic differential equations, 2015. URL https://arxiv.org/abs/1502.05034.

Freddy Bouchet, Jason Laurie, and Oleg Zaboronski. Control and instanton trajectories for random transitions in turbulent flows. *Journal of Physics: Conference Series*, 318(2):022041, dec 2011. doi: 10.1088/1742-6596/318/2/022041. URL https://doi.org/10.1088/1742-6596/318/2/022041.

Freddy Bouchet, Jason Laurie, and Oleg Zaboronski. Langevin dynamics, large deviations and instantons for the quasi-geostrophic model and two-dimensional euler equations. *Journal of Statistical Physics*, 156(6):1066–1092, Sep 2014. ISSN 1572-9613. doi: 10.1007/s10955-014-1052-5. URL https://doi.org/10.1007/s10955-014-1052-5.

Freddy Bouchet, Joran Rolland, and Eric Simonnet. Rare event algorithm links transitions in turbulent flows with activated nucleations. *Phys. Rev. Lett.*, 122:074502, Feb 2019a. doi: 10.1103/PhysRevLett.122.074502. URL https://link.aps.org/doi/10.1103/PhysRevLett.122.074502.

Freddy Bouchet, Joran Rolland, and Jeroen Wouters. Rare event sampling methods. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(8):080402, 2019b. doi: 10.1063/1.5120509. URL https://doi.org/10.1063/1.5120509.

Gregory R. Bowman, Kyle A. Beauchamp, George Boxer, and Vijay S. Pande. Progress and challenges in the automated construction of markov state models for full protein systems. *The Journal of Chemical Physics*, 131(12):124101, 2009. doi: 10.1063/1.3216567. URL https://doi.org/10.1063/1.3216567.

Gregory R Bowman, Vijay S Pande, and Frank Noé. *An introduction to Markov state models and their application to long timescale molecular simulation*, volume 797. Springer Science & Business Media, 2013.

D.S. Broomhead and Gregory P. King. Extracting qualitative dynamics from experimental data. *Physica D: Nonlinear Phenomena*, 20(2):217–236, 1986. ISSN 0167-2789. doi: https://doi.org/10.1016/0167-2789(86)90031-X. URL https://www.sciencedirect.com/science/article/pii/016727898690031X.

Steven L. Brunton, Bingni W. Brunton, Joshua L. Proctor, Eurika Kaiser, and J. Nathan Kutz. Chaos as an intermittently forced linear system. *Nature Communications*, 8(1):19, May 2017. ISSN 2041-1723. doi: 10.1038/s41467-017-00030-8. URL https://doi.org/10.1038/s41467-017-00030-8.

Reid A. Bryson. The paradigm of climatology: An essay. *Bulletin of the American Meteorological Society*, 78(3):449 – 456, 1997. doi: 10.1175/1520-0477(1997)078⟨0449:TPOCAE⟩2.0.CO;2. URL https://journals.ametsoc.org/view/journals/bams/78/3/1520-0477_1997_078_0449_tpocae_2_0_co_2.xml.

R. Buizza, M. Milleer, and T. N. Palmer. Stochastic representation of model uncertainties in the ecmwf ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, 125 (560):2887–2908, 1999. doi: https://doi.org/10.1002/qj.49712556006. URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.49712556006.

Amy Butler, Andrew Charlton-Perez, Daniela I.V. Domeisen, Chaim Garfinkel, Edwin P. Gerber, Peter Hitchcock, Alexey Yu. Karpechko, Amanda C. Maycock, Michael Sigmond, Isla Simpson, and Seok-Woo Son. Chapter 11 - sub-seasonal predictability and the stratosphere. In Andrew W. Robertson and Frédéric Vitart, editors, *Sub-Seasonal to Seasonal Prediction*, pages 223–241. Elsevier, 2019. ISBN 978-0-12-811714-9. doi: https://doi.org/10.1016/B978-0-12-811714-9.00011-5. URL https://www.sciencedirect.com/science/article/pii/B9780128117149000115.

Amy H. Butler and Edwin P. Gerber. Optimizing the definition of a sudden stratospheric warming. *Journal of Climate*, 31(6):2337 – 2344, 2018. doi: 10.1175/JCLI-D-17-0648.1. URL https://journals.ametsoc.org/view/journals/clim/31/6/jcli-d-17-0648.1.xml.

Amy H. Butler, Dian J. Seidel, Steven C. Hardiman, Neal Butchart, Thomas Birner, and Aaron Match. Defining sudden stratospheric warmings. *Bulletin of the American Meteorological Society*, 96(11):1913 – 1928, 2015. doi: 10.1175/BAMS-D-13-00173.1. URL https://journals.ametsoc.org/view/journals/bams/96/11/bams-d-13-00173.1.xml.

Suzana J. Camargo, Michael K. Tippett, Adam H. Sobel, Gabriel A. Vecchi, and Ming Zhao. Testing the performance of tropical cyclone genesis indices in future climates using the hiram model. *Journal of Climate*, 27(24):9171 – 9196, 2014. doi: 10.1175/JCLI-D-13-00505.1. URL https://journals.ametsoc.org/view/journals/clim/27/24/jcli-d-13-00505.1.xml.

Giuseppe Carleo and Matthias Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355(6325):602–606, 2017. doi: 10.1126/science.aag2302. URL https://www.science.org/doi/abs/10.1126/science.aag2302.

Andrew J. Charlton and Lorenzo M. Polvani. A new look at stratospheric sudden warmings. part i: Climatology and modeling benchmarks. *Journal of Climate*, 20(3):449 – 469, 2007. doi: 10.1175/JCLI3996.1. URL https://journals.ametsoc.org/view/journals/clim/20/3/jcli3996.1.xml.

Andrew J. Charlton, Lorenzo M. Polvani, Judith Perlwitz, Fabrizio Sassi, Elisa Manzini, Kiyotaka Shibata, Steven Pawson, J. Eric Nielsen, and David Rind. A new look at stratospheric sudden warmings. part ii: Evaluation of numerical model simulations. *Journal of Climate*, 20(3): 470 – 488, 2007. doi: 10.1175/JCLI3994.1. URL https://journals.ametsoc.org/view/journals/clim/20/3/jcli3994.1.xml.

A. J. Charlton-Perez, L. M. Polvani, J. Austin, and F. Li. The frequency and dynamics of stratospheric sudden warmings in the 21st century. *Journal of Geophysical Research: Atmospheres*, 113(D16), 2008. doi: https://doi.org/10.1029/2007JD009571. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2007JD009571.

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

Jule G. Charney and John G. DeVore. Multiple flow equilibria in the atmosphere and blocking. *Journal of Atmospheric Sciences*, 36(7):1205 – 1216, 1979. doi: 10.1175/1520-0469(1979)036⟨1205:MFEITA⟩2.0.CO;2. URL https://journals.ametsoc.org/view/journals/atsc/36/7/1520-0469_1979_036_1205_mfeita_2_0_co_2.xml.

Ashesh Chattopadhyay, Ebrahim Nabizadeh, and Pedram Hassanzadeh. Analog forecasting of extreme-causing weather patterns using deep learning. *Journal of Advances in Modeling Earth Systems*, 12(2):e2019MS001958, 2020. doi: https://doi.org/10.1029/2019MS001958. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001958. e2019MS001958 10.1029/2019MS001958.

Ashesh Chattopadhyay, Mustafa Mustafa, Pedram Hassanzadeh, Eviatar Bach, and Karthik Kashinath. Towards physically consistent data-driven weather forecasting: Integrating data assimilation with equivariance-preserving deep spatial transformers, 2021. URL https://arxiv.org/abs/2103.09360.

Nan Chen and Andrew J. Majda. 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):033101, 2020. doi: 10.1063/1.5122199. URL https://doi.org/10.1063/1.5122199.

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

John D Chodera and Frank Noé. Markov state models of biomolecular conformational dynamics. *Current Opinion in Structural Biology*, 25:135–144, 2014. ISSN 0959-440X. doi: https://doi.org/10.1016/j.sbi.2014.04.002. URL https://www.sciencedirect.com/science/article/pii/S0959440X14000426. Theory and simulation / Macromolecular machines.

John D. Chodera, William C. Swope, Jed W. Pitera, and Ken A. Dill. Long-time protein folding dynamics from short-time molecular dynamics simulations. *Multiscale Modeling & Simulation*, 5(4):1214–1226, 2006. doi: 10.1137/06065146X. URL https://doi.org/10.1137/06065146X.

Bo Christiansen. Chaos, quasiperiodicity, and interannual variability: Studies of a stratospheric vacillation model. *Journal of the Atmospheric Sciences*, 57(18):3161 – 3173, 2000. doi: 10.1175/1520-0469(2000)057⟨3161:CQAIVS⟩2.0.CO;2. URL https://journals.ametsoc.org/view/journals/atsc/57/18/1520-0469_2000_057_3161_cqaivs_2.0.co_2.xml.

Dim Coumou and Stefan Rahmstorf. A decade of weather extremes. *Nature Climate Change*, 2(7):491–496, Jul 2012. ISSN 1758-6798. doi: 10.1038/nclimate1452. URL https://doi.org/10.1038/nclimate1452.

D. T. Crommelin. Regime transitions and heteroclinic connections in a barotropic atmosphere. *Journal of the Atmospheric Sciences*, 60(2):229 – 246, 2003. 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.

Pierre Del Moral. *Feynman-Kac Formulae*. Springer-Verlag, 2004.

Timothy DelSole and Brian F. Farrell. A stochastically excited linear system as a model for quasi-geostrophic turbulence:analytic results for one- and two-layer fluids. *Journal of Atmospheric Sciences*, 52(14):2531 – 2547, 1995. doi: 10.1175/1520-0469(1995)052⟨2531:ASELSA⟩2.0.CO;2. URL https://journals.ametsoc.org/view/journals/atsc/52/14/1520-0469_1995_052_2531_aselsa_2_0_co_2.xml.

Giovanni Dematteis, Tobias Grafke, and Eric Vanden-Eijnden. Rogue waves and large deviations in deep sea. *Proceedings of the National Academy of Sciences*, 115(5):855–860, 2018. doi: 10.1073/pnas.1710670115. URL https://www.pnas.org/doi/abs/10.1073/pnas.1710670115.

Giovanni Dematteis, Tobias Grafke, Miguel Onorato, and Eric Vanden-Eijnden. Experimental evidence of hydrodynamic instantons: The universal route to rogue waves. *Phys. Rev. X*, 9:041057, Dec 2019. doi: 10.1103/PhysRevX.9.041057. URL https://link.aps.org/doi/10.1103/PhysRevX.9.041057.

Peter Deuflhard, Michael Dellnitz, Oliver Junge, and Christof Schütte. Computation of essential molecular dynamics by subdivision techniques. In Peter Deuflhard, Jan Hermans, Benedict Leimkuhler, Alan E. Mark, Sebastian Reich, and Robert D. Skeel, editors, *Computational Molecular Dynamics: Challenges, Methods, Ideas*, pages 98–115, Berlin, Heidelberg, 1999. Springer Berlin Heidelberg. ISBN 978-3-642-58360-5.

O. Dimdore-Miles, L. Gray, and S. Osprey. Origins of multi-decadal variability in sudden stratospheric warmings. *Weather and Climate Dynamics*, 2(1):205–231, 2021. doi: 10.5194/wcd-2-205-2021. URL https://wcd.copernicus.org/articles/2/205/2021/.

Daniela I. V. Domeisen, Amy H. Butler, Andrew J. Charlton-Perez, Blanca Ayarzagüena, Mark P. Baldwin, Etienne Dunn-Sigouin, Jason C. Furtado, Chaim I. Garfinkel, Peter Hitchcock, Alexey Yu. Karpechko, Hera Kim, Jeff Knight, Andrea L. Lang, Eun-Pa Lim, Andrew Marshall, Greg Roff, Chen Schwartz, Isla R. Simpson, Seok-Woo Son, and Masakazu Taguchi. The role of the stratosphere in subseasonal to seasonal prediction: 2. predictability arising from stratosphere-troposphere coupling. *Journal of Geophysical Research: Atmospheres*, 125(2):e2019JD030923, 2020. doi: https://doi.org/10.1029/2019JD030923. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019JD030923. e2019JD030923 10.1029/2019JD030923.

Daniela I.V. Domeisen. Estimating the frequency of sudden stratospheric warming events from surface observations of the north atlantic oscillation. *Journal of Geophysical Research: Atmospheres*, 124(6):3180–3194, 2019. doi: https://doi.org/10.1029/2018JD030077. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018JD030077.

T. Dunkerton, C-P. F. Hsu, and M. E. McIntyre. Some eulerian and lagrangian diagnostics for a model stratospheric warming. *Journal of Atmospheric Sciences*, 38(4):819 – 844, 1981. doi: 10.1175/1520-0469(1981)038⟨0819:SEALDF⟩2.0.CO;2. URL https://journals.ametsoc.org/view/journals/atsc/38/4/1520-0469_1981_038_0819_sealdf_2_0_co_2.xml.

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

Weinan E. and Eric Vanden-Eijnden. Towards a theory of transition paths. *Journal of Statistical Physics*, 123(3):503, May 2006. ISSN 1572-9613. doi: 10.1007/s10955-005-9003-9. URL https://doi.org/10.1007/s10955-005-9003-9.

Weinan E and Eric Vanden-Eijnden. Transition-path theory and path-finding algorithms for the study of rare events. *Annual Review of Physical Chemistry*, 61(1):391–420, 2010. doi: 10.1146/annurev.physchem.040808.090412. URL https://doi.org/10.1146/annurev.physchem.040808.090412. PMID: 18999998.

Weinan E, Weiqing Ren, and Eric Vanden-Eijnden. Minimum action method for the study of rare events. *Communications on pure and applied mathematics*, 57(5):637–656, 2004.

Weinan E, Tiejun Li, and Eric Vanden-Eijnden. *Applied stochastic analysis*, volume 199. American Mathematical Soc., 2019.

David R. Easterling, Gerald A. Meehl, Camille Parmesan, Stanley A. Changnon, Thomas R. Karl, and Linda O. Mearns. Climate extremes: Observations, modeling, and impacts. *Science*, 289 (5487):2068–2074, 2000. doi: 10.1126/science.289.5487.2068. URL https://www.science.org/doi/abs/10.1126/science.289.5487.2068.

ECMWF. The era-interim reanalysis dataset. Copernicus Climate Change Service (C3S), 2011. available from https://www.ecmwf.int/en/forecasts/datasets/archive-datasets/reanalysis-datasets/era-interim (accessed 2022-02-10).

ECMWF. *IFS Documentation CY43R1 - Part I: Observations*. Number 1 in IFS Documentation. ECMWF, 2016a. doi: 10.21957/7zo03h0ve. URL https://www.ecmwf.int/node/17114.

ECMWF. *IFS Documentation CY43R1 - Part II: Data Assimilation*. Number 2 in IFS Documentation. ECMWF, 2016b. doi: 10.21957/am5dtg9pb. URL https://www.ecmwf.int/node/17115.

ECMWF. *IFS Documentation CY43R1 - Part III: Dynamics and Numerical Procedures*. Number 3 in IFS Documentation. ECMWF, 2016c. doi: 10.21957/m1u2yxwrl. URL https://www.ecmwf.int/node/17116.

ECMWF. *IFS Documentation CY43R1 - Part IV: Physical Processes*. Number 4 in IFS Documentation. ECMWF, 2016d. doi: 10.21957/sqvo5yxja. URL https://www.ecmwf.int/node/17117.

ECMWF. *IFS Documentation CY43R1 - Part V: Ensemble Prediction System*. Number 5 in IFS Documentation. ECMWF, 2016e. doi: 10.21957/6fm80smm. URL https://www.ecmwf.int/node/17118.

ECMWF. *IFS Documentation CY43R1 - Part VI: Technical and Computational Procedures*. Number 6 in IFS Documentation. ECMWF, 2016f. doi: 10.21957/xc3eo8i41. URL https://www.ecmwf.int/node/17119.

ECMWF. *IFS Documentation CY43R1 - Part VII: ECMWF Wave Model*. Number 7 in IFS Documentation. ECMWF, 2016g. doi: 10.21957/18mel2ooj. URL https://www.ecmwf.int/node/17120.

J. G. Esler and N. Joss Matthewman. Stratospheric sudden warmings as self-tuning resonances. part ii: Vortex displacement events. *Journal of the Atmospheric Sciences*, 68(11):2505 – 2523, 2011. doi: 10.1175/JAS-D-11-08.1. URL https://journals.ametsoc.org/view/journals/atsc/68/11/jas-d-11-08.1.xml.

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

Mohammad Farazmand and Themistoklis P. Sapsis. A variational approach to probing extreme events in turbulent dynamical systems. *Science Advances*, 3(9):e1701533, 2017. doi: 10.1126/sciadv.1701533. URL https://www.science.org/doi/abs/10.1126/sciadv.1701533.

Justin Finkel, Dorian S. Abbot, and Jonathan Weare. Path properties of atmospheric transitions: Illustration with a low-order sudden stratospheric warming model. *Journal of the Atmospheric Sciences*, 77(7):2327 – 2347, 2020. doi: 10.1175/JAS-D-19-0278.1. URL https://journals.ametsoc.org/view/journals/atsc/77/7/jasD190278.xml.

Justin Finkel, Robert J. Webber, Edwin P. Gerber, Dorian S. Abbot, and Jonathan Weare. Exploring stratospheric rare events with transition path theory and short simulations, 2021a. URL https://arxiv.org/abs/2108.12727.

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

E. M. Fischer, S. Sippel, and R. Knutti. Increasing probability of record-shattering climate extremes. *Nature Climate Change*, 11(8):689–695, Aug 2021. ISSN 1758-6798. doi: 10.1038/s41558-021-01092-9. URL https://doi.org/10.1038/s41558-021-01092-9.

P.J. Fitzsimmons and Jim Pitman. Kac's moment formula and the feynman–kac formula for additive functionals of a markov process. *Stochastic Processes and their Applications*, 79(1): 117–134, 1999. ISSN 0304-4149. doi: https://doi.org/10.1016/S0304-4149(98)00081-7. URL https://www.sciencedirect.com/science/article/pii/S0304414998000817.

Eric Forgoston and Richard O. Moore. A primer on noise-induced transitions in applied dynamical systems. *SIAM Review*, 60(4):969–1009, 2018. doi: 10.1137/17M1142028. URL https://doi.org/10.1137/17M1142028.

David J. Frame, Suzanne M. Rosier, Ilan Noy, Luke J. Harrington, Trevor Carey-Smith, Sarah N. Sparrow, Dáithí A. Stone, and Samuel M. Dean. Climate change attribution and the economic costs of extreme weather events: a study on damages from extreme rainfall and drought. *Climatic Change*, 162(2):781–797, Sep 2020. ISSN 1573-1480. doi: 10.1007/s10584-020-02729-y. URL https://doi.org/10.1007/s10584-020-02729-y.

Christian Franzke and Andrew J. Majda. Low-order stochastic mode reduction for a prototype atmospheric gcm. *Journal of the Atmospheric Sciences*, 63(2):457 – 479, 2006. doi: 10.1175/JAS3633.1. URL https://journals.ametsoc.org/view/journals/atsc/63/2/jas3633.1.xml.

Mark I. Freidlin and Alexander D. Wentzell. *Random perturbations of dynamical systems*. Springer, 1970.

Hugo Frezat, Julien Le Sommer, Ronan Fablet, Guillaume Balarac, and Redouane Lguensat. A posteriori learning of quasi-geostrophic turbulence parametrization: an experiment on integration steps, 2021.

Gary Froyland and Oliver Junge. Robust fem-based extraction of finite-time coherent sets using scattered, sparse, and incomplete trajectories. *SIAM Journal on Applied Dynamical Systems*, 17(2):1891–1924, 2018. doi: 10.1137/17M1129738. URL https://doi.org/10.1137/17M1129738.

Vera Melinda Gálfi, Valerio Lucarini, and Jeroen Wouters. A large deviation theory-based analysis of heat waves and cold spells in a simplified model of the general circulation of the atmosphere. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(3):033404, mar 2019. doi: 10.1088/1742-5468/ab02e8. URL https://doi.org/10.1088/1742-5468/ab02e8.

Vera Melinda Gálfi, Valerio Lucarini, Francesco Ragone, and Jeroen Wouters. Applications of large deviation theory in geophysical fluid dynamics and climate science. *La Rivista del Nuovo Cimento*, 44(6):291–363, Jun 2021. ISSN 1826-9850. doi: 10.1007/s40766-021-00020-z. URL https://doi.org/10.1007/s40766-021-00020-z.

Chaim I. Garfinkel, Seok-Woo Son, Kanghyun Song, Valentina Aquila, and Luke D. Oman. Stratospheric variability contributed to and sustained the recent hiatus in eurasian winter warming. *Geophysical Research Letters*, 44(1):374–382, 2017. doi: https://doi.org/10.1002/2016GL072035. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016GL072035.

R. Gelaro, R. Buizza, T. N. Palmer, and E. Klinker. Sensitivity analysis of forecast errors and the construction of optimal perturbations using singular vectors. *Journal of the Atmospheric Sciences*, 55(6):1012 – 1037, 1998. doi: 10.1175/1520-0469(1998)055⟨1012:SAOFEA⟩2.0.CO;2. URL https://journals.ametsoc.org/view/journals/atsc/55/6/1520-0469_1998_055_1012_saofea_2.0.co_2.xml.

Ronald Gelaro, Will McCarty, Max J. Suárez, Ricardo Todling, Andrea Molod, Lawrence Takacs, Cynthia A. Randles, Anton Darmenov, Michael G. Bosilovich, Rolf Reichle, Krzysztof Wargan, Lawrence Coy, Richard Cullather, Clara Draper, Santha Akella, Virginie Buchard, Austin Conaty, Arlindo M. da Silva, Wei Gu, Gi-Kong Kim, Randal Koster, Robert Lucchesi, Dagmar Merkova, Jon Eric Nielsen, Gary Partyka, Steven Pawson, William Putman, Michele Rienecker, Siegfried D. Schubert, Meta Sienkiewicz, and Bin Zhao. The modern-era retrospective analysis for research and applications, version 2 (merra-2). *Journal of Climate*, 30(14):5419 – 5454, 2017. doi: 10.1175/JCLI-D-16-0758.1. URL https://journals.ametsoc.org/view/journals/clim/30/14/jcli-d-16-0758.1.xml.

P. Gentine, M. Pritchard, S. Rasp, G. Reinaudi, and G. Yacalis. Could machine learning break the convection parameterization deadlock? *Geophysical Research Letters*, 45(11):5742–5751, 2018. doi: https://doi.org/10.1029/2018GL078202. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018GL078202.

E. P. Gerber, C. Orbe, and L. M. Polvani. Stratospheric influence on the tropospheric circulation revealed by idealized ensemble forecasts. *Geophysical Research Letters*, 36(24), 2009. doi: https://doi.org/10.1029/2009GL040913. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2009GL040913.

Dimitrios Giannakis. Dynamics-adapted cone kernels. *SIAM Journal on Applied Dynamical Systems*, 14(2):556–608, 2015. doi: 10.1137/140954544. URL https://doi.org/10.1137/140954544.

Dimitrios Giannakis. Data-driven spectral decomposition and forecasting of ergodic dynamical systems. *Applied and Computational Harmonic Analysis*, 47(2):338–396, 2019. ISSN 1063-5203. doi: https://doi.org/10.1016/j.acha.2017.09.001. URL https://www.sciencedirect.com/science/article/pii/S1063520317300982.

Dimitrios Giannakis and Andrew J. Majda. Nonlinear laplacian spectral analysis for time series with intermittency and low-frequency variability. *Proceedings of the National Academy of Sciences*, 109(7):2222–2227, 2012. doi: 10.1073/pnas.1118984109. URL https://www.pnas.org/doi/abs/10.1073/pnas.1118984109.

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

Michael Goss, Daniel L Swain, John T Abatzoglou, Ali Sarhadi, Crystal A Kolden, A Park Williams, and Noah S Diffenbaugh. Climate change is increasing the likelihood of extreme autumn wildfire conditions across california. *Environmental Research Letters*, 15(9):094016, aug 2020. doi: 10.1088/1748-9326/ab83a7. URL https://doi.org/10.1088/1748-9326/ab83a7.

Georg A. Gottwald, Daan T. Crommelin, and Christian L. E. Franzke. Stochastic climate theory, 2016. URL https://arxiv.org/abs/1612.07474.

Tobias Grafke and Eric Vanden-Eijnden. Numerical computation of rare events via large deviation theory. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(6):063118, 2019. doi: 10.1063/1.5084025. URL https://doi.org/10.1063/1.5084025.

R. J. Haarsma, M. J. Roberts, P. L. Vidale, C. A. Senior, A. Bellucci, Q. Bao, P. Chang, S. Corti, N. S. Fučkar, V. Guemas, J. von Hardenberg, W. Hazeleger, C. Kodama, T. Koenigk, L. R. Leung, J. Lu, J.-J. Luo, J. Mao, M. S. Mizielinski, R. Mizuta, P. Nobre, M. Satoh, E. Scoccimarro, T. Semmler, J. Small, and J.-S. von Storch. High resolution model intercomparison project (highresmip v1.0) for cmip6. *Geoscientific Model Development*, 9(11):4185–4208, 2016. doi: 10.5194/gmd-9-4185-2016. URL https://gmd.copernicus.org/articles/9/4185/2016/.

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

J Hansen, A Lacis, D Rind, G Russell, P Stone, I Fung, R Ruedy, and J Lerner. Climate sensitivity: Analysis of feedback mechanisms. *feedback*, 1:1–3, 1984.

John Harlim and Haizhao Yang. Diffusion forecasting model with basis functions from qr-decomposition. *Journal of Nonlinear Science*, 28(3):847–872, Jun 2018. ISSN 1432-1467. doi: 10.1007/s00332-017-9430-1. URL https://doi.org/10.1007/s00332-017-9430-1.

K. Hasselmann. Stochastic climate models part i. theory. *Tellus*, 28(6):473–485, 1976. doi: 10.3402/tellusa.v28i6.11316. URL https://doi.org/10.3402/tellusa.v28i6.11316.

P. H. Haynes, M. E. McIntyre, T. G. Shepherd, C. J. Marks, and K. P. Shine. On the "downward control" of extratropical diabatic circulations by eddy-induced mean zonal forces. *Journal of Atmospheric Sciences*, 48(4):651 – 678, 1991. doi: 10.1175/1520-0469(1991)048⟨0651:OTCOED⟩2.0.CO;2. URL https://journals.ametsoc.org/view/journals/atsc/48/4/1520-0469_1991_048_0651_otcoed_2_0_co_2.xml.

Sicheng He, Jing Yang, Qing Bao, Lei Wang, and Bin Wang. Fidelity of the observational/reanalysis datasets and global climate models in representation of extreme precipitation in east china. *Journal of Climate*, 32(1):195 – 212, 2019. doi: 10.1175/JCLI-D-18-0104.1. URL https://journals.ametsoc.org/view/journals/clim/32/1/jcli-d-18-0104.1.xml.

Isaac M. Held. The gap between simulation and understanding in climate modeling. *Bulletin of the American Meteorological Society*, 86(11):1609 – 1614, 2005. doi: 10.1175/BAMS-86-11-1609. URL https://journals.ametsoc.org/view/journals/bams/86/11/bams-86-11-1609.xml.

Luzie Helfmann, Enric Ribera Borrell, Christof Schütte, and Péter Koltai. Extending transition path theory: Periodically driven and finite-time dynamics. *Journal of Nonlinear Science*, 30(6):3321–3366, Dec 2020. ISSN 1432-1467. doi: 10.1007/s00332-020-09652-7. URL https://doi.org/10.1007/s00332-020-09652-7.

Luzie Helfmann, Jobst Heitzig, Péter Koltai, Jürgen Kurths, and Christof Schütte. Statistical analysis of tipping pathways in agent-based models. *The European Physical Journal Special Topics*, 230(16):3249–3271, Oct 2021. ISSN 1951-6401. doi: 10.1140/epjs/s11734-021-00191-0. URL https://doi.org/10.1140/epjs/s11734-021-00191-0.

Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, Adrian Simmons, Cornel Soci, Saleh Abdalla, Xavier Abellan, Gianpaolo Balsamo, Peter Bechtold, Gionata Biavati, Jean Bidlot, Massimo Bonavita, Giovanna De Chiara, Per Dahlgren, Dick Dee, Michail Diamantakis, Rossana Dragani, Johannes Flemming, Richard Forbes, Manuel Fuentes, Alan Geer, Leo Haimberger, Sean Healy, Robin J. Hogan, Elías Hólm, Marta Janisková, Sarah Keeley, Patrick Laloyaux, Philippe Lopez, Cristina Lupu, Gabor Radnoti, Patricia de Rosnay, Iryna Rozum, Freja Vamborg, Sebastien Villaume, and Jean-Noël Thépaut. The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, 2020. doi: https://doi.org/10.1002/qj.3803. URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3803.

Peter Hitchcock and Isla R. Simpson. The downward influence of stratospheric sudden warmings. *Journal of the Atmospheric Sciences*, 71(10):3856 – 3876, 2014. doi: 10.1175/JAS-D-14-0012.1. URL https://journals.ametsoc.org/view/journals/atsc/71/10/jas-d-14-0012.1.xml.

R. N. Hoffman, J. M. Henderson, S. M. Leidner, C. Grassotti, and T. Nehrkorn. 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, 2006. doi: 10.1175/JAS3720.1. URL https://journals.ametsoc.org/view/journals/atsc/63/7/jas3720.1.xml.

James R. Holton and Clifford Mass. Stratospheric vacillation cycles. *Journal of Atmospheric Sciences*, 33(11):2218 – 2225, 1976. doi: 10.1175/1520-0469(1976)033⟨2218:SVC⟩2.0.CO;2. URL https://journals.ametsoc.org/view/journals/atsc/33/11/1520-0469_1976_033_2218_svc_2_0_co_2.xml.

Matthew F. Horan and Thomas Reichler. Modeling seasonal sudden stratospheric warming climatology based on polar vortex statistics. *Journal of Climate*, 30(24):10101 – 10116, 2017. doi: 10.1175/JCLI-D-17-0257.1. URL https://journals.ametsoc.org/view/journals/clim/30/24/jcli-d-17-0257.1.xml.

Guannan Hu, Tamás Bódai, and Valerio Lucarini. Effects of stochastic parametrization on extreme value statistics. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(8):083102, 2019. doi: 10.1063/1.5095756. URL https://doi.org/10.1063/1.5095756.

Chris Huntingford, Terry Marsh, Adam A. Scaife, Elizabeth J. Kendon, Jamie Hannaford, Alison L. Kay, Mike Lockwood, Christel Prudhomme, Nick S. Reynard, Simon Parry, Jason A. Lowe, James A. Screen, Helen C. Ward, Malcolm Roberts, Peter A. Stott, Vicky A. Bell, Mark Bailey, Alan Jenkins, Tim Legg, Friederike E. L. Otto, Neil Massey, Nathalie Schaller, Julia Slingo, and Myles R. Allen. Potential influences on the united kingdom's floods of winter 2013/14. *Nature Climate Change*, 4(9):769–777, Sep 2014. ISSN 1758-6798. doi: 10.1038/nclimate2314. URL https://doi.org/10.1038/nclimate2314.

Masaru Inatsu, Naoto Nakano, Seiichiro Kusuoka, and Hitoshi Mukougawa. Predictability of wintertime stratospheric circulation examined using a nonstationary fluctuation–dissipation relation. *Journal of the Atmospheric Sciences*, 72(2):774 – 786, 2015. doi: 10.1175/JAS-D-14-0088.1. URL https://journals.ametsoc.org/view/journals/atsc/72/2/jas-d-14-0088.1.xml.

Guha Jayachandran, V. Vishal, and Vijay S. Pande. Using massively parallel simulation and markovian models to study protein folding: Examining the dynamics of the villin headpiece. *The Journal of Chemical Physics*, 124(16):164902, 2006. doi: 10.1063/1.2186317. URL https://doi.org/10.1063/1.2186317.

Renzhi Jing and Ning Lin. An environment-dependent probabilistic tropical cyclone model. *Journal of Advances in Modeling Earth Systems*, 12(3):e2019MS001975, 2020. doi: https://doi.org/

10.1029/2019MS001975. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001975. e2019MS001975 2019MS001975.

Martin Jucker. Are sudden stratospheric warmings generic? insights from an idealized gcm. *Journal of the Atmospheric Sciences*, 73(12):5061 – 5080, 2016. doi: 10.1175/JAS-D-15-0353.1. URL https://journals.ametsoc.org/view/journals/atsc/73/12/jas-d-15-0353.1.xml.

Mason Kamb, Eurika Kaiser, Steven L. Brunton, and J. Nathan Kutz. Time-delay observables for koopman: Theory and applications. *SIAM Journal on Applied Dynamical Systems*, 19(2): 886–917, 2020. doi: 10.1137/18M1216572. URL https://doi.org/10.1137/18M1216572.

Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer, 1998.

Alexey Yu. Karpechko, Peter Hitchcock, Dieter H. W. Peters, and Andrea Schneidereit. Predictability of downward propagation of major sudden stratospheric warmings. *Quarterly Journal of the Royal Meteorological Society*, 143(704):1459–1470, 2017. doi: https://doi.org/10.1002/qj.3017. URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3017.

K. Kashinath, M. Mustafa, A. Albert, J-L. Wu, C. Jiang, S. Esmaeilzadeh, K. Azizzadenesheli, R. Wang, A. Chattopadhyay, A. Singh, A. Manepalli, D. Chirila, R. Yu, R. Walters, B. White, H. Xiao, H. A. Tchelepi, P. Marcus, A. Anandkumar, P. Hassanzadeh, and null Prabhat. Physics-informed machine learning: case studies for weather and climate modelling. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379(2194):20200093, 2021. doi: 10.1098/rsta.2020.0093. URL https://royalsocietypublishing.org/doi/abs/10.1098/rsta.2020.0093.

T. Kelder, M. Müller, L. J. Slater, T. I. Marjoribanks, R. L. Wilby, C. Prudhomme, P. Bohlinger, L. Ferranti, and T. Nipen. Using unseen trends to detect decadal changes in 100-year precipitation extremes. *npj Climate and Atmospheric Science*, 3(1):47, Nov 2020. ISSN 2397-3722. doi: 10.1038/s41612-020-00149-4. URL https://doi.org/10.1038/s41612-020-00149-4.

Yuehaw Khoo, Jianfeng Lu, and Lexing Ying. Solving for high-dimensional committor functions using artificial neural networks. *Research in the Mathematical Sciences*, 6(1):1, Oct 2018. ISSN 2197-9847. doi: 10.1007/s40687-018-0160-2. URL https://doi.org/10.1007/s40687-018-0160-2.

Joseph Kidston, Adam A. Scaife, Steven C. Hardiman, Daniel M. Mitchell, Neal Butchart, Mark P. Baldwin, and Lesley J. Gray. Stratospheric influence on tropospheric jet streams, storm tracks and surface weather. *Nature Geoscience*, 8(6):433–440, Jun 2015. ISSN 1752-0908. doi: 10.1038/ngeo2424. URL https://doi.org/10.1038/ngeo2424.

Hye-Mi Kim, Peter J. Webster, Violeta E. Toma, and Daehyun Kim. Predictability and prediction skill of the mjo in two operational forecasting systems. *Journal of Climate*, 27(14):5364 – 5378, 2014. doi: 10.1175/JCLI-D-13-00480.1. URL https://journals.ametsoc.org/view/journals/clim/27/14/jcli-d-13-00480.1.xml.

Junsu Kim, Seok-Woo Son, Edwin P. Gerber, and Hyo-Seok Park. Defining sudden stratospheric warming in climate models: Accounting for biases in model climatologies. *Journal of Climate*, 30(14):5529 – 5546, 2017. doi: 10.1175/JCLI-D-16-0465.1. URL https://journals.ametsoc.org/view/journals/clim/30/14/jcli-d-16-0465.1.xml.

Sookyung Kim, Hyojin Kim, Joonseok Lee, Sangwoong Yoon, Samira Ebrahimi Kahou, Karthik Kashinath, and Mr Prabhat. Deep-hurricane-tracker: Tracking and forecasting extreme climate events. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1761–1769, 2019. doi: 10.1109/WACV.2019.00192.

Andrew D. King, Amy H. Butler, Martin Jucker, Nick O. Earl, and Irina Rudeva. Observed relationships between sudden stratospheric warmings and european climate extremes. *Journal of Geophysical Research: Atmospheres*, 124(24):13943–13961, 2019. doi: https://doi.org/10.1029/2019JD030480. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019JD030480.

Vassili Kitsios and Jorgen S. Frederiksen. Subgrid parameterizations of the eddy–eddy, eddy–mean field, eddy–topographic, mean field–mean field, and mean field–topographic interactions in atmospheric models. *Journal of the Atmospheric Sciences*, 76(2):457 – 477, 2019. doi: 10.1175/JAS-D-18-0255.1. URL https://journals.ametsoc.org/view/journals/atsc/76/2/jas-d-18-0255.1.xml.

Stefan Klus, Feliks Nüske, Péter Koltai, Hao Wu, Ioannis Kevrekidis, Christof Schütte, and Frank Noé. Data-driven model reduction and transfer operator approximation. *Journal of Nonlinear Science*, 28(3):985–1010, Jun 2018. ISSN 1432-1467. doi: 10.1007/s00332-017-9437-7. URL https://doi.org/10.1007/s00332-017-9437-7.

Erik W. Kolstad, Tarjei Breiteig, and Adam A. Scaife. The association between stratospheric weak polar vortex events and cold air outbreaks in the northern hemisphere. *Quarterly Journal of the Royal Meteorological Society*, 136(649):886–893, 2010. doi: https://doi.org/10.1002/qj.620. URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.620.

Marlene Kretschmer, Judah Cohen, Vivien Matthias, Jakob Runge, and Dim Coumou. The different stratospheric influence on cold-extremes in eurasia and north america. *npj Climate and Atmospheric Science*, 1(1):44, Nov 2018a. ISSN 2397-3722. doi: 10.1038/s41612-018-0054-4. URL https://doi.org/10.1038/s41612-018-0054-4.

Marlene Kretschmer, Dim Coumou, Laurie Agel, Mathew Barlow, Eli Tziperman, and Judah Cohen. More-persistent weak stratospheric polar vortex states linked to cold extremes. *Bulletin of the American Meteorological Society*, 99(1):49 – 60, 2018b. doi: 10.1175/BAMS-D-16-0259.1. URL https://journals.ametsoc.org/view/journals/bams/99/1/bams-d-16-0259.1.xml.

Wolfgang Kron, Petra Löw, and Zbigniew W. Kundzewicz. Changes in risk of extreme weather events in europe. *Environmental Science & Policy*, 100:74–83, 2019. ISSN 1462-9011. doi: https://doi.org/10.1016/j.envsci.2019.06.007. URL https://www.sciencedirect.com/science/article/pii/S146290111930142X.

Karin Labitzke. Stratospheric-mesospheric midwinter disturbances: A summary of observed characteristics. *Journal of Geophysical Research: Oceans*, 86(C10):9665–9678, 1981. doi: https://doi.org/10.1029/JC086iC10p09665. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/JC086iC10p09665.

Eric M. Laflamme, Ernst Linder, and Yibin Pan. Statistical downscaling of regional climate model output to achieve projections of precipitation extremes. *Weather and Climate Extremes*, 12: 15–23, 2016. ISSN 2212-0947. doi: https://doi.org/10.1016/j.wace.2015.12.001. URL https://www.sciencedirect.com/science/article/pii/S221209471530058X.

Andrea L. Lang, Kathleen Pegion, and Elizabeth A. Barnes. Introduction to special collection: "bridging weather and climate: Subseasonal-to-seasonal (s2s) prediction". *Journal of Geophysical Research: Atmospheres*, 125(4):e2019JD031833, 2020. doi: https://doi.org/10.1029/2019JD031833. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019JD031833. e2019JD031833 2019JD031833.

A. R. Lawrence, M. Leutbecher, and T. N. Palmer. The characteristics of hessian singular vectors using an advanced data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, 135(642):1117–1132, 2009. doi: https://doi.org/10.1002/qj.447. URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.447.

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

B. Legras and M. Ghil. Persistent anomalies, blocking and variations in atmospheric predictability. *Journal of Atmospheric Sciences*, 42(5):433 – 471, 1985. 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.

Corey Lesk, Pedram Rowhani, and Navin Ramankutty. Influence of extreme weather disasters on global crop production. *Nature*, 529(7584):84–87, Jan 2016. ISSN 1476-4687. doi: 10.1038/nature16467. URL https://doi.org/10.1038/nature16467.

Martin Leutbecher. On ensemble prediction using singular vectors started from forecasts. *Monthly Weather Review*, 133(10):3038 – 3046, 2005. doi: 10.1175/MWR3018.1. URL https://journals.ametsoc.org/view/journals/mwre/133/10/mwr3018.1.xml.

Haoya Li, Yuehaw Khoo, Yinuo Ren, and Lexing Ying. Solving for high dimensional committor functions using neural network with online approximation to derivatives. *arXiv preprint arXiv:2012.06727*, 2020.

Qianxiao Li, Bo Lin, and Weiqing Ren. Computing committor functions for the study of rare events using deep learning. *The Journal of Chemical Physics*, 151(5):054112, 2019. doi: 10.1063/1.5110439. URL https://doi.org/10.1063/1.5110439.

Varavut Limpasuvan, David W. J. Thompson, and Dennis L. Hartmann. The life cycle of the northern hemisphere sudden stratospheric warmings. *Journal of Climate*, 17(13):2584 – 2596, 2004. doi: 10.1175/1520-0442(2004)017⟨2584:TLCOTN⟩2.0.CO;2. URL https://journals.ametsoc.org/view/journals/clim/17/13/1520-0442_2004_017_2584_tlcotn_2.0.co_2.xml.

Kevin K. Lin and Fei Lu. Data-driven model reduction, wiener projections, and the koopman-mori-zwanzig formalism. *Journal of Computational Physics*, 424:109864, 2021. ISSN 0021-9991. doi: https://doi.org/10.1016/j.jcp.2020.109864. URL https://www.sciencedirect.com/science/article/pii/S0021999120306380.

Yuanchao Liu, David P. Hickey, Shelley D. Minteer, Alex Dickson, and Scott Calabrese Barton. Markov-state transition path analysis of electrostatic channeling. *The Journal of Physical Chemistry C*, 123(24):15284–15292, Jun 2019. ISSN 1932-7447. doi: 10.1021/acs.jpcc.9b02844. URL https://doi.org/10.1021/acs.jpcc.9b02844.

Edward N. Lorenz. Deterministic nonperiodic flow. *Journal of Atmospheric Sciences*, 20(2):130 – 141, 1963. 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_020_0130_dnf_2_0_co_2.xml.

Chatipat Lorpaiboon, Erik Henning Thiede, Robert J. Webber, Jonathan Weare, and Aaron R. Dinner. 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, Oct 2020. ISSN 1520-6106. doi: 10.1021/acs.jpcb.0c06477. URL https://doi.org/10.1021/acs.jpcb.0c06477.

Jianfeng Lu and Eric Vanden-Eijnden. Exact dynamical coarse-graining without time-scale separation. *The Journal of Chemical Physics*, 141(4):044109, 2014. doi: 10.1063/1.4890367. URL https://doi.org/10.1063/1.4890367.

Valerio Lucarini and Andrey Gritsun. A new mathematical framework for atmospheric blocking events. *Climate Dynamics*, 54(1):575–598, Jan 2020. ISSN 1432-0894. doi: 10.1007/s00382-019-05018-2. URL https://doi.org/10.1007/s00382-019-05018-2.

Valerio Lucarini, Davide Faranda, Jorge Miguel Milhazes de Freitas, Mark Holland, Tobias Kuna, Matthew Nicol, Mike Todd, Sandro Vaienti, et al. *Extremes and recurrence in dynamical systems*. John Wiley & Sons, 2016.

Dario Lucente, Stefan Duffner, Corentin Herbert, Joran Rolland, and Freddy Bouchet. Machine learning of committor functions for predicting high impact climate events. 2019. doi: 10.48550/ARXIV.1910.11736. URL https://arxiv.org/abs/1910.11736.

Dario Lucente, Corentin Herbert, and Freddy Bouchet. Committor functions for climate phenomena at the predictability margin: The example of el niño southern oscillation in the jin and timmerman model, 2021a. URL https://arxiv.org/abs/2106.14990.

Dario Lucente, Joran Rolland, Corentin Herbert, and Freddy Bouchet. Coupling rare event algorithms with data-based learned committor functions using the analogue markov chain, 2021b.

Chiara Cecilia Maiocchi, Valerio Lucarini, and Andrey Gritsun. Decomposing the dynamics of the lorenz 1963 model using unstable periodic orbits: Averages, transitions, and quasi-invariant sets. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 32(3):033129, 2022. doi: 10.1063/5.0067673. URL https://doi.org/10.1063/5.0067673.

Andrew J. Majda and Di Qi. Strategies for reduced-order models for predicting the statistical responses and uncertainty quantification in complex turbulent dynamical systems. *SIAM Review*, 60(3):491–549, 2018. doi: 10.1137/16M1104664. URL https://doi.org/10.1137/16M1104664.

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

Michael E. Mann, Stefan Rahmstorf, Kai Kornhuber, Byron A. Steinman, Sonya K. Miller, and Dim Coumou. Influence of anthropogenic climate change on planetary wave resonance and extreme weather events. *Scientific Reports*, 7(1):45242, Mar 2017. ISSN 2045-2322. doi: 10.1038/srep45242. URL https://doi.org/10.1038/srep45242.

Andreas Mardt, Luca Pasquali, Hao Wu, and Frank Noé. Vampnets for deep learning of molecular kinetics. *Nature Communications*, 9(1):5, Jan 2018. ISSN 2041-1723. doi: 10.1038/s41467-017-02388-1. URL https://doi.org/10.1038/s41467-017-02388-1.

O. Martius, L. M. Polvani, and H. C. Davies. Blocking precursors to stratospheric sudden warming events. *Geophysical Research Letters*, 36(14), 2009. doi: https://doi.org/10.1029/2009GL038776. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2009GL038776.

Taroh Matsuno. Vertical propagation of stationary planetary waves in the winter northern hemisphere. *Journal of Atmospheric Sciences*, 27(6):871 – 883, 1970. doi: 10.1175/1520-0469(1970)027⟨0871:VPOSPW⟩2.0.CO;2. URL https://journals.ametsoc.org/view/journals/atsc/27/6/1520-0469_1970_027_0871_vpospw_2_0_co_2.xml.

N. Joss Matthewman and J. G. Esler. Stratospheric sudden warmings as self-tuning resonances. part i: Vortex splitting events. *Journal of the Atmospheric Sciences*, 68(11):2481 – 2504, 2011. doi: 10.1175/JAS-D-11-07.1. URL https://journals.ametsoc.org/view/journals/atsc/68/11/jas-d-11-07.1.xml.

Amanda C. Maycock and Peter Hitchcock. Do split and displacement sudden stratospheric warmings have different annular mode signatures? *Geophysical Research Letters*, 42(24):10,943–10,951, 2015. doi: https://doi.org/10.1002/2015GL066754. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015GL066754.

Michael E. McIntyre. How well do we understand the dynamics of stratospheric warmings? *Journal of the Meteorological Society of Japan. Ser. II*, 60(1):37–65, 1982. doi: 10.2151/jmsj1965.60.1_37.

Yilin Meng, Diwakar Shukla, Vijay S. Pande, and Benoît Roux. Transition path theory analysis of c-src kinase activation. *Proceedings of the National Academy of Sciences*, 113(33):9193–9198, 2016. doi: 10.1073/pnas.1602790113. URL https://www.pnas.org/doi/abs/10.1073/pnas.1602790113.

John W. Meriwether and Andrew J. Gerrard. Mesosphere inversion layers and stratosphere temperature enhancements. *Reviews of Geophysics*, 42(3), 2004. doi: https://doi.org/10.1029/2003RG000133. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2003RG000133.

Philipp Metzner, Christof Schütte, and Eric Vanden-Eijnden. Illustration of transition path theory on a collection of simple examples. *The Journal of Chemical Physics*, 125(8):084110, 2006. doi: 10.1063/1.2335447. URL https://doi.org/10.1063/1.2335447.

Philipp Metzner, Christof Schütte, and Eric Vanden-Eijnden. Transition path theory for markov jump processes. *Multiscale Modeling & Simulation*, 7(3):1192–1219, 2009. doi: 10.1137/070699500. URL https://doi.org/10.1137/070699500.

Igor Mezić. Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dynamics*, 41(1):309–325, Aug 2005. ISSN 1573-269X. doi: 10.1007/s11071-005-2824-x. URL https://doi.org/10.1007/s11071-005-2824-x.

Igor Mezić. Analysis of fluid flows via spectral properties of the koopman operator. *Annual Review of Fluid Mechanics*, 45(1):357–378, 2013. doi: 10.1146/annurev-fluid-011212-140652. URL https://doi.org/10.1146/annurev-fluid-011212-140652.

P. Miron, F. J. Beron-Vera, L. Helfmann, and P. Koltai. Transition paths of marine debris and the stability of the garbage patches. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 31 (3):033101, 2021. doi: 10.1063/5.0030535. URL https://doi.org/10.1063/5.0030535.

Vimal Mishra and Harsh L. Shah. Hydroclimatological perspective of the kerala flood of 2018. *Journal of the Geological Society of India*, 92(5):645–650, Nov 2018. ISSN 0974-6889. doi: 10.1007/s12594-018-1079-3. URL https://doi.org/10.1007/s12594-018-1079-3.

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

Mustafa A. Mohamad and Themistoklis P. Sapsis. Sequential sampling strategy for extreme event statistics in nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 115(44):11138–11143, 2018. doi: 10.1073/pnas.1813263115. URL https://www.pnas.org/doi/abs/10.1073/pnas.1813263115.

R. Mureau, Franco Molteni, and T. N. Palmer. Ensemble prediction using dynamically conditioned perturbations. *Quarterly Journal of the Royal Meteorological Society*, 119(510):299–323, 1993. doi: https://doi.org/10.1002/qj.49711951005. URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.49711951005.

G. Myhre, K. Alterskjær, C. W. Stjern, Ø Hodnebrog, L. Marelle, B. H. Samset, J. Sillmann, N. Schaller, E. Fischer, M. Schulz, and A. Stohl. Frequency of extreme precipitation increases extensively with event rareness under global warming. *Scientific Reports*, 9(1):16063, Nov 2019.

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

Philippe Naveau, Alexis Hannart, and Aurélien Ribes. Statistical methods for extreme event attribution in climate science. *Annual Review of Statistics and Its Application*, 7(1):89–110, 2020. doi: 10.1146/annurev-statistics-031219-041314. URL https://doi.org/10.1146/annurev-statistics-031219-041314.

Chigomezyo M. Ngwira, Antti Pulkkinen, M. Leila Mays, Maria M. Kuznetsova, A. B. Galvin, Kristin Simunac, Daniel N. Baker, Xinlin Li, Yihua Zheng, and Alex Glocer. 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, 2013. doi: https://doi.org/10.1002/2013SW000990. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2013SW000990.

Frank Noé and Cecilia Clementi. Collective variables for the study of long-time kinetics from molecular trajectories: theory and methods. *Current Opinion in Structural Biology*, 43:141–147, 2017. ISSN 0959-440X. doi: https://doi.org/10.1016/j.sbi.2017.02.006. URL https://www.sciencedirect.com/science/article/pii/S0959440X17300301. Theory and simulation • Macromolecular assemblies.

Frank Noé and Stefan Fischer. Transition networks for modeling the kinetics of conformational change in macromolecules. *Current Opinion in Structural Biology*, 18(2):154–162, 2008. ISSN 0959-440X. doi: https://doi.org/10.1016/j.sbi.2008.01.008. URL https://www.sciencedirect.com/science/article/pii/S0959440X08000249. Theory and simulation / Macromolecular assemblages.

Frank Noé, Christof Schütte, Eric Vanden-Eijnden, Lothar Reich, and Thomas R. Weikl. Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proceedings of the National Academy of Sciences*, 106(45):19011–19016, 2009. doi: 10.1073/pnas.0905466106. URL https://www.pnas.org/doi/abs/10.1073/pnas.0905466106.

Travis A. O'Brien, William D. Collins, Karthik Kashinath, Oliver Rübel, Suren Byna, Junmin Gu, Hari Krishnan, and Paul A. Ullrich. Resolution dependence of precipitation statistical fidelity in hindcast simulations. *Journal of Advances in Modeling Earth Systems*, 8(2):976–990, 2016. doi: https://doi.org/10.1002/2016MS000671. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016MS000671.

Amee O'Callaghan, Manoj Joshi, David Stevens, and Daniel Mitchell. The effects of different sudden stratospheric warming types on the ocean. *Geophysical Research Letters*, 41(21): 7739–7745, 2014. doi: https://doi.org/10.1002/2014GL062179. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014GL062179.

Paul A. O'Gorman. Sensitivity of tropical precipitation extremes to climate change. *Nature Geoscience*, 5(10):697–700, Oct 2012. doi: 10.1038/ngeo1568. URL https://doi.org/10.1038/ngeo1568.

Paul A. O'Gorman and John G. Dwyer. Using machine learning to parameterize moist convection: Potential for modeling of climate, climate change, and extreme events. *Journal of Advances in Modeling Earth Systems*, 10(10):2548–2563, 2018. doi: https://doi.org/10.1029/2018MS001351. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018MS001351.

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

Paul A. O'Gorman and Tapio Schneider. Scaling of precipitation extremes over a wide range of climates simulated with an idealized gcm. *Journal of Climate*, 22(21):5676 – 5685, 2009. doi: 10.1175/2009JCLI2701.1. URL https://journals.ametsoc.org/view/journals/clim/22/21/2009jcli2701.1.xml.

T. N. Palmer, R. Gelaro, J. Barkmeijer, and R. Buizza. Singular vectors, metrics, and adaptive observations. *Journal of the Atmospheric Sciences*, 55(4):633 – 653, 1998. doi: 10.1175/1520-0469(1998)055⟨0633:SVMAAO⟩2.0.CO;2. URL https://journals.ametsoc.org/view/journals/atsc/55/4/1520-0469_1998_055_0633_svmaao_2.0.co_2.xml.

Tim N Palmer, Roberto Buizza, F Doblas-Reyes, Thomas Jung, Martin Leutbecher, Glenn J Shutts, Martin Steinheimer, and Antje Weisheimer. Stochastic parametrization and model uncertainty. *ECMWF Technical Memoranda*, 2009.

Vijay S. Pande, Kyle Beauchamp, and Gregory R. Bowman. Everything you wanted to know about markov state models but were afraid to ask. *Methods*, 52(1):99–105, 2010. ISSN 1046-2023. doi: https://doi.org/10.1016/j.ymeth.2010.06.002. URL https://www.sciencedirect.com/science/article/pii/S1046202310001568. Protein Folding.

Grigorios A Pavliotis. *Stochastic processes and applications: diffusion processes, the Fokker-Planck and Langevin equations*, volume 60. Springer, 2014.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830, 2011.

Cécile Penland and Prashant D. Sardeshmukh. The optimal growth of tropical sea surface temperature anomalies. *Journal of Climate*, 8(8):1999 – 2024, 1995. doi: 10.1175/1520-0442(1995)

008⟨1999:TOGOTS⟩2.0.CO;2. URL https://journals.ametsoc.org/view/journals/clim/8/8/1520-0442_1995_008_1999_togots_2_0_co_2.xml.

S. Pfahl, P. A. O'Gorman, and E. M. Fischer. Understanding the regional pattern of projected future changes in extreme precipitation. *Nature Climate Change*, 7(6):423–427, Jun 2017. ISSN 1758-6798. doi: 10.1038/nclimate3287. URL https://doi.org/10.1038/nclimate3287.

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

Paul Poli, Hans Hersbach, Dick P. Dee, Paul Berrisford, Adrian J. Simmons, Frédéric Vitart, Patrick Laloyaux, David G. H. Tan, Carole Peubey, Jean-Noël Thépaut, Yannick Trémolet, Elías V. Hólm, Massimo Bonavita, Lars Isaksen, and Michael Fisher. Era-20c: An atmospheric reanalysis of the twentieth century. *Journal of Climate*, 29(11):4083 – 4097, 2016. doi: 10.1175/JCLI-D-15-0556.1. URL https://journals.ametsoc.org/view/journals/clim/29/11/jcli-d-15-0556.1.xml.

PierGianLuca Porta Mana and Laure Zanna. Toward a stochastic parameterization of ocean mesoscale eddies. *Ocean Modelling*, 79:1–20, 2014. ISSN 1463-5003. doi: https://doi.org/10.1016/j.ocemod.2014.04.002. URL https://www.sciencedirect.com/science/article/pii/S1463500314000420.

F. Rabier, E. Klinker, P. Courtier, and A. Hollingsworth. Sensitivity of forecast errors to initial conditions. *Quarterly Journal of the Royal Meteorological Society*, 122(529):121–150, 1996. doi: https://doi.org/10.1002/qj.49712252906. URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.49712252906.

Francesco Ragone and Freddy Bouchet. Computation of extreme values of time averaged observables in climate models with large deviation techniques. *Journal of Statistical Physics*, 179(5):1637–1665, Jun 2020. ISSN 1572-9613. doi: 10.1007/s10955-019-02429-7. URL https://doi.org/10.1007/s10955-019-02429-7.

Francesco Ragone, Jeroen Wouters, and Freddy Bouchet. Computation of extreme heat waves in climate models using a large deviation algorithm. *Proceedings of the National Academy of Sciences*, 115(1):24–29, 2018. ISSN 0027-8424. doi: 10.1073/pnas.1712645115. URL https://www.pnas.org/content/115/1/24.

M. Raissi, P. Perdikaris, and G.E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019. ISSN 0021-9991. doi: https://doi.org/10.1016/j.jcp.2018.10.045. URL https://www.sciencedirect.com/science/article/pii/S0021999118307125.

Thomas Reichler, Junsu Kim, Elisa Manzini, and Jürgen Kröger. A stratospheric connection to atlantic climate variability. *Nature Geoscience*, 5(11):783–787, Nov 2012. ISSN 1752-0908. doi: 10.1038/ngeo1586. URL https://doi.org/10.1038/ngeo1586.

Grant M Rotskoff and Eric Vanden-Eijnden. Learning with rare data: using active importance sampling to optimize objectives dominated by rare events. *Preprint at arXiv https://arxiv.org/abs/2008.06334*, 2020.

Alexander Ruzmaikin, John Lawrence, and Cristina Cadavid. A simple model of stratospheric dynamics including solar variability. *Journal of Climate*, 16(10):1593 – 1600, 2003. doi: 10.1175/1520-0442(2003)016⟨1593:ASMOSD⟩2.0.CO;2. URL https://journals.ametsoc.org/view/journals/clim/16/10/1520-0442_2003_016_1593_asmosd_2.0.co_2.xml.

C. T. Sabeerali, R. S. Ajayamohan, Dimitrios Giannakis, and Andrew J. Majda. Extraction and prediction of indices for monsoon intraseasonal oscillations: an approach based on nonlinear laplacian spectral analysis. *Climate Dynamics*, 49(9):3031–3050, Nov 2017. ISSN 1432-0894. doi: 10.1007/s00382-016-3491-y. URL https://doi.org/10.1007/s00382-016-3491-y.

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

A. A. Scaife, A. Yu. Karpechko, M. P. Baldwin, A. Brookshaw, A. H. Butler, R. Eade, M. Gordon, C. MacLachlan, N. Martin, N. Dunstone, and D. Smith. Seasonal winter forecasts and the stratosphere. *Atmospheric Science Letters*, 17(1):51–56, 2016. doi: https://doi.org/10.1002/asl.598. URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/asl.598.

A. A. Scaife, M. P. Baldwin, A. H. Butler, A. J. Charlton-Perez, D. I. V. Domeisen, C. I. Garfinkel, S. C. Hardiman, P. Haynes, A. Y. Karpechko, E.-P. Lim, S. Noguchi, J. Perlwitz, L. Polvani, J. H. Richter, J. Scinocca, M. Sigmond, T. G. Shepherd, S.-W. Son, and D. W. J. Thompson. Long-range prediction and the stratosphere. *Atmospheric Chemistry and Physics*, 22(4):2601–2623, 2022. doi: 10.5194/acp-22-2601-2022. URL https://acp.copernicus.org/articles/22/2601/2022/.

N Schaller, J Sillmann, J Anstey, E M Fischer, C M Grams, and S Russo. Influence of blocking on northern european and western russian heatwaves in large climate model ensembles. *Environmental Research Letters*, 13(5):054015, may 2018. doi: 10.1088/1748-9326/aaba55. URL https://doi.org/10.1088/1748-9326/aaba55.

Gavin A Schmidt. The physics of climate modeling. *Phys. Today*, 60(1):72–73, 2007.

R. K. Scott and L. M. Polvani. Internal variability of the winter stratosphere. part i: Time-independent forcing. *Journal of the Atmospheric Sciences*, 63(11):2758 – 2776, 2006. doi: 10.1175/JAS3797.1. URL https://journals.ametsoc.org/view/journals/atsc/63/11/jas3797.1.xml.

R. K. Scott, L. M. Polvani, and D. W. Waugh. Internal variability of the winter stratosphere. part ii: Time-dependent forcing. *Journal of the Atmospheric Sciences*, 65(7):2375 – 2388, 2008. doi: 10.1175/2007JAS2619.1. URL https://journals.ametsoc.org/view/journals/atsc/65/7/2007jas2619.1.xml.

Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27 (3):379–423, 1948.

Theodore G. Shepherd, Emily Boyd, Raphael A. Calel, Sandra C. Chapman, Suraje Dessai, Ioana M. Dima-West, Hayley J. Fowler, Rachel James, Douglas Maraun, Olivia Martius, Catherine A. Senior, Adam H. Sobel, David A. Stainforth, Simon F. B. Tett, Kevin E. Trenberth, Bart J. J. M. van den Hurk, Nicholas W. Watkins, Robert L. Wilby, and Dimitri A. Zenghelis. Storylines: an alternative approach to representing uncertainty in physical aspects of climate change. *Climatic Change*, 151(3):555–571, Dec 2018. ISSN 1573-1480. doi: 10.1007/s10584-018-2317-9. URL https://doi.org/10.1007/s10584-018-2317-9.

M. Sigmond, J. F. Scinocca, V. V. Kharin, and T. G. Shepherd. Enhanced seasonal forecast skill following stratospheric sudden warmings. *Nature Geoscience*, 6(2):98–102, Feb 2013. ISSN 1752-0908. doi: 10.1038/ngeo1698. URL https://doi.org/10.1038/ngeo1698.

Michael Sigmond and John F. Scinocca. The influence of the basic state on the northern hemisphere circulation response to climate change. *Journal of Climate*, 23(6):1434 – 1446, 2010. doi: 10.1175/2009JCLI3167.1. URL https://journals.ametsoc.org/view/journals/clim/23/6/2009jcli3167.1.xml.

Jana Sillmann, Thordis Thorarinsdottir, Noel Keenlyside, Nathalie Schaller, Lisa V. Alexander, Gabriele Hegerl, Sonia I. Seneviratne, Robert Vautard, Xuebin Zhang, and Francis W. Zwiers. Understanding, modeling and predicting weather and climate extremes: Challenges and opportunities. *Weather and Climate Extremes*, 18:65–74, 2017. ISSN 2212-0947. doi: https://doi.org/10.1016/j.wace.2017.10.003. URL https://www.sciencedirect.com/science/article/pii/S2212094717300440.

Jana Sillmann, Theodore G. Shepherd, Bart van den Hurk, Wilco Hazeleger, Olivia Martius, Julia Slingo, and Jakob Zscheischler. Event-based storylines to address climate risk. *Earth's Future*, 9(2):e2020EF001783, 2021. doi: https://doi.org/10.1029/2020EF001783. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020EF001783. e2020EF001783 2020EF001783.

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

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

Jeremiah P. Sjoberg and Thomas Birner. Transient tropospheric forcing of sudden stratospheric warmings. *Journal of the Atmospheric Sciences*, 69(11):3420 – 3432, 2012. doi: 10.1175/JAS-D-11-0195.1. URL https://journals.ametsoc.org/view/journals/atsc/69/11/jas-d-11-0195.1.xml.

Jeremiah P. Sjoberg and Thomas Birner. Stratospheric wave–mean flow feedbacks and sudden stratospheric warmings in a simple model forced by upward wave activity flux. *Journal of the Atmospheric Sciences*, 71(11):4055 – 4071, 2014. doi: 10.1175/JAS-D-14-0113.1. URL https://journals.ametsoc.org/view/journals/atsc/71/11/jas-d-14-0113.1.xml.

Adam H. Sobel, Allison A. Wing, Suzana J. Camargo, Christina M. Patricola, Gabriel A. Vecchi, Chia-Ying Lee, and Michael K. Tippett. Tropical cyclone frequency. *Earth's Future*, 9(12):e2021EF002275, 2021. doi: https://doi.org/10.1029/2021EF002275. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021EF002275. e2021EF002275 2021EF002275.

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

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

Floris Takens. Detecting strange attractors in turbulence. In *Dynamical systems and turbulence, Warwick 1980*, pages 366–381. Springer, 1981.

Alexis Tantet, Fiona R. van der Burgt, and Henk A. Dijkstra. An early warning indicator for atmospheric blocking events using transfer operators. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 25(3):036406, 2015. doi: 10.1063/1.4908174. URL https://doi.org/10.1063/1.4908174.

Erik H. Thiede, Dimitrios Giannakis, Aaron R. Dinner, and Jonathan Weare. Galerkin approximation of dynamical quantities using trajectory data. *The Journal of Chemical Physics*, 150(24):244111, 2019. doi: 10.1063/1.5063730. URL https://doi.org/10.1063/1.5063730.

David W. J. Thompson, Mark P. Baldwin, and John M. Wallace. Stratospheric connection to northern hemisphere wintertime weather: Implications for prediction. *Journal of Climate*, 15(12):1421 – 1428, 2002. doi: 10.1175/1520-0442(2002)015⟨1421:SCTNHW⟩2.0.CO;2. URL https://journals.ametsoc.org/view/journals/clim/15/12/1520-0442_2002_015_1421_sctnhw_2.0.co_2.xml.

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996. 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.

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

Anna Trevisan, Francesco Pancotti, and Franco Molteni. Ensemble prediction in a model with flow regimes. *Quarterly Journal of the Royal Meteorological Society*, 127(572):343–358, 2001. doi: https://doi.org/10.1002/qj.49712757206. URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.49712757206.

Om P. Tripathi, Mark Baldwin, Andrew Charlton-Perez, Martin Charron, Jacob C. H. Cheung, Stephen D. Eckermann, Edwin Gerber, David R. Jackson, Yuhji Kuroda, Andrea Lang, Justin McLay, Ryo Mizuta, Carolyn Reynolds, Greg Roff, Michael Sigmond, Seok-Woo Son, and Tim Stockdale. Examining the predictability of the stratospheric sudden warming of january 2013 using multiple nwp systems. *Monthly Weather Review*, 144(5):1935 – 1960, 2016. doi: 10.1175/MWR-D-15-0010.1. URL https://journals.ametsoc.org/view/journals/mwre/144/5/mwr-d-15-0010.1.xml.

Kai-Chih Tseng, Nathaniel C. Johnson, Eric D. Maloney, Elizabeth A. Barnes, and Sarah B. Kapnick. Mapping large-scale climate variability to hydrological extremes: An application of the linear inverse model to subseasonal prediction. *Journal of Climate*, 34(11):4207 – 4225, 2021. doi: 10.1175/JCLI-D-20-0502.1. URL https://journals.ametsoc.org/view/journals/clim/34/11/JCLI-D-20-0502.1.xml.

H. M. van den Dool. A new look at weather forecasting through analogues. *Monthly Weather Review*, 117(10):2230 – 2247, 1989. doi: 10.1175/1520-0493(1989)117⟨2230:ANLAWF⟩2.0.CO;2. URL https://journals.ametsoc.org/view/journals/mwre/117/10/1520-0493_1989_117_2230_anlawf_2_0_co_2.xml.

Geert Jan Van Oldenborgh, Karin Van Der Wiel, Antonia Sebastian, Roop Singh, Julie Arrighi, Friederike Otto, Karsten Haustein, Sihan Li, Gabriel Vecchi, and Heidi Cullen. Attribution of extreme rainfall from hurricane harvey, august 2017. *Environmental Research Letters*, 12(12):124009, 2017.

E. Vanden-Eijnden. *Transition Path Theory*, pages 453–493. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006. ISBN 978-3-540-35273-0. doi: 10.1007/3-540-35273-2_13. URL https://doi.org/10.1007/3-540-35273-2_13.

Eric Vanden-Eijnden. *Transition path theory*, pages 91–100. Advances in Experimental Medicine and Biology. Springer New York LLC, 2014. ISBN 9789400776050. doi: 10.1007/978-94-7-7606-7_7.

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

MWR-D-12-00060.1. URL https://journals.ametsoc.org/view/journals/mwre/141/6/mwr-d-12-00060.1.xml.

F. Vitart, C. Ardilouze, A. Bonet, A. Brookshaw, M. Chen, C. Codorean, M. Déqué, L. Ferranti, E. Fucile, M. Fuentes, H. Hendon, J. Hodgson, H.-S. Kang, A. Kumar, H. Lin, G. Liu, X. Liu, P. Malguzzi, I. Mallas, M. Manoussakis, D. Mastrangelo, C. MacLachlan, P. McLean, A. Minami, R. Mladek, T. Nakazawa, S. Najm, Y. Nie, M. Rixen, A. W. Robertson, P. Ruti, C. Sun, Y. Takaya, M. Tolstykh, F. Venuti, D. Waliser, S. Woolnough, T. Wu, D.-J. Won, H. Xiao, R. Zaripov, and L. Zhang. The subseasonal to seasonal (s2s) prediction project database. *Bulletin of the American Meteorological Society*, 98(1):163 – 173, 2017. doi: 10.1175/BAMS-D-16-0017.1. URL https://journals.ametsoc.org/view/journals/bams/98/1/bams-d-16-0017.1.xml.

Frédéric Vitart and Andrew W. Robertson. The sub-seasonal to seasonal prediction project (s2s) and the prediction of extreme events. *npj Climate and Atmospheric Science*, 1(1):3, Mar 2018. ISSN 2397-3722. doi: 10.1038/s41612-018-0013-0. URL https://doi.org/10.1038/s41612-018-0013-0.

Frédéric Vitart. Evolution of ecmwf sub-seasonal forecast skill scores. *Quarterly Journal of the Royal Meteorological Society*, 140(683):1889–1899, 2014. doi: https://doi.org/10.1002/qj.2256. URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2256.

Zhong Yi Wan, Pantelis Vlachas, Petros Koumoutsakos, and Themistoklis Sapsis. Data-assisted reduced-order modeling of extreme events in complex dynamical systems. *PLOS ONE*, 13 (5):1–22, 05 2018. doi: 10.1371/journal.pone.0197704. URL https://doi.org/10.1371/journal.pone.0197704.

L Wang, S C Hardiman, P E Bett, R E Comer, C Kent, and A A Scaife. What chance of a sudden stratospheric warming in the southern hemisphere? *Environmental Research Letters*, 15 (10):104038, sep 2020a. doi: 10.1088/1748-9326/aba8c1. URL https://doi.org/10.1088/1748-9326/aba8c1.

Xinyang Wang, Joanna Slawinska, and Dimitrios Giannakis. Extended-range statistical enso prediction through operator-theoretic techniques for nonlinear dynamics. *Scientific Reports*, 10 (1):2636, Feb 2020b. ISSN 2045-2322. doi: 10.1038/s41598-020-59128-7. URL https://doi.org/10.1038/s41598-020-59128-7.

Larry Wasserman. *All of statistics*. Springer, New York, 2004.

Jonathan Weare. Particle filtering with path sampling and an application to a bimodal ocean current model. *Journal of Computational Physics*, 228(12):4312–4331, 2009. ISSN 0021-9991. doi: https://doi.org/10.1016/j.jcp.2009.02.033. URL https://www.sciencedirect.com/science/article/pii/S0021999109000801.

Robert J. Webber, David A. Plotkin, Morgan E O'Neill, Dorian S. Abbot, and Jonathan Weare. Practical rare event sampling for extreme mesoscale weather. *Chaos: An Interdisciplinary*

Journal of Nonlinear Science, 29(5):053109, 2019. doi: 10.1063/1.5081461. URL https://doi.org/10.1063/1.5081461.

Christopher J. White, Henrik Carlsen, Andrew W. Robertson, Richard J.T. Klein, Jeffrey K. Lazo, Arun Kumar, Frederic Vitart, Erin Coughlan de Perez, Andrea J. Ray, Virginia Murray, Sukaina Bharwani, Dave MacLeod, Rachel James, Lora Fleming, Andrew P. Morse, Bernd Eggen, Richard Graham, Erik Kjellström, Emily Becker, Kathleen V. Pegion, Neil J. Holbrook, Darryn McEvoy, Michael Depledge, Sarah Perkins-Kirkpatrick, Timothy J. Brown, Roger Street, Lindsey Jones, Tomas A. Remenyi, Indi Hodgson-Johnston, Carlo Buontempo, Rob Lamb, Holger Meinke, Berit Arheimer, and Stephen E. Zebiak. Potential applications of subseasonal-to-seasonal (s2s) predictions. *Meteorological Applications*, 24(3):315–325, 2017. doi: https://doi.org/10.1002/met.1654. URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/met.1654.

T. M. L. Wigley. The effect of changing climate on the frequency of absolute extreme events. *Climatic Change*, 97(1):67, Aug 2009. ISSN 1573-1480. doi: 10.1007/s10584-009-9654-7. URL https://doi.org/10.1007/s10584-009-9654-7.

Yuki Yasuda, Freddy Bouchet, and Antoine Venaille. A new interpretation of vortex-split sudden stratospheric warmings in terms of equilibrium statistical mechanics. *Journal of the Atmospheric Sciences*, 74(12):3915 – 3936, 2017. doi: 10.1175/JAS-D-17-0045.1. URL https://journals.ametsoc.org/view/journals/atsc/74/12/jas-d-17-0045.1.xml.

Shigeo Yoden. Bifurcation properties of a stratospheric vacillation model. *Journal of Atmospheric Sciences*, 44(13):1723 – 1733, 1987a. doi: 10.1175/1520-0469(1987)044⟨1723:BPOASV⟩2.0.CO;2. URL https://journals.ametsoc.org/view/journals/atsc/44/13/1520-0469_1987_044_1723_bpoasv_2_0_co_2.xml.

Shigeo Yoden. Dynamical aspects of stratospheric vacillations in a highly truncated model. *Journal of Atmospheric Sciences*, 44(24):3683 – 3695, 1987b. doi: 10.1175/1520-0469(1987)044⟨3683:DAOSVI⟩2.0.CO;2. URL https://journals.ametsoc.org/view/journals/atsc/44/24/1520-0469_1987_044_3683_daosvi_2_0_co_2.xml.

Janni Yuval, Paul A. O'Gorman, and Chris N. Hill. Use of neural networks for stable, accurate and physically consistent parameterization of subgrid atmospheric processes with good performance at reduced precision. *Geophysical Research Letters*, 48(6):e2020GL091363, 2021. doi: https://doi.org/10.1029/2020GL091363. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020GL091363. e2020GL091363 2020GL091363.

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

Robert Zwanzig. *Nonequilibrium statistical mechanics*. Oxford university press, 2001.