Revealing the statistics of extreme events hidden in short weather forecast data

Justin Finkel[1], Edwin P. Gerber[2], Dorian S. Abbot[3], Jonathan Weare[2]

[1]Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology
[2]Courant Institute of Mathematical Sciences, New York University
[3]Department of Geophysical Sciences, University of Chicago

Key Points:

- Extreme weather risk is inherently difficult to quantify because of data scarcity.
- Subseasonal weather forecast ensembles are an untapped resource for computing statistics of extremes, done here by statistical weighting.
- We characterize stratospheric extremes up to once in 500-year events from 46-day hindcast ensembles across 20 winters.

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

Abstract

Extreme weather events have significant consequences, dominating the impact of climate on society. While high-resolution weather models can forecast many types of extreme events on synoptic timescales, long-term climatological risk assessment is an altogether different problem. A once-in-a-century event takes, on average, 100 years of simulation time to appear just once, far beyond the typical integration length of a weather forecast model. Therefore, this task is left to cheaper, but less accurate, low-resolution or statistical models. But there is untapped potential in weather model output: despite being short in duration, weather forecast ensembles are produced multiple times a week. Integrations are launched with independent perturbations, causing them to spread apart over time and broadly sample phase space. Collectively, these integrations add up to thousands of years of data. We establish methods to extract climatological information from these short weather simulations. Using ensemble hindcasts by the European Center for Medium-range Weather Forecasting (ECMWF) archived in the subseasonal-to-seasonal (S2S) database, we characterize sudden stratospheric warming (SSW) events with multi-centennial return times. Consistent results are found between alternative methods, including basic counting strategies and Markov state modeling. By carefully combining trajectories together, we obtain estimates of SSW frequencies and their seasonal distributions that are consistent with reanalysis-derived estimates for moderately rare events, but with much tighter uncertainty bounds, and which can be extended to events of unprecedented severity that have not yet been observed historically. These methods hold potential for assessing extreme events throughout the climate system, beyond this example of stratospheric extremes.

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 only 20 years of 46-day weather forecasts, we estimate the magnitudes of once-in-500-year events.

1 Introduction

The atmosphere's extreme, irregular behavior is, in some ways, more important to characterize than its typical climatology. A society optimized for average historical weather patterns is highly exposed to damage from extreme heat and cold, flooding, and other natural hazards. Moreover, extremes may respond more sensitively than mean behavior to climate change, an argument supported by elementary statistics (Wigley, 2009), empirical observations (Coumou & 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 & 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 to the development of "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 will more likely appear either non-existent or twice as likely as it really is. This difficulty is present for 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 provide statistics with higher variance due to under-sampling. For example, the Integrated Forecast System (IFS) of the European Center for Medium-Range Weather Forecasts (ECMWF) is one of the most accurate weather models available today, running at high resolutions of $\sim$16-32 km (ECMWF, 2016). The forecasts are skillful, but typically last for a single season or less—far too short a duration to estimate rare event probabilities directly.

However, these forecasts are launched multiple times every week in large parallel ensembles, which can be exploited to bridge the gap from weather to climate timescales.

The key is to include the data from ensemble members in a statistically principled way. Our main contribution in this paper is to introduce methods to achieve this, using the ensemble forecasts archived in the subseasonal-to-seasonal (S2S) project at ECMWF (Vitart et al., 2017).

Specifically, in this work we estimate probabilities of sudden stratospheric warming (SSW) events, in which the winter stratospheric polar vortex rapidly breaks down from its typical state with a strong cyclonic circulation over the winter-hemisphere pole. The associated subsidence of air in the polar stratosphere leads to adiabatic warming, causing lower-stratospheric temperatures to rise up to 40 K or more over a few days (Baldwin et al., 2021). The breakdown of the stratospheric vortex exerts a "downward influence" on tropospheric circulation (Baldwin & Dunkerton, 2001; Baldwin et al., 2003; Hitchcock & 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., 2018). For example, 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 (Sigmond et al., 2013; Butler et al., 2019; Scaife et al., 2022). Pushing this "frontier" of weather forecasting can improve disaster preparation and resource management in the face of meteorological extremes (White et al., 2017; Bloomfield et al., 2021). 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 and other climatological statistics of SSWs (Butler et al., 2015; Gerber et al., 2022).

2 Data and definitions

Fig. 1(a,b) shows the evolution of zonal-mean zonal wind at 10 hPa and 60°N, a standard index for the strength of the stratospheric polar vortex (Butler et al., 2019), which we abbreviate U. Blue timeseries show U through two consecutive winters: (a) 1998-1999, in which an extreme SSW occurred as quantified by the deep drop in U in mid-December, and (b) 2009-2010, when a more mild SSW occurred in February. Both timeseries are superimposed upon the 1959-2019 ERA-5 climatology.

U is typically positive throughout the winter months, characterizing a strong circumpolar jet that forms in the stratosphere during the polar night. The standard definition of an SSW event is that U changes sign (Butler et al., 2015), but it does not capture the range of intensities between events. Clearly, December 1998 exhibited a much stronger breakdown of the vortex 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 barely reversed sign; events of this magnitude occur on average every two years. On the other hand, extraordinary events like December 1998 have only been observed a few times.

We define an SSW as the first decrease in U below a threshold $U^{(\mathrm{th})}$ during the "SSW season" of Nov. 1-Feb. 28. We only count the first event of a winter to exclude the subsequent oscillations of U about $U^{(\mathrm{th})}$ as separate SSW events, without complicating the definition with a minimum separation time as in Charlton and Polvani (2007). The main quantity of interest is the *rate*: the average number of SSW events per year, a number between zero and one. Equivalently, the reciprocal of the rate is called the *return period*: the expected number of years to wait before an event of a given severity. For the standard threshold $U^{(\mathrm{th})} = 0$, the rate is approximately 0.6 (Baldwin et al., 2021), but we will consider a range of severities by varying $U^{(\mathrm{th})}$ down to -52 m/s.

One can estimate the rate with reanalysis by counting the fraction of years with an SSW event. Fig. 2a shows two rate estimates derived from ERA5 (Hersbach et al., 2020) as a function of $U^{(\mathrm{th})}$: the blue points use 61 years of data (1959-2019) while the orange points use only 20 years of data (1996-2015). The corresponding error bars encompass the 50% (thick lines) and 95% (thin lines) confidence intervals of $(X_1+\ldots+X_n)/n$, where the X_i's are independent Bernoulli random variables with success rate given by the estimated rate, and a number of trials equal to 20 (blue) or 61 (orange). Fig. 2b-e shows the corresponding seasonal distribution of events at four selected thresholds, with histograms normalized to have unit area. It may appear inconsistent that the support of the distribution at $U^{(\mathrm{th})} = -8$ is not fully contained in the support of the distribution at $U^{(\mathrm{th})} = 0$; for example, the third blue and orange bins of February have positive weight in panel (c), but zero weight in panel (b). There is no contradiction: although every winter with an SSW at level -8 also must have an SSW at level 0, the weaker threshold is crossed first and is sometimes counted in previous weeks.

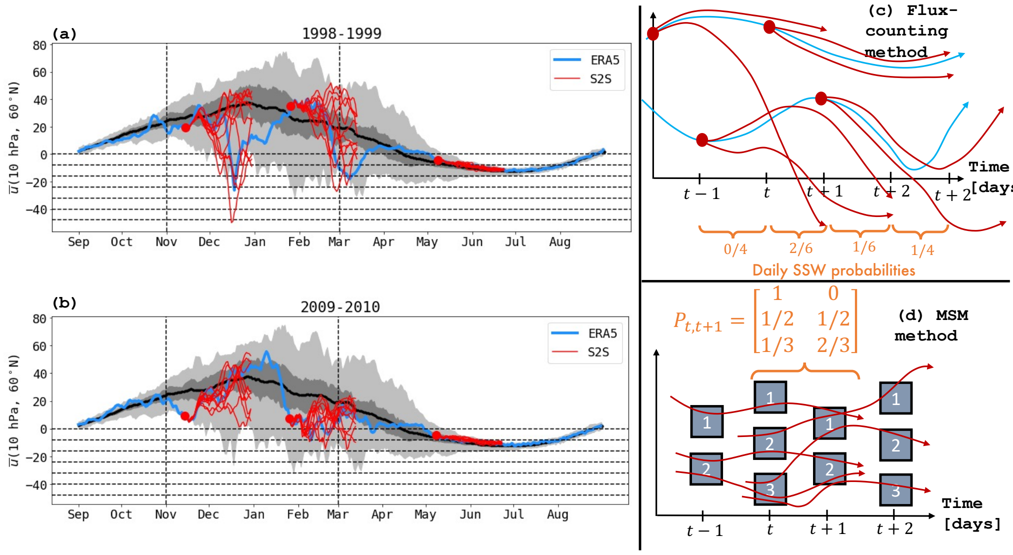

Figure 1. Climatology of polar vortex and illustration of dataset. Black curves show the mean seasonal cycle of $\overline{u}$(10 hPa, 60°N), abbreviated as U, and two gray envelopes show the percentile ranges 25-75 and 0-100, respectively. All statistics are computed with respect to the 61-year ERA-5 dataset between 1959 and 2019. Two individual years are shown in blue: 1998-1999 (a) and 2009-2010 (b). Two ensembles of S2S hindcasts (red) are shown each winter, a small sample from the large S2S dataset of two ensembles *per week* from the ECMWF IFS. Horizontal dashed lines mark several different SSW thresholds $U^{(\mathrm{th})}$ used in this study, including the standard threshold of 0 m/s and several more extreme ones. The time window Nov. 1 - Feb. 28 is marked by vertical dashed lines. When U crosses $U^{(\mathrm{th})}$ downward for the first time within the time window, an SSW has occurred. (c): schematic of the flux-counting method. (d): schematic of the Markov State Model (MSM) method.

The other estimates displayed in Fig. 2 are derived from the S2S dataset, which consists of hindcast trajectories launched in ten-member perturbed ensembles (plus a control member that we omit from our analysis). We use only hindcast data produced by the 2017 version of the ECMWF IFS: that is, integrations initialized from past initial conditions for the 20 years prior, in our case from autumn 1996 to spring 2016 (labeled 1996-2015 in the plots). Each ensemble member has small perturbations applied to its initial conditions, and is integrated forward with stochastically perturbed tendencies (Buizza et al., 1999; Berner et al., 2009). For details on the the model, see Vitart et al. (2017) and ECMWF (2016). The dataset is publically accessible at `https://apps.ecmwf.int/datasets/data/s2s/`.

The total number of days contained therein is roughly

$$20 \text{ years} \times \frac{17 \text{ weeks}}{\text{winter}} \times \frac{2 \text{ ensembles}}{\text{week}} \times \frac{10 \text{ members}}{\text{ensemble}} \times \frac{47 \text{ days}}{\text{member}} = 3.2 \times 10^5 \text{ days} \approx 875 \text{ years} \tag{1}$$

Many of these extra ensemble members reach farther into the negative-U tails than the reanalysis. Thinking of these as alternative realities, we can calculate otherwise inaccessible probabilities.

3 Two estimates of long return times from short trajectories

To take advantage of the S2S data, we have to overcome two complications. 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 875 years. Second, no individual ensemble can directly provide an SSW probability beyond the 46-day time horizon, which is well short of the 120 days between November 1 and February 28 when SSWs are allowed to happen. We cannot use hindcasts directly to estimate the rate, because we need to know what would have unfolded if the 46-day simulation were to continue. The challenge is to make use of the "hanging" trajectory endpoints, such as the eight members of the first ensemble shown in Fig. 1 which do not dip below the threshold. Below, we present two related, but distinct methods: flux-counting and Markov state modeling.

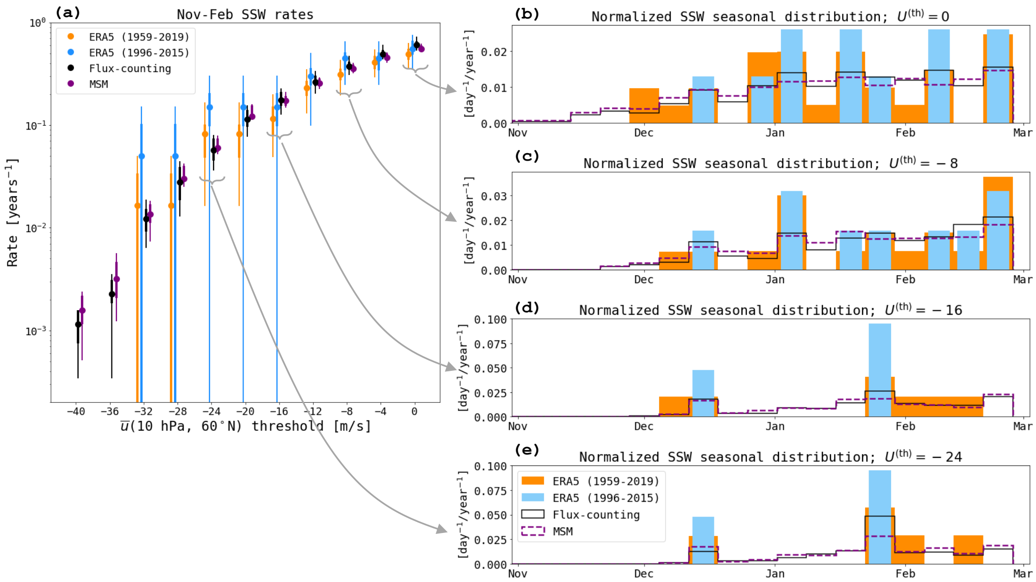

Figure 2. Rate estimates derived from S2S and reanalysis. Left: SSW rate (inverse annual probability of SSW) as a function of zonal wind threshold, $U^{(\mathrm{th})}$, estimated by the four methods described in the text. Error bars indicate 95% confidence intervals. MSM and flux-counting error bars are computed by bootstrapping on entire years of data. ERA5 error bars are computed analytically as the 2.5-97.5 percentile range of the success rate of a binomial random variable with a success probability given by the estimated rate and a number of trials given by the number of years in the record (20 or 61). Error bars going off the bottom of the plot include zero (note the log scale). Right: seasonal distribution of SSW events at four selected thresholds, according to each of four methods. All histograms have a bin width of 7 days and are rescaled to have unit area.

3.1 Flux-counting for direct estimates

The first approach is quite simple, as sketched in Fig. 2c: we compute the probability of an SSW on each day by calculating the fraction of trajectories that cross the threshold on that day (avoiding double counting by keeping track of "active" trajectories as detailed below), and then sum up all the daily probabilities over the season. Formally, we decompose the winter months of interest into a sequence of one-day windows, which is the sampling resolution of S2S:

$$T_0 = \text{Nov } 1, T_0 + 1 = \text{Nov. } 2, \ldots, T_1 = \text{Feb. } 28 \tag{2}$$

and estimate the probability of an SSW separately for each calendar day. By our definition, an SSW can happen at most once per season, to ensure the events are disjoint and have additive probabilities:

$$\text{Rate} = \sum_{t=T_0}^{T_1} \mathbb{P}\{\text{SSW on day } t\} \tag{3}$$

$$= \sum_{t=T_0}^{T_1} \mathbb{P}\Big\{ \min_{T_0 \leq s < t} U(s) > U^{(\text{th})} \text{ and } U(t) \leq U^{(\text{th})} \Big\} \tag{4}$$

The summand can be considered a probability per day of crossing the threshold $U^{(\text{th})}$, i.e., one of the horizontal dashed lines in Fig. 1a,c. It is estimated by averaging over all hindcast trajectories that are "active" on calendar day t, meaning those launched some day between $t-46$ and t. More precisely, if we enumerate the active trajectories by $i \in \mathcal{I}(t) = \{1, ..., N(t)\}$ and denote the i'th trajectory's zonal wind by U_i, then the estimate of daily SSW probability is

$$\mathbb{P}\{\text{SSW on day } t\} = \frac{1}{N(t)} \sum_{i \in \mathcal{I}(t)} \mathbb{I}\Big\{ \min_{T_0 \leq s < t} U_i(s) > U^{(\text{th})} \Big\} \mathbb{I}\Big\{ U_i(t) \leq U^{(\text{th})} \Big\} \tag{5}$$

where $\mathbb{I}$ is an indicator function, equal to 1 if the argument is true and 0 if the argument is false. In words, we count the trajectories that dip below $U^{(\text{th})}$ for the first time on day t, as a fraction of all trajectories that are active on that day. The past of ensemble member i before its initialization date is given by the corresponding reanalysis from which it branched.

Summing up these probabilities from Nov. 1 to Feb. 28, and sweeping over all thresholds $U^{(\text{th})}$, we obtain the black curve in Fig. 2a. Error bars come from a bootstrapping procedure: we apply the estimate (5) to 20 different random 10-year subsets of $\{1996, \ldots, 2015\}$, calculate the 2.5th and 97.5th percentiles of rate estimates, and form the pivotal 95%

bootstrap confidence interval (see Wasserman (2004), chapter 8, for a formal account, although we have modified the procedure by sampling without replacement to maintain independence of different years.)

The average (5) is a sum of *dependent* random variables, with all ensemble members in a given year sharing common history. This increases the variance of the estimator or, in other words, reduces the effective sample size from 875 years. This situation is common in the Monte Carlo simulation for inverse problems. But the error bars make clear that flux-counting enjoys a tremendous advantage over the direct ERA5 estimate. At all thresholds, the flux-counting error bar overlaps with the ERA5 error bar, but is much smaller. This gives us confidence to trust the flux-counting estimate farther into the tail where no ERA5 data are available.

3.2 Markov state model

The second method is more intricate, but delivers more insight into the predictability of SSWs. We construct a *Markov state model* (MSM) (Deuflhard et al., 1999; Pande et al., 2010; Chodera & Noé, 2014) which is sketched in Fig. 2d. On each day t, 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 $t+1$,

$$P_{t,t+1}(j,k) = \mathbb{P}\{\mathbf{X}(t+1) \in S_{t+1,k} | \mathbf{X}(t) \in S_{t,j}\}, \tag{6}$$

by counting the transitions between corresponding boxes. Explicitly,

$$P_{t,t+1}(j,k) = \frac{\sum_{i\in\mathcal{I}(t)} \mathbb{I}\{\mathbf{X}_i(t) \in S_{t,j}\}\mathbb{I}\{\mathbf{X}_i(t+1) \in S_{t+1,k}\}}{\sum_{i\in\mathcal{I}(t)} \mathbb{I}\{\mathbf{X}_i(t) \in S_{t,j}\}} \tag{7}$$

The matrices are row-normalized, which corrects for the redundancy and statistical dependence between ensemble members. This sequence of matrices is the key ingredient that enables all downstream calculations. Choosing the partition of state space is a crucial step which involves a tradeoff: too few clusters will coarsen the dynamics too much, whereas too many clusters will reduce the number of data points in each cluster and thus increase the statistical noise involved in estimating $P_{t,t+1}(j,k)$. There is a lack of general theory on how to construct MSMs, but here we exploit the particular structure of the dataset to validate our algorithmic choices, as explained in the supplement. Here, we focus on conveying the general MSM procedure.

We build the sets $S_{t,i}$ using k-means clustering of the data using the `scikit-learn` package (Pedregosa et al., 2011). As input to k-means, we use a vector of feature Φ consisting of time-delays of U:

$$\Phi(\mathbf{X}(t)) = [U(\mathbf{X}(t)), U(\mathbf{X}(t-1)), \dots, U(\mathbf{X}(t-\delta+1))] \tag{8}$$

where δ days is the number of retained time-delays. This time-delay embedding encodes additional information about the atmospheric state, enabling a model based just on the zonal mean wind at 10 hPa. Heuristically, the embedding captures approximate time-derivatives up to order δ-1. The technique has precedent in climate science (Ghil et al., 2002), and a growing body of theoretical and empirical evidence supports the use of time-delay coordinates as reliable features for encoding dynamical attractors (Takens, 1981; Kamb et al., 2020; Broomhead & King, 1986; Giannakis & Majda, 2012; Brunton et al., 2017; Thiede et al., 2019; Strahan et al., 2021). We have also experimented with richer feature spaces including EOFs of geopotential height, but found it unnecessary.

We find that any δ from 2 to 10 and any number of clusters (denoted M_t) from 50 to 150 gives similar results. In Fig. 2 we display results of a single representative choice of $\delta = 5$ days and $M_t = 150$, along with a shaded 95% confidence interval derived from the pivotal bootstrap procedure (Wasserman, 2004) with 20 independent resamplings of the data (but without replacement). The supplement further explains how we selected these parameters to simultaneously optimize the MSM's fidelity and robustness on a simple performance benchmark. We emphasize that these clusters are not supposed to identify metastable weather regimes in the tradition of, e.g., Michelangeli et al. (1995); rather, they are a discretization of state space meant to represent continuous functions over that space, encoding gradual progress towards an SSW event.

Given the clusters $\{S_{t,j}\}$ and the transition matrices $\{P_{t,t+1}\}$, we can calculate the rate and seasonal distribution of SSW events with the following procedure.

1. Let B denote the set of "weak-vortex" clusters: all (t, j) such that $T_0 \leq t \leq T_1$ and the majority of data points in $S_{t,j}$ have $U < U^{(\text{th})}$. Let A denote the set of "non-winter" clusters: all (t, j) such that $t < T_0$ or $t > T_1$. With this setup, an SSW event is a *transition from A to B*.
2. Compute the *committor probability*,

$$q_t^+(j) = \mathbb{P}\Big\{U(s) \leq U^{(\text{th})} \text{ for some } s \in [t, T_1] \;\Big|\; \mathbf{X}(t) \in S_{t,j}\Big\}, \tag{9}$$

by solving the following terminal/boundary-value problem. By definition, $q^+_{T_1+1}(j) = 0$ for all clusters j at the end of winter, while $q^+_t(j) = 1$ for all $(t, j) \in B$. Stepping backward through time, we have a recursion relation:

$$q^+_t(j) = \sum_{k=1}^{M_{t+1}} P_{t,t+1}(j,k) q^+_{t+1}(k). \tag{10}$$

In words, for a vortex that is initially strong today (t) to break down by Feb. 28 (T_1), it must break down sometime between tomorrow (t+1) and T_1. Hence $q^+_t(j)$ is a weighted combination of $q^+_{t+1}(k)$ for all possible scenarios k for tomorrow. This equation is simply the Kolmogorov Backward Equation in discrete form (E et al., 2019). In this light, viewing Eq. (10) as a discretized partial differential equation, the clusters $\{S_{t,j}\}$ can be seen as members of a finite element basis and $P_{t,t+1}(i, j)$ as stiffness matrices. Indeed, here we use an MSM as a "dynamical Galerkin approximation", a basis expansion approach to computing forecast quantities like the committor probability from short trajectory data that was originally developed for chemistry applications (Thiede et al., 2019; Strahan et al., 2021) and has recently been applied to climate dynamics (Finkel et al., 2021, 2022; Jacques-Dumas et al., 2022).

3. Estimate an empirical probability distribution over clusters at the beginning of winter,

$$\pi_{T_0}(j) = \mathbb{P}\{\mathbf{X}(T_0) \in S_{T_0,j}\} \tag{11}$$

However, in practice, the result is not sensitive to the choice of initial probability distribution. This is because T_0 is early enough in the winter season that the distribution of U is still narrow (see Fig. 1) and the memory of initial conditions is practically erased by the time of the first SSW. We can also propagate π to each day of the season, using the Kolmogorov Forward equation (a.k.a. the Fokker Planck equation) in discrete form:

$$\pi_{t+1}(k) = \sum_{j=1}^{M_t} \pi_t(j) P_{t,t+1}(j,k) \tag{12}$$

4. Compute the rate as the average of committor probabilities on the first day of the SSW season, weighted by the probability distribution π_{T_0}:

$$R = \sum_{j=1}^{M_{T_0}} q^+_{T_0}(j) \pi_{T_0}(j) \tag{13}$$

In words, the probability of SSW in a random year is the sum of probabilities from every possible initial condition, weighted by the probability of that initial condition. Fig. 2 shows in purple the rate according to the MSM , which matches remarkably well with the flux-counting method. Error bars indicate the 95% confidence interval, obtained with the same bootstrapping procedure that we used for flux-counting. In particular, the *entire clustering* procedure is repeated for each 10-year subset of data.

5. Compute the seasonal distribution by decomposing the rate over all possible entrance times to B, rather than exit points (i.e., initial conditions) from A:

$$R = \sum_{t=T_0-1}^{T_1} \sum_{j=1}^{M_t} \sum_{k=1}^{M_{t+1}} q_t^-(j) P_{t,t+1}(j,k) \mathbb{I}\{(t+1,k) \in B\} \tag{14}$$

where $q_t^-(j) = \mathbb{P}\{\text{no SSW has occurred yet between } T_0 \text{ and } t | \mathbf{X}(t) = j\}$ is known as the *backward committor*. The backward committor obeys a recursion analogous to that of q_t^+, but moving backward through time and with a time-reversed transition matrix:

$$q_{t+1}^-(k) = \begin{cases} \sum_{j=1}^{M_t} P_{t,t+1}(j,k) \frac{\pi_t(j)}{\pi_{t+1}(k)} q_t^-(j) & (t,j) \notin A \cup B \\ 0 & (t,j) \in B \\ 1 & (t,j) \in A \end{cases} \tag{15}$$

The purple histogram in Fig. 2b-e is given by the individual summands (in groups of 7, according to the bin width of 7 days).

The committor, defined in step 2 above, measures probabilistic progress towards an SSW event (how likely). To measure *temporal* progress (how soon), we further define the *hitting time* as

$$\tau_t^+ = \min\{s \geq 0 : (t+s, \mathbf{X}(t+s)) \in B\} \tag{16}$$

This is a random variable that tells you the timing of the SSW, depending on the realization of $\mathbf{X}$. We compute two summary statistics of this random variable. First, its cumulative probability mass function $\mathbb{P}\{\tau_t^+ < \sigma\}$ is a *time-limited* version of the committor, which we use to validate our choice of MSM parameters following Benedetti (2010) and Miloshevich et al. (2022) for a standardized measure of prediction skill (see the supplement). Second, the average value of τ_t^+, conditional on the vortex actually breaking down the same winter, is called the *expected lead time*:

$$\eta_t^+ = \mathbb{E}\big[\tau_t^+ | t + \tau_t^+ \leq T_1\big] \tag{17}$$

This is another useful summary statistic to quantify how far away the system is from an SSW event. We displayed a similar quantity in (Finkel et al., 2021, 2022) in the context of an idealized model. The expected lead time can also be computed by recursion with the MSM, but the formula is slightly more involved and left to the supplement.

Let us take a brief aside to reference some mathematical context for the method above. The general framework that we have used to combine committor probabilities to compute rates and other steady-state statistics of rare transitions is *transition path theory* (TPT) (Vanden-Eijnden, 2014). TPT has been applied to molecular dynamics (Noé et al., 2009; Meng et al., 2016; Strahan et al., 2021; Antoszewski et al., 2021), atmospheric and oceanic sciences (Finkel et al., 2020, 2022; Miron et al., 2021, 2022) and social sciences (Helfmann et al., 2021). Though TPT is typically formulated in a time-homogeneous setting, here we have built in explicit time-dependence to deal with the seasonal cycle, similarly to (Helfmann et al., 2020).

Our MSM-based approximation of the committor probability is 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., 2022). Dynamical Galerkin approximation (using a basis different than the one used here) and a short trajectory variant of analogue forecasting are tested on several benchmark problems in (Jacques-Dumas et al., 2022). Formally, the transition operator encoded by the matrix in (6) is related to linear inverse models (Penland & Sardeshmukh, 1995), which have also been used to predict subseasonal extremes (Tseng et al., 2021). Both MSMs and linear inverse models involve finite-dimensional approximations of the transition operator (or Koopman operator for deterministic dynamics) (Mezić, 2013; Mezić, 2005; Klus et al., 2018).

4 Results

4.1 Rate estimates

Fig. 2 compares rate estimates from the MSM and flux-counting methods against the reanalysis rates. We include the ERA-5 estimator based on just 1996-2015 to overlap with the S2S period. For the mild thresholds of $U^{(\mathrm{th})} = 0, -4$ m/s, corresponding to return times of 2-3 years, the MSM and flux-counting estimates agree with both short- and long-term reanalysis estimates. Moving to moderate thresholds of $U^{(\mathrm{th})} \sim -28$ m/s, the MSM and flux-counting rates track somewhat closer with the 61-year estimate (orange), which has slightly lower rates across the board. The S2S data were initialized from the 20-year time period corresponding to the blue curve, but the S2S hindcasts recover the longer-term climatology, despite the (slightly) greater frequency of SSWs of this intensity from the period in which they were initialized.

One can think of these SSW frequencies as the *climatology according to the Integrated Forecast System*, given the boundary conditions of the 1996-2015 period. At least in the "model world" of the IFS, it does not appear that a differences in atmospheric boundary conditions (e.g., sea surface temperatures) caused a systematic increase in intense SSWs between 1996 and 2015; rather the observed increase in SSWs was luck of the draw. It is possible, however, that systematic model error could be obscuring the systematic differences suggested by Reichler et al. (2012) and Dimdore-Miles et al. (2021).

At all levels of $U^{(\mathrm{th})}$, but especially in the negative extremes, the confidence intervals from the two S2S estimates are smaller than those from the ERA5 estimates, thanks to the large amount of S2S data that the MSM and flux-counting methods can exploit. The agreement between MSM and ERA-5 on common events gives us more confidence to trust the MSM on less common events in the negative-$U^{(\mathrm{th})}$ tail, where ERA-5 data are too sparse to give a meaningful rate estimate. Both the direct counting and MSM approaches suggest the potential for events where the vortex becomes so disrupted it spins -40 m/s (stronger than average, but in the opposite direction), albeit only once or twice in a millennium.

Several recent studies have performed the same task of filling out a sparse climate distribution using models (Horan & Reichler, 2017; Kelder et al., 2020), but with uninterrupted long runs of a global climate model. The techniques we have introduced—MSM and flux-counting—offer a novel way to estimate such quantities from short trajectories only, without access to a centennial-scale run of the IFS which the standard estimation method would require. We believe the higher resolution IFS is also more appropriate for capturing the most extreme SSWs.

4.2 Seasonal distribution

Fig. 2b-e illustrates that S2S data also offers an advantage for describing the seasonal distribution of SSW events, an inherently noisier statistic than the full-winter rate estimate because one has to split the data into finer categories. The S2S-derived histograms, in black and purple, are able to bring out seasonal structure that is ambiguous in the reanalysis data directly. For $U^{(\mathrm{th})} = 0$ m/s, there is a gradual increase in SSW frequency from November to January, followed by a plateau in February, consistent with prior studies of seasonality at monthly resolution (Charlton & Polvani, 2007) and supporting the late winter maximum found by Horan and Reichler (2017).

As $U^{(\mathrm{th})}$ becomes more negative, the reanalysis histograms dwindle and degenerate into a few isolated spikes, whereas the S2S histograms become intriguingly bimodal, but retain their smoothness. The S2S histograms show a persisting SSW occurrence throughout February after the second peak, a feature that is also faintly present in the longer reanalysis period (1959-2019), but not at all in the shorter reanalysis period (1996-2015) from which S2S was initialized. Again, the IFS recovers features of the longer-term climatology.

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. The bimodal structure seen in S2S has also been found tentatively in prior studies with both reanalysis and models (Horan & Reichler, 2017; Ayarzagüena et al., 2019), and more robustly in other features of the boreal winter, e.g., the midwinter suppression of Pacific storm activity (Nakamura, 1992). We speculate that the early peak represents Canadian warmings (Meriwether & Gerrard, 2004), which our result suggests may deserve a more decisive classification. 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 & Reichler, 2017).

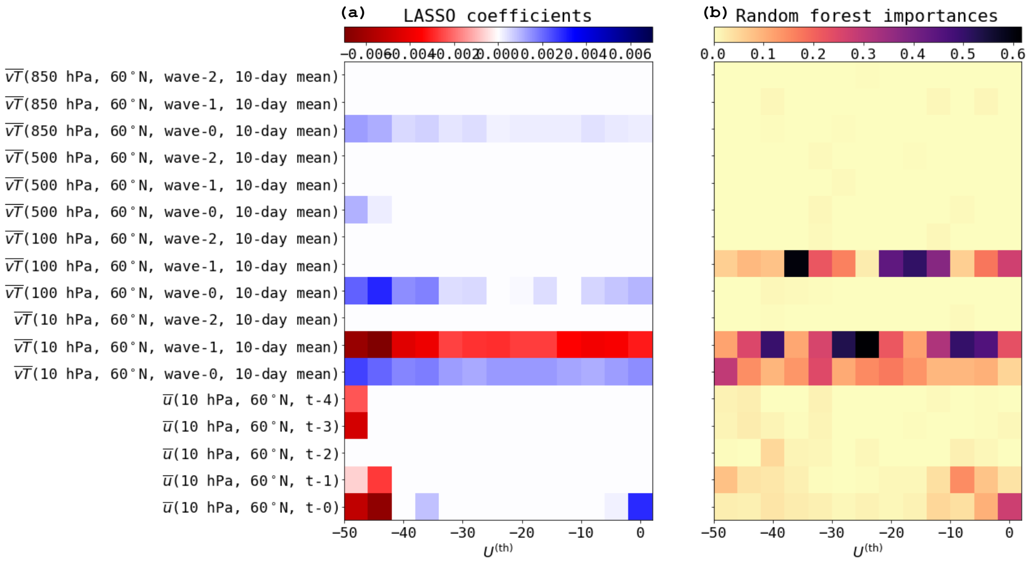

Figure 3. Sparse regression results. Heat maps show the importance of each feature (listed on the vertical axis) for predicting the expected lead time η^+ at a range of zonal wind thresholds $U^{(\mathrm{th})}$ (listed along the horizontal axis). The left-hand heat map shows the LASSO coefficients, and the right-hand panel shows Gini importances from random forest regression. We also used $\overline{u}$ at lower levels than 10 hPa as input features, but found none of them to have any importance, and so omitted them from the figure.

4.3 Statistical predictors of SSWs

Estimates of long return times alone do not provide physical insight into the mechanisms driving the event. The committor probability and expected lead time estimates provided by the MSM approach encode information on the dynamics and predictability of SSW events, and on extreme events in general. These quantities cannot be computed by the flux-counting approach. A number of recent articles have pursued committor probabilities as windows into transitional dynamics, e.g., Miloshevich et al. (2022) for European heat waves and Frishman and Grafke (2022) for the spread of turbulence in a pipe. On SSWs specifically, our own previous studies with a simple SSW model (Finkel et al., 2021, 2022) found through sparse regression that a small set of physical variables could explain key variability in the committor.

Here we analyze the S2S dataset in a similar way, using sparse regression to reveal the main determinants of the committor q_t^+ (how likely is an SSW to occur?) and expected lead time η_t^+ (if it does occur, how soon?), among a large collection of candidate variables including zonal-mean zonal winds and meridional eddy heat fluxes at various time delays, altitudes, and wavenumbers (listed on the left of Fig. 3). We have performed two kinds of regression: linear regression with a sparsity-promoting $L1$ penalty of 0.1, also known as LASSO (Tibshirani, 1996), and random forest regression (Hastie et al., 2009) with 10 trees of depth 3. Both algorithms, as implemented using `scikit-learn`, provide not only predictions of the output variable but also notions of relevance for each input feature: nonzero coefficients in the case of LASSO, and Gini importances in the case of the random forest (Pedregosa et al., 2011). These relevances are of greater interest to us than the raw skill of the regression.

We focus on the early part of the SSW season to connect the results with the rate formula (13). The target variable for regression is $\log(\eta_t^+)$, which guarantees the predicted η_t^+ is positive and also emphasizes variability in small values of η_t^+ (when an SSW is close) rather than large values (when an SSW is distant). The training data consist of trajectory snapshots as inputs and MSM-labeled η_t^+ values as outputs. We include only those snapshots between Nov. 1 and Nov. 30, and strictly outside of sets A and B, where $0 < q_t^+ < 1$. (We also regressed on q_t^+ and discovered similar but more subtle patterns of importance; for brevity we show results only for η_t^+.)

Fig. 3 summarizes the results of regression across all $U^{(\mathrm{th})}$ thresholds. Random forest importances are always nonnegative and represented on a yellow-black color scale, while regression coefficients are signed and represented on a red-blue color scale. Red is associated with a weaker vortex, meaning a shorter lead time. The correlation coefficient R^2 remains between 0.4 and 0.6 for both methods across all thresholds, indicating that these regressions are imperfect expressions of the expected lead time, but do explain a significant part of the variance.

The models illuminate some interesting patterns, some obvious and some surprising. *A priori*, one expects $\overline{u}(10\text{ hPa}, 60°\text{N}, t)$ itself [the bottom listed feature, abbreviated $U(t)$] to dominate the regression, since it defines the event. This is true at a mild threshold of $U^{(\mathrm{th})}$—stronger zonal wind means longer expected lead time, according to the positive coefficient in the panel's lower right corner—but for more extreme thresholds, it is actually the *time-delayed* zonal wind $U(t-1), \dots, U(t-4)$ that is more relevant. Furthermore, the corresponding LASSO coefficients are negative, suggesting that the decrease over time of U is more important than its value today. At the most extreme thresholds, it even appears that *strong* $U(t)$ portends a sooner vortex collapse, suggesting that the most extreme SSW events (those reaching the most *negative* zonal wind) follow from precursor states with anomalously *strong* zonal wind.

Studies with reanalysis and idealized models (e.g., Charlton & Polvani, 2007; Jucker, 2016) have found a similar pattern of strengthening zonal wind, as well as meridional potential vorticity gradient, prior to strong SSW events. These effects are components of preconditioning, wherein the vortex develops a sharper edge and becomes more susceptible to the frequent upward bursts of wave activity emanating from the troposphere (e.g., Albers & Birner, 2014). The presence of the same pattern in S2S is an encouraging signal of physical consistency across the model hierarchy.

Another important set of features is the 10-day averaged meridional heat flux $\overline{vT}$ averaged over 45-75°N, although LASSO and Random Forest regressions emphasize different altitudes and wavenumber components. Both methods agree that the 10 hPa heat flux at wavenumbers 0 and 1 exert strong and competing (statistical) influences on expected lead time: a stronger wavenumber-0 component ($\overline{v}\overline{T}$) means vortex collapse is farther away, while a stronger wavenumber-1 component means vortex collapse is sooner. At lower levels of the atmosphere, the eddy heat fluxes exert significant but diminish-

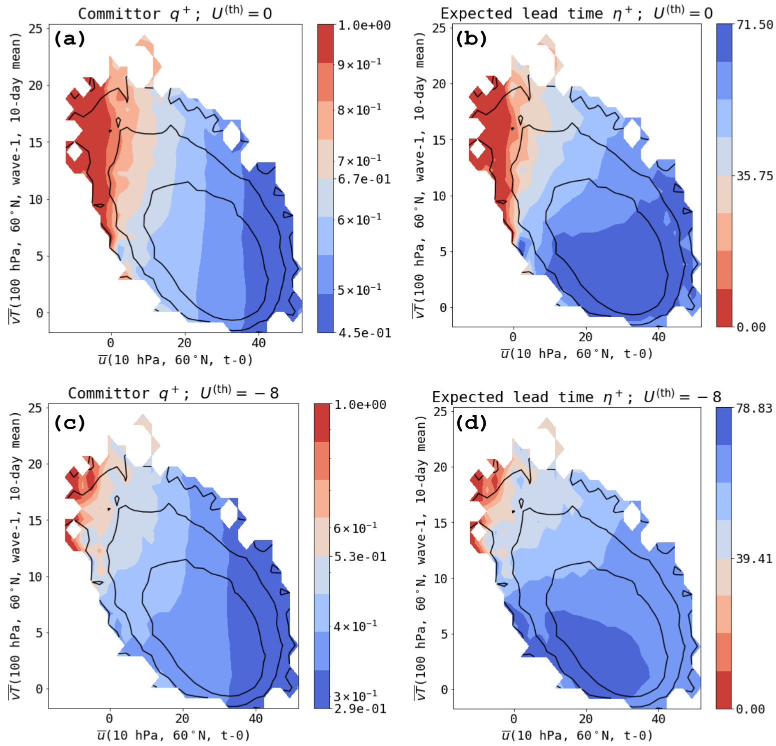

Figure 4. Committor and expected lead time. Aggregating all S2S data from Nov. 1 to Nov. 30, this figure displays the committor (left column) and expected lead time (right column) in shading, as well as the climatological probability density π in black contours, on a logarithmic scale (π has arbitrary units, normalized to have unit integral over state space). Two zonal wind thresholds are considered: 0 m/s (top), and −8 m/s (middle) (bottom). All results are derived from the MSM. We show η^+ and q^+ as functions of two variables only: $\overline{u}$(10 hPa, 60°N, t) and meridional heat flux averaged over 10 days between 45°N and 75°N at 100 hPa and wavenumber 1. The remaining variables are averaged out and weighted by π in this display.

ing influences, although they remain important for the most extreme SSW events (at least according to LASSO).

How can we make sense of all these correlations? One simple method of visualization is to plot the committor and lead time as approximate functions of two variables (averaging over remaining variables). The regression results present us with many possible pairs of important variables. Here we select just one pair: zonal-mean zonal wind at 10 hPa, and wavenumber 1 meridional heat flux at 100 hPa, averaged over the preceding 10 days. The latter feature is assigned high importance by the random forest, though not by LASSO, and is especially interesting as a signal coming from a lower altitude than 10 hPa, possibly related to the two-way influence characteristic of coupled troposphere/stratosphere dynamics. Fig. 4 displays the committor (left column) and expected lead time (right column) as a function of these two variables at two thresholds: $U^{(\mathrm{th})} = 0$ m/s (top), and -8 m/s (bottom). Contours of the climatological probability density π signal which regions are more frequently visited and which ones are rare. We only average over the first month of the SSW season, Nov. 1-Nov. 30, to represent a map of possible "initial conditions" for the winter vortex evolution.

The orientation of contours in phase space reveals a pattern of influence that would be hard to intuit from the regression coefficients alone. At $U^{(\mathrm{th})} = 0$, the q^+ contours run almost perpendicular to the U axis, confirming that the zonal wind itself primarily determines how likely an SSW is for the coming season. But the η^+ contours tell a different side of the story: at stronger U, the contours progressively tilt away from vertical towards horizontal, indicating that the time until SSW depends strongly on heat flux—at least in the regime of strong U, from where an SSW is unlikely to begin with. The influence of heat flux grows more significant as the threshold $U^{(\mathrm{th})}$ is lowered in row 2 of the figure, even for the committor. We infer a general pattern: the 10-hPa zonal wind strength in November determines how likely an SSW is for the coming winter, but when it is rather unlikely, the lower-stratospheric wavenumber-1 heat flux determines when the SSW will happen.

What the phase space images reveal most of all is that q^+ and η^+ are nonlinear functions: the influence of a variable depends on the state of all the other variables. Nonlinear regression methods, such as random forests, are therefore crucial to uncover a complete description. Given the mature wave-mean flow interaction theory of SSWs, there are many other features likely to be as good or better at predicting SSWs. For example, from a long GCM integration, (Jucker & Reichler, 2018) found that meridional potential vorticity gradient and 100-hPa meridional heat flux—representing vortex preconditioning and wave activity respectively—can change SSW probability by roughly an order of magnitude at a one-week lead time, and still significantly at seasonal-scale lead times. At present, we have limited our regression analysis to features that are easy to compute without introducing noise by differentiation. But more specific physical hypotheses can be tested by enlarging the feature space to include the relevant terms. The same principle holds for other extreme events besides SSWs.

5 Discussion

By comparing S2S results with reanalysis, we are measuring the composition of potentially 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, both statistical (from the finite sample size) and systematic (from the projection of forecast functions onto a finite basis in the case of the MSM). 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 (at least in principle) be concretely realized by running the model uninterrupted for millennia, with external climatic parameters sampled from their variability in the 20-year time window of 1996-2015. 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. The S2S dataset is an ensemble of opportunity. 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.

And while the IFS model has proven outstanding in its medium-range forecast skill (Vitart, 2014; Kim et al., 2014; Vitart & Robertson, 2018), it was designed for short forecasts. 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 1996-2015 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 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, as a different way to compare different models' ability to represent extremes.

Boreal SSWs provide an ideal demonstration of our method, providing both moderately and extremely rare events. A natural and intriguing future application is the rate of Southern-hemisphere SSW events, in the spirit of (Jucker et al., 2021), which is postponed to future work for the sake of brevity. The method may be extended to other kinds of extremes as well, though care must be exercised when defining the event (e.g., sets A and B) and choosing features in which to do clustering (for the MSM approach), especially in the case of more spatially localized events.

The rate we estimate from the S2S data set is based on 1996-2015 boundary conditions (sea surface temperatures, CO2), and our MSM method assumes the climate was stationary over this period. Our results indicate that according to the 2017 IFS, 1996-2015 conditions were more similar to 1959-2019 than direct counting of SSW events might suggest. This could mean that the IFS was missing some key climatological variable during that period (Dimdore-Miles et al., 2021). There is, however, substantial uncertaintuy on the impact of global warming on SSWs, even under 4xCO2 forcing (Ayarzagüena et al., 2020). By repeating our analysis on different historical periods, or simulations initialized from climate model integrations under different forcing, one could discern a more decisive signal of forced changes than would be available from raw data. Moreover, the expression for the SSW rate (13) as a "dot product" between a committor and a climatological probability density would allow us to decompose small changes in SSW frequency as changes in these two components separately. A changing probability density π would reflect changes in the slow background conditions, whereas a changing committor q^+ would reflect a change in the system dynamics.

Error source (iii) is the most open to scrutiny and improvement. We have used the short S2S hindcasts directly to validate our parameter choices for the MSM (see the supplement). In a sequence of preceding papers (Finkel et al., 2021, 2022), we have benchmarked the performance of the method on a highly idealized SSW model due to Holton and Mass (1976). Nevertheless, large-scale atmospheric models are a mostly-unexplored frontier for this class of methods.

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 46 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 distribution* which is a flexible but important component in rare event estimation problems (Thiede et al., 2019; Strahan et al., 2021; Finkel et al., 2021). With an executable model, we could initialize secondary and tertiary generations of short trajectories to push into more negative-U 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. (2019) 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. A natural extension of these various techniques would combine elements of active rare event sampling with committor estimation via MSMs. Early developments of such a coupling procedure are presented in Lucente et al. (2022).

6 Conclusion

Extreme weather events present a fundamental challenge to Earth system modeling. Very long 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, due to 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 estimates of the rate and seasonal distribution of extreme SSW events. From just 20 years of data, we obtain probability estimates of events with a 500 year return time, which are so extreme that the vortex is as strong in the easterly direction as its typical westerly climatology. These events have never been observed historically, but can be pieced together using our analysis method.

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 instance, a high-resolution model could be used in ensemble forecast mode, but initialized around a decade at the end of this century provided by a climate model, to understand the impact of global warming on extremes.

7 Open Research

Our analysis is based on publicly available datasets from the European Center for Medium-Range Weather Forecasts (ECMWF, 2022b) and from associated Copernicus Climate Data Store (ECMWF, 2022a). The public Zenodo repository at `https://doi.org/10.5281/zenodo.7675972` contains Python scripts to download the necessary data, reproduce the paper's analysis, and apply the methods to other datasets. Please contact J. F. for help with using the code.

8 Conflicts of interest

The authors declare no conflicts of interest.

Acknowledgments

We extend special thanks to Andrew Charlton-Perez, who suggested the S2S dataset as a case study for the methodology; Simon Lee, who helped familiarize us with the data; Amy Butler for help with the reanalysis; and Martin Jucker and an anonymous reviewer for constructive feedback on earlier versions of this manuscript. 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. was supported by the U.S. DOE, Office of Science, Office of Advanced Scientific Computing Research, Department of Energy Computational Science Graduate Fellowship under Award Number DE-SC0019323 at the time of writing. J. F. also acknowledges continuing support from the MIT Climate Grand Challenge on Weather and Climate Extremes. E.P.G. acknowledges support from the NSF through award OAC-2004572. J.W. acknowledges support from the NSF through award DMS-2054306 and from the Advanced Scientific Computing Research Program within the DOE Office of Science through award DE-SC0020427.

References

Abbot, D. S., Webber, R. J., Hadden, S., Seligman, D., & Weare, J. (2021, dec). Rare event sampling improves mercury instability statistics. *The Astrophysical Journal*, *923*(2), 236. Retrieved from https://dx.doi.org/10.3847/1538-4357/ac2fa8 doi: 10.3847/1538-4357/ac2fa8

AghaKouchak, A., Cheng, L., Mazdiyasni, O., & Farahmand, A. (2014). Global warming and changes in risk of concurrent climate extremes: Insights from the 2014 california drought. *Geophysical Research Letters*, *41*(24), 8847-8852. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014GL062308 doi: https://doi.org/10.1002/2014GL062308

Albers, J. R., & Birner, T. (2014). Vortex preconditioning due to planetary and gravity waves prior to sudden stratospheric warmings. *Journal of the Atmospheric Sciences*, *71*(11), 4028 - 4054. Retrieved from https://journals.ametsoc.org/view/journals/atsc/71/11/jas-d-14-0026.1.xml doi: 10.1175/JAS-D-14-0026.1

Antoszewski, A., Lorpaiboon, C., Strahan, J., & Dinner, A. R. (2021). Kinetics of phenol escape from the insulin r6 hexamer. *The Journal of Physical Chemistry B*, *125*(42), 11637-11649. Retrieved from https://doi.org/10.1021/acs.jpcb.1c06544 (PMID: 34648712) doi: 10.1021/acs.jpcb.1c06544

Ayarzagüena, B., Palmeiro, F. M., Barriopedro, D., Calvo, N., Langematz, U., & Shibata, K. (2019). On the representation of major stratospheric warmings in reanalyses. *Atmospheric Chemistry and Physics*, *19*(14), 9469–9484. Retrieved from https://acp.copernicus.org/articles/19/9469/2019/ doi: 10.5194/acp-19-9469-2019

Ayarzagüena, B., Charlton-Perez, A., Butler, A., Hitchcock, P., Simpson, I., Polvani, L., … Watanabe, S. (2020). 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. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019JD032345 (e2019JD032345 2019JD032345) doi: https://doi.org/10.1029/2019JD032345

Baldwin, M. P., Ayarzagüena, B., Birner, T., Butchart, N., Butler, A. H., Charlton-Perez, A. J., … Pedatella, N. M. (2021). Sudden stratospheric warmings.

Reviews of Geophysics, *59*(1), e2020RG000708. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020RG000708 (e2020RG000708 10.1029/2020RG000708) doi: https://doi.org/10.1029/2020RG000708

Baldwin, M. P., & Dunkerton, T. J. (2001). Stratospheric harbingers of anomalous weather regimes. *Science*, *294*(5542), 581-584. Retrieved from https://www.science.org/doi/abs/10.1126/science.1063315 doi: 10.1126/science.1063315

Baldwin, M. P., Stephenson, D. B., Thompson, D. W. J., Dunkerton, T. J., Charlton, A. J., & O'Neill, A. (2003). Stratospheric memory and skill of extended-range weather forecasts. *Science*, *301*(5633), 636-640. Retrieved from https://www.science.org/doi/abs/10.1126/science.1087143 doi: 10.1126/science.1087143

Benedetti, R. (2010). Scoring rules for forecast verification. *Monthly Weather Review*, *138*(1), 203 - 211. Retrieved from https://journals.ametsoc.org/view/journals/mwre/138/1/2009mwr2945.1.xml doi: 10.1175/2009MWR2945.1

Berner, J., Shutts, G. J., Leutbecher, M., & Palmer, T. N. (2009). 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. Retrieved from https://journals.ametsoc.org/view/journals/atsc/66/3/2008jas2677.1.xml doi: 10.1175/2008JAS2677.1

Bloomfield, H. C., Brayshaw, D. J., Gonzalez, P. L. M., & Charlton-Perez, A. (2021). Sub-seasonal forecasts of demand and wind power and solar power generation for 28 european countries. *Earth System Science Data*, *13*(5), 2259–2274. Retrieved from https://essd.copernicus.org/articles/13/2259/2021/ doi: 10.5194/essd-13-2259-2021

Bouchet, F., Rolland, J., & Wouters, J. (2019). Rare event sampling methods. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *29*(8), 080402. Retrieved from https://doi.org/10.1063/1.5120509 doi: 10.1063/1.5120509

Broomhead, D., & King, G. P. (1986). Extracting qualitative dynamics from experimental data. *Physica D: Nonlinear Phenomena*, *20*(2), 217-236. Retrieved from https://www.sciencedirect.com/science/article/pii/016727898690031X doi: https://doi.org/10.1016/0167-2789(86)90031-X

Brunton, S. L., Brunton, B. W., Proctor, J. L., Kaiser, E., & Kutz, J. N. (2017, May 30). Chaos as an intermittently forced linear system. *Nature Communications*, *8*(1), 19. Retrieved from https://doi.org/10.1038/s41467-017-00030-8 doi: 10.1038/s41467-017-00030-8

Buizza, R., Milleer, M., & Palmer, T. N. (1999). Stochastic representation of model uncertainties in the ecmwf ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, *125*(560), 2887-2908. Retrieved from https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.49712556006 doi: https://doi.org/10.1002/qj.49712556006

Butler, A. H., Charlton-Perez, A., Domeisen, D. I., Garfinkel, C., Gerber, E. P., Hitchcock, P., … Son, S.-W. (2019). Chapter 11 - sub-seasonal predictability and the stratosphere. In A. W. Robertson & F. Vitart (Eds.), *Sub-seasonal to seasonal prediction* (p. 223-241). Elsevier. Retrieved from https://www.sciencedirect.com/science/article/pii/B9780128117149000115 doi: https://doi.org/10.1016/B978-0-12-811714-9.00011-5

Butler, A. H., Seidel, D. J., Hardiman, S. C., Butchart, N., Birner, T., & Match, A. (2015). Defining sudden stratospheric warmings. *Bulletin of the American Meteorological Society*, *96*(11), 1913 - 1928. Retrieved from https://journals.ametsoc.org/view/journals/bams/96/11/bams-d-13-00173.1.xml doi: 10.1175/BAMS-D-13-00173.1

Charlton, A. J., & Polvani, L. M. (2007). A new look at stratospheric sudden warmings. part i: Climatology and modeling benchmarks. *Journal of Climate*, *20*(3), 449 - 469. Retrieved from https://journals.ametsoc.org/view/journals/clim/20/3/jcli3996.1.xml doi: 10.1175/JCLI3996.1

Chattopadhyay, A., Nabizadeh, E., & Hassanzadeh, P. (2020). Analog forecasting of extreme-causing weather patterns using deep learning. *Journal of Advances in Modeling Earth Systems*, *12*(2), e2019MS001958. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001958 (e2019MS001958 10.1029/2019MS001958) doi: https://doi.org/10.1029/2019MS001958

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

Coumou, D., & Rahmstorf, S. (2012, Jul 01). A decade of weather extremes. *Nature Climate Change*, *2*(7), 491-496. Retrieved from https://doi.org/10.1038/nclimate1452 doi: 10.1038/nclimate1452

Deuflhard, P., Dellnitz, M., Junge, O., & Schütte, C. (1999). Computation of essential molecular dynamics by subdivision techniques. In P. Deuflhard, J. Hermans, B. Leimkuhler, A. E. Mark, S. Reich, & R. D. Skeel (Eds.), *Computational molecular dynamics: Challenges, methods, ideas* (pp. 98–115). Berlin, Heidelberg: Springer Berlin Heidelberg.

Dimdore-Miles, O., Gray, L., & Osprey, S. (2021). Origins of multi-decadal variability in sudden stratospheric warmings. *Weather and Climate Dynamics*, *2*(1), 205–231. Retrieved from https://wcd.copernicus.org/articles/2/205/2021/ doi: 10.5194/wcd-2-205-2021

Domeisen, D. I. V., Butler, A. H., Charlton-Perez, A. J., Ayarzagüena, B., Baldwin, M. P., Dunn-Sigouin, E., ... Taguchi, M. (2020). 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. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019JD030923 (e2019JD030923 10.1029/2019JD030923) doi: https://doi.org/10.1029/2019JD030923

E, W., Li, T., & Vanden-Eijnden, E. (2019). *Applied stochastic analysis* (Vol. 199). American Mathematical Soc.

ECMWF. (2016). Ifs documentation cy43r1 - part v: Ensemble prediction system. In *Ifs documentation cy43r1.* Author. Retrieved from https://www.ecmwf.int/node/17118 doi: 10.21957/6fm80smm

ECMWF. (2022a). *Copernicus climate data store* [dataset]. Retrieved from https://cds.climate.copernicus.eu

ECMWF. (2022b). *Ecmwf public datasets* [dataset]. Retrieved from www.ecmwf.int (This service is based on data and products of the European Centre for

Medium-Range Weather Forecasts (ECMWF). This ECMWF data is published under a Creative Commons Attribution 4.0 International (CC BY 4.0). https://creativecommons.org/licenses/by/4.0/. ECMWF does not accept any liability whatsoever for any error or omission in the data, their availability, or for any loss or damage arising from their use.)

Finkel, J., Abbot, D. S., & Weare, J. (2020). Path properties of atmospheric transitions: Illustration with a low-order sudden stratospheric warming model. *Journal of the Atmospheric Sciences*, *77*(7), 2327 - 2347. Retrieved from https://journals.ametsoc.org/view/journals/atsc/77/7/jasD190278.xml doi: 10.1175/JAS-D-19-0278.1

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

Finkel, J., Webber, R. J., Gerber, E. P., Abbot, D. S., & Weare, J. (2022). Data-driven transition path analysis yields a statistical understanding of sudden stratospheric warming events in an idealized model. *Journal of the Atmospheric Sciences*. Retrieved from https://journals.ametsoc.org/view/journals/atsc/aop/JAS-D-21-0213.1/JAS-D-21-0213.1.xml doi: 10.1175/JAS-D-21-0213.1

Fischer, E. M., Sippel, S., & Knutti, R. (2021, Aug 01). Increasing probability of record-shattering climate extremes. *Nature Climate Change*, *11*(8), 689-695. Retrieved from https://doi.org/10.1038/s41558-021-01092-9 doi: 10.1038/s41558-021-01092-9

Frishman, A., & Grafke, T. (2022, Apr). Dynamical landscape of transitional pipe flow. *Phys. Rev. E*, *105*, 045108. Retrieved from https://link.aps.org/doi/10.1103/PhysRevE.105.045108 doi: 10.1103/PhysRevE.105.045108

Gerber, E. P., Martineau, P., Ayarzaguena, B., Barriopedro, D., Bracegirdle, T. J., Butler, A. H., … Taguchi, M. (2022). Extratropical stratosphere–troposphere coupling, in Stratosphere-troposphere Processes and their Role in Climate (SPARC) Reanalysis Intercomparison Project (S-RIP) (chap. 6), M. Fujiwara, G. L. Manney, L. Gray, and J. S. Wright, Eds. *SPARC Report No. 10, WCRP-*

6/2021, 221-264. doi: 10.17874/800dee57d13

Ghil, M., Allen, M. R., Dettinger, M. D., Ide, K., Kondrashov, D., Mann, M. E., … Yiou, P. (2002). Advanced spectral methods for climatic time series. *Reviews of Geophysics*, *40*(1), 3-1-3-41. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2000RG000092` doi: https://doi.org/10.1029/2000RG000092

Giannakis, D., & Majda, A. J. (2012). Nonlinear laplacian spectral analysis for time series with intermittency and low-frequency variability. *Proceedings of the National Academy of Sciences*, *109*(7), 2222-2227. Retrieved from `https://www.pnas.org/doi/abs/10.1073/pnas.1118984109` doi: 10.1073/pnas.1118984109

Goss, M., Swain, D. L., Abatzoglou, J. T., Sarhadi, A., Kolden, C. A., Williams, A. P., & Diffenbaugh, N. S. (2020, aug). Climate change is increasing the likelihood of extreme autumn wildfire conditions across california. *Environmental Research Letters*, *15*(9), 094016. Retrieved from `https://doi.org/10.1088/1748-9326/ab83a7` doi: 10.1088/1748-9326/ab83a7

Hansen, J., Lacis, A., Rind, D., Russell, G., Stone, P., Fung, I., … Lerner, J. (1984). Climate sensitivity: Analysis of feedback mechanisms. *feedback*, *1*, 1–3.

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2). Springer.

Helfmann, L., Heitzig, J., Koltai, P., Kurths, J., & Schütte, C. (2021, Oct 01). Statistical analysis of tipping pathways in agent-based models. *The European Physical Journal Special Topics*, *230*(16), 3249-3271. Retrieved from `https://doi.org/10.1140/epjs/s11734-021-00191-0` doi: 10.1140/epjs/s11734-021-00191-0

Helfmann, L., Ribera Borrell, E., Schütte, C., & Koltai, P. (2020, Dec 01). Extending transition path theory: Periodically driven and finite-time dynamics. *Journal of Nonlinear Science*, *30*(6), 3321-3366. Retrieved from `https://doi.org/10.1007/s00332-020-09652-7` doi: 10.1007/s00332-020-09652-7

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., … Thépaut, J.-N. (2020). The era5 global reanalysis. *Quarterly Jour-*

nal of the Royal Meteorological Society, *146*(730), 1999-2049. Retrieved from https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3803 doi: https://doi.org/10.1002/qj.3803

Hitchcock, P., & Simpson, I. R. (2014). The downward influence of stratospheric sudden warmings. *Journal of the Atmospheric Sciences*, *71*(10), 3856 - 3876. Retrieved from https://journals.ametsoc.org/view/journals/atsc/71/10/jas-d-14-0012.1.xml doi: 10.1175/JAS-D-14-0012.1

Holton, J. R., & Mass, C. (1976). Stratospheric vacillation cycles. *Journal of Atmospheric Sciences*, *33*(11), 2218 - 2225. Retrieved from https://journals.ametsoc.org/view/journals/atsc/33/11/1520-0469_1976_033_2218_svc_2_0_co_2.xml doi: 10.1175/1520-0469(1976)033⟨2218:SVC⟩2.0.CO;2

Horan, M. F., & Reichler, T. (2017). Modeling seasonal sudden stratospheric warming climatology based on polar vortex statistics. *Journal of Climate*, *30*(24), 10101 - 10116. Retrieved from https://journals.ametsoc.org/view/journals/clim/30/24/jcli-d-17-0257.1.xml doi: 10.1175/JCLI-D-17-0257.1

Huntingford, C., Marsh, T., Scaife, A. A., Kendon, E. J., Hannaford, J., Kay, A. L., … Allen, M. R. (2014, Sep 01). Potential influences on the united kingdom's floods of winter 2013/14. *Nature Climate Change*, *4*(9), 769-777. Retrieved from https://doi.org/10.1038/nclimate2314 doi: 10.1038/nclimate2314

Jacques-Dumas, V., van Westen, R. M., Bouchet, F., & Dijkstra, H. A. (2022). Data-driven methods to estimate the committor function in conceptual ocean models. *EGUsphere*, *2022*, 1–35. Retrieved from https://egusphere.copernicus.org/preprints/egusphere-2022-1362/ doi: 10.5194/egusphere-2022-1362

Jucker, M. (2016). Are sudden stratospheric warmings generic? insights from an idealized gcm. *Journal of the Atmospheric Sciences*, *73*(12), 5061 - 5080. Retrieved from https://journals.ametsoc.org/view/journals/atsc/73/12/jas-d-15-0353.1.xml doi: 10.1175/JAS-D-15-0353.1

Jucker, M., & Reichler, T. (2018). Dynamical precursors for statistical prediction of stratospheric sudden warming events. *Geophysical Research Letters*, *45*(23), 13,124-13,132. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018GL080691 doi: https://doi.org/10.1029/

2018GL080691

Jucker, M., Reichler, T., & Waugh, D. W. (2021). How frequent are antarctic sudden stratospheric warmings in present and future climate? *Geophysical Research Letters*, *48*(11), e2021GL093215. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021GL093215` (e2021GL093215 2021GL093215) doi: https://doi.org/10.1029/2021GL093215

Kamb, M., Kaiser, E., Brunton, S. L., & Kutz, J. N. (2020). Time-delay observables for koopman: Theory and applications. *SIAM Journal on Applied Dynamical Systems*, *19*(2), 886-917. Retrieved from `https://doi.org/10.1137/18M1216572` doi: 10.1137/18M1216572

Karpechko, A. Y., Hitchcock, P., Peters, D. H. W., & Schneidereit, A. (2017). Predictability of downward propagation of major sudden stratospheric warmings. *Quarterly Journal of the Royal Meteorological Society*, *143*(704), 1459-1470. Retrieved from `https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3017` doi: https://doi.org/10.1002/qj.3017

Kelder, T., Müller, M., Slater, L. J., Marjoribanks, T. I., Wilby, R. L., Prudhomme, C., … Nipen, T. (2020, Nov 27). Using unseen trends to detect decadal changes in 100-year precipitation extremes. *npj Climate and Atmospheric Science*, *3*(1), 47. Retrieved from `https://doi.org/10.1038/s41612-020-00149-4` doi: 10.1038/s41612-020-00149-4

Kidston, J., Scaife, A. A., Hardiman, S. C., Mitchell, D. M., Butchart, N., Baldwin, M. P., & Gray, L. J. (2015, Jun 01). Stratospheric influence on tropospheric jet streams, storm tracks and surface weather. *Nature Geoscience*, *8*(6), 433-440. Retrieved from `https://doi.org/10.1038/ngeo2424` doi: 10.1038/ngeo2424

Kim, H.-M., Webster, P. J., Toma, V. E., & Kim, D. (2014). Predictability and prediction skill of the mjo in two operational forecasting systems. *Journal of Climate*, *27*(14), 5364 - 5378. Retrieved from `https://journals.ametsoc.org/view/journals/clim/27/14/jcli-d-13-00480.1.xml` doi: 10.1175/JCLI-D-13-00480.1

King, A. D., Butler, A. H., Jucker, M., Earl, N. O., & Rudeva, I. (2019). Observed relationships between sudden stratospheric warmings and european climate extremes. *Journal of Geophysical Research: Atmospheres*, *124*(24), 13943-

13961. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019JD030480 doi: https://doi.org/10.1029/2019JD030480

Klus, S., Nüske, F., Koltai, P., Wu, H., Kevrekidis, I., Schütte, C., & Noé, F. (2018, Jun 01). Data-driven model reduction and transfer operator approximation. *Journal of Nonlinear Science*, *28*(3), 985-1010. Retrieved from https://doi.org/10.1007/s00332-017-9437-7 doi: 10.1007/s00332-017-9437-7

Kolstad, E. W., Breiteig, T., & Scaife, A. A. (2010). 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. Retrieved from https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.620 doi: https://doi.org/10.1002/qj.620

Kretschmer, M., Cohen, J., Matthias, V., Runge, J., & Coumou, D. (2018, Nov 22). The different stratospheric influence on cold-extremes in eurasia and north america. *npj Climate and Atmospheric Science*, *1*(1), 44. Retrieved from https://doi.org/10.1038/s41612-018-0054-4 doi: 10.1038/s41612-018-0054-4

Lucente, D., Rolland, J., Herbert, C., & Bouchet, F. (2022). Coupling rare event algorithms with data-based learned committor functions using the analogue markov chain. *Journal of Statistical Mechanics: Theory and Experiment*, *2022*(8), 083201.

Meng, Y., Shukla, D., Pande, V. S., & Roux, B. (2016). Transition path theory analysis of c-src kinase activation. *Proceedings of the National Academy of Sciences*, *113*(33), 9193-9198. Retrieved from https://www.pnas.org/doi/abs/10.1073/pnas.1602790113 doi: 10.1073/pnas.1602790113

Meriwether, J. W., & Gerrard, A. J. (2004). Mesosphere inversion layers and stratosphere temperature enhancements. *Reviews of Geophysics*, *42*(3). Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2003RG000133 doi: https://doi.org/10.1029/2003RG000133

Mezić, I. (2005, Aug 01). Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dynamics*, *41*(1), 309-325. Retrieved from https://doi.org/10.1007/s11071-005-2824-x doi: 10.1007/s11071-005-2824-x

Mezić, I. (2013). Analysis of fluid flows via spectral properties of the koopman operator. *Annual Review of Fluid Mechanics*, *45*(1), 357-378. Retrieved from `https://doi.org/10.1146/annurev-fluid-011212-140652` doi: 10.1146/annurev-fluid-011212-140652

Michelangeli, P.-A., Vautard, R., & Legras, B. (1995). Weather regimes: Recurrence and quasi stationarity. *Journal of Atmospheric Sciences*, *52*(8), 1237 - 1256. Retrieved from `https://journals.ametsoc.org/view/journals/atsc/52/8/1520-0469_1995_052_1237_wrraqs_2_0_co_2.xml` doi: 10.1175/1520-0469(1995)052⟨1237:WRRAQS⟩2.0.CO;2

Miloshevich, G., Cozian, B., Abry, P., Borgnat, P., & Bouchet, F. (2022). Probabilistic forecasts of extreme heatwaves using convolutional neural networks in a regime of lack of data. *arXiv preprint arXiv:2208.00971*.

Miron, P., Beron-Vera, F. J., Helfmann, L., & Koltai, P. (2021). Transition paths of marine debris and the stability of the garbage patches. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *31*(3), 033101. Retrieved from `https://doi.org/10.1063/5.0030535` doi: 10.1063/5.0030535

Miron, P., Beron-Vera, F. J., & Olascoaga, M. J. (2022). Transition paths of north atlantic deep water. *Journal of Atmospheric and Oceanic Technology*, *39*(7), 959 - 971. Retrieved from `https://journals.ametsoc.org/view/journals/atot/39/7/JTECH-D-22-0022.1.xml` doi: 10.1175/JTECH-D-22-0022.1

Mishra, V., & Shah, H. L. (2018, Nov 01). Hydroclimatological perspective of the kerala flood of 2018. *Journal of the Geological Society of India*, *92*(5), 645-650. Retrieved from `https://doi.org/10.1007/s12594-018-1079-3` doi: 10.1007/s12594-018-1079-3

Myhre, G., Alterskjær, K., Stjern, C. W., Hodnebrog, Ø., Marelle, L., Samset, B. H., … Stohl, A. (2019, Nov 05). Frequency of extreme precipitation increases extensively with event rareness under global warming. *Scientific Reports*, *9*(1), 16063.

Nakamura, H. (1992). Midwinter suppression of baroclinic wave activity in the pacific. *Journal of Atmospheric Sciences*, *49*(17), 1629 - 1642. Retrieved from `https://journals.ametsoc.org/view/journals/atsc/49/17/1520-0469_1992_049_1629_msobwa_2_0_co_2.xml` doi: 10.1175/1520-0469(1992)049⟨1629:MSOBWA⟩2.0.CO;2

Naveau, P., Hannart, A., & Ribes, A. (2020). Statistical methods for extreme event attribution in climate science. *Annual Review of Statistics and Its Application*, *7*(1), 89-110. Retrieved from https://doi.org/10.1146/annurev-statistics-031219-041314 doi: 10.1146/annurev-statistics-031219-041314

Noé, F., Schütte, C., Vanden-Eijnden, E., Reich, L., & Weikl, T. R. (2009). Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proceedings of the National Academy of Sciences*, *106*(45), 19011-19016. Retrieved from https://www.pnas.org/doi/abs/10.1073/pnas.0905466106 doi: 10.1073/pnas.0905466106

O'Gorman, P. A. (2012, Oct 01). Sensitivity of tropical precipitation extremes to climate change. *Nature Geoscience*, *5*(10), 697-700. Retrieved from https://doi.org/10.1038/ngeo1568 doi: 10.1038/ngeo1568

Pande, V. S., Beauchamp, K., & Bowman, G. R. (2010). Everything you wanted to know about markov state models but were afraid to ask. *Methods*, *52*(1), 99-105. Retrieved from https://www.sciencedirect.com/science/article/pii/S1046202310001568 (Protein Folding) doi: https://doi.org/10.1016/j.ymeth.2010.06.002

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... others (2011). Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, *12*, 2825–2830.

Penland, C., & Sardeshmukh, P. D. (1995). The optimal growth of tropical sea surface temperature anomalies. *Journal of Climate*, *8*(8), 1999 - 2024. Retrieved from https://journals.ametsoc.org/view/journals/clim/8/8/1520-0442_1995_008_1999_togots_2_0_co_2.xml doi: 10.1175/1520-0442(1995)008⟨1999:TOGOTS⟩2.0.CO;2

Pfahl, S., O'Gorman, P. A., & Fischer, E. M. (2017, Jun 01). Understanding the regional pattern of projected future changes in extreme precipitation. *Nature Climate Change*, *7*(6), 423-427. Retrieved from https://doi.org/10.1038/nclimate3287 doi: 10.1038/nclimate3287

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

Reichler, T., Kim, J., Manzini, E., & Kröger, J. (2012, Nov 01). A stratospheric connection to atlantic climate variability. *Nature Geoscience*, *5*(11), 783-787. Retrieved from https://doi.org/10.1038/ngeo1586 doi: 10.1038/ngeo1586

Sapsis, T. P. (2021). Statistics of extreme events in fluid flows and waves. *Annual Review of Fluid Mechanics*, *53*(1), 85-111. Retrieved from https://doi.org/10.1146/annurev-fluid-030420-032810 doi: 10.1146/annurev-fluid-030420-032810

Scaife, A. A., Baldwin, M. P., Butler, A. H., Charlton-Perez, A. J., Domeisen, D. I. V., Garfinkel, C. I., ... Thompson, D. W. J. (2022). Long-range prediction and the stratosphere. *Atmospheric Chemistry and Physics*, *22*(4), 2601–2623. Retrieved from https://acp.copernicus.org/articles/22/2601/2022/ doi: 10.5194/acp-22-2601-2022

Shepherd, T. G., Boyd, E., Calel, R. A., Chapman, S. C., Dessai, S., Dima-West, I. M., ... Zenghelis, D. A. (2018, Dec 01). Storylines: an alternative approach to representing uncertainty in physical aspects of climate change. *Climatic Change*, *151*(3), 555-571. Retrieved from https://doi.org/10.1007/s10584-018-2317-9 doi: 10.1007/s10584-018-2317-9

Sigmond, M., Scinocca, J. F., Kharin, V. V., & Shepherd, T. G. (2013, Feb 01). Enhanced seasonal forecast skill following stratospheric sudden warmings. *Nature Geoscience*, *6*(2), 98-102. Retrieved from https://doi.org/10.1038/ngeo1698 doi: 10.1038/ngeo1698

Sillmann, J., Shepherd, T. G., van den Hurk, B., Hazeleger, W., Martius, O., Slingo, J., & Zscheischler, J. (2021). Event-based storylines to address climate risk. *Earth's Future*, *9*(2), e2020EF001783. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020EF001783 (e2020EF001783 2020EF001783) doi: https://doi.org/10.1029/2020EF001783

Strahan, J., Antoszewski, A., Lorpaiboon, C., Vani, B. P., Weare, J., & Dinner, A. R. (2021). Long-time-scale predictions from short-trajectory data: A benchmark analysis of the trp-cage miniprotein. *Journal of Chemical Theory and Computation*, *17*(5), 2948-2963. Retrieved from https://doi.org/10.1021/acs.jctc.0c00933 (PMID: 33908762) doi: 10.1021/acs.jctc.0c00933

Takens, F. (1981). Detecting strange attractors in turbulence. In *Dynamical systems and turbulence, warwick 1980* (pp. 366–381). Springer.

Thiede, E. H., Giannakis, D., Dinner, A. R., & Weare, J. (2019). Galerkin approximation of dynamical quantities using trajectory data. *The Journal of Chemical Physics*, *150*(24), 244111. Retrieved from `https://doi.org/10.1063/1.5063730` doi: 10.1063/1.5063730

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

Tripathi, O. P., Baldwin, M., Charlton-Perez, A., Charron, M., Cheung, J. C. H., Eckermann, S. D., ... Stockdale, T. (2016). Examining the predictability of the stratospheric sudden warming of january 2013 using multiple nwp systems. *Monthly Weather Review*, *144*(5), 1935 - 1960. Retrieved from `https://journals.ametsoc.org/view/journals/mwre/144/5/mwr-d-15-0010.1.xml` doi: 10.1175/MWR-D-15-0010.1

Tseng, K.-C., Johnson, N. C., Maloney, E. D., Barnes, E. A., & Kapnick, S. B. (2021). 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. Retrieved from `https://journals.ametsoc.org/view/journals/clim/34/11/JCLI-D-20-0502.1.xml` doi: 10.1175/JCLI-D-20-0502.1

van den Dool, H. M. (1989). A new look at weather forecasting through analogues. *Monthly Weather Review*, *117*(10), 2230 - 2247. Retrieved from `https://journals.ametsoc.org/view/journals/mwre/117/10/1520-0493_1989_117_2230_anlawf_2_0_co_2.xml` doi: 10.1175/1520-0493(1989)117⟨2230:ANLAWF⟩2.0.CO;2

Vanden-Eijnden, E. (2014). Transition path theory. In *An introduction to markov state models and their application to long timescale molecular simulation* (pp. 91–100). Springer New York LLC. doi: 10.1007/978-94-7-7606-7_7

Van Oldenborgh, G. J., Van Der Wiel, K., Sebastian, A., Singh, R., Arrighi, J., Otto, F., ... Cullen, H. (2017). Attribution of extreme rainfall from hurricane harvey, august 2017. *Environmental Research Letters*, *12*(12), 124009.

Vitart, F. (2014). Evolution of ecmwf sub-seasonal forecast skill scores. *Quarterly Journal of the Royal Meteorological Society*, *140*(683), 1889-1899. Retrieved from https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2256 doi: https://doi.org/10.1002/qj.2256

Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., … Zhang, L. (2017). The subseasonal to seasonal (s2s) prediction project database. *Bulletin of the American Meteorological Society*, *98*(1), 163 - 173. Retrieved from https://journals.ametsoc.org/view/journals/bams/98/1/bams-d-16-0017.1.xml doi: 10.1175/BAMS-D-16-0017.1

Vitart, F., & Robertson, A. W. (2018, Mar 12). The sub-seasonal to seasonal prediction project (s2s) and the prediction of extreme events. *npj Climate and Atmospheric Science*, *1*(1), 3. Retrieved from https://doi.org/10.1038/s41612-018-0013-0 doi: 10.1038/s41612-018-0013-0

Wasserman, L. (2004). *All of statistics.* New York: Springer.

Webber, R. J., Plotkin, D. A., O'Neill, M. E., Abbot, D. S., & Weare, J. (2019). Practical rare event sampling for extreme mesoscale weather. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *29*(5), 053109. Retrieved from https://doi.org/10.1063/1.5081461 doi: 10.1063/1.5081461

White, C. J., Carlsen, H., Robertson, A. W., Klein, R. J., Lazo, J. K., Kumar, A., … Zebiak, S. E. (2017). Potential applications of subseasonal-to-seasonal (s2s) predictions. *Meteorological Applications*, *24*(3), 315-325. Retrieved from https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/met.1654 doi: https://doi.org/10.1002/met.1654

Wigley, T. M. L. (2009, Aug 29). The effect of changing climate on the frequency of absolute extreme events. *Climatic Change*, *97*(1), 67. Retrieved from https://doi.org/10.1007/s10584-009-9654-7 doi: 10.1007/s10584-009-9654-7