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]Committee on Computational and Applied Mathematics, University of Chicago
[2]Courant Institute of Mathematical Sciences, New York University
[3]Department of Geophysical Sciences, University of Chicago

Key Points:

- Extreme weather risk, as measured by rate or return times, is inherently difficult to analyze because of data scarcity.
- Transition path theory reveals climatological statistics of sudden stratospheric warming events from high-fidelity subseasonal forecasts.
- Rates and seasonal distributions of 100-year stratospheric extremes are robustly computed from 47-day hindcast ensembles across 21 winters.

Corresponding author: Justin Finkel, `justinfocus12@gmail.com`

Abstract

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.

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 & 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 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 ∼16-32 km (ECMWF, 2016), the IFS produces skillful ensemble forecasts spanning ∼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 & 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, Cohen, et al., 2018). 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 & Robertson, 2018; A. 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 ∼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 2. To realize this potential requires a method to stitch the short trajectories together, which we outline in section 3 and describe more fully in Supporting Information. Section 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 5 and conclude in section 6.

2 Data and definitions

Fig. 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 circumpolar 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., A. H. 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 non-stationary climate (Fischer et al., 2021).

Figure 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.

To quantify SSW intensity, we vary the the $U_{10,60}$ threshold—henceforth called $U_{10,60}^{(\text{th})}$—from 0 m/s to -35 m/s in 5 m/s increments and consider each case separately. Horan and Reichler (2017) and A. H. Butler and Gerber (2018) have suggested the utility of examining different thresholds, as SSW events form a continuum. Horizontal red lines in Fig. 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 $U_{10,60}^{(\text{th})}$ as separate SSW events. A minimum separation time can also be imposed, as in (Charlton & 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.

figures (dragged).pdf

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 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, 2016). This introduces randomness into the ensemble, causing the members to drift apart over time after the initialization date, as shown in Fig. 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.

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 & Noé, 2014) which is sketched in Fig. 1c. At every time sample $t = 1$ day, 2 days, ..., we partition state space into a disjoint collection of bins $S_{t,1}, S_{t,2}, \dots, S_{t,M_t}$ and approximate the transition probability matrix for each time-step from t to $t+1$,

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

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 (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 $U_{10,60}$:

$$\Phi(\mathbf{X}(t)) = [U_{10,60}(\mathbf{X}(t)), U_{10,60}(\mathbf{X}(t-1)), \dots, U_{10,60}(\mathbf{X}(t-\delta))] \tag{2}$$

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 & King, 1986; Giannakis & 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; Antoszewski 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, 2021a). 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., 2021). Formally, the transition operator encoded by the matrix in (1) is related to linear inverse models (LIMs; Penland & 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.

4 Results

4.1 Rate estimates

Fig. 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. **??**). 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 $U_{10,60}^{(\mathrm{th})}$, 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. 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 $U_{10,60}^{(\mathrm{th})} = -15$ 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 & 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. 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, Coumou, et al., 2018; 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 figures (dragged) 2.pdf

Figure 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 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 dashed line shows the boundary of B. The flux density of $\mathbf{J}_{AB}$ across ∂B gives the seasonal distribution shown in Fig. 4.

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. **??**), they roughly agree on the SSW rates with moderate thresholds of $U^{(\mathrm{th})}_{10,60} = 0$ and $U^{(\mathrm{th})}_{10,60} = -5$ 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 $U^{(\mathrm{th})}_{10,60}$ 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. 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.

figures (dragged) 8.pdf

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 space-time,

$$A = \{(t,\mathbf{x}) : t < \text{Nov. 1 or } t > \text{Feb. 28}\} \tag{3}$$

$$B = \{(t,\mathbf{x}) : \text{Nov. } 1 \leq t \leq \text{Feb. 28, and } U_{10,60}(\mathbf{x}) < U^{(\mathrm{th})}_{10,60}\}. \tag{4}$$

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. 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 and the curves turn from solid to dashed; this is why 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

$$\int_{\text{Nov. 1}}^{\text{Feb. 28}} \mathbf{J}_{AB}\cdot\mathbf{n}\,dt = \frac{\#\text{ SSW events}}{\text{Year}} \tag{5}$$

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,

$$\int_{\text{Dec. 1}}^{\text{Dec. 31}} \mathbf{J}_{AB}\cdot\mathbf{n}\,dt = \frac{\#\text{ Dec. SSW events}}{\text{Year}} \tag{6}$$

This relation allows us to examine more refined details of SSW climatology: the seasonal distribution of events.

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 (A. H. 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). 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. 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. 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. 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, $U_{10,60}^{(\text{th})} = 0$ 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 & Polvani, 2007). At a finer resolution of ~ 10 days, however, the S2S figures (dragged) 4.pdf

Figure 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.

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 & 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 & Reichler, 2017; Ayarzagüena et al., 2019). We speculate that the early peak represents Canadian warmings (Meriwether & 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 subject to all the caveats of our data-driven procedure (see Supporting Information), but merit further investigation with numerical models.

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 & 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 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. 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, 2021a), 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. (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. Many other rare event sampling studies have been performed in fluid dynamics and other complex systems (Simonnet et al., 2021; Hoffman et al., 2006; Weare, 2009; Vanden-Eijnden & Weare, 2013; Bouchet et al., 2014; Chen et al., 2014; Farazmand & Sapsis, 2017; Dematteis et al., 2018; Mohamad & 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., 2021).

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 & Polvani, 2007; Albers & Birner, 2014). Do split-type and displacement-type events have fundamentally different mechanisms and/or different downstream effects? (Matthewman & Esler, 2011; Esler & Matthewman, 2011; O'Callaghan et al., 2014; Maycock & Hitchcock, 2015). Will climate change affect the frequency of SSW, perhaps through arctic amplification? (Charlton-Perez et al., 2008; Garfinkel et al., 2017; Kretschmer, Coumou, et al., 2018). 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 & 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 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). Python scripts to download the necessary data and reproduce the paper's analysis will be made available in a public Zenodo repository at publication time.

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.

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://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

Black, R. X., McDaniel, B. A., & Robinson, W. A. (2006). Stratosphere–troposphere coupling during spring onset. *Journal of Climate*, *19*(19), 4891 - 4901. Retrieved from `https://journals.ametsoc.org/view/journals/clim/19/19/jcli3907.1.xml` doi: 10.1175/JCLI3907.1

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., Laurie, J., & Zaboronski, O. (2014, Sep 01). Langevin dynamics, large deviations and instantons for the quasi-geostrophic model and two-dimensional euler equations. *Journal of Statistical Physics*, *156*(6), 1066-1092. Retrieved from https://doi.org/10.1007/s10955-014-1052-5 doi: 10.1007/s10955-014-1052-5

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., 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., & Gerber, E. P. (2018). Optimizing the definition of a sudden stratospheric warming. *Journal of Climate*, *31*(6), 2337 - 2344. Retrieved from https://journals.ametsoc.org/view/journals/clim/31/6/jcli-d-17-0648.1.xml doi: 10.1175/JCLI-D-17-0648.1

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

Camargo, S. J., Tippett, M. K., Sobel, A. H., Vecchi, G. A., & Zhao, M. (2014). Testing the performance of tropical cyclone genesis indices in future climates using the hiram model. *Journal of Climate*, *27*(24), 9171 - 9196. Retrieved from https://journals.ametsoc.org/view/journals/clim/27/24/jcli-d-13-00505.1.xml doi: 10.1175/JCLI-D-13-00505.1

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

Charlton-Perez, A. J., Polvani, L. M., Austin, J., & Li, F. (2008). The frequency and dynamics of stratospheric sudden warmings in the 21st century. *Journal of Geophysical Research: Atmospheres*, *113*(D16). Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2007JD009571 doi: https://doi.org/10.1029/2007JD009571

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

Chen, N., Giannakis, D., Herbei, R., & Majda, A. J. (2014). An mcmc algorithm for parameter estimation in signals with hidden intermittent instability. *SIAM/ASA Journal on Uncertainty Quantification*, *2*(1), 647-669. Retrieved from `https://doi.org/10.1137/130944977` doi: 10.1137/130944977

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

Dematteis, G., Grafke, T., & Vanden-Eijnden, E. (2018). Rogue waves and large deviations in deep sea. *Proceedings of the National Academy of Sciences*, *115*(5), 855-860. Retrieved from `https://www.pnas.org/doi/abs/10.1073/pnas.1710670115` doi: 10.1073/pnas.1710670115

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

ECMWF. (2011). *The era-interim reanalysis dataset.* Copernicus Climate Change Service (C3S). (available from `https://www.ecmwf.int/en/forecasts/datasets/archive-datasets/reanalysis-datasets/era-interim` (accessed 2022-02-10))

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.)

Esler, J. G., & Matthewman, N. J. (2011). Stratospheric sudden warmings as self-tuning resonances. part ii: Vortex displacement events. *Journal of the Atmospheric Sciences*, *68*(11), 2505 - 2523. Retrieved from https://journals.ametsoc.org/view/journals/atsc/68/11/jas-d-11-08.1.xml doi: 10.1175/JAS-D-11-08.1

Farazmand, M., & Sapsis, T. P. (2017). A variational approach to probing extreme events in turbulent dynamical systems. *Science Advances*, *3*(9), e1701533. Retrieved from https://www.science.org/doi/abs/10.1126/sciadv.1701533 doi: 10.1126/sciadv.1701533

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. (2021a). *Exploring stratospheric rare events with transition path theory and short simulations.* arXiv. Retrieved from https://arxiv.org/abs/2108.12727 doi: 10.48550/ARXIV.2108.12727

Finkel, J., Webber, R. J., Gerber, E. P., Abbot, D. S., & Weare, J. (2021b). 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

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

Garfinkel, C. I., Son, S.-W., Song, K., Aquila, V., & Oman, L. D. (2017). Stratospheric variability contributed to and sustained the recent hiatus in eurasian winter warming. *Geophysical Research Letters*, *44*(1), 374-382. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016GL072035 doi: https://doi.org/10.1002/2016GL072035

Gelaro, R., Buizza, R., Palmer, T. N., & Klinker, E. (1998). Sensitivity analysis of forecast errors and the construction of optimal perturbations using singular vectors. *Journal of the Atmospheric Sciences*, *55*(6), 1012 - 1037. Retrieved from https://journals.ametsoc.org/view/journals/atsc/55/6/1520-0469_1998_055_1012_saofea_2.0.co_2.xml doi: 10.1175/1520-0469(1998)055⟨1012:SAOFEA⟩2.0.CO;2

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

Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., ... von Storch, J.-S. (2016). High resolution model intercomparison project (highresmip v1.0) for cmip6. *Geoscientific Model Development*, *9*(11), 4185–4208. Retrieved from https://gmd.copernicus.org/articles/9/4185/2016/ doi: 10.5194/gmd-9-4185-2016

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.

He, S., Yang, J., Bao, Q., Wang, L., & Wang, B. (2019). 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. Retrieved from https://journals.ametsoc.org/view/journals/clim/32/1/jcli-d-18-0104.1.xml doi: 10.1175/JCLI-D-18-0104.1

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 Journal 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

Hoffman, R. N., Henderson, J. M., Leidner, S. M., Grassotti, C., & Nehrkorn, T. (2006). The response of damaging winds of a simulated tropical cyclone to finite-amplitude perturbations of different variables. *Journal of the Atmospheric Sciences*, *63*(7), 1924 - 1937. Retrieved from `https://journals.ametsoc.org/view/journals/atsc/63/7/jas3720.1.xml` doi: 10.1175/JAS3720.1

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

Jing, R., & Lin, N. (2020). An environment-dependent probabilistic tropical cyclone model. *Journal of Advances in Modeling Earth Systems*, *12*(3), e2019MS001975. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001975` (e2019MS001975 2019MS001975) doi: https://doi.org/10.1029/2019MS001975

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

Kretschmer, M., Coumou, D., Agel, L., Barlow, M., Tziperman, E., & Cohen, J. (2018). More-persistent weak stratospheric polar vortex states linked to cold extremes. *Bulletin of the American Meteorological Society*, *99*(1), 49 - 60. Retrieved from https://journals.ametsoc.org/view/journals/bams/99/1/bams-d-16-0259.1.xml doi: 10.1175/BAMS-D-16-0259.1

Laflamme, E. M., Linder, E., & Pan, Y. (2016). Statistical downscaling of regional climate model output to achieve projections of precipitation extremes. *Weather and Climate Extremes*, *12*, 15-23. Retrieved from https://www.sciencedirect.com/science/article/pii/S221209471530058X doi: https://doi.org/10.1016/j.wace.2015.12.001

Lang, A. L., Pegion, K., & Barnes, E. A. (2020). Introduction to special collection: "bridging weather and climate: Subseasonal-to-seasonal (s2s) prediction". *Journal of Geophysical Research: Atmospheres*, *125*(4), e2019JD031833. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019JD031833 (e2019JD031833 2019JD031833) doi: https://doi.org/10.1029/2019JD031833

Lawrence, A. R., Leutbecher, M., & Palmer, T. N. (2009). The characteristics of hessian singular vectors using an advanced data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, *135*(642), 1117-1132. Retrieved from https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.447 doi: https://doi.org/10.1002/qj.447

Lee, C.-Y., Tippett, M. K., Sobel, A. H., & Camargo, S. J. (2018). An environmentally forced tropical cyclone hazard model. *Journal of Advances in Modeling Earth Systems*, *10*(1), 223-241. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017MS001186 doi: https://doi.org/10.1002/2017MS001186

Leutbecher, M. (2005). On ensemble prediction using singular vectors started from forecasts. *Monthly Weather Review*, *133*(10), 3038 - 3046. Retrieved from https://journals.ametsoc.org/view/journals/mwre/133/10/mwr3018.1.xml doi: 10.1175/MWR3018.1

Lucente, D., Rolland, J., Herbert, C., & Bouchet, F. (2021). *Coupling rare event algorithms with data-based learned committor functions using the analogue markov chain.* arXiv. Retrieved from https://arxiv.org/abs/2110.05050 doi: 10.48550/ARXIV.2110.05050

Matthewman, N. J., & Esler, J. G. (2011). Stratospheric sudden warmings as self-tuning resonances. part i: Vortex splitting events. *Journal of the Atmospheric Sciences*, *68*(11), 2481 - 2504. Retrieved from https://journals.ametsoc.org/view/journals/atsc/68/11/jas-d-11-07.1.xml doi: 10.1175/JAS-D-11-07.1

Maycock, A. C., & Hitchcock, P. (2015). Do split and displacement sudden stratospheric warmings have different annular mode signatures? *Geophysical Research Letters*, *42*(24), 10,943-10,951. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015GL066754 doi: https://doi.org/10.1002/2015GL066754

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

Metzner, P., Schütte, C., & Vanden-Eijnden, E. (2009). Transition path theory for markov jump processes. *Multiscale Modeling & Simulation*, *7*(3), 1192-1219. Retrieved from https://doi.org/10.1137/070699500 doi: 10.1137/070699500

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

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

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

Mohamad, M. A., & Sapsis, T. P. (2018). Sequential sampling strategy for extreme event statistics in nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, *115*(44), 11138-11143. Retrieved from https://www.pnas.org/doi/abs/10.1073/pnas.1813263115 doi: 10.1073/pnas.1813263115

Mureau, R., Molteni, F., & Palmer, T. N. (1993). Ensemble prediction using dynamically conditioned perturbations. *Quarterly Journal of the Royal Meteorological Society*, *119*(510), 299-323. Retrieved from https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.49711951005 doi: https://doi.org/10.1002/qj.49711951005

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.

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'Brien, T. A., Collins, W. D., Kashinath, K., Rübel, O., Byna, S., Gu, J., ... Ullrich, P. A. (2016). Resolution dependence of precipitation statistical fidelity in hindcast simulations. *Journal of Advances in Modeling Earth Systems*, *8*(2), 976-990. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016MS000671 doi: https://doi.org/10.1002/2016MS000671

O'Callaghan, A., Joshi, M., Stevens, D., & Mitchell, D. (2014). The effects of different sudden stratospheric warming types on the ocean. *Geophysical Research Letters*, *41*(21), 7739-7745. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014GL062179 doi: https://doi.org/10.1002/2014GL062179

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

Palmer, T. N., Buizza, R., Doblas-Reyes, F., Jung, T., Leutbecher, M., Shutts, G. J., … Weisheimer, A. (2009). Stochastic parametrization and model uncertainty. *ECMWF Technical Memoranda*.

Palmer, T. N., Gelaro, R., Barkmeijer, J., & Buizza, R. (1998). Singular vectors, metrics, and adaptive observations. *Journal of the Atmospheric Sciences*, *55*(4), 633 - 653. Retrieved from https://journals.ametsoc.org/view/journals/atsc/55/4/1520-0469_1998_055_0633_svmaao_2.0.co_2.xml doi: 10.1175/1520-0469(1998)055⟨0633:SVMAAO⟩2.0.CO;2

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

Poli, P., Hersbach, H., Dee, D. P., Berrisford, P., Simmons, A. J., Vitart, F., … Fisher, M. (2016). Era-20c: An atmospheric reanalysis of the twentieth century. *Journal of Climate*, *29*(11), 4083 - 4097. Retrieved from https://journals.ametsoc.org/view/journals/clim/29/11/jcli-d-15-0556.1.xml doi: 10.1175/JCLI-D-15-0556.1

Rabier, F., Klinker, E., Courtier, P., & Hollingsworth, A. (1996). Sensitivity of forecast errors to initial conditions. *Quarterly Journal of the Royal Meteorological Society*, *122*(529), 121-150. Retrieved from https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.49712252906 doi: https://doi.org/10.1002/qj.49712252906

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

Scaife, A. A., Karpechko, A. Y., Baldwin, M. P., Brookshaw, A., Butler, A. H., Eade, R., … Smith, D. (2016). Seasonal winter forecasts and the stratosphere. *Atmospheric Science Letters*, *17*(1), 51-56. Retrieved from `https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/asl.598` doi: https://doi.org/10.1002/asl.598

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

Simonnet, E., Rolland, J., & Bouchet, F. (2021). Multistability and rare spontaneous transitions in barotropic beta-plane turbulence. *Journal of the Atmospheric Sciences*, *78*(6), 1889 - 1911. Retrieved from `https://journals.ametsoc.org/view/journals/atsc/78/6/JAS-D-20-0279.1.xml` doi: 10.1175/JAS-D-20-0279.1

Sobel, A. H., Wing, A. A., Camargo, S. J., Patricola, C. M., Vecchi, G. A., Lee, C.-Y., & Tippett, M. K. (2021). Tropical cyclone frequency. *Earth's Future*, *9*(12), e2021EF002275. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021EF002275` (e2021EF002275 2021EF002275) doi: https://doi.org/10.1029/2021EF002275

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

Trevisan, A., Pancotti, F., & Molteni, F. (2001). Ensemble prediction in a model with flow regimes. *Quarterly Journal of the Royal Meteorological Society*, *127*(572), 343-358. Retrieved from `https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.49712757206` doi: https://doi.org/10.1002/qj.49712757206

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
Vanden-Eijnden, E., & Weare, J. (2013). Data assimilation in the low noise regime with application to the kuroshio. *Monthly Weather Review*, *141*(6), 1822 - 1841. Retrieved from https://journals.ametsoc.org/view/journals/mwre/141/6/mwr-d-12-00060.1.xml doi: 10.1175/MWR-D-12-00060.1
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.
Weare, J. (2009). Particle filtering with path sampling and an application to a bimodal ocean current model. *Journal of Computational Physics*, *228*(12), 4312-4331. Retrieved from https://www.sciencedirect.com/science/article/pii/S0021999109000801 doi: https://doi.org/10.1016/j.jcp.2009.02.033
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