https://doi.org/10.5194/egusphere-2026-230
Preprint. Discussion started: 6 February 2026

Forecast-based attribution of the role of stratospheric variability in weather extremes

William J. M. Seviour[1], Justin Finkel[2], Philip Rupp[3], Regan Mudhar[1,4], Amy H. Butler[5], Chaim I. Garfinkel[6], Peter Hitchcock[7], Blanca Ayarzagüena[8], Dong-Chan Hong[9], Yu-Kyung Hyun[10], Hera Kim[11], Eun-Pa Lim[12], Daniel De Maeseneire[8], Gabriele Messori[13,14,15], Gerbrand Koren[16], Michael Sigmond[17], Isla R. Simpson[18], and Seok-Woo Son[9]

[1]Department of Mathematics and Statistics, University of Exeter, Exeter, UK
[2]Data Science Institute and Department of Geophysical Sciences, University of Chicago, Chicago, IL, USA
[3]Meteorological Institute Munich, Ludwig-Maximilians University, Munich, Germany
[4]Institute of Earth Surface Dynamics, Université de Lausanne, Lausanne, Switzerland
[5]NOAA Chemical Sciences Laboratory, Boulder, CO, USA
[6]Fredy and Nadine Herrmann Institute of Earth Sciences, Hebrew University, Jerusalem, Israel
[7]Department of Earth and Atmospheric Sciences, Cornell University, Ithaca, NY, USA
[8]Department of Earth Physics and Astrophysics, Facultad de CC. Físicas, Universidad Complutense Madrid, Spain
[9]School of Earth and Environmental Sciences, Seoul National University, Seoul, Republic of Korea
[10]National Institute of Meteorological Sciences, Korea Meteorological Administration, Jeju, Republic of Korea
[11]Department of Atmospheric Sciences, University of Washington, Seattle, WA, USA
[12]Bureau of Meteorology, Melbourne, Australia
[13]Department of Earth Sciences, Uppsala University, Uppsala, Sweden
[14]Swedish Centre for Impacts of Climate Extremes (climes), Uppsala University, Uppsala, Sweden
[15]Department of Meteorology, Stockholm University, Stockholm, Sweden
[16]Copernicus Institute of Sustainable Development, Utrecht University, Utrecht, Netherlands
[17]Canadian Centre for Climate Modelling and Analysis, Environment and Climate Change Canada, Victoria, British Columbia, Canada
[18]NSF National Center for Atmospheric Research, Boulder, CO, USA

Correspondence: William J. M. Seviour (w.seviour@exeter.ac.uk)

Abstract. Variability of the stratospheric polar vortex, particularly its dramatic breakdown during sudden stratospheric warming (SSW) events, has been linked to a number of surface weather extremes. However, attributing the role of stratospheric variability in a specific observed weather extreme, rather than an abstracted class of extremes, has proved highly challenging. Here we use an ensemble of subseasonal forecast simulations from 7 forecast systems participating in the Stratospheric Nudging and Predictable Surface Impacts (SNAPSI) project to carry out this task. By comparing the likelihood of extreme events in free-running forecasts to those with the zonal-mean stratospheric state nudged towards its observed or climatological evolution (while the troposphere is freely-evolving), we are able to calculate the changes in the risk and severity of extremes due to the occurrence, or non-occurrence, of an SSW. We focus on three case-study events: (i) the 2018 boreal SSW and subsequent Eurasian cold air outbreak and snowfall, (ii) the 2019 boreal SSW and subsequent North American cold air outbreak, and (iii) the 2019 austral near-SSW and subsequent Australian heat wave. Through an extreme value statistical analysis, we find in all three cases a significant stratospheric contribution to the risk of relevant weather extremes. In case (i), improving the SSW

https://doi.org/10.5194/egusphere-2026-230
Preprint. Discussion started: 6 February 2026

prediction by nudging as much as doubles the forecast risk of extreme Eurasian cold and UK snow. The differences in risk and severity between experiments nudged to the SSW and to climatology are relatively insensitive to the lead time before the cold air outbreak of case (i). By contrast, in case (ii) this difference only emerges at short lead times before the event, indicating a stratospheric influence on this event that is dependent on the tropospheric state. For case (iii) we find a stronger and more robust stratospheric impact on the severity of the Australian heat wave than on its risk, with the latter being highly sensitive to model bias. The methodology outlined here, including both the experimental design and the semi-parametric approaches for calculating risks, can be applied to attribute several other internal climate system drivers of extreme event risk.

1 Introduction

Following a severe weather event it is regularly asked what factors might have influenced its likelihood or severity. Of such factors, most interest has focussed on the role of anthropogenic climate change and there are now several hundred studies investigating the human influence on a wide range of extremes such as floods, droughts, heat waves, and storms (Stott et al., 2016; Otto, 2017, 2023; Seneviratne et al., 2021; Faranda et al., 2024). The field of extreme event attribution has also evolved to focus not just on weather events themselves, but also on their economic, human health, and ecosystem impacts (e.g. Mitchell et al., 2016; Perkins-Kirkpatrick et al., 2024). Relatively less attention has been paid to attributing the roles of internal climate system processes in extreme event probabilities. Here we will explore the role of one of these processes: polar stratospheric variability.

Methodologies for extreme event attribution can be broadly categorized into either probabilistic or storyline approaches (Noy et al., 2024). The probabilistic approach (Allen, 2003) evaluates changes in the likelihood of an abstracted class of event (such as annual maximum temperature) rather than a specific observed event (such as a heatwave). In contrast, storyline approaches focus on physical mechanisms and the change in magnitude of the particular event (Hoerling et al., 2013), which can be achieved through the use of circulation analogues (Shepherd, 2016), nudged model simulations (van Garderen et al., 2021; Sánchez-Benítez et al., 2022), perturbed reanalysis (Hawkins et al., 2023), or ensemble boosting (Gessner et al., 2021). Both methods have significant drawbacks. For instance, probabilistic approaches may give misleading results if the response of the abstracted class of event is not representative of the given observed event, while storyline approaches cannot determine changes to an event's likelihood, only its magnitude or a plausible range of magnitudes. These differing approaches have led to some seemingly contradictory conclusions (Otto et al., 2012; Hauser et al., 2017), and both are further limited by the inability of climate model simulations to capture the dynamics underlying many extreme events (Bellprat and Doblas-Reyes, 2016). Leach et al. (2021, 2024) proposed that an attribution approach based on medium-to-extended-range forecasts can help mitigate the above deficiencies. They showed that by varying the lead-time of forecasts the approach can move flexibly between a probabilistic framing (at long lead times) and storyline framing (at short lead times), and by using a successful forecast, the ability of the model to capture the given extreme event and its causes can be ensured.

We will here aim to adapt the forecast-based attribution methodology to investigate the role of stratospheric variability in three case studies of extreme weather events. The ensemble of subseasonal forecast experiments performed within the

https://doi.org/10.5194/egusphere-2026-230
Preprint. Discussion started: 6 February 2026

Stratospheric Nudging And Predictable Surface Impacts (SNAPSI) project (Hitchcock et al., 2022) will form the basis of our study. These consist of three core experiments: a standard free-running forecast ensemble, an ensemble in which the zonal-mean stratospheric state is nudged towards observations, and an ensemble in which the zonal-mean stratospheric state is nudged towards time-evolving climatology. In the typical framing of event attribution, we can view the two nudged ensembles as *counterfactual* scenarios in which the future evolution of the polar stratosphere can be predicted with either near-perfect skill or no skill (beyond that provided by climatology). By comparing extreme event probabilities in the counterfactual and standard forecasts, we will aim to answer two related questions:

1. What is the stratospheric contribution to the risk of a given extreme event?
2. How much can improved (or degraded) forecasts of the stratosphere improve (or degrade) the prediction of a given extreme event?

We will find that the answers to these questions depend on the event in question.

It is well established that polar stratospheric variability can be associated with a variety of extreme weather events, including cold air outbreaks, heatwaves, storms, and precipitation extremes (Domeisen and Butler, 2020). This link has been most clearly shown following sudden stratospheric warming (SSW) events (Baldwin et al., 2021), in which the stratospheric polar vortex rapidly breaks down in winter in the Northern Hemisphere and spring in the Southern Hemisphere, the impacts of which may propagate downwards and influence large-scale circulation patterns such as annular modes during the following two months. Several studies have used observational data to identify SSW impacts on specific surface extremes through analyses of weather regimes and dynamical processes such as wave propagation and reflection (Ayarzagüena et al., 2018; Matthias and Kretschmer, 2020; Lim et al., 2021). However, in such studies the lack of counterfactuals makes it impossible to quantify stratospheric impact on the risk or magnitude of a weather extreme, only to infer it qualitatively. Other studies have quantified a stratospheric impact through statistical causal inference (Kretschmer et al., 2018; Huang et al., 2021) or analysis of subseasonal-to-seasonal hindcasts (Spaeth and Birner, 2022; Rao et al., 2025), but these studies calculate an average over many SSW events, rather than the impact of a given event. SSWs vary widely in their tropospheric precursors, magnitude, and morphology (e.g. Charlton and Polvani, 2007) so it is reasonable to expect that their impacts may be similarly varied. This expectation is further motivated by the wide range of surface weather events observed following SSWs. For instance, only around half of boreal SSWs are followed by European cold spells (Hall et al., 2023); likewise, many cold spells occur in the absence of significant stratospheric variability. In summary, while a mean downward impact of SSWs is well established, a quantitative measure of this impact on an event-by-event basis has remained challenging (see further discussion in the review of Butler et al., 2019).

Kautz et al. (2020) address this challenge with an approach that is similar to ours, using nudging of the zonal-mean stratospheric state to quantify its contribution to extreme weather risk. Specifically, they investigate the boreal major SSW of 12 February 2018 (one of the three case studies considered here), using the European Centre for Medium-range Weather Forecasts (ECMWF) Integrated Forecasting System (IFS) (one of the 7 models considered here). They find that the nudged forecasts initialized on 1 February 2018 have about a 9-times higher risk of a Eurasian cold spell than do the standard forecasts (absolute risks of 45% and 5% respectively). As well as considering more case studies and forecast systems, we expand upon Kautz

https://doi.org/10.5194/egusphere-2026-230
Preprint. Discussion started: 6 February 2026

et al. (2020) in two other ways. First, we consider lead-time dependence by initializing all experiments at two dates for each case study. Second, we develop a semi-parametric approach to quantifying risk over a broad range of severity thresholds, in this case using extreme value distributions but easily applicable to other parametric families.

The nudging approach we use here is not limited to studies of SSWs and indeed has been applied to several other aspects of atmospheric variability, such as the quasi-biennial oscillation (Martin et al., 2021), as well as broader climate variability, such as sea-ice concentrations (Audette and Kushner, 2022) and sea-surface temperatures (Amaya et al., 2019). We therefore hope that the general forecast-based "process attribution" methodology described in this study can be adapted for quantifying the roles of other drivers of extreme events, which in turn can aid forecasters and risk analysts, and be used to compare anthropogenic and natural risks.

The remainder of this article is structured as follows. Section 2.1 gives a general description of the SNAPSI experimental protocol and describes the three case study events, while Sect. 2.2 explains the extreme value statistics approach. Section 3.1 then presents the attribution of extreme temperatures to stratospheric variability in each of the case study events. We then briefly investigate the attribution of extreme snow accumulation in Sect. 3.2; the aim being to illustrate how our methodology can be adapted for different variables, and, with this in mind, we limit our discussion here to the case of the southern UK and Ireland following the 12 February 2018 SSW case study. Finally, a discussion and conclusions of our results is given in Sect. 4.

2 Methodology

2.1 Experimental design and case studies

SNAPSI experiments consist of retrospective forecasts of three case study stratospheric polar vortex disruption events: the boreal major SSW of 12 February 2018 (which we refer to as SSWFeb18), the boreal major SSW of 2 January 2019 (SSWJan19), and the austral minor warming of 18 September 2019 (SSWSep19). SSWFeb18 and SSWJan19 are referred to as *major* warmings as the zonal-mean zonal wind at 10 hPa, 60°N reversed to become easterly, and their central dates are those on which the winds reversed. In contrast, this threshold (at 60°S) was not crossed for SSWSep19, although the zonal winds did decelerate rapidly; in SSWSep19 the central date is taken to be that of minimum zonal wind at 10 hPa, 60°S. For each case study we consider three experiments:

free A standard forecast ensemble, which is free-running following initialization.

nudged An ensemble in which stratospheric zonal-mean zonal wind and temperature are relaxed (i.e. nudged) to their observed evolution. The lower boundary of nudging is at 90 hPa, with a ramp to full strength at 50 hPa and above.

control An ensemble in which the relaxation is instead towards climatological mean values, which evolve in time following the seasonal cycle.

https://doi.org/10.5194/egusphere-2026-230
Preprint. Discussion started: 6 February 2026

SSW event	Initialization dates	Time window of interest	Region of interest
SSWFeb18: 12 Feb 2018	25 Jan 2018, 8 Feb 2018	21 Feb - 8 Mar, 2018	$[10°\mathrm{W}, 130°\mathrm{E}] \times [50°\mathrm{N}, 65°\mathrm{N}]$ (Eurasia)
SSWJan19: 2 Jan 2019	13 Dec 2018, 8 Jan 2019	15 Jan - 27 Jan, 2019	$[102°\mathrm{W}, 55°\mathrm{W}] \times [40°\mathrm{N}, 60°\mathrm{N}]$ (NE. N. America)
SSWSep19: 18 Sep 2019	29 Aug 2019, 1 Oct 2019	2 Oct - 15 Nov, 2019	$[112°\mathrm{E}, 154°\mathrm{E}] \times [46°\mathrm{S}, 10°\mathrm{S}]$ (Australia)

Table 1. Case study dates and spatial and temporal regions of interest for analyzing temperature extremes.

Center	Model name
Canadian Centre for Climate Modelling and Analysis (CCCma)	CanESM5
National Center for Atmospheric Research (NCAR)	CESM2-CAM6
Météo France	CNRM-CM6-1
Met Office (UKMO)	UKMO-GloSea6
Korea Meteorological Administration (KMA)	KMA-GloSea6
Seoul National University (SNU)	GRIMs
European Centre for Medium-Range Weather Forecasts (ECMWF)	IFS

Table 2. Modeling centers and model names for the 7 forecast systems analyzed.

The observed evolution and climatology used for nudging are derived from the ERA5 reanalysis (Hersbach et al., 2020). Each experiment is initialized at two different dates for each case study (given in Table 1), and each consists of a 50-member ensemble, generated according to a methodology that varies by forecast system. For example, the IFS generates oceanic and atmospheric perturbations separately, the atmospheric component combining singular vectors and ensemble data assimilation. The same set of perturbations is applied to the free, control, and nudged experiments for fair comparisons. We analyze the output of 7 forecasting systems contributing to SNAPSI, which at the time of analysis had the data necessary (archived on the JASMIN facility; Lawrence et al., 2013) for our study available. These systems are listed in Table 2. A detailed description of the observed dynamical evolution of the three events and of the SNAPSI experimental design is given by Hitchcock et al. (2022), the impact of nudging on upward wave activity is investigated by Ayarzagüena et al. (2025), and an analysis of surface predictability within the SNAPSI experiments is given by Kim et al. (2026). However, to allow for interpretation of our later results, we illustrate the ensemble-mean impact of nudging on the polar mid-stratospheric zonal-mean zonal wind (at 10 hPa, 60°N/S) in Fig. 1.

SSWFeb18 (Fig. 1(a)) is seen to be poorly predicted by the *free* experiment at the early initialization, but well predicted at the late initialization. Indeed, for most models, the early initialization *free* experiment has a stronger polar vortex than the *control*. For SSWJan19 (Fig. 1(b)) , there is a smaller difference between early initialization *free* and *nudged* experiments, with some *free* models successfully predicting an SSW despite being initialized 20 days before it occurred. The late initialization for SSWJan19 comes after the SSW central date (this was chosen due to the longer lag between the SSW and surface anomalies, described further in Hitchcock et al. (2022)), and the subsequent recovery of the polar vortex from this time is well predicted by the *free* experiments, with a large difference between these and *control* until early February. Similarly, for SSWSep19 (Fig.

https://doi.org/10.5194/egusphere-2026-230
Preprint. Discussion started: 6 February 2026

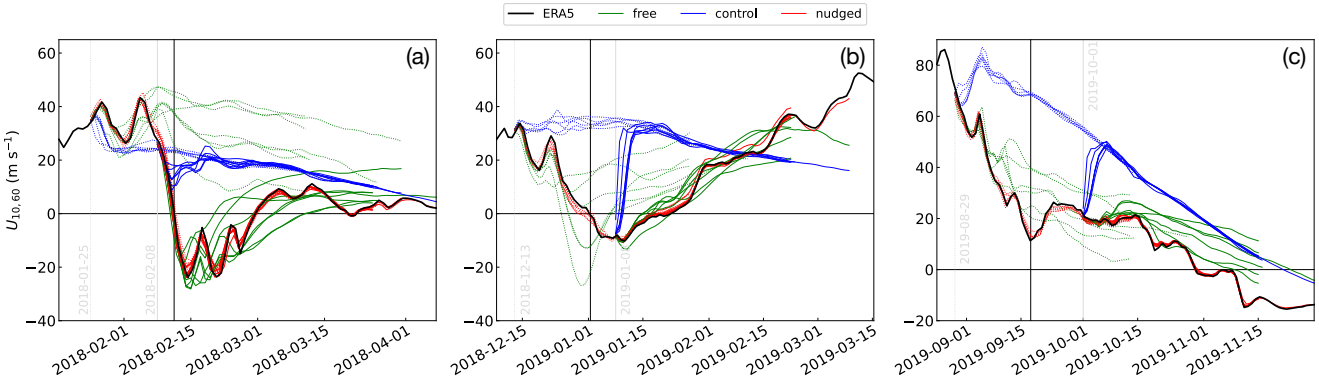

Figure 1. Stratospheric zonal-mean zonal wind evolution for the three case study events of SSWFeb18 (a), SSWJan19 (b) and SSWSep19 (c), in ERA5 reanalysis (black) and the three SNAPSI experiments (colors). Values are shown at 10 hPa and 60°N (a,b) or 60°S (c). Colored lines represent ensemble means of the 7 individual forecast systems in our analysis (excepting CanESM5 in (b) and (c), for which the necessary data was not available). Dashed colored lines are used for the early initialization, solid lines for the late initialization. Solid vertical lines indicate the central date for each event and dashed vertical lines indicate the experiment initialization dates.

1(c)), the differences between *free* and *nudged* experiments are smaller than with either of those and *control*, at least until November, indicating relatively good skill in the prediction of this event.

One motivation for the choice of these three case study events for SNAPSI was the range of surface weather extremes that were observed in the months following them. Here we will describe some key features of these events, with a focus on surface temperature extremes. In the period following SSWFeb18, from late-February to early-March 2018, negative Northern Annular Mode (NAM) and North Atlantic Oscillation (NAO) patterns were associated with anomalously cold temperatures extending over much of Eurasia (Lü et al., 2020). Concurrent weather events included snow over the UK (which we analyze in Sect. 3.2) (Galvin et al., 2019), as well as enhanced precipitation over Iberia (Ayarzagüena et al., 2018), ending an extended period of drought. An analysis of precipitation within SNAPSI experiments of SSWFeb18 is presented in Dai et al. (2025).

In contrast, during the period following the SSW of SSWJan19 the NAM remained near neutral, and temperature anomalies over Eurasia were comparably muted (Butler et al., 2020; Rao et al., 2020). However, more significant temperature anomalies occurred in North America, with an extensive cold spell in late January 2019 covering much of the Midwestern US and eastern Canada, with several daily cold temperature records being set in this region (Lillo et al., 2021). The dynamical evolution of this event has been described on synoptic scales as relating to an equatorward deviation of the extratropical jet stream as well as a tropopause polar vortex (Lillo et al., 2021), and on larger scales as linked to an Alaskan Ridge weather regime. However, in an analysis of SNAPSI experiments, Lee et al. (2025) find little stratospheric impact on the development of this weather regime. Indeed, by contrasting stratospherically- and tropically-nudged seasonal hindcast experiments, Knight et al. (2021) proposed a stronger tropical influence. Nonetheless, a direct attribution of the stratospheric role in the extreme cold spell has not been carried out.

https://doi.org/10.5194/egusphere-2026-230
Preprint. Discussion started: 6 February 2026

SSWSep19 was the most significant disturbance to the austral stratospheric polar vortex since the major warming of 2002 (Lim et al., 2021). Following the central date of 18 September on which zonal mean winds at 10 hPa, 60°S reached a minimum (in the ERA5 data set), a negative Southern Annular Mode (SAM) pattern slowly descended from the mid-stratosphere, reaching the surface from approximately mid-October. The negative SAM pattern has been associated with hot and dry conditions over eastern Australia during late October to December, which in turn contributed to unprecedented destructive wildfires in the region (Lim et al., 2019, 2021). Stratosphere-troposphere coupling within the SSWSep19 SNAPSI experiments is discussed in detail by Feng et al. (2025), including an analysis of wildfire risk.

Motivated by these observed events we restrict our analysis to broad regions that capture the dominant observed temperature extremes for each case study (Fig. 2). For SSWFeb18 (Fig. 2(a)) this region matches exactly that of Kautz et al. (2020), for SSWJan19 (Fig. 2(b)) we select a region containing the most extreme cold anomalies identified by Lillo et al. (2021), and for SSWSep19 (Fig. 2(c)) we select a region containing the entire continent of Australia, in all cases considering only land areas for extreme temperature analysis. We also define a time window for each case study in which to determine temperature extremes. In doing so, we balance a desire to capture events similar in timing to that observed, while also not overly penalizing models that capture the magnitude of events well but have errors in timing. As such, we define periods that follow the central date of each SSW and bracket the most extreme temperatures observed in reanalysis by a margin of at least 5 days, or as wide as possible while remaining fully inside the common forecast period of both early and late initializations—with an exception of SSWSep19, where we extend the horizon beyond the range of the early forecast in order to bracket the highest extremes seen in the late ensemble. The SSWFeb18 ensembles bracket the extremes comfortably, but the SSWJan19 ensembles end before the greatest extremes and the late SSWSep19 ensembles start just barely before the greatest extremes, making them more sensitive to the choice of time interval. Expanded forecast horizons may be a goal for future experiments, and more systematic sensitivity analysis may be a goal for future attribution methods, but the current setup suffices to demonstrate our core methodology. The spatial and temporal boundaries of our below analysis are shown in Table 1, and Fig 2 shows large and relatively spatially coherent anomalies of temperature extremes within the chosen regions, motivating their suitability for area-averaged analysis. The following spatial maps will focus only on the selected sub-regions.

2.2 Extreme value statistical analysis

For the purposes of our analysis we define *severity* as the spatially averaged and then temporally minimized (for SSWFeb18 and SSWJan19) or maximized (for SSWSep19) temperature. We notably consider absolute temperature, not anomaly, which is better connected to some major societal impacts (Sheridan et al., 2019), and on which several downstream impacts (e.g. snow) depend. Crucially, we quantify statistical responses to different nudging interventions using two quantities:

***relative risk* (RR)**, a ratio of probabilities of exceeding some fixed severity, and

***quantile shift* (QS)**, a difference in severity at some fixed exceedance probability.

These two quantities have complementary strengths and weaknesses that allow us to discern robust causal signals that partially overcome model bias.

https://doi.org/10.5194/egusphere-2026-230
Preprint. Discussion started: 6 February 2026

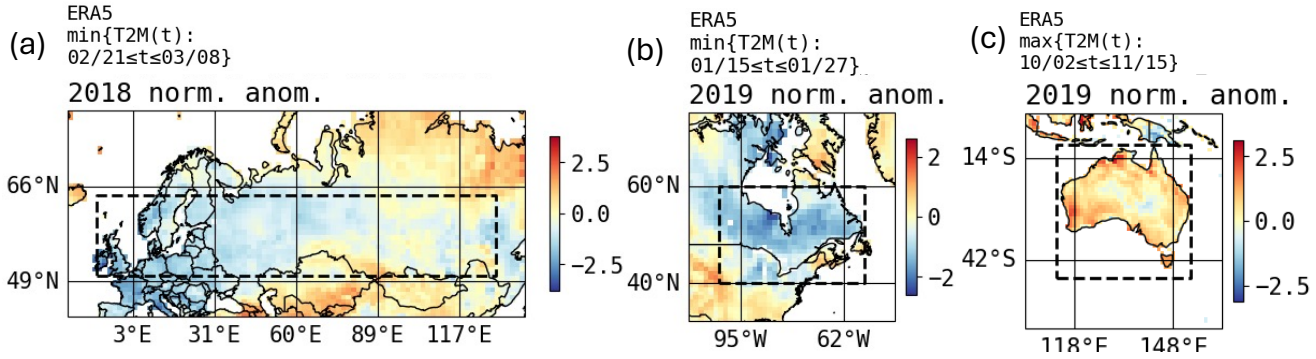

Figure 2. Anomalies of observed (ERA5) near-surface land temperature extremes for the three case studies, during their time windows of interest. Minima are shown for SSWFeb18 (a) and SSWJan19 (b), and maximima for SSWSep19 (c). Values are calculated relative to the climatological mean minimum (a,b) or maximum (c) temperature during the time window of interest at each grid point over 1979-2020, and normalized by the standard deviation over those years. The regions of interest for later area-averaged analysis for each case study are illustrated by dashed black boxes.

We need some notation to summarize our statistical approach. Let g be an index for the data-generating system, either a forecast system or reanalysis, $\in \{\text{CESM2-CAM6}, \ldots, \text{IFS}, \text{ERA5}\}$; e be an index for the experimental nudging protocol $\in \{\textit{control}, \textit{free}, \textit{nudged}\}$ (not applicable to ERA5); i be an index for the initialization date $\in \{\text{early}, \text{late}\}$ (not applicable to ERA5); t be a timestamp within the temporal window $[A, B]$ in which the extremes can occur; m denote the ensemble member, $\in \{1, \ldots, M = 50\}$ (for forecast systems) or $\{1980, ..., 2019\}$ (for ERA5); and $S_{g,e,i,m}(t)$ be the *intensity* simulated by system g, experiment e, initialized from date i, as realized by the mth ensemble member, on day t. When $g = \text{ERA5}$, the subscripts (e, i) are not applicable, and are replaced by $(\cdot, \cdot)$. For all three SSWs, S represents a land-only area-averaged 2-meter temperature $\langle T(t) \rangle$, and then minimized (or maximized for SSWSep19) daily over four six-hourly samples. The areas for averaging and the time intervals of interest are described in the previous section and shown in Table 1.

The format of the ensemble lends itself naturally to extreme value analysis with the method of block maxima (Coles, 2001). For each (g, e, i, m), we define the *severity* as the peak intensity attained, and call it $S^*_{g,e,i,m}$:

$$S^*_{g,e,i,m} = \max_{A \le t \le B} S_{g,e,i,m}(t), \tag{1}$$

and fit a generalized extreme value (GEV) distribution to each ensemble $\{S^*_{e,i,m} : m = 1, \ldots, M\}$, using the method of probability-weighted moments, which tends to be more stable than maximum-likelihood for small sample sizes (Hosking et al., 1985). In other words, we consider $S^*_{g,e,i}$ to be a random variable whose cumulative distribution function (CDF) is defined for any value s as

$$\mathbb{P}\{S^*_{g,e,i} \le s\} = F(s; \mu_{g,e,i}, \sigma_{g,e,i}, \xi_{g,e,i}) \text{ where } F(s; \mu, \sigma, \xi) = \begin{cases} \exp\left[-\left(1 + \xi \frac{s-\mu}{\sigma}\right)_+^{-1/\xi} \right] & \text{if } \xi \neq 0 \\ \exp\left[-\exp\left(-\frac{s-\mu}{\sigma} \right)_+ \right] & \text{if } \xi = 0, \end{cases} \tag{2}$$

https://doi.org/10.5194/egusphere-2026-230
Preprint. Discussion started: 6 February 2026

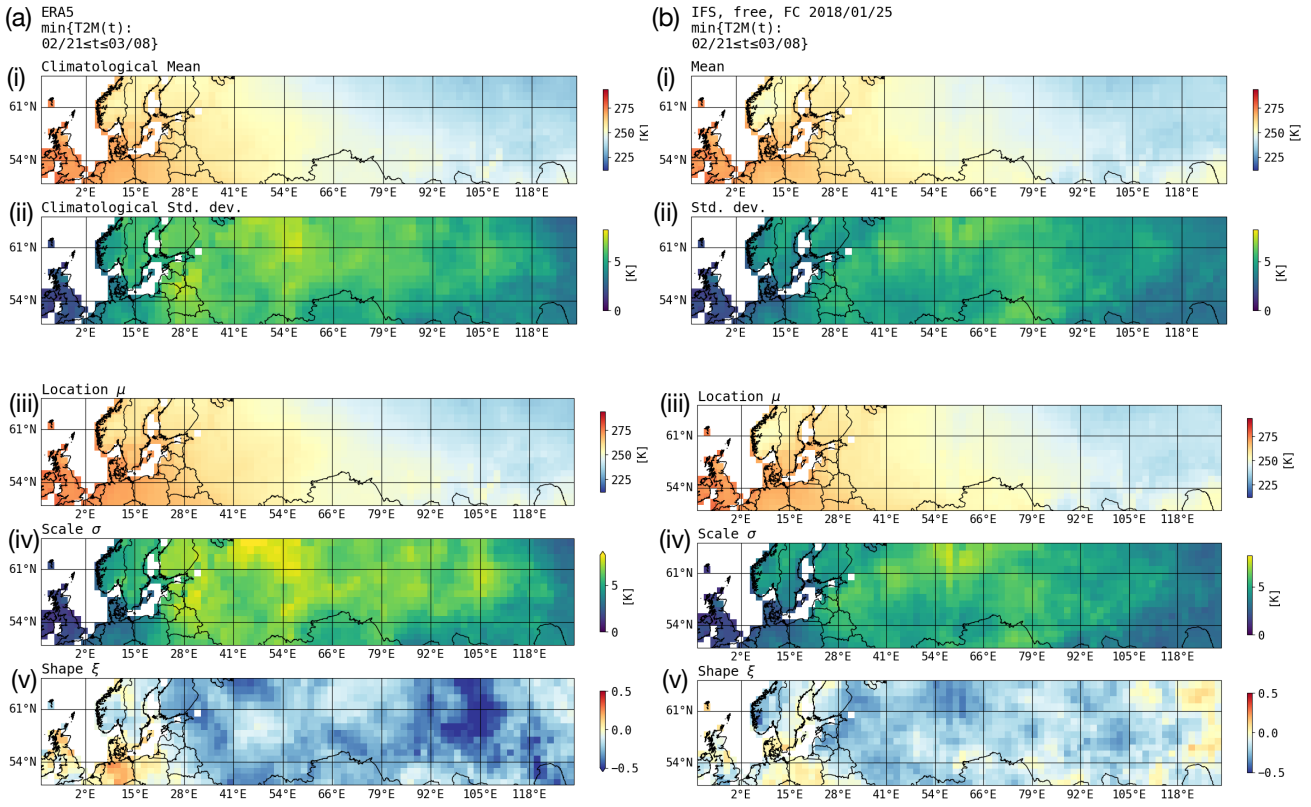

Figure 3. Statistical summary maps of surface temperature minima in the 16-day interval from 21 February. (a) Shows values over the 'ensemble' of years from 1980–2020 in ERA5, (b) shows values from the IFS *free* ensemble for SSWFeb18 initialized on 2018-01-25 (early). Panels (i) show mean minima, (ii) standard deviation of minima, and (iii, iv, v) show GEV parameters of location, scale, and shape respectively. The similarity in spatial patterns between "normal" summary statistics (mean, standard deviation) with analogous GEV summary statistics (location, scale) affirms qualitatively that the GEV distribution is an appropriate parametric choice. The similarity between panels (a) and (b) affirms qualitatively that the early forecast is not highly skillful, but statistically accurate with respect to climatology.

and $(\cdot)_+ := \max(\cdot, 0)$. We will also refer below to the *complementary* CDF, or CCDF, which is the exceedance probability: CCDF $= 1 -$ CDF. The three parameters of the GEV (μ, σ, ξ) are called the location, scale, and shape respectively, which we abbreviate as a vector $\theta = [\mu, \sigma, \xi]$.

We emphatically acknowledge that the GEV is not *a priori* guaranteed to be appropriate, as this would require S^* to be the maximum of a set of random variables that is *large* in number (much more than 2 weeks), *independent* (unlike consecutive days), and *identically distributed* (devoid of any seasonal cycle). However, the GEV was found to capture the distribution of temperature extrema better than a Gaussian, and we use it only as a regularizing tool for the relatively small ensemble size of $M = 50$. It is worth exploring other distribution choices more thoroughly—RR and QS are equally well-defined for any parametric family—but GEV was a natural, illustrative, and sufficiently fitting distribution family for the present demonstration.

https://doi.org/10.5194/egusphere-2026-230
Preprint. Discussion started: 6 February 2026

To provide some context and aid the interpretation of the GEV results to follow, we provide in Fig. 3(a) a visual summary of distributional parameters derived from ERA5. Specifically, we consider the 16-day time interval (as in Table 1) from 21 February (ending on 7 or 8 March depending on the year, inclusive) each year from 1980 to 2019 as a different ensemble member (thus, $M = 40$ in the ERA5 case). We compute the severity for each year, but for geographic illustration, the area averages are performed separately on each of the boxes tiling the domain (restricting the average to land pixels in all boxes). The mean severity and its standard deviation are shown in Figs. 3a(i,ii). Similar maps for SSWJan19 and SSWSep19, over their respective areas of interest, can be found in the supplement.

We then fit a GEV distribution to each grid box, and plot the corresponding parameter maps in Fig. 3(a)(iii,iv,v). The mean and GEV location have a clearly similar spatial pattern as do the standard deviation and GEV scale, confirming that these two GEV parameters represent what their names suggest. The shape parameter (Fig. 3(a)(v)) stands apart as a subtler-to-interpret, but highly influential parameter: a negative shape parameter imposes a bound on severity equal to $\mu - \sigma/\xi$, whereas a positive shape parameter allows for unbounded severity [see Eq. (2)]. The map indicates negative shape nearly everywhere, which is usually the case in GEV analysis of temperature extremes (Huang et al., 2016), despite the lack of any *a priori* bounds on surface temperature (Krakauer, 2024). We also illustrate equivalent distribution parameters for the 2018-01-25 initialization of the *free* ensemble of the IFS model in Fig. 3(b). In general, the IFS model parameters are similar to those from reanalysis indicating that the ensemble is a close match to the climatology of extremes, but has little precision from this initialization date.

https://doi.org/10.5194/egusphere-2026-230
Preprint. Discussion started: 6 February 2026

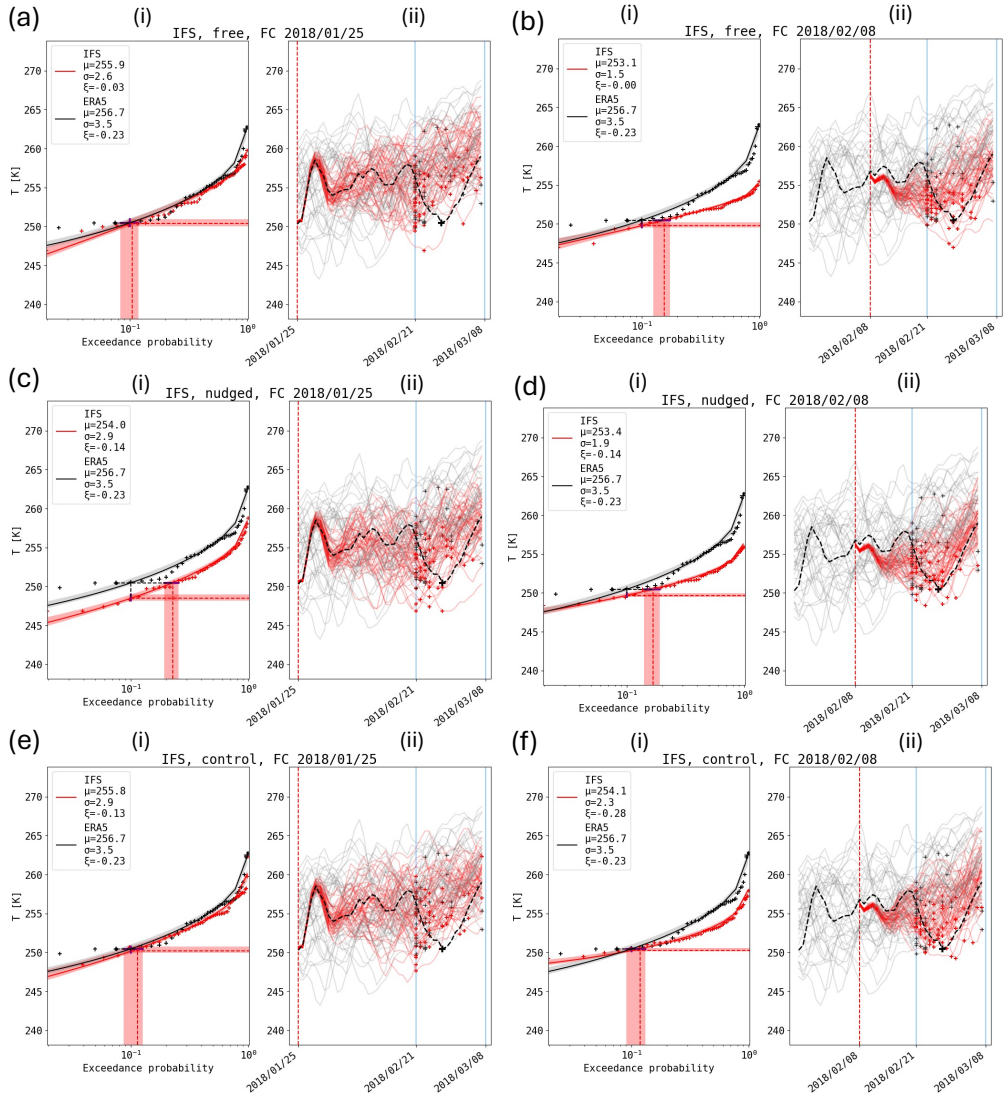

Figure 4. Forecast ensembles for SSWFeb18 produced by the IFS model: severity distributions (i) and timeseries (ii) of daily-minimum surface temperature, T, averaged over the full spatial region in Fig. 3. Left (a,c,e) and right (b,d,f) columns show early and late forecast dates. Top (a,b), middle (c,d), and bottom (e,f) rows show *free*, *nudged*, and *control* experiments. In more detail, panels (ii) display individual ensemble members' evolution as red timeseries, spreading out from the 2018 reanalysis (dashed black line) following the initialization date (red dashed vertical line). Gray lines show other years ("ensemble members" of ERA5) from the 1980-2020 record. Crosses (red for IFS and black for ERA5) mark the minimum of each member during the time span of interest (demarcated with blue vertical lines). These same minima are plotted as crosses in panels (i), versus the empirical exceedance probability (the probability of being even more extreme, in this case colder; e.g., for the 4th-coldest ensemble member out of 50, $\frac{4}{50} = 0.07$). GEV fits and 50% bootstrapped confidence intervals are overlaid. Dashed lines illustrate the calculation of absolute risk and equivalent severity: the observed severity (larger black crosses in panels (ii)) has, according to the forecast ensemble, an exceedance probability given by the red curve's horizontal position at the same vertical level, and an equivalent-risk severity given by the red curve's vertical position at the same exceedance probability as the 2018 event (according to the ERA5 GEV fit, not the empirical value).

https://doi.org/10.5194/egusphere-2026-230
Preprint. Discussion started: 6 February 2026

Although each grid box sub-region has its own severity distribution, the following quantitative risk analysis only considers full-region severities. Note the full-region GEV parameters are not simply the average of the sub-region GEV parameters, and a legitimate concern with the full-region average is lumping together many different regional climatologies, each of which merits a separate localized analysis. We deem the fits shown in the next section sufficiently good to provide robust results and establish the methodology, but a follow-on paper will quantify the extremal behavior across sub-regions and averaging scales.

3 Results

3.1 Extreme temperatures

We now aim to apply our statistical analysis to compare temperature extremes in the three SNAPSI experiments and in ERA5. Figure 4 summarizes full-region risk analysis on the IFS model for SSWFeb18. The left and right panel groups (i and ii) display results for the early and late initializations, respectively, and the three rows display in order results for *free*, *nudged*, and *control* experiments. For each experiment (in red) and for ERA5 (in black) we show, on the left (i), the *empirical risk function* consisting of ordered pairs of points at $(1/m, T_{(m)})$ for $m = 1, \ldots, M$, where $T_{(m)}$ is the temperature (area-averaged and then time-minimized) of the mth ensemble member (or year, for ERA5) in order of increasing severity (i.e., the severity with a $\frac{1}{m}$ chance, empirically, of being equalled or exceeded). We also plot as solid lines the GEV-fitted risk functions and their 50th percentile-bootstrapped confidence intervals (resampling with replacement 1000 times, re-calculating GEV parameters for each, and shading between the 25th and 75th percentiles).

The dashed lines help to explain how each forecast ensemble is boiled down to two numbers for later inter-model comparison. Starting at the ERA5 SSWFeb18 severity (~ 250 K), we draw a horizontal black dashed line through both the black and the red curves, and read off the horizontal positions of intersection to get *absolute risks* (ARs) of exceeding the observed severity. Then, from the ERA5 absolute risk, we draw a vertical dashed black line to the red curve, and read off the vertical position of intersection to get the modeled *equivalent-risk quantile* (EQ): the severity of an equivalent-risk event in the model ensemble. AR and EQ are then projected onto the corresponding axes, along with error bars. Mathematically, following the notation in Sect. 2.2, for any model g, forcing experiment e and initialization i, we define

$$\mathrm{AR}_{g,e,i} = \mathbb{P}\left\{S^*_{g,e,i} \geq S^*_{\mathrm{ERA5},\cdot,\cdot,2018}\right\} = 1 - F(S^*_{\mathrm{ERA5},\cdot,\cdot,2018}; \theta_{g,e,i}) \tag{3}$$

$$\mathrm{EQ}_{g,e,i} = F^{-1}\Big(F(S^*_{\mathrm{ERA5},\cdot,\cdot,2018}; \theta_{\mathrm{ERA5},\cdot,\cdot}); \theta_{g,e,i}\Big) \tag{4}$$

where F^{-1} is the inverse function of F, also known as the quantile function.

In the following, RR is a ratio of two absolute risks while QS is a difference in quantile. For example,

$$\mathrm{RR} = \frac{\mathrm{AR}_{\mathrm{IFS,nudged,2018\text{-}01\text{-}25}}}{\mathrm{AR}_{\mathrm{IFS,free,2018\text{-}01\text{-}25}}} \quad \text{and} \quad \mathrm{QS} = \mathrm{EQ}_{\mathrm{IFS,nudged,2018\text{-}01\text{-}25}} - \mathrm{EQ}_{\mathrm{IFS,free,2018\text{-}01\text{-}25}} \tag{5}$$

measure two aspects of the response to nudging in the IFS model, initialized on 25 Jan 2018 (the early date). RR and QS have complementary advantages and drawbacks: RR is interpretable on a universal scale, no matter the event type or the physical

https://doi.org/10.5194/egusphere-2026-230
Preprint. Discussion started: 6 February 2026

units, but it may be zero or infinite in the presence of model bias. QS is always finite, but has physical units and characteristic magnitudes that depend, for example, on the spatial scale used for averaging.

In order to further illustrate the calculation of the empirical risk function, the right-hand panels (a-f).(ii) of Fig. 4 show time series of intensity (region-averaged temperature) from the model ensemble and from all years of ERA5. The red and black dots show the location of the minimum temperature during the period of interest, lining up with the locations of the points shown in the empirical risk function. The black dashed line shows the ERA5 time series for 2018. We see that the 2018 minumum temperature during this period, reached on 27 February, is the coldest on record for that day of year, and is the third coldest temperature reached over the whole period of interest. It is also apparent that in all 6 model experiment cases this observed severity lies within the ensemble spread, indicating the IFS model's ability to capture events of this magnitude. The same is true for SSWJan19 and SSWSep19, for which plots equivalent to Fig. 4 are shown in the supplement. However, the same is not true across all models and events, as Fig. 6 will show several infinite, or zero, RRs.

At the earlier initialization, the risk function for all three experiments matches relatively closely that for ERA5 (Fig. 4, left column). This both illustrates that the forecasts from this initialization date have little precision beyond climatology (as previously noted in Fig. 3), and also that the "IFS climatology"—interpreted in the sense of an ensemble that has run just long enough to disperse, not a long multi-year single run—is relatively unbiased. The *control* climatology matches ERA5 climatology even better, which makes some heuristic sense because the *control* forcing makes the stratosphere match the ERA5 climatological *mean* by construction; however, nudging toward a single mean value does not imply matching the full distribution better, and might equally be expected to reduce the variance across members as a result of the lack of dispersion in the stratosphere and its influence on the troposphere. In other words, it is not obvious whether *free* or *control* experiments should better match the climatological *distribution*, especially when the climatological mean stratosphere is a sum over both SSW and non-SSW years and might not even be physically realizable. This is an important caveat for interpretation, and a potential improvement area for experimental design. For the later initialization (Fig. 4, right column), the IFS risk function is notably flatter than that of ERA5, indicative of greater forecast precision at the later initialization.

https://doi.org/10.5194/egusphere-2026-230
Preprint. Discussion started: 6 February 2026

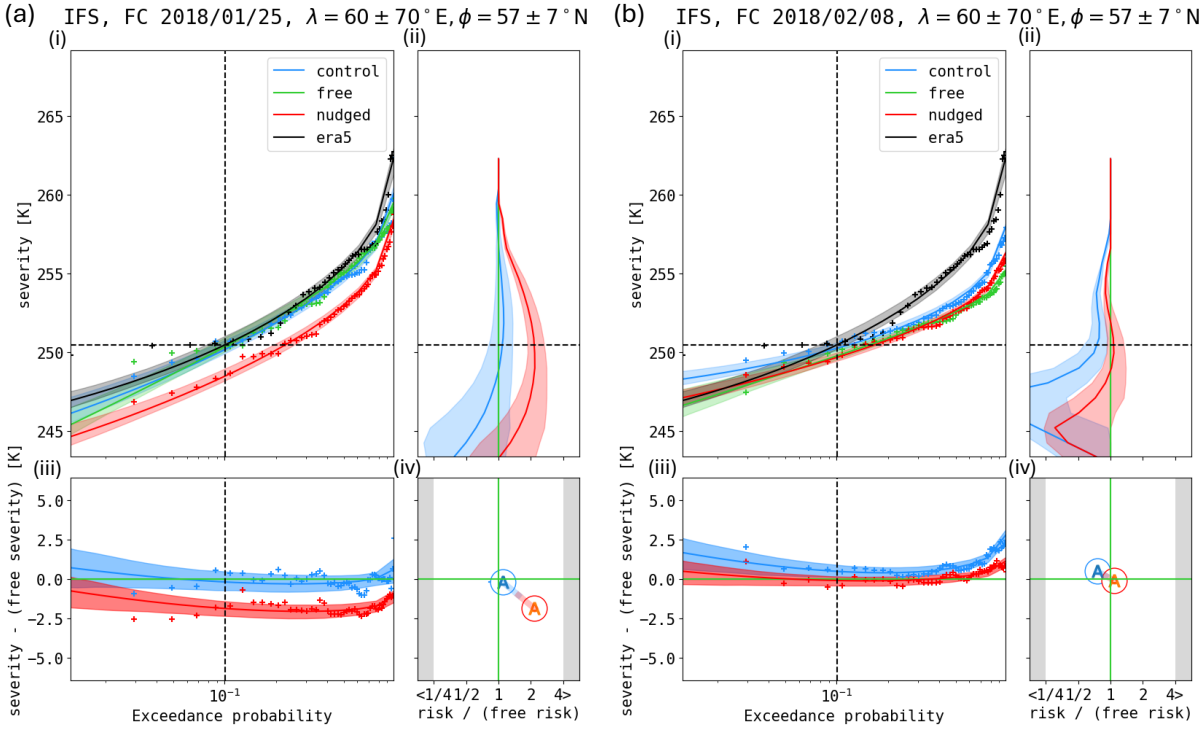

Figure 5. Summary of experimental effects according to IFS for SSWFeb18, for both early (a) and late (b) initialization dates. Panels (i) show severities S^* vs. risk, meaning the probability of an event of even greater severity. Crosses represent empirical risks, solid lines represent GEV-fitted risks, and shaded bands represent 50% confidence intervals of GEV fits based on 1000 bootstrap resamplings. Panels (ii) show RR of the *control* and *nudged* distributions relative to *free*, as a function of severity. Panels (iii) show QSs of the *control* and *nudged* distributions relative to the *free*, as a function of risk, as well as empirical differences (available only when the two ensembles are equal in size). In other words, panels (ii,iii) respectively show horizontal and vertical differences between curves in panels (i). In panels (iv), we summarize each response as a single ordered pair (relative risk, quantile shift), where the reference severity is that observed in ERA5 and the reference quantile is that fitted from the ERA5 GEV at its reference severity. Any ratios $< \frac{1}{4}$ or > 4 would be clipped to the gray shading at the margins (see also Fig. 6).

https://doi.org/10.5194/egusphere-2026-230
Preprint. Discussion started: 6 February 2026

The exceedance probabilities (CCDFs) for the three experiments, and both initialization dates of SSWFeb18 (retaining IFS as our example model), are overlaid in Fig. 5. RR is shown in panels (ii) as a function of severity, and QS is shown in panels (iii) as a function of exceedance probability, which are simply the horizontal ratio and vertical differences between curves in panels (i). The special thresholds given by ERA5 are marked by black dashed lines, whose intersections with RR and QS curves give the nominal RR and QS values plotted as ordered pairs in panels (iv), and later used for inter-model comparison.

In the case of IFS, there is a clear and consistent pattern of influence: nudging toward observations both increases the RR and intensifies the QS relative to the *control* simulation, and the effect is larger for the earlier initialization. Fig. 5(iv) captures this as a downward and rightward displacement of red relative to blue, which shrinks in magnitude at the later date. In the special case that $\xi = 0$ for both experiments and the nominal severity S^* is far into the tail, RR and (log) QS are proportional, since the Gumbel distribution (GEV with $\xi = 0$) is asymptotically exponential and thus the CCDF on a log scale is asymptotically linear:

$$F(s;\mu,\sigma,\xi=0) = \exp\left[-\exp\left(-\frac{s-\mu}{\sigma}\right)\right] \sim 1-\exp\left(-\frac{s-\mu}{\sigma}\right) \implies \log(1-F) \sim -\frac{s-\mu}{\sigma} \text{ as } s \to \infty. \tag{6}$$

However, in general RR and (log) QS are not necessarily proportional, especially if the severity of interest—nominally, the 250.5 K level achieved by ERA5—is near the GEV-implied severity bound where the CCDF must curve sharply. We see this already in the late-initialized *control* experiment in Fig. 5b(i,ii), where the blue curve levels off substantially before exceeding (dipping below) the ERA5 severity, and its relative risk rapidly approaches zero. The implied lower bound on *control* severities is not far off, at $-(\mu - \frac{\sigma}{\xi}) = -\left(-254 - \frac{2.3}{-0.29}\right) = 246$ K (the overall negative sign accounts for considering cold-side extremes). A small positive bias could easily have made the ERA5 severity impossible according to the *control* GEV fit, a conclusion which is physically reasonable but not well-described by relative risk. QS thus provides an important complementary viewpoint.

Let us point out some patterns in Fig. 5 which convey the effects of modulation of the stratospheric state on European cold extremes following SSWFeb18. In order from broad and robust to intricate and more uncertain:

1. Nudging the stratosphere toward the observed SSW intensifies surface cold extremes compared to relaxing the stratosphere toward its climatological mean. The RR (*nudged/free*) exceeds the RR (*control/free*) for the full range of severities in panels (a,b)(ii), and QS (*nudged* − *free*) is more negative than QS (*control* − *free*) in panel (a,b)(iii). Both of those differences exceed the error bar width at, and around, the 2018 ERA5 values, and for both early and late initializations. At the ERA5 severity of 250.5 K, the relative risk is a factor ~ 2 larger for *nudged* than *control* at both initializations; at the ERA5 probability of ~ 0.1, the *nudged* temperature is ~ 2 K and ~ 0.5 K colder than the *control* temperature based on early (panel a) and late (panel b) initializations respectively. IFS thus delivers a clear message that SSWFeb18 disposed the troposphere towards a stronger cold-air outbreak than the climatological average stratospheric evolution would have done.

2. Nudging towards observations (*nudged*) also intensifies the extreme event compared to the *free* simulation, but only for the early initialization. Started from the later date, nudging has a broadly neutral effect because the SSW is already well-predicted in the *free* simulation (Fig. 1(a)). Meanwhile, nudging toward climatology has a neutral effect when started early, and a moderating effect when started late.

https://doi.org/10.5194/egusphere-2026-230
Preprint. Discussion started: 6 February 2026

3. The GEV fit does not capture all the important variation in severities, which is amplified in panels (a,b)(iii) and visible as non-random, oscillatory trends in empirical severities (crosses) about the GEV-derived CCDFs (solid lines). Possible culprits might include sub-asymptotic behavior (where GEV is not the best model), autocorrelations across space and time, and deterministic imprints of common perturbations between *control* and *free* ensembles, as suggested by the similar pattern of scatter between blue and red crosses.

Figures equivalent to Figs. 3, 4 and 5, but for SSWJan19 and SSWSep19 can be found in the supplement. For a succinct inter-model comparison of the three events, Fig. 6 summarizes these results for all three case studies (a: SSWFeb18, b: SSWJan19, and c: SSWSep19) and both initialization dates (i: early and ii: late), reporting for each model and date only the slice of the RR curve evaluated at the ERA5 severity, and the slice of the QS curve evaluated at the ERA5 GEV-fitted risk, as in Figs. 5(a,b)(iv). Error bars are 50th percentile-bootstrapped confidence intervals, as in Fig. 5. A multi-model mean, taken by averaging these two metrics after calculating them for each model, is also shown (large hexagon). The results for SSWFeb18 are relatively consistent across models, and so with those already discussed for IFS. The early initialized *nudged* experiments of all models have both increased risk (on average by a factor of about 2) and increased severity (on average by about 1-2 K) compared to *free* forecasts. The *control* experiments also show (with the exception of CNRM-CM6-1) an increase in risk and severity, though this is more modest, both on average and in most individual models. This might be explained by the fact that the stratospheric polar vortex is on average stronger in the *free* than the *control* experiments for the early initialization (Fig. 1(a)). At the later initialization, *nudged* experiments show little change (relative to *free*) in RR or QS, but *control* experiments show a consistent reduction of both by factors of about 1/2 and 1.5 K respectively.

The other two case studies are not as consistent across models and initializations, which highlights the diversity of SSWs: in surface impacts, in lead-time dependence of predictability, and in how robustly models represent them. In SSWJan19 late nudging towards climatology reduces the risk of cold air outbreaks over North America, and reduces their severity–the same as in SSWFeb18, with close similarity between panels 6(a,b)(ii)–but, unlike SSWFeb18, early nudging or relaxation to climatology in SSWJan19 has a small average effect and little consistency of sign among models. This sensitivity of the stratospheric effect to lead-time suggests state-dependence, a point that will be expanded upon in Sect. 4.

In SSWSep19, nudging towards observations appears to consistently strengthen the severity of hot conditions over Australia in terms of QS, but no consistent or strong signature is seen in terms of RR. Effect sizes are small in both QS and RR. Especially for the late initialization, most RRs are not significantly different from unity, because most models are more biased relative to ERA5 than they are sensitive to different forcing (see Fig. S6 for the example of IFS), which is the very reason why RR is an erratic indicator and QS is needed to extract some meaningful signal. Also, in this case it is not as fair to compare early with late initializations because we chose a target time horizon (through Nov. 15) that extends beyond the range of the early forecast (which ends Oct. 14; see SI Fig. 5). Nevertheless, the larger QS for *nudged* than for *control* in the early initialization forecast, which is consistent between models, reveals the role of the SSW in increasing the temperatures over Australia in the first half of October 2019. The same result in the late initialization forecast further confirms that the SSW did play a role in the observed extremes during spring 2019.

https://doi.org/10.5194/egusphere-2026-230
Preprint. Discussion started: 6 February 2026

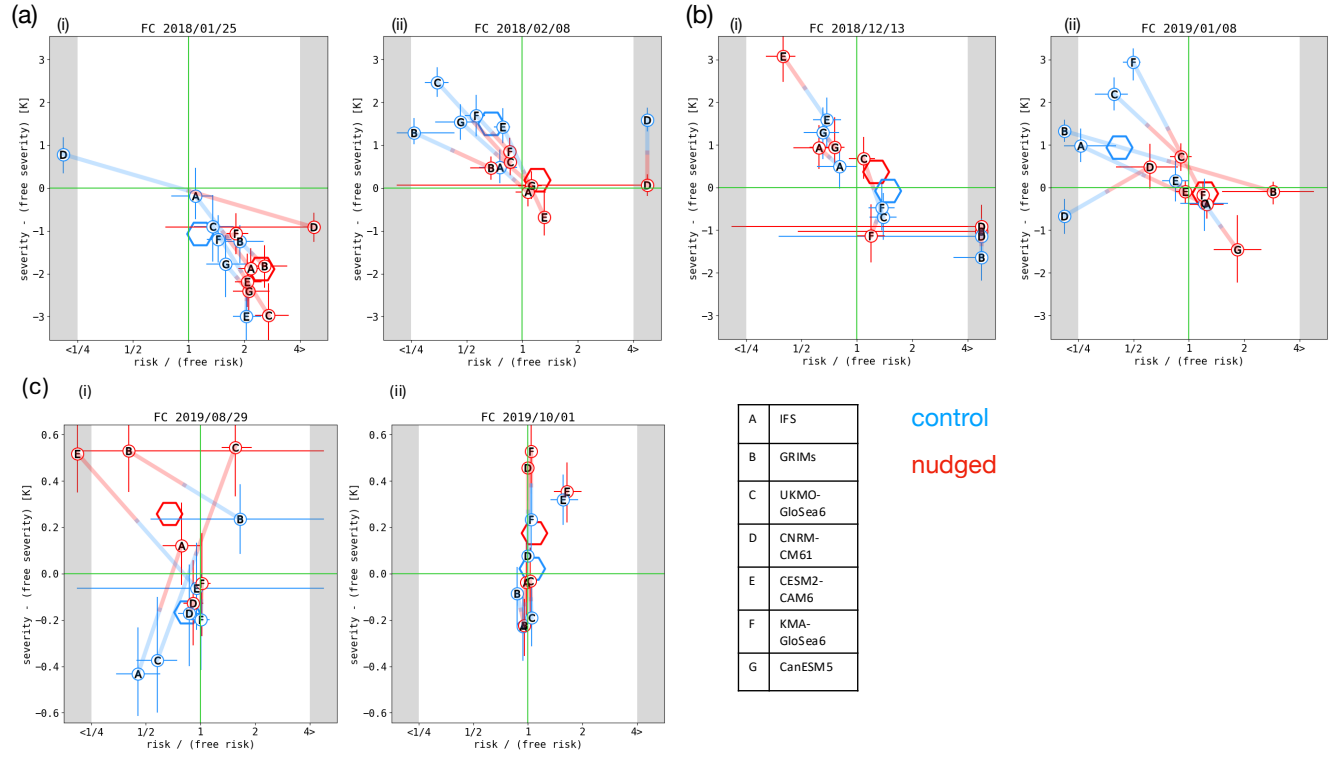

Figure 6. Relative risks (RRs) and quantile shifts (QSs). For each case study of extreme temperatures (a: SSWFeb18, b: SSWJan19, and c: SSWSep19), for each initialization date (i: early and ii: late), for both forcing experiments (blue: *control* and red: *nudged*), and for each model (represented by circled letters, as per the key), the response to forcing is plotted in a two-dimensional space (RR, QS), both defined with respect to the *free* experiment as a baseline (Fig. 5(a,b).(iv) are the IFS-only versions for SSWFeb18). Thick shaded lines connect each model's *free* and *control* responses, to help see patterns in the direction of influence. Multi-model means are represented by the large hexagons, with the relative-risk component being a geometric mean (taken in log space).

In aggregate, we affirm the original hypothesis of the SNAPSI project with regard to surface extremes—that the SSWs significantly affect the predicted risk of extreme surface temperatures—in only some cases: SSWFeb18 (both early and late initializations), SSWJan19 (the late initialization only) and SSWSep19 (both initializations, but with incomplete information for the first, which are truncated too early for a completely fair comparison). The other effects are mixed, and warrant a great deal of further exploration, in particular the geography of impacts, which a following paper will explore.

3.2 Extreme snow

As an illustration of how the attribution methodology described above might be adapted to other variables besides temperature, we will briefly describe an analysis of snow extremes. Specifically, we focus on SSWFeb18 and the extreme snow that accumulated over much of the UK during late February-March 2018 (Galvin et al., 2019). Figure 7 shows snow depth averaged

https://doi.org/10.5194/egusphere-2026-230
Preprint. Discussion started: 6 February 2026

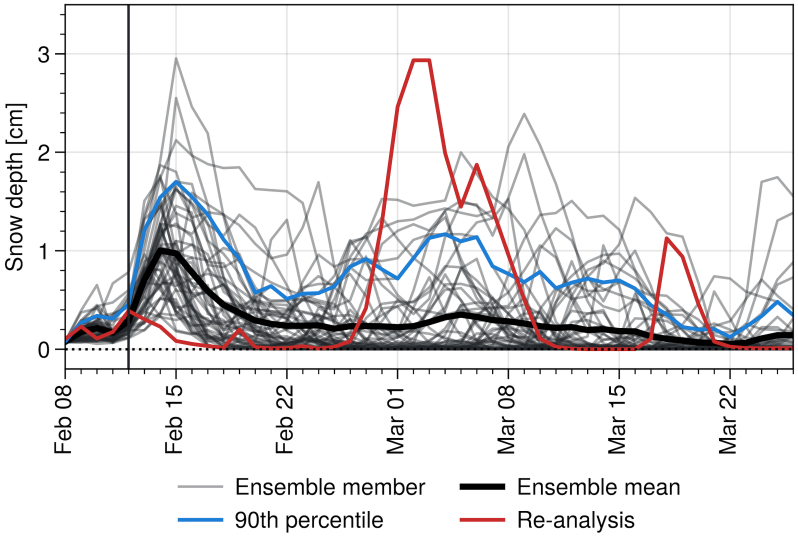

Figure 7. Snow depth averaged over the southern UK and Ireland ($[10°\mathrm{W}, 2°\mathrm{E}] \times [50°\mathrm{N}, 55°\mathrm{N}]$) in 2018, in ERA5 (red), IFS *free* ensemble members (gray). The IFS ensemble mean is in black and the 90th percentile in blue. The vertical black line indicates the SSW onset date.

over land in the box $[10°\mathrm{W}, 2°\mathrm{E}] \times [50°\mathrm{N}, 55°\mathrm{N}]$, constituting the southern UK and Ireland. ERA5 reanalysis (in red) shows a maximum close to 3 cm being reached in early March. The 2018-01-25 initialized *free* forecasts from IFS are shown in gray. No ensemble members simulate an accurate timing and magnitude of this snow event, although there is some indication of an increased probability of high snow in early March (shown by a local maximum in the 90th percentile of the ensemble, in blue). The model also predicts a snow event in mid-February of much greater magnitude than in reanalysis, such that on February 14 all ensemble members have snow depth greater than ERA5.

Figure 8 shows exceedance probabilities (for the same time window as in Table 1) of snow depth for the early initializations of two example models; IFS and CNRM-CM61. As could be seen in the snow depth time series, very few ensemble members simulate snow accumulation maxima close to or above the observed values of about 3 cm (about 5/50 ensemble members in IFS and 2/50 in CNRM-CM61). Second, the parametric fits are much poorer than those for temperature, a problem that is again greater for CNRM-CM61. We hypothesize that the poor GEV fit is due to a very large number of ensemble members having zero (or close to zero) snow during the whole evaluation period (as can be seen in the time series of Fig. 7). Nonetheless, it is apparent that there are systematic differences between the three experiments, with *nudged* giving generally higher exceedance probabilities than *free*, and *control* generally lower probabilities.

Given the poor parametric fits for snow we illustrate a simpler approach to calculating RR based purely on counting ensemble members. Here, RR is simply the ratio of the number of members exceeding a given threshold, and is shown for *nudged/free* and *control/free* in Fig. 9 for 5 of the 7 models for which appropriate snow depth data was available, as well as a multi-model mean. We choose a threshold of 0.8 cm, which is much lower than the observed maximum near 3 cm; this is to avoid large errors or infinite RR caused by dividing by a small number (or zero) when few ensemble members exceed the threshold. An inability

https://doi.org/10.5194/egusphere-2026-230
Preprint. Discussion started: 6 February 2026

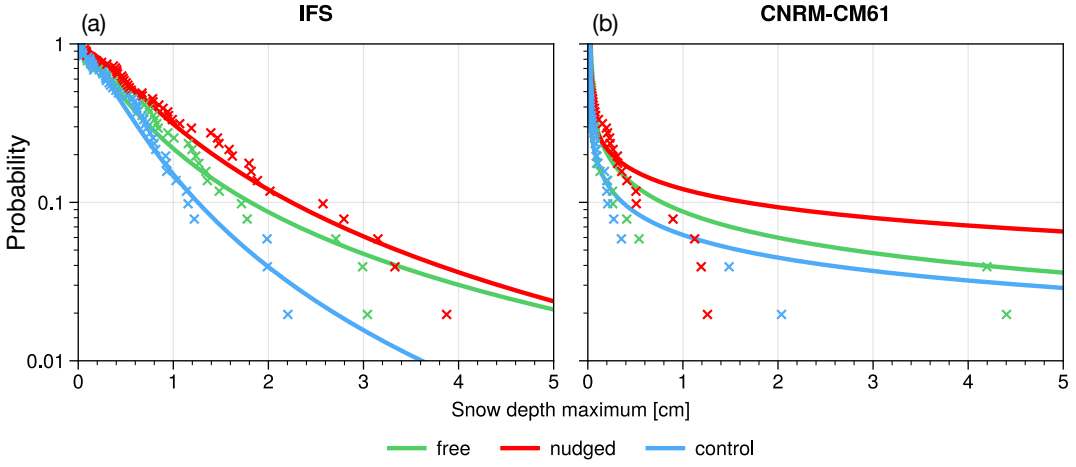

Figure 8. Empirical risk functions, calculated as in Fig. 4(a)(i), but for maximum snow depth over the southern UK and Ireland, for IFS (a) and CNRM-CM6-1 (b). Values are shown for the 2018-01-25 initialization, with the same time window as in Table 1. Crosses indicate risk calculated from individual ensemble members, solid lines show GEV fits.

to quantify RR for thresholds near or beyond the ensemble spread is an inherent drawback of this non-parametric approach. However, Fig. S7 shows similar results (as a multi-model mean) for thresholds from 0.2 cm to 1.8 cm, with qualitatively similar results across this range

Although there is significant spread between models, some consistent results emerge. First, in all models, improving the stratospheric forecast by nudging to reanalysis early (when the SSW is poorly predicted; Fig 1(a)) increases the risk of snow (by on average about a factor of 2), while the sign of the effect of nudging late is mixed among models. Second, nudging to climatology early has a small and inconsistent sign effect, while nudging to climatology late (when the SSW is well predicted) reduces the risk of snow in all models (by on average about a factor of 0.5). These results are broadly consistent with our findings for temperature (noting the larger geographical region considered for temperature), and indicate that, in terms of inter-model consistency, there is a robust impact of the 2018 SSW on the probability of UK and Ireland snow accumulation.

4 Conclusions and discussion

We have compared ensembles of free-running and nudged (to either an observed event or observed climatology) subseasonal forecasts to understand the role of stratospheric variability in the risk of surface weather extremes. Specifically, we set out to address (1) the attribution of extreme event risk to stratospheric variability, and (2) the extent to which improved or degraded stratospheric forecasts impact extreme event prediction. For both tasks we have found that our results depend on the particular case study considered.

For SSWFeb18 we see that an improved stratospheric prediction at an early initialization increases the predicted cold-spell risk by about a factor of 2 and severity by about 2°C, while degrading the stratospheric forecast at a later initialization

https://doi.org/10.5194/egusphere-2026-230
Preprint. Discussion started: 6 February 2026

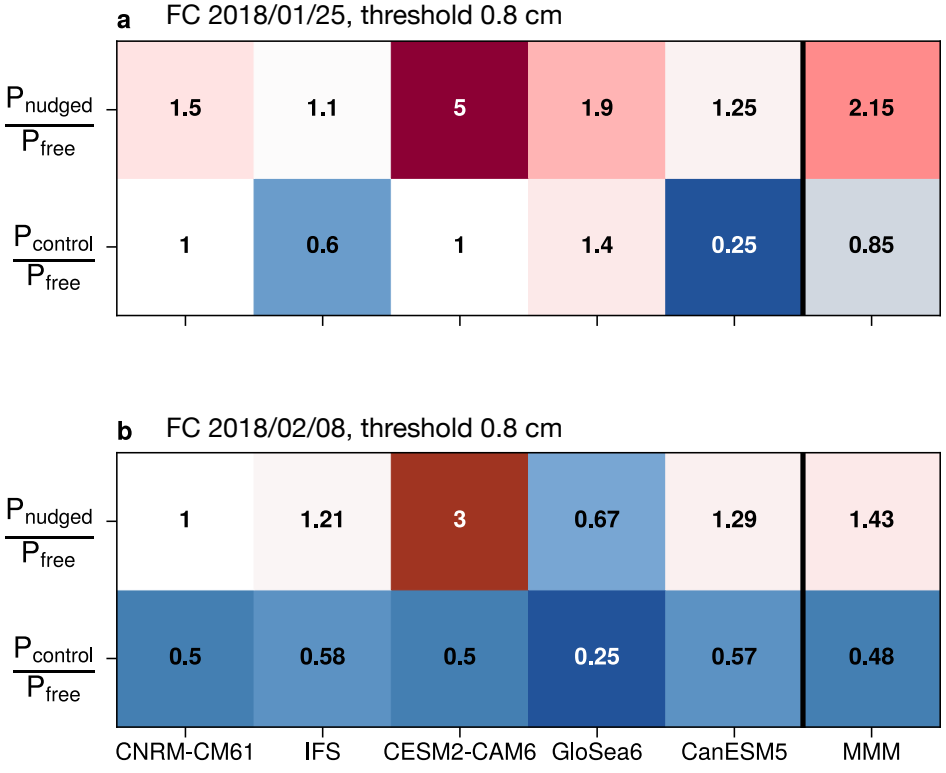

Figure 9. Non-parameteric RR calculated for 5 models (for which appropriate snow depth data was available) for southern UK and Ireland snow depth, with a threshold of 0.8 cm. Values are shown for early (a) and late (b) SSWFeb18 initialization dates, and a multi-model mean, calculated by averaging RR after calculating for individual models, is also shown. The version of GloSea6 shown is UKMO-GloSea6.

decreases the Eurasian cold-spell risk by a similar factor (although there is significant spread among models). Hence we can conclude that the role of the stratosphere in this event remains significant and approximately constant throughout this time period, with a *fraction of attributable risk* given by $\mathrm{FAR} = 1 - 1/\mathrm{RR} = 1 - 1/2 = 0.5$ (e.g. Lott and Stott, 2016). In other words, approximately half of the total risk of the cold spell can be attributed to the stratospheric evolution. In this case, while an improved stratospheric forecast at an early initialization increases the predicted risk, the forecast at the late initialization is sufficiently accurate that further improvement by nudging has little impact. We find similar results for UK and Ireland snow, with a relatively consistent stratospheric influence on event risk throughout. Our RR is smaller than that of Kautz et al. (2020), who used the same region and time period and found a RR for a Eurasian cold spell of about 9 for IFS in their nudged relative to control experiments, higher than we find for any model. A potential explanation for our weaker signal is that Kautz et al. (2020) nudge to persistence rather than climatology in the stratosphere, with forecasts initialized on February 1, which would result in a vortex stronger than the SNAPSI *control* simulation (e.g. Lee et al., 2025). Kautz et al. (2020) also nudge to the full stratospheric evolution, not just the zonal mean, which may further enhance the stratospheric impact. Interestingly, the FAR of

https://doi.org/10.5194/egusphere-2026-230
Preprint. Discussion started: 6 February 2026

0.5 that we find here is close to that found by Spaeth and Birner (2022) for the mean impact of SSWs on Arctic Oscillation extremes.

For SSWJan19 our results are more initialization dependent. At the early initialization, nudging has little impact on the North American cold-spell risk. However, at the late intialization (in this case 6 days after the SSW onset), degrading the stratospheric forecast by nudging to climatology consistently reduces the predicted risk (by about half) and severity (by about 1°C) of the cold spell. It is noteworthy here that, unlike Eurasia, mean temperature anomalies following SSWs are small or positive within our North American region (Domeisen and Butler, 2020), and the Alaskan Ridge weather regime that accompanied the cold spell is not, on average, favored by a weak stratospheric polar vortex (Lee et al., 2019). In an analysis of weather regimes in SNAPSI experiments, Lee et al. (2025) found that the early-initialized *nudged* forecasts for SSWJan19 erroneously favor a negative NAO pattern, more typical of the 'canonical' SSW response (seen in SSWFeb18), with little effect on North American weather regimes. However, for the late-initialized forecasts, the Alaskan Ridge regime is well-predicted, likely developing due to a tropical teleconnection rather than a stratospheric influence (Knight et al., 2021). In this case, *control* experiments see a decreased Alaskan Ridge probability relative to *free* or *nudged* (Lee et al., 2025), a result that is consistent with a strong vortex instead favoring a Pacific Trough regime (Lee et al., 2019). In summary, these results suggest that while the SSWJan19 SSW did not drive the weather regime responsible for the North American cold-spell, it did play a role, once this regime had developed, in its persistence. This is an example of state-dependence in the stratospheric influence on event risk.

For SSWSep19, the severity of the Australian spring 2019 heat wave is higher in *nudged* than *control* in almost all models for both initializations. However the change in RR between experiments is seen here to be a much more variable metric among models than is QS for the early initialization, and for the late initialization RR is near 1 for all models. We suggest that this is due to the greater sensitivity of RR to bias which may be greater for Australia than is the signal caused by nudging for many models. Note, however, that we only infer this potential bias from our results for this single case study (e.g. noting the large difference in ERA5 and IFS empirical risk functions in Fig. S6), and we are not able to fully evaluate model bias. Nonetheless, the more robust effect of nudging increasing heat wave severity (positive QS) is consistent with other studies of the impact of weak vortex events on Australian temperature (Lim et al., 2019) and of this event in particular (Lim et al., 2021; Feng et al., 2025). Specifically, studies have linked weak vortex events to a negative Southern Annular Mode, which in turn is associated with warm and dry, and therefore wildfire-conducive, conditions over eastern Australia. Indeed, in an analysis of SNAPSI experiments Feng et al. (2025) have found an increase in wildfire risk index in *nudged* relative to *control* simulations, a result that is likely related to the increased heat wave severity found here, but that may also have contributions from extreme dry and windy conditions that we do not analyze. Our ability to attribute the Australian heatwave event is limited by the 45-day length of the SNAPSI experiments which, even for the later initialization, do not extend into December, when the most extreme heat wave occurred. Revisiting this event with longer hindcast experiments would be a valuable exercise.

More generally, we have aimed for this paper to illustrate a "process attribution" methodology that can be used to attribute the contribution of aspects of internal climate variability to extreme event risk. In doing so, we have learned three main lessons. First, we have found the S2S models analyzed here to be generally capable of simulating events at the observed severity, which

https://doi.org/10.5194/egusphere-2026-230
Preprint. Discussion started: 6 February 2026

verifies their suitability for studying drivers of the risk of these events, in common with recent climate change attribution studies (Leach et al., 2021, 2024).

Second, we have seen that GEV distributions provide a good fit to temperature extremes, both in their geographic pattern and when spatially averaged. This allows for a parametric approach to be taken, reducing the sensitivity to individual ensemble members. We have analyzed the impact of stratospheric nudging on extremes via two metrics: RR (the fractional change in risk for a given severity) and QS (the absolute change in severity for a given risk). While these two have generally been seen to change together (i.e. an increase in RR being accompanied with a negative QS for cold spells), this is not generally true due to the nonlinear nature of GEV distributions, and for SSWSep19 we saw significant changes in QS without accompanying change in RR. We have also seen that RR, being a ratio, is a much more erratic metric, showing less consistency among models. We therefore propose viewing extreme event changes under the lens of both metrics.

Third, we have seen that our parametric approach may not be suited to all variables. In our analysis of snow depth, a variable with very different characteristics from temperature, we have seen that the GEV distributions do not provide a good fit (Fig. 8). In this case, we illustrate a non-parametric approach, based simply on counting ensemble members exceeding a given threshold. This has the disadvantage of being highly sensitive to individual ensemble members, and being unable to assess risk near or beyond the ensemble spread. It may be possible to adapt the parametric approach to snow and other variables; for instance, by fitting a different distribution, or by removing zero values from the ensemble before fitting.

A significant limitation of our analysis is that we have not been able to carry out bias correction of SNAPSI simulations, due to a lack of historical hindcast simulations in equivalent model versions. This limitation is ameliorated somewhat by the fact that we are analyzing differences or ratios between *nudged, free* and *control* simulations, rather than absolute values; however, since risk functions are non-linear, this does not remove the effect of bias. Indeed, as discussed above, we suggest that larger Australian temperature biases may be responsible for the small relative risk changes seen in SSWSep19. A thorough evaluation of model bias would include not just free-running hindcasts but also nudged and control hindcast experiments.

Some caution must also be exercised in interpreting our results as fully capturing the risk attributable to stratospheric variability. S2S models vary in their strength of stratosphere-troposphere coupling relative to reanalysis (Garfinkel et al., 2025). Indeed, the majority of models under-estimate the correlation between stratospheric polar vortex strength and Eurasian surface temperature (Garfinkel et al., 2026), which might tend to lead to a reduction in the level of attribution. Furthermore, the *control* experiment, which involves nudging to the zonal-mean climatological-mean state, is not necessarily a realizable state of the atmosphere, so may not be suitable as a counterfactual that represents a lack of SSW. Finally, although we have identified several robust signals, our results are quite noisy. Large uncertainties are mostly a result of the 50-member ensemble size of SNAPSI simulations, which, being comparable to the number of years of reanalysis, is only just able to capture risks of the size evaluated here. Future work with larger ensemble sizes, a greater range of SSW (and other) events, and including nudging over other regions (e.g. the tropical troposphere) to capture the effects of other remote drivers of risk would be a valuable extension of this work. A further extension, of value to risk analysts, would be to understand and attribute the stratospheric contribution to not just weather extremes, but also their societal impacts. Motivated by this, Rupp et al. (2025) investigate mortality impacts of SSWFeb18 using SNAPSI experiments, and we plan for a following paper to also investigate energy system impacts.

https://doi.org/10.5194/egusphere-2026-230
Preprint. Discussion started: 6 February 2026

Code and data availability. The ERA5 data used in this study is available from https://cds.climate.copernicus.eu/datasets. The SNAPSI forecasts used in this study are archived by CEDA at https://catalogue.ceda.ac.uk/uuid/0a5a1ce22fb047749e040879efa8e9b5. The code used for all extreme-temperature analysis is available in a public Zenodo repository at https://doi.org/10.5281/zenodo.18157368 (justinfocus12, 2026).

Author contributions. In alphabetical order, RM, JF, PR and WJMS carried out the analysis. JF and WJMS prepared the original manuscript draft. WJMS administered the working group. AHB, PH, and CIG administered the SNAPSI project. D-CH, Y-KH, HK, JK, MS, IRS, and S-WS produced the simulations. BA, AHB, JF, PH, E-PL, DDM, GM, RM, CIG and GK, PR and WJMS contributed to development of the methodology. All authors contributed to review and editing of the manuscript.

Competing interests. Some authors are members of the editorial board of Weather and Climate Dynamics.

Acknowledgements. This work used JASMIN, the UK's collaborative data analysis environment (https://www.jasmin.ac.uk). RM was supported by a NERC GW4+ Doctoral Training Partnership studentship from the Natural Environment Research Council (NE/S007504/1). JF was supported by Schmidt Sciences. CIG was supported by the ISF–NSFC joint research program (Israel Science Foundation grant no. 3065/23) and by the NSF–BSF joint research program (United States–Israel Binational Science Foundation grant no. 2021714). E-PL was supported by the National Environmental Science Program (NESP), funded by the Australian Government Department of Climate Change, Energy, the Environment and Water. DDM was supported by a predoctoral grant program from the Comunidad de Madrid (No. PIPF-2023_ECO-30282). S-WS and D-CH were supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2025-02363044).

https://doi.org/10.5194/egusphere-2026-230
Preprint. Discussion started: 6 February 2026

References

Allen, M.: Liability for climate change, Nature, 421, 891–892, https://doi.org/10.1038/421891a, 2003.

Amaya, D. J., Kosaka, Y., Zhou, W., Zhang, Y., Xie, S.-P., and Miller, A. J.: The North Pacific pacemaker effect on historical ENSO and its mechanisms, Journal of Climate, 32, 7643–7661, https://doi.org/10.1175/JCLI-D-19-0040.1, 2019.

Audette, A. and Kushner, P. J.: Simple hybrid sea ice nudging method for improving control over partitioning of sea ice concentration and thickness, Journal of Advances in Modeling Earth Systems, 14, e2022MS003 180, https://doi.org/10.1029/2022MS003180, 2022.

Ayarzagüena, B., Barriopedro, D., Garrido-Perez, J., Abalos, M., De La Cámara, A., García-Herrera, R., Calvo, N., and Ordóñez, C.: Stratospheric connection to the abrupt end of the 2016/2017 Iberian drought, Geophysical Research Letters, 45, 12–639, https://doi.org/10.1029/2018GL079802, 2018.

Ayarzagüena, B., Butler, A. H., Hitchcock, P., Garfinkel, C. I., Lawrence, Z. D., Ning, W., Rupp, P., Wu, Z., Afargan-Gerstman, H., Calvo, N., et al.: The role of the stratospheric state in upward wave flux prior to Sudden Stratospheric Warmings: a SNAPSI analysis, EGUsphere, 2025, 1–40, https://doi.org/10.5194/egusphere-2025-3611, 2025.

Baldwin, M. P., Ayarzagüena, B., Birner, T., Butchart, N., Butler, A. H., Charlton-Perez, A. J., Domeisen, D. I., Garfinkel, C. I., Garny, H., Gerber, E. P., et al.: Sudden stratospheric warmings, Reviews of Geophysics, 59, e2020RG000 708, https://doi.org/10.1029/2020RG000708, 2021.

Bellprat, O. and Doblas-Reyes, F.: Attribution of extreme weather and climate events overestimated by unreliable climate simulations, Geophysical Research Letters, 43, 2158–2164, https://doi.org/10.1002/2015GL067189, 2016.

Butler, A., Charlton-Perez, A., Domeisen, D. I., Garfinkel, C., Gerber, E. P., Hitchcock, P., Karpechko, A. Y., Maycock, A. C., Sigmond, M., Simpson, I., et al.: Sub-seasonal predictability and the stratosphere, Sub-seasonal to seasonal prediction, pp. 223–241, https://doi.org/10.1016/B978-0-12-811714-9.00011-5, 2019.

Butler, A. H., Lawrence, Z. D., Lee, S. H., Lillo, S. P., and Long, C. S.: Differences between the 2018 and 2019 stratospheric polar vortex split events, Quarterly Journal of the Royal Meteorological Society, 146, 3503–3521, https://doi.org/10.1002/qj.3858, 2020.

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

Coles, S.: An introduction to statistical modeling of extreme values, Springer Series in Statistics, Springer, 1 edn., ISBN 978-1-85233-459-8, https://doi.org/10.1007/978-1-4471-3675-0, 2001.

Dai, Y., Hitchcock, P., Butler, A. H., Garfinkel, C. I., and Seviour, W. J.: Assessing stratospheric contributions to subseasonal predictions of precipitation after the 2018 SSW from SNAPSI, Weather and Climate Dynamics, 6, 841–862, https://doi.org/10.5194/wcd-6-841-2025, 2025.

Domeisen, D. I. and Butler, A. H.: Stratospheric drivers of extreme events at the Earth's surface, Communications Earth & Environment, 1, 59, https://doi.org/10.1038/s43247-020-00060-z, 2020.

Faranda, D., Messori, G., Coppola, E., Alberti, T., Vrac, M., Pons, F., Yiou, P., Saint Lu, M., Hisi, A. N., Brockmann, P., et al.: ClimaMeter: contextualizing extreme weather in a changing climate, Weather and Climate Dynamics, 5, 959–983, 2024.

Feng, K., Rao, J., Garfinkel, C. I., Butler, A. H., Jie, W., Wu, T., Hitchcock, P., Lim, E.-P., Dowdy, A. H., and Seviour, W.: Does stratospheric nudging improve surface predictability? Insights from the 2019 Southern Hemisphere sudden stratospheric warming, npj Clim. Atmos. Sci., 8, https://doi.org/10.1038/s41612-025-01234-2, 2025.

https://doi.org/10.5194/egusphere-2026-230
Preprint. Discussion started: 6 February 2026

Galvin, J., Kendon, M., and McCarthy, M.: Snow cover and low temperatures in February and March 2018., Weather (00431656), 74, https://doi.org/10.1002/wea.3469, 2019.

Garfinkel, C. I., Lawrence, Z. D., Butler, A. H., Dunn-Sigouin, E., Erner, I., Karpechko, A. Y., Koren, G., Abalos, M., Ayarzagüena, B., Barriopedro, D., et al.: A process-based evaluation of biases in extratropical stratosphere–troposphere coupling in subseasonal forecast systems, Weather and Climate Dynamics, 6, 171–195, https://doi.org/10.5194/wcd-6-171-2025, 2025.

Garfinkel, C. I., Lawrence, Z. D., Butler, A. H., Dunn-Sigouin, E., Erner, I., Karpechko, A. Y., Koren, G., Abalos, M., Ayarzagüena, B., Barriopedro, D., Calvo, N., de la Cámara, A., Charlton-Perez, A., Cohen, J., Domeisen, D. I., García-Serrano, J., Hindley, N. P., Jucker, M., Kim, H., Lee, R. W., Lee, S. H., Osman, M., Palmeiro, F. M., Polichtchouk, I., Rao, J., Richter, J. H., Schwartz, C., Son, S.-W., Taguchi, M., Tyrrell, N. L., Wright, C. J., and Wu, R. W.-Y.: Impacts of extratropical stratosphere-troposphere coupling biases on subseasonal predictive skill, in prep., 2026.

Gessner, C., Fischer, E. M., Beyerle, U., and Knutti, R.: Very Rare Heat Extremes: Quantifying and Understanding Using Ensemble Reinitialization, Journal of Climate, 34, 6619 – 6634, https://doi.org/10.1175/JCLI-D-20-0916.1, 2021.

Hall, R. J., Mitchell, D. M., Seviour, W. J., and Wright, C. J.: Surface hazards in North-west Europe following sudden stratospheric warming events, Environmental Research Letters, 18, 064 002, https://doi.org/10.1088/1748-9326/acd0c3, 2023.

Hauser, M., Gudmundsson, L., Orth, R., Jézéquel, A., Haustein, K., Vautard, R., Van Oldenborgh, G. J., Wilcox, L., and Seneviratne, S. I.: Methods and model dependency of extreme event attribution: the 2015 European drought, Earth's Future, 5, 1034–1043, https://doi.org/10.1002/2017EF000612, 2017.

Hawkins, E., Compo, G. P., and Sardeshmukh, P. D.: ESD Ideas: Translating historical extreme weather events into a warmer world, Earth System Dynamics, 14, 1081–1084, https://doi.org/10.5194/esd-14-1081-2023, 2023.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hitchcock, P., Butler, A., Charlton-Perez, A., Garfinkel, C. I., Stockdale, T., Anstey, J., Mitchell, D., Domeisen, D. I., Wu, T., Lu, Y., et al.: Stratospheric Nudging And Predictable Surface Impacts (SNAPSI): a protocol for investigating the role of stratospheric polar vortex disturbances in subseasonal to seasonal forecasts, Geoscientific Model Development, 15, 5073–5092, https://doi.org/10.5194/gmd-15-5073-2022, 2022.

Hoerling, M., Kumar, A., Dole, R., Nielsen-Gammon, J. W., Eischeid, J., Perlwitz, J., Quan, X.-W., Zhang, T., Pegion, P., and Chen, M.: Anatomy of an extreme event, Journal of Climate, 26, 2811–2832, https://doi.org/10.1175/JCLI-D-12-00270.1, 2013.

Hosking, J. R. M., Wallis, J. R., and Wood, E. F.: Estimation of the Generalized Extreme-Value Distribution by the Method of Probability-Weighted Moments, Technometrics, 27, 251–261, https://doi.org/10.1080/00401706.1985.10488049, 1985.

Huang, J., Hitchcock, P., Maycock, A. C., McKenna, C. M., and Tian, W.: Northern hemisphere cold air outbreaks are more likely to be severe during weak polar vortex conditions, Communications Earth & Environment, 2, 147, https://doi.org/10.1038/s43247-021-00215-6, 2021.

Huang, W. K., Stein, M. L., McInerney, D. J., Sun, S., and Moyer, E. J.: Estimating changes in temperature extremes from millennial-scale climate simulations using generalized extreme value (GEV) distributions, Advances in Statistical Climatology, Meteorology and Oceanography, 2, 79–103, https://doi.org/10.5194/ascmo-2-79-2016, 2016.

justinfocus12: justinfocus12/snapsi_analysis: Code for submission: "Forecast-based attribution of the role of stratospheric variability in weather extremes", https://doi.org/10.5281/zenodo.18157368, 2026.

https://doi.org/10.5194/egusphere-2026-230
Preprint. Discussion started: 6 February 2026

Kautz, L.-A., Polichtchouk, I., Birner, T., Garny, H., and Pinto, J. G.: Enhanced extended-range predictability of the 2018 late-winter Eurasian cold spell due to the stratosphere, Quarterly Journal of the Royal Meteorological Society, 146, 1040–1055, https://doi.org/https://doi.org/10.1002/qj.3724, 2020.

Kim, H., Butler, A., Garfinkel, C., Hitchcock, P., Hong, D.-C., Rao, J., Lawrence, Z., Anstey, J., Ayarzagüena, B., Baldwin, M., Charlton-Perez, A., Erner, I., Henderson, S., Hu, D., Hyun, Y.-K., Jia, L., Kang, M.-J., Karpechko, A., Kharin, V., Knight, J., Koren, G., Lang, A., Lim, E.-P., Lin, H., Lee, R., Lee, S., Malguzzi, P., Manney, G., Mastrangelo, D., Muncaster, R., Paquette, C., Park, C.-H., Polichtchouk, I., Polvani, L., Richter, J. H., Seviour, W., Sigmond, M., Simpson, I., Son, S.-W., Specq, D., Stockdale, T., Taguchi, M., and Xiang, B.: Quantification of stratospheric influence on surface predictability: An overview of SNAPSI results, in prep, 2026.

Knight, J., Scaife, A., Bett, P. E., Collier, T., Dunstone, N., Gordon, M., Hardiman, S., Hermanson, L., Ineson, S., Kay, G., et al.: Predictability of European Winters 2017/2018 and 2018/2019: Contrasting influences from the Tropics and stratosphere, Atmospheric Science Letters, 22, e1009, https://doi.org/10.1002/asl.1009, 2021.

Krakauer, N. Y.: It Is Normal: The Probability Distribution of Temperature Extremes, Climate, 12, https://doi.org/10.3390/cli12120204, 2024.

Kretschmer, M., Cohen, J., Matthias, V., Runge, J., and Coumou, D.: The different stratospheric influence on cold-extremes in Eurasia and North America, npj Climate and Atmospheric Science, 1, 44, https://doi.org/10.1038/s41612-018-0054-4, 2018.

Lawrence, B. N., Bennett, V. L., Churchill, J., Juckes, M., Kershaw, P., Pascoe, S., Pepler, S., Pritchard, M., and Stephens, A.: Storing and manipulating environmental big data with JASMIN, in: 2013 IEEE international conference on big data, pp. 68–75, IEEE, https://doi.org/10.1109/BigData.2013.6691556, 2013.

Leach, N. J., Weisheimer, A., Allen, M. R., and Palmer, T.: Forecast-based attribution of a winter heatwave within the limit of predictability, Proceedings of the National Academy of Sciences, 118, e2112087 118, https://doi.org/10.1073/pnas.2112087118, 2021.

Leach, N. J., Roberts, C. D., Aengenheyster, M., Heathcote, D., Mitchell, D. M., Thompson, V., Palmer, T., Weisheimer, A., and Allen, M. R.: Heatwave attribution based on reliable operational weather forecasts, Nature Communications, 15, 4530, https://doi.org/10.1038/s41467-024-48280-7, 2024.

Lee, R. W., Charlton-Perez, A. J., and Lee, S. H.: Stratospheric impacts on weather regimes following the 2018 and 2019 sudden stratospheric warmings, Geophysical Research Letters, 52, e2025GL115 668, https://doi.org/10.1029/2025GL115668, 2025.

Lee, S., Furtado, J., and Charlton-Perez, A.: Wintertime North American weather regimes and the Arctic stratospheric polar vortex, Geophysical Research Letters, 46, 14 892–14 900, https://doi.org/10.1029/2019GL085592, 2019.

Lillo, S. P., Cavallo, S. M., Parsons, D. B., and Riedel, C.: The role of a tropopause polar vortex in the generation of the January 2019 extreme Arctic outbreak, Journal of the Atmospheric Sciences, 78, 2801–2821, https://doi.org/10.1175/JAS-D-20-0285.1, 2021.

Lim, E.-P., Hendon, H. H., Boschat, G., Hudson, D., Thompson, D. W., Dowdy, A. J., and Arblaster, J. M.: Australian hot and dry extremes induced by weakenings of the stratospheric polar vortex, Nature Geoscience, 12, 896–901, https://doi.org/10.1038/s41561-019-0456-x, 2019.

Lim, E.-P., Hendon, H. H., Butler, A. H., Thompson, D. W., Lawrence, Z. D., Scaife, A. A., Shepherd, T. G., Polichtchouk, I., Nakamura, H., Kobayashi, C., et al.: The 2019 Southern Hemisphere stratospheric polar vortex weakening and its impacts, Bulletin of the American Meteorological Society, 102, E1150–E1171, https://doi.org/10.1175/BAMS-D-20-0112.1, 2021.

Lott, F. C. and Stott, P. A.: Evaluating simulated fraction of attributable risk using climate observations, Journal of Climate, 29, 4565–4575, https://doi.org/10.1175/JCLI-D-15-0566.1, 2016.

https://doi.org/10.5194/egusphere-2026-230
Preprint. Discussion started: 6 February 2026

Lü, Z., Li, F., Orsolini, Y. J., Gao, Y., and He, S.: Understanding of European cold extremes, sudden stratospheric warming, and Siberian snow accumulation in the winter of 2017/18, Journal of Climate, 33, 527–545, https://doi.org/10.1175/JCLI-D-18-0861.1, 2020.

Martin, Z., Orbe, C., Wang, S., and Sobel, A.: The MJO–QBO relationship in a GCM with stratospheric nudging, Journal of Climate, 34, 4603–4624, https://doi.org/10.1175/JCLI-D-20-0636.1, 2021.

Matthias, V. and Kretschmer, M.: The influence of stratospheric wave reflection on North American cold spells, Monthly Weather Review, 148, 1675–1690, https://doi.org/10.1175/MWR-D-19-0339.1, 2020.

Mitchell, D., Heaviside, C., Vardoulakis, S., Huntingford, C., Masato, G., Guillod, B. P., Frumhoff, P., Bowery, A., Wallom, D., and Allen, M.: Attributing human mortality during extreme heat waves to anthropogenic climate change, Environmental Research Letters, 11, 074 006, https://doi.org/10.1088/1748-9326/11/7/074006, 2016.

Noy, I., Stone, D., and Uher, T.: Extreme events impact attribution: A state of the art, Cell Reports Sustainability, 1, https://doi.org/10.1016/j.crsus.2024.100101, 2024.

Otto, F. E.: Attribution of weather and climate events, Annual Review of Environment and Resources, 42, 627–646, https://doi.org/10.1146/annurev-environ-102016-060847, 2017.

Otto, F. E.: Attribution of Extreme Events to Climate Change, Annual Review of Environment and Resources, 48, 813–828, https://doi.org/10.1146/annurev-environ-112621-083538, 2023.

Otto, F. E., Massey, N., van Oldenborgh, G. J., Jones, R. G., and Allen, M. R.: Reconciling two approaches to attribution of the 2010 Russian heat wave, Geophysical Research Letters, 39, https://doi.org/10.1029/2011GL050422, 2012.

Perkins-Kirkpatrick, S. E., Alexander, L. V., King, A. D., Kew, S. F., Philip, S. Y., Barnes, C., Maraun, D., Stuart-Smith, R. F., Jézéquel, A., Bevacqua, E., et al.: Frontiers in attributing climate extremes and associated impacts, Frontiers in Climate, 6, 1455 023, https://doi.org/10.3389/fclim.2024.1455023, 2024.

Rao, J., Garfinkel, C. I., and White, I. P.: Predicting the downward and surface influence of the February 2018 and January 2019 sudden stratospheric warming events in subseasonal to seasonal (S2S) models, Journal of Geophysical Research: Atmospheres, 125, e2019JD031 919, https://doi.org/10.1029/2019JD031919, 2020.

Rao, J., Garfinkel, C. I., Zhang, X., and Lu, Q.: Subseasonal to seasonal (s2s) prediction of continental cold following the sudden stratospheric warming in the 2022/23 winter, Climate Dynamics, 63, 75, 2025.

Rupp, P., Fernández, L. V., Seviour, W. J., and Birner, T.: Attribution of the impact of the February 2018 sudden stratospheric warming on mortality in the Nordics and United Kingdom, EGUsphere, 2025, https://doi.org/10.5194/egusphere-2025-4587, 2025.

Sánchez-Benítez, A., Goessling, H., Pithan, F., Semmler, T., and Jung, T.: The July 2019 European heat wave in a warmer climate: Storyline scenarios with a coupled model using spectral nudging, Journal of Climate, 35, 2373–2390, https://doi.org/10.1175/JCLI-D-21-0573.1, 2022.

Seneviratne, S., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Di Luca, A., Vicente-Serrano, S., Wehner, M., and Zhou, B.: Weather and Climate Extreme Events in a Changing Climate. In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, https://doi.org/10.1017/9781009157896.013, 2021.

Shepherd, T. G.: A common framework for approaches to extreme event attribution, Current Climate Change Reports, 2, 28–38, https://doi.org/10.1007/s40641-016-0033-y, 2016.

Sheridan, S. C., Lee, C. C., and Allen, M. J.: The mortality response to absolute and relative temperature extremes, International Journal of Environmental Research and Public Health, 16, 1493, https://doi.org/10.3390/ijerph16091493, 2019.

https://doi.org/10.5194/egusphere-2026-230
Preprint. Discussion started: 6 February 2026

Spaeth, J. and Birner, T.: Stratospheric modulation of Arctic Oscillation extremes as represented by extended-range ensemble forecasts, Weather and Climate Dynamics, 3, 883–903, https://doi.org/10.5194/wcd-3-883-2022, 2022.

Stott, P. A., Christidis, N., Otto, F. E., Sun, Y., Vanderlinden, J.-P., Van Oldenborgh, G. J., Vautard, R., von Storch, H., Walton, P., Yiou, P., et al.: Attribution of extreme weather and climate-related events, Wiley Interdisciplinary Reviews: Climate Change, 7, 23–41, 2016.

van Garderen, L., Feser, F., and Shepherd, T. G.: A methodology for attributing the role of climate change in extreme events: a global spectrally nudged storyline, Natural Hazards and Earth System Sciences, 21, 171–186, https://doi.org/10.5194/nhess-21-171-2021, 2021.