# ~~Case-to-Case Variability in~~ Quantifying the ~~Tropospheric Response~~ tropospheric response to ~~Sudden Stratospheric Warmings Revealed~~ individual sudden stratospheric warmings revealed by ~~Ensemble Re-Forecasts~~an ensemble simulation strategy

Sheena Loeffel[1], Philip Rupp[2], Selina Kiefer[3], Joaquim G. Pinto[3], Thomas Birner[2,1], and Hella Garny[1,2]

[1]Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany
[2]Meteorological Institute Munich, Ludwig-Maximilians-University, Munich, Germany
[3]Institute of Meteorology and Climate Research Troposphere Research (IMKTRO), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

**Correspondence:** Sheena Loeffel (sheena.loeffel@dlr.de)

**Abstract.** Stratospheric extreme events during Northern winter and spring have been shown to sometimes enhance the ~~subseasonal~~ sub-seasonal predictability of large-scale tropospheric circulation patterns such as the North Atlantic oscillation (NAO) and ~~Greenland~~Greenland / European blocking. ~~We aim to quantify the highly variable downward influence of sudden stratospheric warming (SSW) events on the troposphere in numerical simulations. With this aim, we~~ However, it remains unclear whether event-to-event differences in the observed tropospheric evolution after individual sudden stratospheric warmings (SSWs) represent a robust difference in the tropospheric response to the events, or whether such differences in tropospheric evolutions are simply caused by tropospheric variability. To make progress on this question, we robustly quantify the tropospheric response with an ensemble simulation strategy in a controlled model environment. We construct a model climatology using the ICON global numerical weather prediction (NWP) model~~consisting of possible realistic stratosphere-troposphere evolutions of the coupled troposphere-stratosphere system~~ , representing a wide range of realistic stratosphere–troposphere evolutions during winter months~~. The resulting simulations demonstrate clear~~ , but under controlled boundary conditions to exclude confounding factors like teleconnections of tropical origin. The simulations reproduce key aspects of observed stratosphere-troposphere coupling, ~~consistent with observational findings from previous studies. Ensemble re-forecasts, centred around~~ providing a consistent framework to assess event-specific tropospheric responses. We produce spin-off ensembles for selected SSW events~~,~~ ~~reveal significant variability in surface responses and robustly show that on average across SSW events, the lower stratosphere serves as a mediator between the upper/mid-stratosphere and the tropospheric flow. We show that the mean~~ ; the corresponding ensemble means help robustly quantify the tropospheric response to these SSWs. We find pronounced and robust event-to-event differences in the tropospheric response to SSWs~~based on composites and ensemble re-forecasts is heavily case dependent and ties in with the strength of the lower stratospheric anomaly. Our results establish an increased likelihood of developing Greenland blocking with an anomalous lower stratospheric evolution. Moreover, we present indications that the height~~ . We further show that the flow anomalies in the lower stratosphere in the second week are well correlated with the surface response 3-7 weeks after the SSW. Moreover, our results indicate that the formation of wave reflection surfaces ~~can be decisive in~~

establishing within the lower stratosphere may prevent the establishment of persistent lower-stratospheric anomaliesand the associated tropospheric response. Overall, we our controlled model simulations show that individual SSW events SSWs may differ significantly in their likelihood to induce a canonical tropospheric response , and this likelihood can be predicted at the onset of the SSWtropospheric response and that this likelihood is mainly determined by the post-SSW flow evolution within the stratosphere. These results may be relevant for sub-seasonal predictability of surface weather, especially given that the stratospheric part of the response to SSWs is highly predictable.

## 1 Introduction

Sudden stratospheric warmings (SSWs) are a prominent feature of the Earth's atmospheric dynamics and are known to have a significant impact on tropospheric weather. These changes in the tropospheric circulation are typically (i.e., on average over many SSWs) characterised by negative anomalies in the North Atlantic Oscillation (NAO) and Arctic Oscillation (AO), resulting in an equatorward shift of the jet stream and storm tracks in the Northern Hemisphere (further details in Domeisen et al., 2020a). Stratospheric and tropospheric dynamics mostly evolve on substantially different timescales, the former typically within weeks, and the latter within days. For this reason, typically within several days. This separation of timescales implies that stratospheric anomalies associated with SSWs can introduce a degree of persistence into the coupled system, such that the correlation between tropospheric and stratospheric variability is particularly important for weather forecasts running longer than 2 weeks (sub-seasonal and longer). Accurately simulating the fundamental principles governing troposphere-stratosphere interactions can enhance the accuracy of weather sub-seasonal forecasts, as well as extend the corresponding time ranges of predictability.

However, the relationship between surfaceanomalies and extreme stratospheric events, though clearly observed on average, can vary significantly from case to case. In fact, only a small fraction of events - ranging from two-thirds to one-fifth - shows a significant surface signal. While first clearly highlighted that the multi-event composite zonal-mean signature of SSWs propagates downwards and (in some cases) leads to Despite the well-documented average influence of SSWs on tropospheric circulation patterns (see e.g., Baldwin and Dunkerton, 2001), attributing a given tropospheric post-SSW evolution to an individual SSW is not straightforward. This is because any stratospherically induced response is small compared to internal tropospheric variability (see e.g., Oehrlein et al., 2021). As a consequence, a given single-event post-SSW anomaly near the surfaceimpacts, they also presented phases where anomalous stratospheric conditions do not correspond with noticeable changes in the troposphere. This is also shown in the more recent works of , and . However, also showed that persistent lower stratospheric anomalies can trigger a consistent response in the tropospheric flow. even if it strongly resembles the canonical SSW response (e.g., obtained from composite means) does not by itself demonstrate a genuine downward coupling response – the troposphere may have fortuitously produced such a canonical response-like signal independently of the stratospheric evolution (more akin to an apparent response), as shown in e.g., Davis et al. (2022). Likewise, the lack of a single-event post-SSW canonical surface anomaly does not by itself demonstrate a lack of downward coupling – in this case, the full tropospheric evolution may have overwhelmed the (small) actual SSW response (see e.g., Domeisen et al., 2020b). In other

words, for individual observed SSWs one cannot distinguish apparent from genuine surface responses to SSWs (or a lack thereof). As we argue below, ensemble simulations offer a powerful tool to attribute anomalous surface signatures to SSWs.

One powerful approach to attribute anomalous post-SSW surface signatures to that SSW for single events is based on ensemble simulations. Specifically, the ensemble mean for sufficiently large ensembles offers a possibility to average out intrinsic tropospheric variability, thereby providing a proxy for the SSW-induced surface response. By effectively isolating the mean signal, this approach may be applied to multiple individual SSWs, which then also allows one to determine whether the magnitude and structure of the responses vary substantially between individual stratospheric events. Such an ensemble-based approach has previously been applied to characterise the response to SSWs (see e.g., Kautz et al., 2020; Hitchcock and Simpson, 2014). A more recent and comprehensive application is provided by the SNAPSI project, within which ensemble re-forecasts were conducted by multiple model centres, initialised around three different observed SSWs (two in the northern, one in the southern hemisphere) to assess their contribution to surface predictability on sub-seasonal timescales (Hitchcock et al., 2022). Using these ensemble re-forecasts, it was demonstrated that averaging across forecast members enables the isolation of a robust tropospheric signal following SSWs (see e.g. Dai et al., 2025; Hitchcock et al., 2022). The sub-seasonal-to-seasonal (S2S) database also offers the possibility of studying the tropospheric response to observed SSWs with ensemble simulations from multiple modelling centres (see e.g., Karpechko et al., 2018; Rao et al., 2020, 2021; Butler et al., 2020). In a more recent study, Nebel et al. (2024) analysed the tropospheric response to 16 observed SSWs in S2S ensemble re-forecasts and found a varying mid-tropospheric response between events.

However, these approaches are constrained by the limited number of observed SSWs for which sufficiently large ensemble simulations are available (in some cases requiring the combination of ensembles from different forecasting models, which introduces model uncertainty). More importantly, during observed episodes it is hard to disentangle the purely stratospheric influence from confounding factors arising from other teleconnections such as El Niño and the Southern Oscillation (ENSO), the state of the Quasi-biennial Oscillation (QBO) or Madden-Julian Oscillation (MJO) (see e.g. Ma et al., 2024; Elsbury et al., 2024; Yadav et al., 2024; Cagnazzo and Manzini, 2009). Therefore, it is exceedingly hard to fully disentangle the extent to which differences in the ensemble-mean response to individual observed SSWs arise from intrinsic event-to-event variability versus confounding modulating factors.

~~Despite the well-documented probabilistic surface impacts of these events, the precise pathways through which the stratospheric signal is transferred to (and possibly amplified within)the troposphere remain poorly understood.~~ Several theoretical pathways ~~nonetheless~~ have been proposed to explain how SSWs can ~~affect surface weather—~~ influence surface weather, including remote effects of stratospheric wave driving, planetary wave absorption and reflection, and direct effects on baroclinicity and eddies~~.~~ ~~Consequently, a wide range of interaction patterns relating to the downward propagation of stratospheric anomalies into the troposphere have been documented .~~ ; but the exact mechanistic pathway of downward coupling remains elusive (Baldwin et al., 2021). A robust result across multiple studies has been that SSWs which penetrate down to lower stratospheric levels on average persist longer and show stronger tropospheric flow anomalies (e.g., Hitchcock et al., 2013; Gerber et al., 2009; Runde et al., 2016; Karpechko et al., 2017). However, most of these studies have been based on composite averages over many events, rather than ensemble simulations of individual events. Thus, it is not evident whether the lower stratosphere-to-troposphere

relation stems from event-to-event differences (i.e., certain SSWs develop lower stratospheric anomalies, while others do not), or whether it is a result of internal post-SSW atmospheric variability (i.e., varies across different realizations of post-SSW evolution for a specific event). Again, this question can only be clarified with ensemble simulations.

Focussing on individual events, analysed the strong polar vortex winter event in 2019/2020 and highlighted the importance of the lower stratosphere in facilitating the co-evolution of tropospheric and stratospheric extremes. Moreover, Overall, the authors stressed the need of large-ensemble simulations to enable a full characterisation of the coupled extremes in the polar vortex and tropospheric jet seen in early 2020. Using forecast ensembles, found that the observed SSW in February 2018 and its surface weather impact, e.g. through local blocking phenomena , increased the likelihood of colder surface weather. Notably, they were able to partially attribute the occurrence of the 2018 cold spell to the SSW event.

The inherent variability within the troposphere complicates attribution of circulation anomalies to the stratosphere, especially for individual events. For example, some stratospheric events may be followed by the 'expected' tropospheric response by chance, without actual downward coupling. Likewise, some stratospheric eventsmay involve downward coupling without showing a tropospheric response, because the stratospherically-induced response is overwhelmed by other tropospheric anomalies. Therefore, probabilistic analyses presence of substantial intrinsic tropospheric variability, combined with the presence of confounding factors in observed events, motivates the statistical analysis of SSWs using large-ensemble simulations , as used in the studies of , , and , are a necessary method to quantify the stratospherically-induced anomalous circulation in the troposphere. For this reason, this study employs the in a controlled model environment with prescribed boundary conditions. Such controlled ensemble simulations provide an effective framework for isolating ensemble-mean tropospheric signals following individual SSWs, and for investigating their relation to the (lower) stratospheric evolution of flow anomalies.

Motivated by these considerations, we employ the state-of-the-art ICON global numerical weather prediction (NWP) model to generate a comprehensive set of potential wintertime evolutions of the coupled troposphere-stratosphere system and ensemble re-forecasts spin-off ensembles based on selected SSW eventsSSWs, as described in Sec. 2. We present our analyses of the SSW composite mean downward coupling in Sec. 3. We use the ensemble re-forecasts spin-off ensembles to investigate the variability in downward coupling for separateSSW events , individual SSWs (Sec. 4) and its influence on the tropospheric circulation (Sec. 5). With our dedicated simulation setup, we aim to clarify whether individual SSW events consistently alter In particular, we examine whether individual events consistently shift the probability distribution of subsequent tropospheric large-scale circulation patterns, or whether this impact varies significantly substantially from one event to another. Additionally, we evaluate the extent to which SSWs modify the regional tropospheric flow (Sec. 6) and discuss the possible underlying factors responsible for differences in these responses (Sec. 7) before concluding our findings in Sec. 9.

## 2 Model and data set

### 2.1 Simulation setup

We use the ICOsahedral Nonhydrostatic (ICON; version 2.5.0) model of the German weather service (DWD) which runs on a triangular grid at a horizontal resolution of roughly 40 km. Vertically the model is discretised into 90 terrain-following hybrid-

height levels, with a model top at about 75 km. The output is provided on a 1°x1° regular grid on 52 pressure levels at 6-hourly temporal resolutions. More details on the dynamical core are given by Zängl et al. (2015).

The model runs are structured into two types: We create an 'Event Generating Ensemble' (EGE) to construct a model climatology, use for statistical analysis, and to generate a range of ~~SSW events~~SSWs. We identify 57 ~~SSW events~~ SSWs occurring in the EGE. In a second step, we perform a number of 'spin-off ensembles'~~: these re-forecast the evolution associated with selected SSWs generated in the EGE~~, each of which constitutes a distinct initial condition experiment initialised on the day of the onset of a specific SSW generated within the EGE, re-forecasting the evolution associated with that event. Our approach, based on the construction of the set of SSWs obtained from the EGE's internal variability and ~~of the ensemble-average evolution following selected SSW events~~on the analysis of event-specific ensemble distributions under identical boundary conditions, allows us to extract the actual stratospheric impact against internal tropospheric variability. The following describes in detail how the EGE and ~~ensemble re-forecasts~~ spin-off ensembles are designed.

The EGE is created as a free-running 120-member ensemble simulation continuously covering a time period of 8 months from October to May. We constructed the EGE as a set of three time-lagged 40-member ensembles initialised with realistic atmospheric conditions taken from Oct. 1st, 2nd and 3rd 2020 and run through to May 31st 2021. Initial conditions were taken from operational ICON analysis products provided by DWD as individual sets of initial conditions for 40 slightly perturbed ensemble members. Since we did not aim to model the specific conditions of the winter period 2020/21 but rather winter periods under average conditions, we used a climatological sea surface temperature (SST) distribution computed based on ~~re-analysis~~ reanalysis data (ERA5, see below), rather than the SST distribution provided in the initial conditions. The SSTs are varied daily following a climatological profile. The model further uses a climatological ozone distribution, provided as part of the ICON setup. Since the boundary conditions (e.g. SSTs and Ozone) of the EGE run are given as climatological fields, the model will lose all skill for long lead times and the specific year of initialisation becomes irrelevant. This occurs after about 1 month lead time (see Section 3, Fig.1). We can therefore interpret each ensemble member as an alternative realisation of the atmospheric evolution in a winter with fixed boundary conditions. Furthermore, the removal of ~~all~~ inter-annual variability between ensemble members in the boundary conditions provides a more robust ~~statistical basis to analyse the~~ basis for analysing case-to-case differences ~~of stratosphere-troposphere coupling after a SSW,~~ between individual SSWs and their associated tropospheric responses than would be ~~given by~~ possible with a (limited) ~~re-analysis dataset. In particular, all differences between SSW events~~ reanalysis dataset. Identical large-scale initial conditions ensure that each winter in the ensemble experiences the same QBO phase (all members drift consistently from a slight westerly into an easterly phase, as measured at 30 hPa - see Fig. S2 in the Supplement). Furthermore, the model does not internally generate an MJO, thus known sources for teleconnections from the tropics are missing or are identical between ensemble members. Consequently, any differences in the surface response between SSWs in our EGE ~~have to result~~ must arise from differences in the internal dynamical evolution ~~and cannot be a result of different~~ of the system, rather than from differences in external or slowly varying forcings outside the mid-latitudes (e.g. ENSO ~~,~~ or the QBO).

A set of ~~ensemble re-forecast~~ spin-off ensemble simulations is then performed for 18 selected ~~SSW events~~ SSWs identified in the individual members of the EGE. These are initialised with perturbed initial conditions of the corresponding EGE member

at the SSW onset ~~(first day of negative zonal mean zonal wind at 10hPa and 60° North)~~ or the day ~~before (~~before– to minimise resource usage we only store the full atmospheric state every other ~~day)~~ day– and run for 60 days. ~~Each re-forecast~~ Here, $d_0$ denotes the onset date of the SSW, defined as the first day on which the zonal-mean zonal wind at 10 hPa and 60 °N becomes negative (see Sec. 2.2 for details).

Each spin-off ensemble consists of 40 members, with initial ensemble perturbations constructed following the random field perturbation approach described by Magnusson et al. (2009). The random field perturbations use the naturally occurring variability patterns within the climatological dataset of our EGE. To perturb a ~~member of a re-forecast ensemble for a specific event occurring in a certain EGE memberat day~~ spin-off ensemble member for a given SSW with onset date $d_0$, identified in a given EGE member, we compute the difference ~~in atmospheric states between~~ between the atmospheric states on two randomly drawn days, $d_1$ and $d_2$. Here, $d_1$ and $d_2$ are drawn from two distinct EGE members other than the member containing the ~~event~~ SSW and have to be within a $\pm 15$ day window around $d_0$ ~~,~~to reduce seasonal signals. The difference in atmospheric states at those days is then scaled by a tuning factor adjusted to provide adequate error growth over the first few weeks of the model run. The method for determining this factor is outlined in Magnusson et al. (2009). Each perturbation pattern then creates two members of the ~~re-forecast~~ spin-off ensemble by adding and subtracting it from the unperturbed initial conditions field, leading to an unchanged ensemble mean. As a result, each spin-off ensemble samples the range of plausible tropospheric evolutions conditional on the same stratospheric event.

This approach of generating ~~ensemble re-forecasts~~ spin-off ensembles not only guarantees that each ensemble member ~~of a re-forecast~~ captures the chosen SSW ~~event~~ identically, but also provides the basis for an improved statistical characterisation of the possible tropospheric evolutions following a stratospheric event. Crucially, this setup enables a direct comparison of conditional tropospheric response distributions across different SSWs. The key advantage of this ~~ensemble re-forecast~~ spin-off ensemble setup can be distilled to the following: While the constructed EGE setup rules out external confounding factors, we cannot deduce from the EGE alone whether the case-to-case variability in the tropospheric ~~response (and lower stratospheric signal)~~ evolution following SSWs is driven by the tropospheric variability, or by how strongly an SSW ~~event~~ couples downward. This is where the ~~re-forecast~~ spin-off simulations play a pivotal role, as ~~the quantification of the distribution of the response is facilitated~~ they allow for the quantification and inter-comparison of event-specific response distributions through the ensemble ~~setup~~framework.

We use the ERA5 ~~re-analysis dataset~~ reanalysis dataset (Hersbach et al., 2020) from the European Centre for Medium range Weather Forecasts (ECMWF) as ~~the~~ an observational climatological reference ~~when presenting~~ for evaluating our ICON-generated set of possible wintertime evolutions of the coupled troposphere-stratosphere system (i.e., the EGE) in Section 3. The climatology is calculated as the average spanning the years 1980–2019. The ERA5 dataset was obtained as output of the ~~re-analysis~~ reanalysis on a $1° \times 1°$ regular horizontal grid following pressure surfaces and has a 6-hourly temporal resolution.

## 2.2 Definition of metrics

We identify SSWs as the reversal of the ~~zonal mean zonal~~ zonal-mean zonal wind at 10 hPa and 60°N ($U_{60}^{10}$) and define the SSW ~~central date~~ onset as the first day the reversal to an easterly wind occurs, following the criteria introduced in Charlton

and Polvani (2007). Although all events are conditioned on SSW onset and occur within the winter season, weak residual seasonality in radiative vortex recovery may still modulate post-onset zonal winds. However, SSW identification based on absolute wind reversal ($U_{60}^{10} < 0$) constitutes a more stringent criterion earlier in winter. We use the $U_{60}^{10}$ index to encapsulate the strength of ~~SSW events~~ an SSW and require the ~~zonal mean zonal~~ zonal-mean zonal wind between two events to recover and remain westerly for 20 consecutive days or more for these to be considered separate SSWs (e.g. Butler et al., 2017). Final warming events (Butler et al., 2015) are excluded from our analyses.

We use standardized geopotential height (GPH) anomalies averaged over the polar cap (60-90°° N) as a measure of the large-scale zonal circulation in the stratosphere and troposphere when investigating possible stratosphere-troposphere coupling~~, capturing~~ . This metric captures the progression and evolution of ~~(downward propagating ) signals ,~~ downward propagating signals both within the stratosphere ~~as well as~~ and from the stratosphere to ~~surface level. The geopotential height~~ the surface. GPH anomalies are standardised with respect to the 120-member EGE, unless stated otherwise. We define the surface response metric as the 1000 hPa GPH anomaly averaged over weeks 3-7 ~~post central date, and define a~~ following SSW onset, and the lower stratospheric response metric as ~~GPH anomalies~~ the GPH anomaly at 100~~hPa. We refer to a SSW as having~~ hPa. An SSW is classified as exhibiting a lower stratospheric response if the 100 hPa GPH anomaly exceeds $1.5\sigma$ for at least 10 consecutive days ~~,~~ within the first ~~6 weeks after the central warming date.~~ six weeks after SSW onset. These time ranges ~~for the tropospheric surface and lower stratospheric response (week 3-7 and first 6 weeks, respectively)~~ are motivated by the timescales typically ~~observed in~~ associated with the downward propagation of stratospheric ~~mean flow~~ mean-flow anomalies, spanning days to ~~approx. two months .Sensitivity to the choice of time periods in further discussed in Section 5 (see Fig. 9).~~ approximately two months (see e.g., Ding et al., 2023; Scaife et al., 2022; Sigmond et al., 2013), and are further supported by the composite results of this study (e.g., Figs. 3 and 8). We tested the sensitivity to the exact definition of these time ranges, and results are generally robust with respect to variations of a few weeks (not shown). The weeks 3–7 period is therefore chosen to capture the delayed sub-seasonal surface response following SSW onset, while allowing for event-to-event differences in response timing.

To quantify large-scale atmospheric circulation variability we also employ the $\tilde{AO}$ index, which is calculated as the difference in the ~~zonal mean~~ zonal-mean GPH at 1000 hPa between the meridional means over 30–50°° N and 60–90°° N. As detailed in Rupp et al. (2022), this index serves as a proxy for the strength, latitudinal position, and offers the advantage of being computable directly from model output without requiring a long-term climatological baseline for EOF analysis.

To diagnose wave–mean flow interactions in the stratospheric zonal-mean circulation, we employ the Transformed Eulerian Mean (TEM) formulation (Andrews and McIntyre, 1976):

$$\frac{\partial \overline{u}}{\partial t} = f_0 \overline{v}^\star + \frac{1}{\rho_0} \nabla \cdot \mathbf{F} + \overline{\mathbf{X}} \tag{1}$$

The Eliassen-Palm (EP) flux $\mathbf{F}$ is defined in terms of eddy momentum and heat fluxes, and provides a way to portray the origin and impact of waves on the zonal mean flow. In Eq.1, the EP-flux divergence appears as an explicit wave-induced forcing term, quantifying the net momentum transfer from propagating waves to the mean flow. Physically, a convergent EP-flux (positive $\nabla \cdot \mathbf{F}$) corresponds to a westward (decelerating) force on the mean zonal wind due to breaking and dissipation of planetary waves, and is largely responsible for decelerating the stratospheric westerly jet. This formalism elucidates the fact that the

stratospheric momentum balance is fundamentally controlled by wave drag: stratospheric conditions, such as the strength of the polar vortex, are strongly coupled to planetary wave activity in the lower stratosphere, consistent with previous findings (see also, e.g. Birner and Albers, 2017; Dunn-Sigouin and Shaw, 2015).

Blocking detection is done for a latitude band of 60°° N to 75°° N around the Northern Hemisphere using the two dimensional blocking index by Scherrer et al. (2006), resulting in a binary gridpoint-wise blocking occurrence. The index is calculated for every central latitude ($\Phi_0$) between $45°N$ and $75°N$ using the southern and northern GPH gradient at 500 hPa:

$$GHGS = \frac{Z(\Phi_0) - Z(\Phi_s)}{\Phi_0 - \Phi_s} \tag{2}$$

$$GHGN = \frac{Z(\Phi_n) - Z(\Phi_0)}{\Phi_n - \Phi_0} \tag{3}$$

with $\Phi_n = \Phi_0 + 15°N$ and $\Phi_s = \Phi_0 - 15°N$. A given latitude-longitude grid-point is defined as blocked when the following conditions are satisfied: $GHGS > 0$ and $GHGN < -10$ m/°lat. In this paper we show the blocking frequency anomalies (with respect to climatology) averaged over weeks 3-7 after the onset of the SSW~~event~~.

The statistical significance of the difference in means is determined using the one-sided two-sample Student's t-test on a 95% confidence level. We compare the variances of ERA5 and the EGE using a sample-size adjusted Fisher's F-test. To measure the strength of the linear association between metrics presented in this paper we use the standard Pearson correlation coefficient, denoted by $r$.

## 3 Composite mean downward impact of SSWs in the Event-Generating Ensemble

As detailed in Section 2, the 120-member ensemble simulation provides a large set of possible evolutions of the coupled troposphere-stratosphere system during winter months (see Fig. 1). The ICON climatology of stratospheric polar vortices follows that of ERA5 well (Fig. 1), and the number of ~~SSW events~~ SSWs produced by the EGE is similar to that of the ERA5 climatology (Butler et al., 2017): Out of the 120 members, 57 developed SSWs; the majority of which occur in January and February (Fig. 1). The EGE controlled conditions suppress differences between ensemble members in terms of interannual variability (e.g. QBO, MJO, ENSO) despite the independent stratosphere-troposphere evolution. We expect a smaller variability in the EGE dataset. ~~Indeed, the variance is indistinguishable from~~ However, the interannual variability becomes indistinguishable from that of ERA5 after the first month, suggesting an important role of internal variability for the polar vortex. This also ~~suggests~~ indicates that initial condition memory has subsided from November onwards, ~~and one can refer to the~~ such that individual ICON ensemble members can be treated as independent winter evolutions.

The distinct evolution of the stratospheric warming events compared to the modelled climatological progression (Fig. 2a) allows ~~for~~ the potential downward influence of SSWs to be examined: Fig. 2b highlights the significant anomaly in the composite mean tropospheric ~~zonal mean circulation three weeks after the start of the SSW~~ zonal-mean circulation beginning two weeks after SSW onset, consistent with observations and previous work (see e.g., Baldwin and Dunkerton, 2001; Lee et al.,

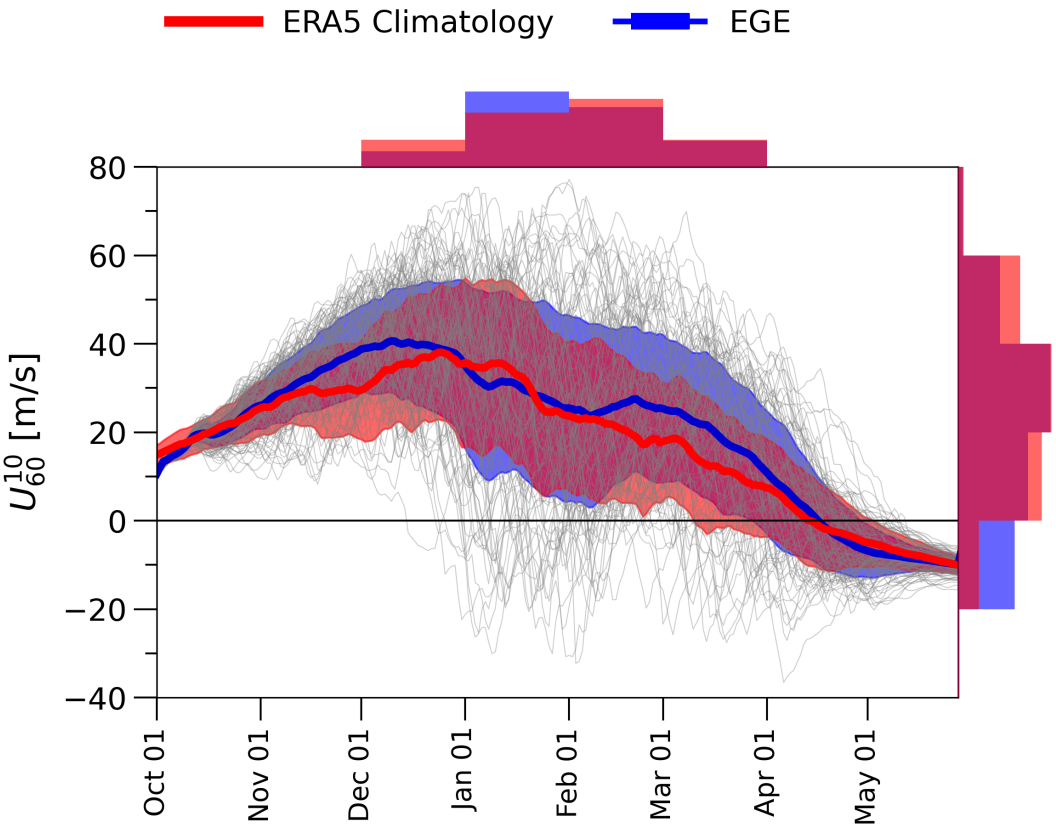


**Figure 1.** Evolution of the zonal-mean zonal wind ~~index~~ $\overline{U}_{60N}^{10hPa}$ at 10 hPa, 60° N in ensemble simulations and ~~re-analysis~~ reanalysis data. ~~Blue:~~ Grey thin lines show the evolution of ~~the EGE for all~~ individual EGE ensemble members~~(thin lines) and~~ , with the EGE ensemble ~~member~~ mean shown in blue (thick line). ~~Red:~~ The ERA5 ~~re-analysis~~ reanalysis climatology for 1980–2019 is shown in red (~~1980-2019,~~ thick line). ~~The shaded~~ Shaded regions indicate ~~the~~ climatological variability ($\pm$1 standard deviation) ~~of~~ for the EGE (blue) and ~~of~~ ERA5 (red). ~~The histograms relate to the EGE simulations (blue) during the~~ Histograms represent winter ~~months only~~ (~~Dec-Mar~~December–March) values only, with EGE shown in blue and ERA5 in red. ~~Top~~ The top panel ~~: Number of~~ shows the SSW ~~events~~ frequency per month, allocated according to the SSW start date~~of the SSW~~. ~~Right~~ The right panel ~~: Distribution~~ shows the distribution of the zonal-mean zonal wind ~~index~~ $\overline{U}_{60N}^{10hPa}$ at 10 hPa, 60° N. All histograms are normalized by the respective dataset size (EGE: 120; ERA5: 40).

2019; Spaeth and Birner, 2022). This tropospheric signal of ~~decreased AO~~ negative $\tilde{AO}$ index anomalies remains statistically significant ~~until five weeks after~~ for up to five to six weeks following the SSW onset. ~~Fig.~~ For the climatological reference shown in Fig. 2, day 0 corresponds to randomly selected winter dates (January–February) within each EGE member, such that the time axis represents lag days relative to a random reference date. Fig. 3 shows the downward progression of the anomalies after ~~weak vortex events~~ SSWs in terms of average polar cap GPH. Selecting only events with a lower stratospheric anomaly highlights that the state of an anomalous lower stratosphere influences the strength of the surface response following an SSW event, consistent with previous work (see e.g. Hitchcock et al. (2013) and others in Section 1). The lower stratospheric anomalies were classified as anomalously strong if the ~~geopotential height (GPH)~~ GPH anomalies at 100 hPa exceeded values of $1.5\sigma$ for at least 10 days within the 6-week period after SSW onset (see Section 2.2). As shown in Fig. 3b, the ~~SSW events~~ SSWs displaying this lower stratospheric signal show a clear increase in the large-scale tropospheric surface anomaly in the weeks following the SSW. Furthermore, the increased surface anomalies differ significantly from the remaining SSW members not satisfying the lower stratospheric threshold.

With the aim of quantifying ~~a possible resulting shift in the probabilities~~ shifts in the probability of anomalous tropospheric ~~zonal mean~~ zonal-mean circulation patterns, ~~the~~ probability density functions (~~PDF) and the distribution of daily data~~ PDFs) and distributions of daily GPH anomalies are compared in Fig.~~4 for the first 6 weeks post~~ 4 for weeks 3-7 following the stratospheric event. Consistent with the findings of Baldwin and Dunkerton (2001), ~~we observe~~ pronounced differences in tropospheric anomalies are evident between the composite means of the SSW group ('SSW'~~, yellow) and the~~ ) and EGE members without ~~stratospheric warming events~~ SSWs ('No SSW'~~, light grey~~). The mean tropospheric GPH anomaly for the SSW ~~cluster lies at~~ group is $0.36 \pm 0.97$ [standardised], ~~whereas the corresponding 'No SSW' groupproduced a mean value of~~ compared to $-0.2 \pm 0.97$ ~~standardised. Comparing~~ for the No-SSW group. Differences in the shapes of the ~~probability density functions between these two groups showcases the increase in~~ PDFs further highlight changes in the likelihood of anomalous ~~mean tropospheric zonal mean circulation regimes: values~~ tropospheric circulation regimes. Values exceeding $+2$ ~~standard-deviations~~ standard deviations at 1000 hPa are at least twice as likely ~~to occur~~ for the 'SSW'-group, ~~in line with the estimation of the fraction of attributable risk shown in the work of~~ consistent with estimates of attributable risk reported by Spaeth and Birner (2022). ~~In particular, values exceeding~~ Conversely, values below $-2$ ~~standard-deviations~~ standard deviations at 1000 hPa are more than twice as likely to occur for the 'No-SSW'-group. ~~Furthermore, the increased tropospheric anomaly of the '~~The enhanced tropospheric anomalies associated with the LS-signal ~~' cluster~~cluster, previously shown in Fig.~~3b is~~ 3b, are also clearly visible in Fig.~~4a. in the form of a distinct right~~ 4a, both as a rightward shift of the PDF (upper panel) ~~as well as~~ and in the box-and-whisker plots (~~bottom~~ lower panel). This shift ~~reaches its peak~~ peaks approximately four weeks after SSW onset (not shown). Surface anomalies exceeding $+2$ ~~standard-deviations~~ standard deviations are more than four times ~~more likely than~~ as likely in the LS-signal group than in the 'No SSW'~~-regimes, while the ensemble members experiencing a SSW but no~~ -regime, while SSWs without a lower stratospheric signal are at least twice ~~more~~ as likely (green). ~~In fact, the latter group (SSW without lower stratospheric anomaly)~~ The latter group behaves similarly to the modelled climatology ~~with respect to~~ in terms of both composite means and ~~tail ends of the respective PDFs. In terms of the lower stratospheric evolution following a SSW event~~the tails of the PDFs. In the lower stratosphere, positive anomalies are ~~more likely to occur in~~

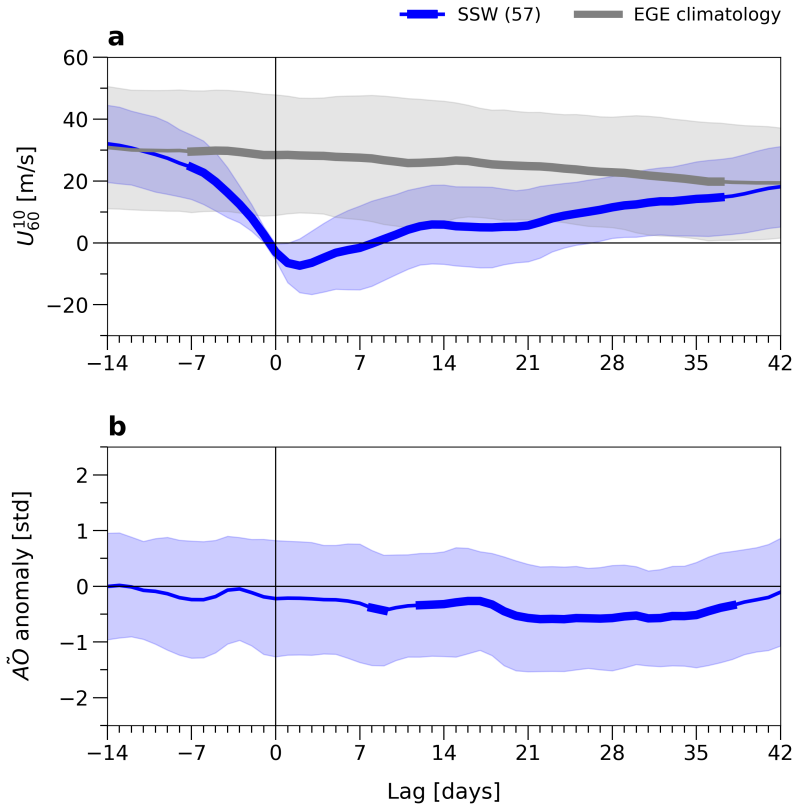


**Figure 2.** Lagged ~~timeseries~~ time series of ~~the~~ (a~~.~~) the zonal-mean zonal ~~mean zonal~~ wind at ~~60°N/~~10 hPa~~and (b.) AO index~~ , 60° N for the EGE climatology (~~thick,~~ grey) and SSW composite (~~thick,~~ blue)~~. The number of EGE members in the SSW composite is shown in brackets~~, and (b) $\tilde{AO}$ anomalies for the ~~individual members are shown with the thin~~ SSW composite (blue~~lines~~). ~~Note that the~~ The climatology is calculated as the EGE ensemble ~~member~~ mean~~of~~ . For the climatological reference, day 0 corresponds to randomly selected winter dates (January–February) within each EGE member, such that the time axis represents lag days relative to a random reference date. ~~The thickened~~ For the SSW composite, day 0 corresponds to the SSW onset and is marked by the thin black vertical line. Thickened line segments indicate ~~statistically significant~~ periods during which the difference in means between the ~~composite~~ SSW composite and climatology ~~(to~~ is statistically significant at the 95 % confidence ~~). Day 0 is taken as the start of the SSW~~level, ~~ie.~~ with shaded envelopes indicating the ~~first day~~ ensemble spread ($\pm 1$ standard deviation) $\tilde{AO}$ anomalies for the ~~zonal mean zonal wind at 60°N/10 hPa becomes negative~~climatology are not shown, ~~and~~ as their climatological mean is ~~marked with the thin black vertical line~~zero. The number of EGE members included in the SSW composite is indicated in brackets

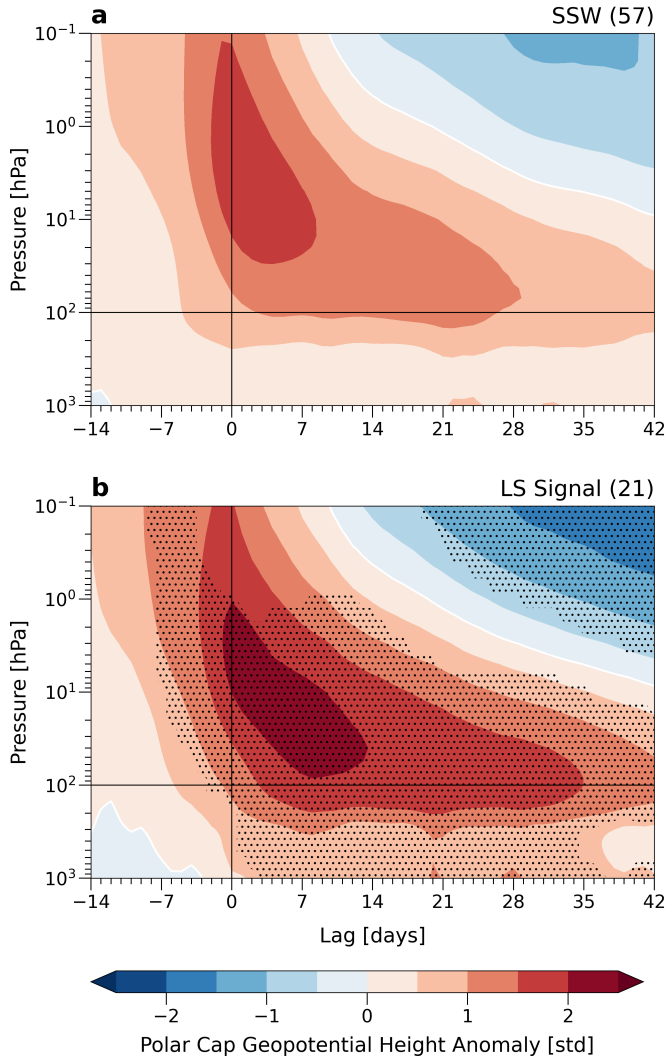


**Figure 3.** Event-based composite of ~~standardised geopotential height~~ GPH anomalies ~~over the polar cap region (60-90°N)~~ of (a~~.~~) all SSWs identified in the EGE and (b~~.~~) ~~only the~~ events with a strong ~~lower stratospheric (~~LS ~~)~~ signal following the SSW. Numbers in the panel headings represent group sizes. Stippling indicates significant differences ~~(to~~ at the 95% confidence ~~)~~ level in cluster means between ~~the~~ events with an LS signal ~~cluster~~ and ~~members~~ those without ~~LS signal~~ (not shown, see Fig. S1). Thin black lines in the vertical and horizontal directions mark the start of the SSW and the 100 hPa pressure level, respectively.

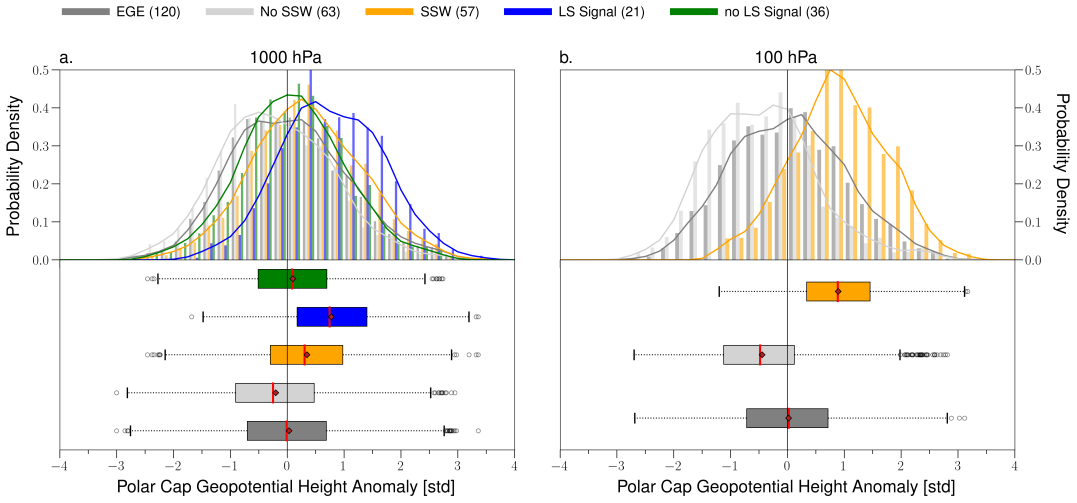


**Figure 4.** Probability density functions (top ~~panel~~panels) and ~~distribution~~ distributions (~~lower panel~~bottom panels) of ~~polar cap geopotential height~~ standardised GPH anomalies for weeks 3-7 ~~post~~ following the event date at (a~~.~~) 1000 hPa and (b~~.~~) 100 hPa from the EGE simulation. Dark grey ~~:~~ shows the climatology, light grey ~~:~~ members without SSWs, and yellow ~~:~~ members with SSWs. ~~Of the~~ Among SSWs, ~~members~~ events with and without a ~~lower stratospheric (LS)~~ lower-stratospheric signal are shown in blue (LS) and green (noLS), respectively. ~~Mean and median shown as red~~ Red diamonds and vertical lines indicate the mean and median, respectively~~. Number of members per group is shown~~ , with sample sizes given in brackets. ~~Note that for~~ Anomalies are defined relative to the EGE climatology (120-member ensemble mean). For the climatology and No-SSW groups, ~~a random day in January or February was selected as~~ the event date corresponds to a randomly selected January–February day. ~~We present~~ Boxplots show the ~~distribution of surface and lower stratospheric anomalies using boxplots, set the~~ interquartile range ~~as the range between the~~ (25th~~and~~ –75th ~~percentile and calculate the minimum and maximum values as~~ percentiles), with whiskers extending to 1.5 times the interquartile range~~below the lower and above the upper quartile, respectively. Any data point exceeding these~~ ; values beyond this range are ~~classed~~ shown as outliers.

~~the lower stratosphere after a SSW event takes place, as can be seen~~ likewise more likely following SSWs, as indicated by the distributional shift shown in Fig. 4b.

Overall, our analyses of the EGE show a downward influence of ~~SSW events~~ SSWs on the tropospheric ~~ensemble-mean flow . This is consistent with previous studies on stratosphere-troposphere coupling, e.g. , some of which classified stratospheric extreme events according to their downward propagation to the troposphere~~flow in the composite mean, in agreement with the well-known signal first described in Baldwin and Dunkerton (2001). Our analyses ~~in particular highlight the~~ further confirm that the EGE ensemble successfully captures the important role of the lower stratosphere for downward coupling, in agreement with previous studies (see Section 1), pointing towards the lower stratosphere acting as a mediator between the upper/mid-stratosphere (SSW~~event~~) and the troposphere. Based on our findings from the composites of SSWs, we utilize the ~~ensemble re-forecasts~~ spin-off ensembles in the next Section to examine whether the development of a lower stratospheric anomaly is

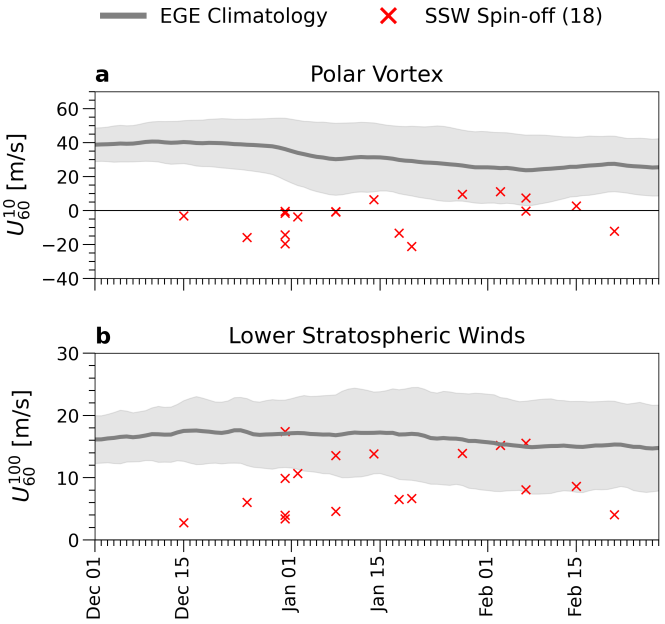


**Figure 5.** Overview of initial stratospheric conditions in terms of (a.) polar vortex strength, defined as the zonal-mean zonal wind at 10 hPa and 60° N, and (b.) lower stratospheric winds~~of~~, defined as the zonal-mean zonal wind at 100 hPa and 60° N, both averaged over the first two weeks after SSW onset, for the spin-off ensemble ~~re-forecast~~ simulations based on ~~SSW events~~ SSWs (red crosses) ~~in~~ during the winter months from December to the end of February. The thick line represents the EGE ensemble ~~member~~ mean, with ±1 standard deviation shown as shading.

more or less likely to develop for individual SSWs, and whether this determines the likelihood for a tropospheric response to develop.

## 4 Selecting SSW events for ~~Ensemble Re-forecasting~~Spin-off Ensembles

The selected ~~re-forecast events~~ SSWs for which we conducted spin-off ensemble simulations, cover the range of possible stratospheric SSW evolutions ~~(also~~ in terms of extremity ~~)~~ and timing throughout the winter months from December to February. For each ~~re-forecast~~spin-off, the initial stratospheric conditions in terms of the $U_{60}^{10}$ index on the day of the selected event are shown in Fig. 5.

In addition to covering a broad range of initial states in the stratosphere, the SSW ~~re-forecast~~ spin-off simulations also capture the range of possible coupled stratosphere-troposphere evolutions (Fig. 6). As expected, the majority lie in the lower right quadrant, as observational data has shown that ~~SSW events~~ SSWs are linked to increased surface GPH anomalies (i.e. negative anomalies on the y-axis and positive anomalies on the x-axis in Fig. 6; see e.g. Kautz et al. (2020)). The upper left and right quadrants capture weak warming events that are followed by negative or positive (i.e. weak to strong) tropospheric signals,

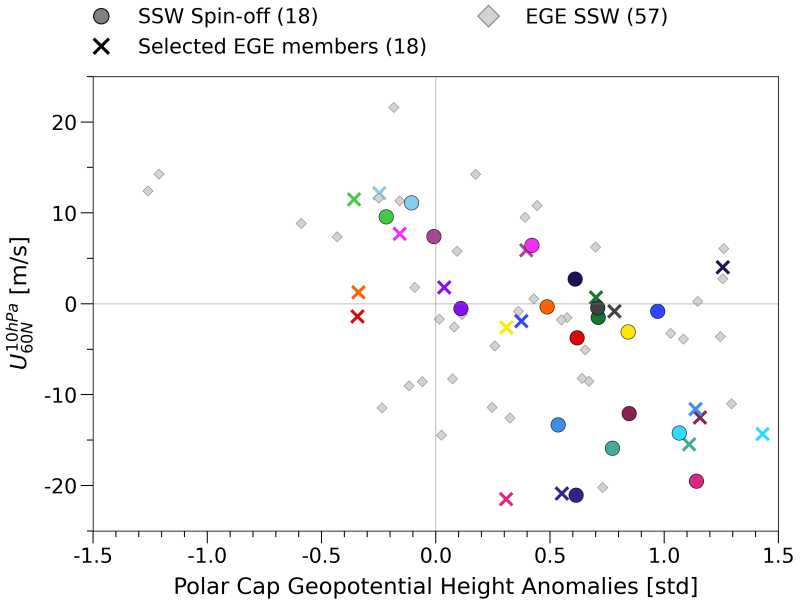


**Figure 6.** Overview of selection of SSW ~~re-forecasts~~ spin-offs (circles, different colours) in terms of the zonal-mean zonal ~~mean zonal~~ wind at 10 hPa, 60°N averaged over the ~~initial~~ first two weeks and the tropospheric surface response index in the weeks after SSW onset. The 57 ~~SSW events~~ SSWs from the event-generating EGE are shown as diamonds. Of these, the EGE members selected for the initialisations of the ~~re-forecasts~~ spin-offs are marked separately with crosses and are shown in the corresponding ~~re-forecast~~ spin-off colour. The number of members per group is shown in the brackets. Note that zonal-mean zonal winds are shown as two-week averages following SSW onset; positive values reflect the beginning of post-SSW recovery in some events.

315 respectively. Figure 6 further contrasts the selected EGE members with the ensemble mean ~~of~~ over the 40 ~~ensemble re-forecasts~~ spin-off ensemble members (coloured crosses and dots). Note that for each selected event, the selected EGE member represents one of the 40 members of the ~~re-forecast and the initial~~ spin-off and the stratospheric state is very similar: The deterministic state lasts up to 10 days and the state is shown here averaged over the first 14 days. We see a greater difference in the values from the selected SSW of the EGE and the ~~re-forecast~~ spin-off ensemble mean in terms of the tropospheric surface response
320 index, as this index captures the surface state several weeks after initialisation once all members have evolved separately, which again emphasises the key advantage of the ~~ensemble re-forecast~~ spin-off ensemble simulation setup.

## 5 Differences in downward coupling in selected SSW events

In Section 3 we showed that our ICON EGE simulations reproduce the known downward influence of ~~SSW events~~ SSWs on the tropospheric flow and established that on average across ~~SSW~~ events, a persistent ~~reduction~~ weakening of the zonal circulation
325 in the lowermost stratosphere following the SSWs increases the likelihood for tropospheric flow anomalies to occur. In this

section we analyse if the forced SSW impact onto the troposphere is strongly related with the 100 hPa stratospheric anomaly, and whether it is case dependent.

Fig. 7 presents the ~~distribution of the polar cap geopotential height anomalies~~ ensemble distribution of standardised GPH anomalies averaged over 3 to 7 weeks after SSW onset at the surface (Fig. 7a.) and lower stratosphere (Fig. 7b.) for all SSW ~~re-forecasts~~spin-offs. The results at 1000 hPa illustrate the varyingly positive surface responses in the ensemble mean for the SSWs, in line with Fig. 6. For this reason, we can conclude that across individual ~~SSW events~~SSWs, some are more likely than others to develop a tropospheric response in the aftermath of the stratospheric event. This is an important result because it suggests that ~~some observed SSW do show the canonical tropospheric~~ the differing tropospheric evolutions observed following SSWs— where some events exhibit the canonical response while others do not~~, which might not entirely be due~~ — may not be attributable solely to internal tropospheric variability (i.e. ~~, by~~ chance). ~~Rather~~Instead, there is predictive ~~power at the day of the SSW on the specific likelihood of this SSW to develop a response~~information available at the time of SSW onset regarding the likelihood of a given SSW developing a tropospheric response. At the same time, the broad distributions of the ensemble response confirm that the tropospheric response to an individual SSW cannot be inferred from a single observed realisation: even for SSWs with a strong ensemble-mean response, some individual members exhibit negative time-averaged tropospheric anomalies.

One might suspect that stronger wind reversal events in the stratosphere have stronger surface impacts, and vice versa, but this is not generally the case as illustrated in Fig.~~8, depicting polar cap geopotential height standardised anomalies~~ 8 for three example events. While the ~~ensemble re-forecast~~ spin-off ensemble simulation ‘SSW 1’ captured both a strong warming event as well as a strong surface response in the ensemble mean (Fig. 8a and d, respectively), the correspondingly weak event in ~~ensemble re-forecast~~ spin-off ‘SSW 3’ (Fig. 8, third column) shows no evidence of downward coupling in the ensemble mean. Another equally weak SSW ~~event~~ in terms of a stratospheric reversal of winds (‘SSW 2’, middle column) develops positive surface anomalies of comparable magnitude with those of the strong event ‘SSW 1’. While one could first assume this difference relates to the strength of the SSW, a clear link between the event strength and surface response is visible only to a limited extent (see also e.g. Fig. 10b). However, the evolution of the lower stratosphere points towards the importance of an anomalous state at 100 hPa for a surface signal to develop (compare ‘SSW2’ to ‘SSW3’). Moving beyond these three examples in Fig. 8 and considering the full set of ~~ensemble re-forecast~~ spin-off ensemble simulations, we indeed observe that the ensemble mean response in the lower stratosphere does differ substantially between the ~~SSW events~~SSWs, cf. Fig. 7b. We corroborate this with Figure 9, in which the ensemble mean GPH anomaly at different times and heights is correlated against the ensemble mean surface response (averaged over week 3-7) across the 18 ~~re-forecasts~~spin-offs. Here, a weaker correlation of the surface response with GPH anomalies at 10 hPa than with those at 100 hPa is found. Most notably, the surface response can be considered highly correlated with the ~~geopotential height~~ GPH anomaly at 100 hPa from week 2 onwards, as the composite correlation coefficient exceeds 0.8 and approaches maximum values. This correlation peak in the lower stratosphere appears before the surface signal develops, and notably is stronger compared to the correlation of the surface signal with anomalies in the upper troposphere (around 400 hPa) .

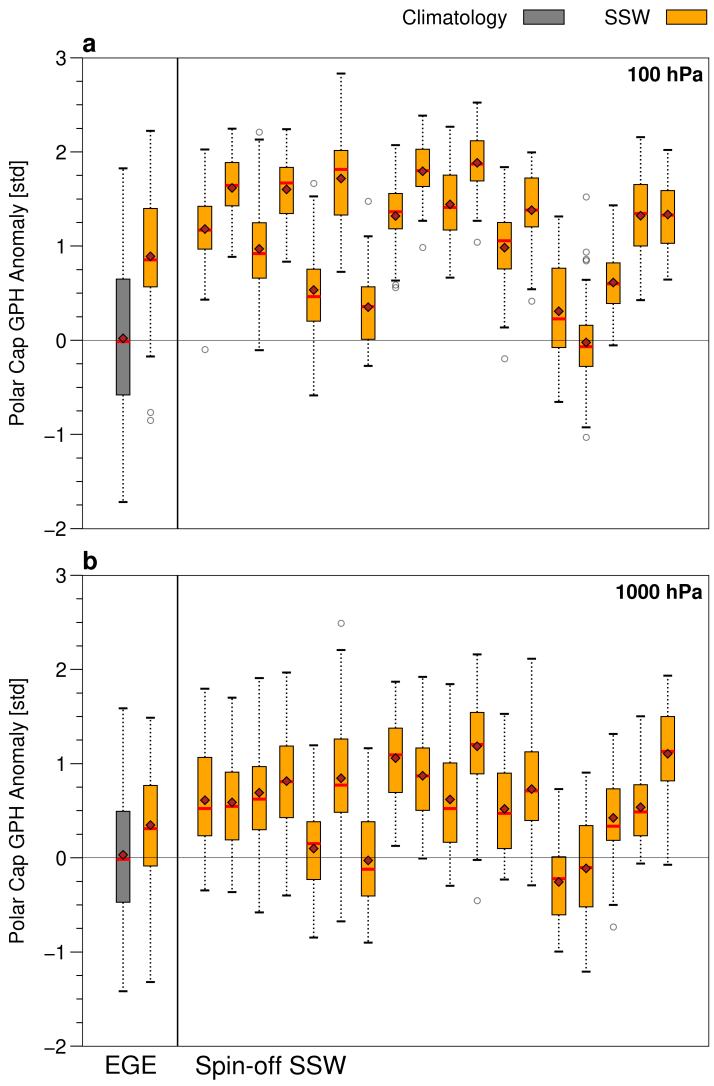


**Figure 7.** Distribution of standardised ~~geopotential height~~ GPH anomalies ~~over the polar cap~~ at (a~~.~~) 100 hPa and ~~1000 hPa~~ (b~~.~~) 1000 hPa, ~~selecting daily values from~~ averaged over weeks 3-7 post-event. ~~Within~~ In each ~~plot: The left~~ panel, the left side shows values for the EGE (as ~~shown~~ in ~~in~~ Fig. 4), while the right ~~panel~~ side shows all ~~re-forecast~~ spin-off simulations. Colours ~~are as in~~ follow Fig. 4: the EGE ensemble mean ~~(ie. EGE climatology)~~ is shown in dark grey~~. Yellow depicts~~ , and yellow denotes the composite mean of the EGE SSWs (left~~panel~~) and the ~~re-forecast~~ spin-off ensemble mean (right~~panel~~). ~~Mean and median shown as red~~ Red diamonds indicate the mean and vertical lines ~~, respectively~~the median. ~~We present~~ Boxplots show the ~~distribution of surface and lower stratospheric anomalies using boxplots, set the~~ interquartile range ~~as the range between the~~ (25th~~and~~ –75th ~~percentile and calculate the minimum and maximum values as~~ percentiles); whiskers extend to 1.5 times the interquartile range~~below the lower and above the upper quartile~~, ~~respectively. Any data point exceeding these values are classed~~ with points beyond classified as outliers.

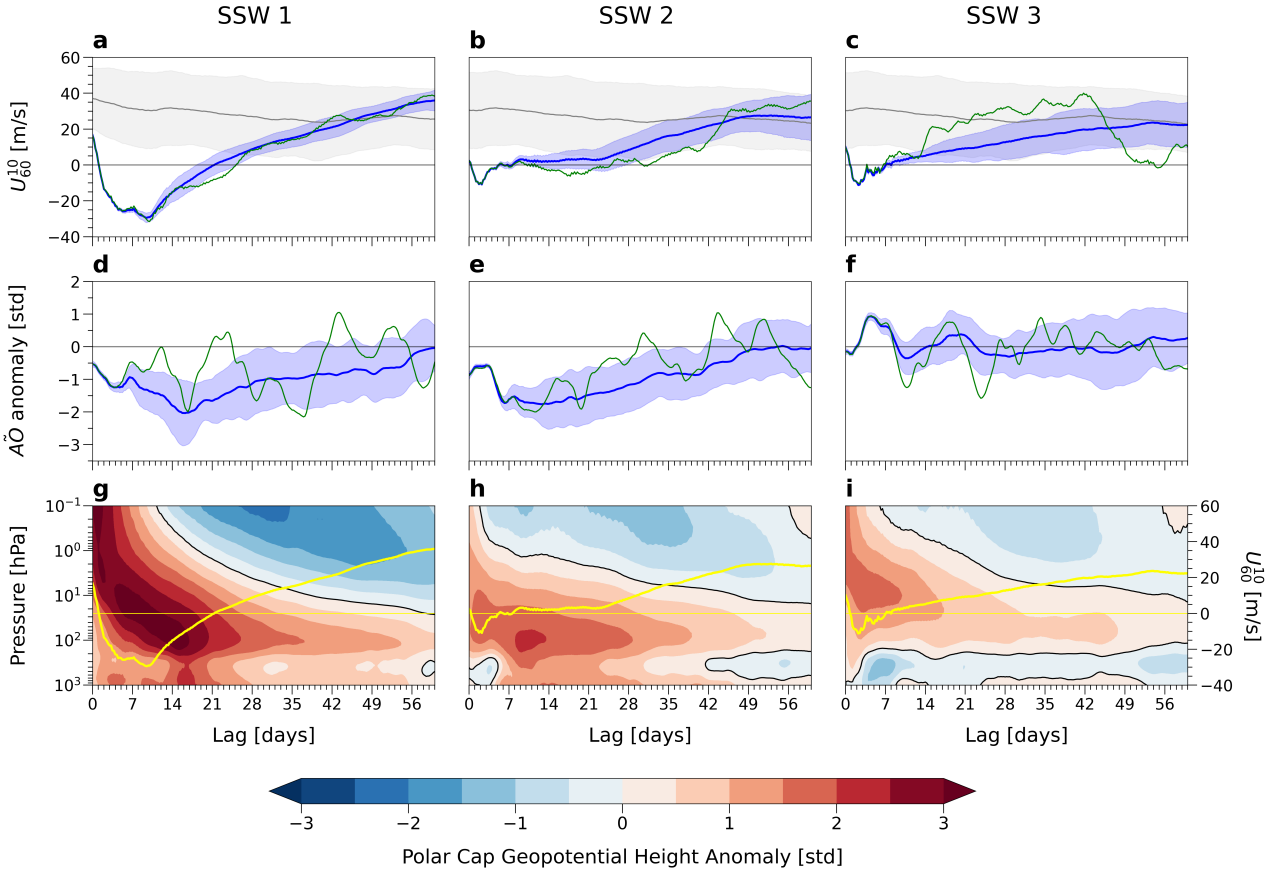


**Figure 8.** Time series of the zonal-mean zonal ~~mean zonal~~ wind at 60°N ~~,~~ and 10 hPa (~~a-c~~a–c)~~and ,~~ AO ~~index~~ anomalies (~~d-f~~d–f)~~of ,~~ and standardised GPH anomalies (g–i) for selected ~~ensemble re-forecast~~ spin-off simulations~~, shown as columns~~. ~~a-f: re-forecasts~~ Each column corresponds to a distinct SSW. In panels a–f, spin-off simulations (blue) are based on selected ~~SSW events~~ SSWs from the EGE (~~control,~~ green). The EGE ~~ensemble~~ climatology (120-ensemble mean) and standard deviation are shown ~~as~~ by the grey line and shading, respectively. ~~g-i: standardised geopotential height anomalies over~~ In panels g–i, the ~~polar cap region (60-90°N). The~~ ensemble ~~member~~ mean ~~of the zonal mean zonal wind~~ zonal-mean zonal-wind at 60°N ~~,~~ and 10 hPa ~~has been added~~ is overlaid as a yellow ~~lines~~line. Day 0 ~~is taken as~~ denotes the ~~start~~ onset of the SSW.

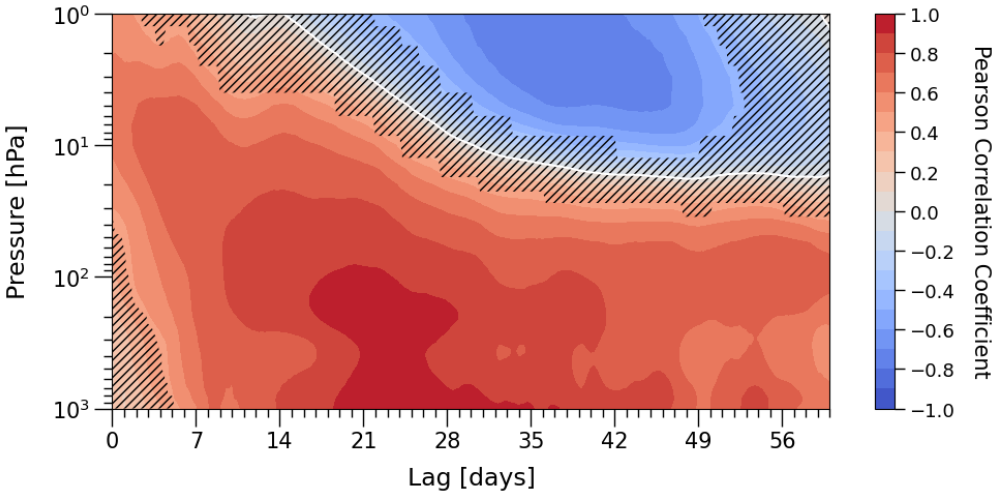


**Figure 9.** 'Dripping paint' plot ~~showing~~ of the spin-off ensemble ~~re-forecast~~ composite Pearson correlation ~~coefficient, calculated from the respective~~ between GPH anomalies ~~correlated with~~ and the week-3–7 averaged 1000 hPa GPH anomalies~~averaged over week 3-7~~. ~~All GPH anomalies calculated over the polar cap region.~~ Stippling ~~indicates regions of statistical insignificance with p-values exceeding 0.05~~ denotes non-significant correlations (~~to 95% confidence~~$p > 0.05$).

Previously we established that the occurrence of a 100 hPa anomaly plays a key role in whether we see downward coupling across the SSW-composite mean (see Fig. 4). Fig. 10 now summarizes and underlines the relationship between the state of the lower stratosphere and the strength of the tropospheric surface signal across events. The lower stratospheric ~~geopotential height anomalies~~ GPH anomalies in the second week after SSW onset display a particularly strong relationship with the surface anomalies averaged over week 3 to 7, with the Pearson correlation coefficient at 0.85 (Fig. 10). In comparison, the relationship between the surface anomalies and initial vortex strength bears a negative correlation coefficient of 0.71, as shown in Fig. 10b. Note that the initial 2-week mean zonal-mean zonal wind is used here to characterise the absolute post-SSW vortex strength for each event, rather than a departure from a fixed seasonal climatology. Consequently, the surface response is more strongly correlated with the lower stratospheric state post-event than with the strength of the initial event itself. ~~Our findings shown in Fig. 10 combined with Fig.9 allow us to further specify that the 100 hPa anomaly characteristic of a given SSW two weeks after the initial event strongly correlates with the surface outcome.~~ This correlation with lower stratospheric anomalies is strong for an extended period post-SSW (see Fig. 9), but becomes strong already in week two after event (Fig. 10). The anomaly in the lower stratosphere in week two is distinct between individual events, and can be predicted quasi-deterministically at the SSW onset (small spread across ensemble, see Fig. 10 and Fig. 7a). Thus, overall we can conclude that individual SSWs are distinct and predictable in their development of a lower stratosphere anomaly, and that this anomaly is closely correlated with the likelihood to develop a zonal-mean tropospheric response thereafter.

Given that the strength of the lower stratospheric anomaly is of critical importance, but not uniquely linked to the SSW strength itself, the question arises which dynamical mechanisms lead to (or prevent) the development of the lower stratospheric

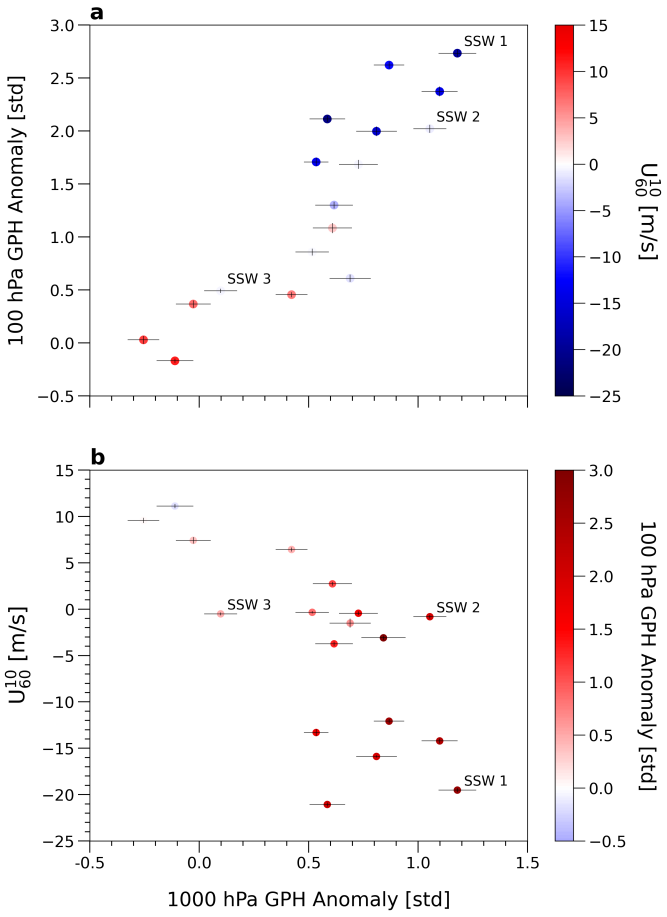


**Figure 10.** Scatter plot showing the correlation of 1000 hPa GPH anomalies ~~over the polar cap region~~ with (a~~.~~) week-2 averaged 100 hPa GPH anomalies ~~over the polar cap region~~ and with (b~~.~~) initial ~~2-week average~~ two weeks averaged polar vortex strength for the ensemble means of all 18 ~~ensemble re-forecasts~~spin-off ensembles, $U_{60}^{10}$. This two-weeks averaged $U_{60}^{10}$ is evaluated relative to the onset of each SSW and is used here as a measure of the absolute post-SSW vortex strength, rather than as a seasonal anomaly. Each ensemble mean point is shaded according to the (a) initial 2-week average polar vortex strength or (~~a.~~b) ~~or polar cap~~ GPH anomalies at 100 hPa~~(b.)~~, as shown with the colour bars. Error bars depict the standard error of the ensemble mean. Values of the correlation coefficient between: (a~~.~~) ~~between~~ 1000 hPa and 100 hPa GPH anomalies: 0.85. (b~~.~~) 1000 hPa GPH anomalies and ~~initial vortex strength~~$U_{60}^{10}$: -0.71

anomaly. ~~This is examined in Section 7.~~ While we do not intend to fully answer this question, we explore possible mechanisms in Section 7. Before, we turn to the question whether the distinct zonal-mean tropospheric response also transfers to distinct regional impacts.

## 6 Differences in the regional tropospheric impact in selected SSW events

Subsequent to the analysis of the ~~zonal mean~~zonal-mean, regional effects are examined to assess whether the differences in ~~zonal mean~~ zonal-mean downward coupling between events are reflected in variations in regional impact. Both the EGE and ~~ensemble re-forecast~~ spin-off ensemble simulations produce a surface response pattern consistent with the well-documented general surface response to ~~SSW events~~ SSWs (Baldwin et al., 2021). This is evident in Fig. 11b-e, showing close to zonally symmetric, positive mean sea level pressure anomalies (MSLP) over the polar cap in weeks 3-7 after SSW onset. The EGE composite signal (Fig. 11b) compares generally well with the composite constructed based on ERA5 (Fig. 11a). Notably, the composite of ~~SSW events from the ensemble re-forecast~~ SSWs from the spin-off ensemble simulations (Fig. 11c) not only exhibits slightly stronger anomalies over the polar cap but also shows enhanced statistical significance compared to the EGE composite. This key advantage of the ~~ensemble re-forecast~~ spin-off ensemble approach presented in this study becomes evident when comparing overall statistically significant regions: the ~~ensemble re-forecast~~ spin-off composite (18 events) covers a broader area and exhibits stronger signals than the composite derived from an equal number of selected EGE events. The 18 ~~re-forecast SSW events~~ spin-off SSWs are further classified based on the anomalous state of the lower stratosphere two weeks after SSW onset, using the 100 hPa GPH anomaly threshold of $1.5\sigma$ (cf. Fig. 9 and Section 2). The surface signal intensifies for events associated with a stronger lower stratospheric state (Fig. 11d), both in terms of signal strength as well as the size of the region covered.

In both ERA5 (Fig. 12a) and ICON (Fig. 12b-e), atmospheric blocking frequency in weeks 3–7 following SSW onset shows positive anomalies over Greenland and surrounding regions~~. The re-forecast~~ , consistent with previous work (Lu et al., 2021; Lee et al., 2019; Beerli and Grams, 2019). The spin-off simulations capture statistically significant increases in Greenland blocking, particularly around the Davis Strait, Greenland, and Baffin Island, with frequency values reaching 0.05 (approximately 4 days per season). The enhanced statistical significance in the ~~re-forecast~~ spin-off SSW composite (Fig. 12c) further underscores the benefits of this simulation approach. Analogous to Fig. 11d, the blocking signal strengthens when selecting ~~SSW events~~ SSWs by lower stratospheric anomalies: cases with strong anomalies show a more than twofold increase in blocking frequency over Greenland compared to the ~~re-forecast~~ spin-off composite, exceeding 0.1 (about 9 days per season) in Fig. 12d.

Given the statistically significant Greenland blocking signal, we compute the correlation between large-scale ~~zonal mean~~ zonal-mean surface anomalies and regional Greenland blocking frequency, finding a strong correlation of 0.84 (Fig.~~13b~~ 13). The correlation between lower stratospheric (100 hPa) anomalies and blocking frequency is slightly lower but remains distinct (0.71). Notably, ~~ensemble re-forecast~~ spin-off events with weak lower stratospheric anomalies ($< 1.5\sigma$, blue shading) and strong anomalies ($> 1.5\sigma$, red shading) show an almost full separation (see Section 8 for more details). These results highlight the near-linear and strong connection between large-scale zonal mean flow anomalies and regional blocking ~~variability~~occurrence, though causality remains uncertain. Individual ~~EGE events~~ realisations of post-SSW evolutions from the EGE do not produce any correlation ~~(see~~ across the SSWs (see Fig.S7 in Supplement), which again emphasizes that an ensemble is needed to ~~bring out~~ robustly quantify the response.

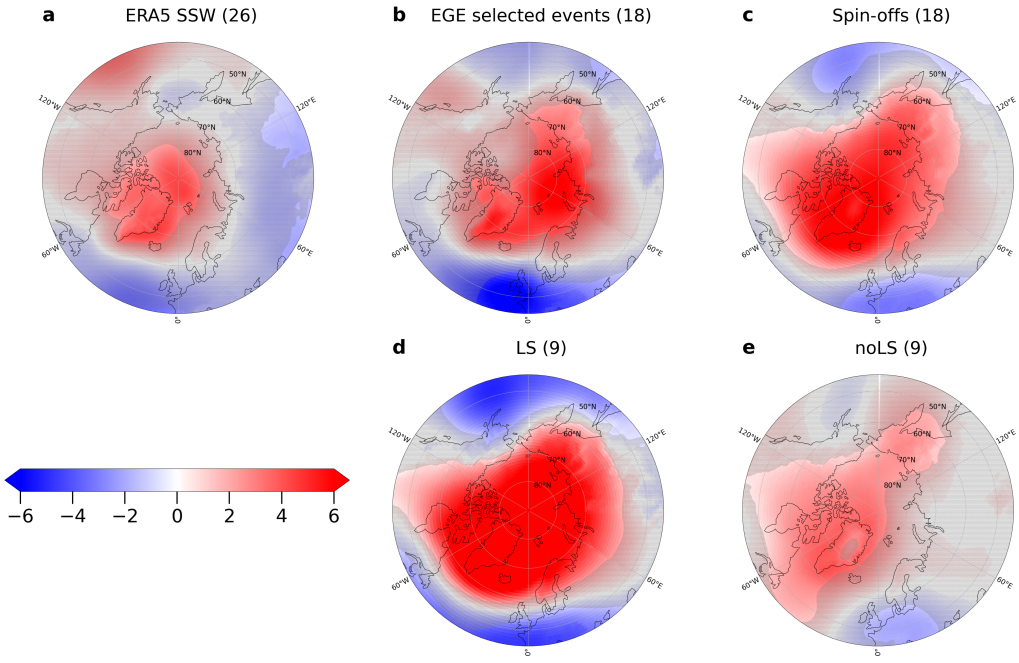


**Figure 11.** Mean sea level pressure anomalies averaged over weeks 3-7 after SSW onset for composites of: (a. ) ERA5 ~~SSW events,~~ SSWs; (b. ) selected EGE ~~events~~ SSWs (18), ~~,~~ ; (c. ~~ensemble re-forecast events~~ ) spin-off SSWs (18), (d. ~~re-forecast~~ ) spin-off SSWs with ~~lower stratospheric~~ a lower-stratospheric anomaly (LS), ~~,~~ ; and (e. ) spin-off SSWs without ~~lower stratospheric~~ lower-stratospheric anomaly development (noLS) ~~development~~ in week 2 after ~~event~~ onset. Stippling indicates regions not ~~of statistical significance (to~~ statistically significant at the 95% confidence ~~)~~level. Maps are shown in a North Polar Stereographic projection ~~of latitudes~~ north of 45°N, centred ~~on~~ at 90°N, with 10° graticule spacing.

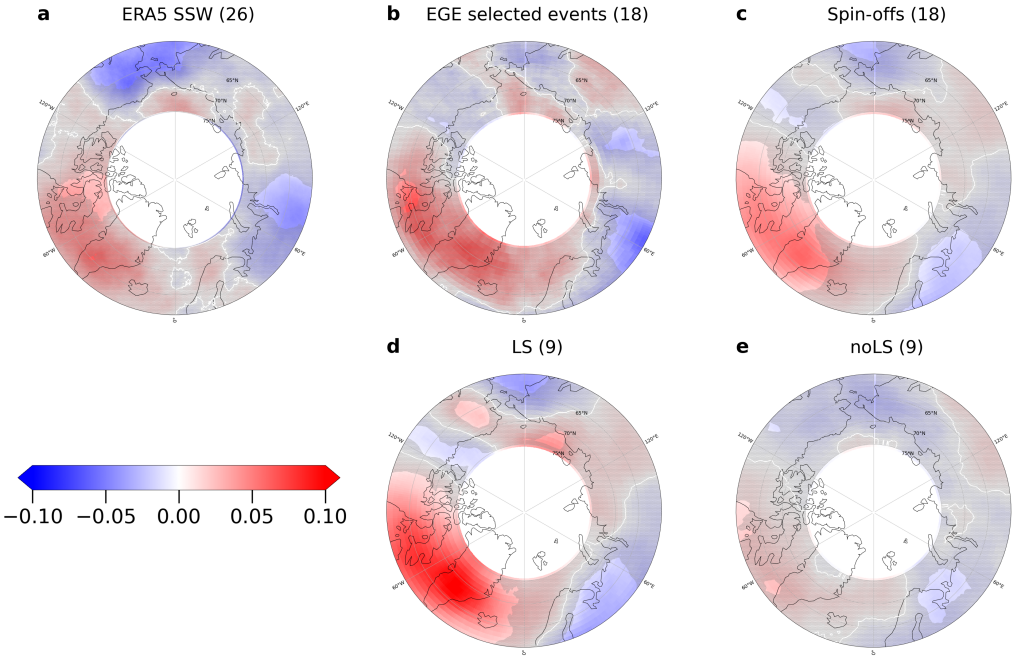


**Figure 12.** ~~Blocking Frequency Anomalies, calculated~~ defined as the ~~average~~ mean number of blocking days ~~in the week 3-7 average~~ averaged over weeks 3–7 after SSW onset. Shown are ~~the anomalies averaged as~~ composite ~~SSW events in~~ anomalies for (a~~.~~) ERA5 ~~re-analysis data~~SSWs, (b~~.~~) selected EGE events, (c~~.~~) spin-off ensemble ~~re-forecast~~ simulations, and ~~ensemble re-forecast simulation~~ spin-off members with ~~and without strong lower stratospheric anomalies post-event~~ (d~~.~~ ; LS) and without (e~~., respectively~~; noLS) strong lower-stratospheric anomalies following the event. Numbers in ~~parenthesis~~ parentheses indicate the number of members ~~per~~ in each composite. ~~Note that the anomaly is~~ Anomalies are calculated ~~as the difference~~ relative to the corresponding ~~climatological~~ climatology for each group (e.g. ~~for a. the anomaly is calculated as the~~ ERA5 ~~climatology subtracted from~~ SSW composites relative to the ~~average number of blocking days per~~ ERA5 ~~SSW event~~climatology). Stippling indicates regions not ~~of statistical significance (to~~ statistically significant at the 95% confidence ~~)~~level.

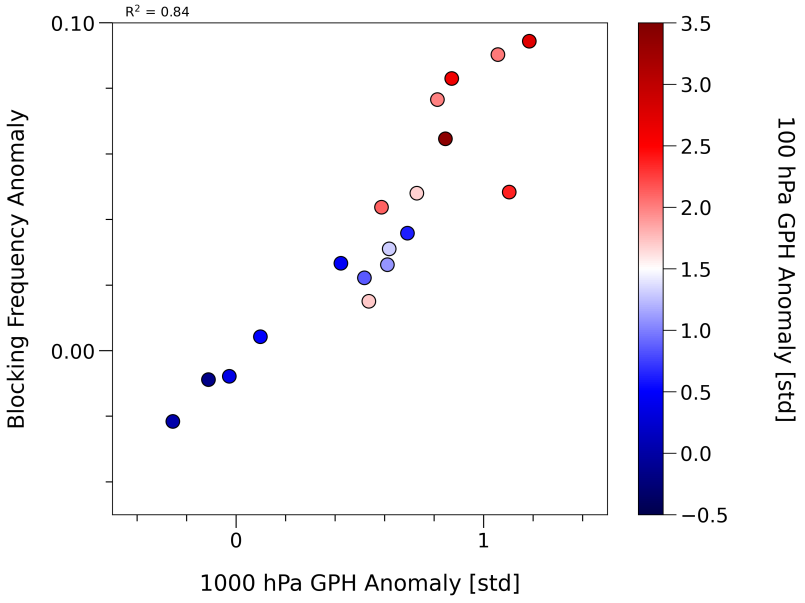


**Figure 13.** Scatter plot showing the correlation between ~~ensemble member mean~~ ensemble-mean 1000 hPa GPH anomalies and the total number of blocking days for the 18 ~~ensemble re-forecast simulations~~ spin-offs over ~~the region~~ 60-75°N ~~,~~ and -90–30°E. ~~Each point is shaded according to~~ Points are coloured by the ~~polar cap GPH anomalies at~~ 100 hPa GPH anomaly in week 2 after SSW onset~~, see colour bar)~~.

## 7 ~~Towards~~ Toward a mechanistic understanding of why not all ~~SSW events~~ SSWs lead to lower stratospheric anomalies

As presented in Section 5, the evolution of an anomalous state in the lower stratosphere is instrumental in whether a surface signal develops~~. We further note that the lower stratospheric zonal mean zonal wind (specifically at 100 hPa) correlates very strongly ($r = -0.91$) with the lower stratospheric geopotential height anomalies at 100 hPa (not shown). Since the stratospheric evolution after a warming event is controlled by the interaction of~~ , but the strength of the lower stratospheric anomaly is not necessarily linearly linked to the strength of the wind reversal at 10 hPa. In the following, we discuss possible mechanisms that might prevent the development of lower stratospheric anomalies on the example of three SSWs. While we have examined diagnostics related to wave reflection across a broader set of events, substantial case-to-case variability and strong temporal dependence make it difficult to define a single objective metric for a systematic classification; we therefore focus here on three representative cases to illustrate possible mechanisms.

For each of these events, we analyse the interaction between planetary waves and the zonal mean flow using the TEM momentum budget, explicitly considering the contributions of wave activity and EP-flux divergence from both WN1 and WN2.

Since the post-SSW stratospheric evolution is governed by interactions between planetary waves and the zonal mean flow, we ~~investigate the momentum budget of the~~ use the transformed Eulerian mean (TEM) ~~equations in order to understand the development of a flow anomaly (or lack thereof) in the lower stratosphere.~~ framework to interpret the evolution of lower stratospheric wind anomalies and to diagnose processes that may favour— or inhibit— the development of lower stratospheric flow anomalies. Consistent with the established understanding of polar vortex dynamics, we find that the lower stratospheric zonal-mean zonal wind is strongly correlated ($r = -0.91$) with GPH anomalies at the same level (not shown), indicating that both diagnostics provide a coherent measure of the evolving vortex state in our framework.

In Section 5, we presented three ~~ensemble re-forecast~~ spin-off simulations: one with both a strong ~~warming event~~ SSW and surface response in the ~~ensemble member mean~~ ensemble-mean (SSW 1~~, Fig.~~; Fig. 8), and two weak ~~SSW events~~ SSWs in terms of wind reversal, ~~whereby~~ of which one showed evidence of downward coupling in the ensemble ~~member~~ mean (SSW 2) and the other did not (SSW 3, Fig. 8). To provide insight in the stratosphere-troposphere evolution following ~~a warming event~~an SSW, key terms ~~for these selected simulations stemming~~ from the TEM momentum budget ~~equation and~~ and corresponding Eliassen-Palm (EP) flux ~~anomalies~~ diagnostics for these simulations are shown in Fig. 14. These ~~terms~~ include the temporal ~~changes in~~ tendency of the background zonal ~~flow (i.e. total wind tendency) and disturbances~~ mean flow and its forcing by large-scale waves, specifically, ~~the contribution of the wavenumber (WN) decomposed~~ quantified by the EP-flux divergence decomposed into wavenumber-1 (WN1~~and~~) and wavenumber-2 (WN2) ~~. We additionally present the~~ contributions. We also show anomalies of the ~~EP-flux~~ WN1 and WN2 EP-flux components. Hatched areas indicate regions where all ensemble members exhibit ~~positive~~ negative EP-fluxes, corresponding to zones of wave reflection. The meridional residual circulation term ~~from~~ in the TEM momentum budget ($f_0\overline{v}^\star$, see Eq. 1 in Sec. 2.2)) is not shown, as it effectively balances the combined ~~contributions from the~~ WN1 and WN2 EP-flux divergence contributions.

Focussing on the ~~zonal mean zonal~~ zonal-mean zonal wind tendencies shown in Fig. 14a-c, we ~~observe~~ find that for the strong event (SSW 1~~(~~; Fig.14a) the initial deceleration at 10~~hPa propagates downwards in typical fashion, and~~ hPa propagates downward in typical manner, with a resulting flow anomaly ~~arises~~ developing at 100~~hPa in~~ hPa within the first two weeks. ~~Regarding~~ For SSW 2 (Fig.14b), the ~~zonal mean zonal~~ zonal-mean zonal wind also decelerates at 100 hPa, albeit less pronounced. In ~~comparison,the initial deceleration in the~~ contrast,for the other weak event (SSW 3~~(~~; Fig.14c)~~with~~ , which exhibits only a weak surface response (cf. Fig.8)~~is interrupted after~~ , the initial deceleration is interrupted a few days ~~following the reversal of winds, and in its place,~~ after SSW onset and is instead replaced by a relatively strong acceleration~~develops~~.

~~Moving on~~ Turning to the interaction of planetary waves and the zonal mean flow (Fig. 14d-o)~~we notice for~~ , SSW 1 ~~the~~ exhibits strong initial upward wave activity flux in WN2 (Fig.14m), which matches the deceleration ~~due to~~ associated with EP-flux divergence in the upper stratosphere~~, whereby the latter~~ . This deceleration propagates downward and reaches the lower stratosphere (100 hPa) by the end of the first week ~~at~~ (lag day 7~~(~~; Fig.14g). ~~We also observe wave~~ Wave reflection in WN1 is initially constrained to the upper stratosphere ~~only throughout the first two weeks, reaching the~~ and only reaches the lower stratosphere (and below) by day 7 (Fig.14j). While the corresponding acceleration in the upper stratosphere (Fig.14d) ~~does roughly match~~ broadly matches the pattern of wave reflection, the strengthening of winds does not ~~reach stratospheric levels~~ extend below the 50 hPa level. ~~Overall~~As a result, the initially decelerated winds in the lower stratosphere ~~only increase~~ recover only slowly, with ~~a weakly positive zonal mean zonal wind tendency for the remaining run time~~ weakly positive zonal-mean wind tendencies persisting at 100~~hPa~~ hPa for the remainder of the integration (Fig.14a). ~~Thus, lower~~ Lower stratospheric anomalies can ~~develop in this case due to the~~ therefore develop due to uninterrupted deceleration by ~~downward progressing~~ downward-propagating planetary wave forcing, ~~making the occurrence~~ with a resulting increased likelihood of strong surface anomalies~~far more likely~~.

The ~~re-forecast~~ spin-off capturing event SSW ~~2~~ 2-- ~~with similarly weak reversal of winds~~ characterised by a weak wind reversal but a strong surface response comparable ~~with the strong~~ to that of SSW 1 ~~event~~ (cf. Fig. 8g)-- ~~presents~~ shows strong initial (and dominant) ~~strong~~ deceleration exerted by ~~the~~ EP-flux divergence in WN1 (Fig.14e), ~~which goes hand in hand with positive anomalies visible~~ accompanied by pronounced positive anomalies in the upward ~~EP-flux~~ WN1 ~~component~~ EP-flux (Fig. 14k). ~~There are also regions~~ Regions of WN1 wave reflection ~~, but limited to the initial week in the upper stratosphere (~~are present but confined to the upper stratosphere during the initial week, corresponding to the ~~reversal of winds) and ensuing weeksshow (moderate )~~ period of wind reversal; during subsequent weeks, moderate anomalous upward wave activity ~~. However, note that the~~ dominates. In contrast, lower stratospheric wave activity in WN2 during the first two weeks ~~in WN2 consists of wave reflection ,~~ is characterised by wave reflection spanning tropospheric and stratospheric ~~(~~levels up to 30~~hPa) pressure levels~~ hPa (Fig.14n). Correspondingly, the acceleration ~~(~~resulting from the ~~wave reflective surface ) seen in~~ reflective surface (Fig.14h~~occurs at levels around the reflection surface at~~ ) is confined to levels around 30~~hPa, and the deceleration of winds~~ hPa, while deceleration at 100~~hPa is predominated~~ hPa remains dominated by WN1 ~~, allowing for the lower stratospheric response to evolve. In other words, overall the~~ forcing. As a result, lower stratospheric winds do not develop a remarkable acceleration despite persistent ~~reflection of~~ WN2 ~~and consequently,~~ wave reflection, with a resulting persistence of WN1-forced anomalies in the lower stratosphere~~forced by WN1 can persist~~.

~~Meanwhile~~In contrast, the weak ~~event~~ SSW 3 ~~shows initial~~ exhibits strongly anomalous upward EP-flux in WN1, exceeding $2\sigma$ during the initial phase (Fig.14l), ~~that corresponds~~ corresponding to the initial deceleration in the upper stratosphere (Fig.14c). ~~The deceleration in WN1 (Fig.14f) propagates down~~ This WN1-induced deceleration propagates downward to the lower stratosphere (100 hPa) during the first week (lag day 7), ~~and~~ but is followed by acceleration at lower stratospheric levels thereafter ~~.~~ (Fig.14f). This pursuing acceleration ~~in WN1~~ can be explained by ~~the areas~~ persistent WN1 wave reflection, as indicated by regions of negative EP-flux ~~as illustrated by the hatchings~~(hatchings) in Fig.14l~~: while initially visible in the troposphere starting mid-week1, planetary~~. While initially confined to the troposphere from mid-week 1 onward, WN1 ~~waves are further reflected in~~ wave reflection extends into the lower stratosphere (particularly between ~~100-50 hPa) in~~ 100 and 50 hPa) during the 2 weeks that follow. As a consequence, the lower stratosphere ~~can neither enter nor continue in~~ neither enters nor maintains an anomalous state (c.f. Fig. 8i)~~and the possibility of developing~~, with a resulting reduced likelihood of a surface response in the ensuing weeks~~becomes much less likely. Note that for~~. For WN2, the first two weeks also show acceleration of ~~the~~ lower stratospheric winds resulting from WN2 EP-flux divergence ~~in WN2~~ (Fig.14i), ~~which goes hand in hand with the observed wave reflection in~~ consistent with the wave reflection occurring during the first week in both the troposphere and lower stratosphere (Fig.14o).

Based on the examples examined here, we hypothesise that the height of the wave reflection surface ~~plays a paramount role in whether an anomalous state~~ can modulate, or even inhibit, the development of anomalous states in the lower stratosphere~~can develop or not. This is indeed the key difference between~~. This key difference is illustrated by the two selected ~~ensemble re-forecast~~ spin-off simulations with comparably weak ~~SSW events: In~~ SSWs: in SSW 3, the wave reflection surface occurs in the lower stratosphere (around 50 hPa and below~~, whereas~~), whereas in SSW 2 ~~develops the reflective surface much higher, namely in the mid to upper~~ it develops much higher in the mid-to-upper stratosphere. The ~~corresponding acceleration of the winds in the lower stratosphere subsequently determines~~ resulting differences in lower stratospheric wind acceleration subsequently determine whether anomalies can develop. ~~For simplicity, we illustrated this based on 3~~ While this mechanism is illustrated here using three exemplary spin-off simulations (out of 18)~~exemplary re-forecasts. We find indications for similar relations between the development of lower stratospheric anomalies with the occurrence /~~, we find indications of similar relationships between lower stratospheric anomaly development and the occurrence and height of wave ~~reflecting surfaces in many of the 18 ensemble re-forecast simulations, yet substantial~~ reflection surfaces in additional cases. However, substantial case-to-case differences in the coupled wave-mean-flow evolution ~~make it difficult to deduce an over-arching mechanism~~ suggest that there may not be a single universal mechanism. We therefore leave it to future studies to further investigate the role and generality of the wave reflection mechanism suggested by these examples.

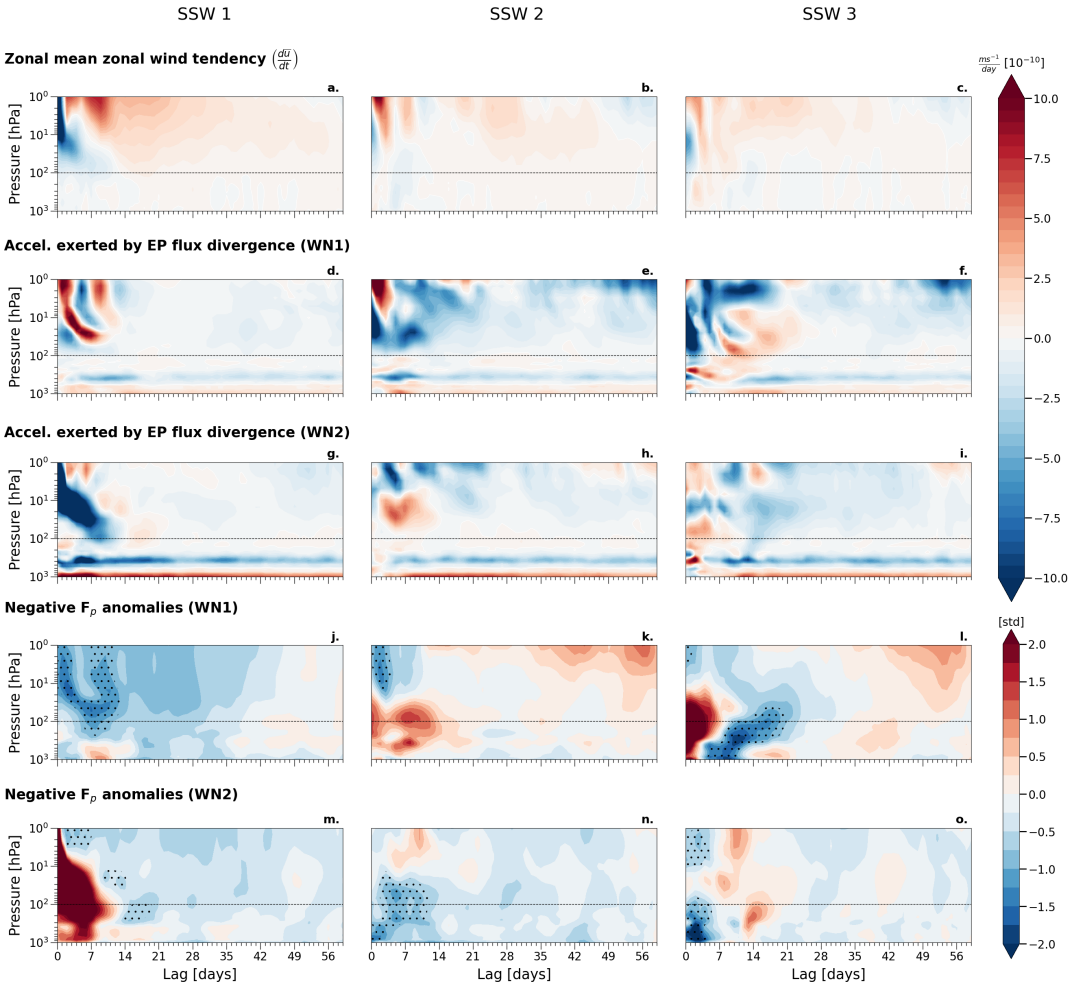


**Figure 14.** Daily means of major terms from the TEM momentum budget ~~of the TEM equations~~ and EP-flux anomalies for the three ~~SSW events~~ SSWs shown in Fig. 8. ~~Rows 1-5:~~ Shown are zonal-mean zonal ~~mean zonal~~ wind tendency (~~a-c~~a–c), acceleration exerted by the EP-flux divergence in WN1 (d-f) and WN2 (g-i), respectively ~~, shown as~~ ($ms^{-1}$ per day), and standardised negative EP-flux anomalies in WN1 (j-l) and WN2 (m-o)~~, respectively~~. ~~Hatchings (j-o) depict~~ Hatching denotes regions of ~~downward propagating~~ downward-propagating waves in the ensemble mean ~~, ie. where the ensemble mean~~ (negative EP-flux~~is positive~~). All ~~metrics shown here~~ fields are ~~calculated~~ averaged over ~~the region~~ 45-75°° N.

## 8 Discussion

The results presented in the previous section indicate that the height of reflective surfaces could play a role in the coupling mechanism by determining whether mean flow anomalies associated with an SSW can progress downward into the lower stratosphere. We further examined both WN1 and WN2 type of events and found no clear distinction in their behaviour in this context. Overall, and in contrast to some recent studies (e.g., Bett et al., 2023), we do not observe a consistent correlation between the type of initial event—whether a split or displacement— and the resulting surface impact (not shown), although it should be noted that the literature on this point is itself inconclusive (see e.g., Maycock and Hitchcock, 2015).

Interestingly, Rupp et al. (2022) found for the early 2020 event that the increase in polar vortex strength following the reflection event influenced the tropospheric flow and thus increased the likelihood of extreme surface values. In other words, the role of the wave reflection event was to modify the stratospheric mean flow in terms of the polar vortex recovery (and subsequent strengthening) after the initial sudden stratospheric deceleration event with subsequent downward coupling, rather than a direct influence of the wave reflection itself on the troposphere. This bears similar tones to our findings, namely that the wave reflection surfaces here play a determining role in the evolution of the (lower) stratospheric mean flow evolution, which in turn leads to a surface response. This is a different role of wave reflection in stratosphere-troposphere coupling compared to studies by e.g. Dunn-Sigouin and Shaw (2020), who discuss a direct downward impact of the wave reflection events themselves.

An inherent aspect of the ~~ensemble re-forecast~~ spin-off ensemble setup used in this study is that the tropospheric initial conditions are only slightly perturbed across the ensemble members, leading to initial condition memory within the troposphere over the first 1 to 2 weeks. Consequently, disentangling the relative contributions of initial tropospheric variability and associated wave pulses to the lower stratospheric evolution outcome remains challenging. However, the absence of a strong correlation ~~between the initial tropospheric conditions and the later surface response (Fig.9) suggests that~~in Fig. 9 suggests that, within this experimental framework, large-scale initial tropospheric circulation ~~patterns may play a less critical role in determining the eventual~~ anomalies, as defined here, are only weakly related to the subsequent surface response following ~~the SSW. It should further be noted that tropospheric variability emerging after week two can effectively be ruled out as a significant factor in the subsequent lower stratospheric~~ an SSW. This is consistent with our experimental design, which effectively excludes externally forced tropospheric anomalies (e.g. SST-related signals), such that potential tropospheric precursor signals are absent by construction and the initial tropospheric state is therefore not a decisive factor for the subsequent stratosphere–troposphere evolution. One possible approach to further disentangle the role of tropospheric versus stratospheric initial condition memory in determining the development of the lower stratospheric response is to run forecasts with different scrambled initial states, an example of which can be seen in the work of Davis et al. (2022). They found that forecasts initialized with scrambled stratospheric initial conditions explained most of the observed surface temperature variability in the month after the SSW. Further analyses revealed that disturbed stratospheric states may be an important feedback on persistent tropospheric surface behaviour, rather than its proximate cause.

Examining the longitudinally resolved tropospheric circulation, we find a clear Greenland blocking signal in weeks 3–7 after SSW onset in the ~~re-forecast~~ spin-off composite mean (Fig. 12), consistent with earlier case-study work (e.g., Kautz

et al., 2022). Ensemble members that develop an anomalously strong lower stratosphere two weeks after the initial event show an amplified surface response, both in the full-field MSLP pattern (Fig. 11) and in blocking frequency (Fig. 12). A further key finding of our study is illustrated in Fig. 13~~b~~, which reveals a strong correlation ($r = 0.84$) between large-scale polar-cap surface GPH anomalies and regional blocking frequencies over Greenland. Combining the above results~~and this~~ , the robust and near-linear relationship identified here implies a clear sequence of events: *(i)* an anomalous lower-stratospheric state modulates the zonal-mean tropospheric flow and favours a pronounced zonal-mean surface response, which in turn *(ii)* increases the likelihood and intensity of regional atmospheric blocking. For Central Europe this sequence implies a higher probability of cold-air outbreaks (due to the enhanced blocking occurrence over Greenland and the adjacent North Atlantic ocean)~~, while over Scandinavia the simulations still indicate an increased chance of cold spells (not shown), even though they do not reproduce the more frequently discussed Scandinavian-blocking configuration~~ . Thus, lower stratospheric anomalies act less as a direct trigger and more as a regulator that makes certain tropospheric states, such as Greenland blocking, statistically more likely across the 18 ~~SSW events~~ SSWs analysed. Determining the actual mechanism of how the lower stratosphere anomaly couples down to the troposphere will require future studies (see e.g., Baldwin et al., 2021, for an overview of studies on this topic).

Note that the ensemble-mean nature of the ~~re-forecast~~ spin-off dataset makes this chain particularly evident; by contrast, the ~~SSW events~~ SSWs from the EGE control runs show essentially no correlation between blocking and large-scale surface signals (~~Fig. 13a~~not shown), highlighting the added value and statistical robustness of our ~~re-forecast~~ spin-off approach.

## 9 Summary and Conclusion

In this study, a dedicated simulation set-up is used to shed light on the question whether, and why, some ~~SSW events~~ SSWs are more likely than others to induce a tropospheric response. Firstly, an ensemble of winter simulations provide a set of ~~SSW events~~ SSWs under controlled boundary conditions, excluding external factors to cause differences between events. Secondly, ~~ensemble re-forecasts~~ spin-off ensembles centred around selected ~~SSW events~~ SSWs set the stage to ascertain whether ~~downward coupling of SSW events is truly~~ the tropospheric response to individual events is distinct, or whether the different observed post-SSW tropospheric evolutions are primarily of random nature.

Constructing this simulation setup has two key benefits: i) Using the free-running large-ensemble simulation to generate a large set of realistic wintertime evolutions with SSWs in a controlled climate, while retaining independent stratosphere-troposphere evolution, rules out the possibility that ~~different external~~ differences in external confounding forcings (e.g. ENSO, QBO) lead to differences between ~~SSW events~~ SSWs in our EGE. In particular, this means that all differences have to result from differences in internal dynamical evolution. ii) The ~~ensemble re-forecasts~~ spin-off ensembles pave the way for ~~an improved~~ a robust statistical characterisation of the ~~possible tropospheric evolutions~~ tropospheric response following stratospheric warming events, ~~as~~ via the quantification of the distribution of ~~the response is~~ possible tropospheric evolutions facilitated through the ensemble setup. Since the simulations are centred around the start date of the SSW ~~event~~ and hence all 40 members

capture the initial ~~SSW~~ event identically, we effectively 'average out' the tropospheric internal variability in the post-event surface response, a key advantage of these ~~ensemble re-forecasts~~spin-off ensembles.

In Section 3, we validated the suitability of this EGE for simulating realistic wintertime evolutions by comparing its performance against ERA5 reanalysis data from 1979–2019. Our evaluation showed that the EGE realistically captures enhanced lower-stratospheric and surface anomalies following major ~~SSW events~~SSWs. Notably, the modelled stratosphere-troposphere coupling closely resembles the characteristics of previously observed events reported in observation-based studies (e.g. Baldwin and Dunkerton, 2001).

To further investigate the variability and predictability of surface responses following SSWs, we conducted ~~ensemble re-forecast~~ spin-off ensemble simulations for 18 carefully selected events from the original EGE dataset. These ~~re-forecasts~~ spin-offs were specifically chosen to represent a realistic range of variability in initial event strengths, lower-stratospheric anomalies, surface responses, and occurrence times within the winter season. Our key findings are outlined in the following paragraphs.

Our analysis of the ~~re-forecast~~ spin-off simulations revealed substantial event-to-event variability in the likelihood of downward coupling, i.e., some SSWs are more likely to lead to a tropospheric response than others. Crucially, we identified a strong correlation between the strength of lower-stratospheric anomalies shortly after the SSW onset and the magnitude of the ensemble average surface signals starting approximately three weeks later. ~~Lower-stratospheric~~ Lower stratospheric anomalies shortly after the SSW ~~were found to be superior predictors of~~ tend to show a stronger relationship with subsequent surface anomalies ~~compared to~~ than the initial intensity of the warming itself ~~. This emphasizes the critical role of the lower stratosphere as a mediator controlling the downward coupling process~~(based on correlations across the 18 event-wise ensemble-mean responses).

Examining the TEM momentum budget highlighted the potential importance of wave reflection events occurring post-SSW for the mean flow evolution. In particular, the height of the reflective surface can influence whether persistent lower-stratospheric anomalies are established: if the reflection occurs within the lower stratosphere, the associated acceleration of mean winds lead to a quick recovery from the SSW-related negative wind anomaly in the lower stratosphere. If the reflection occurs at higher levels, the lower stratospheric anomalies are less affected and might persist, creating conditions that favour coupling to the troposphere. By conducting our experiments ~~in~~ with ensembles (averaging out the noise), we isolated this wave-reflection mechanism and its potential effects on stratosphere-troposphere coupling. However, the relative importance of this mechanism to inhibit the development of lower stratospheric anomalies after SSWs remains to be investigated.

Importantly, we found that individual ~~SSW events~~ SSWs exhibit considerable case-dependent differences in their likelihood of developing the canonical tropospheric response not only in terms of ~~zonal mean~~ zonal-mean circulation, but also on a regional level. Specifically, stronger polar cap surface anomalies after the warming event translate almost linearly to increased blocking frequencies over key regions such as Hudson Bay, Greenland, and the North Atlantic Ocean. These differences in SSW response can be reliably predicted as early as the day of the SSW ~~event~~ itself, with the magnitude of post-event lower-stratospheric anomalies being an effective predictor.

## Data availability

This study used the ERA5 ~~re-analysis~~ reanalysis dataset, details of which can be found on the ECMWF website https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5, and be accessed via the Climate Data Store https://cds.climate.copernicus.eu/. Details pertaining to the ERA5 SSW dataset can be accessed via the Sudden Stratospheric Warming Compendium (https://csl.noaa.gov/groups/csl8/sswcompendium/). Datasets including the $\overline{U}_{60N}^{10hPa}$ time series for for numerical ensembles and ~~re-analysis~~ reanalysis data will be provided in a public repository upon publication.

## Author contributions

SL performed the analyses and prepared the manuscript with contributions from all co-authors. PR performed the simulations. The concept was developed together with all co-authors.

## Competing interests

Some authors are members of the editorial board of journal WCD.


## Acknowledgements

The authors thank the Transregional Collaborative Research Center SFB/TRR 165 “Waves to Weather” funded by the German Research Foundation (DFG) for support. We thank Sebastian Borchert from the Deutscher Wetterdienst (DWD) for providing helpful insights about the ICON setup.

## Financial support

This work was supported by the Transregional Collaborative Research Center SFB/TRR 165 “Waves to Weather” funded by the German Research Foundation (DFG) and the German Aerospace Centre (DLR) in the framework of the Open Access Publishing Program. JGP thanks the AXA Research Fund for support.

## References


Andrews, D. G. and McIntyre, M. E.: Planetary Waves in Horizontal and Vertical Shear: The Generalized Eliassen-Palm Relation and the Mean Zonal Acceleration, Journal of Atmospheric Sciences, 33, 2031 – 2048, https://doi.org/10.1175/1520-0469(1976)033<2031:PWIHAV>2.0.CO;2, 1976.

Baldwin, M. and Dunkerton, T.: Stratospheric Harbingers of Anomalous Weather Regimes, Science, 294, 581–584, https://doi.org/10.1126/science.1063315, 2001.

Baldwin, M. P., Ayarzagüena, B., Birner, T., Butchart, N., Butler, A. H., Charlton-Perez, A. J., Domeisen, D. I. V., Garfinkel, C. I., Garny, H., Gerber, E. P., Hegglin, M. I., Langematz, U., and Pedatella, N. M.: Sudden Stratospheric Warmings, Reviews of Geophysics, 59, e2020RG000 708, https://doi.org/https://doi.org/10.1029/2020RG000708, e2020RG000708 10.1029/2020RG000708, 2021.

Beerli, R. and Grams, C. M.: Stratospheric modulation of the large-scale circulation in the Atlantic–European region and its implications for surface weather events, Quarterly Journal of the Royal Meteorological Society, 145, 3732–3750, https://doi.org/https://doi.org/10.1002/qj.3653, 2019.

Bett, P. E., Scaife, A. A., Hardiman, S. C., Thornton, H. E., Shen, X., Wang, L., and Pang, B.: Using large ensembles to quantify the impact of sudden stratospheric warmings and their precursors on the North Atlantic Oscillation, Weather and Climate Dynamics, 4, 213–228, https://doi.org/10.5194/wcd-4-213-2023, 2023.

Birner, T. and Albers, J. R.: Sudden Stratospheric Warmings and Anomalous Upward Wave Activity Flux, SOLA, 13A, 8–12, https://doi.org/10.2151/sola.13A-002, 2017.

Butler, A. H., Seidel, D. J., Hardiman, S. C., Butchart, N., Birner, T., and Match, A.: Defining Sudden Stratospheric Warmings, Bulletin of the American Meteorological Society, 96, 1913 – 1928, https://doi.org/10.1175/BAMS-D-13-00173.1, 2015.

Butler, A. H., Sjoberg, J. P., Seidel, D. J., and Rosenlof, K. H.: A sudden stratospheric warming compendium, Earth System Science Data, 9, 63–76, https://doi.org/10.5194/essd-9-63-2017, 2017.

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/https://doi.org/10.1002/qj.3858, 2020.

Cagnazzo, C. and Manzini, E.: Impact of the Stratosphere on the Winter Tropospheric Teleconnections between ENSO and the North Atlantic and European Region, Journal of Climate, 22, 1223 – 1238, https://doi.org/10.1175/2008JCLI2549.1, 2009.

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.

Dai, Y., Hitchcock, P., Butler, A. H., Garfinkel, C. I., and Seviour, W. J. M.: Assessing stratospheric contributions to subseasonal predictions of precipitation after the 2018 sudden stratospheric warming from the Stratospheric Nudging And Predictable Surface Impacts (SNAPSI) project, Weather and Climate Dynamics, 6, 841–862, https://doi.org/10.5194/wcd-6-841-2025, 2025.

Davis, N. A., Richter, J. H., Glanville, A. A., Edwards, J., and LaJoie, E.: Limited surface impacts of the January 2021 sudden stratospheric warming, Nature Communications, 13, 1136, 2022.

Ding, X., Chen, G., Zhang, P., Domeisen, D. I. V., and Orbe, C.: Extreme stratospheric wave activity as harbingers of cold events over North America, Communications Earth & Environment, 4, 187, 2023.

Domeisen, D., Butler, A. H., Charlton-Perez, A. J., Ayarzagüena, B., Baldwin, M. P., Dunn-Sigouin, E., Furtado, J. C., Garfinkel, C. I., Hitchcock, P., Karpechko, A. Y., Kim, H., Knight, J., Lang, A. L., Lim, E.-P., Marshall, A., Roff, G., Schwartz, C., Simpson, I. R., Son, S.-W., and Taguchi, M.: The Role of the Stratosphere in Subseasonal to Seasonal Prediction: 1. Predictability of the Stratosphere, Journal

of Geophysical Research: Atmospheres, 125, e2019JD030 920, https://doi.org/https://doi.org/10.1029/2019JD030920, e2019JD030920 10.1029/2019JD030920, 2020a.

Domeisen, D. I. V., Grams, C. M., and Papritz, L.: The role of North Atlantic–European weather regimes in the surface impact of sudden stratospheric warming events, Weather and Climate Dynamics, 1, 373–388, https://doi.org/10.5194/wcd-1-373-2020, 2020b.

Dunn-Sigouin, E. and Shaw, T.: Dynamics of Anomalous Stratospheric Eddy Heat Flux Events in an Idealized Model, Journal of the Atmospheric Sciences, 77, 2187 – 2202, https://doi.org/10.1175/JAS-D-19-0231.1, 2020.

Dunn-Sigouin, E. and Shaw, T. A.: Comparing and contrasting extreme stratospheric events, including their coupling to the tropospheric circulation, Journal of Geophysical Research: Atmospheres, 120, 1374–1390, https://doi.org/https://doi.org/10.1002/2014JD022116, 2015.

Elsbury, D., Butler, A., Peings, Y., and Magnusdottir, G.: Sensitivity of Easterly QBO's Boreal Winter Teleconnections and Surface Impacts to SSWs, Journal of Climate, 37, 3675 – 3688, https://doi.org/10.1175/JCLI-D-23-0395.1, 2024.

Gerber, E. P., Orbe, C., and Polvani, L. M.: Stratospheric influence on the tropospheric circulation revealed by idealized ensemble forecasts, Geophysical Research Letters, 36, https://doi.org/https://doi.org/10.1029/2009GL040913, 2009.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/https://doi.org/10.1002/qj.3803, 2020.

Hitchcock, P. and Simpson, I. R.: The Downward Influence of Stratospheric Sudden Warmings, Journal of the Atmospheric Sciences, 71, 3856 – 3876, https://doi.org/10.1175/JAS-D-14-0012.1, 2014.

Hitchcock, P., Shepherd, T. G., and Manney, G. L.: Statistical Characterization of Arctic Polar-Night Jet Oscillation Events, Journal of Climate, 26, 2096 – 2116, https://doi.org/10.1175/JCLI-D-12-00202.1, 2013.

Hitchcock, P., Butler, A., Charlton-Perez, A., Garfinkel, C. I., Stockdale, T., Anstey, J., Mitchell, D., Domeisen, D. I. V., Wu, T., Lu, Y., Mastrangelo, D., Malguzzi, P., Lin, H., Muncaster, R., Merryfield, B., Sigmond, M., Xiang, B., Jia, L., Hyun, Y.-K., Oh, J., Specq, D., Simpson, I. R., Richter, J. H., Barton, C., Knight, J., Lim, E.-P., and Hendon, H.: 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.

Karpechko, A. Y., Hitchcock, P., Peters, D. H. W., and Schneidereit, A.: Predictability of downward propagation of major sudden stratospheric warmings, Quarterly Journal of the Royal Meteorological Society, 143, 1459–1470, https://doi.org/https://doi.org/10.1002/qj.3017, 2017.

Karpechko, A. Y., Charlton-Perez, A., Balmaseda, M., Tyrrell, N., and Vitart, F.: Predicting Sudden Stratospheric Warming 2018 and Its Climate Impacts With a Multimodel Ensemble, Geophysical Research Letters, 45, 13,538–13,546, https://doi.org/https://doi.org/10.1029/2018GL081091, 2018.

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.

Kautz, L.-A., Martius, O., Pfahl, S., Pinto, J. G., Ramos, A. M., Sousa, P. M., and Woollings, T.: Atmospheric blocking and weather extremes over the Euro - Atlantic sector - a review, Weather and Climate Dynamics, 3, 305–336, https://doi.org/10.5194/wcd-3-305-2022, 2022.

Lee, S., Furtado, J., and Charlton-Perez, A.: Wintertime North American Weather Regimes and the Arctic Stratospheric Polar Vortex, Geophysical Research Letters, 46, https://doi.org/10.1029/2019GL085592, 2019.

Lu, Y., Tian, W., Zhang, J., Huang, J., Zhang, R., Wang, T., and Xu, M.: The Impact of the Stratospheric Polar Vortex Shift on the Arctic Oscillation, Journal of Climate, 34, 4129 – 4143, https://doi.org/10.1175/JCLI-D-20-0536.1, 2021.

Ma, J., Chen, W., Yang, R., Ma, T., and Shen, X.: Downward propagation of the weak stratospheric polar vortex events: the role of the surface arctic oscillation and the quasi-biennial oscillation, Climate Dynamics, 62, 4117–4131, https://doi.org/10.1007/s00382-024-07121-5, 2024.

Magnusson, L., Nycander, J., and Källén, E.: Flow-dependent versus flow-independent initial perturbations for ensemble prediction, Tellus A, 61, 194–209, https://doi.org/https://doi.org/10.1111/j.1600-0870.2008.00385.x, 2009.

Maycock, A. C. and Hitchcock, P.: Do split and displacement sudden stratospheric warmings have different annular mode signatures?, Geophysical Research Letters, 42, 10,943–10,951, https://doi.org/https://doi.org/10.1002/2015GL066754, 2015.

Nebel, D. M., Garfinkel, C. I., Cohen, J., Domeisen, D. I. V., Rao, J., and Schwartz, C.: The Predictability of the Downward Versus Non-Downward Propagation of Sudden Stratospheric Warmings in S2S Hindcasts, Geophysical Research Letters, 51, e2024GL110 529, https://doi.org/https://doi.org/10.1029/2024GL110529, e2024GL110529 2024GL110529, 2024.

Oehrlein, J., Polvani, L. M., Sun, L., and Deser, C.: How Well Do We Know the Surface Impact of Sudden Stratospheric Warmings?, Geophysical Research Letters, 48, e2021GL095 493, https://doi.org/https://doi.org/10.1029/2021GL095493, e2021GL095493 2021GL095493, 2021.

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/https://doi.org/10.1029/2019JD031919, e2019JD031919 2019JD031919, 2020.

Rao, J., Garfinkel, C. I., Wu, T., Lu, Y., Lu, Q., and Liang, Z.: The January 2021 Sudden Stratospheric Warming and Its Prediction in Subseasonal to Seasonal Models, Journal of Geophysical Research: Atmospheres, 126, e2021JD035 057, https://doi.org/https://doi.org/10.1029/2021JD035057, e2021JD035057 2021JD035057, 2021.

Runde, T., Dameris, M., Garny, H., and Kinnison, D.: Classification of stratospheric extreme events according to their downward propagation to the troposphere: Classification of stratospheric events, Geophysical Research Letters, 43, https://doi.org/10.1002/2016GL069569, 2016.

Rupp, P., Loeffel, S., Garny, H., Chen, X., Pinto, J. G., and Birner, T.: Potential Links Between Tropospheric and Stratospheric Circulation Extremes During Early 2020, Journal of Geophysical Research: Atmospheres, 127, e2021JD035 667, https://doi.org/https://doi.org/10.1029/2021JD035667, e2021JD035667 2021JD035667, 2022.

Scaife, A. A., Baldwin, M. P., Butler, A. H., Charlton-Perez, A. J., Domeisen, D. I. V., Garfinkel, C. I., Hardiman, S. C., Haynes, P., Karpechko, A. Y., Lim, E.-P., Noguchi, S., Perlwitz, J., Polvani, L., Richter, J. H., Scinocca, J., Sigmond, M., Shepherd, T. G., Son, S.-W., and Thompson, D. W. J.: Long-range prediction and the stratosphere, Atmospheric Chemistry and Physics, 22, 2601–2623, https://doi.org/10.5194/acp-22-2601-2022, 2022.

Scherrer, S. C., Croci-Maspoli, M., Schwierz, C., and Appenzeller, C.: Two-dimensional indices of atmospheric blocking and their statistical relationship with winter climate patterns in the Euro-Atlantic region, International Journal of Climatology, 26, 233–249, https://doi.org/10.1002/joc.1250, 2006.

Sigmond, M., Scinocca, J. F., Kharin, V. V., and Shepherd, T. G.: Enhanced seasonal forecast skill following stratospheric sudden warmings, Nature Geoscience, 6, 98–102, 2013.

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.

Yadav, P., Garfinkel, C. I., and Domeisen, D. I. V.: The Role of the Stratosphere in Teleconnections Arising From Fast and Slow MJO Episodes, Geophysical Research Letters, 51, e2023GL104 826, https://doi.org/https://doi.org/10.1029/2023GL104826, e2023GL104826 2023GL104826, 2024.

Zängl, G., Reinert, D., Rípodas, P., and Baldauf, M.: The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core, Quarterly Journal of the Royal Meteorological Society, 141, 563–579, https://doi.org/https://doi.org/10.1002/qj.2378, 2015.