Potential links between tropospheric and stratospheric circulation extremes during early 2020

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

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

Key Points:

- Description of the atmospheric evolution in early 2020, characterized by extreme strengths of the mid-latitude jet and the polar vortex
- Analysis of the strat. wave reflection event observed in February 2020 and its influence on strat-trop evolution and predictability
- Quantification of strat-trop coupling and its effects on trop. flow extremes using large-ensemble simulations and clustering approaches

Corresponding author: Philip Rupp, `philip.rupp@lmu.de`

Abstract

February-March 2020 was marked by highly anomalous large-scale circulations in the Northern extratropical troposphere and stratosphere. The Atlantic jet reached extreme strength, linked to some of the strongest and most persistent positive values of the Arctic Oscillation index on record, which provided conditions for extreme windstorms hitting Europe. Likewise, the stratospheric polar vortex reached extreme strength that persisted for an unusually long period. Past research indicated that such circulation extremes occurring throughout the troposphere-stratosphere system are dynamically coupled, although the nature of this coupling is still not fully understood and generally difficult to quantify.

We employ sets of numerical ensemble simulations to provide statistical characterizations of the mutual coupling of the early 2020 circulation extremes. We find an overall robust coupling between tropospheric and stratospheric anomalies: ensemble members with polar vortex exceeding a certain strength tend to exhibit a stronger tropospheric jet and vice versa. Moreover, members exhibiting a breakdown of the stratospheric circulation (associated with a sudden stratospheric warming) tend to lack periods of persistently enhanced tropospheric circulation. A crucial component in determining the stratospheric vortex strength appears to be the vortex geometry, corresponding to states that promote either reflection or dissipation of planetary waves. Despite these lines of evidence for vertical coupling, our simulations underline the role of internal variability within each atmospheric layer. The circulation extremes during early 2020 may be viewed as resulting from a fortuitous alignment of dynamical evolutions within the troposphere and stratosphere, aided by each layer's modification of the other layer's boundary condition.

1 Introduction

1.1 General coupling of troposphere and stratosphere

The dynamical coupling between the troposphere and stratosphere is an important aspect of the climate system with considerable meteorological implications for the surface weather in mid-latitudes. The resulting correlation between tropospheric and stratospheric variability (e.g., Thompson & Wallace, 1998) can be of particular importance for medium-range weather forecasts since the dynamics of the two layers mostly evolves on substantially different time scales (typically days within the troposphere and weeks within the stratosphere). A comprehensive understanding of the fundamental principles of troposphere-stratosphere interactions can help to improve the accuracy of weather forecasts and climate projections and to extend the corresponding predictable time ranges (e.g., Baldwin et al., 2003; Sigmond et al., 2013; Tripathi et al., 2015; Scaife et al., 2016).

Traditionally troposphere-stratosphere coupling considered planetary wave forcing of the stratospheric circulation (Charney & Drazin, 1961). More recently, also downward coupling following (extreme) circulation anomalies in the stratosphere has been documented and is now established observationally (e.g., Baldwin & Dunkerton, 2001; Domeisen & Butler, 2020). More complex combinations of such interactions are possible, with troposphere and stratosphere feeding back onto each other. A useful way to study stratosphere-troposphere (strat-trop) interactions is to analyze stratospheric (tropospheric) extremes and their influence on the troposphere (stratosphere). Examples of such extremes are periods with persistently strengthened/weakened circulation or (shorter) periods with extremely anomalous circulation (usually exceeding several climatological standard deviations). Note that coupling mechanisms related to extremes might not necessarily behave symmetrically for anomalously weak or strong

atmospheric circulations and thus a corresponding distinction is potentially important (e.g., Domeisen et al., 2020).

1.2 Downward influence of periods with strong or weak polar vortex

Much of the existing literature has focused on tropospheric signals following stratospheric polar vortex breakdowns associated with sudden stratospheric warming (SSW) events (Tomassini et al., 2012; Kautz et al., 2020; Karpechko et al., 2017; Charlton-Perez et al., 2018). Such conditions with weakened stratospheric circulation are often preceded by interactions of the zonal mean flow with vertically propagating planetary waves leading to an anomalous warming over the pole and a corresponding rapid deceleration of stratospheric winds (Matsuno, 1971; Baldwin et al., 2021). Note that periods with weakened polar vortex are not necessarily associated with a complete breakdown of the stratospheric circulation (i.e., a conventional SSW), but can also occur following sudden stratospheric deceleration (SSD) events, where stratospheric winds weaken substantially but remain predominantly westerly.

It has been shown that the downward influence of periods with weak polar vortex (or correspondingly SSWs) on the troposphere can lead to an equatorward shift of the mid-latitude jet. Such jet shifts typically correspond to negative anomalies of hemispheric scale tropospheric metrics like the Northern Annular Mode (NAM) and the Arctic Oscillation (AO), often used to quantify the strength and position of the tropospheric mid-latitude jet. A typical downward influence of a weak stratospheric polar vortex can therefore be thought of as weakening the tropospheric circulation. Several mechanisms for the tropospheric response following SSWs have been proposed and studied over the past decades, including synoptic scale tropospheric eddy feedback (Domeisen et al., 2013; Hitchcock & Simpson, 2014; Wittman et al., 2007; Rupp & Birner, 2021), interactions involving planetary waves (Song & Robinson, 2004; Shaw et al., 2010) and the downward control principle (Haynes et al., 1991). Besides an influence on the tropospheric large scale circulation, SSWs and periods with weak polar vortex have been shown to be associated with more regional tropospheric signatures, including a preference for negative values of the Northern Atlantic Oscillation (NAO, e.g., Domeisen, 2019; Beerli & Grams, 2019), effects on the occurrence of tropospheric weather regimes (e.g., Woollings et al., 2010; Domeisen et al., 2020) and an increase in likelihood of extreme surface events like cold-spells (e.g., Thompson et al., 2002; Kolstad et al., 2010; Kautz et al., 2020).

Studies investigating the tropospheric response to a strong polar vortex are relatively sparse and thus the robustness of the downward coupling of strong stratospheric flow anomalies and their influence on extended-range predictability are not yet fully understood. Generally (extended) periods with a relatively strong vortex can couple to the troposphere and influence the corresponding large-scale circulation. Limpasuvan et al. (2005), for example, found the tropospheric NAM to be anomalously positive following periods with a strong polar vortex, corresponding to a general strengthening of the tropospheric jet. They further note various differences in the details like timing and strength of the tropospheric response to a strong polar vortex compared to the equivalent weak vortex phenomenon. Periods of strong tropospheric jet stream are typically characterized with an enhanced number of intense cyclones over the region (e.g., Pinto et al., 2009). During the winter 2013/2014, for example, Western Europe was affected by an unusually high number of storms associated with a long-enduring intensified and more zonal North Atlantic jet stream (Matthews et al., 2014; ?, ?). Over the British Isles, this series of storms lead to enduring wet and windy conditions, leading to widespread floods (Huntingford et al., 2014; Priestley et al., 2017). The occurrence of multiple and related cyclones over a certain area is know as cyclone clustering (Mailier et al., 2006; Dacre & Pinto, 2020). Although many authors find the two stratospheric situations (strong/weak vortex) to generally have a similar (but

opposite) tropospheric response (Baldwin & Dunkerton, 2001; Thompson et al., 2002; Tripathi et al., 2015), certain differences in response could in principle be expected, e.g., based on asymmetries in the potential for strat-trop wave coupling during strong and weak vortex conditions (further discussed in Section 7).

1.3 Mechanisms leading to a strong or weak polar vortex

The dynamical processes involved in causing periods with weak or strong polar vortex typically involve either equatorward or anomalously strong poleward planetary-scale eddy heat flux, respectively, which further corresponds to anomalies in the vertical wave activity flux F_p (also referred to as vertical component of the Eliassen-Palm flux). Stratospheric conditions are therefore strongly coupled to (lower) stratospheric planetary wave activity (Dunn-Sigouin & Shaw, 2015; Birner & Albers, 2017) and the associated anomalies in vertical propagation of planetary waves. A period of anomalously strong upward wave flux (corresponding to poleward heat flux) is likely associated with a subsequent SSW or SSD event. Periods of anomalously weak upward wave flux or even downward wave flux, on the other hand, are likely associated with a subsequent strengthening of the polar vortex. In some situations polar vortex conditions can change in a way that leads to the formation of a reflecting surface in the upper stratosphere and a corresponding downward reflection of previously upward propagating waves. These phenomena are often referred to as wave reflection events (WRE) and have been shown to couple to subsequent periods of strong polar vortex.

Harnik and Lindzen (2001) studied the formation of reflective surfaces for meridional and vertical propagation in the Southern hemispheric stratosphere and showed that the WREs caused by these surfaces can strongly influence the structure and strength of the (Southern) polar vortex on both seasonal and daily time scales. Their work also stressed the importance of both meridional and vertical reflective surfaces for the dynamical evolution during a WRE. Shortly after, Perlwitz and Harnik (2003) provided observational evidence that wave reflections also occur in the Northern hemisphere, predominantly in the period January to March. By now it has been established that the formation of (persistent) reflective surfaces often follows (short) pulses of upward propagating waves and a corresponding strong deceleration of upper stratospheric winds (Harnik & Lindzen, 2001; Perlwitz & Harnik, 2003; Harnik, 2009). Following this idea, the state of the stratosphere can often be classified as reflective or non-reflective (based on the vertical gradient of zonal mean zonal wind in the upper stratosphere). Perlwitz and Harnik (2004) showed that Northern hemispheric winter months in a reflective state are mostly characterized by downward wave coupling, while non-reflective months are more often characterized by zonal-mean coupling. Recent work by Dunn-Sigouin and Shaw (2015) study the relationship between extreme stratospheric events with vertical coupling of planetary waves and the strength of the polar vortex. They show that the polar vortex experiences a deceleration or acceleration during upward or downward wave events, respectively. They also put forward that tropospheric impacts resulting from planetary wave events are larger in magnitude than those following a strong vortex event. Note that most of the studies investigating the cause and effect of WREs focus on events involving only a single wave number (usually wave number 1), but in principle individual planetary wave numbers can behave differently and hence atmospheric states combining reflective and non-reflective characteristics are possible (as discussed in Section 4).

1.4 The importance of studying individual cases and large ensembles

Although the tropospheric response to anomalously strong or weak stratospheric circulations seems to be robust in composite studies with averages over many cases (either periods with anomalous winds or following corresponding deceleration or acceleration events), the variability on a case-to-case basis is high in terms of details like

strength or timing of the tropospheric response, and periods of anomalous stratospheric conditions without any apparent tropospheric signal are not uncommon (Baldwin & Dunkerton, 2001; Runde et al., 2016; Domeisen & Butler, 2020). One aspect that seems to be strongly correlated to whether anomalous stratospheric conditions or events couple down to the troposphere is the response of the lower stratosphere, where particularly cases with persistent lower stratospheric anomalies produce a coherent tropospheric response (e.g., Hitchcock & Simpson, 2014; Runde et al., 2016; Karpechko et al., 2017). As discussed in Section 7, the importance of lower stratospheric circulation anomalies can potentially be interpreted as a modification of internal tropospheric dynamics via a change in the corresponding upper boundary condition of the (purely tropospheric) system.

The highly probabilistic nature of strat-trop interaction generally increases the subtlety and difficulties in finding coherent mechanisms for the downward propagation of stratospheric anomalies into the troposphere and allows for a wide range of possible interactions and behaviors (see, e.g., Rao et al., 2020; Domeisen et al., 2020; Domeisen & Butler, 2020). Hence, caution is required in particular when analyzing specific cases in a deterministic framework. Taguchi (2008) found that the hypotheses of SSWs either preceding or following tropospheric blocking events could not be supported based on a comprehensive statistical analysis of 49 winter periods, although they note that individual cases can show pronounced signs of an apparent coupling. As an example, Wang and Chen (2010) studied the cold winter of 2009/10 and hypothesized a stratospheric involvement based on an extremely negative surface NAM signal following a stratospheric deceleration event, particularly emphasizing the relatively weak and short-lived nature of the stratospheric anomalies in this case. On the other hand, Kautz et al. (2020) used an ensemble-simulation approach to analyze the SSW that occurred in early 2018 and its potential contribution to the Eurasian cold spell observed during that winter. They attributed an increased likelihood for cold surface temperatures to the SSW, but emphasized caution to interpret the corresponding correlation as a probabilistic influence rather than direct causal link. Such probabilistic analyzes with large-ensemble simulations (as used in the present study) are a useful way to quantify the robust signal of stratospheric downward influence, which tends to be weak compared to the internal variability of the troposphere. Ensemble studies are further necessary to fully sample the range of possible atmospheric behavior and strat-trop interactions, especially when it comes to the study of extreme situations.

Detailed studies of individual cases of downward influence are crucial to complement composite studies with averages over many events due to the large variability in the strat-trop coupling signal. In particular, studying the potential downward influence of extreme stratospheric circulation anomalies is essential to fully understand the dynamical links between troposphere and stratosphere. The winter period of 2019/20 was one of the most extreme ever recorded in terms tropospheric and stratospheric dynamics, with both layers experiencing periods of extremely strong large-scale circulation. The AO index and the stratospheric polar vortex strength showed anomalous persistence and reached extreme positive anomalies. Different authors have presented evidence for a strong apparent correlation between stratospheric and tropospheric anomalies, predominantly on monthly to seasonal time scales (Lawrence et al., 2020; Lee et al., 2020). Particularly, the occurrence of a WRE in mid February was suggested to have contributed to the formation of extreme and persistent tropospheric circulation anomalies in February and March due to destructive interference of the reflected wave with the climatological standing wave pattern in the troposphere. Besides a potential stratospheric downward influence, Hardiman et al. (2020) provided evidence that tropospheric internal dynamics (in particular an anomalously strong Indian ocean dipole) contributed to the formation of the persistent tropospheric circulation anomalies of early 2020 and their anomalously high predictability (some aspects were predictable at seasonal lead times).

1.5 Structure of this study

The present study aims to analyze the potential links between tropospheric and stratospheric extremes in the large scale zonal circulation during January to March 2020, using a combination of re-analysis and large-ensemble approaches. The paper is structured as follows. Section 2 provides an overview of the numerical model and dataset used. Section 3 discusses the evolution of tropospheric and stratospheric circulation on different scales during early 2020 within re-analysis data and ensemble mean forecasts. In particular we identify multiple periods of extreme atmospheric behavior, including a pronounced WRE in early February. Hence, in Section 4 we give a detailed description of the dynamics associated with this WRE and analyze its potential contribution to strengthening the polar vortex in the following weeks. Section 5 then identifies potential indications for two-way strat-trop coupling, in particular a downward influence of the WRE and the associated strengthened vortex in February. In contrast to the downward influence of stratospheric flow anomalies induced due to a specific stratospheric behavior like the WRE, Section 6 studies the downward coupling of generally anomalous stratospheric circulation like observed during late February and early March. Section 7 then discusses the findings of this paper before, in Section 8, we summarize our main conclusions.

2 Model and dataset

2.1 ICON ensemble simulations

We use the comprehensive 'ICOsahedral Nonhydrostatic' (ICON) model of the German weather service (DWD) to perform a series of ensemble forecast simulations (for a detailed description of the dynamical core see Zängl et al., 2015). The model uses a triangular grid with roughly 40 km horizontal resolution and 90 terrain-following height levels in the vertical up to a height of 75 km (with about 40 levels in the extratropical troposphere). The output is interpolated onto a 1° regular grid following 52 levels of constant pressure. The time step used within the dynamical core of the model runs is 1 minute, with output given as 6-hourly data (although most diagnostics are shown as daily means for easier interpretation).

All simulations are initialized with realistic atmospheric conditions taken from operational ICON analysis products provided by DWD as individual sets of initial conditions for a 40 member ensemble (with slightly perturbed initial conditions for a given initialization date). In order to increase the ensemble size, and thus obtain more robust statistics, we combine several ensemble runs of slightly varying initialization time (with time difference of 1-3 days) to form 'time lagged ensembles'. This approach was first proposed as a method to increase the efficiency of the (deterministic) operational forecast by Hoffman and Kalnay (1983). It has been shown that (time) lagged ensembles can provide an increased prediction skill for the ensemble mean (likely due to the increase in ensemble size) for medium range forecasts (e.g., Dalchér et al., 1988; Branković et al., 1990). However, we do not primarily intend to use the lagged ensemble approach to improve ensemble mean evolution. Instead, the small temporal offset of the different ensembles are taken to be equivalent to small perturbations of the initial conditions, with the intention of improving the sampling of the underlying probability distribution of possible atmospheric states in the probabilistic period of the model simulation. A similar approach has previously been used by other authors in probabilistic studies on different time scales ranging from days to months (e.g., Katsafados et al., 2014; Vogel et al., 2014).

Note that the numerical model used in the present study includes parametrization schemes for non-resolved orographic (Lott & Miller, 1997) and non-orographic (Orr et al., 2010) gravity waves, respectively. Further, sea surface temperatures are

initialized with data taken from the analysis products mentioned above and varied in time according to daily increments based on the linear interpolation of a monthly mean climatology.

2.2 ERA5 re-analysis data

The ERA5 re-analysis dataset (Hersbach et al., 2020) of the European Centre for Medium range Weather Forecasts (ECMWF) is used as the representation of atmospheric state and behavior during the winter period of 2019/20 and in a climatological sense, with climatology being calculated as the inter-annual average of the years 2000-2019. The dataset used is given as a direct output of the re-analysis on a $1°x1°$ regular horizontal grid following pressure surfaces with temporal resolution of 6h. To reduce any high frequency variability the climatology has been smoothed using a 30-day running mean. Throughout the manuscript we interpret diagnostics based on re-analysis data as observations and thus the true state of the atmosphere.

3 Description and predictability of the extremely anomalous periods in early 2020

3.1 Strategy and definition of metrics

We start the analysis of extreme tropospheric and stratospheric behavior during the early months of the year 2020 by providing a general description of the observed (according to re-analysis data) evolution of the atmosphere (also see Lawrence et al., 2020; Lee et al., 2020) and how it is represented by the ensemble mean of model simulations with varying initialization date. When interpreting the ensemble mean of a given model integration, it can be useful to distinguish between early forecast times, which are strongly influenced by the initial conditions, and later times, where most of the memory of initial conditions is lost. Correspondingly the nature of predictive skill of an ensemble simulation will generally change with progressing lead time. Throughout this manuscript we associate the ability of the model to reproduce a feature of the observed atmospheric evolution within the ensemble mean with deterministic prediction skill (short lead times), since in this case the corresponding feature is generally reproduced by the majority of the individual ensemble members. In contrast, features of the observed atmospheric evolution that are not well captured by the ensemble mean (long lead times) might still be captured probabilistically (i.e., the feature is reproduced by a fraction of ensemble members). The time scales for the deterministic prediction range of an ensemble forecast depend on the particular phenomenon of interest - for the dynamical evolution of large scale circulations typical time ranges are several days within the troposphere and a few weeks within the stratosphere. Note that the transition from time periods with predominantly deterministic prediction skill to periods with mostly probabilistic prediction skill might occur gradually and not necessarily at a well defined point in time. In general the ensemble mean of model simulations tends to converge towards the (model) climatology for long simulation times (as the model transitions out of the deterministic range) since any memory of the initial conditions is lost.

For most of the following analyzes we focus on two indices to characterize and quantify the large scale zonal circulation of troposphere and stratosphere, respectively. We define the U_{10}^{60} index as zonally averaged zonal wind at 60° North and 10 hPa. It is an index typically used to determine the strength and zonal symmetry of the stratospheric polar vortex, in particular to define SSWs (with $U_{10}^{60} < 0$). We further define the $\tilde{AO}$ index as the difference in the zonal mean geopotential height at 1000 hPa of the meridional mean between 30-50° North and 60-90° North. This definition is motivated by the spatial empirical orthogonal function (EOF) 'loading' pattern of the Arctic Oscillation (AO) index and can be used as a quantitative measure of the

strength, position and zonal symmetry of the tropospheric mid-latitude jet. The precise definition of $\tilde{AO}$ was chosen so that it provides a strong climatological correlation with the AO index (within ERA5 re-analysis data). However, in contrast to the AO index, the $\tilde{AO}$ index can be calculated from the available model simulations without the need for a climatological dataset that would be required to perform the EOF analysis.

3.2 Zonal mean signatures of tropospheric and stratospheric circulation extremes

Figures 1a and b show the evolution of U_{10}^{60} and $\tilde{AO}$ during early 2020 according to re-analysis data and the ensemble mean of numerical simulations with varying initialization date (mid January to early February). Both the stratospheric and tropospheric circulation are dominated by periods of extreme and persistent positive anomalies compared to the climatology.

A prominent feature in the observed evolution of the stratospheric index (U_{10}^{60}, see Fig. 1a) is a period with persistently large values spanning roughly from February 10th to March 20th. The U_{10}^{60} index reaches values up to about three times the climatological value during that period and deviations from the climatology temporarily exceed the equivalent of two standard deviations. Prior to this period of elevated values, the observed U_{10}^{60} index shows signs of a (weak) SSD, with a drop of about 25 m/s within one week from about 45 m/s on January 30th to roughly 20 m/s on February 6th (note that U_{10}^{60} drops to roughly the climatological value). However, the stratospheric polar vortex does not stay in this weakened state, but immediately recovers and shows record-high values over the course of the following weeks (e.g., $U_{10}^{60} \approx 60$ m/s on February 17th). The period of persistently extreme positive anomaly ends around March 20th, again showing signs of a SSD with the U_{10}^{60} index dropping more than 30 m/s within about 7 days.

Various numerical model simulations initialized in January and early February are able to capture some aspects of the described evolution of the polar vortex strength. However, we find a high sensitivity of (deterministic) predictive skill to the initialization date. In particular the observed sequence of events in the stratosphere, including a SSD and subsequent strengthening of the polar vortex (due to a WRE; see Section 4) in early February, is not (fully) reproduced by ensembles with lead times exceeding about two weeks (Figure 1). Based on these qualitative differences in prediction skill we can separate the ensembles into three groups with distinct differences in U_{10}^{60} evolution depending on their initialization date: pre-SSD, SSD-onset and post-SSD.

Ensembles within the **pre-SSD** group are initialized well before the SSD (e.g., January 19th and 22nd) and therefore do not capture the SSD in early February and the subsequent strengthening of the polar vortex. The ensembles in the **SSD-onset** group, on the other hand, are initialized just before the onset of the SSD (e.g., January 25th and 28th). Hence, these experiments capture the SSD deterministically but the corresponding ensemble means do not show clear signs of a pronounced recovery phase. In contrast, ensembles in the **post-SSD** group are initialized towards the end or after the SSD (e.g., January 30th or February 1st and 4th) and deterministically capture many of the aspects of the observed evolution of the U_{10}^{60} index, including a pronounced recovery phase in early and mid February. Note that in particular ensembles capturing part of the vortex recovery in their respective ensemble mean evolution (i.e., post-SSD ensembles) show a strong persistence in their anomalous polar vortex strength. The distinct differences in behavior of ensembles in the three groups and the corresponding differences in the downward influence on the troposphere are further discussed in Section 5.

In terms of the evolution of the observed tropospheric large scale circulation (Figure 1b) we find two pronounced periods with extremely large positive $\tilde{AO}$ index,
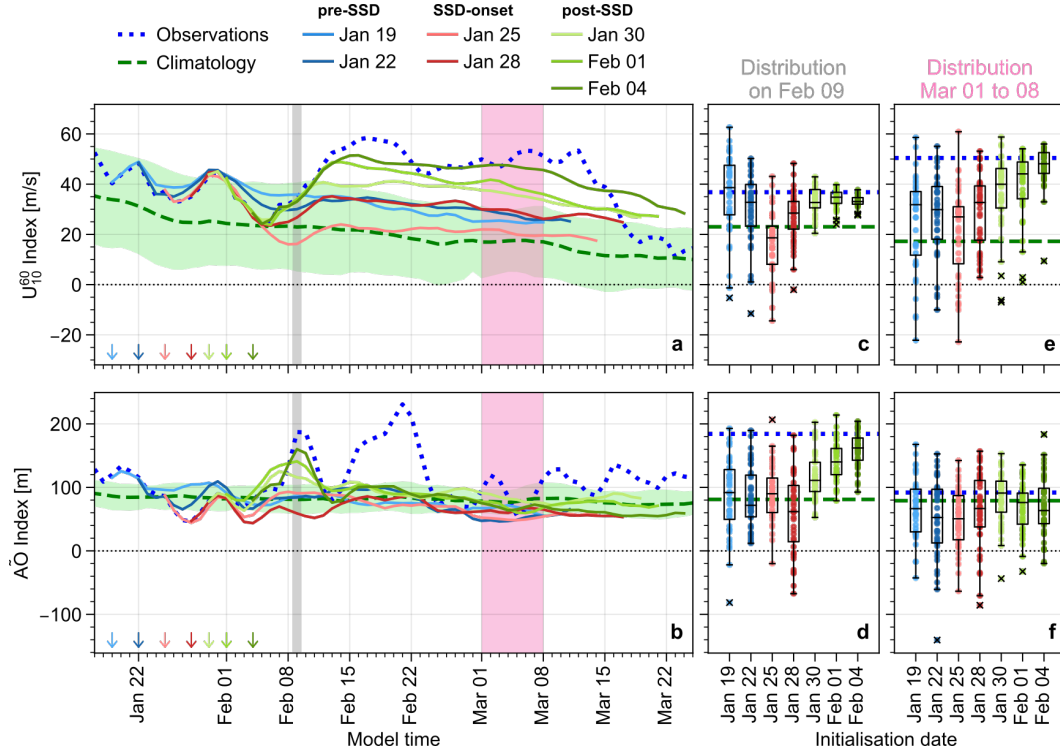
Figure 1. Left column: Evolution of U_{10}^{60} (a) and $\tilde{AO}$ (b) indices in ensemble simulations and re-analysis data. Solid lines: ensemble means for different initialization dates (indicated by arrows on the x-axis). Dotted and dashed lines: re-analysis evolution for early 2020 and climatology, respectively, with corresponding shading indicating the climatological variability in terms of one standard deviation. Right columns: member distribution in U_{10}^{60} (top) and $\tilde{AO}$ (bottom) on February 9th (c and d) and averaged over March 1st to 8th (e and f), compared to climatology and re-analysis observations.

peaking at February 9th and 21st with values around two times the climatological value and deviations from the climatology corresponding to up to four and six standard deviations, respectively. The extreme period around Feb. 9th is (to some extent) represented deterministically within the post-SSD ensemble simulations, while earlier initializations (pre-SSD and SSD-onset) do not seem to capture this event in their ensemble means. Note that the ensemble means of experiments in the post-SSD group also deterministically predict (some aspects of) the observed evolution of the U_{10}^{60} index (in terms of a strong recovery phase following a SSD; see Fig. 1a), suggesting a correlation between these tropospheric and stratospheric events.

The $\tilde{AO}$ index shows persistent positive anomalies during most of February and March, exceeding two climatological standard deviations multiple times. It should be emphasized that the $\tilde{AO}$ index (or equivalently the AO index) has a typical decorrelation time scale of about 1 or 2 weeks (e.g., Gerber et al., 2008) and thus the prolonged persistent period of almost entirely positive climatological anomaly in February and March can be regarded as an extreme period in a persistence sense. The (in re-analysis data) observed $\tilde{AO}$ index drops slightly below climatology at the end of March (not shown) and stays anomalously low for the following weeks, consistent with a drop in AO index towards climatology (see Lawrence et al., 2020). This drop is potentially correlated to the substantial weakening of the polar vortex in mid March.

Figures 1c+d and e+f further emphasize a potential coupling between the stratosphere and troposphere beyond a correlation of ensemble mean behavior. Panel 1c shows how the member distribution of $\tilde{AO}$ on February 9th changes from a generally broad distribution spanning the climatological mean to a narrow distribution spanning the observed value as the initialization date approaches February 9th (and thus the SSD with subsequent vortex strengthening is captured entirely deterministically). However, the predicted U_{10}^{60} index changes rather rapidly with lead time (compared to the typical stratospheric time scale of several weeks). In particular the ensemble initialized on January 25th shows a large fraction of members that predict a weak vortex. This is consistent with the introduced prediction of systematic SSDs (and for some members associated wave reflection events) within the ensembles around the corresponding lead times.

Figures 1e+f show the evolution of $\tilde{AO}$ and U_{10}^{60} indices during the later stages of the ensemble simulations (averaged over March 1st to 8th), when the forecast is well beyond the deterministic period and the system has only limited initial conditions memory remaining. Note that pre-SSD and SSD-onset ensembles (i.e., initialized before about Jan. 30th) correspond to rather broad distributions of the stratospheric index that show a significant amount of members with weak polar vortices (e.g. $U_{10}^{60} < 0$ m/s), while later initializations correspond to much narrower distributions centered around a rather strong polar vortex (large U_{10}^{60}). The member distributions of the tropospheric index also show pre-SSD and SSD-onset ensembles to include a substantial fraction of members with weak tropospheric flow (e.g., $\tilde{AO} < 0$ m), while post-SSD ensembles correspond to relatively strong tropospheric flows only. This is further discussed in Section 6.

3.3 Regional signatures of tropospheric and stratospheric circulation extremes

As previously discussed by Lawrence et al. (2020) and Lee et al. (2020), the weather conditions at the surface showed very anomalous characteristics from early February onwards, particularly over the North Atlantic and Eurasia. Figure 2 provides an overview of the observed surface conditions and the structure of the polar vortex and the mid-latitude jet in mid/late February. This period is characterized by an extremely large positive $\tilde{AO}$ index peaking on February 21st (cf. Figure 1). Hence we

find strong mean sea level pressure anomalies over the polar cap, and to a lesser extent over the North Atlantic and North Pacific (Fig. 2a). The associated strong meridional pressure gradients are consistent with enhanced westerlies (cf. Fig. 2d). The strong westerlies were further associated with a large number of intense storms traveling over the North Atlantic towards the British Isles and Western Europe, whose impact on precipitation is clearly visible in a south-west/north-east rainband across the North Atlantic towards the British Isles and western Europe (Fig. 2b), leading locally to wet and windy conditions during this period. Figure 2c displays a large warm anomaly of 2m temperature which covers almost entire Eurasia at higher latitudes. This positive anomaly, which is intense and persistent, develops in mid February and lasts until March (see Fig. **??** in the Supplement). We further discuss these regional anomalies and their representation in numerical model simulations in Section 5.3.
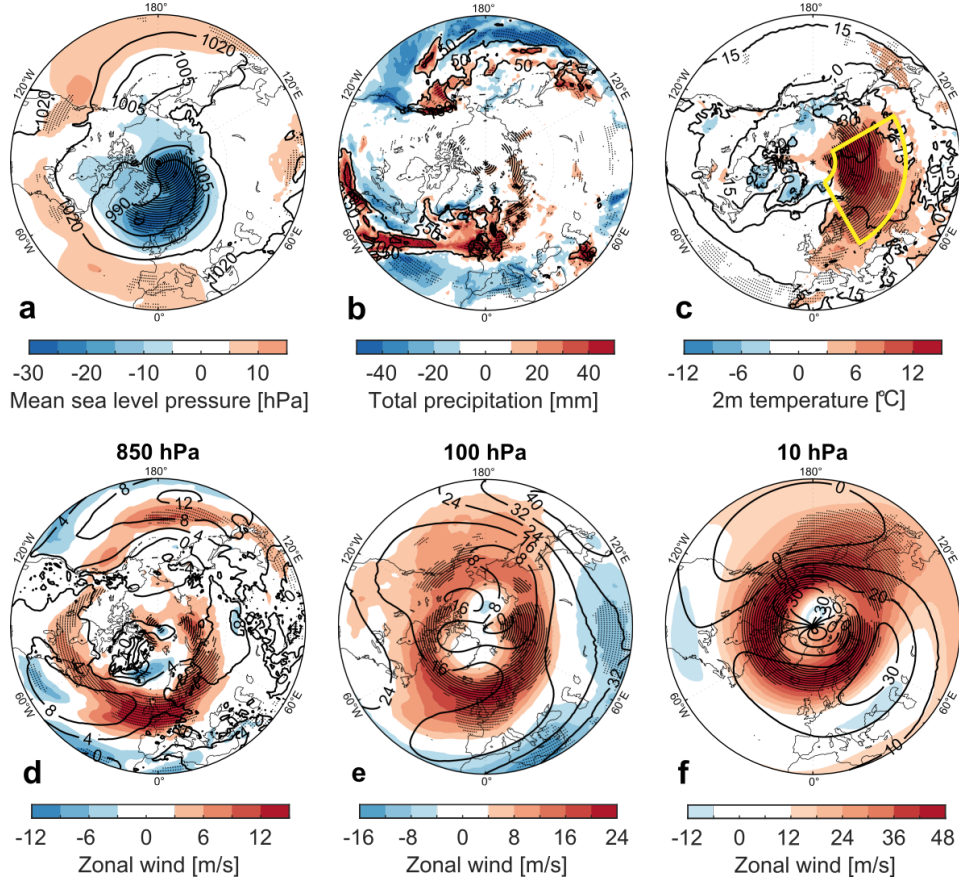

Figure 2. The full re-analysis fields (contours) and climatological anomalies (shading) of mean sea level pressure (a), total precipitation (b), 2m temperature (c) and zonal wind at 850hPa (d), 100hPa (e) and 10hPa (f) averaged over February 15th to 25th. Values above the 95th and below the 5th climatological percentile are dotted. The outline in panel c visualizes the region used for clustering in Figure 12.

Next, we analyze in more detail the structure of the polar vortex and the mid-latitude jet during February. Figures 2d to f show the zonal wind anomalies in the troposphere and the lower and upper stratosphere, respectively. After the SSD event in early February, the polar vortex recovered quickly to an extremely strong state. Indeed, the polar night jet is highly symmetric and zonal wind speeds exceed the 95th percentile of the climatological distribution over most of the polar cap (Fig. 2f).

In the lower stratosphere (100hPa), the jet over the North Atlantic and Eurasia is significantly enhanced compared to the beginning of February. There are negative zonal wind anomalies over the Mediterranean and Mongolia, reducing the gradients between high and low latitudes. Many of the characteristics of the 10hPa and 850hPa winds are also captured at 100hPa, indicating a strat-trop coupling. In the lower troposphere, the eddy-driven jet extends north-westward to western Europe and the North Sea, resulting in wet and stormy weather in these regions (cf. Fig. **??**b). Over the Pacific sector, the jet stream is slightly shifted northward compared to climatology, but is less anomalous than the Atlantic jet stream. The evolution of both tropospheric and stratospheric fields from January to March (not shown) provides evidence that the anomalous signals develop during or after the recovery from SSD.

4 Wave reflection event in early February and its impact on the stratospheric circulation

In Section 3.2 we find indications that the evolution of the stratospheric polar vortex during early February 2020 was dominated by a SSD, followed by a pronounced recovery of the polar vortex with persistent and extreme positive stratospheric circulation anomalies that lasted for several weeks. In this section we show that this recovery of the polar vortex was associated with a wave reflection event (WRE). Recall that both types of events, SSDs and WREs, are characterized by anomalous (planetary) wave-interactions between the troposphere and stratosphere and are therefore in particular associated with anomalous vertical wave activity flux F_p in the lower stratosphere.

The evolution of F_p during early 2020 is shown in Fig. 3 for zonal planetary wave numbers 1 and 2. Wave reflection events can be identified as periods of persistent downward wave activity fluxes preceded by corresponding upward fluxes. We find a clear indication for a WRE in both wave numbers in early February, with initially pronounced upward wave fluxes turning downward around February 10th for wave number 1 and around February 5th for wave number 2. In both instances, the subsequent downward propagating waves reach tropospheric pressure levels (below ca. 200hPa). Associated with these downward fluxes we find strongly enhanced (horizontal) momentum fluxes (not shown) in the troposphere during the wave reflection event around Feb. 10th, consistent with the observed strengthening of the mid-latitude jet and corresponding extreme values in $\tilde{AO}$ (see Fig. 1). Figure 3 further shows the period of enhanced polar vortex strength following the WRE (late February and early March) to be generally characterized by relatively weak tropospheric wave forcing and consequently overall weak upward wave fluxes in the (lower) stratosphere, in particular for wave number 1.

The dynamics associated with the WRE in early February, to some extent, can be understood in terms of linear wave theory and wave-mean flow interaction near reflective surfaces. Reflective surfaces are typically associated with regions of negative refractive index (see, e.g., Albers and Birner (2014) for details on how to compute the refractive index), which generally prohibit Rossby wave propagation and hence allow for the reflection of these waves. As shown in Fig 4, the refractive index exhibited an anomalous structure in both early and mid/late February compared to periods with a vortex more similar to climatology (e.g., as in Feb 2019, shown in Fig. 4a for comparison), including pronounced regions of negative refractive index in the mid-latitude and polar stratosphere. In particular, we find two regions in early February at around 40-50°N/100-5hPa and around 60-80°N/above 20hPa, acting as meridional and upper boundaries of a vertically confined wave guide at about 60°N. This wave guide further intensifies in mid/late February, making the stratospheric state even more favorable for WREs and potentially helping to maintain the strengthened polar vortex.
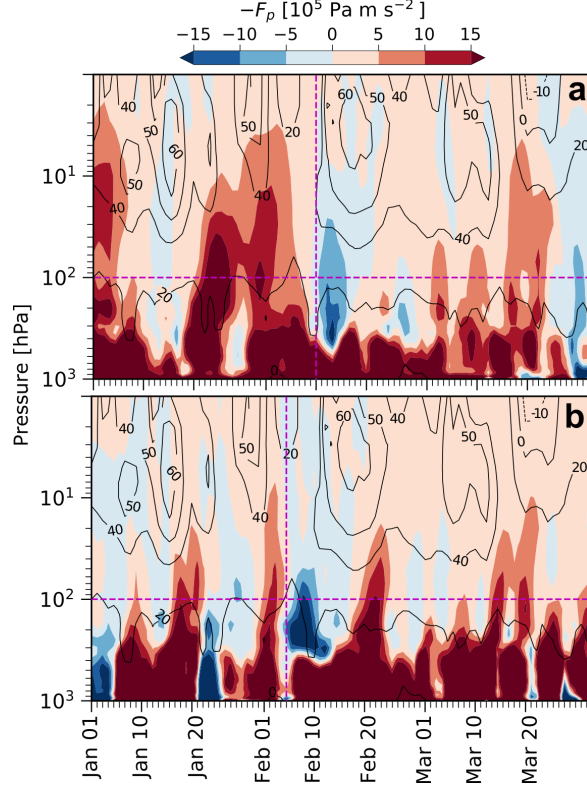
Figure 3. Decomposition into wave numbers 1 (a) and 2 (b) of the vertical component of the negative wave activity flux $(-F_p)$ from re-analysis averaged over $45°$ to $75°$ North. Positive values $(-F_p > 0$, shown in red) correspond to upward wave propagation, while negative values $(-F_p < 0$, shown in blue) correspond to downward wave propagation. Contours depict the zonal mean zonal wind at $60°$ North. Horizontal dashed lines indicate the 100 hPa level, vertical dashed lines mark the time of sign reversal in F_p at 100hPa (Feb. 10th and 5th, respectively).
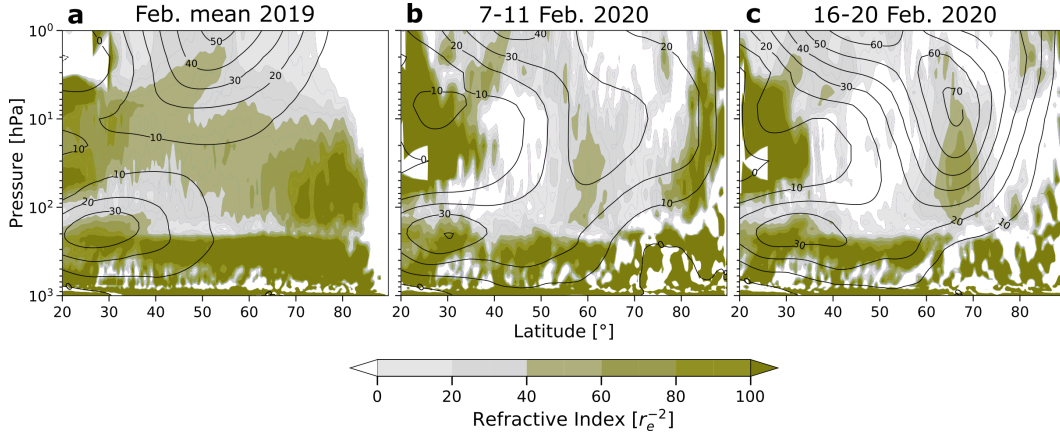

Figure 4. Zonal mean (quasi-geostrophic) refractive index n^2 (shading) and zonal mean zonal wind (contours) in re-analysis data, averaged over different time periods. The refractive index has been calculated for stationary waves with zonal wave number 1 and zero intrinsic phase speed and is non-dimensionalized with the Earth's radius of r_e.
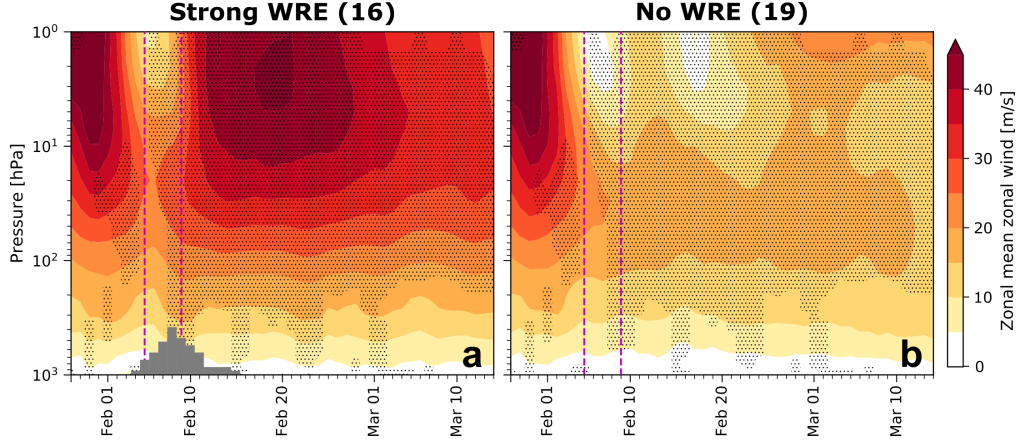
Figure 5. Evolution of zonal mean zonal wind at $60°$ North for clusters with strong wave reflection during the period February 5th to 17th and without (persistent) wave reflection in a lagged SSD-onset ensemble (initialized Jan. 25th and 28th). The stippling indicates significant differences (to 95% confidence) in cluster means and numbers in the panel headings represent cluster sizes. Vertical dashed lines mark the times of observed WREs in wave numbers 1 and 2. The histogram at the bottom of panel a indicates the number of WRE occurrences (wave numbers 1 and 2 counted individually) in the strong WRE cluster, with a peak of 13 events on Feb. 8th.

In order to investigate the effect of the mid February wave reflection on the stratospheric zonal circulation, the members of SSD-onset ensembles (initialized January 25th and 28th) are clustered based on whether they experience a strong and persistent WRE in early/mid February or do not show any signs of a WRE. Here we define (strong) WREs as periods with downward wave propagation ($-F_p < 0$ at 100hPa) in both wave number 1 and 2 for at least 4 consecutive days[1] between February 5th and 17th. In contrast we classify a member as not experiencing a WRE in early/mid February if all coherent periods of downward wave propagation (in wave number 1 or 2) last for less than 3 consecutive days. We further define a cluster of (remaining) members with weak WREs, where the length of periods of downward flux exceed 2 consecutive days in at least one wave number, but does not exceed 4 consecutive in both. Note that the observed wave fluxes satisfy the condition for a strong WRE (Fig. 3) in early February.

The zonal mean zonal wind at $60°$ North is depicted in Figure 5 for clusters with strong WRE and no WRE. While both clusters show indications for a SSD experienced at the beginning of February, the ensemble members with strong wave reflection display a substantial recovery in the stratosphere thereafter, with a similarly strong and persistent stratospheric polar vortex as seen in the re-analysis data (e.g., wind speeds exceeding 40m/s at 10hPa in late February; see Fig. 1). A much weaker stratospheric polar vortex is seen for the ensemble members with no WRE, suggesting a strong coupling of (lower) stratospheric wave fluxes to the polar vortex strength.

The substantial differences in polar vortex strength between the clusters are also visible in Fig. 6a, which depicts the stratospheric response relative to the onset of the

---

[1] The periods do not have to occur simultaneously for the both wave numbers within the selection period, consistent with the observed WRE in Fig. 3.
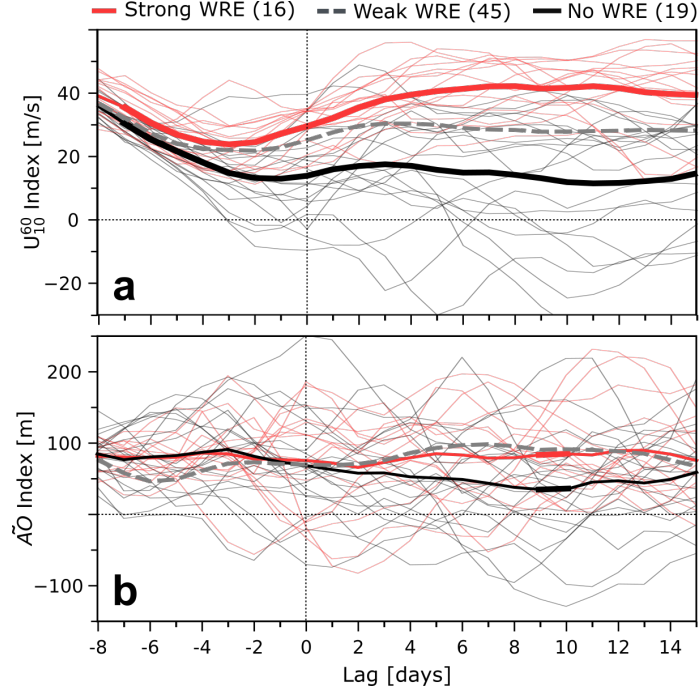
Figure 6. Lagged composite evolution of U_{10}^{60} and $\tilde{AO}$ indices for clusters with strong wave reflection and without reflection, respectively, for a lagged SSD-onset ensemble (initialized Jan. 25th and 28th). The event day is defined as the first day of sign-change in either the wave number 1 or 2 component of F_p at 100 hPa for the wave reflection cluster, and as February 10th for the no wave reflection cluster (the day of sign-change in the wave number 1 component of F_p at 100 hPa in re-analysis data). The cluster mean is shown for all clusters, individual members are also shown for the strong and no WRE clusters. Thick line segments indicate a statistically significant difference in means between clusters with strong and no WRE (to 95% confidence). Numbers in the legend indicate cluster sizes.

wave reflection (time of first sign reversal in F_p in either wave number). It can further be seen that none of the members in the strong WRE cluster shows signs of a SSW (in terms of $U_{10}^{60} < 0$) and the inter-member variability in this cluster is relatively small. In contrast, members in the cluster without WRE span a large range of vortex strengths following the event, including several members with pronounced SSW. Note here that the weak WRE cluster mean also shows a relatively weak recovery in polar vortex strength following the SSD in early February, and consequently moderate stratospheric zonal wind speeds in the following weeks compared to the clusters with strong WRE and no WRE.

While significant differences in the cluster means are visible in the stratosphere, the tropospheric signal is less robust. We find a systematic difference in the $\tilde{AO}$ values between the cluster with strong WRE and the cluster without WRE for about two weeks following the event. This difference in $\tilde{AO}$ appears to be mostly associated with the weakening of the tropospheric circulation following the weakening of the vortex due to SSWs, rather than the strengthening of it due to WREs. Correspondingly, we do not find any difference in $\tilde{AO}$ between the clusters with strong and weak WREs. Overall, these results indicate that the mid February WRE in the stratosphere on its own did not exhibit a direct downward influence on the tropospheric zonal mean circulation. We address this once more in Section 7.

5 Downward coupling following the wave reflection event in early February

5.1 Sensitivity of predictability and coupling on initialization period

As shown in Section 3, February 2020 was characterized by two periods of extreme $\tilde{AO}$ index (peaking February 9th and 21st) as well as an abrupt increase of the polar vortex strength following a SSD in early February, leading to extremely elevated values of U_{10}^{60} that persisted for over a month. We further show in Section 4 that this strengthening of the polar vortex in early February was associated with a WRE. March 2020, on the other hand, was mostly characterized by persistently positive values of both $\tilde{AO}$ and U_{10}^{60}. This section aims to analyze the coupling between the anomalous tropospheric ($\tilde{AO}$) and stratospheric (U_{10}^{60}) development in the period following the WRE. In particular we analyze whether the tropospheric extremes observed in February were more likely to occur due to the WRE directly, or secondly due to the anomalously strong and persistent polar vortex.

First, we examine whether the tropospheric $\tilde{AO}$ anomalies in mid/late February are linked to the polar vortex strength in general. To do so, we consider a very large ensemble of 320 members utilizing 8 initialization dates between late January and early February, hence including ensembles from all three groups (pre-SSD, SSD-onset and post-SSD) defined in Section 3. In Figure 7 ensemble members have been separated into two clusters based on whether they do or do not show extreme values in $\tilde{AO}$ in mid/late February, where extremes are defined as $\tilde{AO}$ exceeding 140m for at least one day between February 15th and 25th (the threshold of 140m roughly corresponds to the 85th-percentile of the member distribution at any given time step; the time period in mid/late February corresponds to a period with extremely increased observed $\tilde{AO}$ and lies sufficiently far outside of the deterministic period). The two clusters not only show significantly different $\tilde{AO}$ values during mid/late February (by definition), but also several days before and after this period (see Figs. 8b, d). In particular we find significantly increased $\tilde{AO}$ around February 9th, i.e., during the time of the first observed $\tilde{AO}$ extreme and the stratospheric wave reflection event (note that for some of the ensembles included in Figure 7 early February lies within the deterministic period, hence cluster means should be interpreted with care near the beginning of the shown period).
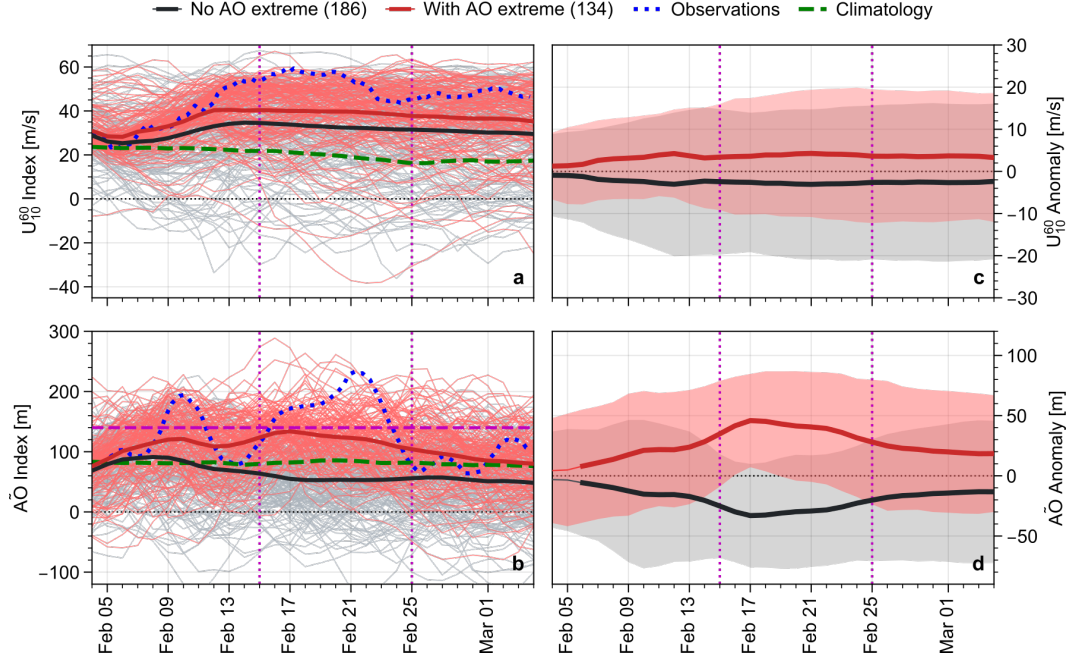
Figure 7. Evolution of U_{10}^{60} (top) and $\tilde{AO}$ (bottom) indices for a lagged ensemble with 8 initialization dates between January 16th and February 4th. Members are clustered based on whether $\tilde{AO}$ does or does not exceed 140m for at least one day between February 15th and 25th with cluster sizes indicated with bracketed numbers. Subplots a and b show full indices, thin lines represent individual members and thick lines represent the respective cluster means. Subplots c and d show the deviation from the ensemble mean, where lines represent the cluster means, shading indicates cluster variability in terms of one standard deviation and thick line-segments indicate periods where the difference in cluster means is statistically significant (to 95% confidence). The dotted and dashed lines show re-analysis data for 2020 and the climatology, respectively. Vertical dotted lines indicate the time period for the clustering criterion, the horizontal dashed line in subplot b indicates the threshold of 140m. Bracketed numbers in the legend indicate the number of members in the respective cluster (out of 320 in total).

Figures 7a and c also show significant differences between the two clusters in terms of the U_{10}^{60} index. Since the clusters have been formed purely based on the tropospheric evolution (with and without $\tilde{AO}$ extreme in mid/late Feb.), any stratospheric differences suggest a coupling of the troposphere and stratosphere. The stratospheric signal within the lagged ensemble shown in Fig. 7 consists of a relatively weak (about 8m/s) but very persistent positive anomaly in terms of U_{10}^{60} cluster difference. Note in particular that within the cluster with $\tilde{AO}$ extreme in mid February, only a very small fraction of members experiences SSWs over the course of the integration.

After having established the link of stratospheric anomalies and tropospheric extremes in a general sense by averaging over all available ensembles in Fig. 7, we now turn to the question how the coupling of the strat-trop flow evolved over the period of interest. Therefore, we analyze clusters differences in $\tilde{AO}$ and U_{10}^{60} for each individual ensemble with varying initialization date (Fig. 8). In particular, the evolution of U_{10}^{60} is consistent with the grouping of these individual ensembles into the three groups (pre-SSD, SSD-onset and post-SSD) defined in Section 3. Note that although the variability within individual ensembles can be very large, circulation anomalies evolve consistently within the respective groups. For pre-SSD ensembles the cluster with $\tilde{AO}$ extreme shows a relative strengthening of the stratospheric circulation (positive difference in U_{10}^{60} cluster mean) in early February compared to the cluster without $\tilde{AO}$ extreme. This circulation anomaly persists well into mid March. In SSD-onset ensembles the positive mean U_{10}^{60} cluster difference is limited to about a week prior to the $\tilde{AO}$ extreme around Feb 21, i.e. the period of strong vortex recovery in the observations. This is consistent with essentially all SSD-onset ensemble members predicting the SSD, but only part of them the recovery. Post-SSD ensembles, on the other hand, do not show any significant difference in U_{10}^{60} between the clusters throughout the simulation. The lack of difference is due to the fact that essentially all ensemble members simulate a strong and persistent polar vortex (see Fig. 1). Thus, SSD-onset ensemble members are most suitable to study the coupling between stratospheric and tropospheric flow for the specific evolution in February 2020, and they will be further analyzed in Section 5.2.

The distinctly different stratospheric evolution of ensembles seen in Figures 1 and 8 is consistent with a change in predictive skill for the clustering period with advancing initialization date. As discussed in Section 4, the stratospheric dynamics in early 2020 was characterized by a SSD around Feb. 6th induced by upwards propagating planetary waves, followed by a stratospheric WRE leading to an anomalously strong polar vortex for several weeks. Depending on initialization date, a fraction of ensemble members in a given ensemble is able to predict either the entire sequence of stratospheric events (SSD followed by WRE and persistently strengthened vortex) or only certain aspects of this sequence. This behavior is summarized in Fig. 9, showing an increased probability for a weak polar vortex to occur in early February for initialization dates in the SSD-onset period (around January 25th). In contrast, ensembles initialized in the post-SSD period predict a generally strong polar vortex in early/mid February, consistent with an increased fraction of members that predict a WRE.

Fig. 9 also shows the chance for extreme events with increased $\tilde{AO}$ to occur on February 9th and 21st, i.e., the periods with observed extreme values. The likelihood for extreme $\tilde{AO}$ events on February 9th seems to increase during the post-SSD period (consistent with the initialization date approaching the time of the observed $\tilde{AO}$ extreme), although only after the chance for a WRE in early February increases. The chance for an $\tilde{AO}$ extreme on Feb. 21st is generally small (consistent with the long lead times) with ensembles initialized in the SSD-onset period appearing to show a low likelihood in particular, potentially indicating a downward influence of the generally weakened polar vortex characterizing these simulations. Further, the chance for an $\tilde{AO}$ extreme on Feb. 21st seems generally higher for post-SSD ensembles compared to
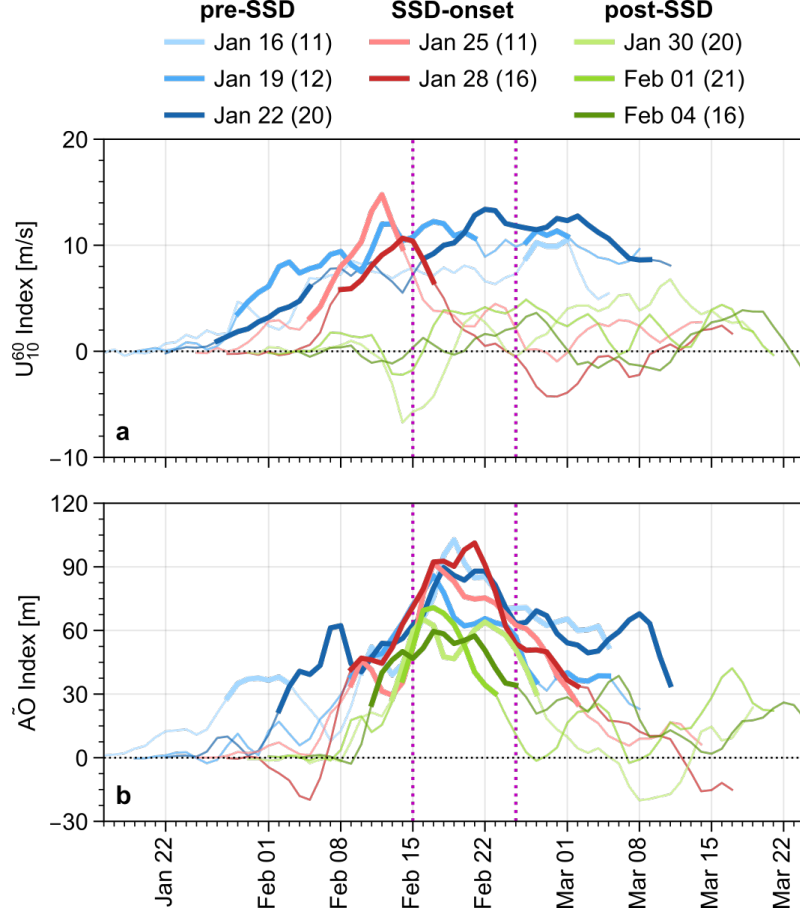
Figure 8. Difference in U_{10}^{60} (top) and $\tilde{AO}$ (bottom) indices between clusters from ensembles with varying initialization date between January 16th and February 4th. Clusters were formed based on whether $\tilde{AO}$ does or does not exceed 140m for at least one day between February 15th and 25th. Thick line-segments indicate periods for which the difference in cluster means is statistically significant (to 95% confidence). Vertical dotted lines indicate the time period for the selection criterion. Bracketed numbers in the legend indicate the number of ensemble members (out of 40 per ensemble) that exceed the threshold.
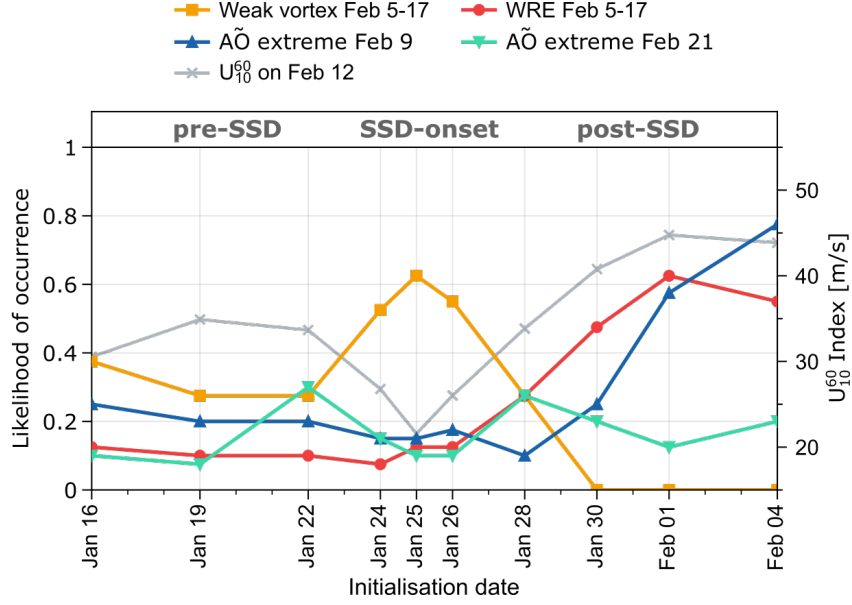
Figure 9. Different probabilities for atmospheric behavior (in terms of member fraction) as function of initialization date. Shown are the likelihood for extreme events with $\tilde{AO} > 140$m on Feb. 9th and 21st, for a weak polar vortex with $U_{10}^{60} < 15$m/s any time during Feb. 5th and 17th and for a WRE ($F_p > 0$ at 100hPa for at least 5 consecutive days in wave number components 1 and 2, see Section 4) to occur during Feb. 5th and 17th. Also shown is the ensemble mean of U_{10}^{60} (corresponding to the left y-axis, all other metrics correspond to the right y-axis).

pre-SSD ensembles, further suggesting a potential downward coupling of the strengthened polar vortices in post-SSD members (cf. Fig. 1). However, the likelihood for an extreme event on Feb. 21st in Figure 9 should be interpreted with caution since it is not fully clear to what extent any differences go beyond natural climatological variability or to what extent the increased likelihood for ('transition') initializations Jan. 22nd and 28th is significant.

5.2 Evolution in pre-SSD ensembles

Figures 1 and 7 suggest a downward influence of a strengthened stratospheric polar vortex (following a WRE) on the troposphere and hence a contribution to the formation of observed extreme $\tilde{AO}$ values in late February. The extent to which this potential downward influence is captured in ensemble members depends on the initialization time (pre-SSD, SSD-onset or post-SSD periods; see Figures 8 and 9). The pre-SSD ensembles do not capture the specific stratospheric evolutions of the SSD and WRE, but still show a statistical relation of the occurrence of tropospheric extremes to the stratospheric polar vortex strength. To analyze this potential coupling in more detail, Figures 10a and b show the clustering of a lagged pre-SSD ensemble (combining initializations from January 16th, 19th and 22nd) into clusters with and without $\tilde{AO}$ extreme in late February (also compare to Fig. 8). Despite being clustered purely based on the tropospheric $\tilde{AO}$ index, the clusters show a pronounced and significant difference in U_{10}^{60}.

However, if we perform the same clustering, but exclude all members that experience a weak polar vortex at any point in time (Figures 10b and d), the stratospheric difference between both clusters disappears and both clusters show relatively large
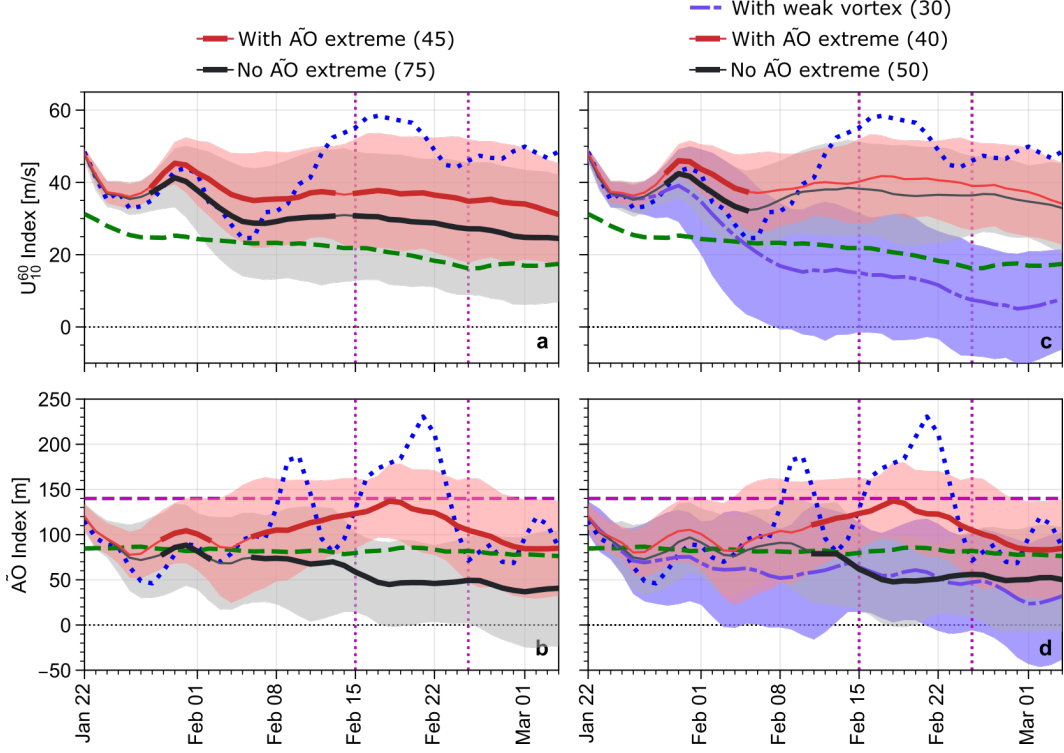
Figure 10. Cluster mean evolution in $\tilde{AO}$ and U_{10}^{60} indices for clusters with and without $\tilde{AO}$ extreme between February 15th and 25th (same criterion as used in Fig. 7) in a lagged pre-SSD ensemble (initialized January 16th, 19th and 22nd). The left column shows clusters based on all 120 ensemble members, right column shows clusters where 30 members that exhibit a weak polar vortex (defined as $U_{10}^{60} \leq 5$ m/s at any time during the 50 day integration) have been removed from the ensemble and clusters with/without $\tilde{AO}$ extreme have been formed based on the remaining 90 members.

values of U_{10}^{60} throughout February, while the tropospheric anomaly in late February remains almost unchanged. This behavior of the clusters suggests that the strat-trop coupling signal seen in Figures 10a and b (i.e., the correlation of stratospheric and tropospheric circulation anomalies) is (almost entirely) caused by members predicting a weak polar vortex rather than by members that predict a particularly strong vortex. Correspondingly, the downward influence of strong vortex events can to some extent be interpreted as a lack of downward influence from weak vortices in these members.

Further note that the cluster with $\tilde{AO}$ extreme has significantly increased $\tilde{AO}$ around February 9th (the period with the first observed $\tilde{AO}$ extreme and prior to the clustering period) compared to the cluster without $\tilde{AO}$ extreme if we do not exclude members with weak vortex (Fig. 10b). This cluster difference in $\tilde{AO}$ during early February seems to be reduced if weak vortices are excluded from the ensemble (Fig. 10d). If we interpret the difference between members in U_{10}^{60} in mid/late February as a proxy for the nature and strength of strat-trop wave coupling in early February (e.g., WRE vs. no WRE; see Figures 6 and 9) the correlation between vortex behavior and $\tilde{AO}$ index on Feb. 9th is potentially indication for strat-trop coupling via modification of wave-mean-interactions and therefore a direct downward influence of the WRE early February (or the corresponding lack of a strong SSD; further discussed in Section 7).

We find a similar tropospheric signal when clustering pre-SSD ensemble members based on the strength of their stratospheric zonal circulation in early February (Fig. **??** in the Supplement). Members with a stronger polar vortex in February correspond to a generally increased $\tilde{AO}$ index. However, clusters associated with a generally strong or moderate polar vortex strength do not show a clear difference in terms of $\tilde{AO}$ in mid/late February. This is consistent with Fig. 10 in the sense that it is rather the lack of the occurrence of a weak vortex (event) than a particular strong vortex that explains the tropospheric signal. As such, this indicates an asymmetry in the downward influence of strong versus weak polar vortex situations, which will be further discussed in Section 7.

5.3 Evolution in SSD-onset ensembles

The evolution of the stratospheric polar vortex and the downward coupling changes in some respect when considering a lagged ensemble initialized during the SSD-onset period (e.g., January 25th and 28th). These ensembles mostly capture the (weak) SSD in early February but not fully capture the WRE (as discussed earlier in this section and in Sections 3 and 4). The corresponding ensemble members therefore differ substantially in terms of if and how strong the polar vortex recovers following the SSD. Figure 11 shows three clusters of ensemble members clustered based on their mean U_{10}^{60} index between February 8th and 15th. All three clusters experience a pronounced drop in stratospheric circulation in early Feb, but the U_{10}^{60} index of the cluster with strong vortex quickly recovers and reaches values exceeding 45m/s in mid February, while the U_{10}^{60} index of the cluster with a weak vortex keeps dropping and stays persistently below 20m/s for February and early March. This behavior of the different clusters is consistent with the idea that some members capture the wave reflection event in mid February, while others do not. (Also note that the U_{10}^{60} evolution of the three clusters is similar to the cluster composite formed based on vertical wave fluxes seen in Fig. 6.)

The distinct evolution of the stratospheric circulation in the three clusters allows us to study its influence on the tropospheric zonal mean circulation (Fig. 11b), which does not show a significant difference between the clusters up to about Feb. 15th (the end of the period used for cluster definition and towards the end of the wave reflection event). However, during the second half of February and early March the clusters differ significantly in their $\tilde{AO}$ indices, with the strong vortex cluster showing a strengthened
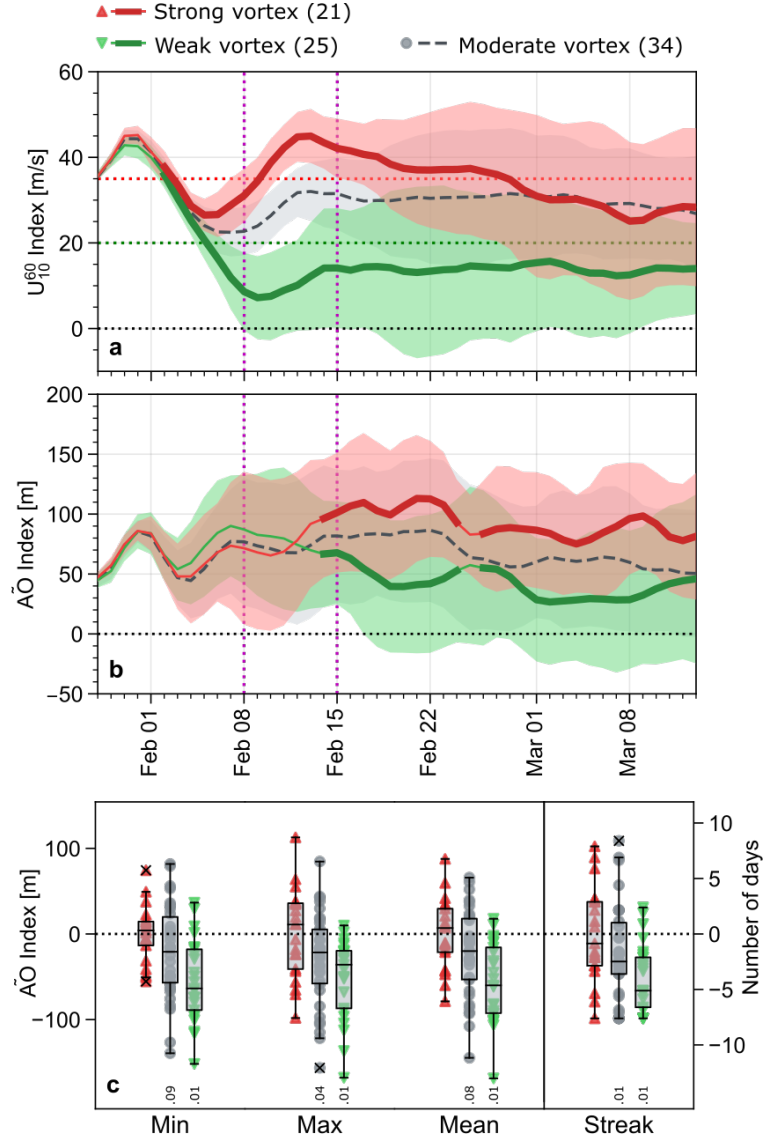
Figure 11. Evolution of U_{10}^{60} (a) and $\tilde{AO}$ (b) indices for a lagged SSD-onset ensemble (initialized on January 25th and 28th). Members are separated into 3 clusters based on whether their U_{10}^{60} index averaged between February 8th and 15th is above 35m/s (strong vortex), between 20m/s and 35m/s (moderate vortex) or below 20m/s (weak vortex). The curves show the cluster means, with thick line segments for the clusters with strong and weak vortex indicate statistical significant (to 95% confidence) differences between the corresponding cluster means. Vertical dashed lines indicate the time period used for the clustering and the horizontal dashed lines in subplot a indicate the thresholds of 20m/s and 35m/s and numbers in the legend represent cluster sizes. Panel c shows the normalized member distributions of the three clusters for the minimum, maximum and mean $\tilde{AO}$ between Feb. 15th and 25th, as well as the maximum number of consecutive days with positive $\tilde{AO}$ anomaly between Feb. 15th and Mar. 14th. All member distributions are normalized so zero corresponds to the mean of the cluster with strong vortex and numbers at the bottom show the p-values for significance tests assuming the mean of clusters with medium and weak vortex to be zero.

tropospheric circulation compared to the weak vortex cluster. Note that the cluster with moderate vortex also shows moderate values of $\tilde{AO}$, even during early March, when the clusters with strong and moderate vortex are almost identical in terms of mean U_{10}^{60} index. The correlation between the stratospheric and tropospheric jets is further emphasized in Fig. 11c, which shows the (normalized) member distributions of various characteristics of the $\tilde{AO}$ evolution of the three clusters defined based on U_{10}^{60}. We find a clear difference between the three clusters in all characteristics. In particular the cluster with strong vortex shows significantly increased values in persistence and extreme $\tilde{AO}$ metrics compared to the cluster with moderate vortex. Note here that the difference between members with strong and moderate stratospheric flows seems (in general) much less robust than the difference between weak and moderate flows (cf. Figures 6 and 10, and **??** in the Supplement). As further discussed in Section 7 the lack of a robust and clear coupling signal (in particular between situations with strong and moderate vortex) emphasizes the probabilistic nature of the coupling and indicates an asymmetric downward influence of positive and negative anomalies in polar vortex strength.
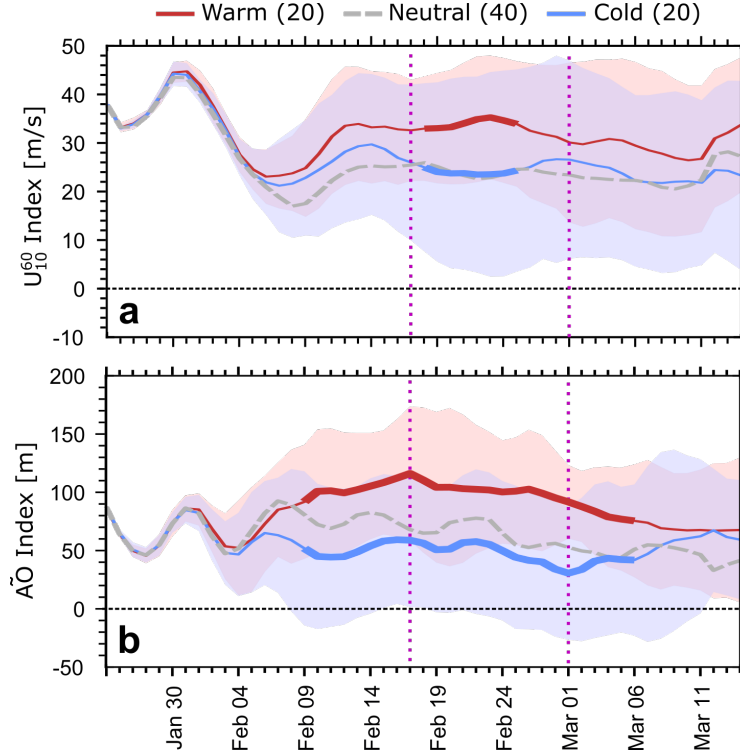

Figure 12. Evolution of U_{10}^{60} (a) and $\tilde{AO}$ (b) indices for a lagged SSD-onset ensemble (initialized on January 25th and 28th). Members are separated into 3 clusters based on whether their 2-meter temperature averaged over 30-120°E/50-80°N (see Fig. 2c) and Feb. 17th to March 1st lies below the 25th percentile (cold cluster), between the 25th and 75th percentile (neutral cluster) or exceeds the 75th percentile (warm cluster); corresponding cluster thresholds correspond to -1°C and +1.9°C. The curves show the cluster means, shadings show one standard deviation and thick line segments for the warm and cold clusters indicate statistically significant (to 95% confidence) differences between the corresponding cluster means. Vertical dashed lines indicate the time period used for the clustering and numbers in the legend represent cluster sizes.

As discussed in Section 3.3, the observed tropospheric conditions during the period with strengthened vortex (late Feb. and early Mar.) were characterized by a persistent warm anomaly in surface temperature over the Eurasian continent. We find a corresponding tropospheric regional temperature signal in certain members of the SSD-onset ensembles. Figure 12 shows that members with increased 2-meter temperature over Eurasia in late February are generally associated with increased values of $\tilde{AO}$ and correspondingly a stronger mid-latitude jet than members with lower temperature. In addition, warm cluster members show increased values in U_{10}^{60}. The correlation of the surface temperature signal and the polar vortex strength further highlights the strat-trop interaction in early 2020. It should be noted that the difference in $\tilde{AO}$ index between warm and cold clusters seems to develop slightly before the difference in temperature develops (Fig. 12 and Fig. **??** in the Supplement), and both anomalies develop about a week prior to the start of the clustering period.

The clustering approaches used in this section strongly rely on the underlying ensemble to sufficiently sample a range of possible evolutions of the stratospheric and/or tropospheric circulation. A similar analysis of post-SSD ensembles is therefore much less meaningful, as corresponding members do systematically capture the WRE in early February and therefore produce a generally strengthened polar vortex in the following weeks.

6 Influence of the polar vortex strength on tropospheric extreme events

We showed that the particular stratospheric evolution in 2020 with the SSD followed by a WRE and a persistently strong polar vortex potentially contributed to persistent positive $\tilde{AO}$ anomalies in February and March (based on the SSD-onset ensembles, Sec. 5.3), but also found a more general relation between the stratospheric polar vortex strength and $\tilde{AO}$ extremes (based on pre-SSD ensembles, Sec. 5.2). This sections aims to analyze this general stratospheric influence on $\tilde{AO}$ extremes in situations with a persistently strong or weak polar vortex, irrespective of the specific evolution in February 2020. In particular, we analyze whether persistent stratospheric anomalies impact the persistence, strength and likelihood of tropospheric extreme events. We first perform the analysis for a lagged ensemble initialized in mid February, hence covering the time period of interest in March with persistently increased observed $\tilde{AO}$ index. We later show that our results are insensitive to the initialization date.

To obtain a general idea of how the tropospheric circulation is influenced by persistent stratospheric anomalies (and vice versa), we analyze three sets of two distinct clusters in terms of the difference in their cluster means. In each set members of a lagged ensemble are separated into two clusters based on whether they do or do not correspond to a persistently strong zonal circulation in either the troposphere, the lower stratosphere or the middle/upper stratosphere, corresponding to the three indices $\tilde{AO}$, U_{100}^{60} (defined as zonal mean zonal wind at 60° North and 100hPa) and U_{10}^{60}, respectively. The differences in cluster means of the three sets are shown in Fig. 13. We find that even though persistent positive anomalies in the middle/upper stratosphere show a significant response in the lower stratosphere, they do not appear to induce any significant signal in the troposphere. In a similar way persistent circulation anomalies in the troposphere lead to a strengthening of the lower stratospheric flow, but not the middle/upper stratospheric flow. Consistently, anomalies within the lower stratosphere show a significant signal in both the troposphere and the middle/upper stratosphere.

Figure 13 suggests there is no direct influence of persistent (middle/upper) stratospheric flow anomalies on the tropospheric circulation in terms of ensemble means. However, the stratospheric state can still have an effect on individual tropospheric extreme events, which is not apparent in averages over many events as they, for example, might happen at different times or correspond to a combination of positive or
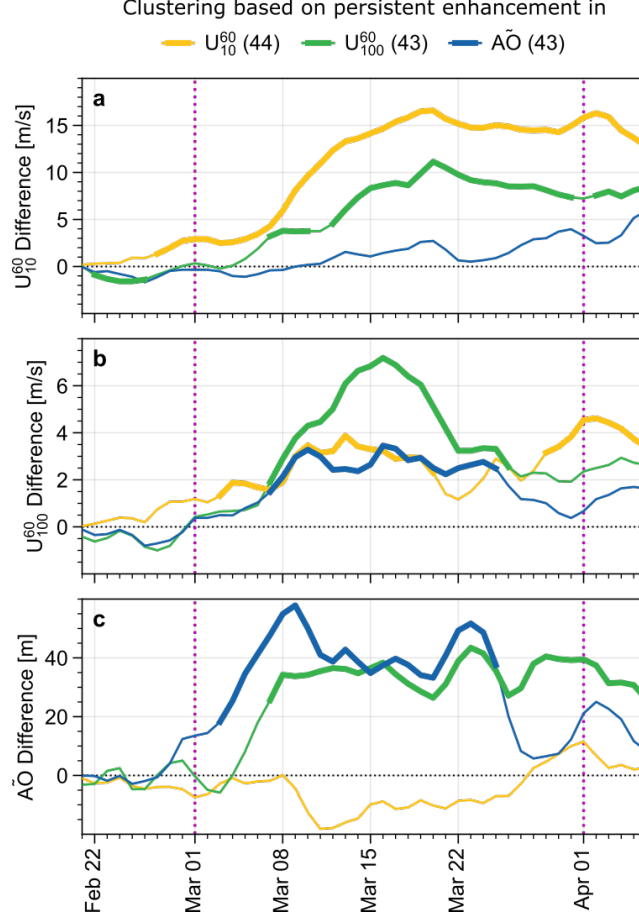
Figure 13. Evolution of the difference in U_{10}^{60}, U_{100}^{60} and $\tilde{AO}$ cluster means in a lagged post-SSD ensemble (initialized February 18th and 21st) for three sets of two clusters (represented by different line colors). The clusters are formed based on one of three criteria, distinguishing members that do or do not show persistently enhanced values in either of the three plotted indices. The three different sets of clusters are based on members that have $U_{10}^{60} \geq 35$m/s for at least 20 consecutive days, $U_{100}^{60} \geq 20$m/s for at least 20 consecutive days or $\tilde{AO} \geq 80$m for at least 10 consecutive days within the time period March 1st to April 1st. Thick line segments indicate periods for which the cluster means of each set are significantly different in the respective index (to 95% confidence) and numbers in brackets show how many members (out of 80) satisfy the respective criterion.

negative tropospheric flow anomalies. To study the influence of stratospheric conditions on certain tropospheric extremes, Figure 14a shows a composite of tropospheric events with substantially increased $\tilde{AO}$ index during March. Consistent with Fig. 13, we do not find a significant signal in the middle/upper stratosphere correlated to tropospheric extremes, but the lower stratosphere shows slightly elevated values in U_{100}^{60} for roughly a week following the tropospheric extreme (indicative of upward coupling). Figure 13 indicates that flow anomalies in the lower stratosphere can couple to anomalies in both troposphere and middle/upper stratosphere. Any positive downward influence of a strengthened middle/upper stratospheric circulation on the troposphere should therefore be associated with strong lower stratospheric circulation anomalies, while the downward influence of weakened middle/upper stratospheric circulation on positive tropospheric flow anomalies should be characterized by weakened (or even negative) lower stratospheric anomalies. Correspondingly, if we divide the group of events in Fig. 14a into two clusters based on whether the lower stratosphere corresponds to a generally positive or negative ensemble anomaly following the event (Figures 14b and c), we find both clusters to show substantial differences in terms of the evolution of their stratospheric and tropospheric circulation.

In particular we find the positive $\tilde{AO}$ extremes to be significantly stronger and more persistent in the case with positive stratospheric flow anomalies. Note that the two clusters already differ significantly in terms of middle/upper stratospheric circulation more than a week prior to the event date and thus before cluster differences in $\tilde{AO}$ develop. This difference in U_{10}^{60} prior to the start of the event can be interpreted in the way that such tropospheric (extreme) events do occur in situations with both strong or weak polar vortex simply due to internal tropospheric variability, but that the probability distribution and evolution of these events can be significantly modified by the state of the stratosphere at the time of their occurrence. Further, Figure 14 indicates that stratospheric anomalies, in return, respond to the changes in tropospheric circulation, leading to a mutual amplification of the flow anomalies in both layers.

The general modification of tropospheric events by the stratospheric state is a rather robust result within our set of ensembles. Table 1 summarizes the modification of positive and negative $\tilde{AO}$ extreme events by the stratospheric state as an average over multiple ensembles[2]. The event persistence is essentially estimated in terms of the e-folding time of a Gaussian profile fitted to the $\tilde{AO}$ anomaly time series at positive lags of each individual event in the respective group[3].

In situations with an anomalously strong lower stratospheric flow (positive U_{100}^{60} anomaly) we find positive $\tilde{AO}$ extreme events (strong tropospheric circulation) to be stronger and longer-lasting. Furthermore, these positive $\tilde{AO}$ anomaly events are more frequent than in situations with weak polar vortex, which is consistent with an increase in event magnitude due to a shift in probability distribution when considering the definition of events based on a fixed $\tilde{AO}$ threshold. In return, when considering periods with an anomalously low $\tilde{AO}$ index (weak tropospheric circulation), we find that a strengthened stratospheric polar vortex has a suppressing influence, with the corresponding (negative) tropospheric events occurring less often, being weaker and lasting considerably shorter.

---

[2] Corresponding values for individual ensembles can be found in the Supplement as Figure S4.

[3] To be precise the persistence time scale is given by the parameter τ calculated via a least-squares fit of the regression model $\sqrt{-\ln \tilde{AO}_a(\lambda)/\tilde{AO}_a(0)} = \lambda/\tau + \epsilon(\lambda)$, where λ represents time lag, $\epsilon(\lambda)$ is the regression error function, $\tilde{AO}_a(\lambda)$ is the ensemble anomaly of the $\tilde{AO}$ index and the regression is performed for all positive lags $\lambda \geq 0$ for which $\tilde{AO}(\lambda) > 0$. Other ways to estimate the event persistence gave qualitatively similar results.
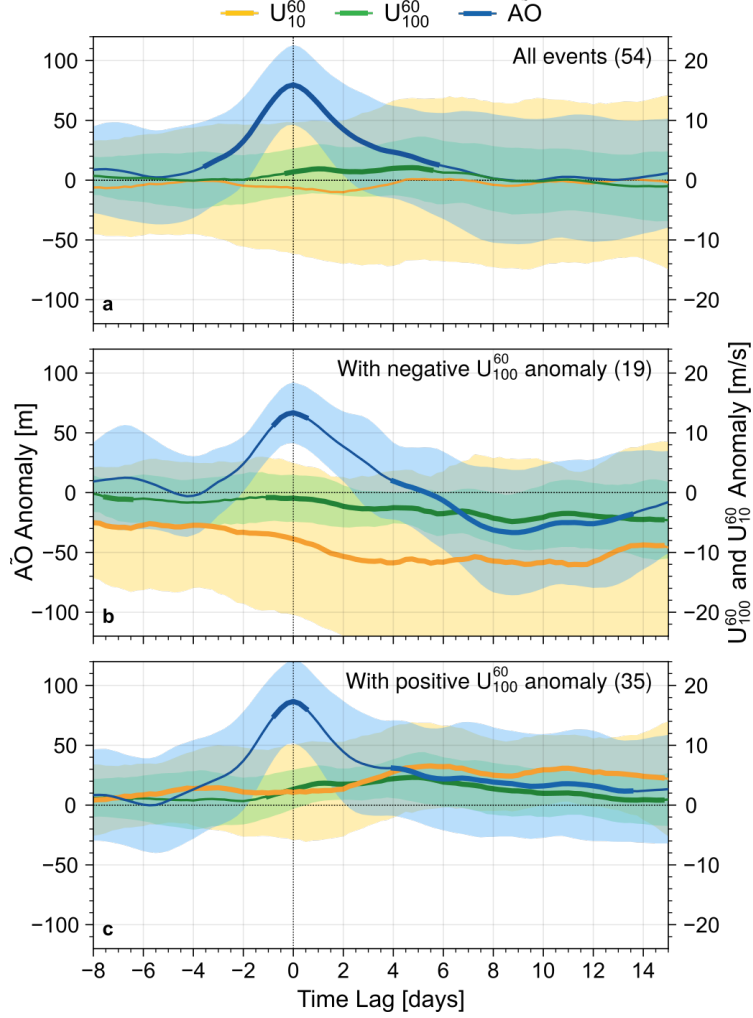
Figure 14. Composite evolution of $\tilde{AO}$, U_{100}^{60} and U_{10}^{60} anomalies for 54 extreme $\tilde{AO}$ events in a lagged ensemble initialized on February 18th and 21st. Events are defined as global maximum in $\tilde{AO}$ anomaly between March 9th and 23rd under the condition that $\tilde{AO} \geq 120$m at the event date. Panel a shows all 54 events, while panels b and c show the events separated into two complementary clusters (19 and 35 members) that either have a positive or a negative U_{100}^{60} anomaly (when averaged over the 10 days following the event). The thick line-segments in panel a indicate periods where the composite mean of a respective index-anomaly is significantly different from zero, and in panels b and c periods with significant difference between the two cluster means (to 95% confidence in all cases).

Table 1. Different properties of $\tilde{AO}$ extreme events averaged over 18 ensembles (720 members in total) with (unevenly spaced) initialization dates between January 10h and March 12th. Positive/negative events are defined via the time of maximum/minimum $\tilde{AO}$ anomaly satisfying $\tilde{AO}$ ≥ 100m (positive) or $\tilde{AO} \leq 60$m (negative) any time between 15 and 38 days after initialization. Events have further been separated into two clusters with either positive or negative U_{100}^{60} ensemble anomaly (when averaged over the 10 days following the event). Shown are the number of events in the respective clusters, the cluster mean of e-folding decay time of the $\tilde{AO}$ anomaly for the events and the magnitude of the event in terms of $\tilde{AO}$ anomaly.

	All events	Positive U_{100}^{60} anomaly	Negative U_{100}^{60} anomaly
Positive $\tilde{AO}$ events			
Number of events	35	24	11
Persistence time [days]	2.6	4.3	1.9
Event magnitude [m]	80	86	65
Negative $\tilde{AO}$ events			
Number of events	36	15	21
Persistence time [days]	2.6	1.8	4.5
Event magnitude [m]	-76	-60	-87

Qualitatively, similar characteristics of strat-trop coupling as seen in Fig. 14 can be found when analyzing cluster composites based SSDs (not shown). In particular, our analysis shows no significant deviation from the ensemble mean in $\tilde{AO}$ when averaging over all SSD events occurring in a large lagged ensemble covering February and March 2020. However, we find a significant correlation with the tropospheric jet strength when distinguishing members that are associated with either anomalously strong or weak zonal circulations in the lower stratosphere during/after the SSD. Members associated with an overall strong (lower) stratospheric circulation exhibit a general strengthening of the tropospheric circulation for more than a week following the event. Consistently strong vortex members show a significant increase in magnitude and likelihood of (positive) $\tilde{AO}$ extremes (and a corresponding decrease in magnitude and likelihood of negative $\tilde{AO}$ extremes) compared to members with anomalously weak lower stratospheric circulation.

7 Discussion

We have shown that the stratospheric evolution in early 2020 was characterized by strong wave-mean flow interaction, in particular a SSD in early February followed by a WRE that led up to a strong persistent polar vortex. In how far this behavior is captured by ensemble predictions strongly depends on the initialization date, and correspondingly the stratospheric evolution was difficult to predict during certain times (see Fig. 1). This behavior is consistent with the 'bifurcation' of wave propagation properties at critical lines within the stratosphere (e.g., Perlwitz & Graf, 2001; Noguchi et al., 2016), e.g., during the SSD (and subsequent WRE) in early February. Such bifurcations can then lead to a range of extremely different potential evolutions of the stratospheric circulation, depending on the associated interaction of upward propagating waves with the zonal mean flow. As an example, ensembles initialized close to the onset of the SSD (around January 25th) capture the tropospheric wave forcing in early February and the beginning of the corresponding SSD deterministically. However, for certain members of these ensembles the upward wave flux can cause the

stratosphere to enter a 'reflecting state', where subsequent upward propagating waves are reflected (i.e., WREs occur), while other members enter (or stay in) an 'absorbing state', allowing subsequent waves to interact with the mean flow and further decelerate stratospheric winds. Consistently, SSD-onset ensembles predict a large likelihood for a weak vortex or even a SSW in early/mid February (Fig. 9). Note that WREs can in principle occur for a single planetary wave number, or for multiple wave numbers (as was the case in early 2020) and the evolution of the polar vortex strength is generally sensitive to the precise details of the reflection. Correspondingly, members within the SSD-onset ensembles show a spectrum of polar vortex evolutions, spanning the two extreme cases with either the upward wave flux being absorbed completely in the stratosphere (leading to a strong SSW) or a (quasi-simultaneous) reflection of all planetary wave components (leading to a strong strengthening of the vortex).

Ensembles initialized during the post-SSD period (after about January 30th) produce relatively accurate (deterministic) predictions of both the recovery of the polar vortex in mid February (including extremely strong and persistent stratospheric winds for the following weeks) and the tropospheric extreme around February 9th (Fig. 1 and 9). In contrast, pre-SSD and SSD-onset ensembles only capture these events probabilistically (e.g., Figures 5 and 9). We further find pre-SSD ensemble members that do not experience a SSW in February to have generally increased values in $\tilde{AO}$ around February 9th (Fig. 10). The timing of events and deterministic/probabilistic model behavior suggests a stratospheric contribution to the formation of the corresponding tropospheric extreme via planetary wave-driving. In particular the destructive interference of a downward propagating planetary wave with the tropospheric standing wave pattern following a WRE was suggested by various authors as (partially) responsible for creating tropospheric extremes in early 2020 (Lawrence et al., 2020; Lee et al., 2020). However, Figure 11 does not show any significant differences in tropospheric zonal mean zonal flow during the recovery period of the polar vortex (which is a proxy for the WRE) in early February. Anomalies in the tropospheric zonal mean flow develop only after substantial stratospheric flow anomalies are formed, suggesting no direct downward wave-coupling of stratosphere and troposphere in relation with the WRE, but a persistent downward influence of the subsequent state with a persistently strengthened polar vortex (e.g., Figure 6). In addition to zonal mean metrics, we also studied the (direct) response of the planetary wave components of the tropospheric geopotential height field (not shown), but could not identify a robust downward influence of the WRE.

The systematic correlation in predicted behavior between troposphere and stratosphere suggests a coupling between the two layers. In particular, the low predictability of the strong stratospheric winds in mid February suggests an influence of the troposphere via wave forcing (e.g., Figures 5 or 9). The persistently strong polar vortex in February and March, in turn, increased the likelihood of extreme tropospheric large-scale circulation that has been observed around February 21st. However, the strength and robustness of the corresponding downward influence of positive stratospheric circulation anomalies seems to strongly depend on the strength of the anomaly and potentially other factors related to the current state of the atmosphere. Figures 11 and 10a and b indicate a significant difference in tropospheric circulation strength ($\tilde{AO}$ index) in late February and early March between model realizations with a strong or weak stratosphere in various ensembles initialized mid to late January. This significant difference in $\tilde{AO}$ between strong and weak vortex members, however, seems to be largely attributed to the robust downward coupling of negative stratospheric zonal flow anomalies following weak vortex periods (or in extreme cases SSWs), as can be seen in Figures 10a and b. Consistently, the downward influence of strengthened polar vortices (or similarly tropospheric difference between ensemble members with strong and more moderate stratospheric circulations) seems to be less robust in these ensemble simulations (e.g., Fig. 11 and Fig. S2 in the Supplement).

We also find the changes in stratospheric zonal mean circulation to be connected (likely via the tropospheric zonal mean circulation) to regional surface temperature signals (cf. Figure 12). In particular ensembles members that experience a relatively strong polar vortex reproduce the warm anomaly over the Eurasian continent that has been observed in late February. However, although our ensemble analysis reveals a consistent coupling between stratospheric circulation anomalies and tropospheric response, individual members show a large variety of troposphere-stratosphere evolutions.

The lack of a robust tropospheric signal between strong and moderate stratospheric circulations in our ensemble simulations emphasizes the probabilistic nature of strat-trop coupling. Our results show a more robust downward influence of negative stratospheric flow anomalies than for positive anomalies. One potential reason for the asymmetric downward influence of positive or negative anomalies in polar vortex strength is the asymmetric effect of corresponding circulation anomalies on the vertical propagation of planetary Rossby waves. Following the simple linear criterion originally derived by Charney and Drazin (1961), waves can only propagate vertically in regions with zonal wind u satisfying the two inequalities $0 < u$ and $u < U_c$, where generally U_c is a function of the potential vorticity gradient, the meridional and zonal wave numbers and potentially other state characteristics. The condition $0 < u$ forms a distinct lower boundary of the propagation-permitting range of u and is (by definition) met following SSWs for at least part of the vortex. Hence, the potential for strat-trop wave interactions is in general severely limited during weak vortex periods. The upper limit for wave propagation ($u < U_c$), on the other hand, is not a distinct fixed value of the zonal wind, but is flow-dependent, and also dependent on the wave number. As such, it does not form such a clear upper boundary of the propagation permitting regime, which is consistent with a potentially asymmetric behavior of strat-trop coupling for weak versus strong polar vortices, as well as with the typically very large wind speeds required to substantially limit the vertical propagation of (planetary) waves into the stratosphere.

The regime of a strong polar vortex that largely inhibits planetary wave propagation can, for example, be seen in the Southern hemisphere during the respective cold season (Plumb, 1989). The Southern hemisphere polar vortex reaches typical wind speeds (at 60° South and 10hPa) of about 90m/s in August. These strong winds are consistent with typically low planetary wave amplitudes in the stratosphere and a resulting low variability of the vortex. In particular, Plumb (1989) finds a pronounced local minimum in planetary wave amplitudes in the stratosphere during July and August despite the lack of an equally pronounced minimum in the tropospheric wave forcing, indicating an increased opacity for vertically propagating waves due to the strong winds of the Southern hemispheric polar vortex. The polar vortex in the Northern hemisphere, in contrast, exhibits typical climatological wind speeds (at 60° North and 10hPa) of about 30m/s, consistent with a large variability and relatively large wave amplitudes in the stratosphere throughout the winter. Vortex conditions similar to the Southern hemisphere are almost never reached in the Northern hemisphere. The direct comparison between the hemispheres emphasizes the need for excessively large positive flow anomalies in the stratosphere in order to substantially change the permeability for upward propagating waves, while rather weak negative anomalies are often sufficient to produce a predominantly westward stratospheric flow (e.g., with $U_{10}^{60} < 0$) and thus completely alter the ability of wave-interaction between the troposphere and stratosphere. Such changes to the zonal flow in the (lower) stratosphere and the corresponding effects on strat-trop wave coupling and with it tropospheric wave-behavior can then potentially be interpreted as a modification of the upper boundary condition of the troposphere and hence as downward influence on the internal evolution of the tropospheric circulation.

The downward influence of the strengthened stratospheric polar vortex in late February and early March potentially contributed to the extremely strong tropospheric zonal circulation observed in late February and the persistently strengthened circulation in March (see Fig. 1)and associated surface anomalies (Fig. 2). As shown in Fig. 14, wind anomalies in the (lower) stratosphere can affect the likelihood, magnitude and persistence of tropospheric circulation extremes (in terms of $\tilde{AO}$ anomaly). However, this downward influence should be regarded as a modification of probabilities (e.g., via modification of the tropospheric upper boundary condition) rather than a direct forcing of tropospheric anomalies as it strongly relies on tropospheric internal variability and still requires tropospheric anomaly events to develop internally. Table 1 shows that strong stratospheric flow anomalies generally strengthen positive tropospheric events, but weaken negative events (and vice versa for weak stratospheric winds). This can be interpreted as a shift in probability distribution of tropospheric events towards more positive magnitudes and longer durations during periods with strengthened stratospheric winds. Similarly, the $\tilde{AO}$ probability distribution is shifted towards more negative anomalies during phases with weakened stratospheric winds and hence positive extreme events will tend to have a weaker magnitude and be less persistent. When averaging over many (e.g., positive) tropospheric events one might not find a significant stratospheric signal since one accounts for situations in which these events occur during times with a strengthened and weakened stratosphere (due to internal variability). Note that the downward influence of stratospheric flow anomalies can potentially also be interpreted as a direct modification of the tropospheric events (which themselves are formed primarily due to internal tropospheric variability). Thus, if a positive tropospheric event happens to occur in a period with strengthened stratospheric winds it will likely be modified due to downward coupling and thus be stronger in magnitude and more persistent. If, however, the same positive tropospheric event occurs during a phase with weakened stratospheric winds, it will generally be suppressed, leading to a weaker magnitude and a shorter event duration. Based on the analyses presented in this study it remains unclear whether an interpretation in terms of a more active modification is adequate.

The sensitivity of tropospheric extremes on the general stratospheric state underlines the general robustness of strat-trop coupling, but also its highly probabilistic nature. Note that Table 1 further indicates the importance of lower stratospheric wind anomalies as part of the downward coupling that has been identified previously by other authors. For the type of coupling described above, any changes in (lower) stratospheric winds can be interpreted as changes in flow restriction at the upper boundary of the (purely) tropospheric system, which then modify the corresponding internal flow evolution. The downward modification of tropospheric events by the (lower) stratospheric state is a rather robust process in early 2020 (see Fig. S4 in the Supplement), further suggesting a strong coupling of observed stratospheric and tropospheric extreme events and extreme persistence. However, it is not clear whether this coupling process was anomalously strong during the winter period of 2020 or if a similarly strong general downward influence can be observed in other years.

8 Summary and Conclusions

Eurasian weather in early 2020 was characterized by anomalously high temperatures and increased precipitation. These anomalous conditions were consistent with tropospheric wind anomalies and an associated shift in the mid-latitude jet (and corresponding Atlantic storm tracks). Further study of re-analysis data and numerical ensemble simulations suggested that the tropospheric extreme conditions were coupled to an extremely strong and zonally symmetric Northern hemispheric polar vortex in the stratosphere.

The substantial strengthening of the polar vortex in early February was mainly associated with a pronounced wave reflection event with persistently strong stratospheric winds for the following several weeks. This wave reflection event was particularly pronounced as it involved a quasi-simultaneous reflection of planetary waves with both zonal wave numbers 1 and 2. The event was preceded by a stratospheric sudden deceleration event, substantially decelerating upper stratospheric winds. The corresponding change in vertical curvature of the zonal wind (and consistent changes in refractive index) created a vertically confined wave guide in the extratropical stratosphere and thus ideal conditions for wave reflection. Being the result of highly non-linear wave-mean flow interactions, WREs are difficult to predict. Therefore early February was a period of low stratospheric predictability, with predictions ranging from a strong recovery of the polar vortex to an increased likelihood of SSWs.

Our analysis did not show indication for a direct coupling of the downwards-reflected wave to the tropospheric zonal mean circulation. We presented evidence that the observed increase in polar vortex strength following the reflection event had an influence on the tropospheric flow and increased the likelihood of the formation of extreme values and the persistently positive phase of the $\tilde{AO}$ index in February and March (and the associated occurrence of extreme weather conditions in the troposphere). We further showed that during periods with a generally strong (weak) stratospheric polar vortex tropospheric extreme events corresponding to a strengthened (weakened) mid-latitude jet tend to occur more often and are generally stronger and more persistent. At the same time positive (negative) $\tilde{AO}$ extremes in the troposphere couple upwards and strengthen (weaken) the polar vortex, allowing for a mutual amplification of the two layers. We further find the downward influence of positive stratospheric zonal circulation anomalies to be overall less robust than the corresponding influence of negative anomalies.

Although the observed extreme circulation anomalies in troposphere and stratosphere appear to be broadly connected, a detailed statistical analysis of large-ensemble simulations revealed a wide range of possible troposphere-stratosphere evolutions. This emphasizes the fortuitous nature of such rare circulation anomalies as observed during early 2020. Our results emphasize the generally robust dynamical coupling between stratosphere and troposphere in both directions, which for individual events nevertheless involves a large range of possible outcomes.

References

Albers, J. R., & Birner, T. (2014). Vortex preconditioning due to planetary and gravity waves prior to sudden stratospheric warmings. *Journal of the Atmospheric Sciences*, *71*(11), 4028–4054.

Baldwin, M. P., Ayarzagüena, B., Birner, T., Butchart, N., Butler, A. H., Charlton-Perez, A. J., ... others (2021). Sudden stratospheric warmings. *Reviews of Geophysics*, *59*(1), e2020RG000708.

Baldwin, M. P., & Dunkerton, T. J. (2001). Stratospheric harbingers of anomalous

weather regimes. *Science*, *294*(5542), 581–584.

Baldwin, M. P., Stephenson, D. B., Thompson, D. W., Dunkerton, T. J., Charlton, A. J., & O'Neill, A. (2003). Stratospheric memory and skill of extended-range weather forecasts. *Science*, *301*(5633), 636–640.

Beerli, R., & Grams, C. M. (2019). 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*(725), 3732–3750.

Birner, T., & Albers, J. R. (2017). Sudden stratospheric warmings and anomalous upward wave activity flux. *Sola*, *13*(Special_Edition), 8–12.

Branković, Č., Palmer, T., Molteni, F., Tibaldi, S., & Cubasch, U. (1990). Extended-range predictions with ecmwf models: Time-lagged ensemble forecasting. *Quarterly Journal of the Royal Meteorological Society*, *116*(494), 867–912.

Charlton-Perez, A. J., Ferranti, L., & Lee, R. W. (2018). The influence of the stratospheric state on north atlantic weather regimes. *Quarterly Journal of the Royal Meteorological Society*, *144*(713), 1140–1151.

Charney, J., & Drazin, P. (1961). Propagation of planetary scale waves from the lower atmosphere to the upper atmosphere. *J Geophys Res*, *66*, 83–109.

Dacre, H. F., & Pinto, J. G. (2020). Serial clustering of extratropical cyclones: a review of where, when and why it occurs. *npj Climate and Atmospheric Science*, *3*(1), 1–10.

Dalchér, A., Kalnay, E., & Hoffman, R. N. (1988). Medium range lagged average forecasts. *Monthly Weather Review*, *116*(2), 402–416.

Domeisen, D. I. (2019). Estimating the frequency of sudden stratospheric warming events from surface observations of the north atlantic oscillation. *Journal of Geophysical Research: Atmospheres*, *124*(6), 3180–3194.

Domeisen, D. I., & Butler, A. H. (2020). Stratospheric drivers of extreme events at the earth's surface. *Communications Earth & Environment*, *1*(1), 1–8.

Domeisen, D. I., Butler, A. H., Charlton-Perez, A. J., Ayarzagüena, B., Baldwin, M. P., Dunn-Sigouin, E., . . . others (2020). The role of the stratosphere in subseasonal to seasonal prediction: 2. predictability arising from stratosphere-troposphere coupling. *Journal of Geophysical Research: Atmospheres*, *125*(2), e2019JD030923.

Domeisen, D. I., Sun, L., & Chen, G. (2013). The role of synoptic eddies in the tropospheric response to stratospheric variability. *Geophysical research letters*, *40*(18), 4933–4937.

Dunn-Sigouin, E., & Shaw, T. A. (2015). Comparing and contrasting extreme stratospheric events, including their coupling to the tropospheric circulation. *Journal of Geophysical Research: Atmospheres*, *120*(4), 1374-1390. doi: https://doi.org/10.1002/2014JD022116

Gerber, E. P., Polvani, L. M., & Ancukiewicz, D. (2008). Annular mode time scales in the intergovernmental panel on climate change fourth assessment report models. *Geophysical Research Letters*, *35*(22).

Hardiman, S. C., Dunstone, N. J., Scaife, A. A., Smith, D. M., Knight, J. R., Davies, P., . . . Greatbatch, R. J. (2020). Predictability of european winter 2019/20: Indian ocean dipole impacts on the nao. *Atmospheric Science Letters*, *21*(12), e1005.

Harnik, N. (2009). Observed stratospheric downward reflection and its relation to upward pulses of wave activity. *Journal of Geophysical Research: Atmospheres*, *114*(D8). Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2008JD010493 doi: https://doi.org/10.1029/2008JD010493

Harnik, N., & Lindzen, R. S. (2001). The effect of reflecting surfaces on the vertical structure and variability of stratospheric planetary waves. *Journal of the Atmospheric Sciences*, *58*(19), 2872 - 2894. Retrieved from

https://journals.ametsoc.org/view/journals/atsc/58/19/1520-0469
_2001_058_2872_teorso_2.0.co_2.xml doi: 10.1175/1520-0469(2001)058⟨2872:
TEORSO⟩2.0.CO;2

Haynes, P., McIntyre, M., Shepherd, T., Marks, C., & Shine, K. P.   (1991).   On the
"downward control" of extratropical diabatic circulations by eddy-induced
mean zonal forces. *Journal of Atmospheric Sciences*, *48*(4), 651–678.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J.,
. . . others (2020). The era5 global reanalysis. *Quarterly Journal of the Royal
Meteorological Society*, *146*(730), 1999–2049.

Hitchcock, P., & Simpson, I. R.   (2014).   The downward influence of stratospheric
sudden warmings. *Journal of the Atmospheric Sciences*, *71*(10), 3856–3876.

Hoffman, R. N., & Kalnay, E.   (1983).   Lagged average forecasting, an alternative
to monte carlo forecasting. *Tellus A: Dynamic Meteorology and Oceanography*,
35(2), 100–118.

Huntingford, C., Marsh, T., Scaife, A. A., Kendon, E. J., Hannaford, J., Kay, A. L.,
. . . others (2014). Potential influences on the united kingdom's floods of winter
2013/14. *Nature Climate Change*, *4*(9), 769–777.

Karpechko, A. Y., Hitchcock, P., Peters, D. H., & Schneidereit, A.   (2017).   Pre-
dictability of downward propagation of major sudden stratospheric warmings.
Quarterly Journal of the Royal Meteorological Society, *143*(704), 1459–1470.

Katsafados, P., Papadopoulos, A., Varlas, G., Papadopoulou, E., & Mavromatidis,
E.   (2014).   Seasonal predictability of the 2010 russian heat wave.   *Natural
Hazards and Earth System Sciences*, *14*(6), 1531–1542.

Kautz, L.-A., Polichtchouk, I., Birner, T., Garny, H., & Pinto, J. G.   (2020).   En-
hanced extended-range predictability of the 2018 late-winter eurasian cold spell
due to the stratosphere. *Quarterly Journal of the Royal Meteorological Society*,
146(727), 1040–1055.

Kolstad, E. W., Breiteig, T., & Scaife, A. A.   (2010).   The association between
stratospheric weak polar vortex events and cold air outbreaks in the northern
hemisphere.   *Quarterly Journal of the Royal Meteorological Society*, *136*(649),
886–893.

Lawrence, Z. D., Perlwitz, J., Butler, A. H., Manney, G. L., Newman, P. A., Lee,
S. H., & Nash, E. R. (2020). The remarkably strong arctic stratospheric polar
vortex of winter 2020: Links to record-breaking arctic oscillation and ozone
loss. *Journal of Geophysical Research: Atmospheres*, *125*(22), e2020JD033271.

Lee, S. H., Lawrence, Z. D., Butler, A. H., & Karpechko, A. Y.   (2020).   Seasonal
forecasts of the exceptional northern hemisphere winter of 2020.   *Geophysical
Research Letters*, *47*(21), e2020GL090328.

Limpasuvan, V., Hartmann, D. L., Thompson, D. W., Jeev, K., & Yung, Y. L.
(2005). Stratosphere-troposphere evolution during polar vortex intensification.
Journal of Geophysical Research: Atmospheres, *110*(D24).

Lott, F., & Miller, M. J. (1997). A new subgrid-scale orographic drag parametriza-
tion: Its formulation and testing.   *Quarterly Journal of the Royal Meteorologi-
cal Society*, *123*(537), 101–127.

Mailier, P. J., Stephenson, D. B., Ferro, C. A., & Hodges, K. I.   (2006).   Serial clus-
tering of extratropical cyclones. *Monthly weather review*, *134*(8), 2224–2240.

Matsuno, T. (1971). A dynamical model of the stratospheric sudden warming. *Jour-
nal of Atmospheric Sciences*, *28*(8), 1479–1494.

Matthews, T., Murphy, C., Wilby, R. L., & Harrigan, S. (2014). Stormiest winter on
record for ireland and uk. *Nature Climate Change*, *4*(9), 738–740.

Noguchi, S., Mukougawa, H., Kuroda, Y., Mizuta, R., Yabu, S., & Yoshimura, H.
(2016). Predictability of the stratospheric polar vortex breakdown: An ensem-
ble reforecast experiment for the splitting event in january 2009.   *Journal of
Geophysical Research: Atmospheres*, *121*(7), 3388–3404.

Orr, A., Bechtold, P., Scinocca, J., Ern, M., & Janiskova, M. (2010). Improved mid-

dle atmosphere climate and forecasts in the ecmwf model through a nonoro-graphic gravity wave drag parameterization. *Journal of climate*, *23*(22), 5905–5926.

Perlwitz, J., & Graf, H.-F. (2001). Troposphere-stratosphere dynamic coupling under strong and weak polar vortex conditions. *Geophysical Research Letters*, *28*(2), 271–274.

Perlwitz, J., & Harnik, N. (2003). Observational evidence of a stratospheric influence on the troposphere by planetary wave reflection. *Journal of Climate*, *16*(18), 3011 - 3026. Retrieved from https://journals.ametsoc.org/view/journals/clim/16/18/1520-0442_2003_016_3011_oeoasi_2.0.co_2.xml doi: 10.1175/1520-0442(2003)016⟨3011:OEOASI⟩2.0.CO;2

Perlwitz, J., & Harnik, N. (2004). Downward coupling between the stratosphere and troposphere: The relative roles of wave and zonal mean processes. *Journal of Climate*, *17*(24), 4902 - 4909. Retrieved from https://journals.ametsoc.org/view/journals/clim/17/24/jcli-3247.1.xml doi: 10.1175/JCLI-3247.1

Pinto, J. G., Zacharias, S., Fink, A. H., Leckebusch, G. C., & Ulbrich, U. (2009). Factors contributing to the development of extreme north atlantic cyclones and their relationship with the nao. *Climate dynamics*, *32*(5), 711–737.

Plumb, R. A. (1989). On the seasonal cycle of stratospheric planetary waves. *Pure and applied geophysics*, *130*(2), 233–242.

Priestley, M. D., Pinto, J. G., Dacre, H. F., & Shaffrey, L. C. (2017). The role of cyclone clustering during the stormy winter of 2013/2014. *Weather*, *72*(7), 187–192.

Rao, J., Garfinkel, C. I., & White, I. P. (2020). 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*(2).

Runde, T., Dameris, M., Garny, H., & Kinnison, D. (2016). Classification of stratospheric extreme events according to their downward propagation to the troposphere. *Geophysical Research Letters*, *43*(12), 6665–6672.

Rupp, P., & Birner, T. (2021). Tropospheric eddy feedback to different stratospheric conditions in idealised baroclinic life cycles. *Weather and Climate Dynamics*, *2*(1), 111–128.

Scaife, A., Karpechko, A. Y., Baldwin, M., Brookshaw, A., Butler, A., Eade, R., ... others (2016). Seasonal winter forecasts and the stratosphere. *Atmospheric Science Letters*, *17*(1), 51–56.

Shaw, T. A., Perlwitz, J., & Harnik, N. (2010). Downward wave coupling between the stratosphere and troposphere: The importance of meridional wave guiding and comparison with zonal-mean coupling. *Journal of Climate*, *23*(23), 6365–6381.

Sigmond, M., Scinocca, J., Kharin, V., & Shepherd, T. (2013). Enhanced seasonal forecast skill following stratospheric sudden warmings. *Nature Geoscience*, *6*(2), 98–102.

Song, Y., & Robinson, W. A. (2004). Dynamical mechanisms for stratospheric influences on the troposphere. *Journal of the atmospheric sciences*, *61*(14), 1711–1725.

Taguchi, M. (2008). Is there a statistical connection between stratospheric sudden warming and tropospheric blocking events? *Journal of Atmospheric Sciences*, *65*(4), 1442–1454.

Thompson, D. W., Baldwin, M. P., & Wallace, J. M. (2002). Stratospheric connection to northern hemisphere wintertime weather: Implications for prediction. *Journal of Climate*, *15*(12), 1421–1428.

Thompson, D. W., & Wallace, J. M. (1998). The arctic oscillation signature in the wintertime geopotential height and temperature fields. *Geophysical research*

letters, *25*(9), 1297–1300.

Tomassini, L., Gerber, E. P., Baldwin, M. P., Bunzel, F., & Giorgetta, M.   (2012). The role of stratosphere-troposphere coupling in the occurrence of extreme winter cold spells over northern europe. *Journal of advances in modeling Earth Systems*, *4*(4).

Tripathi, O. P., Charlton-Perez, A., Sigmond, M., & Vitart, F.   (2015).   Enhanced long-range forecast skill in boreal winter following stratospheric strong vortex conditions. *Environmental Research Letters*, *10*(10), 104007.

Vogel, H., Förstner, J., Vogel, B., Hanisch, T., Mühr, B., Schättler, U., & Schad, T.   (2014).   Time-lagged ensemble simulations of the dispersion of the eyjaf-jallajökull plume over europe with cosmo-art.   *Atmospheric Chemistry and Physics*, *14*(15), 7837–7845.

Wang, L., & Chen, W.   (2010).   Downward arctic oscillation signal associated with moderate weak stratospheric polar vortex and the cold december 2009.   *Geophysical Research Letters*, *37*(9).

Wittman, M. A., Charlton, A. J., & Polvani, L. M.   (2007).   The effect of lower stratospheric shear on baroclinic instability.   *Journal of the atmospheric sciences*, *64*(2), 479–496.

Woollings, T., Charlton-Perez, A., Ineson, S., Marshall, A., & Masato, G.   (2010). Associations between stratospheric variability and tropospheric blocking. *Journal of Geophysical Research: Atmospheres*, *115*(D6).

Zängl, G., Reinert, D., Rípodas, P., & Baldauf, M.   (2015).   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*(687), 563–579.