Geophysical Research Letters

RESEARCH LETTER
10.1029/2019GL083346

Special Section:
Bridging Weather and Climate: Subseasonal-to-Seasonal (S2S) Prediction

Predictability of Northern Hemisphere Final Stratospheric Warmings and Their Surface Impacts

Amy H. Butler[1,2], Andrew Charlton-Perez[3], Daniela I.V. Domeisen[4], Isla R. Simpson[5], and Jeremiah Sjoberg[6]

[1]Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, CO, USA, [2]National Oceanic and Atmospheric Administration, Chemical Sciences Division, Boulder, CO, USA, [3]Department of Meteorology, University of Reading, Reading, UK, [4]Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland, [5]Climate and Global Dynamics Laboratory, National Center for Atmospheric Research, Boulder, CO, USA, [6]COSMIC/UCAR, Boulder, CO, USA

Abstract Each spring, the climatological westerly winds of the Northern Hemisphere (NH) stratospheric polar vortex turn easterly as the stratospheric equator-to-pole temperature gradient relaxes. The timing of this event is dictated both by the annual return of sunlight to the pole and by dynamical influences from the troposphere. Here we consider the predictability of NH final stratospheric warmings in multimodel hindcasts from the Subseasonal to Seasonal project database. We evaluate how well the Subseasonal to Seasonal prediction systems perform in capturing the timing of final warmings. We compare the predictability of early warmings (which are more strongly driven by wave forcing) and late warmings and find that late warmings are more skillfully predicted at longer lead times. Finally, we find significantly increased predictive skill of NH near-surface temperature anomalies at Weeks 3–4 lead times following only early final warmings.

Key Points:
- Final stratospheric warmings that occur in early spring are less predictable than those that occur in late spring
- The interannual variability in the timing of the final warming is well predicted more than 4 weeks in advance
- Early final warmings provide a source of surface climate predictive skill at Weeks 3–4 in Northern Hemisphere spring

Supporting Information:
- Supporting Information S1

Correspondence to:
A. H. Butler,
amy.butler@noaa.gov

Citation:
Butler, A. H., Charlton-Perez, A., Domeisen, D. I. V., Simpson, I. R., & Sjoberg, J. (2019). Predictability of Northern Hemisphere final stratospheric warmings and their surface impacts. *Geophysical Research Letters*, *46*, 10,578–10,588. https://doi.org/10.1029/2019GL083346

Received 15 APR 2019
Accepted 18 AUG 2019
Accepted article online 22 AUG 2019
Published online 2 SEP 2019

1. Introduction

Wintertime circulation in the extratropical stratosphere is dominated by a strong, westerly polar vortex. The annual evolution of the polar vortex is predominantly dictated by the seasonal cycle of radiative heating. In the Northern Hemisphere (NH), the polar vortex develops at the end of August, reaches maximum westerly zonal wind speeds in January, and returns to its summer easterly state between March and May. If this seasonality were driven by radiative processes alone, there would be little variation in the timing of the formation and the breakup of the polar vortex each year. This is true for the formation date of the vortex, which has varied by 9 days at the most during the period 1979–2017. In contrast, wave mean flow dynamics play a significant role in the variability of the stratospheric polar vortex and in the timing of its final reversal to its easterly summertime state—the so-called "final warming" (FW)—which has a spread of more than 2 months over this same period (Waugh & Rong, 2002; Waugh et al., 1999).

Dynamic disruptions of the midwinter polar vortex, which are caused by a combination of the breaking of tropospherically generated planetary waves (Polvani & Waugh, 2004; Sjoberg & Birner, 2012) and internal resonance of the vortex (Domeisen et al., 2018; Esler & Matthewman, 2011; Matthewman & Esler, 2011), are associated with rapid warming in the polar stratosphere and thus are termed "sudden stratospheric warmings" (SSWs). These events can have significant and persistent impacts on surface climate for days to weeks afterward, and implications for predictability on Subseasonal to Seasonal (S2S) time scales (Butler et al., 2019; Sigmond et al., 2013). These surface impacts often project onto the negative phase of the North Atlantic Oscillation (NAO; Butler et al., 2017; Charlton-Perez et al., 2018; Domeisen, 2019), representing an equatorward shift of the North Atlantic storm track. Recent studies have indicated that midwinter SSWs are generally only predictable at lead times of 10–15 days (Karpechko, 2018; Taguchi, 2018), though there is large variability in the predictability between events, and probabilistic prediction may be possible at weeks or even a season ahead of time (Scaife et al., 2016).

FWs of the stratospheric polar vortex are also followed by surface anomalies in temperature and mean sea level pressure but do not always project well onto the wintertime NAO pattern (Black & McDaniel, 2007;

Black et al., 2006). It is likely FW events contribute to extratropical surface climate predictability in spring (Hardiman et al., 2011; Sun & Robinson, 2009). Given that a FW occurs every year, one important facet of these events is their timing. FWs that occur in early spring lead to a more negative NAO-like pattern in March and April relative to FWs that occur in late spring (Ayarzagüena & Serrano, 2009; Li et al., 2012). While studies have examined FWs in the Southern Hemisphere in relation to predictive skill (Byrne & Shepherd, 2018; Lim et al., 2018), and Jia et al. (2017) find higher seasonal predictive skill related to the stratosphere in spring compared to winter, we are not aware of any study that has quantified predictive skill of NH surface climate at subseasonal time scales following early and late FWs. We seek to quantify the predictability of FWs and their influence on the predictability of surface climate using state-of-the-art S2S prediction systems.

2. Methods

The S2S project database (Vitart et al., 2017) comprises hindcasts of the 1981–2016 period from 11 operational forecasting centers. We consider six of these models (CMA, ECCC, ECMWF, JMA, NCEP, and UKMO). Note that CMA and ECCC have lower model tops and lower stratospheric resolution and so may not fully capture stratospheric variability (A. J. Charlton-Perez et al., 2013). These prediction systems offer a large number of hindcasts, with ensemble sizes ranging from 2–11 and initializations occurring as often as every day down to 3 times per month. Details about the hindcast period and ensemble size for each prediction system can be found in supporting information Table S1. We use the full record available for each system, rather than a common period, in order to maximize the number of FWs in each case. S2S hindcasts are bias-corrected by removing the lead time-dependent climatology from the ensemble-mean of each model over its full record, leaving out the year being bias corrected from the climatology. For looking at the predictability of FWs, the reanalysis climatology is added back to the bias-corrected zonal wind anomalies and then the FWs in the prediction systems are calculated.

We use European Centre for Medium-Range Weather Forecasts Re-Analysis (ERA)-interim reanalysis data (Dee et al., 2011) from 1981–2016 to determine observed FW dates and for verification. Because there is no standard definition for FWs (Butler et al., 2015), and a variety of methods have been used in the literature, here we use daily-mean zonal-mean zonal winds at 60°N and 10 hPa, both because this metric allows consistency with the common definition for midwinter SSWs (Charlton & Polvani, 2007) and because the output is available from the S2S database for all considered model systems. Following Butler and Gerber (2018), the FW occurs on the last date during the 1 July to 30 June period of each year on which the zonal winds reverse and do not return to westerly for more than 10 consecutive days. The resulting dates (supporting information Table S2) generally agree with previously published FW dates using slightly different metrics and reanalysis (Hu et al., 2014, 2015). The mean climatological NH FW date in ERA-interim over the period 1981–2016 is 15 April. There are 19 FWs that occur before 15 April, defined as "early" warmings, and 17 FWs that occur on or after, defined as "late" warmings. The earliest FW date is 5 March 2016, and the latest FW date is 12 May 1981. Note there is some decadal variability in the timing of FWs (Waugh et al., 1999), with more early FWs occurring in the 1990s (a period of few midwinter SSWs) and more late FWs occurring in the 2000s (a period of many midwinter SSWs; Ivy et al., 2014). This agrees with Hu et al. (2014) who found that winters with midwinter SSWs tend to be followed by late FWs and winters with no SSWs tend to be followed by early FWs.

For section 3.1, the same definition for FWs applied to ERA-interim above is also applied to the bias-corrected S2S hindcasts initialized between 1 February and the date of the observed FW every year. We do not consider initialization dates after the observed FW because then the model will be initialized with easterly zonal winds in the stratosphere. Note that if the zonal wind reverses less than 10 days from the end of the hindcast, it is still counted as a predicted FW event.

Our method of classifying early and late events using the climatological mean FW date is a technique used in previous studies (Ayarzagüena & Serrano, 2009; Wei et al., 2007) that is designed to separate predominantly wave-driven events from predominantly radiative-driven events in a simple manner. This is a different technique than used in Hardiman et al. (2011), which classified FWs by their vertical evolution (zonal wind reversals at 10 hPa first compared to reversals at 1 hPa first). However, since data output at 1 hPa is not available for the S2S database, we need a technique based only on 10-hPa data. The analysis in section 3.1 confirms that our method succeeds at separating dynamic events from radiative ones, though there

are likely exceptions to this rule (i.e., some late events as defined here may also exhibit dynamic warming characteristics). We have verified that our results are not sensitive to the choice of 15 April ±10 days.

As a reference for hindcasts initialized during FWs, we create "control" hindcasts, in which for each initialization closest to the observed FW date, we sample all other years in the record for that date in which the stratospheric winds (10 hPa, 60°N) are westerly (because of this requirement, note that the control population size for the late FWs is much smaller than the control for early FWs). The control hindcasts thus have many more samples than the FW hindcasts (though many years are repeated) and also have the same distribution, in terms of seasonality, as the FW hindcasts.

Significance is assessed for composites and the anomaly correlation coefficient (ACC) via Monte Carlo sampling. For each date that enters the FW composite, we randomly sample that forecast date from any of the years in the period of analysis to build up a Monte Carlo sample that consists of the same number of forecasts on the same dates as present in the FW composite but from randomly chosen years. The composite mean and ACC statistics are obtained for this new sample, and this is repeated for 100 such samples. Where the actual composite or ACC values lie outside of ±2 standard deviations of the distribution of Monte Carlo samples, it is considered significant.

3. Results

3.1. Observed and Simulated Differences in Early and Late FWs and Their Predictability

We first highlight observed differences between early and late FWs (Figures 1a–1d) and examine the ability of the forecast models to capture these features (Figures 1e and 1f). The stratospheric zonal winds (Figures 1a and 1c) prior to early FWs are significantly stronger than prior to late FWs by definition—early FWs are defined to occur when the climatological westerly winds are stronger. However, the tendency of the zonal winds in the ±5-day period about the event date is larger and the magnitude of the easterlies after the event is greater for early FWs, suggesting these events are more sudden. The change from westerlies to easterlies following late FWs occurs more gradually, indicative of radiative relaxation as the dominant process. Note that the zonal winds are *anomalously* weak (strong) in the week prior to early (late) FWs (not shown, but in agreement with, e.g., Ayarzagüena & Serrano, 2009), an indication of anomalous wave forcing prior to early events. This is quantified by the anomalous lower stratospheric eddy heat flux (Figures 1b and 1d), which is larger for early FWs than for late FWs, consistent with the idea that early breakups are more wave driven (Hu et al., 2014; Wei et al., 2007). The eddy heat flux anomaly stays elevated 5 days before and after early FWs, similar to midwinter SSWs (Sjoberg & Birner, 2012).

To evaluate how well the S2S models simulate the differences between early and late FWs, we look at the predicted distribution of zonal winds and eddy heat flux anomalies (spanning from 20 days prior to 30 days after observed early and late FWs) based on all forecasts that were initialized within the window that was 3–4 weeks earlier (Figure 1e and 1f). For example, at day = 0, we are showing the distribution for the actual FW date from all the forecasts initialized between 3–4 weeks prior. The models capture well the difference in zonal winds between early and late FWs, showing significantly stronger and more variable zonal winds prior to early FWs compared to late FWs, as observed. Note that at 3- to 4-week lead times the S2S prediction systems predict early FW dates that are on average up to 5 days later than observed.

The ability of the S2S prediction systems to correctly simulate the zonal wind differences between early and late FWs is mostly a reflection of the models' ability to capture the seasonal evolution of the polar vortex. The ability of the S2S systems to predict differences in the lower stratospheric eddy heat flux anomaly for early and late FWs is less apparent (Figure 1f). The eddy heat flux anomalies are more variable and shifted toward slightly more positive values in the −10 to +5 days surrounding early FWs compared to late FWs, as observed. Still, the somewhat weaker response in the eddy heat flux anomaly prior to early FWs compared to observations indicates that the models have difficulty deterministically predicting 3–4 weeks ahead of time any particular wave pulse driving the more rapid deceleration of the vortex to its easterly state.

Next we consider how well the interannual variability of the timing of the FW is predicted (Figure 2). The ERA-interim "day of year" (DOY) time series of the FW shows clearly the early FWs in 2005 and 2016 and the general trend toward later FWs during the 1992–2009 period (Hu et al., 2014; Karpechko et al., 2005). Comparing to the ensemble-mean DOY time series in the S2S prediction systems at different lead times suggests that the timing of certain FW events is more difficult to predict than others. For example, almost all models predicted the 2005 early FW to occur later than observed, and in general the DOY of early FWs

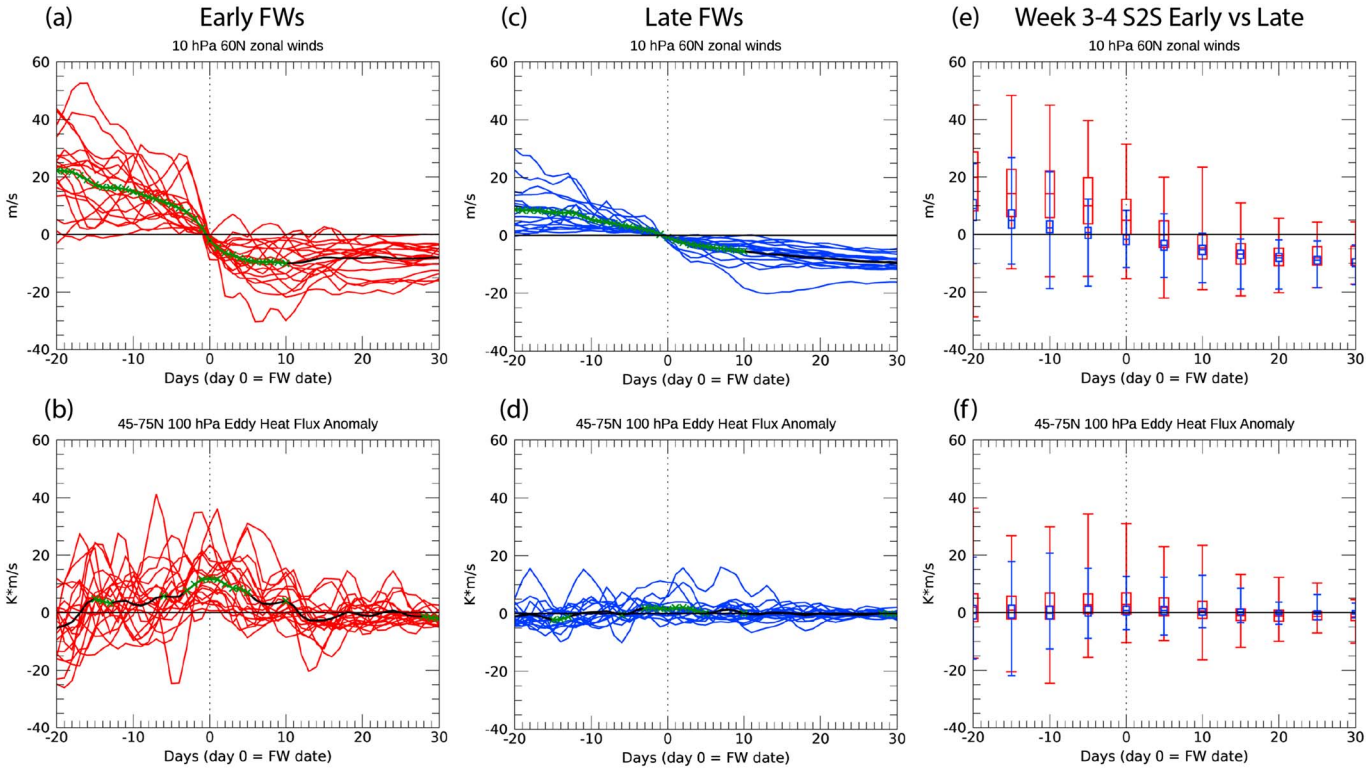

Figure 1. (top row) Zonal-mean zonal winds (m/s) at 60°N and 10-hPa and (bottom row) 45–75°N and 100-hPa eddy heat flux anomaly for (a, b) observed early FWs, (c, d) observed late FWs, and (e, f) early versus late FWs in the S2S prediction systems. In (a–d), the black line is the composite mean and is green with x's where a *t* test difference in means between the early and late FWs is significant at the $p \leq 0.05$ level. In (e) and (f), the box and whiskers are based on Weeks 3–4 forecasts of Days −20, −15, −10, −5, 0, 5, 10, 15, 20, 25, and 30 around the observed early (red) and late (blue) FWs. FW = final warming; S2S = Subseasonal to Seasonal.

shows greater ensemble spread and root-mean-square error (RMSE) than late FWs (Figures 2b and 2c). A comparison of ensemble spread (Figure 2b) to the RMSE (Figure 2c) gives an estimate of ensemble underdispersion or overdispersion (Fortin et al., 2014; Wilks, 2010). For example, the ensemble spread in the UKMO prediction system is smaller at longer leads than the RMSE, suggesting possible underdispersion.

One key result is that the prediction systems are able to generally capture whether the FW will be early or late, even 4 weeks ahead of time; the prediction skill for the DOY time series is not highly dependent on lead time. The correlations between the observed and predicted DOY time series also does not strongly depend on lead time (not shown), and are greater than $r = 0.9$ for most lead times and prediction systems. The ensemble spread and RMSE of the DOY time series generally increases for longer lead times as expected (Figures 2b and 2c), but this increase is not consistent across lead times; in some prediction systems such as NCEP and ECMWF, the RMSE of the DOY time series peaks at Week 2 (this may reflect noise due to relatively small samples). Overall, the difference in RMSE depends more strongly on the prediction system (and sampling period) than on lead time, with CMA and UKMO showing larger RMSE (greater than 8 days) at Weeks 3–4 compared to the other four prediction systems.

The fact that the general timing of the FW is not strongly dependent on lead time suggests that the timing is regulated by some seasonal to decadal influence. Previous studies have suggested a role of the Quasi-biennial Oscillation (QBO), El Niño–Southern Oscillation (ENSO), the North Pacific Gyre Oscillation, or solar cycle variability (Camp & Tung, 2007; Haigh & Roscoe, 2009; Hu et al., 2018; Salby & Callaghan, 2007). We find that over the 1981–2016 period, out of 12 El Niño winters (defined by the DJF Oceanic Niño Index from the NOAA Climate Prediction Center), 9 were followed by early FWs and only 3 by late FWs. Out of 12 La Niña winters, 8 were followed by late FWs and only 4 by early FWs (an equal number, 6, of early and late FWs follow 12 ENSO-neutral winters). This brief analysis suggests some relationship, but given the short record, it is not clear whether this interannual relationship might be dominated by ENSO teleconnections to the stratosphere (Domeisen et al., 2019) or is a sampling artifact. Note that during the 1981–2016 period,

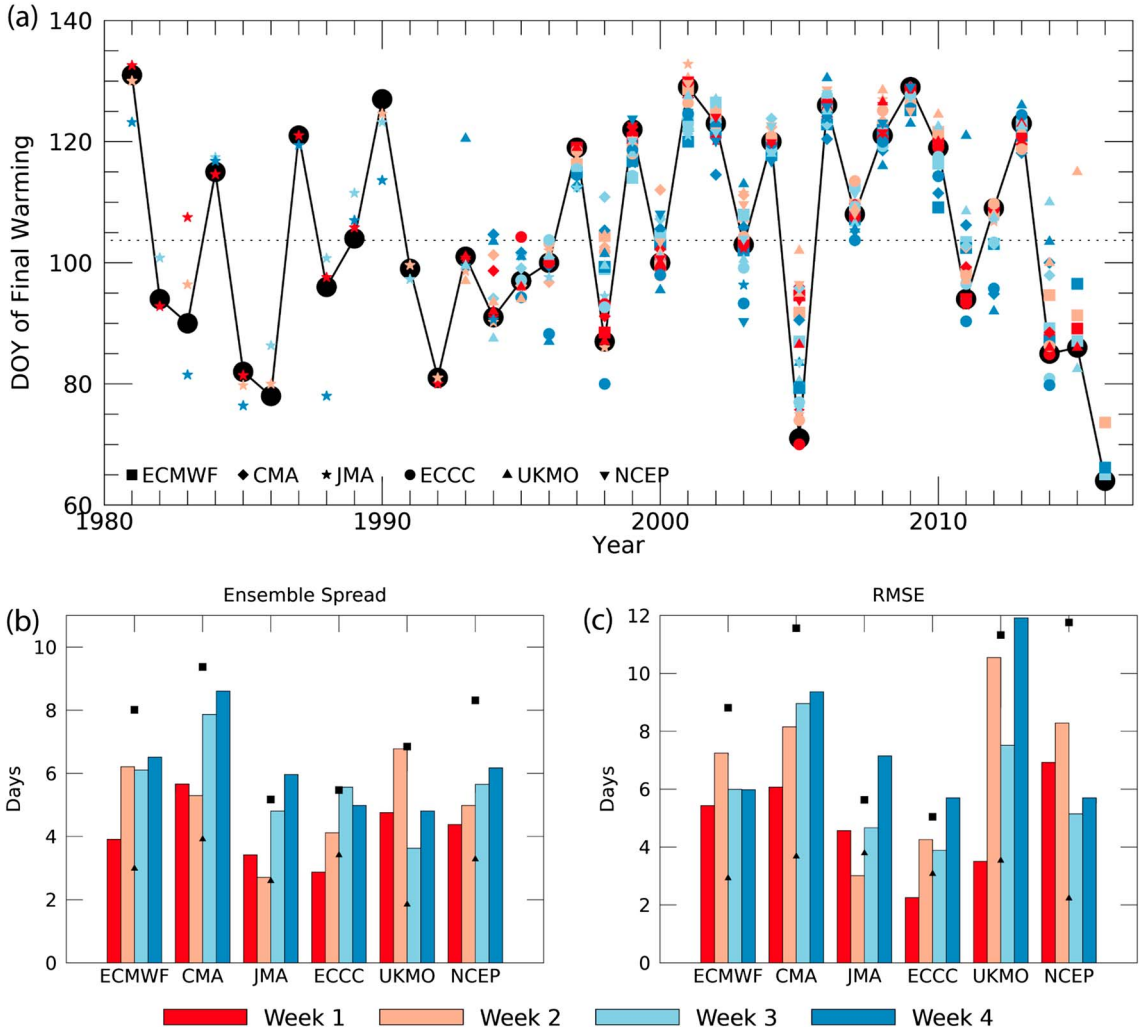

Figure 2. (a) Time series of the day of year (DOY) of the final warming (FW) in European Centre for Medium-Range Weather Forecasts Re-Analysis-interim (black dots) and the Subseasonal to Seasonal prediction systems (symbols), shaded according to different lead times before the observed FW. The number of symbols for each year is an indication of the number of prediction systems with data during that time period; see supporting information Table S1. The dashed line shows the mean FW date of 15 April. (b) Ensemble spread (square root of the average ensemble variance; Fortin et al., 2014) and (c) RMSE of the ensemble-mean DOY time series for Weeks 1–4 lead times for different prediction systems. For (b) and (c), the ensemble spread and RMSE averaged over all lead times is shown as a black square for early FWs and a black triangle for late FWs. RMSE = root-mean-square error.

more SSWs occurred during La Niña than El Niño (Garfinkel et al., 2019). Of the nine El Niño winters that were followed by early FWs, only two had midwinter SSWs. On the other hand, of the eight La Niña winters that were followed by late FWs, seven had midwinter SSWs. Thus, it is likely that the occurrence of midwinter SSWs, which is related to both QBO and ENSO (Butler & Polvani, 2011; Garfinkel & Hartmann, 2010; Garfinkel et al., 2018) but also internal atmospheric variability, ultimately dictates the timing of radiative recovery of the vortex to its westerly state and any subsequent ability of the vortex to reverse in a FW (Hu et al., 2014).

If we now consider only those times in which the prediction systems *accurately* (within ±3 days) predict the observed FW (Figure 3a), we find on average 45% of ensemble members are able to predict the observed FW at lead times of up to 20 days (note that the false alarm rates, in which a FW was predicted but did not occur, are on average about 18%—see supporting information for a description). Even out to 30-day lead times, some systems have >25% of ensemble members accurately predicting the FW, which is higher than the predictive skill at these lead times for most midwinter SSWs (Karpechko, 2018; Taguchi, 2018; A. Y. Tripathi et al., 2014).

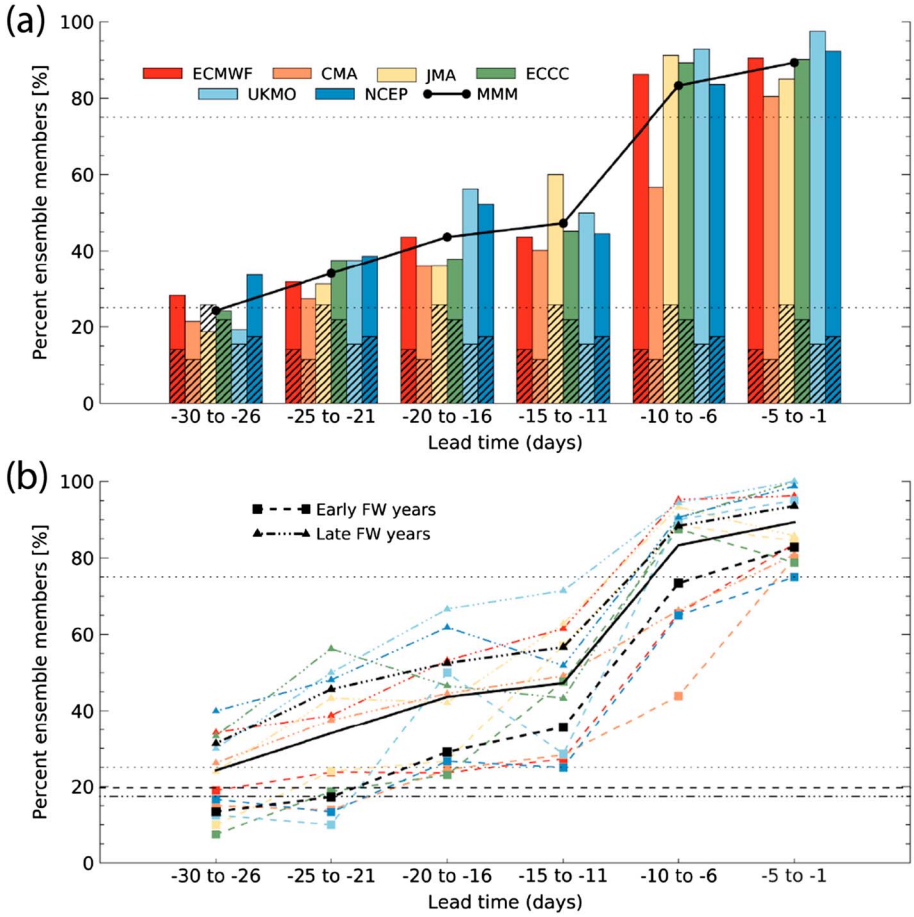

Figure 3. (a) Percent of ensemble members in each prediction system (color bars) that detect a final warming within ±3 days of the observed FW, as a function of lead time (averaged within 5-day bins), where the solid black line is the multimodel mean. Black patterned bars indicate the false alarm rate for each model averaged over lead times of 1–30 days. (b) The same as in (a) but subdivided into early (dashed lines, squares) and late FW years (dashed dot lines, triangles). The black lines show the multimodel means (the line with no symbols is the same line for all years in panel (a)), and light colored lines correspond to different models. The bold horizontal lines are the multimodel mean rates of false alarms averaged over lead times of 1–30 days. In both (a) and (b), the 25% and 75% thresholds are indicated by dotted lines. FW = final warming.

To better understand this result, we consider separately the predictability of early versus late FWs (Figure 3b). Early FWs are less predictable at longer lead times than late FWs. For example, at 11- to 15-day lead times, more than 55% of ensemble members accurately detect late FWs, compared to 35% of members accurately detecting early FWs. Thus, the predictive time scales of early FWs more closely match those of other dynamic stratospheric events such as midwinter SSWs (A. Y. Karpechko, 2018; Taguchi, 2018). The decreased predictability at longer lead times for wave-driven events is associated with the deterministic prediction limits of synoptic tropospheric variability (Buizza & Leutbecher, 2015; Domeisen et al., 2017) that drive these events in the stratosphere. The late FWs are, perhaps unsurprisingly, more predictable at longer lead times, given that the dominant process is the dependable radiative relaxation of the meridional temperature gradient as winter turns to spring. There may also be some inherent advantage to predicting late FWs, as the likelihood of the model forecasting a FW will increase if the FW has not occurred yet as spring progresses.

3.2. Can FWs Influence Surface Predictive Skill?

Hardiman et al. (2011) find a more negative NAO-like pattern in March and April for years with FWs that occur at 10 hPa first compared to years that occur at 1 hPa first, and Ayarzagüena and Serrano (2009) find corresponding shifts in the North Atlantic storm track for early versus late FW years. We find no significant observed differences in the NAO index between early and late FWs when compositing around the FW

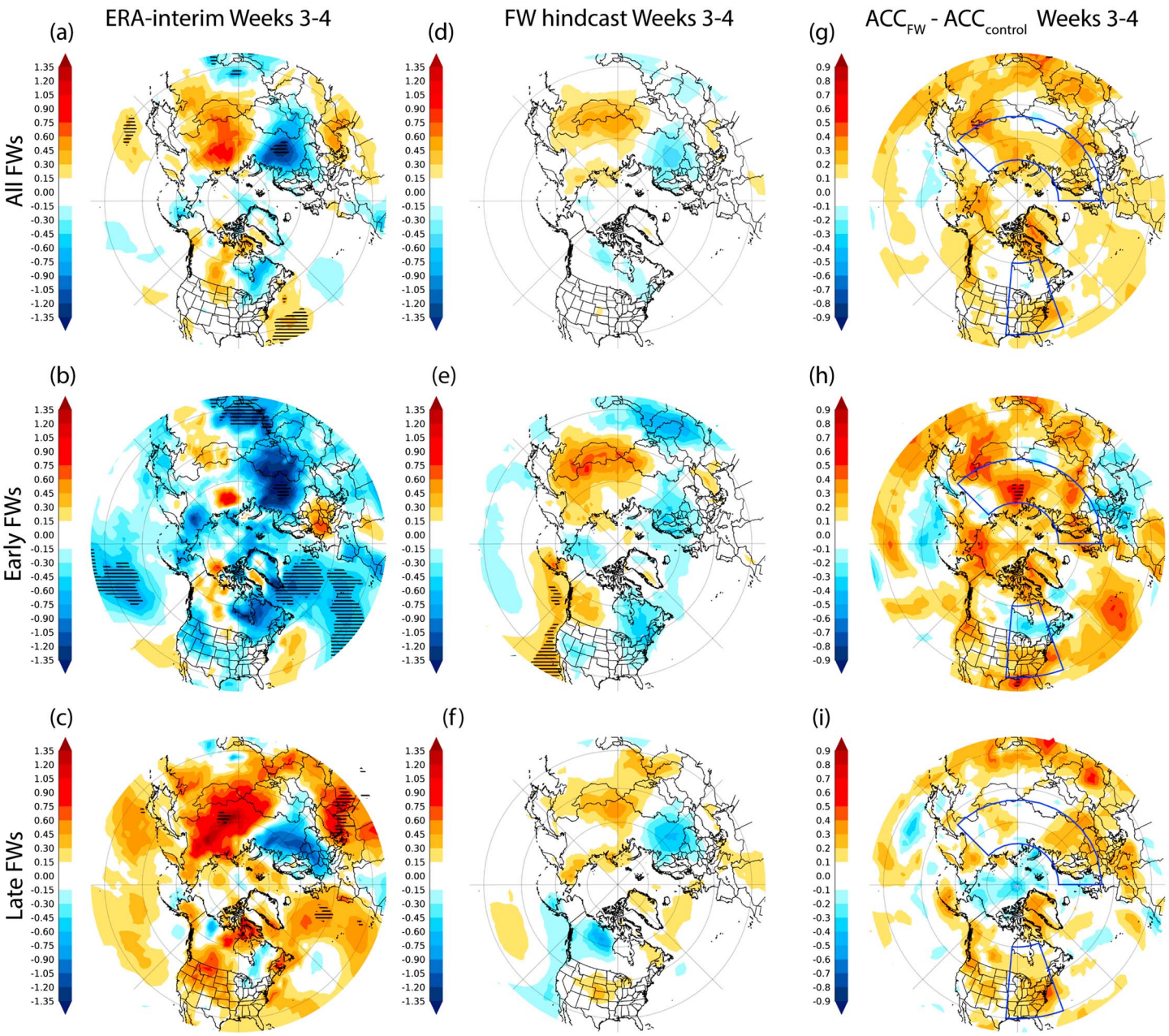

Figure 4. (a–c) Composite T2m anomalies (K) for ERA-interim, (d–f) multimodel composite T2m anomalies (K), and (g–i) the difference in the anomaly correlation coefficient for FW hindcasts and control hindcasts, averaged Weeks 3–4 after (top) all FWs, (middle) early FWs, and (bottom) late FWs. Hatching depicts regions that are significant according to the Monte Carlo test outlined in section 2. T2m = 2-m temperature; ERA = European Centre for Medium-Range Weather Forecasts Re-Analysis; FW = final warming; ACC = anomaly correlation coefficient.

date (supporting information Figure S1). However, the springtime NAO is confined poleward compared to the wintertime NAO, and the centers of action over Iceland and the Azores may shift location seasonally (Domeisen et al., 2017; Hurrell, 1995; Portis et al., 2001). Thus, the pattern change after both early and late FWs is "NAO-like" but may not project well onto the wintertime NAO pattern that we are using here.

To examine the impact of the FW and its timing on predictive skill of springtime surface climate, we first find the model initialization closest to the observed FW date for every year in each model and then composite the surface temperature anomalies for Weeks 3–4. Comparisons to ERA-interim are also made based on the closest initialization date (i.e., not necessarily the observed FW date) for each model and each year.

As a benchmark for the quantitative impact of the FW at the surface, Figure 4a shows that the composite 2-m temperatures (T2m) for ERA-interim averaged for the Weeks 3–4 period after all FWs are anomalously cold over northern Europe and eastern North America and anomalously warm over northern Asia and the western Atlantic. However, there are some noticeable differences for early versus late FWs (Figures 4b and 4c). Early FWs are followed 3–4 weeks later by anomalously cold temperatures throughout most of the NH, particularly over the North Atlantic, central Pacific, and western Asia. Late FWs, on the other hand, are followed by anomalously warm temperatures across most of the NH (though the anomalies are generally not significant). The exception is over Europe where anomalous cold prevails. Identical plots for individual models (supporting information Figures S2–S7) show general agreement with the multimodel mean (Figure 4), even though they cover different time periods, suggesting that at least to first order these composite differences are not dominated by, for example, sampling of surface temperature trends, though that could be playing some role.

The differences in the early and late FW response, particularly over the Pacific sector, may in part be a reflection of ENSO, given that most early FWs occur during El Niño years and most late FWs during La Niña years (see section 3.1). The signature of ENSO in the S2S prediction systems (Figures 4d–4f) at Weeks 3–4 lead times on predicted T2m temperature anomalies is also evident for both early and late FWs; note the equal and opposite response over the Pacific-North American region for early versus late FWs that is typical of El Niño versus La Niña (Domeisen et al., 2019). Indeed, compositing the hindcasts initialized closest to the observed FW separately for El Niño and La Niña years (supporting information Figures S8a and S8b) shows patterns very similar to those in Figures 4e and 4f. However, one interesting point is that the prediction systems capture the dipole in temperature anomalies across Eurasia for both early and late FWs. The S2S hindcast composite for all FWs thus provides a rough estimate of the stratospheric component of the response to FWs, as the ENSO-related component cancels out over the Pacific (a similar pattern is recovered when all El Niño plus La Niña years are composited together, Figure S8d). While composite analysis has been used to isolate the tropospheric and stratospheric components of the ENSO response in reanalysis data and AMIP-type simulations (Butler et al., 2014; Jiménez-Esteve & Domeisen, 2018; Polvani et al., 2017), here it is particularly interesting because any signal at 3- to 4-week lead time in the S2S prediction systems likely results from processes with memory on those time scales or longer. The Pacific sector ENSO component is thus not unexpected, but the Eurasian component is evidence of the S2S models predicting a stratospheric influence, via the FW, on surface temperatures 3–4 weeks later.

To quantify the influence on surface predictive skill, we calculate the ACC for both the FW and control hindcasts for the average of Weeks 3–4 T2m temperature anomalies, for the hindcast initialized closest to the observed FW date. The ACC in the control forecasts for Weeks 3–4 is large and positive over the Pacific sector and near 0 elsewhere (not shown). Because the control forecasts are initialized on the same dates as observed FWs but different years, the skill in the control forecasts is likely dominated almost entirely by ENSO, with little stratospheric influence.

Because the ACC in the control forecasts is dominated by ENSO, the difference in ACC between the Weeks 3–4 FW forecasts and the control forecasts should mostly isolate the skill increase associated with FWs (Figures 4g–4i). Overall, the ACC is marginally enhanced over most of the extratropics after the model is initialized with a FW, relative to the control runs, with the greatest increases in skill seen over the North Atlantic and Eurasian regions. Averaged over the entire hemisphere (supporting information Table S3), the ACC increase is statistically significant, as determined by Monte Carlo testing of the same FW dates but with randomized years (which suggests the skill is not arising from just coincidental timing of the FW dates with some seasonal shift in predictive skill). However, when skill is calculated separately for early and late FWs, it is apparent that the greatest increase in Weeks 3–4 skill follows early FWs, particularly over the Atlantic-Eurasian region, though insignificant increases in ACC also occur following late FWs (compare Figures 4h and 4i). This is also true for most individual models (Figures S2–S7).

Why is the increase in predictive skill greater following early FWs than late FWs? One reason could be that because early FWs induce stronger and more sudden stratospheric changes, their surface impacts are stronger and more persistent. Another reason could be that because they occur more often in El Niño years, the tropospheric pathway of El Niño already induces a tendency toward a negative NAO-like state over the North Atlantic (Jiménez-Esteve & Domeisen, 2018) and is thus more receptive to the FW signal. Finally, it is possible that since there is a FW every year, the greatest increase in skill occurs for early FWs when the

model is not necessarily predicting a stratospheric signal, than for late FWs when the model climatologically expects the stratospheric change to occur. We also note that the response after late FWs may lack significance simply because the control forecasts are based on only a few years that still have westerly winds in late spring.

4. Conclusions

Using the S2S project database, we have examined both the predictability of FWs and their influence on predictive skill of surface climate at 3- to 4-week lead times.

On 3- to 4-week time scales, the S2S prediction systems are not able to predict the enhanced anomalous heat flux associated with early FWs, and in general, the models predict early FWs to occur later than observed. Overall, early FWs are less predictable at longer leads compared to late FWs. Additionally, we find that ensemble-mean hindcasts from S2S prediction systems are able to capture whether the FW will be early or late even 4 weeks ahead of time. We note a relationship of early FWs occurring more frequently after El Niño winters and late FWs more frequently after La Niña winters, but the occurrence of midwinter SSWs seems to play a primary role in the timing of the FW.

Finally, we find that the Weeks 3–4 predictive skill of T2m anomalies (as measured by ACC) is enhanced throughout the NH, and particularly over the North Atlantic-Eurasian region, for hindcasts initialized at the onset of early FWs relative to control hindcasts. Nonsignificant, but positive, increases in ACC are also found following late FWs. This result could have important implications for S2S forecasting, as it suggests that in particular more dynamic early FWs, but to some extent late FWs, could provide an additional source of Weeks 3–4 predictive skill in spring.

Acknowledgments
We acknowledge the scientific guidance of the World Climate Research Programme to motivate this work, coordinated in the framework of SPARC. This work is based on S2S data. S2S is a joint initiative of the World Weather Research Programme (WWRP) and the World Climate Research Programme (WCRP). The original S2S database is hosted at ECMWF as an extension of the TIGGE database (https://www.ecmwf.int/en/research/projects/s2s). D. D. is supported by the Swiss National Science Foundation through Grant PP00P2_170523.

References

Ayarzagüena, B., & Serrano, E. (2009). Monthly characterization of the tropospheric circulation over the Euro-Atlantic area in relation with the timing of stratospheric final warmings. *Journal of Climate*, *22*(23), 6313–6324. https://doi.org/10.1175/2009JCLI2913.1

Black, R. X., & McDaniel, B. A. (2007). The dynamics of Northern Hemisphere stratospheric final warming events. *Journal of the Atmospheric Sciences*, *64*(8), 2932–2946. https://doi.org/10.1175/JAS3981.1

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

Buizza, R., & Leutbecher, M. (2015). The forecast skill horizon. *Quarterly Journal of the Royal Meteorological Society*, *141*(693), 3366–3382. https://doi.org/10.1002/qj.2619

Butler, A., Charlton-Perez, A., Domeisen, DanielaI. V., Garfinkel, C., Gerber, E. P., Hitchcock, P., et al. (2019). Chapter 11—Sub-seasonal predictability and the stratosphere. In A. W. Robertson, & F. Vitart (Eds.), *Sub-seasonal to seasonal prediction* (pp. 223–241). Amsterdam, Netherlands: Elsevier. https://doi.org/10.1016/B978-0-12-811714-9.00011-5

Butler, A. H., & Gerber, E. P. (2018). Optimizing the definition of a sudden stratospheric warming. *Journal of Climate*, *31*(6), 2337–2344. https://doi.org/10.1175/JCLI-D-17-0648.1

Butler, A. H., Gerber, E. P., Mitchell, D., & Seviour, W. J. M. (2014). New efforts in developing a standard definition for sudden stratospheric warmings. *SPARC Newsletter*, *43*, 23–24.

Butler, A. H., & Polvani, L. M. (2011). El Niño, La Niña, and stratospheric sudden warmings: A reevaluation in light of the observational record. *Geophysical Research Letters*, *38*, L13807. https://doi.org/10.1029/2011GL048084

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

Butler, A. H., Sjoberg, J. P., Seidel, D. J., & Rosenlof, K. H. (2017). A sudden stratospheric warming compendium. *Earth System Science Data*, *9*(1), 63–76. https://doi.org/10.5194/essd-9-63-2017

Byrne, N. J., & Shepherd, T. G. (2018). Seasonal persistence of circulation anomalies in the Southern Hemisphere stratosphere, and its implications for the troposphere. *Journal of Climate*, *31*(9), 3467–3483. https://doi.org/10.1175/JCLI-D-17-0557.1

Camp, C. D., & Tung, K.-K. (2007). The influence of the solar cycle and QBO on the late-winter stratospheric polar vortex. *Journal of the Atmospheric Sciences*, *64*(4), 1267–1283. https://doi.org/10.1175/JAS3883.1

Charlton, A. J., & Polvani, L. M. (2007). A new look at stratospheric sudden warmings. Part I: Climatology and modeling benchmarks. *Journal of Climate*, *20*(3), 449–469.

Charlton-Perez, A. J., Baldwin, M. P., Birner, T., Black, R. X., Butler, A. H., Calvo, N., et al. (2013). On the lack of stratospheric dynamical variability in low-top versions of the CMIP5 models. *Journal of Geophysical Research: Atmospheres*, *118*, 2494–2505. https://doi.org/10.1002/jgrd.50125

Charlton-Perez, A., L., F., & R.W., L. (2018). The influence of the stratospheric state on North Atlantic weather regimes. *Quarterly Journal of the Royal Meteorological Society*, *144*, 1140–1151. https://doi.org/10.1002/qj.3280

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597. https://doi.org/10.1002/qj.828/abstract

Domeisen, D. I. V. (2019). Estimating the frequency of sudden stratospheric warming events from surface observations of the North Atlantic Oscillation. *Journal of Geophysical Research: Atmospheres*, *124*, 3180–3194. https://doi.org/10.1029/2018JD030077

Domeisen, D. I. V., Badin, G., & Koszalka, I. M. (2017). How predictable are the Arctic and North Atlantic Oscillations? Exploring the variability and predictability of the Northern Hemisphere. *Journal of Climate*, *31*, 997–1014. https://doi.org/10.1175/JCLI-D-17-0226.1

Domeisen, D. I. V., Garfinkel, C. I., & Butler, A. H. (2019). The teleconnection of El Niño Southern Oscillation to the stratosphere. *Reviews of Geophysics*, *57*, 5–47. https://doi.org/10.1029/2018RG000596

Domeisen, D. I. V., Martius, O., & Jiménez-Esteve, B. (2018). Rossby wave propagation into the Northern Hemisphere stratosphere: The Role of Zonal Phase Speed. *Geophysical Research Letters*, *45*, 2064–2071. https://doi.org/10.1002/2017GL076886

Esler, J. G., & Matthewman, N. J. (2011). Stratospheric sudden warmings as self-tuning resonances. Part II: Vortex displacement events. *Journal of the Atmospheric Sciences*, *68*(11), 2505–2523. https://doi.org/10.1175/JAS-D-11-08.1

Fortin, V., Abaza, M., Anctil, F., & Turcotte, R. (2014). Why should ensemble spread match the RMSE of the ensemble mean? *Journal of Hydrometeorology*, *15*(4), 1708–1713. Retrieved from http://www.jstor.org/stable/24914540

Garfinkel, C. I., & Hartmann, D. L. (2010). Influence of the quasi-biennial oscillation on the North Pacific and El Niño teleconnections. *Journal of Geophysical Research*, *115*, D20116. https://doi.org/10.1029/2010JD014181

Garfinkel, C. I., Schwartz, C., Butler, A. H., Domeisen, D. I. V., Son, S.-W., & White, I. P. (2019). Weakening of the teleconnection from El Niño Southern Oscillation to the Arctic stratosphere over the past few decades: What can be learned from subseasonal forecast models? *Journal of Geophysical Research: Atmospheres*, *124*, 7683–7696. https://doi.org/10.1029/2018JD029961

Garfinkel, C. I., Schwartz, C., Domeisen, D. I. V., Son, S.-W., Butler, A. H., & White, I. P. (2018). Extratropical atmospheric predictability from the quasi-biennial oscillation in subseasonal forecast models. *Journal of Geophysical Research: Atmospheres*, *123*, 7855–7866. https://doi.org/10.1029/2018JD028724

Haigh, J. D., & Roscoe, H. K. (2009). The final warming date of the Antarctic polar vortex and influences on its interannual variability. *Journal of Climate*, *22*(22), 5809–5819. https://doi.org/10.1175/2009JCLI2865.1

Hardiman, S. C., Butchart, N., Charlton-Perez, A. J., Shaw, T. A., Akiyoshi, H., Baumgaertner, A., et al. (2011). Improved predictability of the troposphere using stratospheric final warmings. *Journal of Geophysical Research*, *116*, D18113. https://doi.org/10.1029/2011JD015914

Hu, J., Li, T., & Xu, H. (2018). Relationship between the North Pacific Gyre Oscillation and the onset of stratospheric final warming in the Northern Hemisphere. *Climate Dynamics*, *51*, 3061–3075. https://doi.org/10.1007/s00382-017-4065-3

Hu, J., Ren, R., & Xu, H. (2014). Occurrence of winter stratospheric sudden warming events and the seasonal timing of spring stratospheric final warming. *Journal of the Atmospheric Sciences*, *71*, 2319–2334. https://doi.org/10.1175/JAS-D-13-0349.1

Hu, J., Ren, R., Xu, H., & Yang, S. (2015). Seasonal timing of stratospheric final warming associated with the intensity of stratospheric sudden warming in preceding winter. *Science China Earth Sciences*, *58*(4), 615–627. https://doi.org/10.1007/s11430-014-5008-z

Hurrell, J. W. (1995). Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. *Science*, *269*(5224), 676 LP–679. https://doi.org/10.1126/science.269.5224.676

Ivy, D. J., Solomon, S., & Thompson, D. W. J. (2014). On the identification of the downward propagation of Arctic stratospheric climate change over recent decades. *Journal of Climate*, *27*, 2789–2799. https://doi.org/10.1175/JCLI-D-13-00445.1

Jia, L., Yang, X., Vecchi, G., Gudgel, R., Delworth, T., Fueglistaler, S., et al. (2017). Seasonal prediction skill of northern extratropical surface temperature driven by the stratosphere. *Journal of Climate*, *30*, 4463–4475. https://doi.org/10.1175/JCLI-D-16-0475.1

Jiménez-Esteve, B., & Domeisen, D. I. V. (2018). The tropospheric pathway of the ENSO-North Atlantic teleconnection. *Journal of Climate*, *31*, 4563–4584. https://doi.org/10.1175/JCLI-D-17-0716.1

Karpechko, A. Y. (2018). Predictability of sudden stratospheric warmings in the ECMWF extended-range forecast system. *Monthly Weather Review*, *146*(4), 1063–1075. https://doi.org/10.1175/MWR-D-17-0317.1

Karpechko, A., Kyrö, E, & Knudsen, B. M. (2005). Arctic and Antarctic polar vortices 1957–2002 as seen from the ERA-40 reanalyses. *Journal of Geophysical Research*, *110*, D21109. https://doi.org/10.1029/2005JD006113

Li, L., Li, C., Pan, J., & Tan, Y. (2012). On the differences and climate impacts of early and late stratospheric polar vortex breakup. *Advances in Atmospheric Sciences*, *29*(5), 1119–1128. https://doi.org/10.1007/s00376-012-1012-4

Lim, E.-P., Hendon, H. H., & Thompson, D. W. J. (2018). Seasonal evolution of stratosphere-troposphere coupling in the Southern Hemisphere and implications for the predictability of surface climate. *Journal of Geophysical Research: Atmospheres*, *123*, 12,002–12,016. https://doi.org/10.1029/2018JD029321

Matthewman, N. J., & Esler, J. G. (2011). Stratospheric sudden warmings as self-tuning resonances. Part I: Vortex splitting events. *Journal of the Atmospheric Sciences*, *68*(11), 2481–2504. https://doi.org/10.1175/JAS-D-11-07.1

Polvani, L. M., Sun, L., Butler, A. H., Richter, J. H., & Deser, C. (2017). Distinguishing stratospheric sudden warmings from ENSO as key drivers of wintertime climate variability over the North Atlantic and Eurasia. *Journal of Climate*, *30*(6), 1959–1969. https://doi.org/10.1175/JCLI-D-16-0277.1

Polvani, L. M., & Waugh, D. W. (2004). Upward wave activity flux as a precursor to extreme stratospheric events and subsequent anomalous surface weather regimes. *Journal of Climate*, *17*(18), 3548–3554. https://doi.org/10.1175/1520-0442(2004)017h3548:UWAFAAi2.0.CO;2

Portis, D. H., Walsh, J. E., El Hamly, M., & Lamb, P. J. (2001). Seasonality of the North Atlantic Oscillation. *Journal of Climate*, *14*, 2069–2078.

Salby, M. L., & Callaghan, P. F. (2007). Influence of planetary wave activity on the stratospheric final warming and spring ozone. *Journal of Geophysical Research*, *112*, D20111. https://doi.org/10.1029/2006JD007536

Scaife, A. A., Karpechko, A. Y., Baldwin, M. P., Brookshaw, A., Butler, A. H., Eade, R., et al. (2016). Seasonal winter forecasts and the stratosphere. *Atmospheric Science Letters*, *17*, 51–56. https://doi.org/10.1002/asl.598

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

Sjoberg, J. P., & Birner, T. (2012). Transient tropospheric forcing of sudden stratospheric warmings. *Journal of the Atmospheric Sciences*, *69*(11), 3420–3432. https://doi.org/10.1175/JAS-D-11-0195.1

Sun, L., & Robinson, W. A. (2009). Downward influence of stratospheric final warming events in an idealized model. *Geophysical Research Letters*, *36*, L03819. https://doi.org/10.1029/2008GL036624

Taguchi, M. (2018). Comparison of subseasonal-to-seasonal model forecasts for major stratospheric sudden warmings. *Journal of Geophysical Research: Atmospheres*, *123*, 210–231. https://doi.org/10.1029/2018JD028755

Tripathi, O. P., Baldwin, M., Charlton-Perez, A., Charron, M., Eckermann, S. D., Gerber, E., et al. (2014). The predictability of the extratropical stratosphere on monthly time-scales and its impact on the skill of tropospheric forecasts. *Quarterly Journal of the Royal Meteorological Society*, *141*, 987–1003. https://doi.org/10.1002/qj.2432

Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., et al. (2017). The Sub-seasonal to Seasonal Prediction (S2S) Project Database. *Bulletin of the American Meteorological Society*, *98*, 162–173. https://doi.org/10.1175/BAMS-D-16-0017.1

Waugh, D. W., Randel, W. J., Pawson, S., Newman, P. A., & Nash, E. R. (1999). Persistence of the lower stratospheric polar vortices. *Journal of Geophysical Research*, *104*(D22), 27,191–27,201. https://doi.org/10.1029/1999JD900795

Waugh, D. W., & Rong, P.-P. (2002). Interannual variability in the decay of lower stratospheric arctic vortices. *Journal of the Meteorological Society of Japan. Series II*, *80*(4B), 997–1012. https://doi.org/10.2151/jmsj.80.997

Wei, K., Chen, W., & Huang, R. H. (2007). Dynamical diagnosis of the breakup of the stratospheric polar vortex in the Northern Hemisphere. *Science in China, Series D: Earth Sciences*, *50*(9), 1369–1379. https://doi.org/10.1007/s11430-007-0100-2

Wilks, D. S. (2010). On the reliability of the rank histogram. *Monthly Weather Review*, *139*(1), 311–316. https://doi.org/10.1175/2010MWR3446.1