The role of the stratosphere in subseasonal to seasonal prediction part I: predictability of the stratosphere

Article

Accepted Version

The role of the stratosphere in subseasonal to seasonal prediction
Part I: Predictability of the stratosphere

Daniela I.V. Domeisen[1], Amy H. Butler[2,3], Andrew J. Charlton-Perez[4], Blanca Ayarzagüena[5,6], Mark P. Baldwin[7], Etienne Dunn-Sigouin[8], Jason C. Furtado[9], Chaim I. Garfinkel[10], Peter Hitchcock[11], Alexey Yu. Karpechko[12], Hera Kim[13], Jeff Knight[14], Andrea L. Lang[15], Eun-Pa Lim[16], Andrew Marshall[16], Greg Roff[16], Chen Schwartz[10], Isla R. Simpson[17], Seok-Woo Son[13], Masakazu Taguchi[18]

[1]Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland
[2]Cooperative Institute for Research in Environmental Sciences, Boulder, CO, USA
[3]National Oceanic and Atmospheric Administration, Chemical Sciences Division, USA
[4]University of Reading, Reading, UK
[5]Universidad Complutense de Madrid, Madrid, Spain
[6]Instituto Geociencias, CSIC-UCM, Spain
[7]University of Exeter, Exeter, UK
[8]Geophysical Institute, U. Bergen and Bjerknes Centre, Bergen, Norway
[9]School of Meteorology, University of Oklahoma, USA
[10]Fredy and Nadine Herrmann Institute of Earth Sciences, Hebrew University of Jerusalem, Israel
[11]Cornell University, Ithaca, NY, USA
[12]Finnish Meteorological Institute, Finland
[13]Seoul National University, South Korea
[14]MetOffice Hadley Centre, Exeter, Devon, UK
[15]University at Albany, State University of New York, USA
[16]Bureau of Meteorology, Australia
[17]Climate and Global Dynamics Laboratory, NCAR, USA
[18]Aichi University of Education, Japan

Key Points:

- High-top models have more skill in the stratosphere and the troposphere compared to low-top models.
- Extreme stratospheric events are predictable at one- to two- week lead times in S2S models.
- SSW events tend to be less predictable than strong vortex events or final warming events.

Corresponding author: Daniela Domeisen, ETH Zurich, Universitätstrasse 16, 8092 Zürich, Switzerland, daniela.domeisen@env.ethz.ch

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1029/2019JD030920

Abstract

The stratosphere has been identified as an important source of predictability for a range of processes on subseasonal to seasonal (S2S) timescales. Knowledge about S2S predictability within the stratosphere is however still limited. This study evaluates to what extent predictability in the extratropical stratosphere exists in hindcasts of operational prediction systems in the S2S database. The stratosphere is found to exhibit extended predictability as compared to the troposphere. Prediction systems with higher stratospheric skill tend to also exhibit higher skill in the troposphere. The analysis also includes an assessment of the predictability for stratospheric events, including early and mid-winter sudden stratospheric warming (SSW) events, strong vortex events, and extreme heat flux events for the Northern Hemisphere, and final warming events for both hemispheres. Strong vortex events and final warming events exhibit higher levels of predictability as compared to SSW events. In general, skill is limited to the deterministic range of one to two weeks. High-top prediction systems overall exhibit higher stratospheric prediction skill as compared to their low-top counterparts, pointing to the important role of stratospheric representation in S2S prediction models.

1 Introduction

The winter stratosphere is dominated by strong westerly circumpolar winds in the extratropics of both hemispheres, which exhibit maximum variability from December-March in the Northern Hemisphere (NH) and from October-December in the Southern Hemisphere (SH) (R. A. Plumb, 1989; Thompson & Wallace, 2000). This variability, which is larger in the Northern Hemisphere, is linked to dynamical extreme events. The most prominent events are so-called *major sudden stratospheric warming* (SSW) events. These occur in the polar NH on average every second winter (A. H. Butler, Sjoberg, Seidel, & Rosenlof, 2017; Charlton & Polvani, 2007) and are associated with a disruption of the polar vortex, reversing the climatological westerly winds to easterlies in mid-winter. Temperatures at a height of 30 km can increase by around 50°C within a few days during these events, and the troposphere tends to respond with an anomalously persistent negative signature of the Northern Annular Mode (NAM) and the North Atlantic Oscillation (NAO) (Baldwin & Dunkerton, 2001; Charlton-Perez, Ferranti, & Lee, 2018; D. I. V. Domeisen, 2019; Karpechko, Hitchcock, Peters, & Schneidereit, 2017). In the SH, only one major SSW event has been observed to date, in September 2002 (e.g. Charlton, O'Neill, Lahoz, & Berrisford, 2005; Newman & Nash, 2005; Taguchi, Masakazu, 2014). In addition, minor stratospheric warming events in the SH can also significantly impact the Southern Annular Mode (SAM) and the associated surface climate (e.g. E. P. Lim, Hendon, & Thompson, 2018).

In the NH, the polar vortex can also significantly weaken early in the season. *Early winter weak vortex events* occur before wind speeds peak in the stratosphere, are strongly influenced by the transient development of the vortex into winter, and can precondition the vortex for midwinter variability for both the Northern (Albers & Birner, 2014; Ayarzagüena, Langematz, & Serrano, 2011; Limpasuvan, V, Thompson, D, & Hartmann, D L, 2004) and Southern Hemispheres (Ivy et al., 2017). Early vortex weakening events can potentially influence early winter surface climate, e.g. in NH winter 2016/17 (Tyrrell, Karpechko, Uotila, & Vihma, 2019), despite the fact that they generally do not meet the criteria for major mid-winter SSWs. These events can exhibit zonal wind speeds of less than 10 ms^{-1} for more than a week at 60°N and 10 hPa and can exhibit easterly zonal mean winds at latitudes poleward of 60°N, which can lead to similar surface impacts as major SSWs (A. H. Butler & Gerber, 2018).

Occasionally, the vortex strengthens significantly in so-called *strong polar vortex events* (e.g. Limpasuvan, Hartmann, Thompson, Jeev, & Yung, 2005) in boreal winter or austral spring. Strong polar vortex events occur when the winter polar vortex intensifies significantly above climatology, and these events generally have opposite impacts to mid-winter SSWs on surface weather (i.e., in the NH (SH) the surface influence projects onto the positive phase of the NAO (SAM)). Strong vortex events have been found to increase surface predictability (Tripathi, Charlton-Perez, Sigmond, & Vitart, 2015).

In addition, shorter-lived events, so-called *wave reflection* and *negative heat flux events* can also impact the entire atmospheric column and often precede strong vortex events (Dunn-Sigouin & Shaw, 2015; Perlwitz & Harnik, 2003). Extreme stratospheric wave-1 negative heat flux events are coupled with significant changes in the tropospheric circulation, in particular, they are followed by a poleward shift of the North Atlantic jet consistent with a positive phase of the NAO (Dunn-Sigouin & Shaw, 2015; Lubis, Matthes, Omrani, Harnik, & Wahl, 2016; Shaw & Perlwitz, 2013; Shaw, Perlwitz, & Weiner, 2014). The tropospheric response following negative heat flux events can be reproduced in dry dynamical core experiments if the stratosphere is nudged to the observed event evolution and the troposphere is freely evolving (Dunn-Sigouin & Shaw, 2018).

At the end of winter, the polar vortex collapses to easterlies in a *final stratospheric warming* event in spring (R. Black, McDaniel, & Robinson, 2006; R. X. Black & McDaniel, 2007). While final warmings are typically induced by the radiative relaxation of the equator-to-pole temperature gradient as sunlight returns to the pole, they can also be dynamically induced by wave breaking in a manner similar to mid-winter SSWs (Hardiman et al., 2011; Hu, Ren, & Xu, 2014; Hu, Ren, Yu, & Xu, 2014). Final warmings can exhibit different surface impacts than mid-winter SSWs in the NH (Ayarzagüena & Serrano, 2009; Hardiman et al., 2011). In the SH the downward impact of the final warming tends to manifest in the tropospheric SAM (e.g., E. Gerber et al., 2010; E. P. Lim et al., 2018; Seviour et al., 2014; Son, Purich, Hendon, Kim, & Polvani, 2013; Thompson & Solomon, 2005), which drives variations in surface climate throughout the SH (Bandoro, Solomon, Donohoe, Thompson, & Santer, 2014; E. P. Lim et al., 2018). This indicates that a skillful prediction of the SH stratospheric polar vortex in spring can provide an early warning for the polarity of the surface SAM and associated SH climate in spring to summer, beyond the SAM's typical two-week decorrelation time scale (A. G. Marshall, Hudson, Wheeler, Hendon, & Alves, 2011).

The above described extreme events in the stratosphere remain difficult to predict deterministically despite significant progress in stratospheric representation, including higher model lids and increased stratospheric resolution (e.g. A. H. Butler et al., 2016). In idealized dynamical core models in ensemble mode, SSWs can on average be deterministically predicted 10 days in advance (E. P. Gerber, Orbe, & Polvani, 2009). For more complex prediction systems these predictive lead times are similar (Tripathi et al., 2016; Tripathi, Baldwin, et al., 2015) but can vary widely between different SSW events (Karpechko, 2018; A. Marshall & Scaife, 2010; Noguchi et al., 2016; Taguchi, 2018; Taguchi, Masakazu, 2016).

Given the influence of the stratosphere on surface weather during NH winter and SH spring and the implied added predictability on S2S timescales (e.g. Baldwin et al., 2003; Scaife et al., 2016) it is crucial to understand the dynamics and predictability of the stratosphere itself. Due to the different mechanisms for the above described events there are reasons to expect different timescales of vortex evolution - and hence different predictability - for example during weak versus strong vortex events (Limpasuvan et al., 2005; Limpasuvan, V et al., 2004) in addition to the different surface impacts previously mentioned. Only recently, via the World Climate Research Program (WCRP) and World Weather Research Program (WWRP) S2S project, has an intercomparison of a large number of state-of-the-art operational S2S prediction systems with stratospheric output been made possible. Here, we evaluate the predictability of the extratropical stratosphere of both hemispheres using this database, while the second part of this study (D. I. Domeisen et al., 2019, hereafter Part II) investigates the influence of the stratosphere on the predictability of surface climate with a focus on the NH. Section 2 describes the S2S database and our methodology, including the definition of stratospheric extreme events (section 2.3). Section 3 evaluates the predictability of the winter stratosphere relative to the troposphere, while Section 4 considers the predictability of stratospheric extreme events. Section 5 provides a summary and discussion of the results.

2 Methodology

2.1 Data

The focus of this study will be the analysis of hindcasts from the subseasonal to seasonal forecast project database (Vitart et al., 2017). The database is a repository of forecast and hindcast data from 11 different operational subseasonal forecast systems. The focus of this study is on the hindcast data, since it spans a broad range of different stratospheric states, at the expense of the large ensemble sizes characteristic of the real-time forecasts. Nine of the eleven systems are analyzed in detail in this study. Two models (KMA and HMCR) had to be excluded due to data issues. Table 1 lists the model systems included in our analysis along with specific details of each system and its output availability. The hindcast period differs substantially between different ensemble prediction systems due to their operational strategy. For the majority of the analysis in this study, the period 1996-2010, over which hindcasts are available for most prediction systems, is used. Not all analyses in this study are able to employ all prediction systems, e.g. due to the differing length of the hindcasts or the different time periods for which hindcasts are available, hence different sections may use a more limited set of models or a different hindcast period depending on the specific requirements of a particular analysis. An effort has been made to include as many models as possible into every analysis. Exceptions to the data listed in Table 1 will be noted.

Table 1. Details of the prediction systems considered in this study, based on the data available at the time of analysis. '×' indicates high-top models throughout this study, here referring to a top model level above 0.1 hPa and a stratospheric resolution with several levels above 1 hPa. ALI refers to the BoM data assimilation scheme. Differing numbers of ensemble members for UKMO were used in this study, depending on the members available at the time of data acquisition for each section.

Prediction system	Initialization	Hindcast period	Ensemble size
BoM	ERA-interim/ALI	1981-2013	33
CMA	NCEP-NCAR R1	1994-2014	4
ECCC	ERA-interim	1995-2014	4
ECMWF$^{\times}$	ERA-interim	1997-2016	11
JMA$^{\times}$	JRA-55	1981-2010	5
CNRM-Meteo$^{\times}$	ERA-interim	1993-2014	15
CNR-ISAC	ERA-interim	1981-2010	1
NCEP$^{\times}$	CFSR	1999-2010	4
UKMO$^{\times}$	ERA-interim	1993-2015	2-7

There are several ways in which the design of the prediction systems is important to consider when thinking about their ability to forecast the stratosphere. Of primary importance are the vertical resolution of the atmospheric model component, and the height of the model top level. Figure 1 shows the spacing of model levels for the nine systems

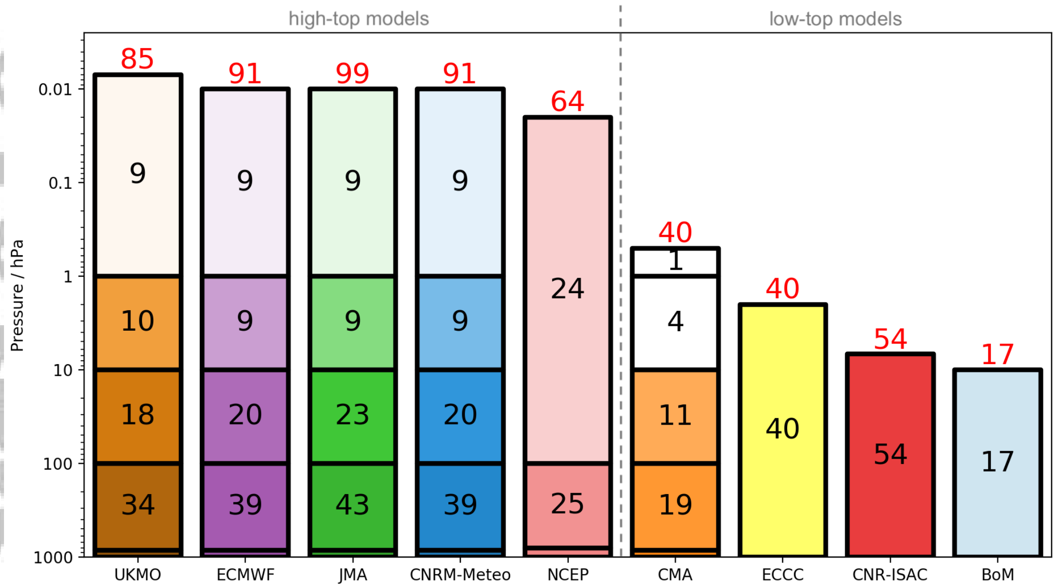


Figure 1. Schematic representation of model vertical resolution for all S2S prediction systems used in this study. Each block represents the pressure range indicated on the y-axis. The number of model levels in each range is shown numerically. The shading in each box is proportional to the average level spacing [in kilometers] in that region of the atmosphere. The red number at the top of each bar shows the total number of levels in each model. The dashed line indicates the separation between high- and low-top models (see Table 1).

considered. The prediction systems are divided into two broad groups, i.e., high-top models (as defined in Table 1), which fully represent the stratosphere (ECMWF, UKMO, JMA, NCEP and CNRM-Meteo), and low-top models (ECCC, CMA, CNR-ISAC and BoM). Note that the prediction systems are initialized with different reanalysis products in the atmosphere, i.e. JRA-55 (Kobayashi et al., 2015), ERA-Interim (Dee et al., 2011), NCEP-NCAR R1 (Kalnay et al., 1998), and CFSR (Saha et al., 2010) as indicated in Table 1. This may lead to differences in the models' performance in the stratosphere. The detailed performance of different reanalysis products in the stratosphere has been reviewed by the SPARC Reanalysis Intercomparison Project (e.g. Long, Fujiwara, Davis, Mitchell, & Wright, 2017). In this study, we verify all hindcasts against ERA-Interim reanalysis. While this could be biased against systems initialized with a different reanalysis, in most cases sampling variability will be much larger than variability between reanalysis products (E. P. Gerber & Martineau, 2018).

2.2 Skill Measures

In this study, skill is evaluated according to a range of measures that are commonly used in the literature. One common metric is the *correlation coefficient* r given by

$$r = \frac{\sum_{t=1}^{T}(X_{mod} - C_{mod})(X_{obs} - C_{obs})}{\sqrt{\sum_{t=1}^{T}(X_{mod} - C_{mod})^2 \cdot \sum_{t=1}^{T}(X_{obs} - C_{obs})^2}} \tag{1}$$

where X is a time-dependent variable, and the subscripts *mod* and *obs* denote the model ensemble mean and the reanalysis dataset, respectively. C_{mod} is the lead time dependent model climatology, over the same period of time as the observed climatology C_{obs}. T is the number of events or time steps for which r is evaluated.

To evaluate the spatial skill of the anomaly pattern, we use the *anomaly correlation coefficient* (e.g. Table 2 and Figs. 2 and 3):

$$ACC = \frac{\sum_{t=1}^{T}\sum_{s=1}^{S} w \cdot (X_{mod} - C_{mod})(X_{obs} - C_{obs})}{\sqrt{\sum_{t=1}^{T}\sum_{s=1}^{S} w \cdot (X_{mod} - C_{mod})^2 \cdot \sum_{t=1}^{T}\sum_{s=1}^{S} w \cdot (X_{obs} - C_{obs})^2}}. \qquad (2)$$

Spatial weighting by the cosine of latitude w and spatial averaging over S grid spaces is applied as an additional summation over the covariance and variance terms separately. This formulation of the ACC allows an *a posteriori* removal of systematic errors in the model hindcasts. In this study, the ACC and r are computed for the ensemble mean for each prediction system as a function of forecast lead time. The multi-model mean is the averaged correlation from all prediction systems. A skill level of 0.6 is used as a threshold to compare the different models, consistent with other studies of seasonal and subseasonal predictability.

A further measure that has recently been introduced by Eade et al. (2014) is the *ratio of predictable components* (RPC), a property of ensemble hindcasts comparing the size of a predicted signal to that expected from their correlation coefficient:

$$RPC = \frac{r \cdot \sigma_{tot}}{\sigma_{mod}} \qquad (3)$$

with r as defined in equation (1). σ_{mod} is the standard deviation of the model ensemble mean, and σ_{tot} is the total variance in the ensemble, where σ_{tot} uses all ensemble members and start dates for each lead time. Thereby, the RPC is the ratio of the correlation coefficient multiplied by the standard deviation across all years and ensemble members (the variability we would expect the ensemble mean to contain given the correlation) to the standard deviation of the year-to-year variations in the ensemble mean (the variability we actually obtain from the system). RPC = 1 indicates that a forecast system perfectly reflects the predictability of the observed system. Eade et al. (2014) showed that we expect an ensemble prediction system that is over-confident to have RPC < 1 and one that is under-confident to have RPC > 1. For RPC > 1 the system has less ensemble mean amplitude than expected by the correlation of the ensemble mean with the observations (i.e., the ACC). This is found for many prediction systems on seasonal timescales and likely reveals deficiencies in the model (e.g., O'Reilly, Weisheimer, Woollings, Gray, and MacLeod (2018)).

2.3 Classification of Stratospheric Events

We investigate the predictability of extreme events in the polar stratosphere in section 4. Here we briefly describe how we classify these stratospheric events.

Early winter weak vortex event. Weak (i.e., less than -1σ from the ERA-interim daily climatological mean) zonal mean zonal winds at 60°N and 10 hPa that persist for at least a week beginning in the month of November. There are 4 of these events in the 1996-2010 period in ERA-interim.

Strong polar vortex event. Strong polar vortex events are defined as periods when zonal mean zonal winds at 60°N and 10 hPa exceed a threshold value. Following Tripathi, Charlton-Perez, et al. (2015) we use the 80th percentile of ERA-Interim November to March (NDJFM) winds over the 1980-2012 period, which is 41.2 m/s. We define the start of the event as the date when the winds exceed the threshold for the first time. This condition is set to ensure that the forecasts are not initiated during a strong polar vortex. An event must last for at least two days and events must be separated by at least 30 days. During the period 1996-2010, there are 12 strong polar vortex events.

Mid-winter SSW event. Though there are several possible definitions for a SSW event (A. H. Butler et al., 2015), here we base our analysis on zonal mean zonal wind reversals at 60°N and 10 hPa (Charlton & Polvani, 2007), as listed in Table 2 of A. H. Butler et al. (2017) for ERA-Interim (December - February (DJF) events only). During the 1996-2010 period, there are 11 mid-winter SSW events.

Negative heat flux events. Negative heat flux events are defined by extreme values of the daily zonal mean wave-1 meridional heat flux ($\overline{v'T'}_{k=1}$, where k denotes the zonal wave number) computed from daily mean values of the meridional wind v and temperature T, and averaged from 60°-90°N at 50 hPa during January - March (JFM), as in Dunn-Sigouin and Shaw (2015). Negative events are identified when the 5-day running mean high latitude heat flux drops below the 5th percentile of the climatological distribution from reanalysis (-13.5 K ms^{-1}). The central date of the events is defined at the day of minimum high latitude heat flux, and events must be separated by a minimum of 15 days. 10 events are identified from 1996-2010 (Table 1 in Dunn-Sigouin and Shaw (2015)).

Final stratospheric warming events. The final warming is defined as the last date prior to June 30 (December 31) of each year when the ERA-Interim daily mean zonal mean zonal winds at 10 hPa and 60° latitude in the NH (SH) turn easterly and do not return to westerly for more than 10 consecutive days (A. H. Butler & Gerber, 2018). The final warming typically occurs around mid-April in the NH and mid-November in the SH at the 10 hPa level. This same definition is used for model runs initialized between February 1st (September 1st for the SH) and the date of the observed final warming. Note that if the zonal wind reverses less than 10 days from the end of the forecast, it is counted as a predicted final warming, although the criterion of not returning to westerlies cannot be evaluated in this case. Because there is a final warming every spring, there are 14 observed events from 1997-2010. The climatological mean final warming date from ERA-Interim (over the longer 1981-2016 period) is April 15 in the NH and November 20 in the SH.

3 Evaluation of the Baseline Prediction Skill in the Stratosphere and the Troposphere

The main purpose of this study is to investigate how well the prediction systems in the database simulate the predictability in the stratosphere and troposphere on subseasonal timescales. As a first step we characterize the baseline skill present in the prediction systems in the stratosphere and troposphere.

The stratosphere and the troposphere have different characteristics when it comes to persistence and predictability. Large-scale variability in the stratosphere has significantly longer decorrelation timescales than the troposphere (Baldwin et al., 2003; E. Gerber et al., 2010; E. Gerber, Polvani, & Ancukiewicz, 2008; Simpson, Hitchcock, Shepherd, & Scinocca, 2011). The extent to which the decorrelation timescale is determined primarily by radiative timescales or a combination of radiative and dynamical processes is uncertain (Charlton-Perez & O'Neill, 2010; Hitchcock, Shepherd, Yoden, Taguchi, & Noguchi, 2013). The longer decorrelation timescales in the stratosphere result in enhanced prediction skill at subseasonal timescales in the stratosphere compared to the skill in the troposphere (Zhang, Shin, Dool, & Cai, 2013).

Table 2 and Figure 2 show the prediction skill (equation 2) at 50 and 500hPa (defined here by the ACC, see equation 2), characterizing the model predictability in the middle stratosphere and the middle troposphere, respectively. The ACC decreases more slowly in the stratosphere than in the troposphere. All the prediction systems, even those with a poor stratospheric representation, are able to capture the enhanced prediction skill in the stratosphere compared to the troposphere. The predictability limit is defined as

Model	NH						SH					
	Annual		DJF		JJA		Annual		DJF		JJA	
	50hPa	500hPa	50hPa	500hPa	50hPa	500hPa	50hPa	500hPa	50hPa	500hPa	50hPa	500hPa
BoM	*10.1*	*6.0*	*12.2*	*6.8*	*5.3*	*5.1*	*8.8*	5.7	*9.4*	*5.8*	*7.6*	5.7
CMA	*10.9*	*5.2*	*11.7*	*6.0*	7.4	*4.7*	*9.0*	*3.9*	11.1	*4.4*	*7.2*	*3.7*
ECCC	15.5	8.3	17.4	9.2	11.2	7.5	13.3	7.9	14.5	8.2	11.4	7.9
ECMWF$^{\times}$	**17.9**	**9.0**	**20.5**	**10.1**	**12.1**	**8.0**	**14.8**	**8.5**	**15.5**	**8.6**	**12.9**	**8.6**
CNR-ISAC	12.0	6.9	*12.9*	7.3	9.1	6.6	10.7	6.7	11.6	6.8	9.4	6.6
JMA$^{\times}$	16.4	8.5	18.3	9.5	11.8	7.7	13.1	7.9	12.5	7.8	11.1	7.9
CNRM-Meteo$^{\times}$	14.2	7.3	16.4	8.0	10.2	6.6	13.4	7.1	**15.0**	7.2	11.5	7.2
NCEP$^{\times}$	14.3	7.8	17.6	8.7	8.4	7.0	12.3	7.2	13.7	7.3	10.4	7.2
UKMO$^{\times}$	15.1	8.1	17.2	9.0	11.0	7.4	12.8	7.5	13.8	7.5	11.4	7.5
MMM	14.0±2.4	7.5±1.2	16.0±2.9	8.3±1.3	9.6±2.2	6.7±1.1	12.0±1.9	6.9±1.3	13.0±1.9	7.1±1.2	10.3±1.8	6.9±1.4

Table 2. Maximum forecast lead time (i.e., predictability limit in days) determined by the lead time when the ACC drops below 0.6, based on the period 1999-2010 for 30° - 90°N and S, respectively. Values that fall below one standard deviation of the MMM are italicized; values that fall above one standard deviation of the MMM are bolded. $\times$ indicates high-top models.

the day when the ACC drops below 0.6. In the troposphere, the daily ACC drops below 0.6 typically at lead times of 6-8 days in both hemispheres regardless of the season. In the stratosphere of both hemispheres, the predictability limit extends to 12 days or longer in DJF. Although the stratospheric predictability limit is shorter in boreal summer, it is still longer than tropospheric predictive timescales. The only exception is BoM in boreal summer which shows comparable prediction skills for the stratosphere and the troposphere. This is likely caused by an unrealistic stratosphere in this prediction system (Y. Lim, Son, Marshall, Hendon, & Seo, 2019). There is notable variation in the stratospheric prediction skill among the prediction systems, with those with little stratospheric variation such as BoM and CMA having reduced prediction skill as compared to the multi-model average. In particular, the average of the high-top models (indicated by $\times$) for DJF in the NH is 18 days, while it is 13.6 days for the low-top models. While evaluating these results it has to be kept in mind that the hemispheres are not fully symmetric. The enhanced persistence of stratospheric and tropospheric variability that can arise due to stratospheric events occurs during mid-winter (December to February) and spring (March to May) in the NH and during spring to early summer (October to December) for the SH (E. P. Lim et al., 2018; Simpson et al., 2011). The SH stratosphere in December - February (DJF) tends to be more predictable than its NH counterpart in June - August (JJA), likely due to the later break-up of the polar vortex in the SH, leading to enhanced predictability in the SH. On the other hand, the NH stratosphere in DJF is more predictable than its SH counterpart in JJA. One possible reason for this are the stronger remote influences in the Northern Hemisphere winter that affect the stratosphere in winter. For the stratosphere, models also often show strongly enhanced predictability for periods of weeks to months after extreme stratospheric events such as SSW events, which are absent in the SH stratosphere in JJA.

It is further found that the stratospheric prediction skill is highly correlated with tropospheric prediction skill. Figure 3 shows a scatter plot for the prediction skill shown in Figure 2 and Table 2. A significant linear relationship across nine prediction systems is found, indicating that the models with a better prediction skill in the stratosphere also exhibit a better tropospheric prediction skill. From this analysis it is however not possible to infer any causality. In particular, the available model data does not allow us to distinguish if the better tropospheric prediction of high-top models is indeed due to a better resolved stratosphere, which might improve tropospheric predictability, or if prediction systems with a higher stratospheric resolution also exhibit better tropospheric predictions due to a better representation of processes unrelated to the stratosphere, or a combination of both.

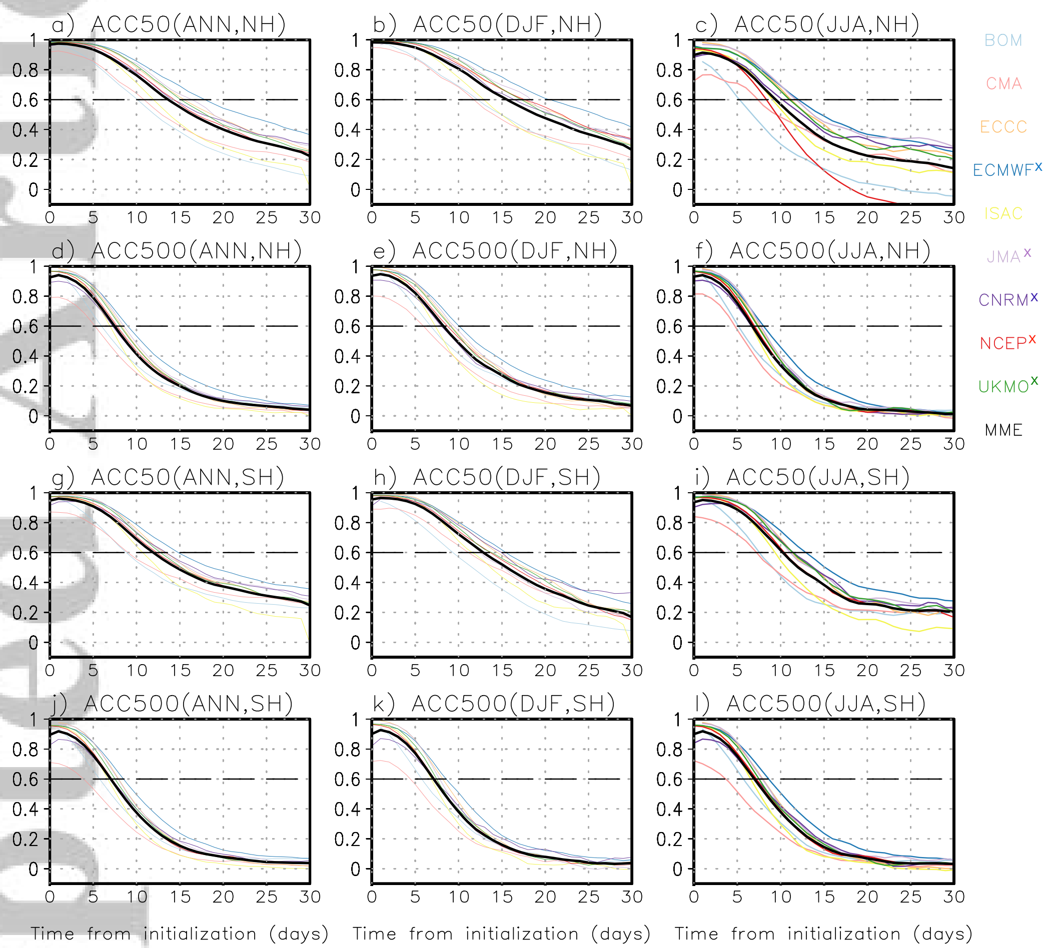


Figure 2. ACC for geopotential height for the area (a-f) north of 30°N and (g-l) south of 30°S. For both hemispheres, the ACC is examined at 50 hPa (a-c, g-i) and 500 hPa (d-f, j-l) as a function of lead time [days]. The results for JJA and DJF are plotted separately for the period common to all prediction systems. Different colors denote individual prediction systems and the black bold line indicates the multi-model mean, which is computed by averaging the ACC values of all prediction systems. '×' indicates high-top models.

While many prediction systems show appreciable skill in simulating large-scale NH winter stratospheric anomalies, they do so with a small signal-to-noise ratio (the so-called 'signal-to-noise paradox' (Scaife & Smith, 2018)). For the subseasonal prediction systems in the S2S database there is evidence that the same problem is also present, at least at lags beyond the limit of predictability in the troposphere. To diagnose signal-to-noise problems in the prediction systems, we examine the RPC diagnostic (Section 2.2, equation 3) and its behavior as a function of lead time and pressure level for the NH winter stratosphere (Fig. 4). For all systems, the RPC starts close to 1.0, indicating, as expected, no initial signal-to-noise problem, but the RPC then subsequently grows larger than 1.0, indicating under-confident forecasts and a signal-to-noise issue. In the troposphere, the speed of this growth and the ultimate level of RPC varies between the systems, but an onset at around 10-20 days is typical, leading to the RPC reaching values of about 1.5-3.0. Note this is similar to the level found at the seasonal timescale, and the positive values indicate under-confidence of the prediction systems (i.e., the prediction systems underestimate the predictability of the observations). In the stratosphere, the RPC is found to grow more slowly than in the troposphere. This is consistent with, but not obviously a result of, the higher predictive skill in the stratosphere. Despite the slower onset, the eventual values of the RPC attained in the stratosphere still tend to be large, in many systems equaling (e.g., CMA, NCEP) or exceeding (e.g. BoM) those reached in the troposphere. Other systems do not appear to be integrated sufficiently long for the signal-to-noise paradox to develop in the stratosphere, e.g., JMA.

Overall, the results show that all systems in the S2S project possess the signal-to-noise paradox as a feature of their predictions. Note that the skill derived in this section is possibly dependent on the ensemble size of the forecasting systems. This has e.g. been shown to yield a difference for the tropospheric winter circulation on seasonal timescales (Athanasiadis et al., 2017).

4 Predicting Stratospheric Events

We now turn to prediction on S2S timescales in the extratropical stratosphere. In particular, this section analyzes the predictability of stratospheric extreme events that can subsequently influence surface climate on S2S timescales, as discussed in Part II of this study.

Polar vortex events that influence surface climate include early and major mid-winter SSW events, strong vortex events, negative heat flux events, and final warming events. These extreme events, which are defined in section 2.3, have different characteristics and potentially different predictability. For example, for SSW events, anomalously large wave breaking is followed by strongly non-linear wave-mean flow interaction that can lead to quickly developing changes in the circulation. For strong vortex events, anomalously weak wave breaking gives way to slow radiative processes that slowly drive the circulation towards radiative equilibrium and hence a strong vortex. Negative heat flux events are associated with reflection (a reversible process), which is different from wave breaking (an irreversible process), and hence different predictability timescales could be expected.

Here we compare the predictability of these events during a common period 1996-2010. Five prediction systems (CMA, ECCC, ECMWF, JMA, and UKMO) were used in the analysis of all types of events for the NH to form the multi-model mean (black line in Fig. 5); additional modeling systems (BoM, CNR-ISAC, and CNRM-Meteo) were considered in some cases where data was available, but are not included in the multi-model mean. NCEP is not considered for this analysis as its period of hindcasts begins in 1999. Note that only 2 ensemble members from UKMO were available for some initialization dates at the time of data acquisition for this section. The data is first bias-corrected by

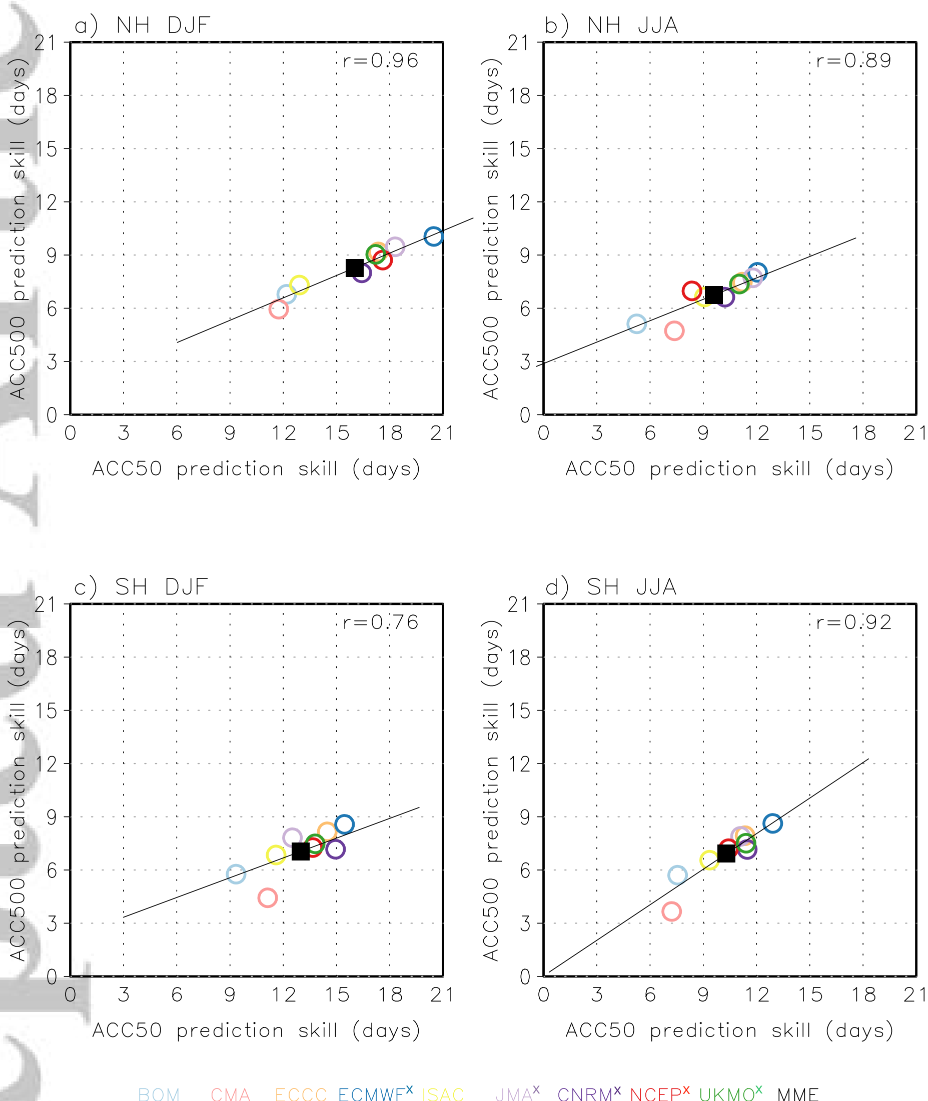


Figure 3. Scatter plot showing the predictability limit (the day for which the ACC crosses 0.6) of geopotential height (a-b) north of 30°N and (c-d) south of 30°S for each model at 50hPa vs. 500hPa for DJF (left) and JJA (right). The average for all prediction systems is shown as the black square. A linear fit to the data points is shown as the solid line. The correlation coefficient between the prediction skill at 50 hPa and 500 hPa is indicated in the upper-right corner of each panel. '×' indicates high-top models.

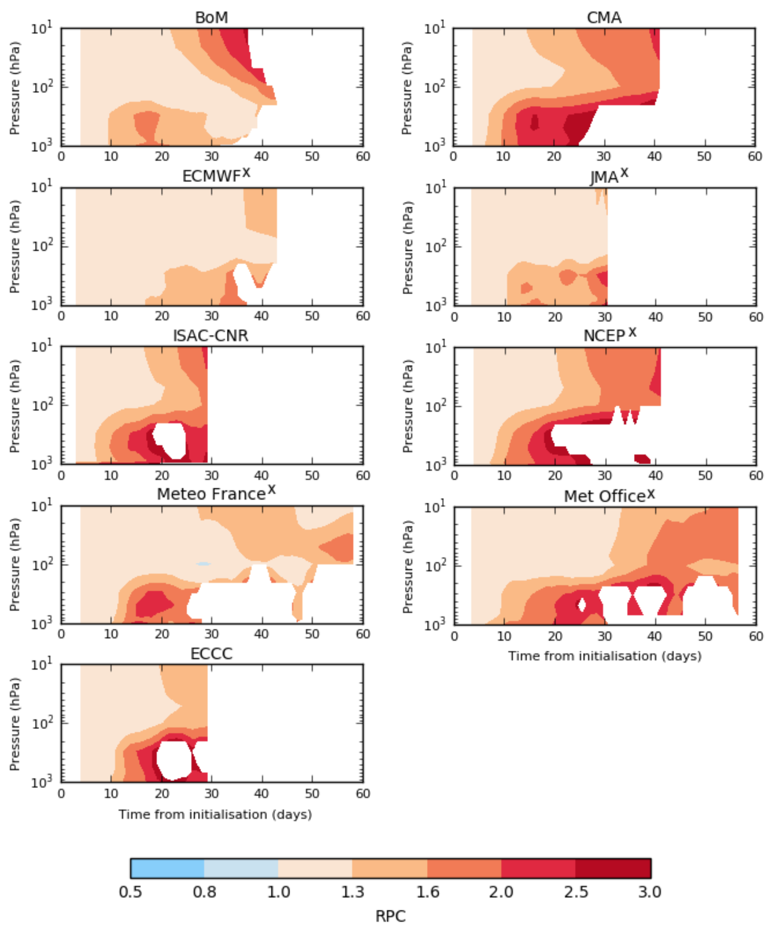


Figure 4. RPC (equation 3) for each prediction system as a function of lead time and height for DJF. Below 100 hPa the RPC is calculated for the zonal means of zonal wind at 60°N for the North Atlantic-European sector between 90°W and 60°E. Above 100hPa the same diagnostic calculated for the entire latitude circle is used. Before calculating the RPC, the data are aggregated into 7-day running means. These two aspects are necessary so that a reliable RPC can be obtained. As the correlation r and the ensemble mean become small, the RPC becomes ill defined, resulting in very noisy estimates. To avoid potentially misleading noise, the plot is masked where the correlation with observations is less than 0.2. For full zonal means at daily resolution the tropospheric correlation is always less than 0.2 after about 20 days, making it impossible to trace the growth of the RPC. '×' indicates high-top models.

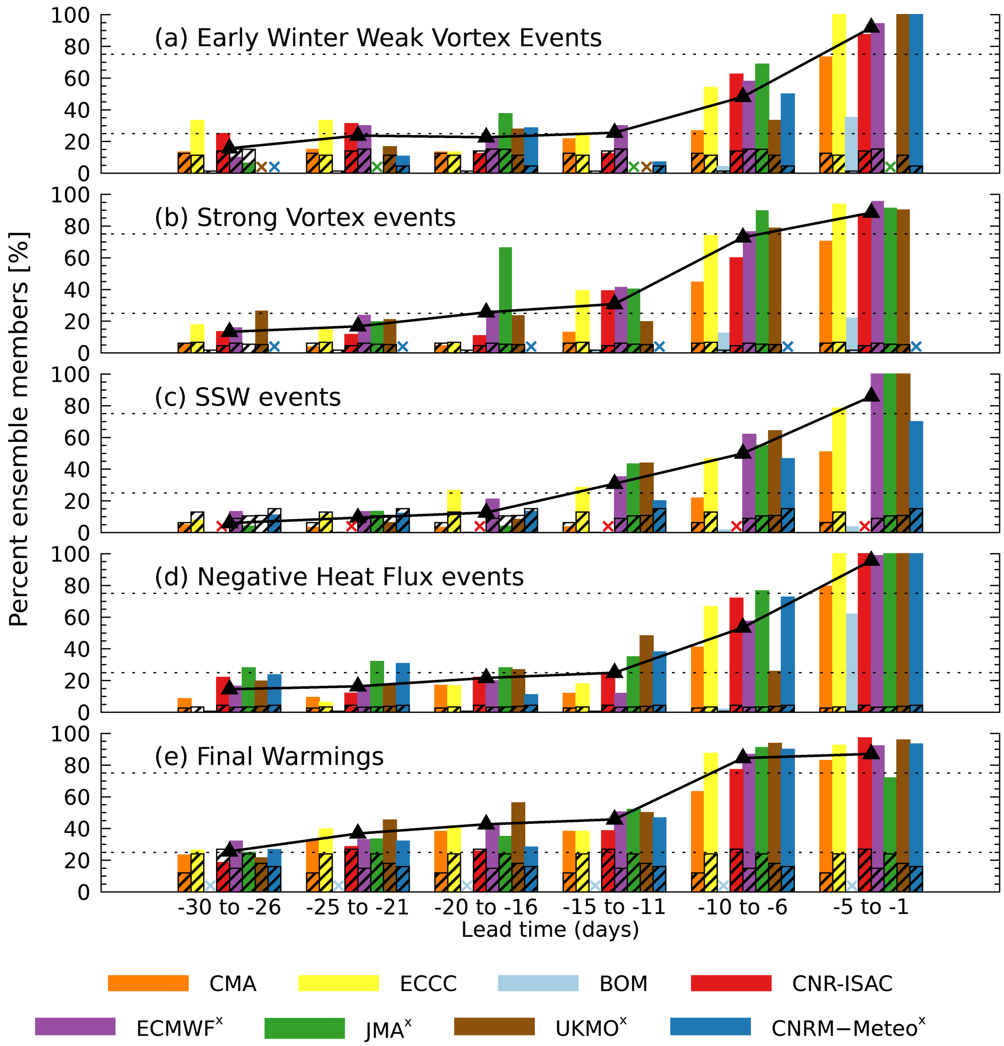


Figure 5. The average across all events of the percentage of ensemble members as a function of lead time [days] that detect the event within ± 3 days of the observed event for (a) early stratospheric warming events, (b) strong polar vortex events, (c) SSW events, (d) negative heat flux events, and (e) final warming events. The black line shows the multi-model mean based on 5 prediction systems (CMA, ECCC, ECMWF, JMA, and UKMO). Dotted lines show where 25% and 75% of ensemble members detect the event. '×' marks the high-top models in the legend. Where a prediction system was not used for the analysis or where there were not enough available ensemble members (at least 10 members were required for a given lead time range) is marked by an × in the color of the prediction system. Patterned black bars give the "false alarm rate" (events that were predicted but not detected at the given lead times).

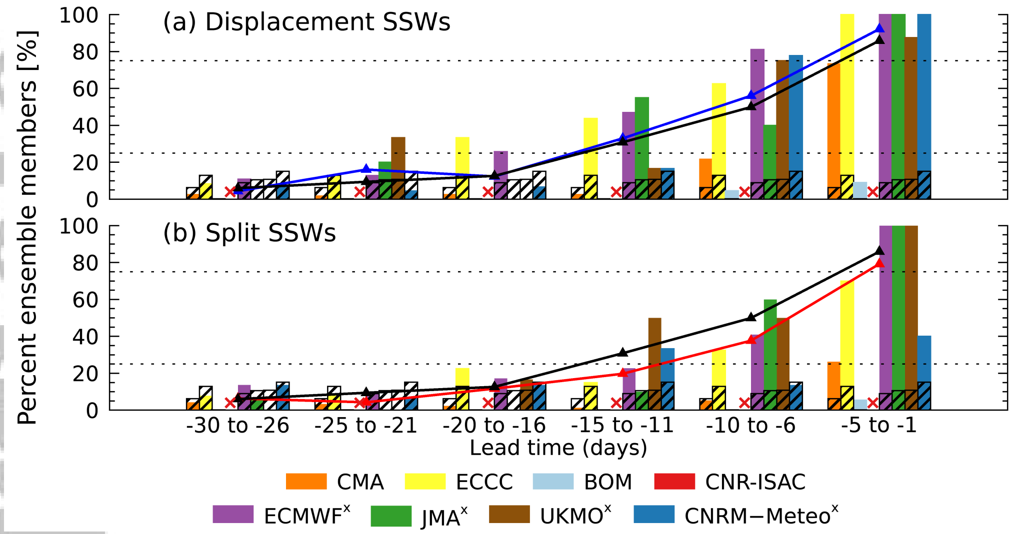


Figure 6. Same as Fig. 5 but for SSW events separated into (a) displacement and (b) split events. The black line corresponds to the multi-model mean from Figure 5c, the blue / red lines indicate the multi-model mean for the displayed events only. A student t-test of the differences between the detection of splits and displacements gives the following p-values for lead times from left to right: [0.6948,0.0279,0.7550,0.357,0.0925,0.3740]. The false alarm rates shown by the black patterned bars are for all SSW events, as in Fig. 5c.

removing the model climatology (leaving the year to be corrected out) and then adding back ERA-interim climatology. The bias-correction had the strongest influence on the detection of strong vortex and negative heat flux events at long-leads (not shown). In particular, after bias-correction, a smaller percentage of members across prediction systems detected strong vortex events at long lead times (suggesting an overestimation of these events in the model climatology), and a greater percentage of detected negative heat flux events at long lead times (suggesting an underestimation of these events in model climatology, in agreement with results from the the Coupled Model Intercomparison Project Phase 5 (CMIP5) models (Shaw et al., 2014, Fig. 5)).

Figure 5 shows the percentage of ensemble members for each prediction system that detects the observed event within $\pm$ 3 days of its actual date, for lead times averaged over 5-day periods prior to the event, which occurs on day 0. The bin length is chosen as a balance between having sufficient hindcasts in each bin for each event while resolving the lead times before each event. The "false alarm rate" is the percentage of members that predict an event to occur within a 1-30 day lead time when no event was observed. The comparison of the hit rate with the false alarm rate in Fig. 5 provides a measure of the predictive skill.

Below, we describe the differences in the predictability between the different types of polar vortex events. The results should be prefaced by a number of caveats: 1) not all prediction systems produce a hindcast in each time bin for each event; 2) the number of ensemble members varies across prediction systems; 3) the number of events is generally small, due to the short period covered by the hindcasts; 4) hindcast data from different model versions of a given model are sometimes used; 5) the $\pm$ 3-day window is an arbitrary choice which could matter for the accuracy in the detection of the events shown here; 6) the false alarm rates are used as a baseline for skill but the prediction systems could over- or underestimate these events, even after bias-correction; and 7) the percentage of ensemble members forecasting an event is only one metric for the assessment of predictability, and may be less reliable for models with a small number of ensemble members at a given lead time. Other skill evaluation techniques (such as in Karpechko (2018)) return similar but not identical results.

Four early winter weak vortex events events (one each in 1996, 2000, 2005, and 2009) are evaluated in the common S2S period. Each of these instances is associated with at least one ensemble member from the S2S hindcasts forecasting a major SSW in November, while other ensemble members miss the event entirely by forecasting vortex intensification. We find that fewer than 50% of ensemble members accurately detect early warming events prior to 6-10 days from the observed event, but almost all capture the event within 5 days (Figure 5a). The multi-model mean rises above the false alarm rate at lags up to 25 days from the event, suggesting some skill at longer leads. Two low-top systems, BoM and CMA, have difficulty predicting early winter weak vortex events even 5 days ahead of time, but two other low-top systems, ECCC and CNR-ISAC, perform similarly to high-top models at most lead times (and even slightly better at long lead times).

Accurate detection of strong polar vortex events (Figure 5b) becomes highly probable (i.e., greater than 75%) up to 10 days before the event. Two exceptions are BoM and CMA. CMA has, on average, relatively low probability (about 70%) of detection even at lead times less than 5 days before the events. BoM clearly has problems with forecasting a strong polar vortex event, which is likely due to a lack of stratospheric resolution in this model. JMA indicates the most skill at 6-20 day leads, but overall all systems (with the exception of BoM and CMA) perform similarly. At lead times longer than 15 days, the forecasted probability of detecting an event is between 5-60%, which typically exceeds the averaged 30-day lead time false alarm rates. The enhanced detection of the event relative to the false alarm rate may indicate some skill even at lead times of 30 days.

Previous studies (e.g., E. P. Gerber et al., 2009; Karpechko, 2018; Karpechko, Perez, Balmaseda, Tyrrell, & Vitart, 2018) have found predictability limits for major mid-winter SSWs of around 10-20 days. Here we find similar results for the S2S prediction systems (Figure 5c). While the percentage of ensemble members detecting an event does exceed false alarm rates at lead times of up to 15 days for most prediction systems, less than 10% of members detect SSW events at long leads (greater than 25 days), and predictions do not exceed 50% of members until lead times of 10 days or less. Even at lead times of 5 days, a few of the prediction systems (CMA, BoM, and CNRM-Meteo) show 80% or less of members detecting the observed SSW. These results generally agree with previous estimates of SSW deterministic predictability (Karpechko, 2018; Tripathi, Baldwin, et al., 2015), and indicate that predictability of such a major non-linear transition can be limited by both the predictability of Rossby wave propagation and their interaction with the stratospheric mean state (R. Plumb, 1981).

One more interesting implication of mid-winter SSW events is the *type* of SSW that occurs. In a common classification, there are two major types of mid-winter SSW events: (1) “split” events, for which the polar vortex splits into two separate vortices, and (2) “displacement” events, for which the polar vortex is distorted and displaced off the pole (e.g., Charlton & Polvani, 2007). Taguchi (2018) provides an analysis of the predictability in the S2S hindcasts of 5 SSW events (Dec 1998, Dec 2001, Jan 2009, Jan 2013 in the NH and Sep 2002 in the SH), showing that the vortex split SSWs (i.e., 2002, 2009, 2013) were more difficult to forecast than the displacements (1998, 2001). Here, we extend that analysis by considering the predictability of 11 NH mid-winter SSW events in ERA-Interim during the 1996-2010 period. A separate analysis separating split and displacement events for this larger number of events, i.e., 6 displacements and 5 split events (Figure 6), confirms the results from Taguchi (2018), that is, that displacement events tend to be more predictable than split events, especially at lead times of 1-2 weeks, though given the limited number of events this difference has limited statistical significance. While this points to potentially different mechanisms in the precursors and causes of these events (e.g. D. I. V. Domeisen, Martius, & Jiménez-Esteve, 2018; Esler & Matthewman, 2011; Martius, Polvani, & Davies, 2009; Matthewman & Esler, 2011), it will have to be further investigated if this difference is indeed robust and what the reasons for these differences are.

Next, we consider the predictability of negative eddy heat flux events (Figure 5d). Mukougawa, Noguchi, Kuroda, Mizuta, and Kodera (2017) used an ensemble forecast model to show that the predictive lead time of a March 2007 negative heat flux event was one week. Extending the analysis to multiple extreme negative stratospheric heat flux events, here we find that the multi-model mean exhibits predictive skill at lead times of up to 30 days. The performance again varies between prediction systems, with JMA and CNRM-Meteo showing the highest skill at long leads, and BOM and CMA showing weaker skill at most leads.

Finally, we find that the predictability of final warmings is higher for longer lead times compared to other events (Figure 5e). However the false alarm rate is also larger than for other events since the prediction systems climatologically must predict a final warming every year. The detection rate rises above the false alarm rate at lead times of up to 25 days. Note also that this particular period (1996-2010) comprises 10 “late” (i.e., after April 15th) final warmings and only 4 “early” (i.e., before April 15th) final warmings. This is relevant since late final warmings are more predictable at longer lead times than early, dynamically-driven final warmings, which show predictability more similar to mid-winter SSW events (A. Butler, Charlton-Perez, Domeisen, Simpson, & Sjoberg, 2019).

We now perform the same analysis for the SH to obtain the model skill for predicting the timing of the final stratospheric warming events in the SH using the same approach as for the NH discussed above. In the SH, the maximum variability of the polar vortex is found in spring in the upper stratosphere when the stratospheric polar night jet seasonally weakens and becomes more susceptible to wave forcing from the troposphere (Byrne & Shepherd, 2018; Graversen, RG & Christiansen, B, 2003; Kuroda & Kodera, 1998; E. P. Lim et al., 2018; Randel, W, 1988; Sheshadri, A & Plumb, R A, 2016; Shiotani & Hirota, 1985; Thompson & Wallace, 2000). Anomalous weakening and warming (strengthening and cooling) of the SH spring polar vortex generally leads to an earlier (later) final warming event (Byrne & Shepherd, 2018; Shiotani, Shimoda, & Hirota, 1993).

Figure 7 assesses the skill of the sub-seasonal forecasting systems in predicting final warming events in the SH. All models show skill (relative to the false alarm rate at these leads, given in black bars), even out to lead times of 30 days. As for the NH, the high-top models tend to show the highest skill, though it is notable that several low-top models such as CNR-ISAC and ECCC show significant skill for all lead times. In comparison to the NH final warmings, the false alarm rates tend to be smaller in the SH, and predictability (the percentage of ensemble members predicting the correct date in comparison to the false alarm rate) can be found for longer lead times: while in the NH, the prediction rate falls below the false alarm rate as early as at lead times of 16 to 20 days before the event for several models, this is not the case for any model in the SH out to 30 days before the final warming event. The multi-model mean predictability is similar to the NH, though it decays faster for lead times of 6 to 10 days, while it remains high for these lead times in the NH. Overall, this indicates a higher predictability of the final warming events at short lead times for the NH, but higher predictability for long lead times of 3-4 weeks for the SH. The predictability at longer lead times in the SH might arise due to the smaller variability in the timing of the SH final warming compared to the NH, despite the observed trend in the timing of the final warming due to ozone variability and trends (R. X. Black & McDaniel, 2007; Sheshadri, A & Plumb, R A, 2016; Thompson et al., 2011). Given that almost all models use non-interactive or climatological ozone, this demonstrated forecast skill to predict the timing of the SH final warm-

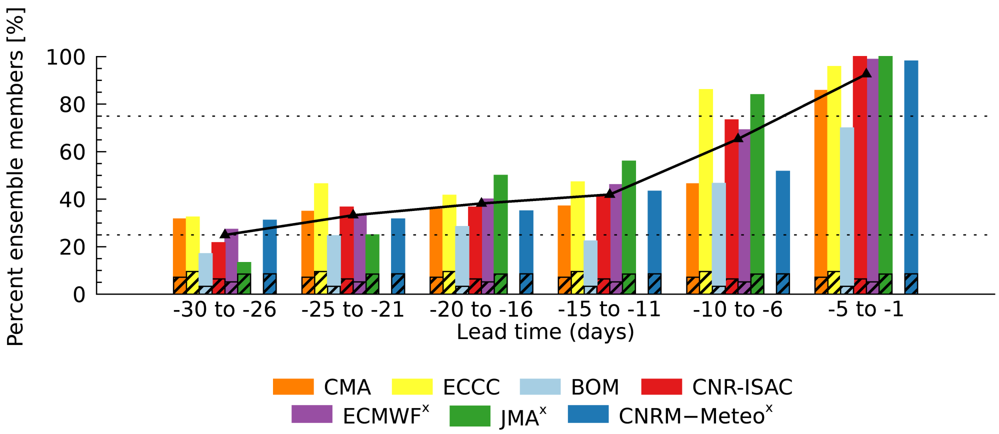


Figure 7. Same as Fig. 5e but for final warming events in the Southern Hemisphere. The false alarm rates are shown by the black patterned bars. The black line shows the multi-model average over all prediction systems displayed here.

ing indicates that dynamical processes are the dominant drivers of predictability for the final warming, but there is scope for further improvement of forecast skill by including prognostic ozone (e.g. Seviour et al., 2014).

While it is difficult to directly compare the predictability of different types of events, given the differences in the number of events and their time of occurrence in each case, in general we can conclude the following:

(a) Models with poorer stratospheric resolution or a low model top such as e.g. CMA and BoM show a weaker performance in predicting stratospheric events. Note that BoM's top level below the model lid is at 10 hPa, so using metrics based on 10 hPa output may not be physically meaningful for this prediction system because of strong damping of wave driven processes by the deep sponge layer. However, ECCC, despite its low model top (see Figure 1), has a predictability of stratospheric events that is comparable to models with a well-resolved stratosphere.

(b) The probability of accurately detecting the observed event increases as lead time decreases, and becomes large (greater than 75%) at lead times of up to 10 days before the events. The probability of accurately detecting the observed event has less dependence on lead time between 30 and 15 days before the event. For these lead times, forecast probability is between 5-50%, with some types of events exhibiting longer-lead predictability than others. Strong vortex events and final warmings appear somewhat more predictable at longer leads than SSW events, which hints at the different mechanisms causing these events. The lower predictability of SSW events is likely linked to their more dynamical and wave-driven nature, while more gradual and/or radiatively driven processes, e.g. strong vortex or late final warming events, tend to be more predictable (A. Butler et al., 2019). While we here provide a first look at the overall predictability of these events in the S2S database, more work will have to be done to fully understand the factors that drive some events to be more predictable than others.

5 Discussion and Outlook

In this study, we have examined the predictability in the stratosphere using the sub-seasonal prediction systems from the S2S database (Vitart et al., 2017). These systems provide important operational guidance for prediction on S2S timescales, so it is important to understand the processes that give rise to predictability, including those that involve the stratosphere. This study focuses on evaluating the predictability of the stratosphere itself, as extreme events in the stratosphere can have significant impacts on the predictability of surface weather, which is investigated in Part II of this study (D. I. Domeisen et al., 2019).

Overall, the stratosphere exhibits longer predictability timescales as compared to the troposphere, as exemplified by the slower decrease in the prediction skill in comparison to the troposphere. For most models, predictability beyond two weeks is typical in the stratosphere. In addition, the stratosphere exhibits a slower growth of the signal-to-noise problem as compared to the troposphere. The stratosphere also exhibits a range of extreme events, however, stratospheric extreme events themselves tend not to be predictable beyond deterministic timescales and exhibit similar predictability to tropospheric weather. This is in particular the case for sudden stratospheric warming events, which are predicted by up to 50% of the ensemble members in all models out to only about a week. Events that are less abrupt in nature, such as late final warming events and strong vortex events tend to be more predictable, with up to 50% of the ensemble members predicting the occurrence of the event 2 weeks in advance (see also: A. Butler et al., 2019). Final warming events in the SH tend to be more predictable than those in the NH.

Due to the limited representation of ozone on the S2S models, it is not possible to assess the role of ozone on predictability using the current set of models. Given the possible influence of ozone on the dynamical evolution of the stratosphere in both hemispheres (Ivy et al., 2017; Ivy, Solomon, & Rieder, 2016; Keeble, Braesicke, Abraham, Roscoe, & Pyle, 2014; Rieder, Chiodo, Fritzer, Wienerroither, & Polvani, 2019; Seviour et al., 2014; Solomon, Haskins, Ivy, & Min, 2014), an improved representation of stratospheric ozone might further increase the predictability of the stratosphere on sub-seasonal and longer time scales. Significant differences can be found in the predictability of stratospheric events between high-top and low-top models, with the high-top models exhibiting significantly higher predictability of stratospheric extreme events as compared to low-top models. Note that here, high-top refers to models with both a high model top and an improved stratospheric resolution.

It should be noted that the estimates of skill in the prediction of various parameters in this study are dependent on the frequency and ensemble size of the hindcasts in the S2S database. Ensemble size has been shown to have a marked influence on the skill of ensemble forecasting of the mid-latitude winter circulation (e.g. Athanasiadis et al., 2017), with larger ensembles tending to be more skillful. Operational requirements within the centres contributing to the S2S dataset frequently mean that hindcast ensemble sizes are considerably smaller than those of operational forecasts. As a result, when the same systems are used to produce forecasts in real-time, they may have levels of skill that exceed those shown here. It might be reasonable to assume, therefore, that the skill shown here is a lower limit for the skill of real-time operational forecasts. In a similar way, our results cannot be used to infer the relative performance of the underlying models within the prediction systems, as any differences in skill measures may be a result of differences in their ensemble size and initialisation strategy rather than the model itself.

Overall, this study shows a clear dependence of S2S prediction skill on the season and the type of extreme event in the stratosphere for all models. In addition, a clear difference in predictability between high-top and low-top models can be observed, with a significantly better prediction of stratospheric extreme events in high-top models. While this study provides an overview of the prediction skill available in the S2S database, further detailed studies of S2S prediction skill for the stratosphere will be necessary in order to assess the full range of stratospheric predictability, especially with further stratospheric data becoming available in future versions of the S2S database.

References

Albers, J., & Birner, T. (2014). Vortex preconditioning due to planetary and gravity waves prior to sudden stratospheric warmings. *Journal of Atmospheric Sciences*, *71*, 4028–4054. doi: 10.1175/JAS-D-14-0026.1

Athanasiadis, P. J., Bellucci, A., Scaife, A. A., Hermanson, L., Borrelli, A., MacLachlan, C., … Gualdi, S. (2017). A Multisystem View of Wintertime NAO Seasonal Predictions. *Journal of Climate*, *30*(4), 1461–1475.

Ayarzagüena, B., Langematz, U., & Serrano, E. (2011). Tropospheric forcing of the stratosphere: A comparative study of the two different major stratospheric warmings in 2009 and 2010. *Journal of Geophysical Research*, *116*(D18), D18114.

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*.

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. J., Dunkerton, T. J., Charlton, A. J., & O'Neill, A. (2003). Stratospheric memory and skill of extended-range weather forecasts. *Science*, *301*, 636–640.

Bandoro, J., Solomon, S., Donohoe, A., Thompson, D. W. J., & Santer, B. D. (2014). Influences of the Antarctic Ozone Hole on Southern Hemispheric Summer Climate Change. *Journal of Climate*, *27*(16), 6245–6264.

Black, R., McDaniel, B., & Robinson, W. A. (2006). Stratosphere-troposphere coupling during spring onset. *Journal of Climate*, *19*, 4891–4901.

Black, R. X., & McDaniel, B. A. (2007). Interannual variability in the Southern Hemisphere circulation organized by stratospheric final warming events. *Journal of the Atmospheric Sciences*, *64*(8), 2968–2974.

Butler, A., Charlton-Perez, A., Domeisen, D., Simpson, I., & Sjoberg, J. (2019). Predictability of Northern Hemisphere final stratospheric warmings and their surface impacts. *Geophysical Research Letters*, *43*. doi: 10.1029/2019GL083346

Butler, A. H., Arribas, A., Athanassiadou, M., Baehr, J., Calvo, N., Charlton-Perez, A., … Yasuda, T. (2016). The Climate-system Historical Forecast Project: Do stratosphere-resolving models make better seasonal climate predictions in boreal winter? *Quarterly Journal of the Royal Meteorological Society*, *142*, 1413–1427.

Butler, A. H., & Gerber, E. P. (2018). Optimizing the definition of a sudden stratospheric warming. *Journal of Climate*, *31*(6), 2337-2344.

Butler, A. H., Seidel, D. J., Hardiman, S. C., Butchart, N., Birner, T., & Match, A. (2015). Defining sudden stratospheric warmings. *Bull. Amer. Meteor. Soc.*, 1–16.

Butler, A. H., Sjoberg, J. P., Seidel, D. J., & Rosenlof, K. H. (2017). A sudden stratospheric warming compendium. *Earth Syst. Sci. Data*, *9*, 63–76.

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.

Charlton, A., O'Neill, A., Lahoz, W., & Berrisford, P. (2005). The splitting of the stratospheric polar vortex in the Southern Hemisphere, September 2002: Dynamical evolution. *Journal of the Atmospheric Sciences: Special Issue on the Southern Hemisphere Sudden Stratospheric Warming of 2002*, *62*, 590–602.

Charlton, A., & Polvani, L. M. (2007). A new look at stratospheric sudden warm-

ings. Part I: Climatology and modeling benchmarks. *Journal of Climate*, *20*(3), 449–469.
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.
Charlton-Perez, A. J., & O'Neill, A. (2010). On the Sensitivity of Annular Mode Dynamics to Stratospheric Radiative Time Scales. *Journal of Climate*, *23*(2), 476–484.
Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., ... Vitart, F. (2011). The ERAInterim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*, 553–597.
Domeisen, D. I., Butler, A. H., Charlton-Perez, A. J., Ayarzagüena, B., Baldwin, M. P., Dunn-Sigouin, E., ... Taguchi, M. (2019). The role of the stratosphere in subseasonal to seasonal prediction. Part II: Predictability arising from stratosphere-troposphere coupling. *Journal of Geophysical Research: Atmospheres*.
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*. doi: http://doi.org/10.1029/2018JD030077
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*(4), 2064–2071. doi: http://doi.org/10.1002/2017GL076886
Dunn-Sigouin, E., & Shaw, T. (2018). Dynamics of Extreme Stratospheric Negative Heat Flux Events in an Idealized Model. *Journal of the Atmospheric Sciences*, *75*(10), 3521–3540.
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.
Eade, R., Smith, D., Scaife, A., Wallace, E., Dunstone, N., Hermanson, L., & Robinson, N. (2014). Do seasonal-to-decadal climate predictions underestimate the predictability of the real world? *Geophysical Research Letters*, *41*(15), 5620–5628.
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*, 2505–2523.
Gerber, E., Baldwin, M. P., Akiyoshi, H., Austin, J., Bekki, S., Braesicke, P., ... Dhomse, S. (2010). Stratosphere-troposphere coupling and annular mode variability in chemistry-climate models. *Journal of Geophysical Research: Atmospheres (1984–2012)*, *115*(D3).
Gerber, E., Polvani, L., & Ancukiewicz, D. (2008). Annular mode time scales in the Intergovernmental Panel on Climate Change Fourth Assessment Report models. *Geophysical Research Letters*.
Gerber, E. P., & Martineau, P. (2018). Quantifying the variability of the annular modes: reanalysis uncertainty vs. sampling uncertainty. *Atmospheric Chemistry And Physics*, *18*(23), 17099–17117.
Gerber, E. P., Orbe, C., & Polvani, L. M. (2009). Stratospheric influence on the tropospheric circulation revealed by idealized ensemble forecasts. *Geophysical Research Letters*, *36*, L24801, doi:10.1029/2009GL040913.
Graversen, RG, & Christiansen, B. (2003). Downward propagation from the stratosphere to the troposphere: A comparison of the two hemispheres. *Journal of Geophysical Research*, *108*, 4780.
Hardiman, S. C., Butchart, N., Charlton-Perez, A. J., Shaw, T. A., Akiyoshi, H.,

Baumgaertner, A., … Shibata, K. (2011). Improved predictability of the troposphere using stratospheric final warmings. *Journal of Geophysical Research*, *116*(D18), 6313.

Hitchcock, P., Shepherd, T. G., Yoden, S., Taguchi, M., & Noguchi, S. (2013). Lower-Stratospheric Radiative Damping and Polar-Night Jet Oscillation Events. *Journal of the Atmospheric Sciences*, *70*(5), 1391–1408.

Hu, J. G., Ren, R. C., & Xu, H. M. (2014). Occurrence of Winter Stratospheric Sudden Warming Events and the Seasonal Timing of Spring Stratospheric Final Warming. *Journal of Atmos. Sci.*, *71*, 23192334.

Hu, J. G., Ren, R. C., Yu, Y. Y., & Xu, H. M. (2014). The boreal spring stratospheric final warming and its interannual and interdecadal variability. *Sci. China Earth Sci.*, *57*, 710-718.

Ivy, D. J., Hilgenbrink, C., Kinnison, D., Plumb, R. A., Sheshadri, A., Solomon, S., … Hilgenbrink, C. (2017). Observed Changes in the Southern Hemispheric Circulation in May. *Journal of Climate*, *30*(2), 527–536.

Ivy, D. J., Solomon, S., & Rieder, H. E. (2016). Radiative and Dynamical Influences on Polar Stratospheric Temperature Trends. *dx.doi.org*, *29*(13), 4927–4938.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., … Joseph, D. (1998). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, *77*, 437–470.

Karpechko, A. Y. (2018). Predictability of Sudden Stratospheric Warmings in the ECMWF Extended-Range Forecast System. *Monthly Weather Review*, *146*(4), 1063–1075.

Karpechko, A. Y., Hitchcock, P., Peters, D. H. W., & Schneidereit, A. (2017). Predictability of downward propagation of major sudden stratospheric warmings. *Quarterly Journal of the Royal Meteorological Society*, *104*, 30937.

Karpechko, A. Y., Perez, A. C., Balmaseda, M., Tyrrell, N., & Vitart, F. (2018). Predicting Sudden Stratospheric Warming 2018 and its Climate Impacts with a Multi-Model Ensemble. *Geophysical Research Letters*, 2018GL081091.

Keeble, J., Braesicke, P., Abraham, N. L., Roscoe, H. K., & Pyle, J. A. (2014, December). The impact of polar stratospheric ozone loss on Southern Hemisphere stratospheric circulation and climate. *Atmospheric Chemistry And Physics*, *14*(24), 13705–13717.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., … Takahashi, K. (2015). The JRA-55 reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan. Ser. II*, *93*(1), 5-48. doi: 10.2151/jmsj.2015-001

Kuroda, Y., & Kodera, K. (1998). Interannual variability in the troposphere and stratosphere of the southern hemisphere winter. *Journal of Geophysical Research*, *103*(D12), 13787.

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*(21), 12,002–12,016.

Lim, Y., Son, S.-W., Marshall, A., Hendon, H. H., & Seo, K.-H. (2019). Influence of the QBO on MJO prediction skill in the subseasonal-to-seasonal prediction models. *Climate Dynamics*. doi: https://doi.org/10.1007/s00382-019-04719-y

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

Limpasuvan, V, Thompson, D, & Hartmann, D L. (2004). The life cycle of the Northern Hemisphere sudden stratospheric warmings. *Journal of the Atmospheric Sciences*, *17*, 2584–2596.

Long, C. S., Fujiwara, M., Davis, S., Mitchell, D. M., & Wright, C. J. (2017). Climatology and interannual variability of dynamic variables in multiple reanalyses evaluated by the SPARC Reanalysis Intercomparison Project (S-RIP). *Atmospheric Chemistry and Physics*, *17*(23), 14593–14629.

Lubis, S. W., Matthes, K., Omrani, N.-E., Harnik, N., & Wahl, S. (2016). Influence of the Quasi-Biennial Oscillation and Sea Surface Temperature Variability on Downward Wave Coupling in the Northern Hemisphere. *J. Atmos. Sci.*, *73*(5), 1943–1965.

Marshall, A., & Scaife, A. A. (2010). Improved predictability of stratospheric sudden warming events in an atmospheric general circulation model with enhanced stratospheric resolution. *Journal of Geophysical Research*, *115*, D16114, doi:10.1029/2009JD012643.

Marshall, A. G., Hudson, D., Wheeler, M. C., Hendon, H. H., & Alves, O. (2011). Simulation and prediction of the Southern Annular Mode and its influence on Australian intra-seasonal climate in POAMA. *Climate Dynamics*, *38*(11-12), 2483–2502.

Martius, O., Polvani, L. M., & Davies, H. (2009). Blocking precursors to stratospheric sudden warming events. *Geophysical Research Letters*, *36*, L14806.

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*, 2481–2504.

Mukougawa, H., Noguchi, S., Kuroda, Y., Mizuta, R., & Kodera, K. (2017). Dynamics and Predictability of Downward-Propagating Stratospheric Planetary Waves Observed in March 2007. *Journal of the Atmospheric Sciences*, *74*(11), 3533–3550.

Newman, P. A., & Nash, E. R. (2005). The unusual Southern Hemisphere stratosphere winter of 2002. *Journal of the Atmospheric Sciences*, *62*, 614–628.

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

O'Reilly, C. H., Weisheimer, A., Woollings, T., Gray, L., & MacLeod, D. (2018). The importance of stratospheric initial conditions for winter North Atlantic Oscillation predictability and implications for the signal-to-noise paradox. *Quarterly Journal of the Royal Meteorological Society*.

Perlwitz, J., & Harnik, N. (2003). Observational evidence of a stratospheric influence on the troposphere by planetary wave reflection. *Journal of Climate*, *16*, 3011–3026. doi: 10.1175/1520-0442(2003)016⟨3011:OEOASI⟩2.0.CO;2

Plumb, R. (1981). Instability of the distorted polar night vortex: A theory of stratospheric warmings. *Journal of the Atmospheric Sciences*, *38*(11), 2514–2531.

Plumb, R. A. (1989). On the Seasonal Cycle of Stratospheric Planetary Waves. *PAGEOPH*, *130*(2/3), 233–242.

Randel, W. (1988). The seasonal evolution of planetary waves in the Southern Hemisphere stratosphere and troposphere. *Quarterly Journal of the Royal Meteorological Society*, *114*(484), 1385–1409.

Rieder, H. E., Chiodo, G., Fritzer, J., Wienerroither, C., & Polvani, L. M. (2019). Is interactive ozone chemistry important to represent polar cap stratospheric temperature variability in Earth-System Models? *Environmental Research Letters*, *14*(4).

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., ... Goldberg, M. (2010). The Ncep Climate Forecast System Reanalysis. *Bulletin of the American Meteorological Society*, *91*(8), 1015–1057.

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

Scaife, A. A., & Smith, D. (2018). A signal-to-noise paradox in climate science. *npj Climate and Atmospheric Science*, *1*(1), 130.

Seviour, W. J. M., Hardiman, S. C., Gray, L. J., Butchart, N., MacLachlan, C., Scaife, A. A., & Seviour, W. J. M. (2014). Skillful Seasonal Prediction of the Southern Annular Mode and Antarctic Ozone. *Journal of Climate*, *27*(19), 7462–7474.
Shaw, T. A., & Perlwitz, J. (2013). The Life Cycle of Northern Hemisphere Downward Wave Coupling between the Stratosphere and Troposphere. *Journal of Climate*, *26*(5), 1745–1763.
Shaw, T. A., Perlwitz, J., & Weiner, O. (2014). Troposphere-stratosphere coupling: Links to North Atlantic weather and climate, including their representation in CMIP5 models. *Journal of Geophysical Research Atmospheres*, *119*, 5864–5880.
Sheshadri, A, & Plumb, R A. (2016). Sensitivity of the surface responses of an idealized AGCM to the timing of imposed ozone depletion-like polar stratospheric cooling. *Geophysical Research Letters*, *43*(5), 2330–2336.
Shiotani, M., & Hirota, I. (1985). Planetary wave-mean flow interaction in the stratosphere: A comparison between northern and southern hemispheres. *Quarterly Journal of the Royal Meteorological Society*, *111*(468), 309–334.
Shiotani, M., Shimoda, N., & Hirota, I. (1993). Interannual variability of the stratospheric circulation in the southern hemisphere. *Quarterly Journal of the Royal Meteorological Society*, *119*(511), 531–546.
Simpson, I. R., Hitchcock, P., Shepherd, T. G., & Scinocca, J. F. (2011). Stratospheric variability and tropospheric annular mode timescales. *Geophys. Res. Lett.*, *38*, L20806.
Solomon, S., Haskins, J., Ivy, D. J., & Min, F. (2014). Fundamental differences between Arctic and Antarctic ozone depletion. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(17), 6220–6225.
Son, S.-W., Purich, A., Hendon, H. H., Kim, B.-M., & Polvani, L. M. (2013). Improved seasonal forecast using ozone hole variability? *Geophysical Research Letters*, *40*(23), 6231–6235.
Taguchi, M. (2018). Comparison of Subseasonal-to-Seasonal Model Forecasts for Major Stratospheric Sudden Warmings. *Journal of Geophysical Research-Atmospheres*, *123*(18), 10,231–10,247.
Taguchi, Masakazu. (2014). Predictability of Major Stratospheric Sudden Warmings of the Vortex Split Type: Case Study of the 2002 Southern Event and the 2009 and 1989 Northern Events. *J. Atmos. Sci.*, *71*(8), 2886–2904.
Taguchi, Masakazu. (2016). Connection of predictability of major stratospheric sudden warmings to polar vortex geometry. *Atmospheric Science Letters*, *17*(1), 33–38.
Thompson, D. W. J., & Solomon, S. (2005). Recent stratospheric climate trends as evidenced in radiosonde data: Global structure and tropospheric linkages. *Journal of Climate*, *18*, 4785–4795.
Thompson, D. W. J., Solomon, S., Kushner, P. J., England, M. H., Grise, K. M., & Karoly, D. J. (2011). Signatures of the Antarctic ozone hole in Southern Hemisphere surface climate change. *Nature Geoscience*, *4*(11), 741–749.
Thompson, D. W. J., & Wallace, J. M. (2000). Annular Modes in the Extratropical Circulation. Part I: Month-to-Month Variability. *Journal of Climate*, *13*(5), 1000–1016.
Tripathi, O. P., Baldwin, M. P., Charlton-Perez, A., Charron, M., Cheung, J. C., Eckermann, S. D., ... Stockdale, T. (2016). Examining the predictability of the Stratospheric Sudden Warming of January 2013 using multiple NWP systems. *Monthly Weather Review*, *144*, 1935–1960.
Tripathi, O. P., Baldwin, M. P., Charlton-Perez, A., Charron, M., Eckermann, S. D., Gerber, E., ... Son, S.-W. (2015). 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*(689),

987–1003.
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), 1–8.
Tyrrell, N. L., Karpechko, A. Y., Uotila, P., & Vihma, T. (2019, March). Atmospheric Circulation Response to Anomalous Siberian Forcing in October 2016 and its Long-Range Predictability. *Geophysical Research Letters*, *104*(D24), 30,937–11.
Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., ... Zhang, L. (2017). The Subseasonal to Seasonal (S2S) Prediction Project Database. *Bulletin of the American Meteorological Society*, *98*(1), 163–173.
Zhang, Q., Shin, C.-S., Dool, H., & Cai, M. (2013). CFSv2 prediction skill of stratospheric temperature anomalies. *Climate Dynamics*, *41*(7-8), 2231–2249.

Acknowledgments

The S2S model data was obtained from the ECMWF data portal at https://apps.ecmwf.int/datasets/data/s2s/. The ERA-Interim Reanalysis data was obtained from the ECMWF data portal at https://apps.ecmwf.int/datasets/data/interim-full-daily/. This work was initiated by the Stratospheric Network for the Assessment of Predictability (SNAP), an activity of SPARC within the World Climate Research Programme (WCRP). We acknowledge the scientific guidance of the WCRP to motivate this work, coordinated in the framework of SPARC.

Funding by the Swiss National Science Foundation to D.D. through project PP00P2_170523 is gratefully acknowledged. B.A. was funded by "Ayudas para la contratación de personal postdoctoral en formación en docencia e investigación en departamentos de la UCM" from Universidad Complutense de Madrid. C.I.G and C.S. were supported by a European Research Council starting Grant under the European Unions Horizon 2020 research and innovation programme (Grant agreement no. 677756). A.Y.K. was funded by the Academy of Finland (grants #286298 and #319397). The work by M.T. was supported by the JSPS Grant-in-Aid for Scientific Research (C) 15K05286. A.L.L. contributed as part of the NOAA/MAPP S2S Prediction Task Force and was supported by NOAA Grant NA16OAR4310068 and NSF Award 1547814. S.S. was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (Ministry of Science and ICT) (2017R1E1A1A01074889).

Figure.

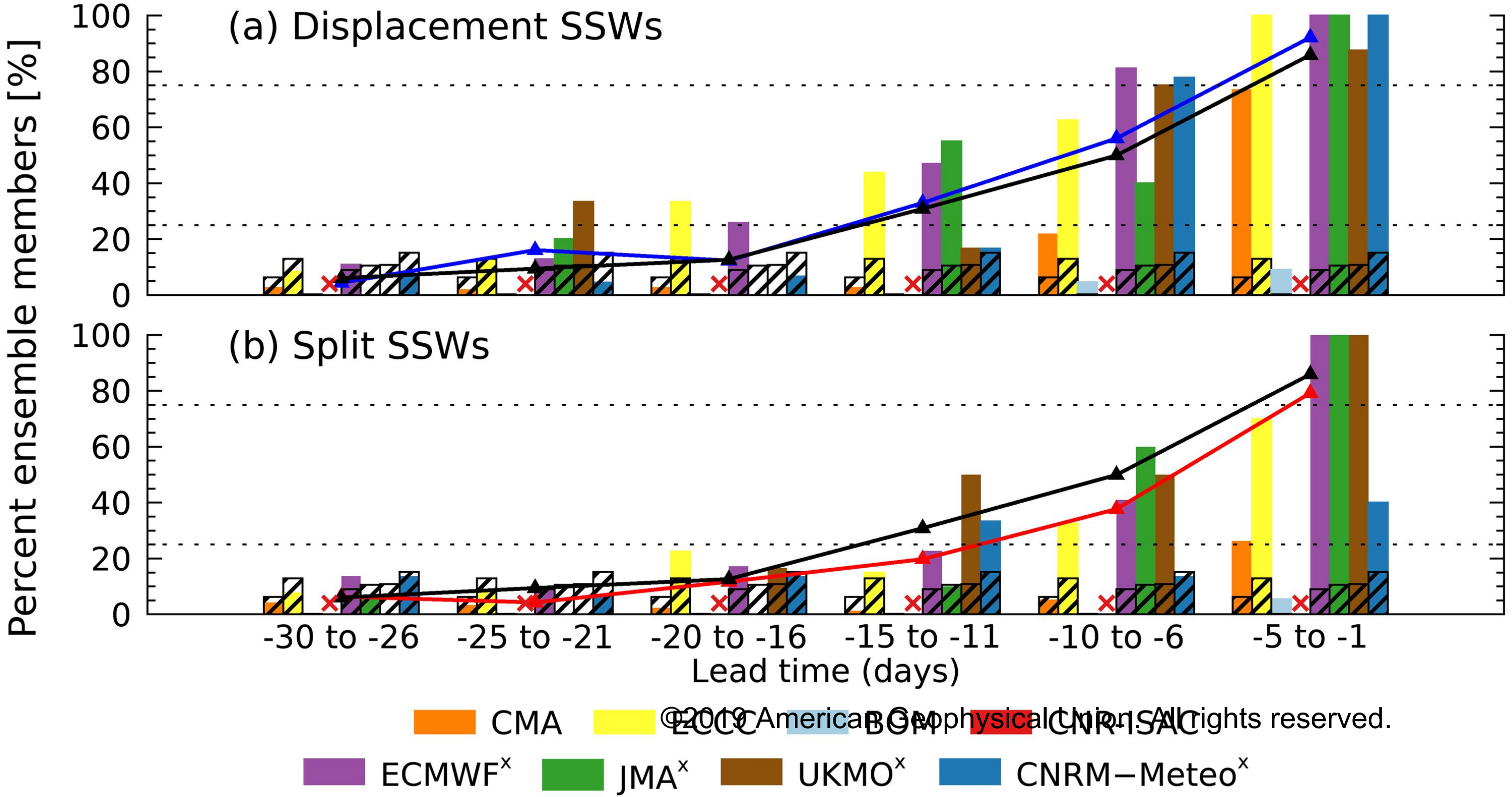
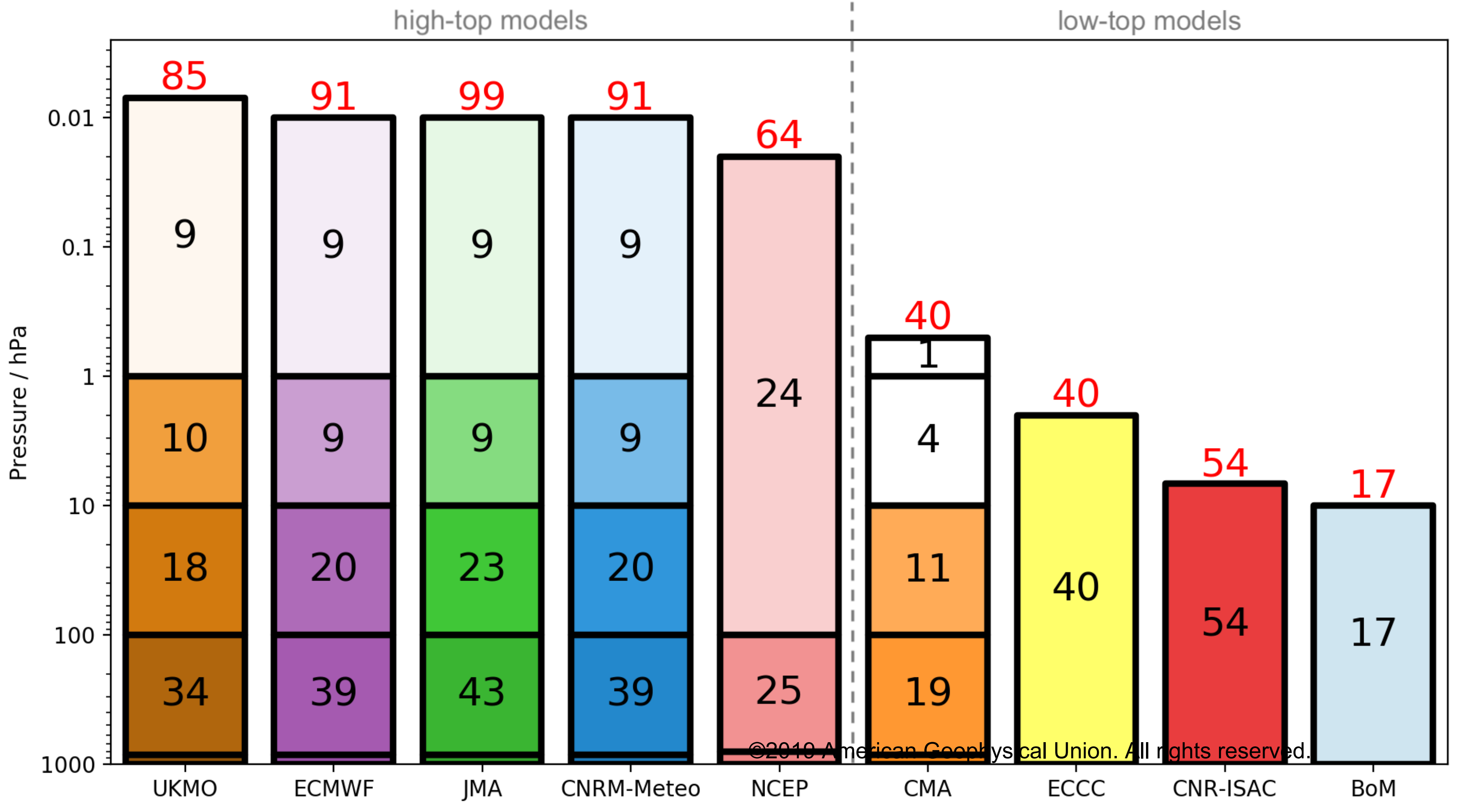

(a) Displacement SSWs
(b) Split SSWs
Percent ensemble members [%]
0
20
40
60
80
100
-30 to -26
-25 to -21
-20 to -16
-15 to -11
-10 to -6
-5 to -1
Lead time (days)
CMA
ECCC
BOM
CNR-ISAC
ECMWF^x
JMA^x
UKMO^x
CNRM−Meteo^x

Figure.



high-top models
low-top models
Pressure / hPa
0.01
0.1
1
10
100
1000
85
91
99
91
64
40
40
54
17
9
9
9
9
24
1
4
10
9
9
9
18
20
23
20
11
40
54
17
34
39
43
39
25
19
UKMO
ECMWF
JMA
CNRM-Meteo
NCEP
CMA
ECCC
CNR-ISAC
BoM

Figure.

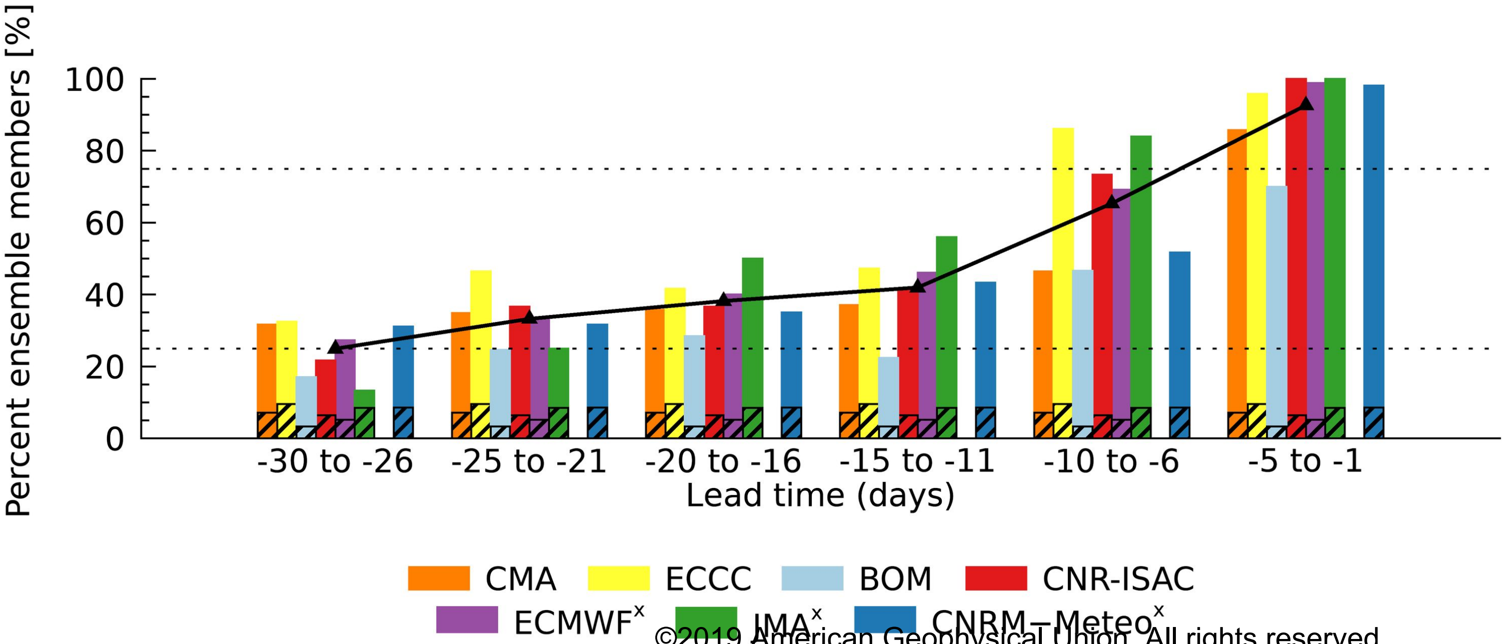
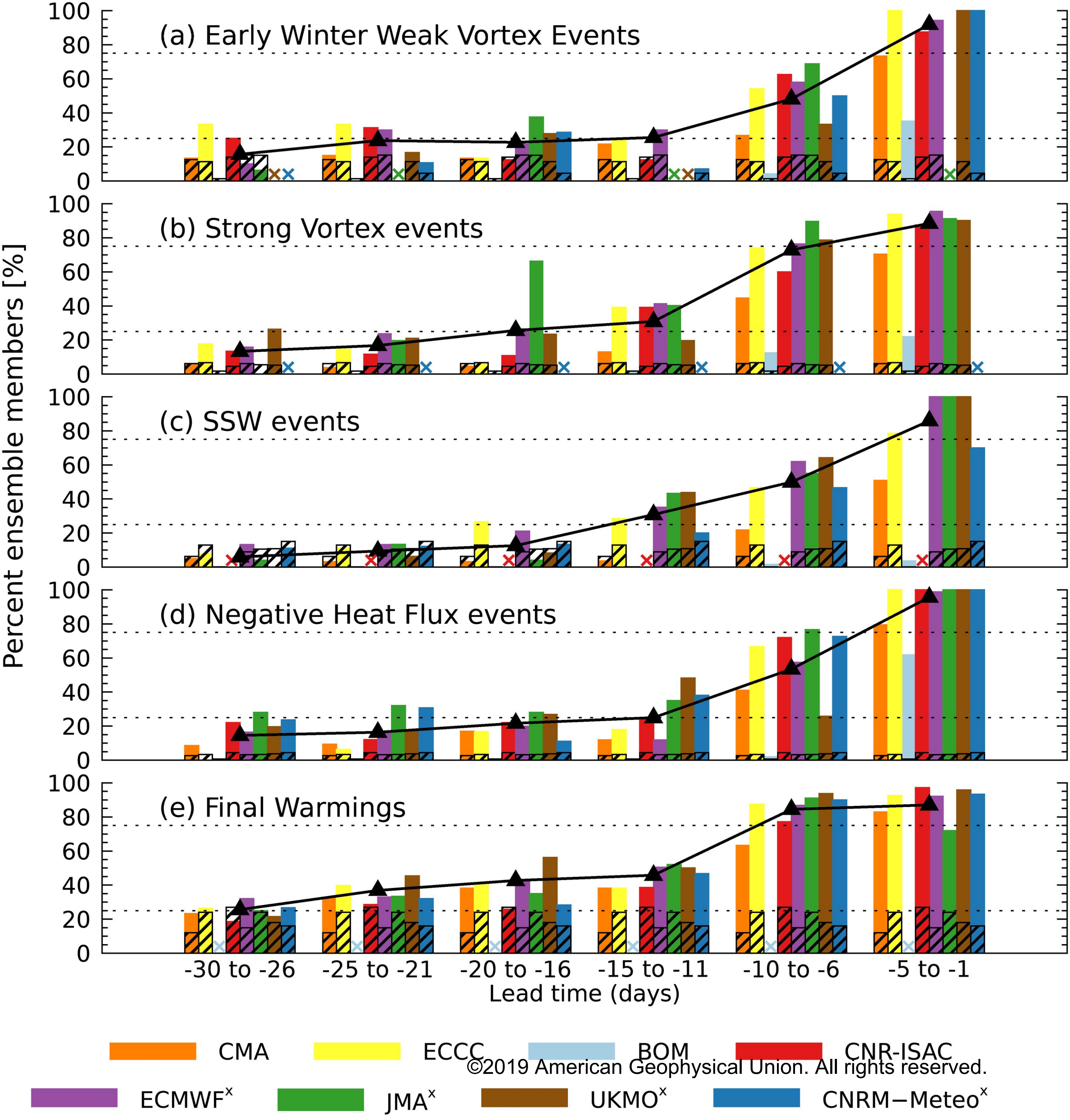

Percent ensemble members [%]
100
80
60
40
20
0
-30 to -26
-25 to -21
-20 to -16
-15 to -11
-10 to -6
-5 to -1
Lead time (days)
CMA
ECCC
BOM
CNR-ISAC
ECMWF^x
JMA^x
CNRM−Meteo^x

Figure.



(a) Early Winter Weak Vortex Events
(b) Strong Vortex events
(c) SSW events
(d) Negative Heat Flux events
(e) Final Warmings
Percent ensemble members [%]
0
20
40
60
80
100
-30 to -26
-25 to -21
-20 to -16
-15 to -11
-10 to -6
-5 to -1
Lead time (days)
CMA
ECCC
BOM
CNR-ISAC
ECMWF^x
JMA^x
UKMO^x
CNRM−Meteo^x

Figure.

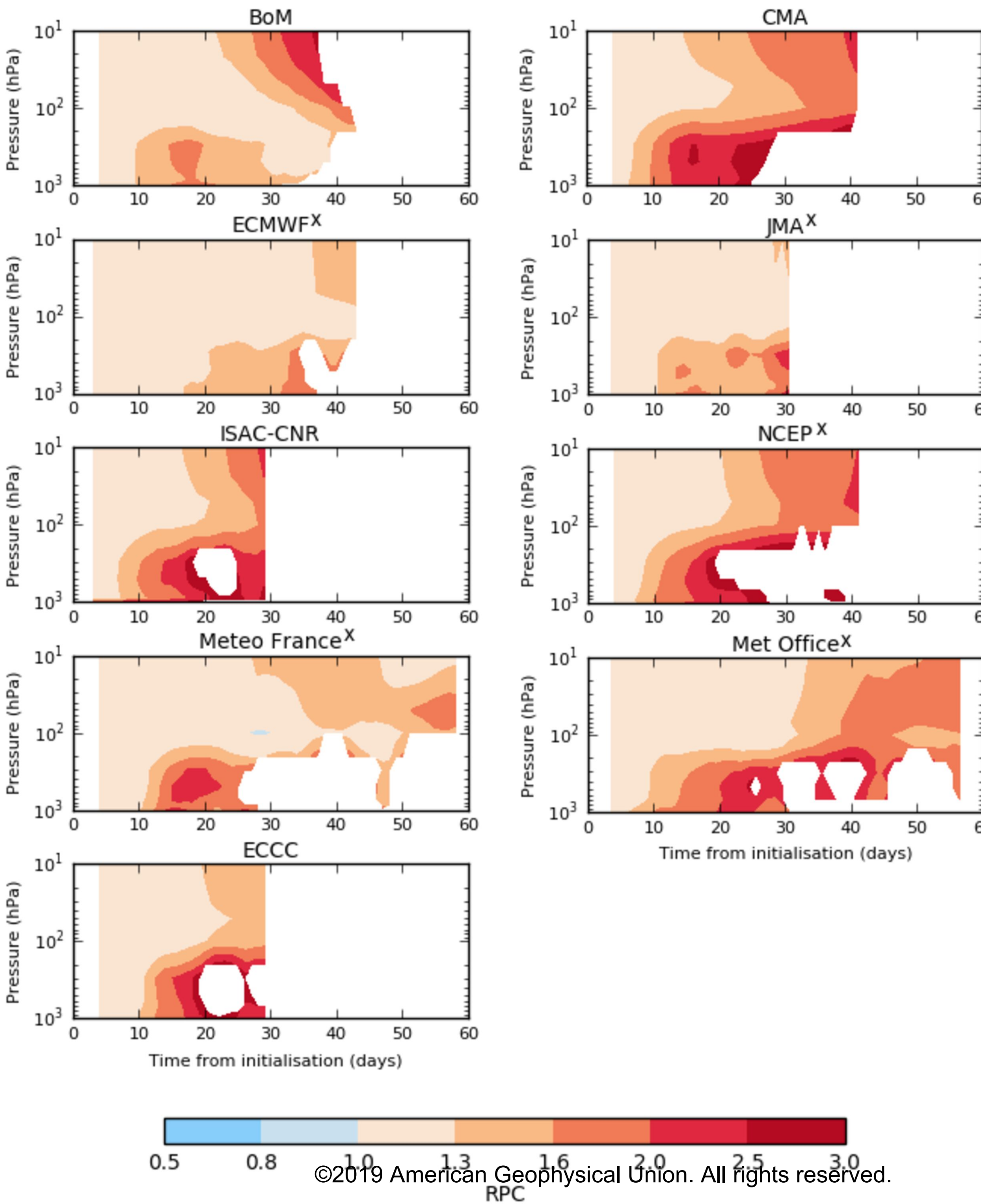

BoM
CMA
ECMWF^x
JMA^x
ISAC-CNR
NCEP^x
Meteo France^x
Met Office^x
ECCC
Pressure (hPa)
Time from initialisation (days)
0.5
0.8
1.0
1.3
1.6
2.0
2.5
3.0
RPC

Figure.

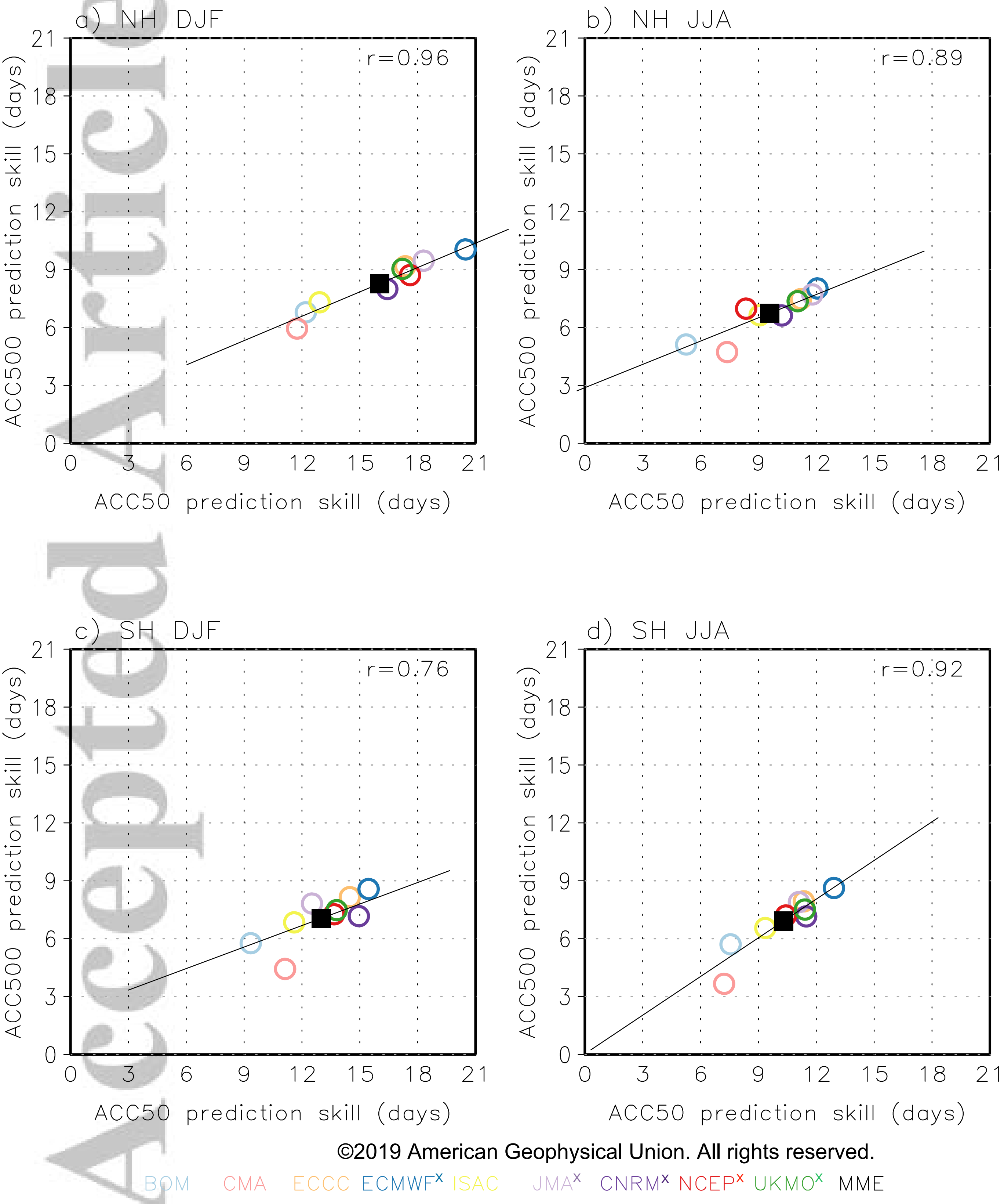
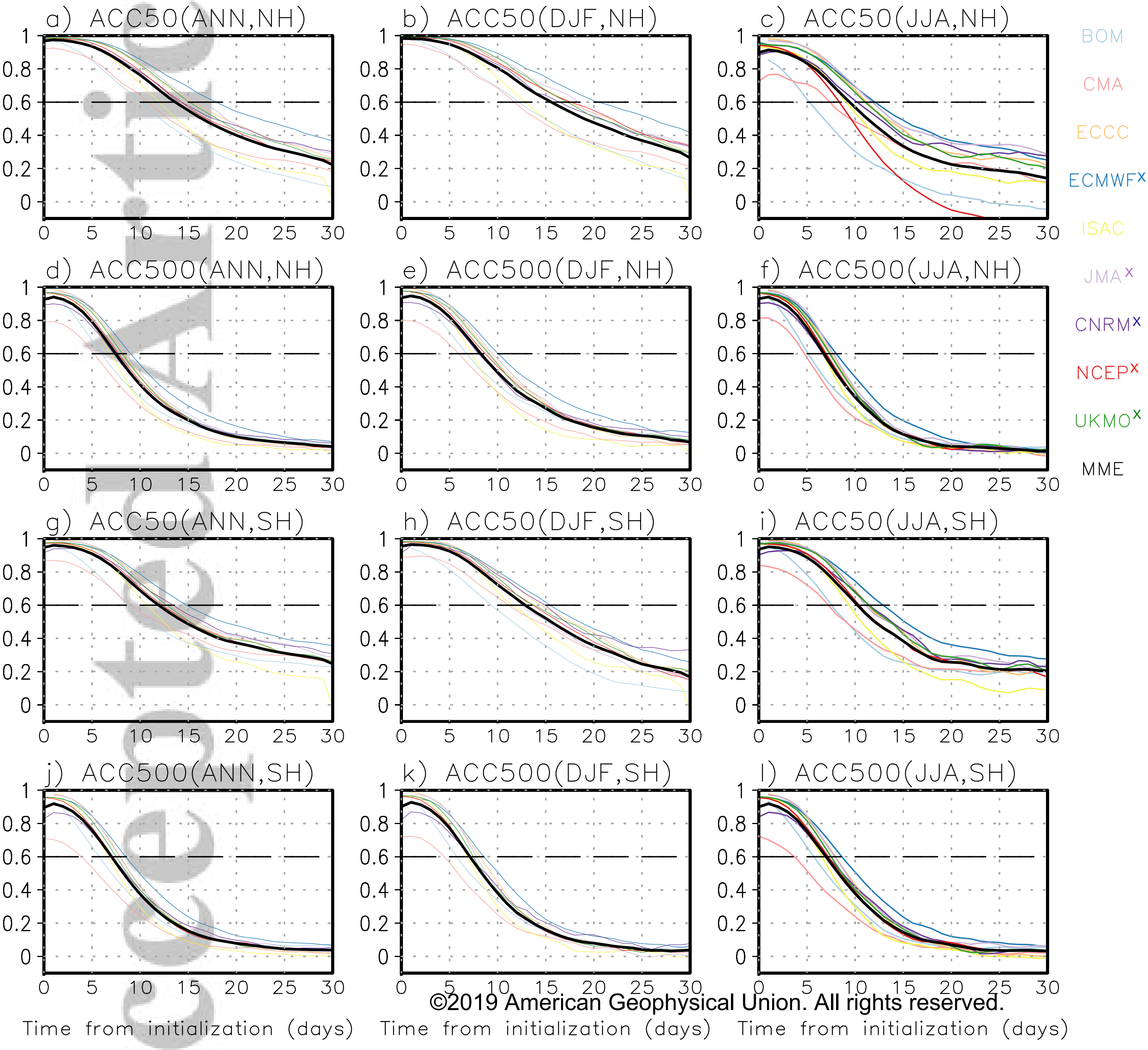

a) NH DJF
b) NH JJA
c) SH DJF
d) SH JJA
r=0.96
r=0.89
r=0.76
r=0.92
ACC500 prediction skill (days)
ACC50 prediction skill (days)
0
3
6
9
12
15
18
21
BOM CMA ECCC ECMWF^x ISAC JMA^x CNRM^x NCEP^x UKMO^x MME

Figure.



a) ACC50(ANN,NH)
b) ACC50(DJF,NH)
c) ACC50(JJA,NH)
d) ACC500(ANN,NH)
e) ACC500(DJF,NH)
f) ACC500(JJA,NH)
g) ACC50(ANN,SH)
h) ACC50(DJF,SH)
i) ACC50(JJA,SH)
j) ACC500(ANN,SH)
k) ACC500(DJF,SH)
l) ACC500(JJA,SH)
BOM
CMA
ECCC
ECMWFx
ISAC
JMAx
CNRMx
NCEPx
UKMOx
MME
Time from initialization (days)