The role of the stratosphere in subseasonal to seasonal prediction part II: predictability arising from stratosphere - troposphere coupling

Article

Accepted Version

To link to this article DOI: http://dx.doi.org/10.1029/2019jd030923

Publisher: American Geophysical Union

The role of the stratosphere in subseasonal to seasonal prediction
Part II: Predictability arising from stratosphere - troposphere coupling

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:

- Tropospheric precursors of SSW events are better represented for the North Pacific than for Eurasia.
- Teleconnections from the tropics add probabilistic skill but are only represented by a few models.
- Weak and strong vortex events in the NH stratosphere can contribute to surface skill 3-4 weeks later.

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:10.1029/2019JD030923

Abstract

The stratosphere can have a significant impact on winter surface weather on subseasonal to seasonal (S2S) timescales. This study evaluates the ability of current operational S2S prediction systems to capture two important links between the stratosphere and troposphere: (1) changes in probabilistic prediction skill in the extratropical stratosphere by precursors in the tropics and the extratropical troposphere and (2) changes in surface predictability in the extratropics after stratospheric weak and strong vortex events. Probabilistic skill exists for stratospheric events when including extratropical tropospheric precursors over the North Pacific and Eurasia, though only a limited set of models captures the Eurasian precursors. Tropical teleconnections such as the Madden-Julian Oscillation, the Quasi-Biennial Oscillation, and El Niño Southern Oscillation increase the probabilistic skill of the polar vortex strength, though these are only captured by a limited set of models. At the surface, predictability is increased over the USA, Russia, and the Middle East for weak vortex events, but not for Europe, and the change in predictability is smaller for strong vortex events for all prediction systems. Prediction systems with poorly resolved stratospheric processes represent this skill to a lesser degree. Altogether, the analyses indicate that correctly simulating stratospheric variability and stratosphere-troposphere dynamical coupling are critical elements for skillful S2S wintertime predictions.

1 Introduction

Subseasonal to seasonal (S2S) predictions of surface climate, generally referring to lead times of two weeks to two months, represent important information for a wide range of sectors including agriculture, insurance, finance, governmental and municipal planning for a range of applications, e.g. for crop planning, disaster readiness, and energy (e.g. Beerli, Wernli, & Grams, 2017; C. J. White et al., 2017). However, the predictability of both Northern and Southern Hemisphere mid-latitudes is limited and decreases considerably after about a week. Although the theoretical limit of short-term weather forecasts is close to 3 weeks (Buizza & Leutbecher, 2015; D. I. V. Domeisen, Badin, & Koszalka, 2018; F. Zhang et al., 2019), weather predictions beyond 2 weeks have traditionally been challenging, as unpredictable 'weather noise' is large compared to the signals that are obtained with an ensemble initial-value approach. Nevertheless, for the prediction on timescales of weeks to months, there exist recent promising improvements in prediction skill. For winter, some facets of the extratropical Northern Hemisphere (NH) circulation such as the North Atlantic Oscillation (NAO; e.g., Hurrell, Kushnir, & Visbeck, 2001; Walker, 1928) are predictable to some degree with seasonal prediction systems (Baker, Shaffrey, Sutton, Weisheimer, & Scaife, 2018; Dobrynin et al., 2018; L'Heureux et al., 2017; Scaife, Arribas, et al., 2014; Stockdale, Molteni, & Ferranti, 2015).

One prospect for enhancing predictive skill of surface climate on S2S timescales is the extratropical winter stratosphere (e.g., Butler et al. (2018); Gerber et al. (2012); Scaife et al. (2016)), which exhibits longer characteristic timescales (Baldwin et al., 2003; Gerber et al., 2010) and hence predictability (Q. Zhang, Shin, Dool, & Cai, 2013) as compared to the troposphere, as shown in the first part of this study (D. I. Domeisen et al., 2019, hereafter Part I). In particular, extreme events in the extratropical stratosphere can have impacts that descend to the lower stratosphere (Hitchcock, Shepherd, Taguchi, Yoden, & Noguchi, 2013; R. A. Plumb & Semeniuk, 2003) and in some cases all the way down to the surface, where they can lead to changes in variability on subseasonal timescales in both the Northern (Baldwin & Dunkerton, 1999, 2001; Butler et al., 2018) and the Southern Hemisphere (E.-P. Lim et al., 2019). The mechanisms of downward influence of the stratosphere onto the troposphere are a topic of active research (D. I. V. Domeisen, Sun, & Chen, 2013; Douville, 2009; Dunn-Sigouin & Shaw, 2018; C. I. Garfinkel, Waugh, & Gerber, 2013; Hitchcock & Simpson, 2014, 2016; Simpson, Blackburn, & Haigh, 2009, 2012; K. L. Smith & Scott, 2016; Y. Song & Robinson, 2004); for a summary of the mechanisms see Kidston et al. (2015); Tripathi, Baldwin, et al. (2015). In particular, the North Atlantic and Eurasia are strongly impacted by stratospheric extremes, with surface temperature anomalies on the order of several °C for days to weeks after a stratospheric event (Butler et al., 2018; Butler, Sjoberg, Seidel, & Rosenlof, 2017). Due to this downward coupling from the stratosphere it has been suggested that the stratosphere may be able to increase the predictability of surface weather (Butler et al., 2016; Scaife et al., 2016; Sigmond, Scinocca, Kharin, & Shepherd, 2013). Several single-model studies found an increase in prediction skill for forecasts that were initialized during sudden stratospheric warming (SSW) events or with an improved stratospheric representation for various tropospheric fields such as the Northern Annular Mode (NAM, e.g., D. W. J. Thompson & Wallace, 2000), with a focus on the North Atlantic sector and hence the NAO, as well as surface temperatures (Kuroda, 2008; Marshall & Scaife, 2010; Sigmond et al., 2013). For example, the major SSW event in February 2018 has been suggested to have led to persistent cold weather over large parts of Europe in late February and early March after an otherwise mild winter (Karpechko, Perez, Balmaseda, Tyrrell, & Vitart, 2018), as well as anomalously wet conditions over southwestern Europe (Ayarzagüena et al., 2018). Like the 2018 event, up to two thirds of SSW events are followed by anomalous tropospheric weather patterns that can remain persistent for several weeks (Charlton-Perez, Ferranti, & Lee, 2018; D. I. V. Domeisen, 2019; Karpechko, Hitchcock, Peters, & Schneidereit, 2017; Simpson, Hitchcock, Shepherd, & Scinocca, 2011; I. White et al., 2018). The prospects of using the stratosphere for enhanced predictability at the surface on sub-seasonal to seasonal timescales is not limited to SSW events, as impacts on surface weather are also expected for other types of polar stratospheric extreme events such as strong vortex events (Tripathi, Charlton-Perez, Sigmond, & Vitart, 2015) and final warming events (Butler, Perez, Domeisen, Simpson, & Sjoberg, 2019; Hardiman et al., 2011).

While skillful deterministic forecasts of the above described extreme stratospheric events are limited to lead times of no more than 10 to 15 days (see Part I), the probability of occurrence of these events during a given winter can be modified through remote impacts that affect polar vortex strength. A range of studies argue for precursors to SSW events in the extratropical troposphere (Davies, 1981; Kolstad & Charlton-Perez, 2010; Schneidereit et al., 2017) such as atmospheric blocking (Ayarzagüena, Langematz, & Serrano, 2011; Martius, Polvani, & Davies, 2009; Nishii, Nakamura, & Orsolini, 2011; Quiroz, 1986; Woollings, Charlton-Perez, Ineson, Marshall, & Masato, 2010), Arctic sea ice (Kim et al., 2014; Sun, Deser, & Tomas, 2015; P. Zhang et al., 2018), Eurasian snow cover (Cohen & Entekhabi, 1999), and precursors in the extratropical lower stratosphere (Albers & Birner, 2014; de la Camara et al., 2017; D. I. V. Domeisen, Martius, & Jiménez-Esteve, 2018; Polvani & Waugh, 2004; Stockdale et al., 2015). The strength of the polar vortex can further be modified through remote impacts from the tropics, i.e. by El Niño Southern Oscillation (ENSO) (Butler et al., 2016; Butler & Polvani, 2011; Butler, Polvani, & Deser, 2014; D. I. V. Domeisen et al., 2015; C. I. Garfinkel & Hartmann, 2007; Ineson & Scaife, 2009; Manzini, Giorgetta, Esch, Kornblueh, & Roeckner, 2006; Polvani, Sun, Butler, Richter, & Deser, 2017; K. Song & Son, 2018), for a summary see D. I. V. Domeisen, Garfinkel, and Butler (2019), tropical convection related to the Madden-Julian Oscillation (C. I. Garfinkel, Benedict, & Maloney, 2014; C. I. Garfinkel, Feldstein, Waugh, Yoo, & Lee, 2012; Kang & Tziperman, 2017), and the Quasi-Biennial Oscillation (QBO) through the Holton-Tan effect (Holton & Tan, 1980): Easterly winds in the tropical lower stratosphere associated with an easterly QBO (eQBO) have been suggested to lead to a weakened stratospheric vortex through modifications in wave propagation and breaking in the surf zone (Andrews, Martin B et al., 2019; C. I. Garfinkel et al., 2018; C. I. Garfinkel, Shaw, Hartmann, & Waugh, 2012; O'Reilly, Weisheimer, Woollings, Gray, & MacLeod, 2018; Richter, Deser, & Sun, 2015; Scaife, Athanassiadou, et al., 2014). These tropical modes of variability can also have a direct effect on the extratropical troposphere without a stratospheric pathway (B. J. Hoskins & Ambrizzi, 1993; Li, Li, Jin, & Zhao, 2015; Scaife et al., 2017), while for ENSO it has been shown that the stratospheric influence, if present, tends to dominate over the tropospheric pathway (Butler et al., 2014; Jiménez-Esteve & Domeisen, 2018).

We use subseasonal model hindcasts from operational prediction systems to evaluate the role of stratosphere - troposphere coupling in the NH with respect to the influence of precursors to stratospheric events (Section 3) and potential changes in predictability of surface weather given stratospheric variability (Section 4). Section 2 gives a brief introduction to the database and the methodology (for more details see Part I). Section 5 provides a discussion of the results.

2 Methodology

2.1 Data

We use hindcast data from the S2S forecast project containing 11 different operational subseasonal forecast systems (Vitart et al., 2017). Table 1 (repeated from Part I) provides an overview over the models used in this study (further details about the models can be found in Part I). Event definitions are given in sections 3 and 4.

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.

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	3

Due to the large differences in ensemble size, time period, and model specifics, the exact datasets or selection of models may vary depending on the analysis or application in this study, depending on the specific requirements of different parts of the analysis in terms of e.g. lead times or available time periods. Different numbers of ensemble members for BoM were used in this analysis, depending on the number of members available at the time of data acquisition.

ERA-interim (Dee et al., 2011) is used for comparison to the model data. Note that not all models are initialized from the same reanalysis dataset (Table 1). For the reanalysis data, anomalies are defined relative to the daily climatological seasonal cycle. For the forecasts, the anomalies are defined relative to the model climatology at an equivalent lead time for all forecasts initialized on the same date of the year. No smoothing has been applied to the climatology.

2.2 Skill Measures

Skill is evaluated according to the following skill measures. If the variable X is not averaged spatially, e.g., in Figure 5, the *correlation coefficient* (r), or correlation skill score, is 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 the subscripts *mod* and *obs* denote the model ensemble mean and the reanalysis dataset of the variable X, 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 samples for which r is being evaluated (e.g. Table 2).

To evaluate the spatial skill of the anomaly pattern as in Fig. 6, the spatial weighting by cosine of latitude w and spatial averaging over S grid points is applied as an additional summation over the covariance and variance terms separately, i.e.,

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

By removing the lead time - dependent climatology from the hindcasts, we *a posteriori* remove systematic errors in the model hindcasts. In this study, r and ACC are computed for the ensemble mean X_{mod} for each prediction system at lead times of 3-4 weeks. The multi-model mean correlation is the averaged correlation over all prediction systems.

We also use the root mean square error (RMSE), which is defined as the root mean square difference between forecast anomalies and observed anomalies averaged over T samples:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{T}([X_{mod} - C_{mod}] - [X_{obs} - C_{obs}])^2}{T}} \tag{3}$$

3 Precursors and Remote Influences on the Northern Hemisphere Stratosphere

As shown in Part I, extreme stratospheric events tend to be difficult to forecast on subseasonal timescales. However, there exist precursors and remote connections to stratospheric events that tend to affect the strength of the polar vortex and thereby the probability of occurrence of these events. These are assessed in the following two sections.

3.1 Precursors in the Extratropical Northern Hemisphere Troposphere

SSW events are often preceded by anomalously strong vertical propagation of waves into the extratropical stratosphere, and favorable tropospheric circulation patterns exist that promote such wave generation (e.g. Bao, Ming, Tan, Xin, Hartmann, Dennis L, & Ceppi, Paulo, 2017; Charlton & Polvani, 2007; Cohen & Jones, 2011; D. I. V. Domeisen, 2019; C. I. Garfinkel, Hartmann, & Sassi, 2010; Jucker & Reichler, 2018; Kolstad & Charlton-Perez, 2010; Martius et al., 2009; I. White et al., 2018). Note that not all SSW events are preceded by significant tropospheric anomalies and there are a range of internal stratospheric processes that have been suggested to give rise to SSW events (Birner & Albers, 2017; de la Camara et al., 2017; D. I. V. Domeisen, Martius, & Jiménez-Esteve, 2018; Esler & Matthewman, 2011; Matthewman & Esler, 2011; R. Plumb, 1981). If precursors exist, they have been suggested to be present for several weeks before the occurrence

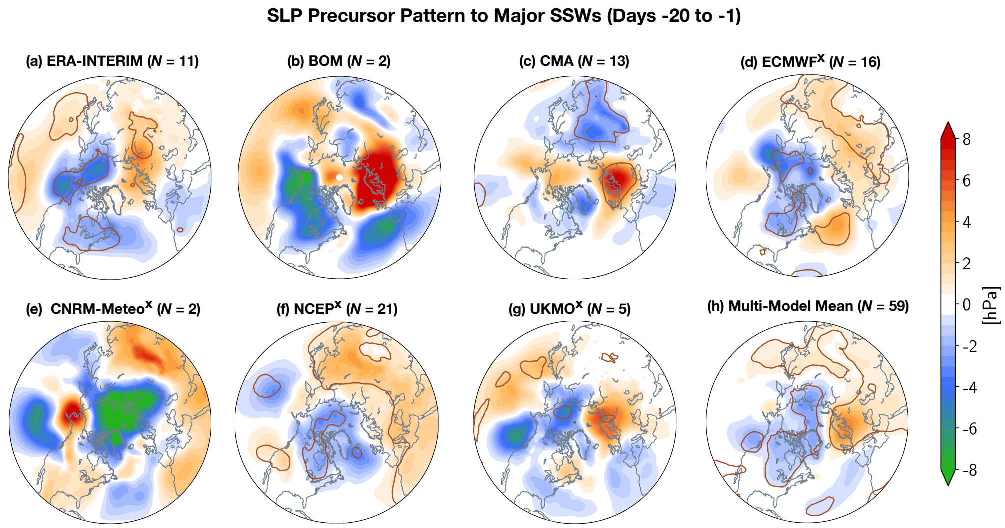


Figure 1. (a) NH SLP anomalies [hPa] averaged over days 1 to 20 before mid-winter SSW events for 1996-2010 in ERA-interim. (b)-(g) As in (a), but for the ensemble mean SLP anomaly composite for simulated mid-winter SSW events in six of the S2S prediction systems considered here (see text for details). Each model composite represents the mean of individual ensemble members. (h) As in (a) but for the multi-model mean. Areas enclosed by solid brown lines denote where the composite mean of each panel is significantly different from zero [$p < 0.05$] as determined by a two-tailed Student's t-test. The sample size for each composite is given in the title of the panel. '$\times$' indicates high-top models.

of a SSW event, thus making them useful to infer stratospheric variability and even contribute to the probabilistic predictability of stratospheric events at lead times of several weeks. As such, evaluating these precursor patterns in the S2S prediction systems serves as a measure to benchmark their ability to predict stratospheric variability on S2S timescales.

Figure 1 illustrates the sea level pressure (SLP) anomalies up to 20 days before a mid-winter SSW event occurs in the NH. As in Part I, mid-winter SSW events are defined based on a zonal mean zonal wind reversal at 60°N and 10 hPa (Charlton & Polvani, 2007). The events considered for reanalysis are the ones in Table 2 of Butler et al. (2017) for ERA-interim, but here only events for December - February (DJF) between 1996-2010 are considered ($N = 11$). For the models, we use the same criterion as for reanalysis for identifying major mid-winter SSW events for each ensemble member. However, because of the limited length of the hindcasts and the fact that we are looking at lagged composites, we can only consider mid-winter SSW events that occur *at least* 20 days into a hindcast run, allowing us to look back as far as 20 days for the precursor patterns within the same hindcast period. Performing the analysis for days -25 to -5 or days -30 to -1 yields sample sizes that become too small for analysis. The composites are generated by averaging SLP for days -20 to -1 before the SSW event for both the reanalysis data and for simulated SSW events. These composites are then averaged over all SSW events for reanalysis and over all ensemble members within each prediction system to form an ensemble-mean picture. Only prediction systems with at least two identified mid-winter SSW events are considered in this analysis. The reanalysis composite (Fig. 1a) shows three distinct features: (1) anomalous ridging in central Asia and extending into northern Europe (though only statistically significant in central Asia); (2) an intensified Gulf of Alaska Low and

Pacific High, corresponding to the positive phase of the North Pacific Oscillation (e.g. Rogers, 1981); and (3) anomalously low SLP across central and northeastern North America. The dominant features in both the North Pacific and over Eurasia have been documented in the literature both in models and different reanalysis products (e.g. D. I. V. Domeisen, 2019; Furtado, Cohen, Butler, Riddle, & Kumar, 2015; C. I. Garfinkel et al., 2010; Karpechko et al., 2018; Kolstad & Charlton-Perez, 2010; Peings, 2019; I. White et al., 2018) and they can manifest as an amplification of the climatological planetary-scale wave pattern through wave interference (K. L. Smith & Kushner, 2012). An amplification of the climatological wave structure, especially over the Pacific sector, thus provides increased wave forcing and easterly momentum to the westerly flow in the stratosphere, increasing the chances of a SSW event.

The SLP anomaly precursors in the individual prediction systems show substantial differences as compared to reanalysis (Figs. 1b-h). The SLP precursor to mid-winter SSW events in the multi-model mean (Fig. 1h) consists of negative anomalies in the Gulf of Alaska and central North America and positive anomalies over the Europe. Ridging over central Asia is less well captured. Examining the prediction systems individually, all of them (except for CNRM-Meteo, Fig. 1e) feature positive SLP anomalies across Scandinavia / northern Europe and extending into Asia, though significance of this feature differs between the prediction systems. The North Pacific SLP anomalies are less well replicated in the individual systems, with the UKMO model showing the closest similarity to reanalysis (though statistically insignificant). The North American negative SLP anomalies seen in the reanalysis plot are also less common in individual models, though ECMWF (Fig. 1d) and NCEP (Fig. 1f) appear to reproduce a similar feature. Note that these two models are also the ones with the two largest sample sizes for their composites (16 and 21, respectively), thus strongly influencing the multi-model mean composite (Fig. 1h).

While the above analysis provides insight into precursor structures in the prediction systems before they produce a SSW, it does not provide information about predictability. Therefore, a similar analysis to that shown in Fig. 1 (but for days -30 to -5 before the event) was performed using the *observed* major SSW event dates in the model hindcasts (i.e., finding model hindcasts corresponding to SSW events recorded in reanalysis; Fig. S1). Some of the same SLP precursors identified in Fig. 1 are reproduced for the composites based on the reanalysis-identified SSW events. In reanalysis (Fig. S1a), anomalous ridging across northern Europe and extending into Asia and an intensified Aleutian Low and Pacific High are apparent. All prediction systems reproduce the negative SLP anomalies near the Aleutians, though with a large range in both strength and location (Figs. S1b-j). The NCEP ensemble-mean composite (Fig. S1i) captures well the amplitude of the SLP anomalies across the North Pacific and Scandinavia. The multi-model mean (Fig. S1k) also captures the importance of negative SLP anomalies in the North Pacific and the European-centered positive SLP anomaly, though the ridge over Siberia is less well captured. Overall, the general similarities between the SLP precursor patterns for both simulated and observed mid-winter SSW events within the prediction systems make these patterns useful for subseasonal forecasts of stratospheric variability. Note that since the SSW dates are based on reanalysis data (i.e. the threshold for reanalysis was used to determine which SSW dates to use in the models), the model composites may include predictions that may not have met the criterion for a SSW event. Interestingly, the figure shows that precursor structures at the surface are nevertheless present in the model systems, although these may not necessarily have led to fulfilling the threshold for a SSW event. This indicates the importance of internal variability in the stratosphere, which to a large extent determines the effect that tropospheric wave forcing has on the stratospheric flow (Albers & Birner, 2014; de la Camara et al., 2017).

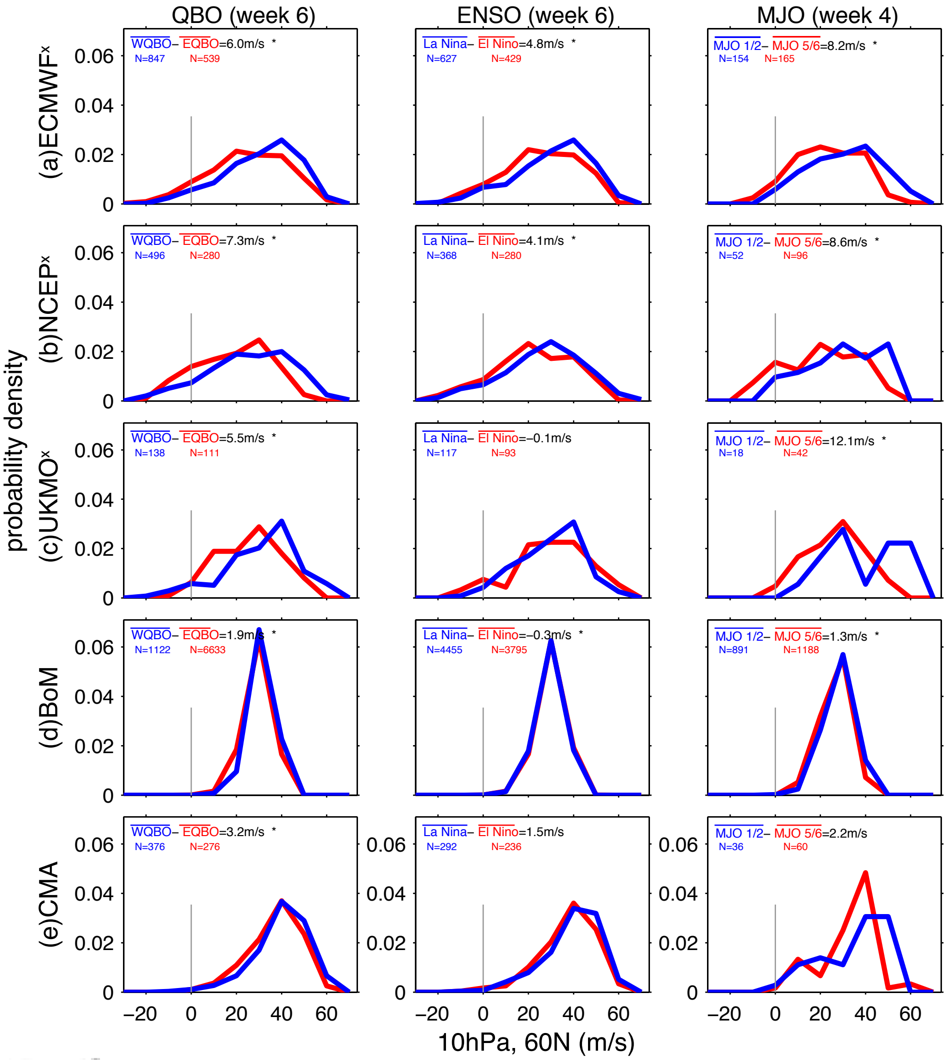


Figure 2. Probability density of zonal mean zonal wind at 10 hPa, 60°N for hindcasts initialized in November and December. Red (blue) lines indicate hindcasts initialized during (left) eQBO (wQBO), (center) El Niño (La Niña) conditions, and (right) MJO phases 5/6 (1/2). All histograms are normalized for comparison. No smoothing is applied. The vertical line indicates zero zonal wind speed. Each panel indicates the difference in the means [ms^{-1}] between the considered phases (top left corner). * indicates values that differ significantly from zero [$p < 0.05$] as given by a Students t-test. High-top models are indicated by an ×. N indicates the sample size for each category.

3.2 Tropical Precursors

The extratropical stratosphere is affected by remote influences from the tropics. These so-called *teleconnections* can affect the strength of the stratospheric polar vortex and thereby the probability of occurrence of stratospheric events such as SSWs. Examples of teleconnections from the tropics with a strong influence on the extratropical stratosphere are the Madden-Julian Oscillation (MJO) (C. I. Garfinkel, Feldstein, et al., 2012), the Quasi-Biennial Oscillation (QBO) (Holton & Tan, 1980), and El Niño Southern Oscillation (ENSO) (D. I. V. Domeisen et al., 2019).

The models used for this part of the analysis are the ones that exhibit lead times long enough to fully exploit these teleconnections, i.e. ECMWF, NCEP, UKMO, BoM, and CMA. The time periods used for the analysis correspond to the last full week available for all models (week 6) for the QBO and ENSO, and the fourth week after MJO phases 1/2 and 5/6 following C. I. Garfinkel, Feldstein, et al. (2012). The hindcasts are those initialized in November and December from Table 1 of C. I. Garfinkel et al. (2018), which overlaps the dates chosen in this paper nearly completely.

The left column of Figure 2 shows the probability density function (PDF) for zonal wind at 10 hPa and 60°N for opposite QBO phases in order to assess whether the prediction systems capture the Holton-Tan effect. The QBO phase is defined by averaging the zonal mean zonal wind at 50 hPa from 5°S - 5°N over the first three days of the hindcast. This metric is categorized as eQBO (wQBO) if the QBO winds are less (more) than -(+)3 ms^{-1}. Note that, for the most part, these prediction systems do not internally generate a QBO, and lose the QBO signal within a few weeks after initialization (Butler et al., 2016; C. I. Garfinkel et al., 2018; Y. Lim, Son, Marshall, Hendon, & Seo, 2019), but the initial conditions are expected to be sufficient to influence the NH polar vortex on subseasonal timescales. The three prediction systems with a more highly resolved stratosphere (ECMWF, NCEP, UKMO) simulate a stronger weakening of the zonal winds at 10 hPa and 60°N for eQBO in week 6 (36 to 42 days after initialization; after C. I. Garfinkel et al., 2018) than those with a more poorly resolved stratosphere.

El Niño conditions in the tropical Pacific have been shown to lead to a weakened stratospheric vortex (D. I. V. Domeisen et al., 2019; García-Herrera, Calvo, Garcia, & Giorgetta, 2006; C. I. Garfinkel & Hartmann, 2007; Manzini et al., 2006), while La Niña tends to be associated with a strengthening, though this connection is less robust (Iza, Calvo, & Manzini, 2016; Polvani et al., 2017). The second column of Figure 2 shows the PDF of zonal wind at 10 hPa and 60°N for week six (days 36 to 42) after initialization for November and December hindcasts initialized during El Niño and La Niña. Monthly mean sea surface temperature anomalies in the Niño3.4 region from ERSSTv5 data (Huang et al., 2017) exceeding ±0.5°C are used to categorize the ENSO phase. The ECMWF and NCEP forecasting systems simulate a weakening of stratospheric zonal winds for El Niño as compared to La Niña (C. Garfinkel et al., 2019).

The phase of the MJO with enhanced convection in the far-West Pacific (phases 5/6 as defined by the real-time multivariate MJO index of Wheeler & Hendon, 2004) more often precedes weak vortex events at 4-week lags than the opposite phases 1/2 with reduced convection in this region and enhanced convection in the Indian Ocean (C. I. Garfinkel et al., 2014; C. I. Garfinkel, Feldstein, et al., 2012; Kang & Tziperman, 2017; Schwartz & Garfinkel, 2017). Figure 2 (right column) shows the PDF for zonal mean zonal winds at 10hPa and 60°N for days 22 to 28 (week 4) following these respective phases for all initialization dates in November and December. As with ENSO and the QBO, the prediction systems with a well-resolved stratosphere also simulate a weakening of the vortex following MJO phases 5/6 (after C. I. Garfinkel & Schwartz, 2017).

When comparing to MERRA reanalysis data (Rienecker et al., 2011), for the QBO and for the MJO, the model simulated effects are somewhat weaker than for reanalysis, even for the high-top models (Fig. S2). C. I. Garfinkel et al. (2018) show that the model spread encompasses the observed response for the QBO, so there is no evidence that models are systematically biased, even if the ensemble mean response is too weak. For ENSO, the observed effect is opposite to that in models (and also opposite to the observed response in the period before the S2S hindcasts); the mismatch between observations and the S2S models for ENSO is analyzed in detail in C. Garfinkel et al. (2019).

Finally, the probability of easterly winds in the polar stratosphere tends to increase if the hindcast is initialized during eQBO, El Niño, or MJO phase 6 (e.g., the ECMWF system shows an increase in the probability for easterly winds by 66% for eQBO vs wQBO, by 30% for El Niño vs La Niña, and by 139% for MJO phases 5/6 vs phases 1/2). While the variability between models is large, these changes in probability could potentially be used to formulate probabilistic predictions of SSW events at time lags where deterministic prediction is not possible according to the analysis in Part I.

4 Predicting the Downward Coupling to the Troposphere

This section analyzes the potential of the S2S prediction systems to reproduce and predict the downward impact of mid-winter stratospheric events onto the surface, with a focus on weak and strong polar vortex events in the Northern Hemisphere.

4.1 Arctic surface anomalies

The strength of the stratospheric polar vortex and its associated potential vorticity anomalies are linked to polar cap surface pressure anomalies through a vertical movement of the polar tropopause (Ambaum & Hoskins, 2002). Thus, polar sea level pressure is a suitable variable for studying tropospheric predictability arising from the stratosphere. Moreover, these surface pressure anomalies are relevant for near-surface weather and even for Arctic sea ice distribution and motion (Kwok, 2000; K. Smith, Polvani, & Tremblay, 2018). In addition, polar pressure anomalies also have implications at midlatitudes, because they can project onto the tropospheric NAO pattern. This surface impact can lead to lagged changes in the near-surface temperature or upper tropospheric winds (Baldwin, 2001; D. Thompson & Wallace, 1998; D. Thompson, Wallace, & Hegerl, 2000).

The stratospheric signal is here characterized by the averaged anomalies over the polar cap of pressure at fixed heights, defined by a metric of the stratospheric variability based on daily 100 hPa temperature averaged over 65°-90°N, denoted the *ST100 index* (Baldwin, Birner, & Ayarzagüena, 2019). We regress the anomalous polar cap pressure for the atmospheric column on the standardized ST100 index in January-March for ERA-interim reanalysis (Fig. 3a, black line) for the period 1999-2010. The pressure anomalies exhibit two maxima, one in the lower stratosphere (around 16km) and the other close to the surface. The latter denotes a strengthened stratospheric signal at lower levels as compared to other tropospheric levels (Baldwin et al., 2019). The vertical structure in Figure 3a is not expected only from mass moving into and out of the polar cap in the stratosphere. For example, during a SSW, mass is moved into the polar cap in the stratosphere, where the air descends and warms adiabatically. In the lower stratosphere (around 16km) pressure increases by 2hPa. Above that level, the pressure increment (ΔP) has to decrease because the ambient pressure drops off below 4hPa. Below the stratospheric maximum, ΔP would be 2hPa if mass were prevented from flowing out of the polar region below that level, as in a cylinder at 65°N with impermeable walls. Moreover, the flux of mass into the polar cap is almost zero in the lowermost stratosphere. Given that the impermeable walls do not exist, as the air descends from 16km in the lowermost stratosphere, it is not confined to the polar cap, and it "leaks" out of the polar cap below the levels with injection of mass (see Ambaum and Hoskins (2002) for a discussion of the potential vorticity dynamics of this situation). This explains the existence of the first maximum of pressure anomalies, but not the second one at Earth's surface, where we would expect a minimum instead. However, below the tropopause in Fig. 3a, the polar pressure anomalies increase, with a second maximum at the surface. The only explanation for this near-surface maximum of the stratospheric signal is the action of additional tropospheric processes that amplify the signal close to the surface. In particular, changes in low-level heat flux (Baldwin et al., 2019; Limpasuvan, V, Thompson, D, & Hartmann, D L, 2004) and temperature advection (Baldwin et al., 2019; D. Thompson et al., 2000) lead to temperature anomalies over the polar cap that induce pressure anomalies (B. Hoskins, McIntyre, & Robertson, 1985). The surface pressure anomalies ultimately are responsible for the mass movement into the polar cap that is synchronised with mass movement in the stratosphere. The net effect is that the surface pressure signal, e.g. for the NAM, is much larger than would be expected based solely on movement of mass within the stratosphere.

The lagged regression of the anomalous polar pressure at different levels on the standardized ST100 index reveals important aspects of the timing of the tropospheric feedbacks involved in the surface pressure amplification (Fig. 3b). The stratospheric-induced Arctic surface pressure anomalies (blue line; lagged regression of anomalous polar pressure at 0 km onto the ST100 index) peak at a lag of around +3 days with respect to the stratospheric anomaly. Thus, the stratosphere leads the surface signal. Moreover, the anomalies persist up to 60 days, longer than the stratospheric signal itself (orange line; lagged regression of anomalous polar pressure at 15 km onto the ST100 index). The tropospheric-only part of the signal (green line; lagged regression of the difference in anomalous polar pressure at 8 km and 0 km onto the ST100 index) also lags the stratospheric signal.

A similar analysis is now performed with the S2S systems to judge their skill in representing the impact of the stratospheric state on Arctic surface anomalies and particularly, characterize up to which lead times they show an effect of the stratosphere on polar surface weather. In this case regressions of pressure anomalies on the standardized ST100 index were computed separately for all S2S systems. To build the ST100 index and compute the instantaneous regression on polar pressure of Fig. 3a only the data for 24h time steps of all available hindcast initialization dates in JFM of the 1999-2010 period are considered. The results indicate that the polar tropospheric amplification of the stratospheric signal is present in all S2S prediction systems and maximizes near the surface (Fig. 3a, colored lines). Regarding the lagged regressions of pressure anomalies on the standardized ST100 index (Fig. 3c), the computation differs slightly between the S2S systems and the reanalysis: For each S2S system, the anomalous polar pressure is calculated for every 24h time step from 24h to 768h with respect to the initialization time and regressed onto the ST100 index (computed for all 24h time steps). Finally, the regression from each system is averaged over all ensemble members and then over all prediction systems.

In the prediction systems, the surface amplification also peaks at a positive lag of around +3 days (Fig. 3c), but it decays more slowly than in the reanalysis. This is consistent with the quicker decay of the troposphere-only signal (0-8 km) in reanalysis as compared to the S2S systems mean (i.e., the reanalysis lies below the S2S system mean $\pm$ 1.5 standard deviations after 20 days, see green line in Fig. 3c). As expected, the spread among prediction systems grows, in general, with forecast lead time. It is particularly large for the surface response after a lag of 8 days (blue shading), but it does not grow much further after that.

Several reasons might explain the models' deviations from reanalysis and the intermodel spread, i.e. the relatively short study period (1999-2010) or model biases. To test both possibilities we repeated the analysis considering all data available for each S2S system separately as shown in Table 1 (Fig. S3). The short data record might be responsible for the noisy result: when extending the period to 1980-2016 for ERA-interim, the results become smoother (Fig. S3a). The same result is obtained when including the presatellite period (not shown). Moreover, the inter-model spread is also reduced with respect to Fig. 3c, in particular for the surface pressure results. However, even if we consider a longer period of study, the prediction systems still show discrepancies among themselves. For instance, whereas high-top model systems (JMA, UKMO, or ECMWF) depict a comparable magnitude of the stratospheric signal in the lowermost stratosphere and near the surface from lag +4 days, systems with lower stratospheric resolution (BoM and CMA) predict a stronger surface signal. In these latter cases, the tropospheric part of the signal (green line) is similar to that of other systems or reanalysis. Thus, the misrepresented processes in these models should relate to the stratospheric signal itself (as is the case with CMA, Fig. S3d) or the coupling between the stratosphere and the troposphere.

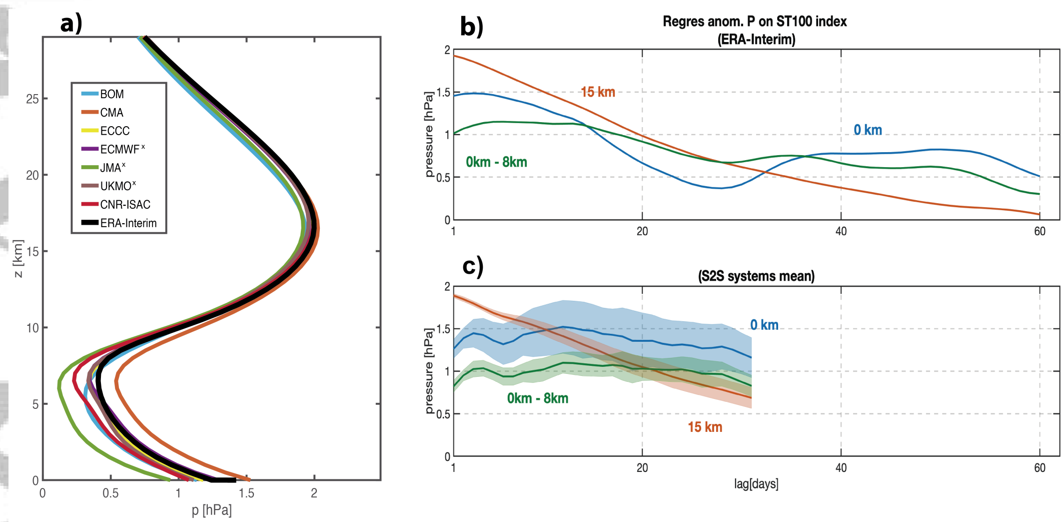


Figure 3. (a) Regression of Arctic (65°N - pole) pressure anomalies (hPa) as a function of height on the standardized ST100 index associated with one standard deviation of the ST100 index for ERA-interim (black line) and the hindcasts from the S2S prediction systems (colored lines) for the period 1999 - 2010. (b,c) Lagged regression between the standardized ST100 index and Arctic mean pressure anomalies at 15 km (orange), sea level (blue), and the difference between sea level and 8km (green) for (b) ERA-interim and (c) the S2S prediction systems associated with one standard deviation of the ST100 index. The regression based on the model predictions is first averaged over ensemble members and then over the different prediction systems (i.e., the multi-model average). '×' indicates high-top models. Shading corresponds to 1.5 standard errors around the multi-system mean.

4.2 Prediction of the Conditions Following Stratospheric Events

Stratospheric events can have a significant surface impact in the extratropical Northern Hemisphere. This is here quantified as the 2-meter temperature anomalies for weeks 3-4 following weak and strong vortex events (Figure 4). Weak and strong vortex states are determined based on the strength of the zonal mean zonal wind at 60°N, 10 hPa in the reanalysis using the following criteria:

$$\text{weak vortex} = \overline{u}_{60N,10hPa} < 5ms^{-1} \tag{4}$$

$$\text{strong vortex} = \overline{u}_{60N,10hPa} > 40ms^{-1} \tag{5}$$

where the overbar denotes the zonal mean. These thresholds were chosen to be close to the ones used in Tripathi, Charlton-Perez, et al. (2015), except that here the thresholds are relaxed in order to allow for sufficient event statistics due to the limited common period covered in the S2S prediction systems. A sensitivity test varying the thresholds by $5ms^{-1}$ does not yield qualitative differences. The forecast anomalies are compared to those of a control population of forecasts determined separately for the weak and strong vortex cases. For example, for each weak vortex event, the control is taken from the same day of the year for all other years within the dataset provided it does not fall into the weak or strong category. For example, for BoM, which covers 1981 to 2013, the first observed weak vortex state by the criterion (4) occurred on 6th Feb 1981. Of the 6th Feb forecast initializations of the remaining years in the 1981-2013 period, 21 had a vortex state that was not characterized as weak or strong according to the criteria (4)-(5), so those 21 forecasts initialized on 6th Feb were added to the control population. This was repeated for each subsequent weak vortex state giving rise to the large control samples listed in Table 2. The control forecasts have roughly the same distribution in terms of seasonality as the weak forecasts. Note that for the BoM prediction system, only the first 24 of the 33 members were used in this analysis (see Methods section). Otherwise, all forecasts within the December to March season are used and we consider the average over weeks 3-4 of the forecasts. It should be noted that for models that have frequent initializations there may be multiple forecasts that are initialized over the course of a particular stratospheric event and so the individual forecasts are not entirely independent, but the same will be true for the accompanying control forecasts.

The surface anomalies following weak vortex events are strongest over Eurasia and northeastern Canada, with cold anomalies over Siberia, Scandinavia, and northern Greenland, and warm anomalies over Alaska, northeastern Canada, the Middle East, and northern Africa (Fig. 4a). The anomalies in the prediction systems appear smoother due to the larger sample size, but overall the anomaly patterns are well represented (Fig. 4b). The main differences exist in the magnitude of the anomalies: warm anomalies are generally stronger in ERA-interim for both weak and strong vortex events. The cold anomalies in strong vortex events are of the same order for the reanalysis and the multi-model mean (Fig. 4c,d), while the cold anomaly over Eurasia after weak vortex events extends further west over Eurasia in the multi-model mean compared to reanalysis.

We consider the dependence of forecast skill on vortex initialization state using the definitions of weak and strong vortex states described above. The use of these definitions of vortex strength increases the sample size of forecasts characterized as WEAK compared to objective definitions of SSW events, but comparison will be made for forecasts initialized on the SSW dates defined in Part I. For this comparison, we define the SSW forecasts as the first forecast that is initialized on or after the SSW onset date and define the CONTROL forecasts as the forecasts for the same day of the year for all other years within the dataset for which a SSW does not occur. This sampling method differs slightly from that used in Sigmond et al. (2013) in that a slightly different definition of SSW dates is used, and instead of only using the forecasts from the year before and after the SSW year as control, we make use of the equivalent date from all years of the dataset that do not contain a SSW during the winter. Note that, unlike for WEAK and STRONG, only one forecast initialization date is used, per event, considerably reducing the sample size. The number of events sampled as WEAK, STRONG or SSW and their associated controls are listed in Table 2.

Figure 5 shows the difference in skill in 2m temperature between the WEAK/STRONG forecasts and their associated controls, considering both the correlation coefficient r (equation 1) and RMSE (equation 3) as defined in section 2.2. The largest differences based on vortex initialization state are found for the correlation in the case of weak vortex events, although these differences do not represent a uniform increase in skill over Northern Hemisphere land regions. Regions that show an apparent increase in skill are Eastern Rus-

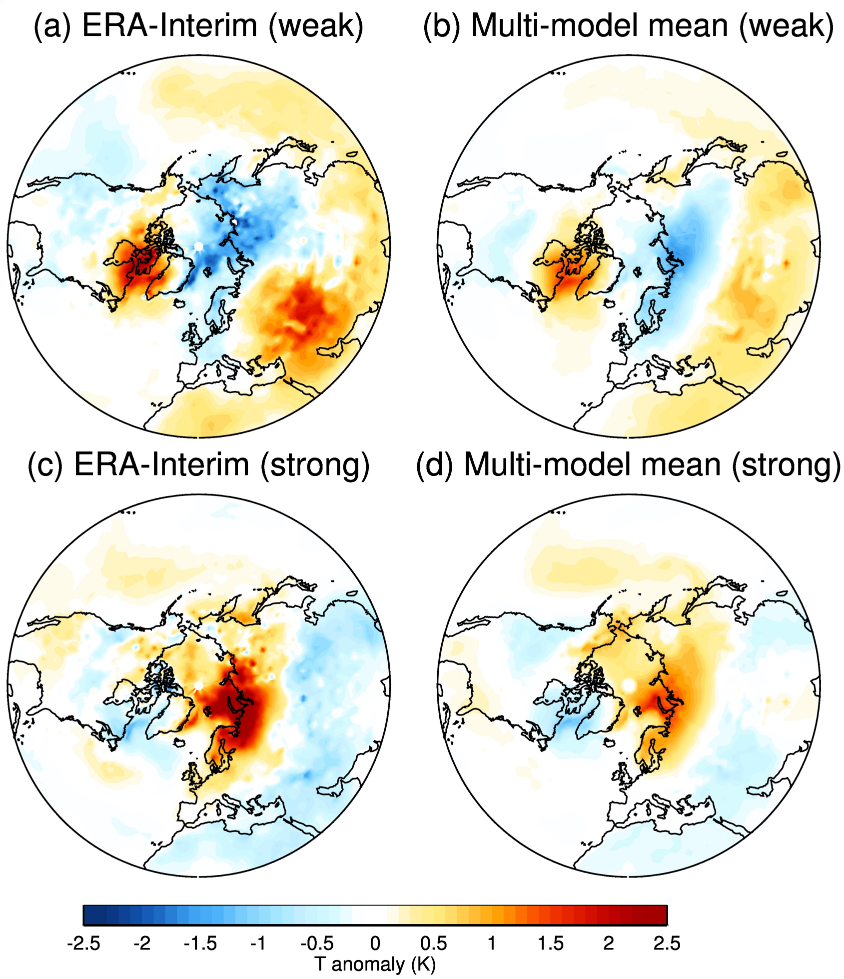


Figure 4. Composite 2m temperature anomalies (K) for weeks 3-4 for (top) weak vortex states and (bottom) strong vortex states. (b)/(d) show the multi-model mean for forecasts initialized during weak/strong vortex states. (a)/(c) shows the equivalent anomalies for ERA-interim where each date present in the multi-model mean in (b)/(d) has been given an equivalent weighting. The individual prediction systems for (b) are shown in Figure S4.

sia, the Middle East and the central USA. Given the anomalies associated with weak vortex states shown in Fig. 4 the increased skill over Eastern Russia and the Middle East is not too surprising since these are regions where weak vortex events are accompanied by substantial temperature anomalies that the forecast systems are capable of capturing. The central USA is characterized by much weaker negative temperature anomalies in association with weak vortex events, although the sign is consistent between ERA-interim and the forecast systems and so this may be giving rise to the enhancement in skill. These three regions are also characterized by a reduction in RMSE.

The extent to which these increases in skill are significant and consistent across the models can be assessed from Fig. 6a,b, where the change in ACC ((equ. 2) as defined in section 2.2) along with uncertainties are presented for these regions. Over Russia, the central USA and the Middle East, the models are rather consistent in showing an increase in ACC during the weak vortex events (Fig. 6a) although this increase is only significant for roughly half of the models in each region. The models are also rather consis-

Table 2. Number of forecasts going into WEAK and STRONG vortex categories, the number of forecasts classified as SSW forecasts, and the associated number of control forecasts. '×' indicates high-top models.

Model	Weak	Weak_Control	Strong	Strong_Control	SSW	SSW_Control
BoM[†]	107	2278	198	3592	18	288
CMA	351	4741	557	6763	12	120
ECCC	39	365	126	1202	12	96
ECMWF[×]	103	1274	127	1382	12	84
CNR-ISAC	100	1901	186	2933	17	238
JMA[×]	58	1089	86	1401	17	255
UKMO[×]	51	737	91	1167	12	132

[†] Here, only 24 members of BoM were used.

tent in showing a reduction in RMSE in Russia and the central USA, but they are less consistent in this measure for the Middle East.

A notable region of reduced skill during weak vortex events arises over Europe (Fig. 5c). While we cannot directly relate the change in skill shown in Fig. 5c to the comparison of the composites in Fig. 4, they are, at least, consistent in that the region of reduced skill over Europe during weak vortex events is a region where the model and reanalysis WEAK composites differ (Figs. 4a,b). The forecast systems suggest that the zero line of surface temperature anomalies roughly cuts through central Europe with cold anomalies to the North and warm anomalies to the South (Fig. 4b), with some variability between individual models (Fig. S4). The ERA-interim composite, however, shows the zero line further north with warm anomalies extending northward from the Middle-East into eastern Europe/western Russia. As a result, the ERA-interim and S2S forecast anomalies differ in sign in this region. Without more verification dates, it is difficult to determine whether this is just because the WEAK vortex composite in ERA-interim is impacted by other unrelated variability, or whether the canonical temperature anomalies that accompany weak vortex events in the real world are different to those in the model. Indeed, only 3 out of the 8 models suggest this reduction in skill is significant (Fig. 6a).

For vortex initializations during strong vortex states there is less consistency among the models on the change in forecast skill (Fig. 6c and d). The only possible exceptions are that for RMSE, almost all the models suggest a reduction in RMSE and hence increased skill over Russia and Europe.

Finally, to provide a comparison with the results of Sigmond et al. (2013), the anomalous skill associated with initialization during SSW events is summarized in Figs. 6e,f. Again, the models are somewhat consistent in showing an increase in ACC over Russia, central USA and the Middle East after SSWs and a decrease over Europe, although there is less consistency than for the WEAK vortex events, presumably due to the limited sample size. There is also less consistency for the RMSE, with the central USA being the only region where the majority of models exhibit a reduced RMSE. That being said, the limited sample size for this assessment leads to very large uncertainty ranges.

As a final comparison with previous work and to summarize the skill associated with weak and strong vortex events in the S2S models, the analysis is repeated for the NAM index at 100 and 1000 hPa. The NAM index is calculated by projecting daily anomalies from each ensemble member onto the NAM loading pattern computed as the first empirical orthogonal function of ERA-interim zonal mean geopotential height between 20° - 90°N. An identical method to that used for 2m temperature for selecting forecasts initialised during weak, neutral and strong vortex is used. The skill of forecasts from weak and strong initializations is compared to a representative control forecast for each state separately as above.

For the lower stratosphere, there is a clear and robust gain in correlation for both weak and strong vortex events in almost all models with the exception of CNRM-Meteo (Fig. 7). In contrast, differences in RMSE are generally small and not significant. For the NAM at 1000 hPa, differences in correlation are smaller and in some models not significant. Notably, the UKMO model shows a large gain in correlation skill at 1000 hPa, particularly for weak vortex events. As at 100 hPa, differences in RMSE are not significant for any of the forecasting systems. The results of the skill calculations for the NAM index are consistent with the results of Sigmond et al. (2013) and Tripathi, Charlton-Perez, et al. (2015) showing modest but significant gains in correlation for both weak and strong vortex events.

5 Discussion and Outlook

In this study, we have examined the predictability arising from stratosphere-troposphere coupling in the operational S2S prediction systems contained within the S2S database (Vitart et al., 2017). We have investigated the notion that the probabilistic prediction of stratospheric events can be enhanced using remote effects from the troposphere and the tropics, and that the coupling between the stratosphere and the troposphere can lead to enhanced predictability of surface weather on S2S timescales.

In more detail, precursors to extratropical stratospheric variability in the extratropical and tropical troposphere and the tropical stratosphere are expected to lead to enhanced, probabilistic predictability for extratropical stratospheric extreme events. The S2S models represent the large-scale anomaly patterns generally observed in the troposphere before sudden stratospheric warming events, though with a weaker amplitude as compared to reanalysis, and with a better representation of the strengthening of the Aleutian low in the Pacific as opposed to the ridging anomalies over Eurasia. In addition to extratropical tropospheric anomalies, the potential of probabilistic predictability on S2S timescales is suggested by teleconnections from tropical phenomena such as the QBO, ENSO, and the MJO. Several high-top S2S models are able to represent the weakening of the polar vortex depending on the phase of these tropical precursors.

Once a stratospheric extreme event occurs, it can be long-lived in the lower stratosphere and have an impact on the troposphere. The S2S models successfully represent the extra-tropical tropospheric response to stratospheric signals throughout the tropospheric column, and the multi-model mean of the S2S systems successfully represents the surface temperature anomaly response after weak and strong vortex events at 3-4 week lead times. Since the surface impact of stratospheric events is long-lived, the exact timing of the stratospheric event, which is more difficult to forecast (see Part I), tends to be less crucial for the duration of tropospheric effects, however it may be important for the onset of anomalous weather. Although remote influences from the tropics also affect tropospheric weather directly, many of these teleconnections have a pathway through the stratosphere, and the stratosphere can therefore act as a modulation and as an additional source for S2S prediction. Despite the significant surface impact of the stratosphere, enhanced predictability of 2m temperature anomalies linked to weak and strong vortex events, and in particular for SSW events, is more difficult to show. For several regions we cannot demonstrate enhanced predictability, at least in part because of the limited record available for hindcast verification, as well as due to some of the models not capturing the correct response locally. Overall, a strong reduction in forecast error and an increase in skill at lead times of 3-4 weeks can be observed over Russia, the USA, and the Middle East after weak vortex events, but not for Europe. For strong vortex events, the increase in predictability is overall less pronounced in these regions, but Europe tends to be better predicted than after weak vortex events. Initializations at the time of SSW events (instead of weak vortex events) show a much higher variability between prediction systems, likely due to the smaller number of available events, with some models showing a decrease in skill / increase in error. Predictions of the NAM index at the surface show a more consistent increase in skill for most models. This suggests that while 2m temperature tends to be difficult to forecast, the prediction of large-scale patterns has skill that could be used to forecast different fields for individual forecasting systems (e.g. Scaife, Arribas, et al., 2014). Further research will have to be conducted to investigate the model differences and to further validate the change in skill for different lead times.

The findings of this study confirm that the stratosphere represents a potentially important ingredient for S2S prediction in winter, despite the difficulty of showing increased predictability for several regions, in particular over Europe. Prediction systems that only include a limited representation of the stratosphere perform more poorly than prediction systems with a better representation of the stratosphere, confirming the results from Butler et al. (2016); Kawatani et al. (2019). This indicates that any effort to make S2S predictions for the extratropical regions of both hemispheres will likely benefit from including a properly represented stratosphere.

These results should be used as a motivation to include a more complete representation of the stratosphere in S2S model predictions and to include information on stratospheric levels in databases used for sharing S2S predictions. An improved representation of the stratosphere, including a better representation of critical physics, and an improved long-range prediction of the stratosphere itself (see Part I) may significantly benefit the prediction of surface weather. While the here presented model intercomparison and assessment is able to give a broad overview of the currently available skill related to the stratosphere, more detailed studies with respect to the documented phenomena and processes involved will have to be performed.

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

Ambaum, M. H. P., & Hoskins, B. (2002). The nao troposphere-stratosphere connection. *Journal of Climate*, *15*, 1969–1978.

Andrews, Martin B, Knight, Jeff R, Scaife, Adam A, Lu, Yixiong, Wu, Tongwen, Gray, Lesley J, & Schenzinger, Verena. (2019). Observed and simulated teleconnections between the stratospheric Quasi-Biennial Oscillation and Northern Hemisphere winter atmospheric circulation. *Journal of Geophysical Research-Atmospheres*, *124*(3), 1219–1232.

Ayarzagüena, B., Barriopedro, D., Perez, J. M. G., Abalos, M., de la Camara, A., Herrera, R. G., ... Ordóñez, C. (2018). Stratospheric Connection to the Abrupt End of the 2016/2017 Iberian Drought. *Geophysical Research Letters*, *45*(22), 12,639–12,646.

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.

Baker, L. H., Shaffrey, L. C., Sutton, R. T., Weisheimer, A., & Scaife, A. A. (2018). An Intercomparison of Skill and Overconfidence/Underconfidence of the Wintertime North Atlantic Oscillation in Multimodel Seasonal Forecasts. *Geophysical Research Letters*, *45*(15), 7808–7817.

Baldwin, M. P. (2001). Annular modes in global daily surface pressure. *Geophysical*

Research Letters.

Baldwin, M. P., Birner, T., & Ayarzagüena, B. (2019). Tropospheric amplification of stratospheric variability. *EGU General Assembly Conference, Vienna, Austria*.

Baldwin, M. P., & Dunkerton, T. J. (1999). Propagation of the Arctic Oscillation from the stratosphere to the troposphere. *Journal of Geophysical Research*, *104*(D24), 30937.

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.

Bao, Ming, Tan, Xin, Hartmann, Dennis L, & Ceppi, Paulo. (2017). Classifying the tropospheric precursor patterns of sudden stratospheric warmings. *Geophysical Research Letters*, *44*(15), 8011–8016.

Beerli, R., Wernli, H., & Grams, C. M. (2017). Does the lower stratosphere provide predictability for month-ahead wind electricity generation in Europe? *Quarterly Journal of the Royal Meteorological Society*, *143*(709), 3025–3036.

Birner, T., & Albers, J. R. (2017). Sudden Stratospheric Warmings and Anomalous Upward Wave Activity Flux. *SOLA*, *13A*(Special Edition), 8–12.

Buizza, R., & Leutbecher, M. (2015). The forecast skill horizon. *Quarterly Journal of the Royal Meteorological Society*, *141*(693), 3366–3382.

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., Charlton-Perez, A., Domeisen, D. I. V., Garfinkel, C., Gerber, E. P., Hitchcock, P., . . . Son, S.-W. (2018). Sub-seasonal predictability and the stratosphere. In A. W. Robertson & F. Vitart (Eds.), *Sub-seasonal to seasonal prediction* (p. 585). Elsevier.

Butler, A. H., Perez, A. C., Domeisen, D. I. V., Simpson, I. R., & Sjoberg, J. (2019, September). Predictability of Northern Hemisphere Final Stratospheric Warmings and Their Surface Impacts. *Geophysical Research Letters*, *43*, 23.

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

Butler, A. H., Polvani, L. M., & Deser, C. (2014). Separating the stratospheric and tropospheric pathways of El Niño - Southern Oscillation teleconnections. *Environmental Research Letters*, *9(2)*. doi: 10.1088/1748-9326/9/2/024014

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.

Charlton, A., & 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., 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.

Cohen, J., & Entekhabi, D. (1999). Eurasian snow cover variability and northern hemisphere climate predictability. *Geophysical Research Letters*, *26*(3), 345–348.

Cohen, J., & Jones, J. (2011). Tropospheric Precursors and Stratospheric Warmings. *Journal of Climate*, *24*(24), 6562–6572.

Davies, H. C. (1981). An Interpretation of Sudden Warmings in Terms of Potential Vorticity. *Journal of the Atmospheric Sciences*, *38*(2), 427–445.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., ... F, V. (2011). The ERA-interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137(656)*, 553-597.

de la Camara, A., Albers, J. R., Birner, T., Garcia, R. R., Hitchcock, P., Kinnison, D. E., & Smith, A. K. (2017). Sensitivity of Sudden Stratospheric Warmings to Previous Stratospheric Conditions. *Journal of the Atmospheric Sciences*, *74*(9), 2857–2877.

Dobrynin, M., Domeisen, D. I. V., Müller, W. A., Bell, L., Brune, S., Bunzel, F., ... Baehr, J. (2018). Improved Teleconnection-Based Dynamical Seasonal Predictions of Boreal Winter. *Geophysical Research Letters*, *44*(9-10), 2723.

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 I: Predictability of the stratosphere. *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., Badin, G., & Koszalka, I. M. (2018). How Predictable Are the Arctic and North Atlantic Oscillations? Exploring the Variability and Predictability of the Northern Hemisphere. *Journal of Climate*, *31*(3), 997–1014. doi: http://doi.org/10.1175/JCLI-D-17-0226.1

Domeisen, D. I. V., Butler, A. H., Fröhlich, K., Bittner, M., Müller, W., & Baehr, J. (2015). Seasonal predictability over Europe arising from El Niño and stratospheric variability in the MPI-ESM Seasonal Prediction System. *Journal of Climate*, *28*(1), 256–271. doi: http://doi.org/10.1175/JCLI-D-14-00207.1

Domeisen, D. I. V., Garfinkel, C. I., & Butler, A. H. (2019). The Teleconnection of El Niño Southern Oscillation to the Stratosphere. *Reviews of Geophysics*. doi: 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*(4), 2064–2071. doi: http://doi.org/10.1002/2017GL076886

Domeisen, D. I. V., Sun, L., & Chen, G. (2013). The role of synoptic eddies in the tropospheric response to stratospheric variability. *Geophysical Research Letters*, *40*, 1–5. doi: http://doi.org/10.1002/grl.50943

Douville, H. (2009). Stratospheric polar vortex influence on Northern Hemisphere winter climate variability. *Geophysical Research Letters*, *36*(18).

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.

Esler, J., & Matthewman, N. (2011). Stratospheric Sudden Warmings as Self-Tuning Resonances. Part II: Vortex Displacement Events. *Journal of the Atmospheric Sciences*, *68*, 2505–2523.

Furtado, J. C., Cohen, J. L., Butler, A. H., Riddle, E. E., & Kumar, A. (2015). Eurasian snow cover variability and links to winter climate in the CMIP5 models. *Climate Dynamics*, *45*(9-10), 2591–2605.

García-Herrera, R., Calvo, N., Garcia, R. R., & Giorgetta, M. A. (2006). Propagation of ENSO temperature signals into the middle atmosphere: A comparison of two general circulation models and ERA-40 reanalysis data. *Journal of Geophysical Research*, *111(D6)*, D06101. doi:10.1029/2005JD006061.

Garfinkel, C., Schwartz, C., Domeisen, D., Butler, A. H., Son, S.-W., & White, I. (2019). Weakening of the teleconnection of El Nino-Southern Oscillation to the Arctic stratosphere over the past few decades: What can be learned from subseasonal forecast models? *J. Geophys. Res. - Atmospheres*.

Garfinkel, C. I., Benedict, J. J., & Maloney, E. D. (2014). Impact of the MJO on the boreal winter extratropical circulation. *Geophysical Research Letters*, *41*, 6055–6062.

Garfinkel, C. I., Feldstein, S. B., Waugh, D. W., Yoo, C., & Lee, S. (2012). Observed connection between stratospheric sudden warmings and the Madden-Julian Oscillation. *Geophysical Research Letters*, *39*(18).

Garfinkel, C. I., & Hartmann, D. L. (2007). Effects of the El Niño - Southern Oscillation and the Quasi-Biennial Oscillation on polar temperatures in the stratosphere. *J. Geophys. Res.*, *115*, D20116.

Garfinkel, C. I., Hartmann, D. L., & Sassi, F. (2010). Tropospheric precursors of anomalous northern hemisphere stratospheric polar vortices. *Journal of Climate*, *23*(12), 3282–3299.

Garfinkel, C. I., & Schwartz, C. (2017). MJO-Related Tropical Convection Anomalies Lead to More Accurate Stratospheric Vortex Variability in Subseasonal Forecast Models. *Geophysical Research Letters*, *44*(19), 10,054 - 10,062. doi: 10.1002/2017GL074470

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*(15), 7855–7866.

Garfinkel, C. I., Shaw, T. A., Hartmann, D. L., & Waugh, D. W. (2012). Does the Holton-Tan mechanism explain how the Quasi-Biennial Oscillation modulates the arctic polar vortex? *Journal of the Atmospheric Sciences*, *69*, 1713–1733.

Garfinkel, C. I., Waugh, D. W., & Gerber, E. P. (2013). The effect of tropospheric jet latitude on coupling between the stratospheric polar vortex and the troposphere. *Journal of Climate*, *26*(6). doi: 10.1175/JCLI-D-12-00301.1

Gerber, E. P., 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. P., Butler, A. H., Calvo, N., Charlton-Perez, A., Giorgetta, M., Manzini, E., … Watanabe, S. (2012). Assessing and Understanding the Impact of Stratospheric Dynamics and Variability on the Earth System. *Bulletin of the American Meteorological Society*, *93*(6), 845–859.

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., Taguchi, M., Yoden, S., & Noguchi, S. (2013). Lower-stratospheric radiative damping and Polar-night Jet Oscillation events. *Journal of the Atmospheric Sciences*, *70*, 1391–1408.

Hitchcock, P., & Simpson, I. R. (2014). The Downward Influence of Stratospheric Sudden Warmings. *J. Atmos. Sci.*, *71*(10), 3856-3876.

Hitchcock, P., & Simpson, I. R. (2016). Quantifying Eddy Feedbacks and Forcings in the Tropospheric Response to Stratospheric Sudden Warmings. *Journal of the Atmospheric Sciences*, *73*(9), 3641–3657.

Holton, J. R., & Tan, H. C. (1980). The influence of the equatorial Quasi-Biennial Oscillation on the global circulation at 50 mb. *Journal of the Atmospheric Sciences*, *37(10)*, 2200-2208.

Hoskins, B., McIntyre, M., & Robertson, A. (1985). On the use and significance of isentropic potential vorticity maps. *Quarterly Journal of the Royal Meteorological Society*, *111*, 877—946.

Hoskins, B. J., & Ambrizzi, T. (1993). Rossby-Wave Propagation on a Realistic Longitudinally Varying Flow. *Journal of the Atmospheric Sciences*, *50*(12), 1661–

1671.

Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., ... Zhang, H. (2017). Extended Reconstructed Sea Surface Temperature, Version 5 (ERSSTv5): Upgrades, Validations, and Intercomparisons. *Journal of Climate*, *30*, 8179–8205.

Hurrell, J. W., Kushnir, Y., & Visbeck, M. (2001). The North Atlantic Oscillation. *Science*, *291*(5504), 603-605.

Ineson, S., & Scaife, A. (2009). The role of the stratosphere in the European climate response to El Niño. *Nature Geoscience*, *2*(1), 32–36. doi: 10.1038/ngeo381

Iza, M., Calvo, N., & Manzini, E. (2016). The Stratospheric Pathway of La Niña. *Journal of Climate*, *29*(24), 8899–8914.

Jiménez-Esteve, B., & Domeisen, D. I. V. (2018). The Tropospheric Pathway of the ENSO-North Atlantic Teleconnection. *Journal of Climate*, *31*, 4563–4584.

Jucker, M., & Reichler, T. (2018). Dynamical Precursors for Statistical Prediction of Stratospheric Sudden Warming Events. *Geophysical Research Letters*, *96*(2), 1.

Kang, W., & Tziperman, E. (2017). More frequent sudden stratospheric warming events due to enhanced MJO forcing expected in a warmer climate. *Journal of Climate*, *30*(21), 8727–8743.

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.

Kawatani, Y., Hamilton, K., Gray, L., Osprey, S., Watanabe, S., & Yamashita, Y. (2019). The effects of a well-resolved stratosphere on the simulated boreal winter circulation in a climate model. *Journal of Atmospheric Sciences*. doi: doi:10.1175/JAS-D-18-0206.1,inpress

Kidston, J., Scaife, A. A., Hardiman, S. C., Mitchell, D. M., Butchart, N., Baldwin, M. P., & Gray, L. J. (2015). Stratospheric influence on tropospheric jet streams, storm tracks and surface weather. *Nature Geoscience*, *8*(6), 433–440.

Kim, B.-M., Son, S.-W., Min, S.-K., Jeong, J.-H., Kim, S.-J., Zhang, X., ... Yoon, J.-H. (2014). Weakening of the stratospheric polar vortex by Arctic sea-ice loss. *Nature Communications*, *5*(1), 4646.

Kolstad, E. W., & Charlton-Perez, A. J. (2010). Observed and simulated precursors of stratospheric polar vortex anomalies in the Northern Hemisphere. *Climate Dynamics*, *37*(7-8), 1443–1456.

Kuroda, Y. (2008). Role of the stratosphere on the predictability of medium-range weather forecast: A case study of winter 2003-2004. *Geophysical Research Letters*, *35*(19).

Kwok, R. (2000). Recent changes in Arctic Ocean sea ice motion associated with the North Atlantic Oscillation. *Geophysical Research Letters*, *27*, 775—778.

L'Heureux, M. L., Tippett, M. K., Kumar, A., Butler, A. H., Ciasto, L. M., Ding, Q., ... Johnson, N. C. (2017). Strong relations between ENSO and the Arctic Oscillation in the North American Multimodel Ensemble. *Geophysical Research Letters*, *44*. doi: 10.1002/2017GL074854

Li, Y., Li, J., Jin, F. F., & Zhao, S. (2015, August). Interhemispheric Propagation of Stationary Rossby Waves in a Horizontally Nonuniform Background Flow. *Journal of the Atmospheric Sciences*, *72*(8), 3233–3256.

Lim, E.-P., Hendon, H. H., Boschat, G., Hudson, D., Thompson, D. W. J., Dowdy, A. J., & Arblaster, J. M. (2019). Australian hot and dry extremes induced by weakenings of the stratospheric polar vortex. *Nature Geoscience*, *12*(11), 896–901.

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

Manzini, E., Giorgetta, M. A., Esch, M., Kornblueh, L., & Roeckner, E. (2006). The influence of sea surface temperatures on the northern winter stratosphere: Ensemble simulations with the MAECHAM5 model. *Journal of Climate*, *19(16)*, 3863–3881.

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).

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.

Nishii, K., Nakamura, H., & Orsolini, Y. J. (2011). Geographical Dependence Observed in Blocking High Influence on the Stratospheric Variability through Enhancement and Suppression of Upward Planetary-Wave Propagation. *Journal of Climate*, *24*(24), 6408–6423.

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*.

Peings, Y. (2019). Ural Blocking as a driver of early winter stratospheric warmings. *Geophysical Research Letters*, 2019GL082097–18.

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., & Semeniuk, K. (2003). Downward migration of extratropical zonal wind anomalies. *Journal of Geophysical Research: Atmospheres (1984–2012)*, *108*(D7).

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. doi: 10.1175/JCLI-D-16-0277.1

Polvani, L. M., & Waugh, D. (2004). Upward wave activity flux as a precursor to extreme stratospheric events and subsequent anomalous surface weather regimes. *Journal of Climate*, *17*, 3548–3554.

Quiroz, R. S. (1986). The association of stratospheric warmings with tropospheric blocking. *Journal of Geophysical Research-Atmospheres*.

Richter, J. H., Deser, C., & Sun, L. (2015). Effects of stratospheric variability on El Niño teleconnections. *Environmental Research Letters*, *10*(12), 124021.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., ... Woollen, J. (2011). MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *Journal of Climate*, *24*(14), 3624–3648.

Rogers, J. C. (1981). The North Pacific Oscillation. *Journal of Climatology*, *1*(1), 39–57.

Scaife, A. A., Arribas, A., Blockley, E., Scaife, A. A., Arribas, A., Blockley, E., ... Williams, A. (2014). Skillful long-range prediction of European and North American winters. *Geophysical Research Letters*, *41*, 2514–2519.

Scaife, A. A., Athanassiadou, M., Andrews, M., Arribas, A., Baldwin, M. P., Dunstone, N., ... Williams, A. (2014). Predictability of the Quasi-Biennial Oscillation and its Northern winter teleconnection on seasonal to decadal timescales. *Geophysical Research Letters*, *41*(5), 1752–1758.

Scaife, A. A., Comer, R. E., Dunstone, N. J., Knight, J. R., Smith, D. M., MacLachlan, C., ... Slingo, J. (2017). Tropical rainfall, Rossby waves and regional winter climate predictions. *Quarterly Journal of the Royal Meteorological Society*, *143*(702), 1–11.

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.

Schneidereit, A., Peters, D. H. W., Keller, J. H., Teubler, F., Martius, O., Grams, C. M., ... Riemer, M. (2017). Enhanced Tropospheric Wave Forcing of Two Anticyclones in the Prephase of the January 2009 Major Stratospheric Sudden Warming Event. *Monthly Weather Review*, *145*(5), 1797–1815.

Schwartz, C., & Garfinkel, C. I. (2017). Relative roles of the MJO and stratospheric variability in North Atlantic and European winter climate. *Journal of Geophysical Research: Atmospheres*, *122*(8), 4184–4201.

Sigmond, M., Scinocca, J. F., Kharin, V. V., & Shepherd, T. G. (2013). Enhanced seasonal forecast skill following stratospheric sudden warmings. *Nature Geoscience*, *6(1)*. doi: 10.1038/ngeo1698

Simpson, I. R., Blackburn, M., & Haigh, J. D. (2009). The Role of Eddies in Driving the Tropospheric Response to Stratospheric Heating Perturbations. *Journal of the Atmospheric Sciences*, *66*(5), 1347–1365.

Simpson, I. R., Blackburn, M., & Haigh, J. D. (2012). A Mechanism for the Effect of Tropospheric Jet Structure on the Annular Mode-Like Response to Stratospheric Forcing. *Journal of the Atmospheric Sciences*, *69*(7), 2152–2170.

Simpson, I. R., Hitchcock, P., Shepherd, T. G., & Scinocca, J. F. (2011). Stratospheric variability and tropospheric annular mode timescales. *Geophys. Res. Lett.*, *38*, L20806.

Smith, K., Polvani, L., & Tremblay, L. (2018). The impact of stratospheric circulation extremes on minimum Arctic sea ice extent. *Journal of Climate*, *31*, 7169–7183.

Smith, K. L., & Kushner, P. J. (2012). Linear interference and the initiation of extratropical stratosphere-troposphere interactions. *Journal of Geophysical Research*, *117*(D13). doi: 10.1029/2012JD017587

Smith, K. L., & Scott, R. K. (2016). The role of planetary waves in the tropospheric jet response to stratospheric cooling. *Geophysical Research Letters*, *43*(6), 2904–2911.

Song, K., & Son, S.-W. (2018). Revisiting the ENSO–SSW Relationship. *Journal of Climate*, *31*(6), 2133–2143.

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

Stockdale, T. N., Molteni, F., & Ferranti, L. (2015). Atmospheric initial conditions and the predictability of the Arctic Oscillation. *Geophysical Research Letters*, *42*(4), 1173–1179.

Sun, L., Deser, C., & Tomas, R. A. (2015). Mechanisms of Stratospheric and Tropospheric Circulation Response to Projected Arctic Sea Ice Loss. *Journal of Climate*, *28*, 7824–7845.

Thompson, D., & Wallace, J. (1998). Observed linkages between Eurasian surface air temperature, the North Atlantic Oscillation, Arctic sea level pressure and the stratospheric polar vortex. *Geophysical Research Letters*, *25*, 1297–1300.

Thompson, D., Wallace, J. M., & Hegerl, G. C. (2000). Annular modes in the extratropical circulation. Part II: Trends. *Journal of Climate*, *13*(5), 1018–1036.

Thompson, D. W. J., & Wallace, J. (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., 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.

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.

Walker, G. (1928). World weather. *Quarterly Journal of the Royal Meteorological Society*, *54*(226), 79–87.

Wheeler, M. C., & Hendon, H. H. (2004). An all-season real-time multivariate mjo index: Development of an index for monitoring and prediction. *Monthly Weather Review*, *132*(8), 1917–1932. doi: 10.1175/1520-0493(2004)132⟨1917:AARMMI⟩2.0.CO;2

White, C. J., Carlsen, H., Robertson, A. W., Klein, R. J. T., Lazo, J. K., Kumar, A., ... Zebiak, S. E. (2017). Potential applications of subseasonal-to-seasonal (S2S) predictions. *Meteorological Applications*, *24*(3), 315–325.

White, I., Garfinkel, C. I., Gerber, E. P., Jucker, M., Aquila, V., & Oman, L. D. (2018). The Downward Influence of Sudden Stratospheric Warmings: Association with Tropospheric Precursors. *Journal of Climate*, *32*(1), 85–108.

Woollings, T., Charlton-Perez, A., Ineson, S., Marshall, A. G., & Masato, G. (2010). Associations between stratospheric variability and tropospheric blocking. *Journal of Geophysical Research: Atmospheres (1984–2012)*, *115*(D6).

Zhang, F., Sun, Y. Q., Magnusson, L., Buizza, R., Lin, S.-J., Chen, J.-H., & Emanuel, K. (2019, April). What Is the Predictability Limit of Midlatitude Weather? *Journal of the Atmospheric Sciences*, *76*(4), 1077–1091.

Zhang, P., Wu, Y., Simpson, I. R., Smith, K. L., Zhang, X., De, B., & Callaghan, P. (2018). A stratospheric pathway linking a colder Siberia to Barents-Kara Sea sea ice loss. *Science advances*, *4*(7), eaat6025.

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/. MERRA-2 Reanalysis data was obtained from NASA at https://disc.gsfc.nasa.gov/. 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).

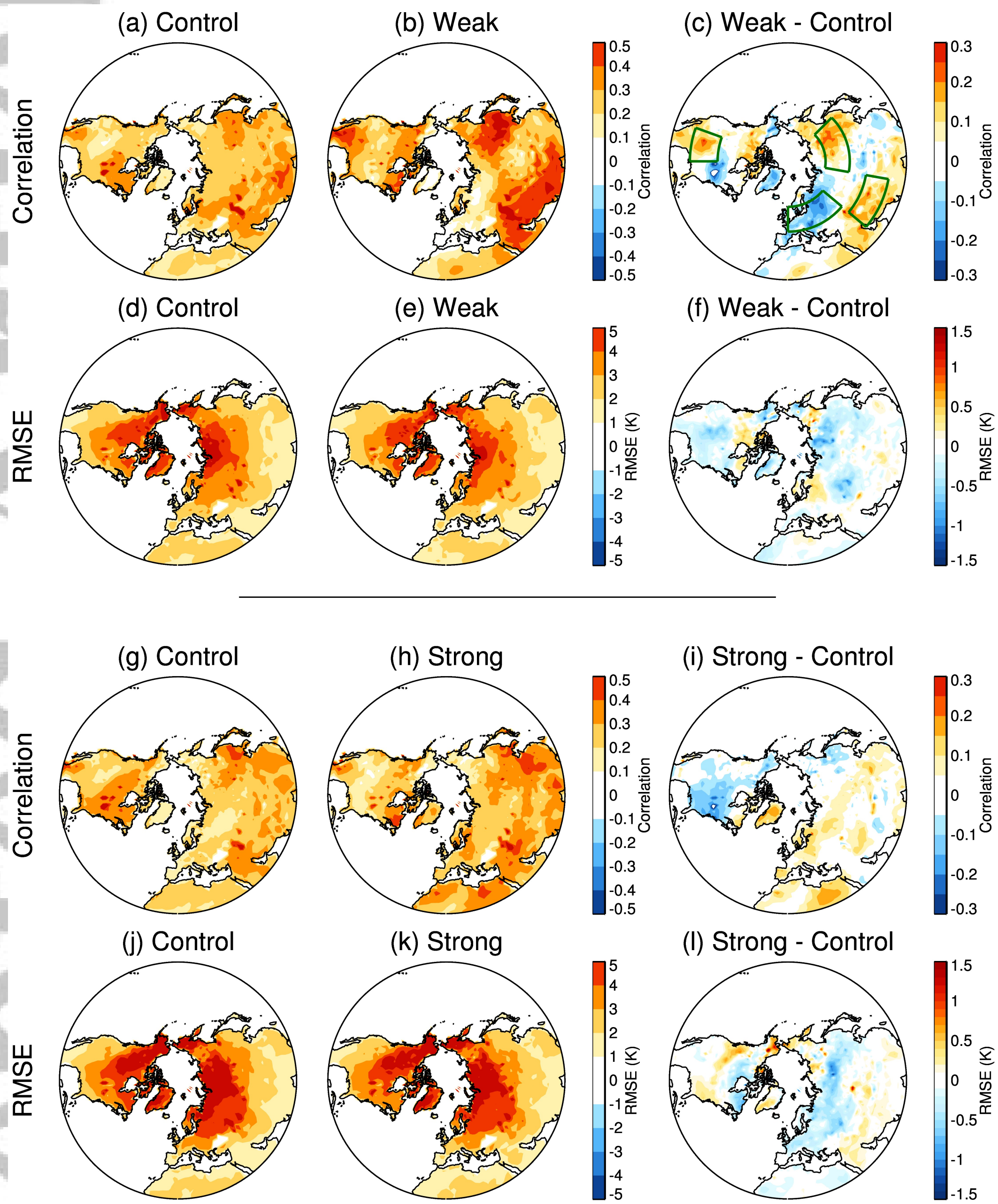


Figure 5. Multi-model mean correlation (see equation 1) and RMSE computed for 2m temperature. (a)-(f) The difference in skill between WEAK and Control forecasts for (top) correlation coefficient and (bottom) RMSE. (left) shows Control forecasts, (middle) shows WEAK vortex forecasts and (right) shows the difference between WEAK and Control forecasts. (g)-(l) are as (a)-(f) but for STRONG vortex initializations. The green boxes in (c) depict the averaging regions used in Fig. 6.

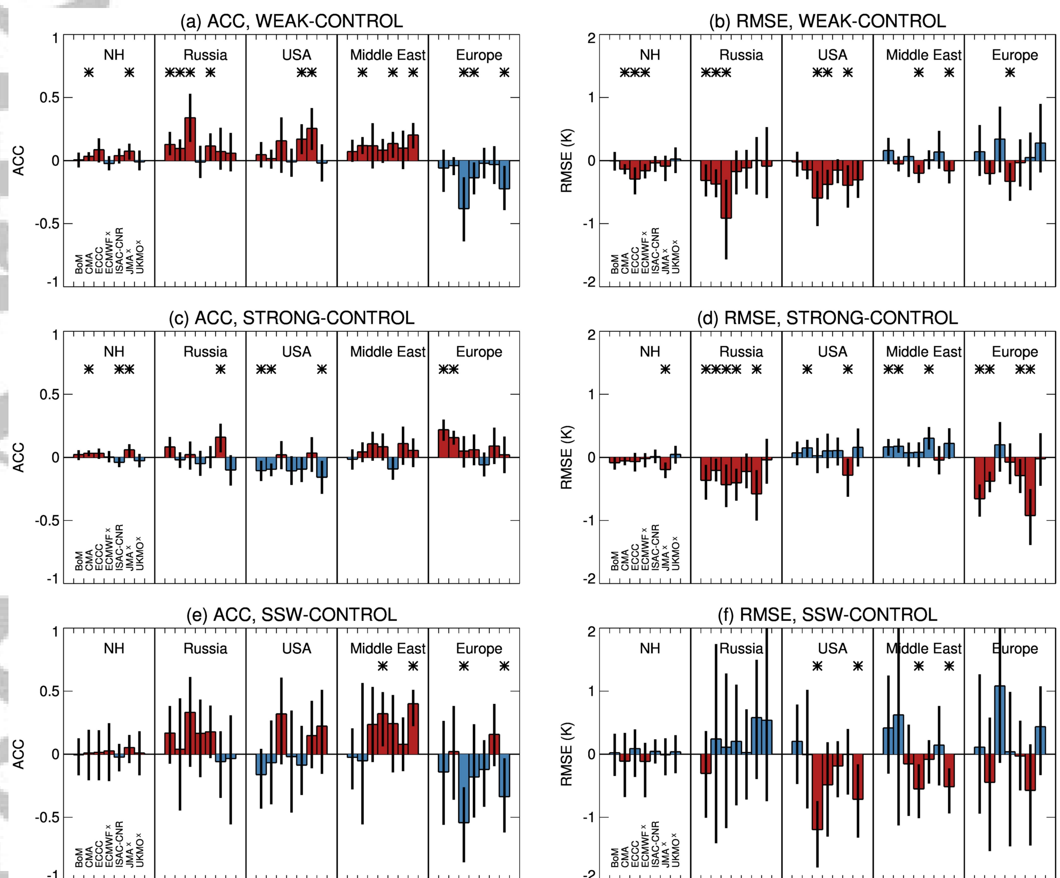


Figure 6. (left) ACC (equation 1) and (right) RMSE (equation 3) for 2m temperature for (top) the difference between WEAK vortex initializations and Control forecasts, (middle) the difference between STRONG vortex initializations and Control forecasts and (bottom) the difference between SSW initializations and Control forecasts. The regions considered (depicted by the green boxes in Fig. 5c) are as follows: NH = the area average from 30°-90°N, Russia = 80°-135°E, 50°-65°N, USA=250°-270°E, 30°-45°N, Middle-East=50°-80°E, 28°-40°N and Europe=0°-50°E, 45°-60°N. Red bars indicate an improvement and blue bars depict a degradation. The error bars indicate the 2.5th to 97.5th percentile range of the difference determined via bootstrapping for WEAK/STRONG/SSW forecasts and Control forecasts with replacement, 200 times to obtain 200 estimates of the skill difference. Asterisks indicate cases where this error bar does not encompass zero, i.e., cases where the difference is significant [$p < 0.05$] using a 2-sided test. '×' indicates high-top models.

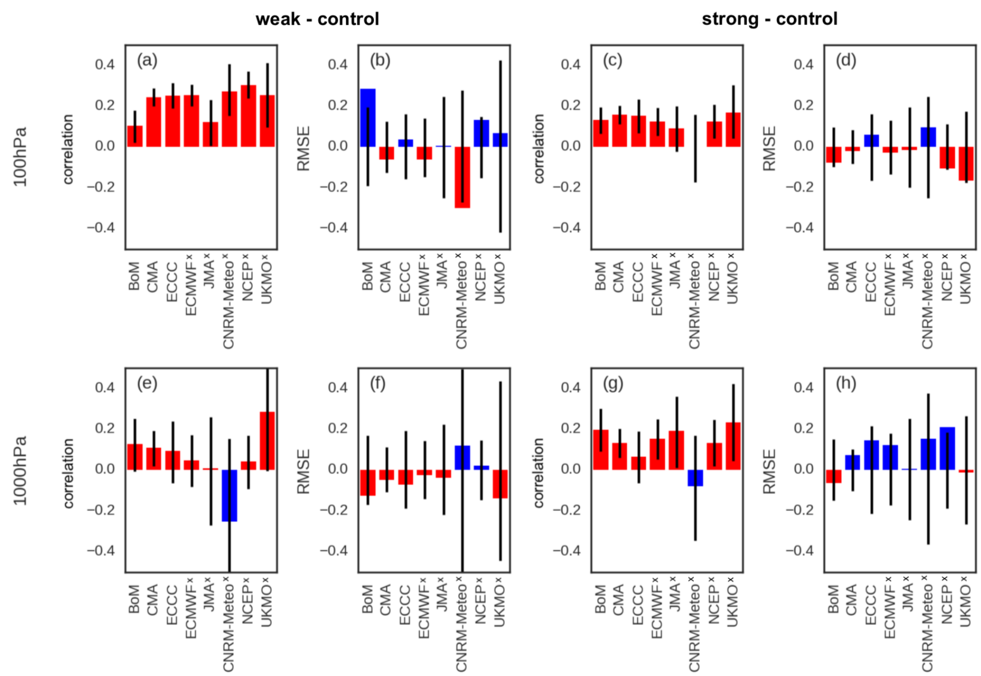
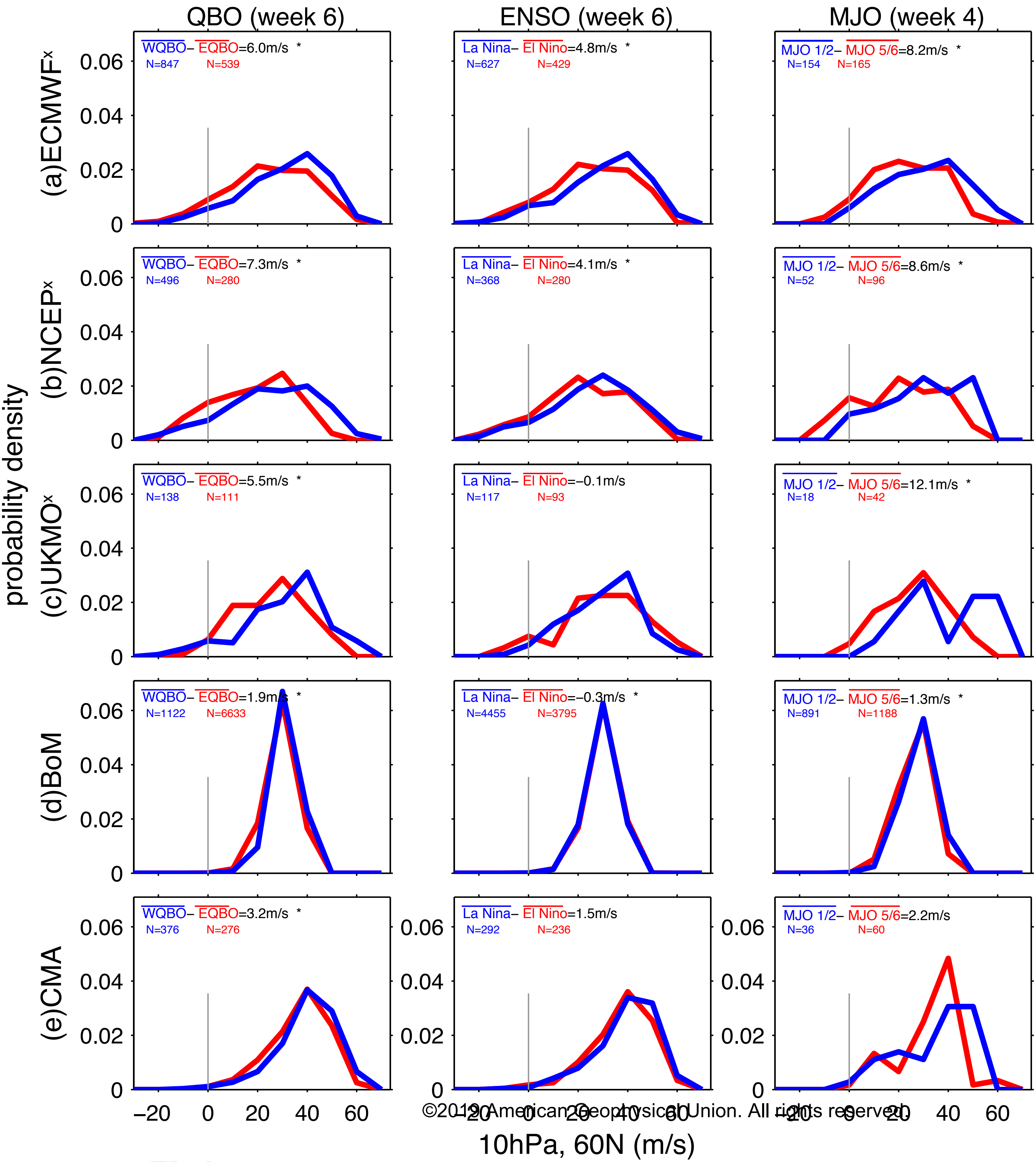


Figure 7. Differences in skill for forecasts initialized during weak (a,b,e,f) and strong vortex (c,d,g,h) for the NAM index at 100 hPa (top) and 1000 hPa (bottom) for the correlation coefficient (equation 1) (a,c,e,g) and RMSE (equation 3) (b,d,f,h). Where the difference represents an improvement (degradation) in skill the bar is plotted in red (blue). Confidence intervals ($p < 0.05$, estimated from a 10,000 bootstrap sample with replacement) are shown in black lines. All metrics are calculated for the average NAM for weeks 3 and 4. Note that for this analysis, model data was not available for CNR-ISAC and so this model is not included. '×' indicates high-top models.

Figure.



QBO (week 6)
ENSO (week 6)
MJO (week 4)
(a)ECMWFˣ
(b)NCEPˣ
(c)UKMOˣ
(d)BoM
(e)CMA
probability density
WQBO– EQBO=6.0m/s *
N=847 N=539
La Nina– El Nino=4.8m/s *
N=627 N=429
MJO 1/2– MJO 5/6=8.2m/s *
N=154 N=165
WQBO– EQBO=7.3m/s *
N=496 N=280
La Nina– El Nino=4.1m/s *
N=368 N=280
MJO 1/2– MJO 5/6=8.6m/s *
N=52 N=96
WQBO– EQBO=5.5m/s *
N=138 N=111
La Nina– El Nino=–0.1m/s
N=117 N=93
MJO 1/2– MJO 5/6=12.1m/s *
N=18 N=42
WQBO– EQBO=1.9m/s *
N=1122 N=6633
La Nina– El Nino=–0.3m/s *
N=4455 N=3795
MJO 1/2– MJO 5/6=1.3m/s *
N=891 N=1188
WQBO– EQBO=3.2m/s *
N=376 N=276
La Nina– El Nino=1.5m/s
N=292 N=236
MJO 1/2– MJO 5/6=2.2m/s
N=36 N=60
0.06
0.04
0.02
0
–20
20
40
60
10hPa, 60N (m/s)

Figure.

a)
BOM
CMA
ECCC
ECMWF x
JMA x
UKMO x
CNR-ISAC
ERA-Interim
z [km]
p [hPa]
0
5
10
15
20
25
0
0.5
1
1.5
2
b)
Regres anom. P on ST100 index
(ERA-Interim)
pressure [hPa]
15 km
0 km
0km - 8km
0
0.5
1
1.5
2
1
20
40
60
c)
(S2S systems mean)
pressure [hPa]
0 km
0km - 8km
15 km
lag[days]
0
0.5
1
1.5
2
1
20
40
60

Figure.

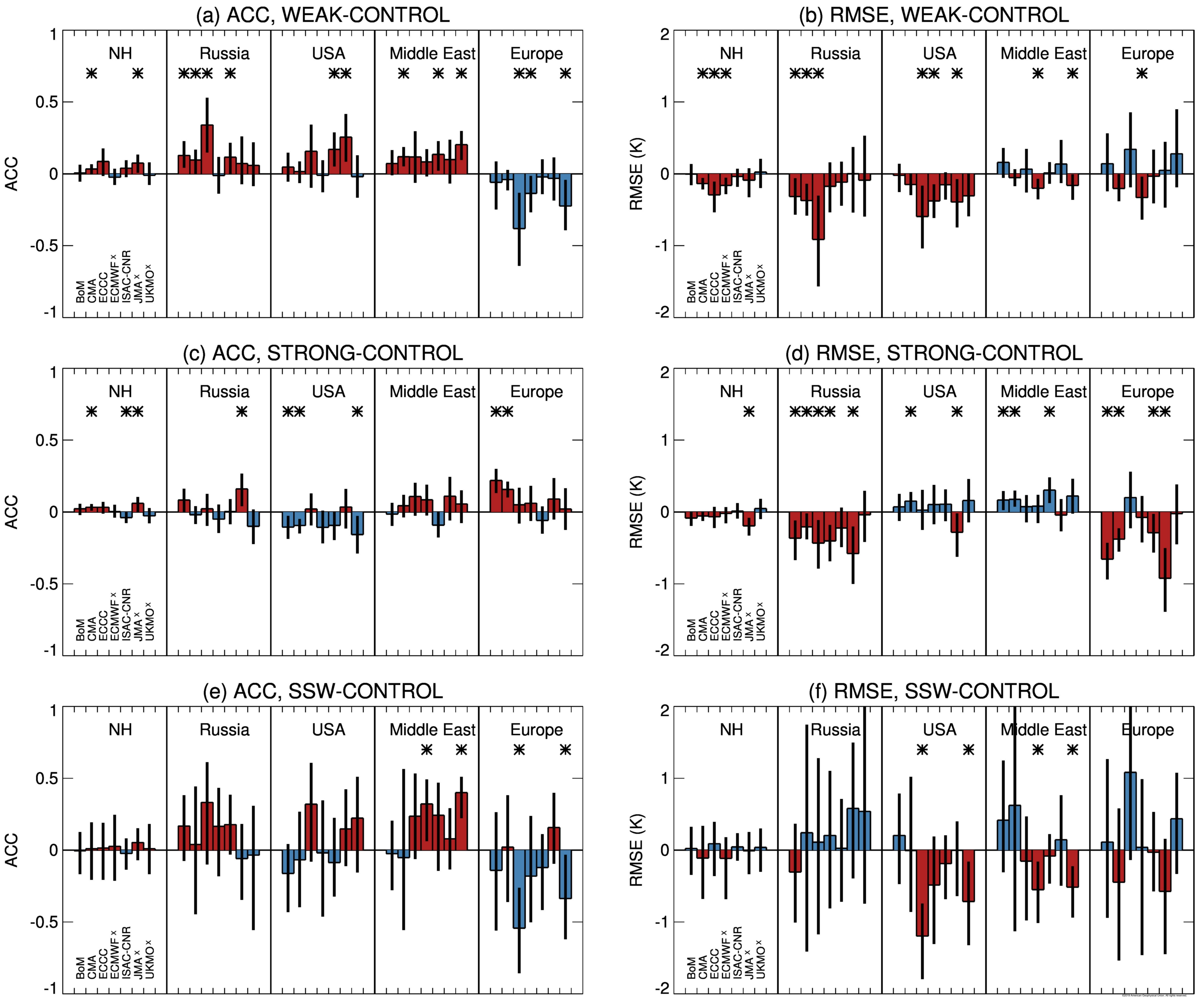

(a) ACC, WEAK-CONTROL
(b) RMSE, WEAK-CONTROL
(c) ACC, STRONG-CONTROL
(d) RMSE, STRONG-CONTROL
(e) ACC, SSW-CONTROL
(f) RMSE, SSW-CONTROL
ACC
RMSE (K)
NH
Russia
USA
Middle East
Europe
BoM
CMA
ECCC
ECMWF ×
ISAC-CNR
JMA ×
UKMO ×

Figure.

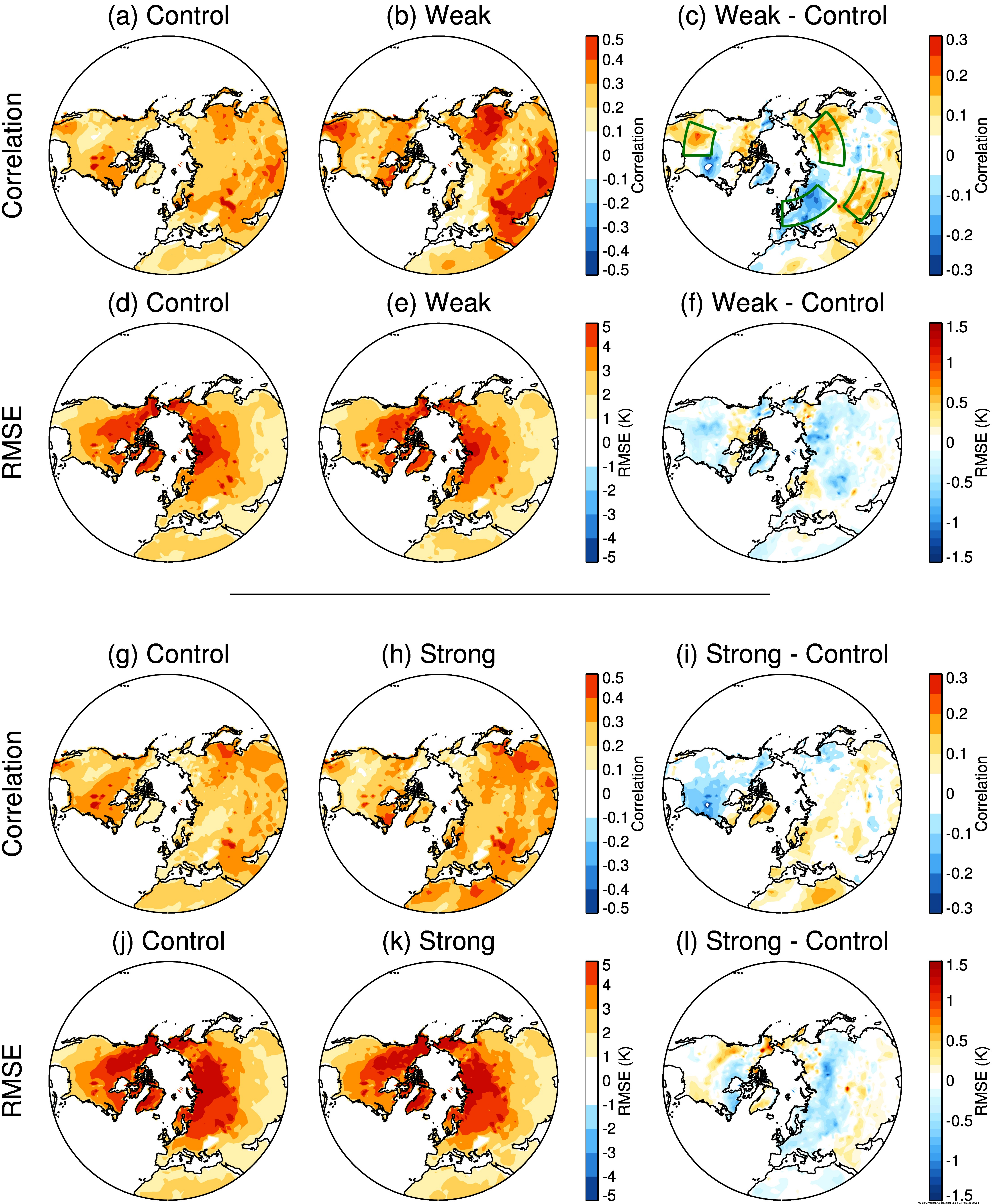
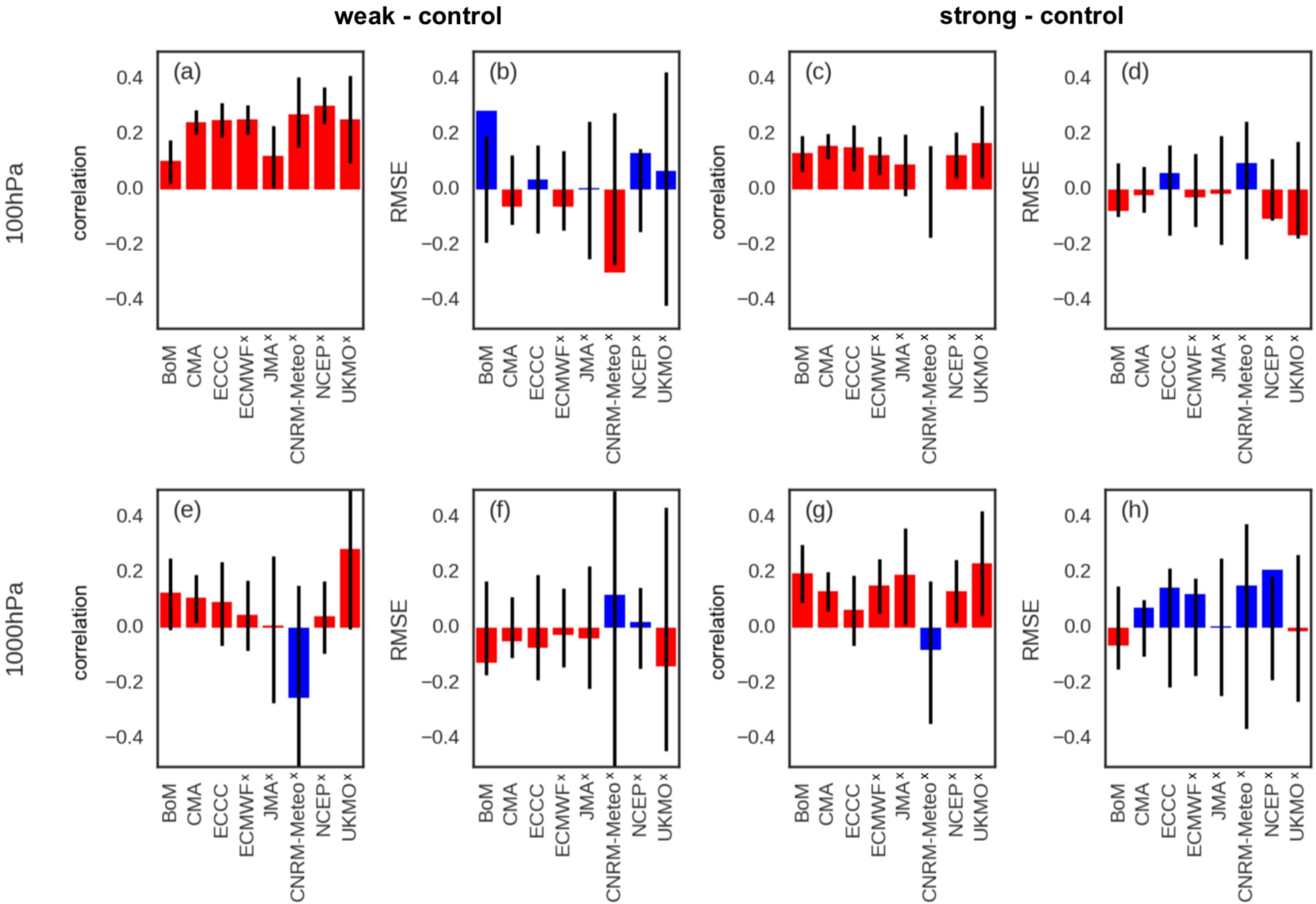

(a) Control
(b) Weak
(c) Weak - Control
Correlation
(d) Control
(e) Weak
(f) Weak - Control
RMSE
RMSE (K)
(g) Control
(h) Strong
(i) Strong - Control
Correlation
(j) Control
(k) Strong
(l) Strong - Control
RMSE
RMSE (K)
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
5
4
3
2
1
0
-1
-2
-3
-4
-5
1.5
1
0.5
0
-0.5
-1
-1.5

Figure.



weak - control
strong - control
100hPa
1000hPa
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
correlation
RMSE
0.4
0.2
0.0
−0.2
−0.4
BoM
CMA
ECCC
ECMWF x
JMA x
CNRM-Meteo x
NCEP x
UKMO x

Figure.

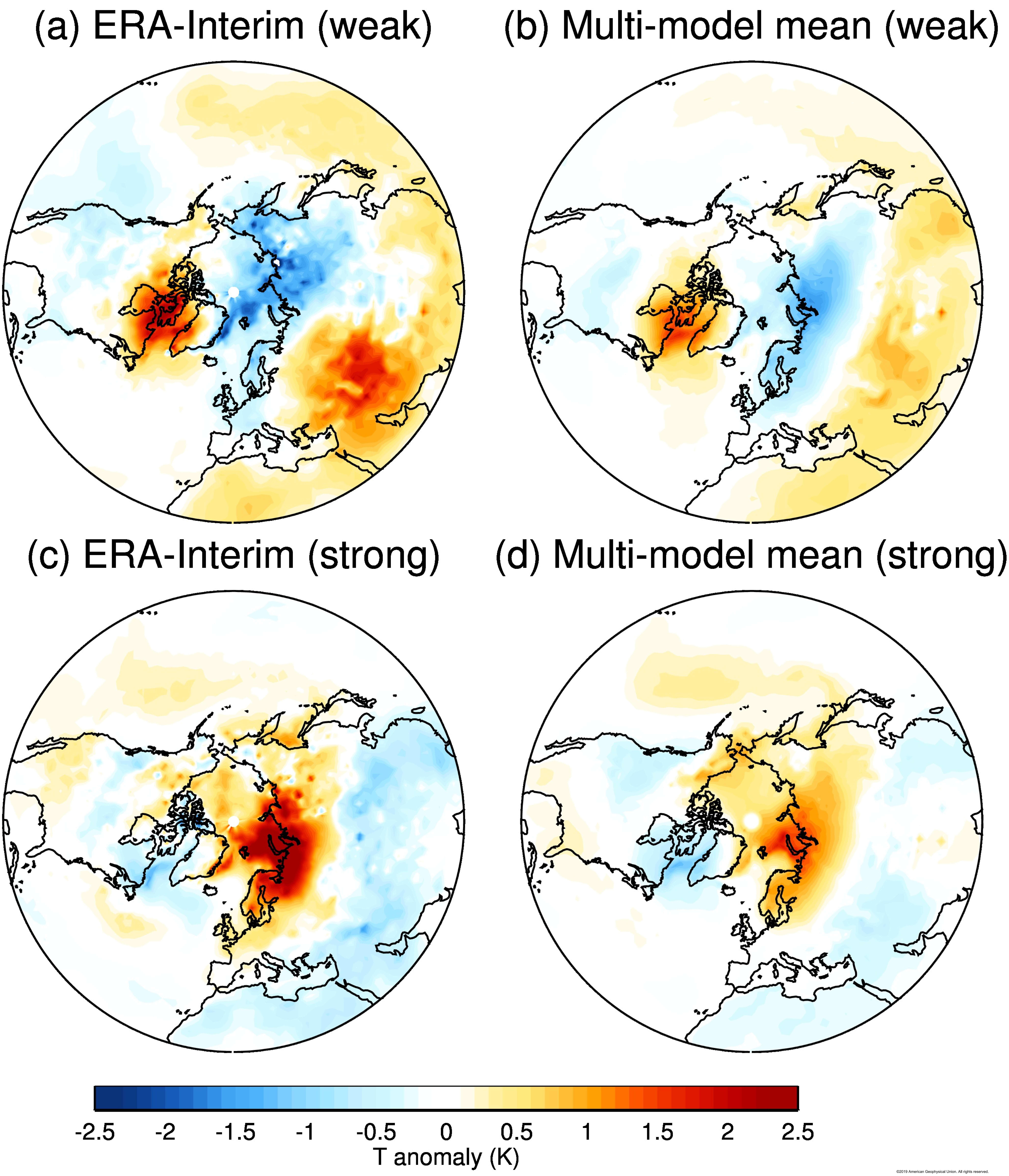

(a) ERA-Interim (weak)
(b) Multi-model mean (weak)
(c) ERA-Interim (strong)
(d) Multi-model mean (strong)
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
T anomaly (K)

Figure.

SLP Precursor Pattern to Major SSWs (Days -20 to -1)