

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

Article

Accepted Version

Domeisen, D. I. V., Butler, A. H., Charlton-Perez, A. J. ORCID: <https://orcid.org/0000-0001-8179-6220>, Ayarzagüena, B., Baldwin, M. P., Dunn-Sigouin, E., Furtado, J. C., Garfinkel, C. I., Hitchcock, P., Karpechko, A. Y., Kim, H., Knight, J., Lang, A. L., Lim, E.-P., Marshall, A., Roff, G., Schwartz, C., Simpson, I. R., Son, S.-W. and Taguchi, M. (2020) The role of the stratosphere in subseasonal to seasonal prediction part II: predictability arising from stratosphere - troposphere coupling. *Journal of Geophysical Research: Atmospheres*, 125 (2). e2019JD030923. ISSN 2169-8996 doi: <https://doi.org/10.1029/2019jd030923> Available at <https://centaur.reading.ac.uk/87609/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

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

Publisher: American Geophysical Union

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

The role of the stratosphere in subseasonal to seasonal prediction

Part II: Predictability arising from stratosphere - troposphere coupling

Daniela I.V. Domeisen¹, Amy H. Butler^{2,3}, Andrew J. Charlton-Perez⁴, Blanca Ayarzagüena^{5,6}, Mark P. Baldwin⁷, Etienne Dunn-Sigouin⁸, Jason C. Furtado⁹, Chaim I. Garfinkel¹⁰, Peter Hitchcock¹¹, Alexey Yu. Karpechko¹², Hera Kim¹³, Jeff Knight¹⁴, Andrea L. Lang¹⁵, Eun-Pa Lim¹⁶, Andrew Marshall¹⁶, Greg Roff¹⁶, Chen Schwartz¹⁰, Isla R. Simpson¹⁷, Seok-Woo Son¹³, Masakazu Taguchi¹⁸

¹Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland

²Cooperative Institute for Research in Environmental Sciences, Boulder, CO, USA

³National Oceanic and Atmospheric Administration, Chemical Sciences Division, USA

⁴University of Reading, Reading, UK

⁵Universidad Complutense de Madrid, Madrid, Spain

⁶Instituto Geociencias, CSIC-UCM, Spain

⁷University of Exeter, Exeter, UK

⁸Geophysical Institute, U. Bergen and Bjerknes Centre, Bergen, Norway

⁹School of Meteorology, University of Oklahoma, USA

¹⁰Fredy and Nadine Herrmann Institute of Earth Sciences, Hebrew University of Jerusalem, Israel

¹¹Cornell University, Ithaca, NY, USA

¹²Finnish Meteorological Institute, Finland

¹³Seoul National University, South Korea

¹⁴MetOffice Hadley Centre, Exeter, Devon, UK

¹⁵University at Albany, State University of New York, USA

¹⁶Bureau of Meteorology, Australia

¹⁷Climate and Global Dynamics Laboratory, NCAR, USA

¹⁸Aichi 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, Universitätsstrasse 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:10.1029/2018JD027000

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 mech-

88 anisms see Kidston et al. (2015); Tripathi, Baldwin, et al. (2015). In particular, the North
 89 Atlantic and Eurasia are strongly impacted by stratospheric extremes, with surface tem-
 90 perature anomalies on the order of several °C for days to weeks after a stratospheric event
 91 (Butler et al., 2018; Butler, Sjoberg, Seidel, & Rosenlof, 2017). Due to this downward
 92 coupling from the stratosphere it has been suggested that the stratosphere may be able
 93 to increase the predictability of surface weather (Butler et al., 2016; Scaife et al., 2016;
 94 Sigmond, Scinocca, Kharin, & Shepherd, 2013). Several single-model studies found an
 95 increase in prediction skill for forecasts that were initialized during sudden stratospheric
 96 warming (SSW) events or with an improved stratospheric representation for various tropo-
 97 spheric fields such as the Northern Annular Mode (NAM, e.g., D. W. J. Thompson
 98 & Wallace, 2000), with a focus on the North Atlantic sector and hence the NAO, as well
 99 as surface temperatures (Kuroda, 2008; Marshall & Scaife, 2010; Sigmond et al., 2013).
 100 For example, the major SSW event in February 2018 has been suggested to have led to
 101 persistent cold weather over large parts of Europe in late February and early March af-
 102 ter an otherwise mild winter (Karpechko, Perez, Balmaseda, Tyrrell, & Vitart, 2018),
 103 as well as anomalously wet conditions over southwestern Europe (Ayarzagüena et al.,
 104 2018). Like the 2018 event, up to two thirds of SSW events are followed by anomalous
 105 tropospheric weather patterns that can remain persistent for several weeks (Charlton-
 106 Perez, Ferranti, & Lee, 2018; D. I. V. Domeisen, 2019; Karpechko, Hitchcock, Peters, &
 107 Schneidereit, 2017; Simpson, Hitchcock, Shepherd, & Scinocca, 2011; I. White et al., 2018).
 108 The prospects of using the stratosphere for enhanced predictability at the surface on sub-
 109 seasonal to seasonal timescales is not limited to SSW events, as impacts on surface weather
 110 are also expected for other types of polar stratospheric extreme events such as strong
 111 vortex events (Tripathi, Charlton-Perez, Sigmond, & Vitart, 2015) and final warming
 112 events (Butler, Perez, Domeisen, Simpson, & Sjoberg, 2019; Hardiman et al., 2011).

113 While skillful deterministic forecasts of the above described extreme stratospheric
 114 events are limited to lead times of no more than 10 to 15 days (see Part I), the proba-
 115 bility of occurrence of these events during a given winter can be modified through re-
 116 mote impacts that affect polar vortex strength. A range of studies argue for precursors
 117 to SSW events in the extratropical troposphere (Davies, 1981; Kolstad & Charlton-Perez,
 118 2010; Schneidereit et al., 2017) such as atmospheric blocking (Ayarzagüena, Langematz,
 119 & Serrano, 2011; Martius, Polvani, & Davies, 2009; Nishii, Nakamura, & Orsolini, 2011;
 120 Quiroz, 1986; Woollings, Charlton-Perez, Ineson, Marshall, & Masato, 2010), Arctic sea
 121 ice (Kim et al., 2014; Sun, Deser, & Tomas, 2015; P. Zhang et al., 2018), Eurasian snow
 122 cover (Cohen & Entekhabi, 1999), and precursors in the extratropical lower stratosphere
 123 (Albers & Birner, 2014; de la Camara et al., 2017; D. I. V. Domeisen, Martius, & Jiménez-
 124 Esteve, 2018; Polvani & Waugh, 2004; Stockdale et al., 2015). The strength of the po-
 125 lar vortex can further be modified through remote impacts from the tropics, i.e. by El
 126 Niño Southern Oscillation (ENSO) (Butler et al., 2016; Butler & Polvani, 2011; Butler,
 127 Polvani, & Deser, 2014; D. I. V. Domeisen et al., 2015; C. I. Garfinkel & Hartmann, 2007;
 128 Ineson & Scaife, 2009; Manzini, Giorgetta, Esch, Kornblueh, & Roeckner, 2006; Polvani,
 129 Sun, Butler, Richter, & Deser, 2017; K. Song & Son, 2018), for a summary see D. I. V. Domeisen,
 130 Garfinkel, and Butler (2019), tropical convection related to the Madden-Julian Oscilla-
 131 tion (C. I. Garfinkel, Benedict, & Maloney, 2014; C. I. Garfinkel, Feldstein, Waugh, Yoo,
 132 & Lee, 2012; Kang & Tziperman, 2017), and the Quasi-Biennial Oscillation (QBO) through
 133 the Holton-Tan effect (Holton & Tan, 1980): Easterly winds in the tropical lower strato-
 134 sphere associated with an easterly QBO (eQBO) have been suggested to lead to a weak-
 135 ened stratospheric vortex through modifications in wave propagation and breaking in the
 136 surf zone (Andrews, Martin B et al., 2019; C. I. Garfinkel et al., 2018; C. I. Garfinkel,
 137 Shaw, Hartmann, & Waugh, 2012; O'Reilly, Weisheimer, Woollings, Gray, & MacLeod,
 138 2018; Richter, Deser, & Sun, 2015; Scaife, Athanassiadou, et al., 2014). These tropical
 139 modes of variability can also have a direct effect on the extratropical troposphere with-
 140 out a stratospheric pathway (B. J. Hoskins & Ambrizzi, 1993; Li, Li, Jin, & Zhao, 2015;
 141 Scaife et al., 2017), while for ENSO it has been shown that the stratospheric influence,

142 if present, tends to dominate over the tropospheric pathway (Butler et al., 2014; Jiménez-
143 Esteve & Domeisen, 2018).

144 We use subseasonal model hindcasts from operational prediction systems to evalu-
145 ate the role of stratosphere - troposphere coupling in the NH with respect to the in-
146 fluence of precursors to stratospheric events (Section 3) and potential changes in pre-
147 dictability of surface weather given stratospheric variability (Section 4). Section 2 gives
148 a brief introduction to the database and the methodology (for more details see Part I).
149 Section 5 provides a discussion of the results.

150 2 Methodology

151 2.1 Data

152 We use hindcast data from the S2S forecast project containing 11 different oper-
153 ational subseasonal forecast systems (Vitart et al., 2017). Table 1 (repeated from Part
154 I) provides an overview over the models used in this study (further details about the mod-
155 els 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 [×]	ERA-interim	1997-2016	11
JMA [×]	JRA-55	1981-2010	5
CNRM-Meteo [×]	ERA-interim	1993-2014	15
CNR-ISAC	ERA-interim	1981-2010	1
NCEP [×]	CFSR	1999-2010	4
UKMO [×]	ERA-interim	1993-2015	3

156
157 Due to the large differences in ensemble size, time period, and model specifics, the
158 exact datasets or selection of models may vary depending on the analysis or application
159 in this study, depending on the specific requirements of different parts of the analysis
160 in terms of e.g. lead times or available time periods. Different numbers of ensemble mem-
161 bers for BoM were used in this analysis, depending on the number of members available
162 at the time of data acquisition.

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

169

2.2 Skill Measures

170

171

172

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

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

173

174

175

176

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).

177

178

179

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.,

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

180

181

182

183

184

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.

185

186

187

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:

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

188

189

3 Precursors and Remote Influences on the Northern Hemisphere Stratosphere

190

191

192

193

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.

194

3.1 Precursors in the Extratropical Northern Hemisphere Troposphere

195

196

197

198

199

200

201

202

203

204

205

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

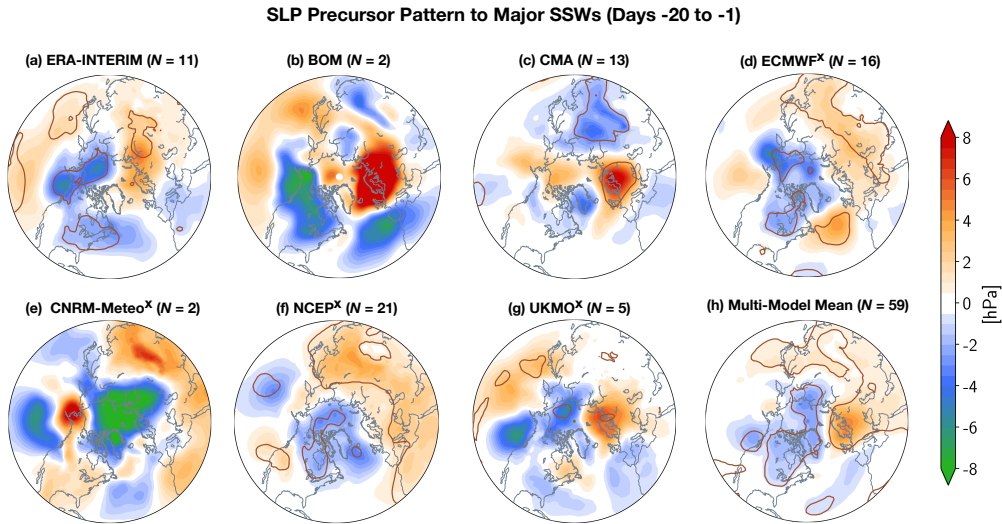


Figure 1. (a) NH SLP anomalies [hPa] averaged over days 1 to 20 before mid-winter SSW events for 1996-2010 in ERA-interim. (b)-(g) As in (a), but for the ensemble mean SLP anomaly composite for simulated mid-winter SSW events in six of the S2S prediction systems considered here (see text for details). 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. 'x' indicates high-top models.

206 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.

210 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

229 Pacific High, corresponding to the positive phase of the North Pacific Oscillation (e.g.
230 Rogers, 1981); and (3) anomalously low SLP across central and northeastern North Amer-
231 ica. The dominant features in both the North Pacific and over Eurasia have been doc-
232 umented in the literature both in models and different reanalysis products (e.g. D. I. V. Domeisen,
233 2019; Furtado, Cohen, Butler, Riddle, & Kumar, 2015; C. I. Garfinkel et al., 2010; Karpechko
234 et al., 2018; Kolstad & Charlton-Perez, 2010; Peings, 2019; I. White et al., 2018) and they
235 can manifest as an amplification of the climatological planetary-scale wave pattern through
236 wave interference (K. L. Smith & Kushner, 2012). An amplification of the climatolog-
237 ical wave structure, especially over the Pacific sector, thus provides increased wave forc-
238 ing and easterly momentum to the westerly flow in the stratosphere, increasing the chances
239 of a SSW event.

240 The SLP anomaly precursors in the individual prediction systems show substan-
241 tial differences as compared to reanalysis (Figs. 1b-h). The SLP precursor to mid-winter
242 SSW events in the multi-model mean (Fig. 1h) consists of negative anomalies in the Gulf
243 of Alaska and central North America and positive anomalies over the Europe. Ridging
244 over central Asia is less well captured. Examining the prediction systems individually,
245 all of them (except for CNRM-Meteo, Fig. 1e) feature positive SLP anomalies across Scan-
246 dinavia / northern Europe and extending into Asia, though significance of this feature
247 differs between the prediction systems. The North Pacific SLP anomalies are less well
248 replicated in the individual systems, with the UKMO model showing the closest simi-
249 larity to reanalysis (though statistically insignificant). The North American negative SLP
250 anomalies seen in the reanalysis plot are also less common in individual models, though
251 ECMWF (Fig. 1d) and NCEP (Fig. 1f) appear to reproduce a similar feature. Note that
252 these two models are also the ones with the two largest sample sizes for their compos-
253 ites (16 and 21, respectively), thus strongly influencing the multi-model mean compos-
254 ite (Fig. 1h).

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

Accepted Article

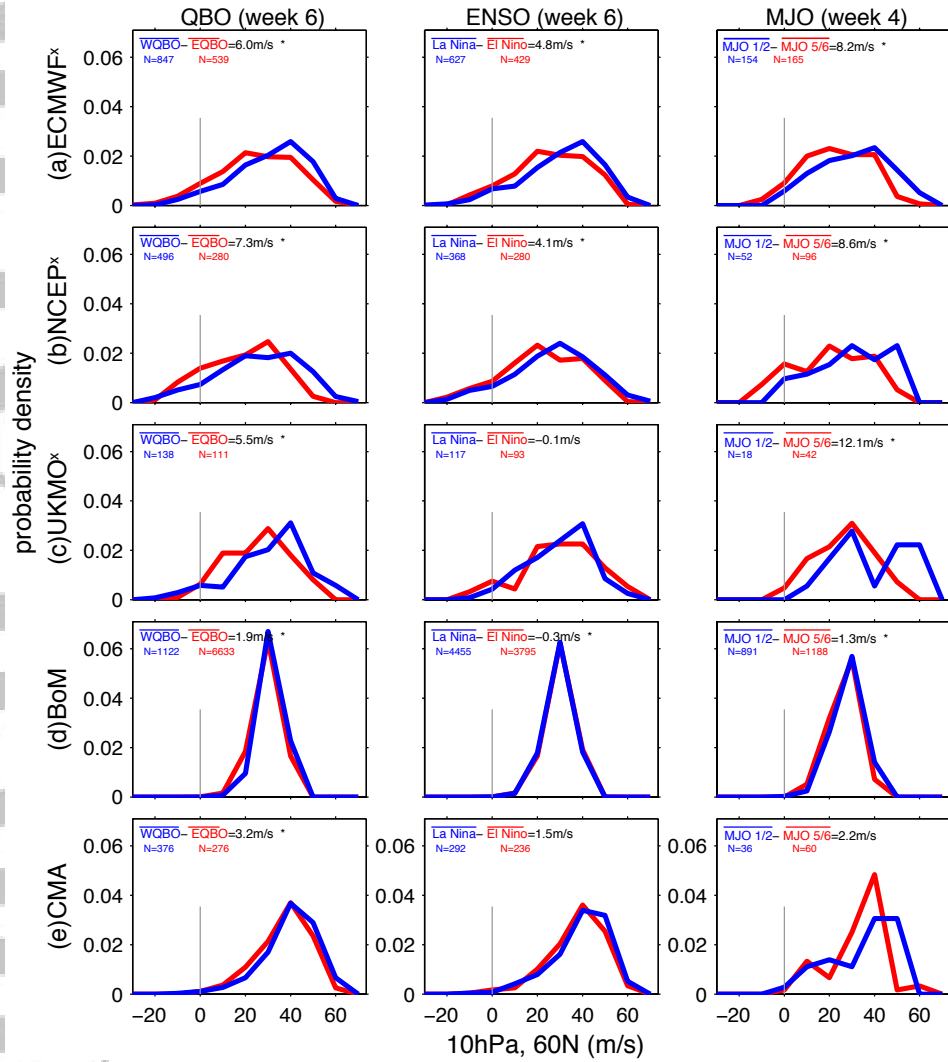


Figure 2. Probability density of zonal mean zonal wind at 10 hPa, 60°N for hindcasts initialized in November and December. Red (blue) lines indicate hindcasts initialized during (left) eQBO (wQBO), (center) El Niño (La Niña) conditions, and (right) MJO phases 5/6 (1/2). All histograms are normalized 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 \times . 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 $\pm 3 \text{ ms}^{-1}$. 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 $\pm 0.5^\circ\text{C}$ 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,

331 even for the high-top models (Fig. S2). C. I. Garfinkel et al. (2018) show that the model
 332 spread encompasses the observed response for the QBO, so there is no evidence that mod-
 333 els are systematically biased, even if the ensemble mean response is too weak. For ENSO,
 334 the observed effect is opposite to that in models (and also opposite to the observed re-
 335 sponse in the period before the S2S hindcasts); the mismatch between observations and
 336 the S2S models for ENSO is analyzed in detail in C. Garfinkel et al. (2019).

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

344 4 Predicting the Downward Coupling to the Troposphere

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

348 4.1 Arctic surface anomalies

349 The strength of the stratospheric polar vortex and its associated potential vortic-
 350 ity anomalies are linked to polar cap surface pressure anomalies through a vertical move-
 351 ment of the polar tropopause (Ambaum & Hoskins, 2002). Thus, polar sea level pres-
 352 sure is a suitable variable for studying tropospheric predictability arising from the strato-
 353 sphere. Moreover, these surface pressure anomalies are relevant for near-surface weather
 354 and even for Arctic sea ice distribution and motion (Kwok, 2000; K. Smith, Polvani, &
 355 Tremblay, 2018). In addition, polar pressure anomalies also have implications at mid-
 356 latitudes, because they can project onto the tropospheric NAO pattern. This surface im-
 357 pact can lead to lagged changes in the near-surface temperature or upper tropospheric
 358 winds (Baldwin, 2001; D. Thompson & Wallace, 1998; D. Thompson, Wallace, & Hegerl,
 359 2000).

360 The stratospheric signal is here characterized by the averaged anomalies over the
 361 polar cap of pressure at fixed heights, defined by a metric of the stratospheric variabil-
 362 ity based on daily 100 hPa temperature averaged over 65°-90°N, denoted the *ST100 in-*
 363 *dex* (Baldwin, Birner, & Ayarzagüena, 2019). We regress the anomalous polar cap pres-
 364 sure for the atmospheric column on the standardized ST100 index in January-March for
 365 ERA-interim reanalysis (Fig. 3a, black line) for the period 1999-2010. The pressure anoma-
 366 lies exhibit two maxima, one in the lower stratosphere (around 16km) and the other close
 367 to the surface. The latter denotes a strengthened stratospheric signal at lower levels as
 368 compared to other tropospheric levels (Baldwin et al., 2019). The vertical structure in
 369 Figure 3a is not expected only from mass moving into and out of the polar cap in the
 370 stratosphere. For example, during a SSW, mass is moved into the polar cap in the strato-
 371 sphere, where the air descends and warms adiabatically. In the lower stratosphere (around
 372 16km) pressure increases by 2hPa. Above that level, the pressure increment (ΔP) has
 373 to decrease because the ambient pressure drops off below 4hPa. Below the stratospheric
 374 maximum, ΔP would be 2hPa if mass were prevented from flowing out of the polar re-
 375 gion below that level, as in a cylinder at 65°N with impermeable walls. Moreover, the
 376 flux of mass into the polar cap is almost zero in the lowermost stratosphere. Given that
 377 the impermeable walls do not exist, as the air descends from 16km in the lowermost strato-
 378 sphere, it is not confined to the polar cap, and it “leaks” out of the polar cap below the
 379 levels with injection of mass (see Ambaum and Hoskins (2002) for a discussion of the po-
 380 tential vorticity dynamics of this situation). This explains the existence of the first max-

381 inum of pressure anomalies, but not the second one at Earth's surface, where we would
382 expect a minimum instead. However, below the tropopause in Fig. 3a, the polar pres-
383 sure anomalies increase, with a second maximum at the surface. The only explanation
384 for this near-surface maximum of the stratospheric signal is the action of additional tro-
385 pospheric processes that amplify the signal close to the surface. In particular, changes
386 in low-level heat flux (Baldwin et al., 2019; Limpasuvan, V, Thompson, D, & Hartmann,
387 D L, 2004) and temperature advection (Baldwin et al., 2019; D. Thompson et al., 2000)
388 lead to temperature anomalies over the polar cap that induce pressure anomalies (B. Hoskins,
389 McIntyre, & Robertson, 1985). The surface pressure anomalies ultimately are respon-
390 sible for the mass movement into the polar cap that is synchronised with mass movement
391 in the stratosphere. The net effect is that the surface pressure signal, e.g. for the NAM,
392 is much larger than would be expected based solely on movement of mass within the strato-
393 sphere.

394 The lagged regression of the anomalous polar pressure at different levels on the stan-
395 dardized ST100 index reveals important aspects of the timing of the tropospheric feed-
396 backs involved in the surface pressure amplification (Fig. 3b). The stratospheric-induced
397 Arctic surface pressure anomalies (blue line; lagged regression of anomalous polar pres-
398 sure at 0 km onto the ST100 index) peak at a lag of around +3 days with respect to the
399 stratospheric anomaly. Thus, the stratosphere leads the surface signal. Moreover, the
400 anomalies persist up to 60 days, longer than the stratospheric signal itself (orange line;
401 lagged regression of anomalous polar pressure at 15 km onto the ST100 index). The tropospheric-
402 only part of the signal (green line; lagged regression of the difference in anomalous pol-
403 ar pressure at 8 km and 0 km onto the ST100 index) also lags the stratospheric signal.

404 A similar analysis is now performed with the S2S systems to judge their skill in rep-
405 resenting the impact of the stratospheric state on Arctic surface anomalies and partic-
406 ularly, characterize up to which lead times they show an effect of the stratosphere on pol-
407 ar surface weather. In this case regressions of pressure anomalies on the standardized
408 ST100 index were computed separately for all S2S systems. To build the ST100 index
409 and compute the instantaneous regression on polar pressure of Fig. 3a only the data for
410 24h time steps of all available hindcast initialization dates in JFM of the 1999-2010 pe-
411 riod are considered. The results indicate that the polar tropospheric amplification of the
412 stratospheric signal is present in all S2S prediction systems and maximizes near the sur-
413 face (Fig. 3a, colored lines). Regarding the lagged regressions of pressure anomalies on
414 the standardized ST100 index (Fig. 3c), the computation differs slightly between the S2S
415 systems and the reanalysis: For each S2S system, the anomalous polar pressure is cal-
416 culated for every 24h time step from 24h to 768h with respect to the initialization time
417 and regressed onto the ST100 index (computed for all 24h time steps). Finally, the re-
418 gression from each system is averaged over all ensemble members and then over all pre-
419 diction systems.

420 In the prediction systems, the surface amplification also peaks at a positive lag of
421 around +3 days (Fig. 3c), but it decays more slowly than in the reanalysis. This is con-
422 sistent with the quicker decay of the troposphere-only signal (0-8 km) in reanalysis as
423 compared to the S2S systems mean (i.e., the reanalysis lies below the S2S system mean
424 ± 1.5 standard deviations after 20 days, see green line in Fig. 3c). As expected, the spread
425 among prediction systems grows, in general, with forecast lead time. It is particularly
426 large for the surface response after a lag of 8 days (blue shading), but it does not grow
427 much further after that.

428 Several reasons might explain the models' deviations from reanalysis and the inter-
429 model spread, i.e. the relatively short study period (1999-2010) or model biases. To test
430 both possibilities we repeated the analysis considering all data available for each S2S sys-
431 tem separately as shown in Table 1 (Fig. S3). The short data record might be respon-
432 sible for the noisy result: when extending the period to 1980-2016 for ERA-interim, the
433 results become smoother (Fig. S3a). The same result is obtained when including the pre-

434 satellite period (not shown). Moreover, the inter-model spread is also reduced with re-
 435 spect to Fig. 3c, in particular for the surface pressure results. However, even if we con-
 436 sider a longer period of study, the prediction systems still show discrepancies among them-
 437 selves. For instance, whereas high-top model systems (JMA, UKMO, or ECMWF) de-
 438 pict a comparable magnitude of the stratospheric signal in the lowermost stratosphere
 439 and near the surface from lag +4 days, systems with lower stratospheric resolution (BoM
 440 and CMA) predict a stronger surface signal. In these latter cases, the tropospheric part
 441 of the signal (green line) is similar to that of other systems or reanalysis. Thus, the mis-
 442 represented processes in these models should relate to the stratospheric signal itself (as
 443 is the case with CMA, Fig. S3d) or the coupling between the stratosphere and the tro-
 444 posphere.

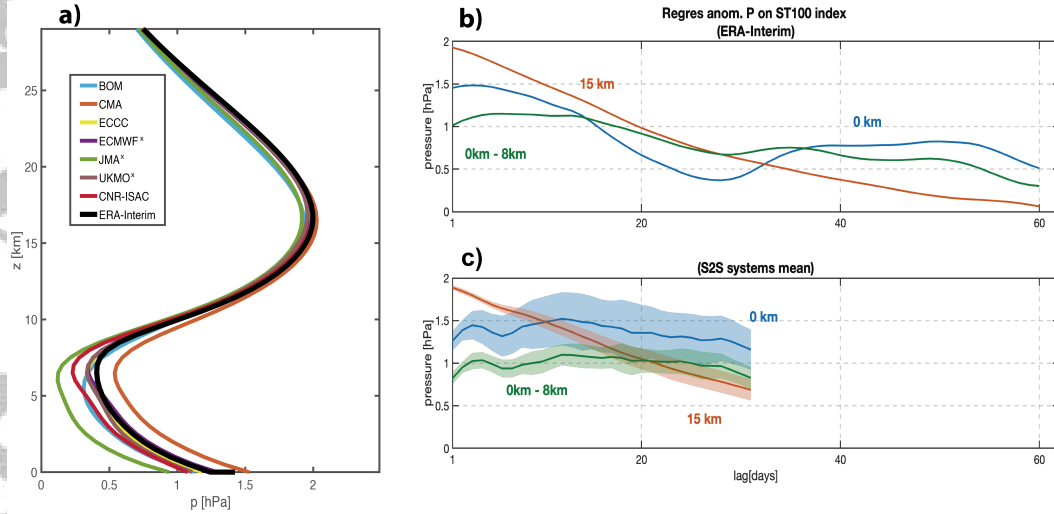


Figure 3. (a) Regression of Arctic (65°N - pole) pressure anomalies (hPa) as a function of height on the standardized ST100 index associated with one standard deviation of the ST100 index for ERA-interim (black line) and the hindcasts from the S2S prediction systems (colored lines) for the period 1999 - 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

445 Stratospheric events can have a significant surface impact in the extratropical North-
 446 ern Hemisphere. This is here quantified as the 2-meter temperature anomalies for weeks
 447 3-4 following weak and strong vortex events (Figure 4). Weak and strong vortex states
 448 are determined based on the strength of the zonal mean zonal wind at 60°N , 10 hPa in
 449 the reanalysis using the following criteria:
 450

$$\text{weak vortex} = \bar{u}_{60\text{N},10\text{hPa}} < 5\text{m.s}^{-1} \quad (4)$$

$$\text{strong vortex} = \bar{u}_{60\text{N},10\text{hPa}} > 40\text{m.s}^{-1} \quad (5)$$

452 where the overbar denotes the zonal mean. These thresholds were chosen to be close to
 453 the ones used in Tripathi, Charlton-Perez, et al. (2015), except that here the thresholds
 454 are relaxed in order to allow for sufficient event statistics due to the limited common pe-
 455 riod covered in the S2S prediction systems. A sensitivity test varying the thresholds by
 456 5ms^{-1} does not yield qualitative differences. The forecast anomalies are compared to
 457 those of a control population of forecasts determined separately for the weak and strong
 458 vortex cases. For example, for each weak vortex event, the control is taken from the same
 459 day of the year for all other years within the dataset provided it does not fall into the
 460 weak or strong category. For example, for BoM, which covers 1981 to 2013, the first ob-
 461 served weak vortex state by the criterion (4) occurred on 6th Feb 1981. Of the 6th Feb
 462 forecast initializations of the remaining years in the 1981-2013 period, 21 had a vortex
 463 state that was not characterized as weak or strong according to the criteria (4)-(5), so
 464 those 21 forecasts initialized on 6th Feb were added to the control population. This was
 465 repeated for each subsequent weak vortex state giving rise to the large control samples
 466 listed in Table 2. The control forecasts have roughly the same distribution in terms of
 467 seasonality as the weak forecasts. Note that for the BoM prediction system, only the first
 468 24 of the 33 members were used in this analysis (see Methods section). Otherwise, all
 469 forecasts within the December to March season are used and we consider the average over
 470 weeks 3-4 of the forecasts. It should be noted that for models that have frequent initial-
 471 izations there may be multiple forecasts that are initialized over the course of a partic-
 472 ular stratospheric event and so the individual forecasts are not entirely independent, but
 473 the same will be true for the accompanying control forecasts.

474 The surface anomalies following weak vortex events are strongest over Eurasia and
 475 northeastern Canada, with cold anomalies over Siberia, Scandinavia, and northern Green-
 476 land, and warm anomalies over Alaska, northeastern Canada, the Middle East, and north-
 477 ern Africa (Fig. 4a). The anomalies in the prediction systems appear smoother due to
 478 the larger sample size, but overall the anomaly patterns are well represented (Fig. 4b).
 479 The main differences exist in the magnitude of the anomalies: warm anomalies are gen-
 480 erally stronger in ERA-interim for both weak and strong vortex events. The cold anom-
 481 alies in strong vortex events are of the same order for the reanalysis and the multi-model
 482 mean (Fig. 4c,d), while the cold anomaly over Eurasia after weak vortex events extends
 483 further west over Eurasia in the multi-model mean compared to reanalysis.

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

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

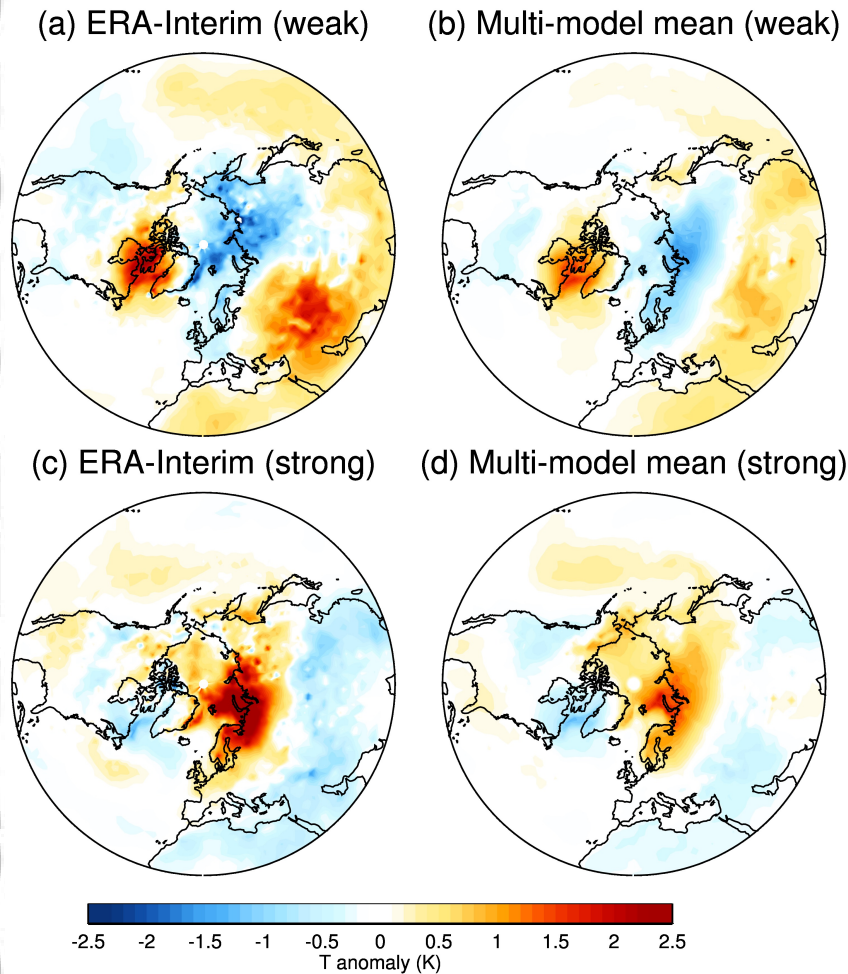


Figure 4. Composite 2m temperature anomalies (K) for weeks 3-4 for (top) weak vortex states and (bottom) strong vortex states. (b)/(d) show the multi-model mean for forecasts initialized during weak/strong vortex states. (a)/(c) shows the equivalent anomalies for ERA-interim where each date present in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

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

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

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

521 A notable region of reduced skill during weak vortex events arises over Europe (Fig.
522 5c). While we cannot directly relate the change in skill shown in Fig. 5c to the compar-
523 ison of the composites in Fig. 4, they are, at least, consistent in that the region of re-
524 duced skill over Europe during weak vortex events is a region where the model and re-
525 analysis WEAK composites differ (Figs. 4a,b). The forecast systems suggest that the zero
526 line of surface temperature anomalies roughly cuts through central Europe with cold anoma-
527 lies to the North and warm anomalies to the South (Fig. 4b), with some variability be-
528 tween individual models (Fig. S4). The ERA-interim composite, however, shows the zero
529 line further north with warm anomalies extending northward from the Middle-East into
530 eastern Europe/western Russia. As a result, the ERA-interim and S2S forecast anoma-
531 lies differ in sign in this region. Without more verification dates, it is difficult to deter-
532 mine whether this is just because the WEAK vortex composite in ERA-interim is im-
533 pacted by other unrelated variability, or whether the canonical temperature anomalies
534 that accompany weak vortex events in the real world are different to those in the model.
535 Indeed, only 3 out of the 8 models suggest this reduction in skill is significant (Fig. 6a).

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

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

548 As a final comparison with previous work and to summarize the skill associated with
549 weak and strong vortex events in the S2S models, the analysis is repeated for the NAM
550 index at 100 and 1000 hPa. The NAM index is calculated by projecting daily anoma-
551 lies from each ensemble member onto the NAM loading pattern computed as the first

552 empirical orthogonal function of ERA-interim zonal mean geopotential height between
553 20° - 90° N. An identical method to that used for 2m temperature for selecting forecasts
554 initialised during weak, neutral and strong vortex is used. The skill of forecasts from weak
555 and strong initializations is compared to a representative control forecast for each state
556 separately as above.

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

567 **5 Discussion and Outlook**

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

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

585 Once a stratospheric extreme event occurs, it can be long-lived in the lower strato-
586 sphere and have an impact on the troposphere. The S2S models successfully represent
587 the extra-tropical tropospheric response to stratospheric signals throughout the tropo-
588 spheric column, and the multi-model mean of the S2S systems successfully represents
589 the surface temperature anomaly response after weak and strong vortex events at 3-4
590 week lead times. Since the surface impact of stratospheric events is long-lived, the ex-
591 act timing of the stratospheric event, which is more difficult to forecast (see Part I), tends
592 to be less crucial for the duration of tropospheric effects, however it may be important
593 for the onset of anomalous weather. Although remote influences from the tropics also
594 affect tropospheric weather directly, many of these teleconnections have a pathway through
595 the stratosphere, and the stratosphere can therefore act as a modulation and as an ad-
596 ditional source for S2S prediction. Despite the significant surface impact of the strato-
597 sphere, enhanced predictability of 2m temperature anomalies linked to weak and strong
598 vortex events, and in particular for SSW events, is more difficult to show. For several
599 regions we cannot demonstrate enhanced predictability, at least in part because of the
600 limited record available for hindcast verification, as well as due to some of the models
601 not capturing the correct response locally. Overall, a strong reduction in forecast error
602 and an increase in skill at lead times of 3-4 weeks can be observed over Russia, the USA,

603 and the Middle East after weak vortex events, but not for Europe. For strong vortex events,
 604 the increase in predictability is overall less pronounced in these regions, but Europe tends
 605 to be better predicted than after weak vortex events. Initializations at the time of SSW
 606 events (instead of weak vortex events) show a much higher variability between predic-
 607 tion systems, likely due to the smaller number of available events, with some models show-
 608 ing a decrease in skill / increase in error. Predictions of the NAM index at the surface
 609 show a more consistent increase in skill for most models. This suggests that while 2m
 610 temperature tends to be difficult to forecast, the prediction of large-scale patterns has
 611 skill that could be used to forecast different fields for individual forecasting systems (e.g.
 612 Scaife, Arribas, et al., 2014). Further research will have to be conducted to investigate
 613 the model differences and to further validate the change in skill for different lead times.

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

622 These results should be used as a motivation to include a more complete represen-
 623 tation of the stratosphere in S2S model predictions and to include information on strato-
 624 spheric levels in databases used for sharing S2S predictions. An improved representa-
 625 tion of the stratosphere, including a better representation of critical physics, and an im-
 626 proved long-range prediction of the stratosphere itself (see Part I) may significantly ben-
 627 efit the prediction of surface weather. While the here presented model intercomparison
 628 and assessment is able to give a broad overview of the currently available skill related
 629 to the stratosphere, more detailed studies with respect to the documented phenomena
 630 and processes involved will have to be performed.

631 References

- 632 Albers, J., & Birner, T. (2014). Vortex preconditioning due to planetary and gravi-
 633 ty waves prior to sudden stratospheric warmings. *Journal of Atmospheric Sci-*
 634 *ences*, *71*, 4028–4054. doi: 10.1175/JAS-D-14-0026.1
- 635 Ambaum, M. H. P., & Hoskins, B. (2002). The nao troposphere-stratosphere connec-
 636 tion. *Journal of Climate*, *15*, 1969–1978.
- 637 Andrews, Martin B, Knight, Jeff R, Scaife, Adam A, Lu, Yixiong, Wu, Tongwen,
 638 Gray, Lesley J, & Schenzinger, Verena. (2019). Observed and simulated tele-
 639 connections between the stratospheric Quasi-Biennial Oscillation and Northern
 640 Hemisphere winter atmospheric circulation. *Journal of Geophysical Research-*
 641 *Atmospheres*, *124*(3), 1219–1232.
- 642 Ayarzagüena, B., Barriopedro, D., Perez, J. M. G., Abalos, M., de la Camara, A.,
 643 Herrera, R. G., ... Ordóñez, C. (2018). Stratospheric Connection to the
 644 Abrupt End of the 2016/2017 Iberian Drought. *Geophysical Research Letters*,
 645 *45*(22), 12,639–12,646.
- 646 Ayarzagüena, B., Langematz, U., & Serrano, E. (2011). Tropospheric forcing of
 647 the stratosphere: A comparative study of the two different major stratospheric
 648 warmings in 2009 and 2010. *Journal of Geophysical Research*, *116*(D18),
 649 D18114.
- 650 Baker, L. H., Shaffrey, L. C., Sutton, R. T., Weisheimer, A., & Scaife, A. A. (2018).
 651 An Intercomparison of Skill and Overconfidence/Underconfidence of the Win-
 652 tertime North Atlantic Oscillation in Multimodel Seasonal Forecasts. *Geophys-*
 653 *ical Research Letters*, *45*(15), 7808–7817.
- 654 Baldwin, M. P. (2001). Annular modes in global daily surface pressure. *Geophysical*

Research Letters.

- 655
656 Baldwin, M. P., Birner, T., & Ayarzagüena, B. (2019). Tropospheric amplification
657 of stratospheric variability. *EGU General Assembly Conference, Vienna, Aus-*
658 *tria.*
- 659 Baldwin, M. P., & Dunkerton, T. J. (1999). Propagation of the Arctic Oscillation
660 from the stratosphere to the troposphere. *Journal of Geophysical Research,*
661 *104(D24), 30937.*
- 662 Baldwin, M. P., & Dunkerton, T. J. (2001). Stratospheric harbingers of anomalous
663 weather regimes. *Science, 294(5542), 581–584.*
- 664 Baldwin, M. P., Stephenson, D. B., Thompson, D. W. J., Dunkerton, T. J., Charl-
665 ton, A. J., & O’Neill, A. (2003). Stratospheric memory and skill of extended-
666 range weather forecasts. *Science, 301, 636–640.*
- 667 Bao, Ming, Tan, Xin, Hartmann, Dennis L, & Ceppi, Paulo. (2017). Classifying the
668 tropospheric precursor patterns of sudden stratospheric warmings. *Geophysical*
669 *Research Letters, 44(15), 8011–8016.*
- 670 Beerli, R., Wernli, H., & Grams, C. M. (2017). Does the lower stratosphere provide
671 predictability for month-ahead wind electricity generation in Europe? *Quar-*
672 *terly Journal of the Royal Meteorological Society, 143(709), 3025–3036.*
- 673 Birner, T., & Albers, J. R. (2017). Sudden Stratospheric Warmings and Anomalous
674 Upward Wave Activity Flux. *SOLA, 13A(Special Edition), 8–12.*
- 675 Buizza, R., & Leutbecher, M. (2015). The forecast skill horizon. *Quarterly Journal*
676 *of the Royal Meteorological Society, 141(693), 3366–3382.*
- 677 Butler, A. H., Arribas, A., Athanassiadou, M., Baehr, J., Calvo, N., Charlton-Perez,
678 A., ... Yasuda, T. (2016). The Climate-system Historical Forecast Project:
679 Do stratosphere-resolving models make better seasonal climate predictions in
680 boreal winter? *Quarterly Journal of the Royal Meteorological Society, 142,*
681 *1413–1427.*
- 682 Butler, A. H., Charlton-Perez, A., Domeisen, D. I. V., Garfinkel, C., Gerber, E. P.,
683 Hitchcock, P., ... Son, S.-W. (2018). Sub-seasonal predictability and the
684 stratosphere. In A. W. Robertson & F. Vitart (Eds.), *Sub-seasonal to seasonal*
685 *prediction* (p. 585). Elsevier.
- 686 Butler, A. H., Perez, A. C., Domeisen, D. I. V., Simpson, I. R., & Sjoberg, J. (2019,
687 September). Predictability of Northern Hemisphere Final Stratospheric Warm-
688 ings and Their Surface Impacts. *Geophysical Research Letters, 43, 23.*
- 689 Butler, A. H., & Polvani, L. (2011). El Niño, La Niña, and stratospheric sudden
690 warmings: A reevaluation in light of the observational record. *Geophysical Re-*
691 *search Letters, 38.*
- 692 Butler, A. H., Polvani, L. M., & Deser, C. (2014). Separating the stratospheric and
693 tropospheric pathways of El Niño - Southern Oscillation teleconnections. *Envi-*
694 *ronmental Research Letters, 9(2).* doi: 10.1088/1748-9326/9/2/024014
- 695 Butler, A. H., Sjoberg, J. P., Seidel, D. J., & Rosenlof, K. H. (2017). A sudden
696 stratospheric warming compendium. *Earth Syst. Sci. Data, 9, 63–76.*
- 697 Charlton, A., & Polvani, L. M. (2007). A new look at stratospheric sudden warm-
698 ings. Part I: Climatology and modeling benchmarks. *Journal of Climate,*
699 *20(3), 449–469.*
- 700 Charlton-Perez, A. J., Ferranti, L., & Lee, R. W. (2018). The influence of the strato-
701 spheric state on North Atlantic weather regimes. *Quarterly Journal of the*
702 *Royal Meteorological Society, 144(713), 1140–1151.*
- 703 Cohen, J., & Entekhabi, D. (1999). Eurasian snow cover variability and northern
704 hemisphere climate predictability. *Geophysical Research Letters, 26(3), 345–*
705 *348.*
- 706 Cohen, J., & Jones, J. (2011). Tropospheric Precursors and Stratospheric Warmings.
707 *Journal of Climate, 24(24), 6562–6572.*
- 708 Davies, H. C. (1981). An Interpretation of Sudden Warmings in Terms of Potential
709 Vorticity. *Journal of the Atmospheric Sciences, 38(2), 427–445.*

- 710 Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., ...
 711 F, V. (2011). The ERA-interim reanalysis: Configuration and performance of
 712 the data assimilation system. *Quarterly Journal of the Royal Meteorological*
 713 *Society*, *137*(656), 553–597.
- 714 de la Camara, A., Albers, J. R., Birner, T., Garcia, R. R., Hitchcock, P., Kinnison,
 715 D. E., & Smith, A. K. (2017). Sensitivity of Sudden Stratospheric Warmings
 716 to Previous Stratospheric Conditions. *Journal of the Atmospheric Sciences*,
 717 *74*(9), 2857–2877.
- 718 Dobrynin, M., Domeisen, D. I. V., Müller, W. A., Bell, L., Brune, S., Bunzel, F.,
 719 ... Baehr, J. (2018). Improved Teleconnection-Based Dynamical Seasonal
 720 Predictions of Boreal Winter. *Geophysical Research Letters*, *44*(9-10), 2723.
- 721 Domeisen, D. I., Butler, A. H., Charlton-Perez, A. J., Ayarzagüena, B., Baldwin,
 722 M. P., Dunn-Sigouin, E., ... Taguchi, M. (2019). The role of the stratosphere
 723 in subseasonal to seasonal prediction. Part I: Predictability of the stratosphere.
 724 *Journal of Geophysical Research: Atmospheres*.
- 725 Domeisen, D. I. V. (2019). Estimating the Frequency of Sudden Stratospheric
 726 Warming Events from Surface Observations of the North Atlantic Oscillation.
 727 *Journal of Geophysical Research - Atmospheres*. doi: [http://doi.org/10.1029/
 728 2018JD030077](http://doi.org/10.1029/2018JD030077)
- 729 Domeisen, D. I. V., Badin, G., & Koszalka, I. M. (2018). How Predictable Are the
 730 Arctic and North Atlantic Oscillations? Exploring the Variability and Pre-
 731 dictability of the Northern Hemisphere. *Journal of Climate*, *31*(3), 997–1014.
 732 doi: <http://doi.org/10.1175/JCLI-D-17-0226.1>
- 733 Domeisen, D. I. V., Butler, A. H., Fröhlich, K., Bittner, M., Müller, W., & Baehr, J.
 734 (2015). Seasonal predictability over Europe arising from El Niño and strato-
 735 spheric variability in the MPI-ESM Seasonal Prediction System. *Journal of*
 736 *Climate*, *28*(1), 256–271. doi: <http://doi.org/10.1175/JCLI-D-14-00207.1>
- 737 Domeisen, D. I. V., Garfinkel, C. I., & Butler, A. H. (2019). The Teleconnection of
 738 El Niño Southern Oscillation to the Stratosphere. *Reviews of Geophysics*. doi:
 739 [10.1029/2018RG000596](http://doi.org/10.1029/2018RG000596)
- 740 Domeisen, D. I. V., Martius, O., & Jiménez-Esteve, B. (2018). Rossby Wave
 741 Propagation into the Northern Hemisphere Stratosphere: The Role of
 742 Zonal Phase Speed. *Geophysical Research Letters*, *45*(4), 2064–2071. doi:
 743 <http://doi.org/10.1002/2017GL076886>
- 744 Domeisen, D. I. V., Sun, L., & Chen, G. (2013). The role of synoptic eddies in
 745 the tropospheric response to stratospheric variability. *Geophysical Research*
 746 *Letters*, *40*, 1–5. doi: <http://doi.org/10.1002/grl.50943>
- 747 Douville, H. (2009). Stratospheric polar vortex influence on Northern Hemisphere
 748 winter climate variability. *Geophysical Research Letters*, *36*(18).
- 749 Dunn-Sigouin, E., & Shaw, T. (2018). Dynamics of Extreme Stratospheric Negative
 750 Heat Flux Events in an Idealized Model. *Journal of the Atmospheric Sciences*,
 751 *75*(10), 3521–3540.
- 752 Esler, J., & Matthewman, N. (2011). Stratospheric Sudden Warmings as Self-Tuning
 753 Resonances. Part II: Vortex Displacement Events. *Journal of the Atmospheric*
 754 *Sciences*, *68*, 2505–2523.
- 755 Furtado, J. C., Cohen, J. L., Butler, A. H., Riddle, E. E., & Kumar, A. (2015).
 756 Eurasian snow cover variability and links to winter climate in the CMIP5
 757 models. *Climate Dynamics*, *45*(9-10), 2591–2605.
- 758 García-Herrera, R., Calvo, N., Garcia, R. R., & Giorgetta, M. A. (2006). Propaga-
 759 tion of ENSO temperature signals into the middle atmosphere: A comparison
 760 of two general circulation models and ERA-40 reanalysis data. *Journal of*
 761 *Geophysical Research*, *111*(D6), D06101. doi:10.1029/2005JD006061.
- 762 Garfinkel, C., Schwartz, C., Domeisen, D., Butler, A. H., Son, S.-W., & White, I.
 763 (2019). Weakening of the teleconnection of El Niño-Southern Oscillation to
 764 the Arctic stratosphere over the past few decades: What can be learned from

- 765 subseasonal forecast models? *J. Geophys. Res. - Atmospheres*.
- 766 Garfinkel, C. I., Benedict, J. J., & Maloney, E. D. (2014). Impact of the MJO on
767 the boreal winter extratropical circulation. *Geophysical Research Letters*, *41*,
768 6055–6062.
- 769 Garfinkel, C. I., Feldstein, S. B., Waugh, D. W., Yoo, C., & Lee, S. (2012). Observed
770 connection between stratospheric sudden warmings and the Madden-Julian Os-
771 cillation. *Geophysical Research Letters*, *39*(18).
- 772 Garfinkel, C. I., & Hartmann, D. L. (2007). Effects of the El Niño - Southern
773 Oscillation and the Quasi-Biennial Oscillation on polar temperatures in the
774 stratosphere. *J. Geophys. Res.*, *115*, D20116.
- 775 Garfinkel, C. I., Hartmann, D. L., & Sassi, F. (2010). Tropospheric precursors
776 of anomalous northern hemisphere stratospheric polar vortices. *Journal of*
777 *Climate*, *23*(12), 3282–3299.
- 778 Garfinkel, C. I., & Schwartz, C. (2017). MJO-Related Tropical Convection Anoma-
779 lies Lead to More Accurate Stratospheric Vortex Variability in Subseasonal
780 Forecast Models. *Geophysical Research Letters*, *44*(19), 10,054 - 10,062. doi:
781 10.1002/2017GL074470
- 782 Garfinkel, C. I., Schwartz, C., Domeisen, D. I. V., Son, S.-W., Butler, A. H., &
783 White, I. P. (2018). Extratropical Atmospheric Predictability From the Quasi-
784 Biennial Oscillation in Subseasonal Forecast Models. *Journal of Geophysical*
785 *Research-Atmospheres*, *123*(15), 7855–7866.
- 786 Garfinkel, C. I., Shaw, T. A., Hartmann, D. L., & Waugh, D. W. (2012). Does the
787 Holton-Tan mechanism explain how the Quasi-Biennial Oscillation modulates
788 the arctic polar vortex? *Journal of the Atmospheric Sciences*, *69*, 1713–1733.
- 789 Garfinkel, C. I., Waugh, D. W., & Gerber, E. P. (2013). The effect of tropospheric
790 jet latitude on coupling between the stratospheric polar vortex and the tropo-
791 sphere. *Journal of Climate*, *26*(6). doi: 10.1175/JCLI-D-12-00301.1
- 792 Gerber, E. P., Baldwin, M. P., Akiyoshi, H., AUSTIN, J., Bekki, S., Braesicke, P.,
793 ... Dhomse, S. (2010). Stratosphere-troposphere coupling and annular mode
794 variability in chemistry-climate models. *Journal of Geophysical Research:*
795 *Atmospheres (1984–2012)*, *115*(D3).
- 796 Gerber, E. P., Butler, A. H., Calvo, N., Charlton-Perez, A., Giorgetta, M., Manzini,
797 E., ... Watanabe, S. (2012). Assessing and Understanding the Impact of
798 Stratospheric Dynamics and Variability on the Earth System. *Bulletin of the*
799 *American Meteorological Society*, *93*(6), 845–859.
- 800 Hardiman, S. C., Butchart, N., Charlton-Perez, A. J., Shaw, T. A., Akiyoshi, H.,
801 Baumgaertner, A., ... Shibata, K. (2011). Improved predictability of the tropo-
802 sphere using stratospheric final warmings. *Journal of Geophysical Research*,
803 *116*(D18), 6313.
- 804 Hitchcock, P., Shepherd, T. G., Taguchi, M., Yoden, S., & Noguchi, S. (2013).
805 Lower-stratospheric radiative damping and Polar-night Jet Oscillation events.
806 *Journal of the Atmospheric Sciences*, *70*, 1391–1408.
- 807 Hitchcock, P., & Simpson, I. R. (2014). The Downward Influence of Stratospheric
808 Sudden Warmings. *J. Atmos. Sci.*, *71*(10), 3856–3876.
- 809 Hitchcock, P., & Simpson, I. R. (2016). Quantifying Eddy Feedbacks and Forcings in
810 the Tropospheric Response to Stratospheric Sudden Warmings. *Journal of the*
811 *Atmospheric Sciences*, *73*(9), 3641–3657.
- 812 Holton, J. R., & Tan, H. C. (1980). The influence of the equatorial Quasi-Biennial
813 Oscillation on the global circulation at 50 mb. *Journal of the Atmospheric Sci-*
814 *ences*, *37*(10), 2200–2208.
- 815 Hoskins, B., McIntyre, M., & Robertson, A. (1985). On the use and significance of
816 isentropic potential vorticity maps. *Quarterly Journal of the Royal Meteorolog-*
817 *ical Society*, *111*, 877–946.
- 818 Hoskins, B. J., & Ambrizzi, T. (1993). Rossby-Wave Propagation on a Realistic Lon-
819 gitudinally Varying Flow. *Journal of the Atmospheric Sciences*, *50*(12), 1661–

- 1671.
- 820 Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore,
821 J. H., . . . Zhang, H. (2017). Extended Reconstructed Sea Surface Temper-
822 ature, Version 5 (ERSSTv5): Upgrades, Validations, and Intercomparisons.
823 *Journal of Climate*, *30*, 8179–8205.
- 824 Hurrell, J. W., Kushnir, Y., & Visbeck, M. (2001). The North Atlantic Oscillation.
825 *Science*, *291*(5504), 603–605.
- 826 Ineson, S., & Scaife, A. (2009). The role of the stratosphere in the European climate
827 response to El Niño. *Nature Geoscience*, *2*(1), 32–36. doi: 10.1038/ngeo381
- 828 Iza, M., Calvo, N., & Manzini, E. (2016). The Stratospheric Pathway of La Niña.
829 *Journal of Climate*, *29*(24), 8899–8914.
- 830 Jiménez-Esteve, B., & Domeisen, D. I. V. (2018). The Tropospheric Pathway of the
831 ENSO-North Atlantic Teleconnection. *Journal of Climate*, *31*, 4563–4584.
- 832 Jucker, M., & Reichler, T. (2018). Dynamical Precursors for Statistical Prediction
833 of Stratospheric Sudden Warming Events. *Geophysical Research Letters*, *96*(2),
834 1.
835
- 836 Kang, W., & Tziperman, E. (2017). More frequent sudden stratospheric warming
837 events due to enhanced MJO forcing expected in a warmer climate. *Journal of*
838 *Climate*, *30*(21), 8727–8743.
- 839 Karpechko, A. Y., Hitchcock, P., Peters, D. H. W., & Schneiderit, A. (2017). Pre-
840 dictability of downward propagation of major sudden stratospheric warmings.
841 *Quarterly Journal of the Royal Meteorological Society*, *104*, 30937.
- 842 Karpechko, A. Y., Perez, A. C., Balmaseda, M., Tyrrell, N., & Vitart, F. (2018).
843 Predicting Sudden Stratospheric Warming 2018 and its Climate Impacts with
844 a Multi-Model Ensemble. *Geophysical Research Letters*, 2018GL081091.
- 845 Kawatani, Y., Hamilton, K., Gray, L., Osprey, S., Watanabe, S., & Yamashita, Y.
846 (2019). The effects of a well-resolved stratosphere on the simulated boreal
847 winter circulation in a climate model. *Journal of Atmospheric Sciences*. doi:
848 doi:10.1175/JAS-D-18-0206.1,inpress
- 849 Kidston, J., Scaife, A. A., Hardiman, S. C., Mitchell, D. M., Butchart, N., Bald-
850 win, M. P., & Gray, L. J. (2015). Stratospheric influence on tropospheric jet
851 streams, storm tracks and surface weather. *Nature Geoscience*, *8*(6), 433–440.
- 852 Kim, B.-M., Son, S.-W., Min, S.-K., Jeong, J.-H., Kim, S.-J., Zhang, X., . . . Yoon,
853 J.-H. (2014). Weakening of the stratospheric polar vortex by Arctic sea-ice
854 loss. *Nature Communications*, *5*(1), 4646.
- 855 Kolstad, E. W., & Charlton-Perez, A. J. (2010). Observed and simulated precu-
856 rors of stratospheric polar vortex anomalies in the Northern Hemisphere. *Cli-*
857 *mate Dynamics*, *37*(7-8), 1443–1456.
- 858 Kuroda, Y. (2008). Role of the stratosphere on the predictability of medium-range
859 weather forecast: A case study of winter 2003-2004. *Geophysical Research Let-*
860 *ters*, *35*(19).
- 861 Kwok, R. (2000). Recent changes in Arctic Ocean sea ice motion associated with the
862 North Atlantic Oscillation. *Geophysical Research Letters*, *27*, 775–778.
- 863 L’Heureux, M. L., Tippett, M. K., Kumar, A., Butler, A. H., Ciasto, L. M., Ding,
864 Q., . . . Johnson, N. C. (2017). Strong relations between ENSO and the Arctic
865 Oscillation in the North American Multimodel Ensemble. *Geophysical Research*
866 *Letters*, *44*. doi: 10.1002/2017GL074854
- 867 Li, Y., Li, J., Jin, F. F., & Zhao, S. (2015, August). Interhemispheric Propagation
868 of Stationary Rossby Waves in a Horizontally Nonuniform Background Flow.
869 *Journal of the Atmospheric Sciences*, *72*(8), 3233–3256.
- 870 Lim, E.-P., Hendon, H. H., Boschhat, G., Hudson, D., Thompson, D. W. J., Dowdy,
871 A. J., & Arblaster, J. M. (2019). Australian hot and dry extremes induced
872 by weakenings of the stratospheric polar vortex. *Nature Geoscience*, *12*(11),
873 896–901.
- 874 Lim, Y., Son, S.-W., Marshall, A., Hendon, H. H., & Seo, K.-H. (2019). Influence

- 875 of the QBO on MJO prediction skill in the subseasonal-to-seasonal prediction
 876 models. *Climate Dynamics*. doi: <https://doi.org/10.1007/s00382-019-04719-y>
- 877 Limpasuvan, V, Thompson, D, & Hartmann, D L. (2004). The life cycle of the
 878 Northern Hemisphere sudden stratospheric warmings. *Journal of the Atmo-*
 879 *spheric Sciences*, *17*, 2584–2596.
- 880 Manzini, E., Giorgetta, M. A., Esch, M., Kornbluh, L., & Roeckner, E. (2006). The
 881 influence of sea surface temperatures on the northern winter stratosphere: En-
 882 semble simulations with the MAECHAM5 model. *Journal of Climate*, *19*(16),
 883 3863–3881.
- 884 Marshall, A., & Scaife, A. A. (2010). Improved predictability of stratospheric sudden
 885 warming events in an atmospheric general circulation model with enhanced
 886 stratospheric resolution. *Journal of Geophysical Research*, *115*(D16114).
- 887 Martius, O., Polvani, L. M., & Davies, H. (2009). Blocking precursors to strato-
 888 spheric sudden warming events. *Geophysical Research Letters*, *36*, L14806.
- 889 Matthewman, N. J., & Esler, J. G. (2011). Stratospheric Sudden Warmings as
 890 Self-Tuning Resonances. Part I: Vortex Splitting Events. *Journal of the Atmo-*
 891 *spheric Sciences*, *68*, 2481–2504.
- 892 Nishii, K., Nakamura, H., & Orsolini, Y. J. (2011). Geographical Dependence Ob-
 893 served in Blocking High Influence on the Stratospheric Variability through
 894 Enhancement and Suppression of Upward Planetary-Wave Propagation. *Jour-*
 895 *nal of Climate*, *24*(24), 6408–6423.
- 896 O’Reilly, C. H., Weisheimer, A., Woollings, T., Gray, L., & MacLeod, D. (2018).
 897 The importance of stratospheric initial conditions for winter North Atlantic
 898 Oscillation predictability and implications for the signal-to-noise paradox.
 899 *Quarterly Journal of the Royal Meteorological Society*.
- 900 Peings, Y. (2019). Ural Blocking as a driver of early winter stratospheric warmings.
 901 *Geophysical Research Letters*, 2019GL082097–18.
- 902 Plumb, R. (1981). Instability of the distorted polar night vortex: A theory of strato-
 903 spheric warmings. *Journal of the Atmospheric Sciences*, *38*(11), 2514–2531.
- 904 Plumb, R. A., & Semeniuk, K. (2003). Downward migration of extratropical zonal
 905 wind anomalies. *Journal of Geophysical Research: Atmospheres (1984–2012)*,
 906 *108*(D7).
- 907 Polvani, L. M., Sun, L., Butler, A. H., Richter, J. H., & Deser, C. (2017). Dis-
 908 tinguishing Stratospheric Sudden Warmings from ENSO as Key Drivers of
 909 Wintertime Climate Variability over the North Atlantic and Eurasia. *Journal*
 910 *of Climate*, *30*(6), 1959–1969. doi: 10.1175/JCLI-D-16-0277.1
- 911 Polvani, L. M., & Waugh, D. (2004). Upward wave activity flux as a precursor
 912 to extreme stratospheric events and subsequent anomalous surface weather
 913 regimes. *Journal of Climate*, *17*, 3548–3554.
- 914 Quiroz, R. S. (1986). The association of stratospheric warmings with tropospheric
 915 blocking. *Journal of Geophysical Research-Atmospheres*.
- 916 Richter, J. H., Deser, C., & Sun, L. (2015). Effects of stratospheric variability on El
 917 Niño teleconnections. *Environmental Research Letters*, *10*(12), 124021.
- 918 Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., ...
 919 Woollen, J. (2011). MERRA: NASA’s Modern-Era Retrospective Analysis for
 920 Research and Applications. *Journal of Climate*, *24*(14), 3624–3648.
- 921 Rogers, J. C. (1981). The North Pacific Oscillation. *Journal of Climatology*, *1*(1),
 922 39–57.
- 923 Scaife, A. A., Arribas, A., Blockley, E., Scaife, A. A., Arribas, A., Blockley, E., ...
 924 Williams, A. (2014). Skillful long-range prediction of European and North
 925 American winters. *Geophysical Research Letters*, *41*, 2514–2519.
- 926 Scaife, A. A., Athanassiadou, M., Andrews, M., Arribas, A., Baldwin, M. P., Dun-
 927 stone, N., ... Williams, A. (2014). Predictability of the Quasi-Biennial Oscilla-
 928 tion and its Northern winter teleconnection on seasonal to decadal timescales.
 929 *Geophysical Research Letters*, *41*(5), 1752–1758.

- 930 Scaife, A. A., Comer, R. E., Dunstone, N. J., Knight, J. R., Smith, D. M., MacLach-
 931 lan, C., ... Slingo, J. (2017). Tropical rainfall, Rossby waves and regional
 932 winter climate predictions. *Quarterly Journal of the Royal Meteorological*
 933 *Society*, *143*(702), 1–11.
- 934 Scaife, A. A., Karpechko, A. Y., Baldwin, M. P., Brookshaw, A., Butler, A. H.,
 935 Eade, R., ... Smith, D. (2016). Seasonal winter forecasts and the stratosphere.
 936 *Atmospheric Science Letters*, *17*(1), 51–56.
- 937 Schneidereit, A., Peters, D. H. W., Keller, J. H., Teubler, F., Martius, O., Grams,
 938 C. M., ... Riemer, M. (2017). Enhanced Tropospheric Wave Forcing of Two
 939 Anticyclones in the Prephase of the January 2009 Major Stratospheric Sudden
 940 Warming Event. *Monthly Weather Review*, *145*(5), 1797–1815.
- 941 Schwartz, C., & Garfinkel, C. I. (2017). Relative roles of the MJO and strato-
 942 spheric variability in North Atlantic and European winter climate. *Journal of*
 943 *Geophysical Research: Atmospheres*, *122*(8), 4184–4201.
- 944 Sigmond, M., Scinocca, J. F., Kharin, V. V., & Shepherd, T. G. (2013). Enhanced
 945 seasonal forecast skill following stratospheric sudden warmings. *Nature Geo-*
 946 *science*, *6*(1). doi: 10.1038/ngeo1698
- 947 Simpson, I. R., Blackburn, M., & Haigh, J. D. (2009). The Role of Eddies in Driving
 948 the Tropospheric Response to Stratospheric Heating Perturbations. *Journal of*
 949 *the Atmospheric Sciences*, *66*(5), 1347–1365.
- 950 Simpson, I. R., Blackburn, M., & Haigh, J. D. (2012). A Mechanism for the Effect
 951 of Tropospheric Jet Structure on the Annular Mode-Like Response to Strato-
 952 spheric Forcing. *Journal of the Atmospheric Sciences*, *69*(7), 2152–2170.
- 953 Simpson, I. R., Hitchcock, P., Shepherd, T. G., & Scinocca, J. F. (2011). Strato-
 954 spheric variability and tropospheric annular mode timescales. *Geophys. Res.*
 955 *Let.*, *38*, L20806.
- 956 Smith, K., Polvani, L., & Tremblay, L. (2018). The impact of stratospheric cir-
 957 culation extremes on minimum Arctic sea ice extent. *Journal of Climate*, *31*,
 958 7169–7183.
- 959 Smith, K. L., & Kushner, P. J. (2012). Linear interference and the initiation of ex-
 960 tratropical stratosphere-troposphere interactions. *Journal of Geophysical Re-*
 961 *search*, *117*(D13). doi: 10.1029/2012JD017587
- 962 Smith, K. L., & Scott, R. K. (2016). The role of planetary waves in the tropospheric
 963 jet response to stratospheric cooling. *Geophysical Research Letters*, *43*(6),
 964 2904–2911.
- 965 Song, K., & Son, S.-W. (2018). Revisiting the ENSO–SSW Relationship. *Journal of*
 966 *Climate*, *31*(6), 2133–2143.
- 967 Song, Y., & Robinson, W. (2004). Dynamical mechanisms for stratospheric in-
 968 fluences on the troposphere. *Journal of the Atmospheric Sciences*, *61*, 1711–
 969 1725.
- 970 Stockdale, T. N., Molteni, F., & Ferranti, L. (2015). Atmospheric initial conditions
 971 and the predictability of the Arctic Oscillation. *Geophysical Research Letters*,
 972 *42*(4), 1173–1179.
- 973 Sun, L., Deser, C., & Tomas, R. A. (2015). Mechanisms of Stratospheric and Tropo-
 974 spheric Circulation Response to Projected Arctic Sea Ice Loss. *Journal of Cli-*
 975 *mate*, *28*, 7824–7845.
- 976 Thompson, D., & Wallace, J. (1998). Observed linkages between Eurasian surface
 977 air temperature, the North Atlantic Oscillation, Arctic sea level pressure and
 978 the stratospheric polar vortex. *Geophysical Research Letters*, *25*, 1297–1300.
- 979 Thompson, D., Wallace, J. M., & Hegerl, G. C. (2000). Annular modes in the extra-
 980 tropical circulation. Part II: Trends. *Journal of Climate*, *13*(5), 1018–1036.
- 981 Thompson, D. W. J., & Wallace, J. (2000). Annular modes in the extratropical cir-
 982 culation. Part I: Month-to-Month Variability. *Journal of Climate*, *13*(5), 1000–
 983 1016.
- 984 Tripathi, O. P., Baldwin, M. P., Charlton-Perez, A., Charron, M., Eckermann, S. D.,

- 985 Gerber, E., ... Son, S.-W. (2015). The predictability of the extratropical
 986 stratosphere on monthly time-scales and its impact on the skill of tropospheric
 987 forecasts. *Quarterly Journal of the Royal Meteorological Society*, 141(689),
 988 987–1003.
- 989 Tripathi, O. P., Charlton-Perez, A., Sigmond, M., & Vitart, F. (2015). Enhanced
 990 long-range forecast skill in boreal winter following stratospheric strong vortex
 991 conditions. *Environmental Research Letters*, 10(10), 1–8.
- 992 Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., ...
 993 Zhang, L. (2017). The Subseasonal to Seasonal (S2S) Prediction Project
 994 Database. *Bulletin of the American Meteorological Society*, 98(1), 163–173.
- 995 Walker, G. (1928). World weather. *Quarterly Journal of the Royal Meteorological
 996 Society*, 54(226), 79–87.
- 997 Wheeler, M. C., & Hendon, H. H. (2004). An all-season real-time multivariate
 998 mjo index: Development of an index for monitoring and prediction. *Monthly
 999 Weather Review*, 132(8), 1917–1932. doi: 10.1175/1520-0493(2004)132<1917:
 1000 AARMMI>2.0.CO;2
- 1001 White, C. J., Carlsen, H., Robertson, A. W., Klein, R. J. T., Lazo, J. K., Kumar,
 1002 A., ... Zebiak, S. E. (2017). Potential applications of subseasonal-to-seasonal
 1003 (S2S) predictions. *Meteorological Applications*, 24(3), 315–325.
- 1004 White, I., Garfinkel, C. I., Gerber, E. P., Jucker, M., Aquila, V., & Oman, L. D.
 1005 (2018). The Downward Influence of Sudden Stratospheric Warmings: Associa-
 1006 tion with Tropospheric Precursors. *Journal of Climate*, 32(1), 85–108.
- 1007 Woollings, T., Charlton-Perez, A., Ineson, S., Marshall, A. G., & Masato, G. (2010).
 1008 Associations between stratospheric variability and tropospheric blocking. *Jour-
 1009 nal of Geophysical Research: Atmospheres (1984–2012)*, 115(D6).
- 1010 Zhang, F., Sun, Y. Q., Magnusson, L., Buizza, R., Lin, S.-J., Chen, J.-H., &
 1011 Emanuel, K. (2019, April). What Is the Predictability Limit of Midlatitude
 1012 Weather? *Journal of the Atmospheric Sciences*, 76(4), 1077–1091.
- 1013 Zhang, P., Wu, Y., Simpson, I. R., Smith, K. L., Zhang, X., De, B., & Callaghan, P.
 1014 (2018). A stratospheric pathway linking a colder Siberia to Barents-Kara Sea
 1015 sea ice loss. *Science advances*, 4(7), eaat6025.
- 1016 Zhang, Q., Shin, C.-S., Dool, H., & Cai, M. (2013). CFSv2 prediction skill of strato-
 1017 spheric temperature anomalies. *Climate Dynamics*, 41(7-8), 2231–2249.

1018 Acknowledgments

1019 The S2S model data was obtained from the ECMWF data portal at
 1020 <https://apps.ecmwf.int/datasets/data/s2s/>. The ERA-interim Reanalysis data was
 1021 obtained from the ECMWF data portal at
 1022 <https://apps.ecmwf.int/datasets/data/interim-full-daily/>. MERRA-2 Reanalysis
 1023 data was obtained from NASA at
 1024 <https://disc.gsfc.nasa.gov/>. This work was initiated by the Stratospheric Network
 1025 for the Assessment of Predictability (SNAP), an activity of SPARC within the World
 1026 Climate Research Programme (WCRP). We acknowledge the scientific guidance of the
 1027 WCRP to motivate this work, coordinated in the framework of SPARC.

1028 Funding by the Swiss National Science Foundation to D.D. through project PP00P2_170523
 1029 is gratefully acknowledged. B.A. was funded by "Ayudas para la contratación de per-
 1030 sonal postdoctoral en formación en docencia e investigación en departamentos de la UCM"
 1031 from Universidad Complutense de Madrid. C.I.G and C.S. were supported by a Euro-
 1032 pean Research Council starting Grant under the European Unions Horizon 2020 research
 1033 and innovation programme (Grant agreement no. 677756). A.Y.K. was funded by the
 1034 Academy of Finland (grants #286298 and #319397). The work by M.T. was supported
 1035 by the JSPS Grant-in-Aid for Scientific Research (C) 15K05286. A.L.L. contributed as
 1036 part of the NOAA/MAPP S2S Prediction Task Force and was supported by NOAA Grant
 1037 NA16OAR4310068 and NSF Award 1547814. S.S. was supported by the National Re-

1038 search Foundation of Korea (NRF) grant funded by the Korean government (Ministry
1039 of Science and ICT) (2017R1E1A1A01074889).

Accepted Article

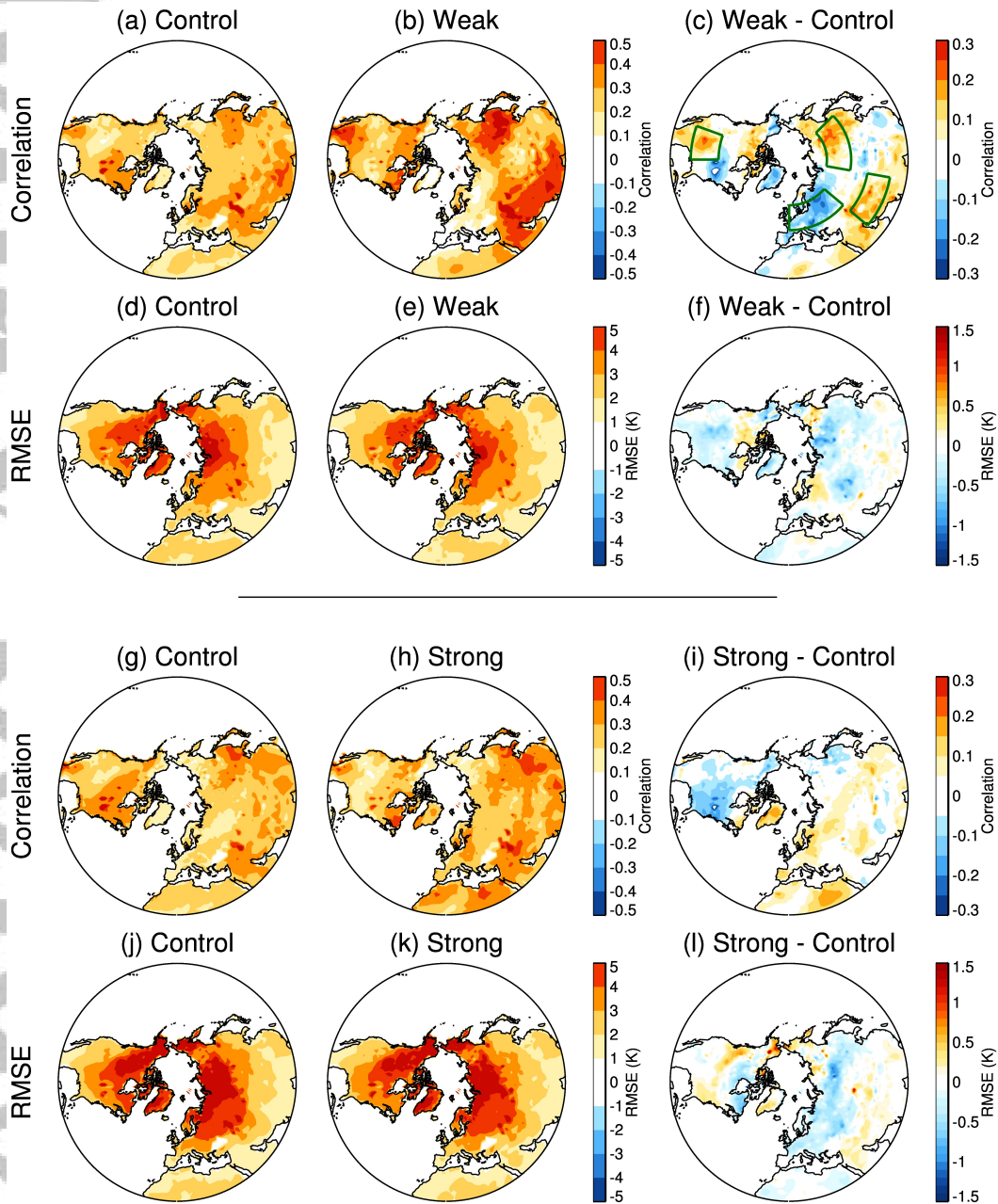


Figure 5. Multi-model mean correlation (see equation 1) and RMSE computed for 2m temperature. (a)-(f) The difference in skill between WEAK and Control forecasts for (top) correlation coefficient and (bottom) RMSE. (left) shows Control forecasts, (middle) shows WEAK vortex forecasts and (right) shows 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.

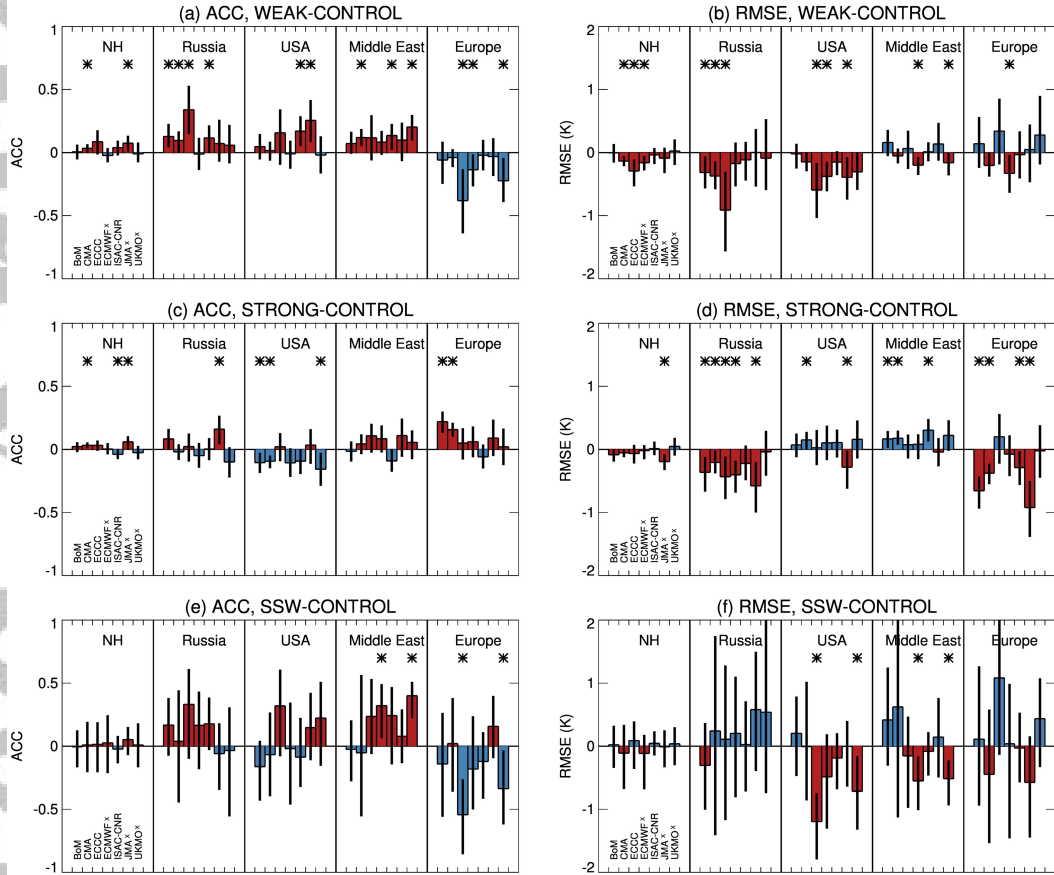


Figure 6. (left) ACC (equation 1) and (right) RMSE (equation 3) for 2m temperature for (top) the difference between WEAK vortex initializations and Control forecasts, (middle) the difference between STRONG vortex initializations and Control forecasts and (bottom) the difference between SSW initializations 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. 'x' indicates high-top models.

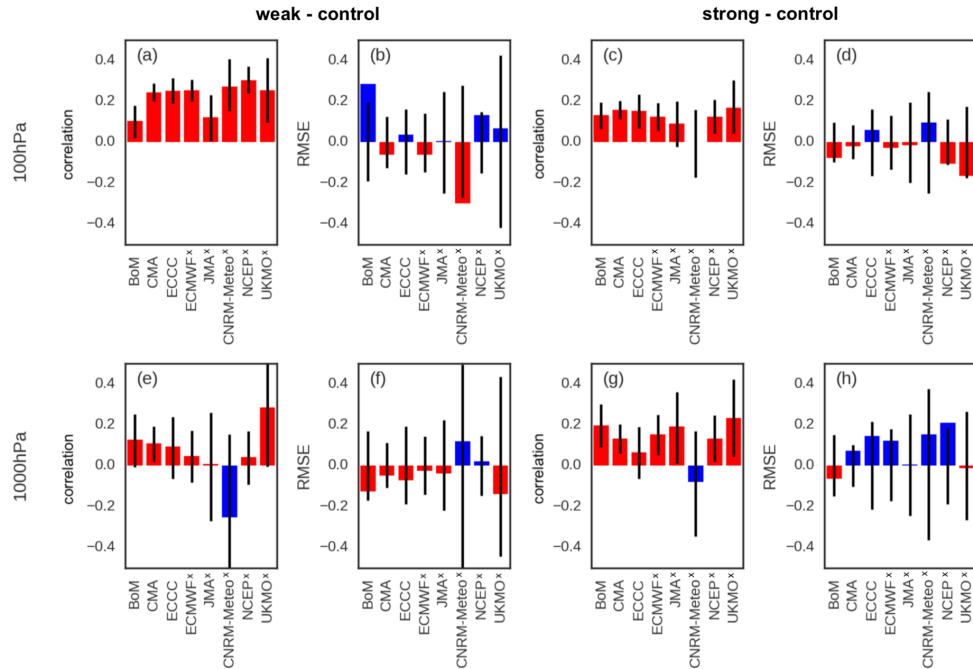


Figure 7. Differences in skill for forecasts initialized during weak (a,b,e,f) and strong vortex (c,d,g,h) for the NAM index at 100 hPa (top) and 1000 hPa (bottom) for the correlation coefficient (equation 1) (a,c,e,g) and RMSE (equation 3) (b,d,f,h). Where the difference represents an improvement (degradation) 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. 'x' indicates high-top models.

Figure.

Accepted Article

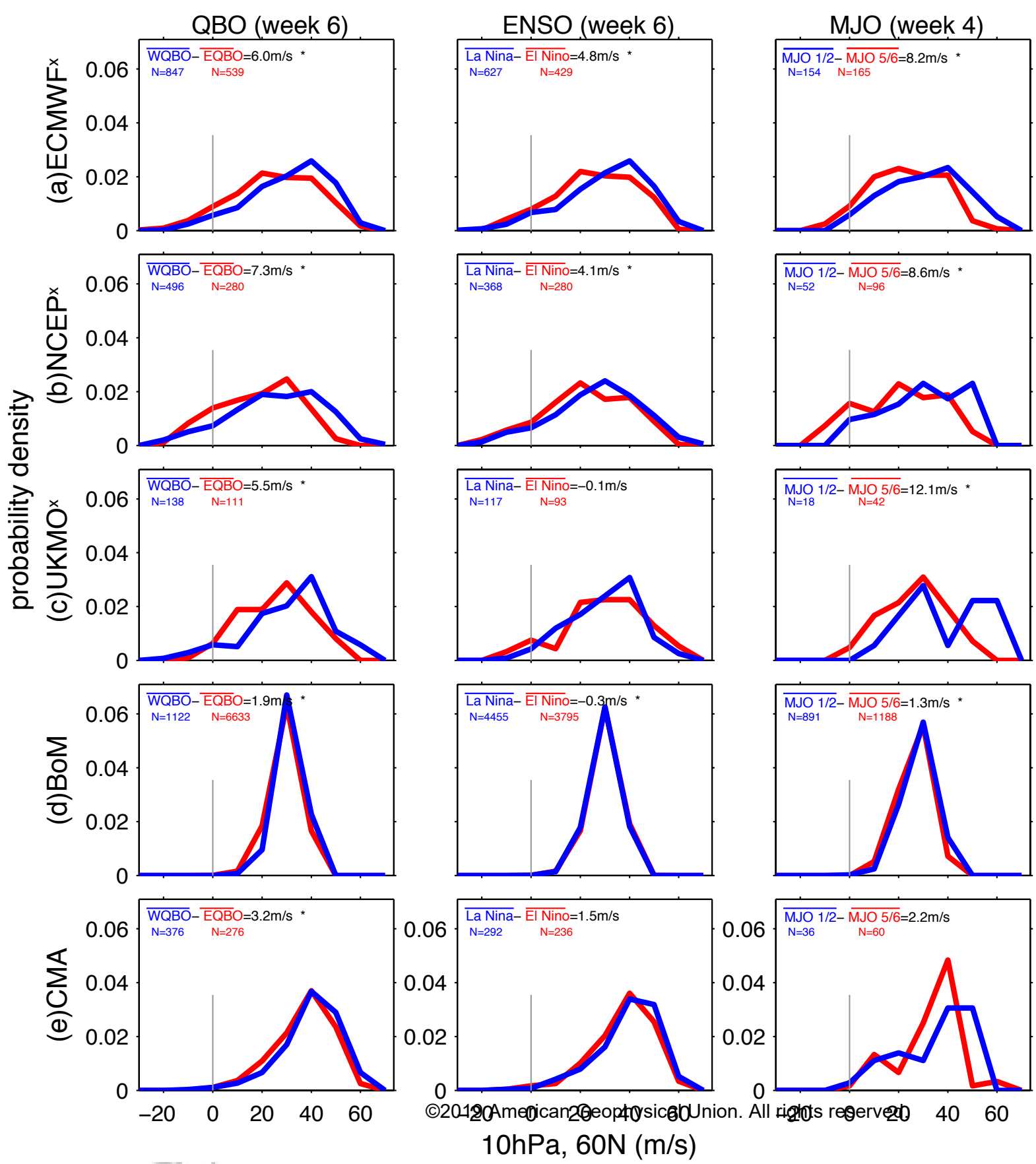


Figure.

Accepted Article

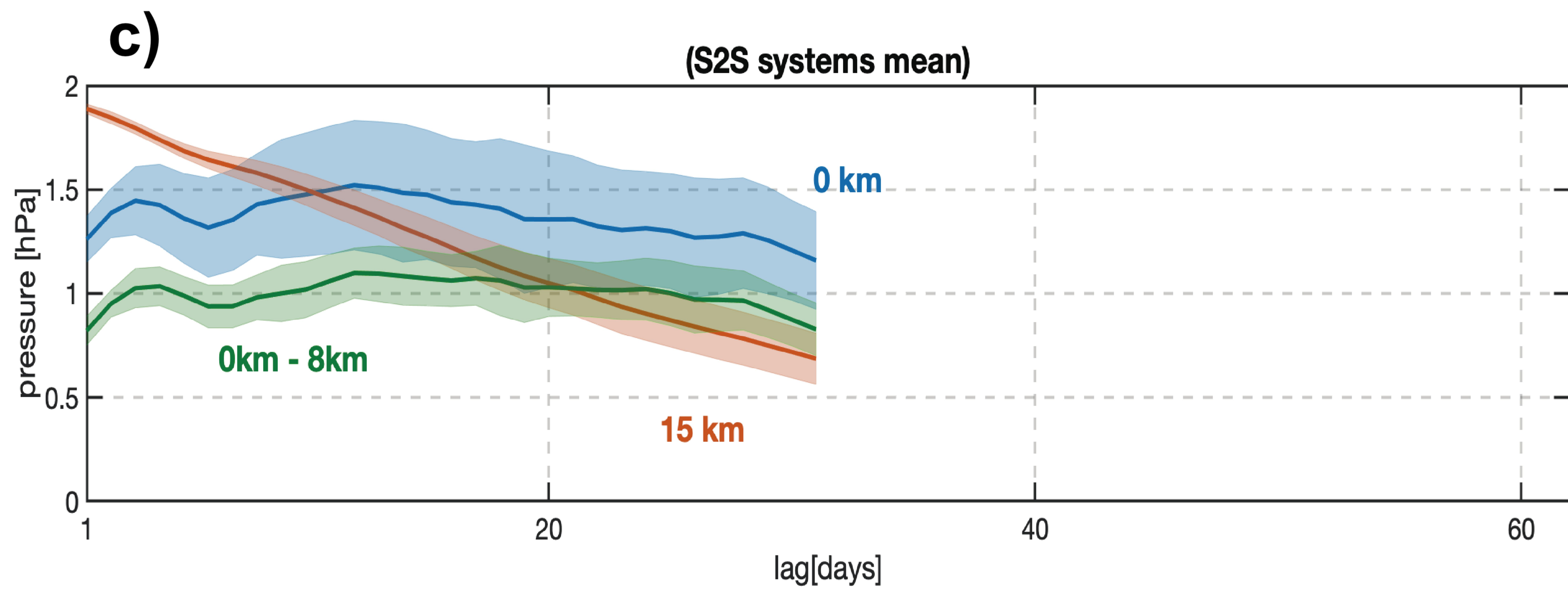
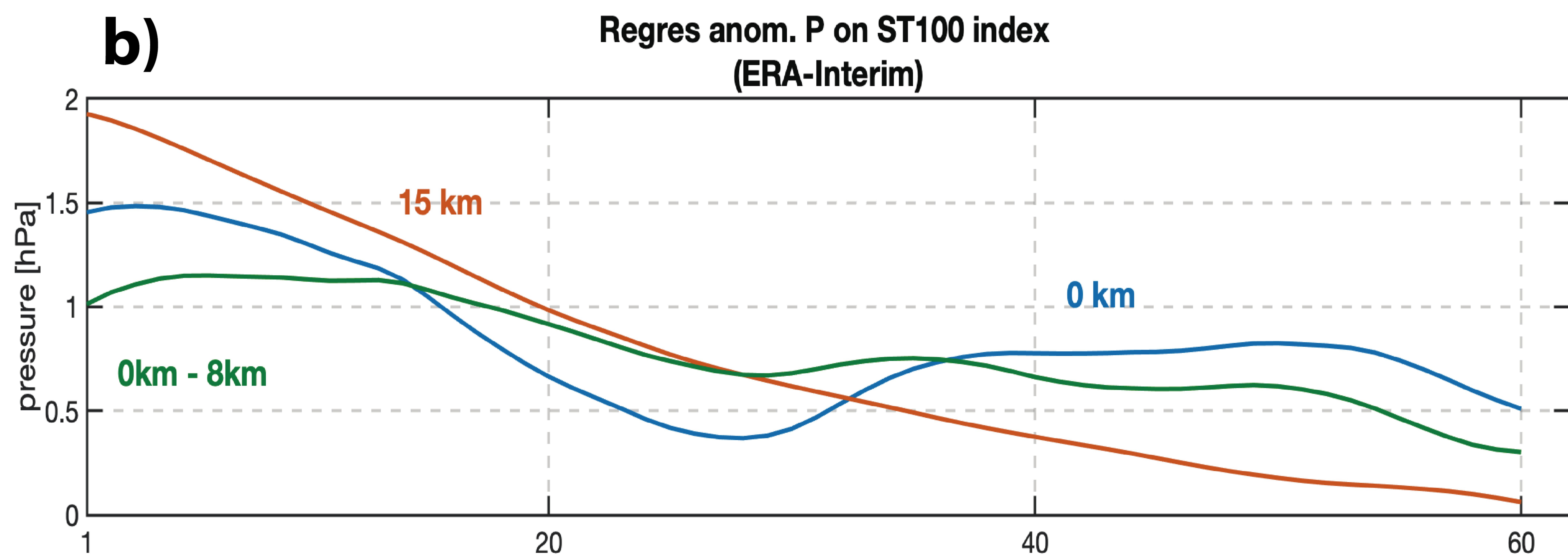
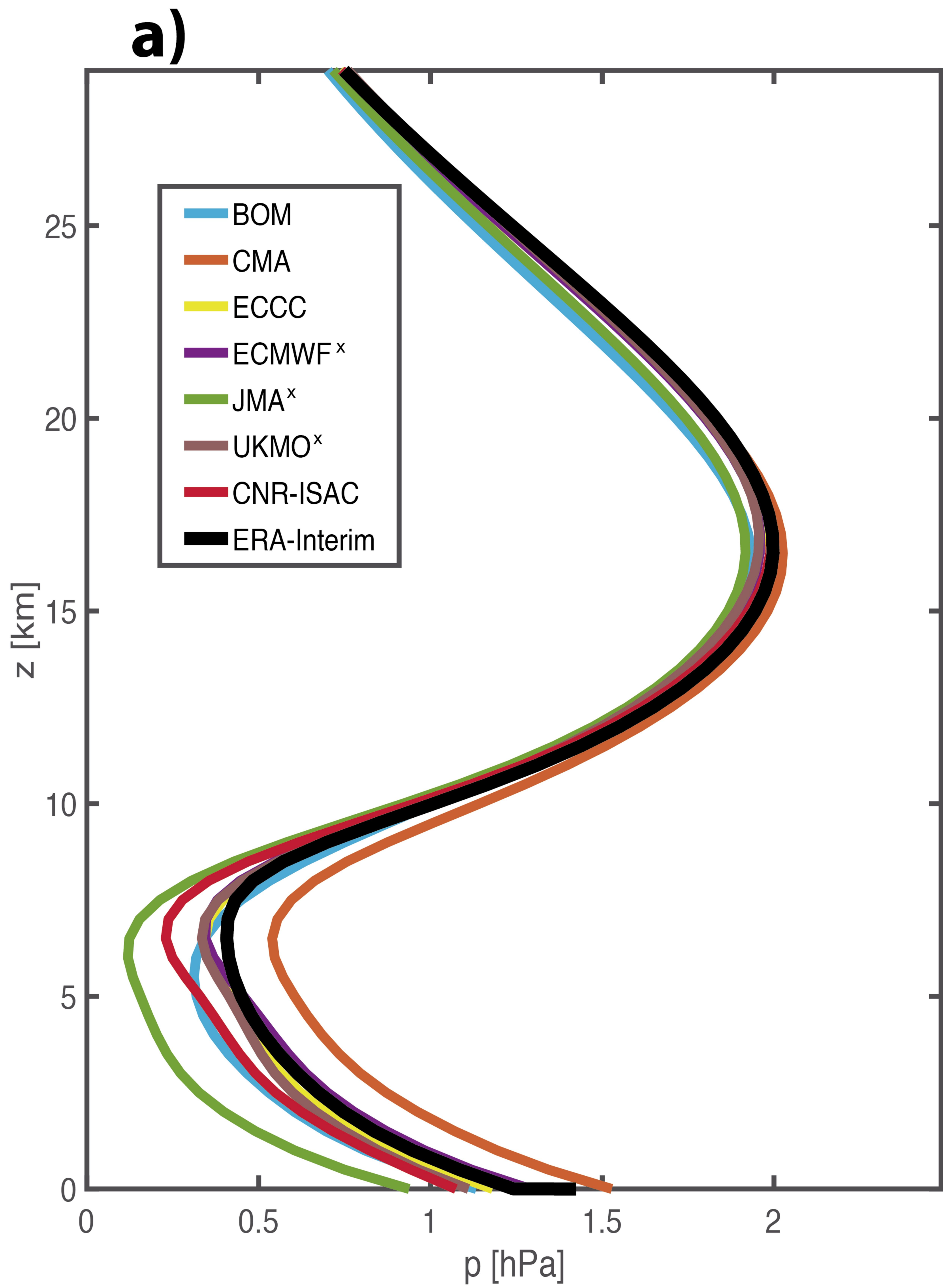
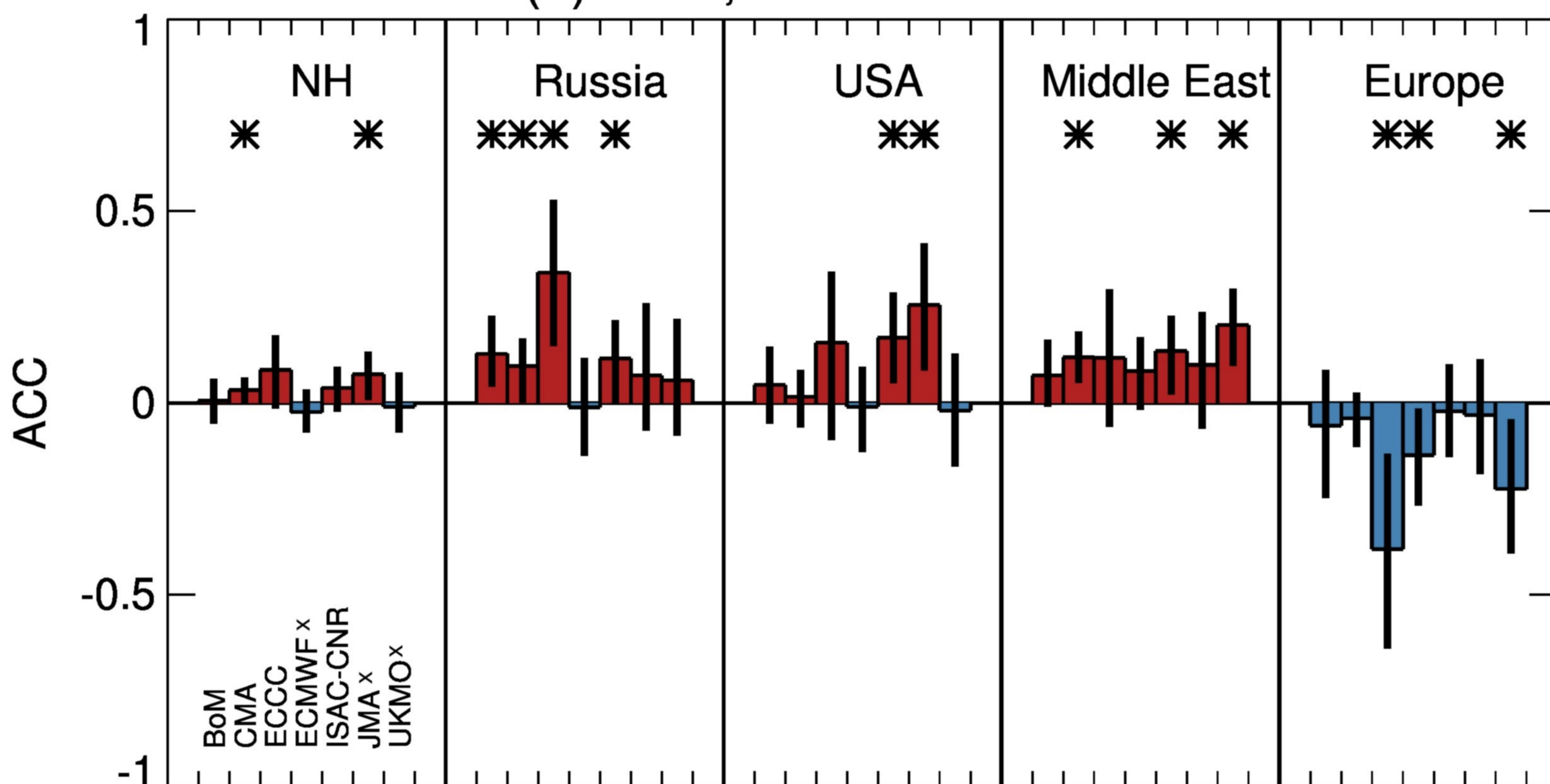


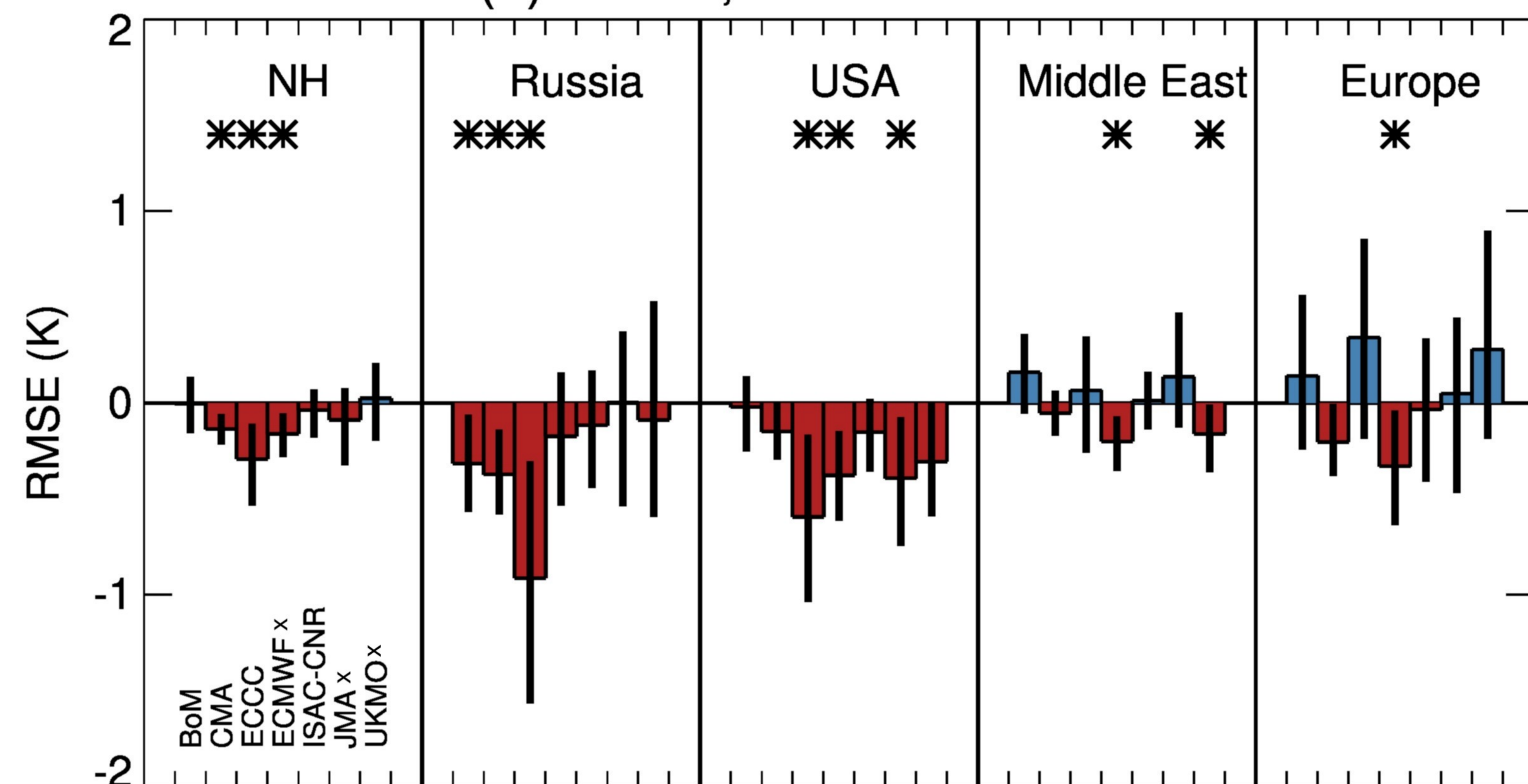
Figure.

Accepted Article

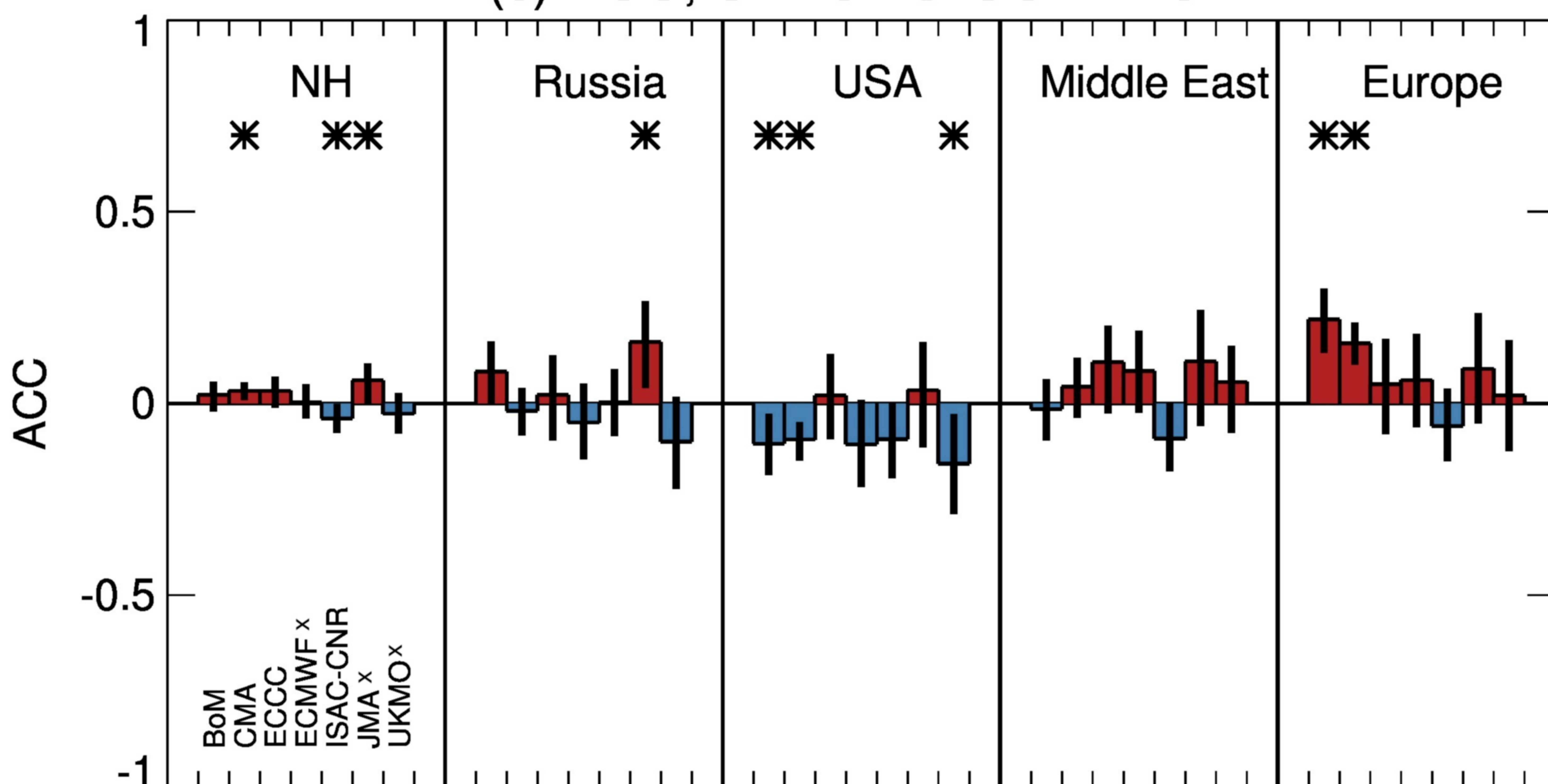
(a) ACC, WEAK-CONTROL



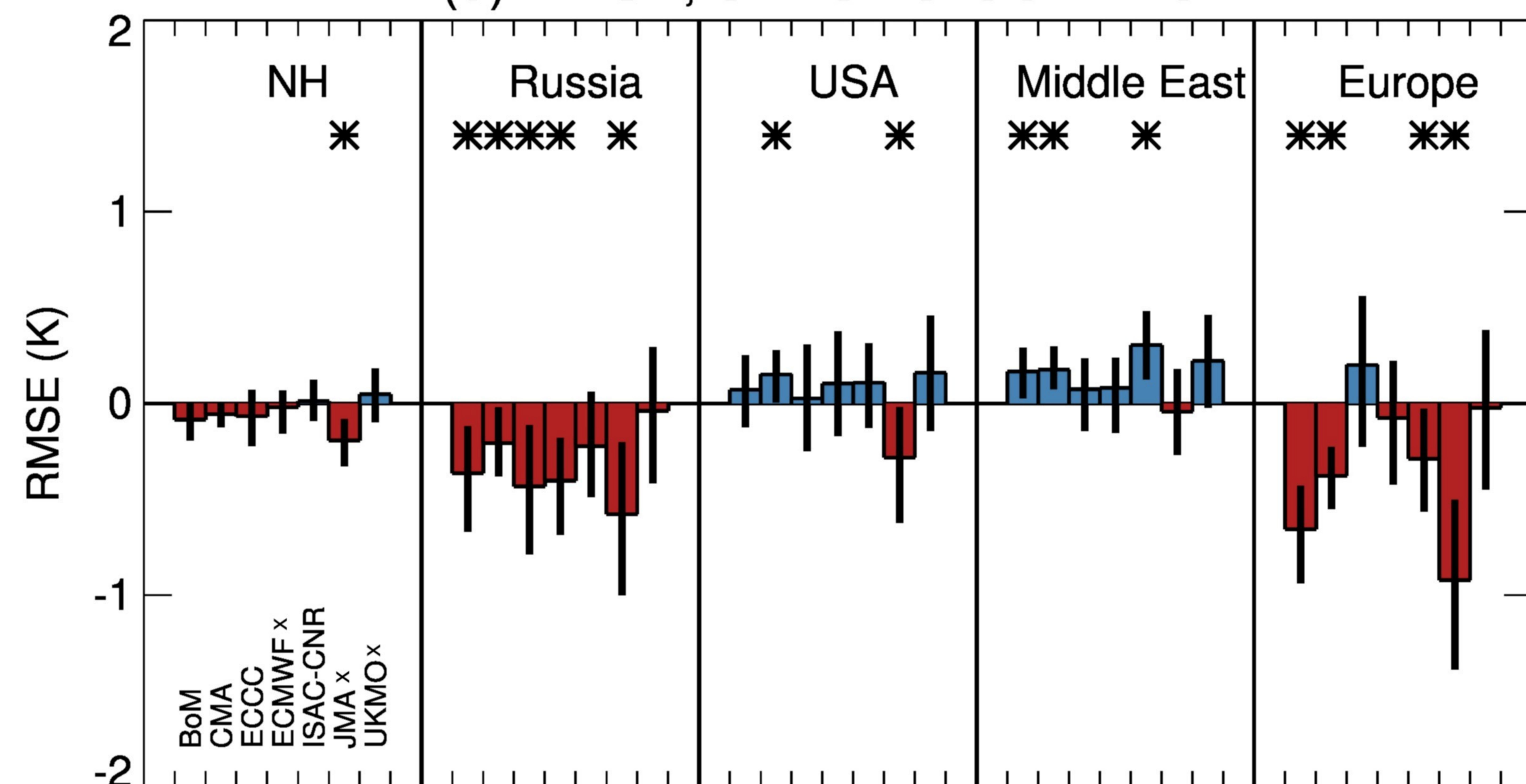
(b) RMSE, WEAK-CONTROL



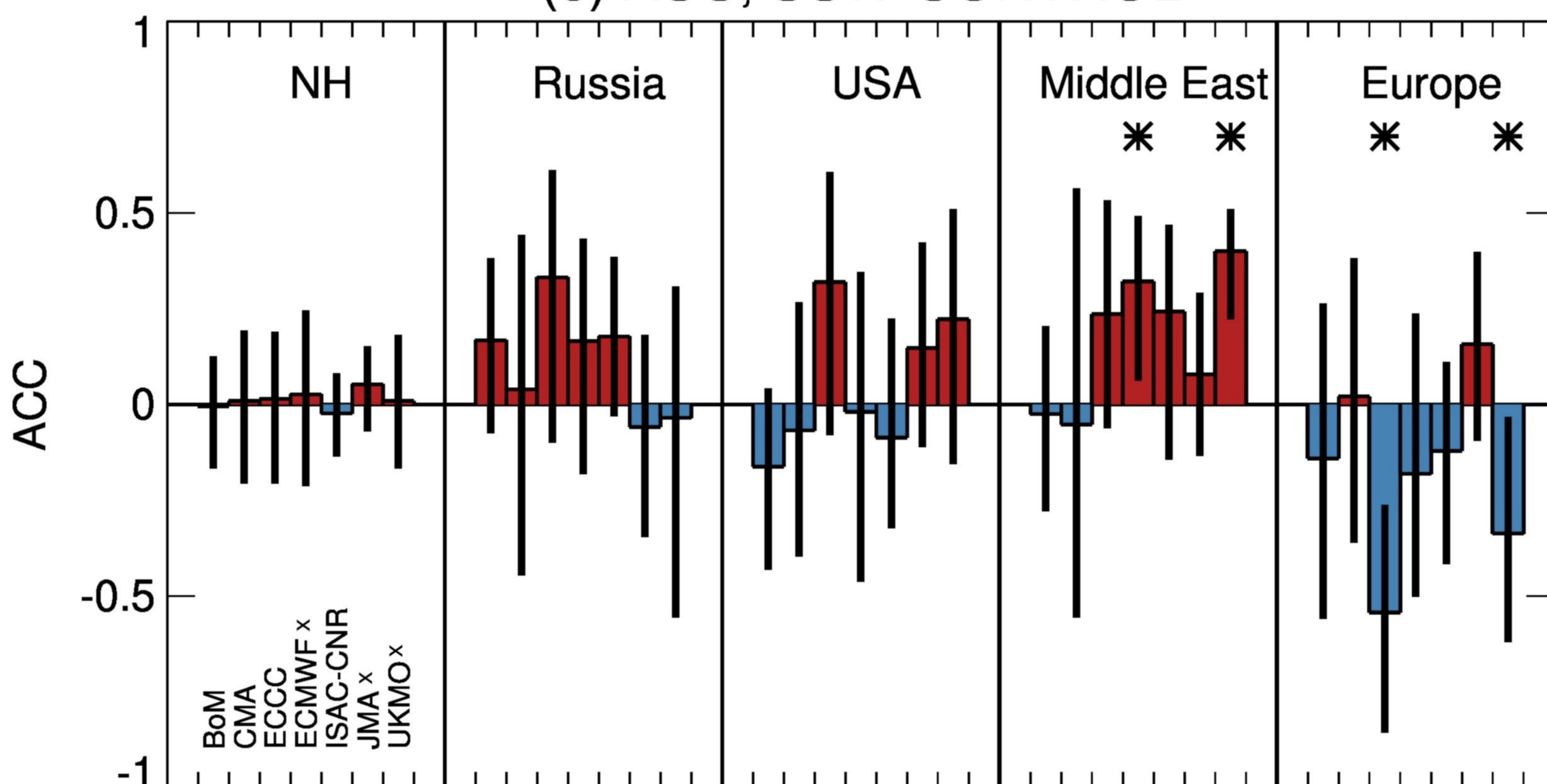
(c) ACC, STRONG-CONTROL



(d) RMSE, STRONG-CONTROL



(e) ACC, SSW-CONTROL



(f) RMSE, SSW-CONTROL

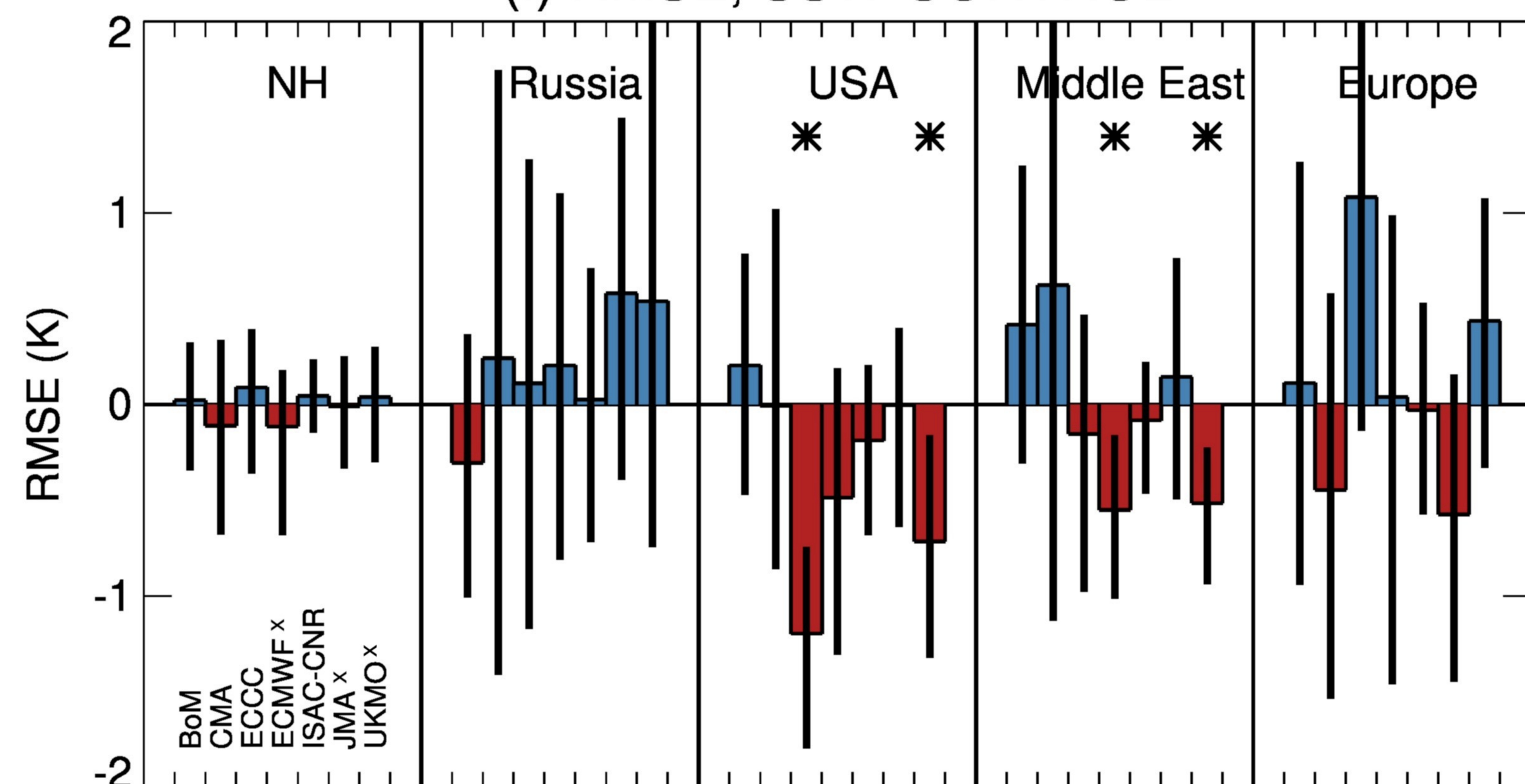


Figure.

Accepted Article

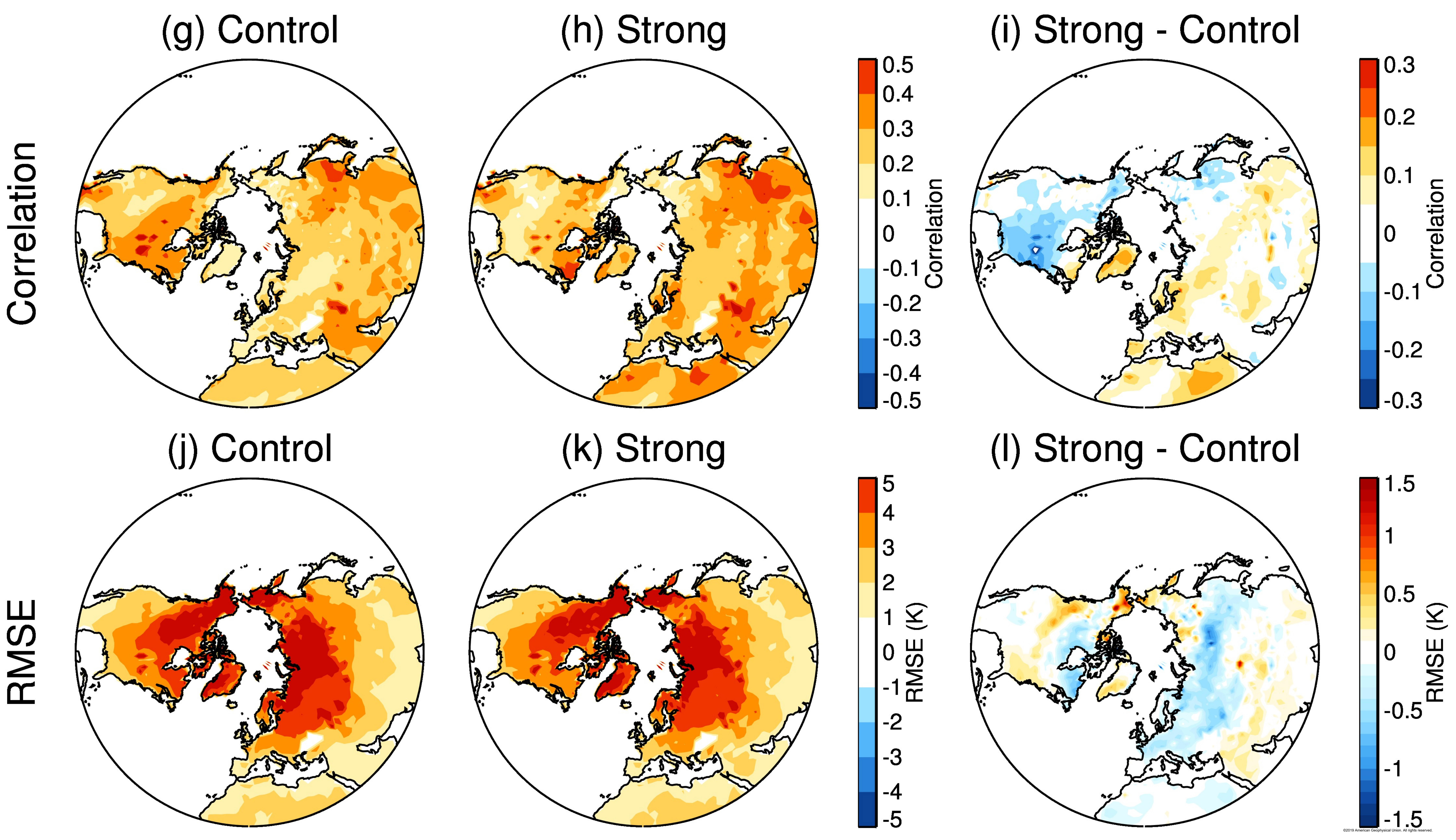
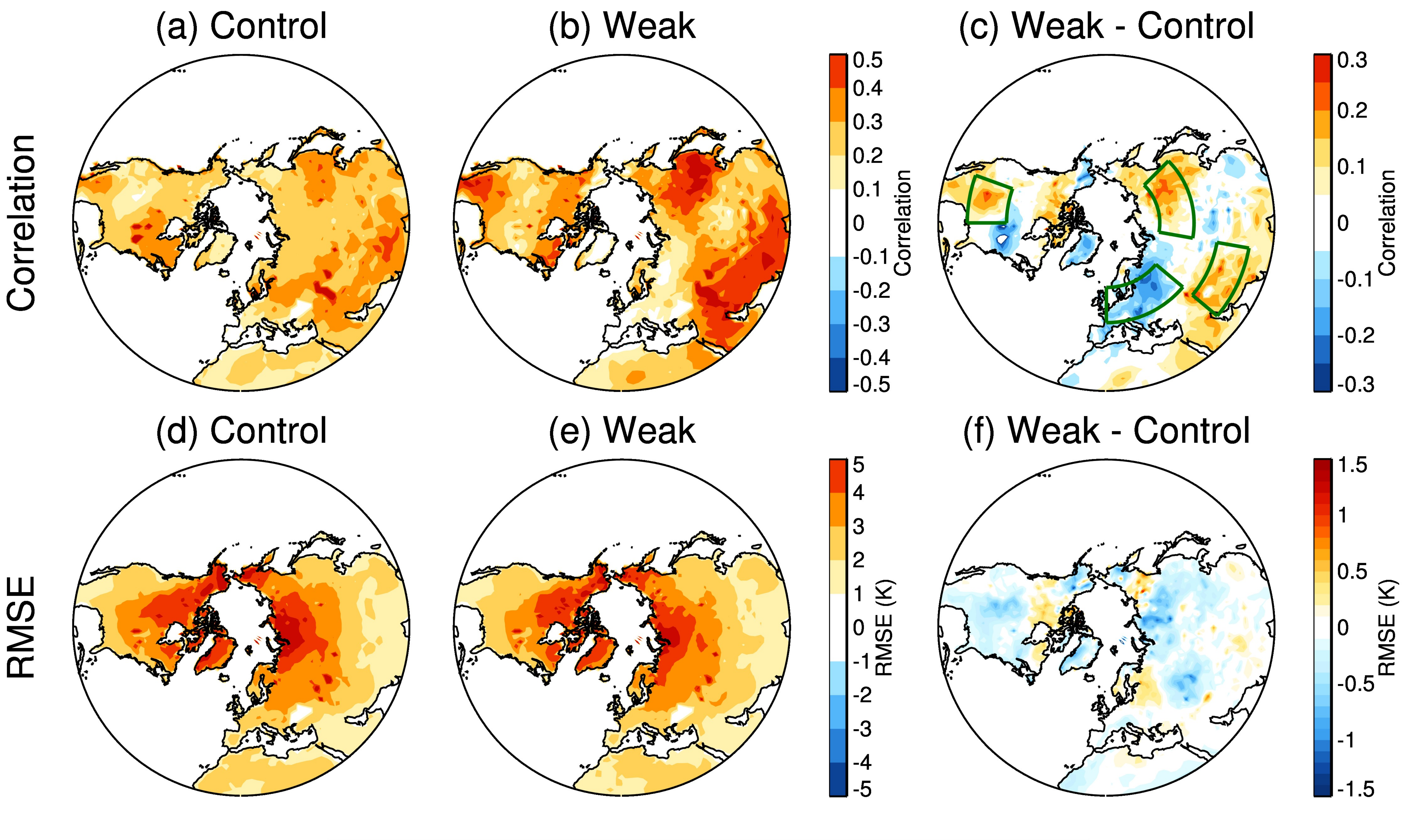


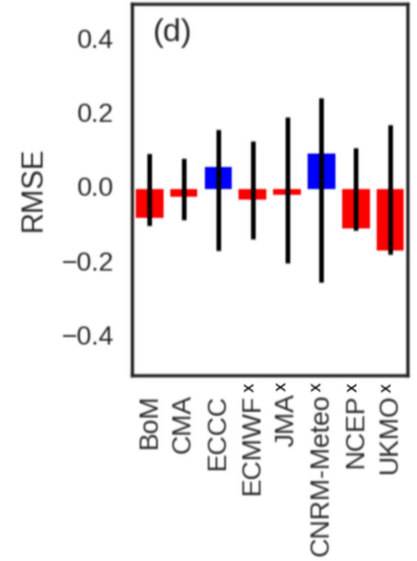
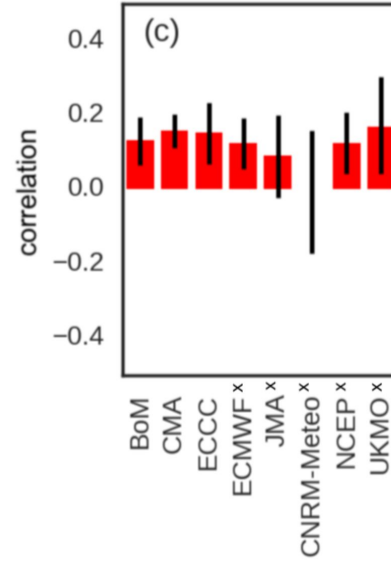
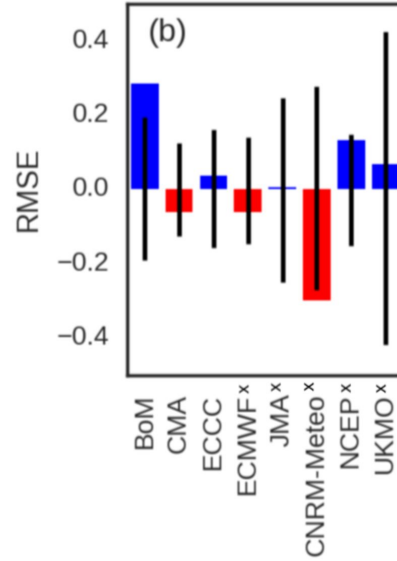
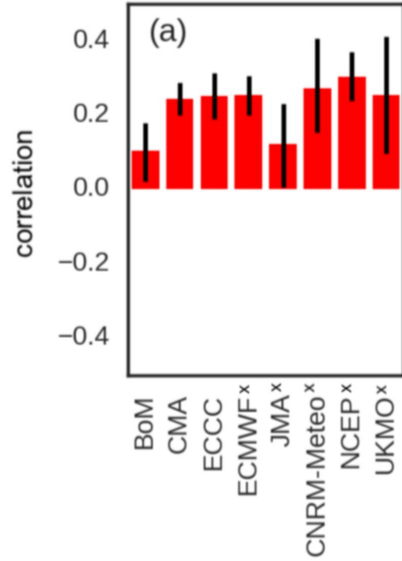
Figure.

Accepted Article

weak - control

strong - control

100hPa



1000hPa

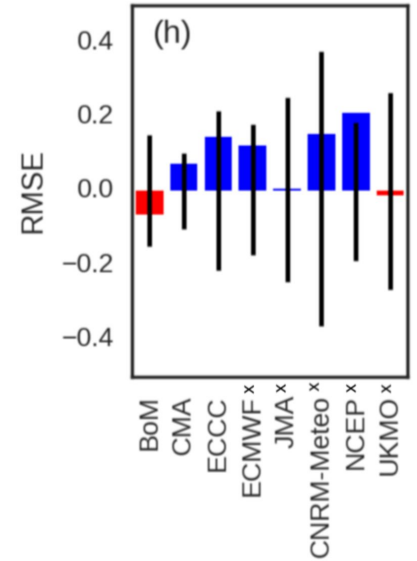
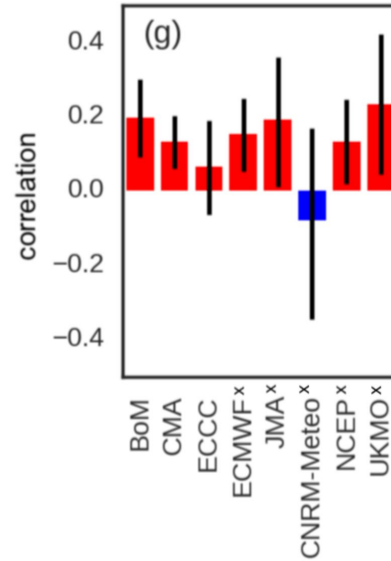
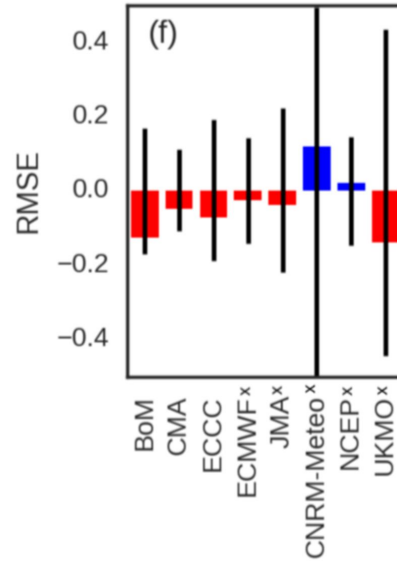
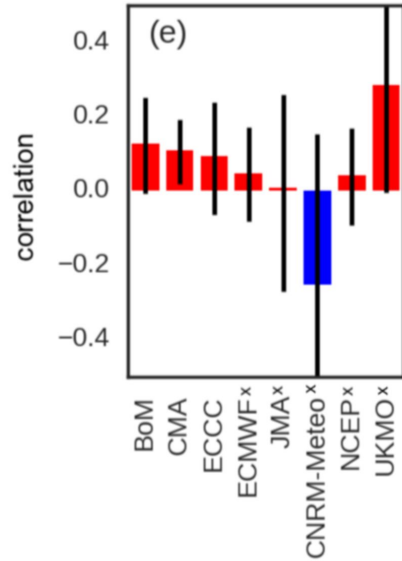
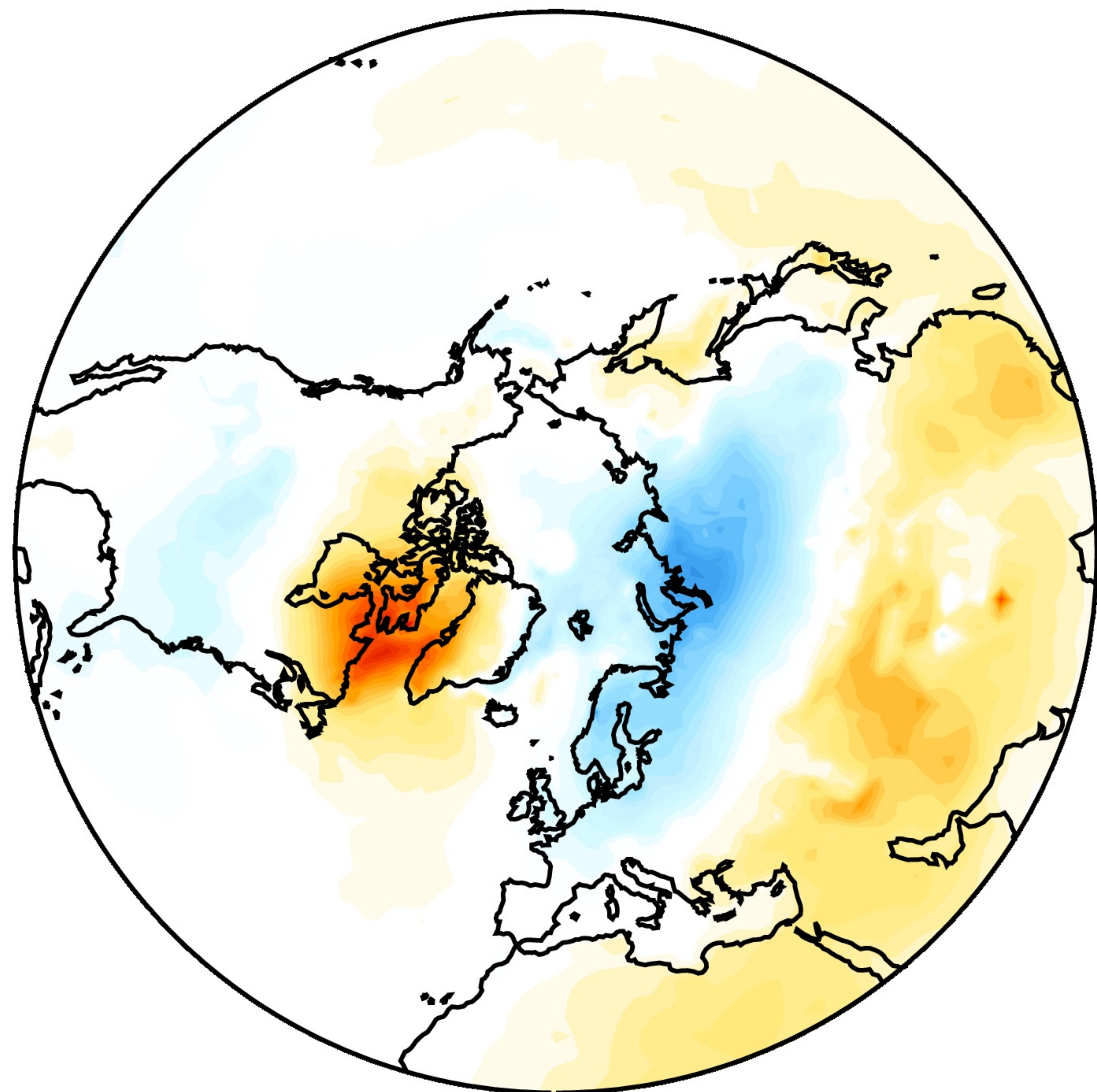
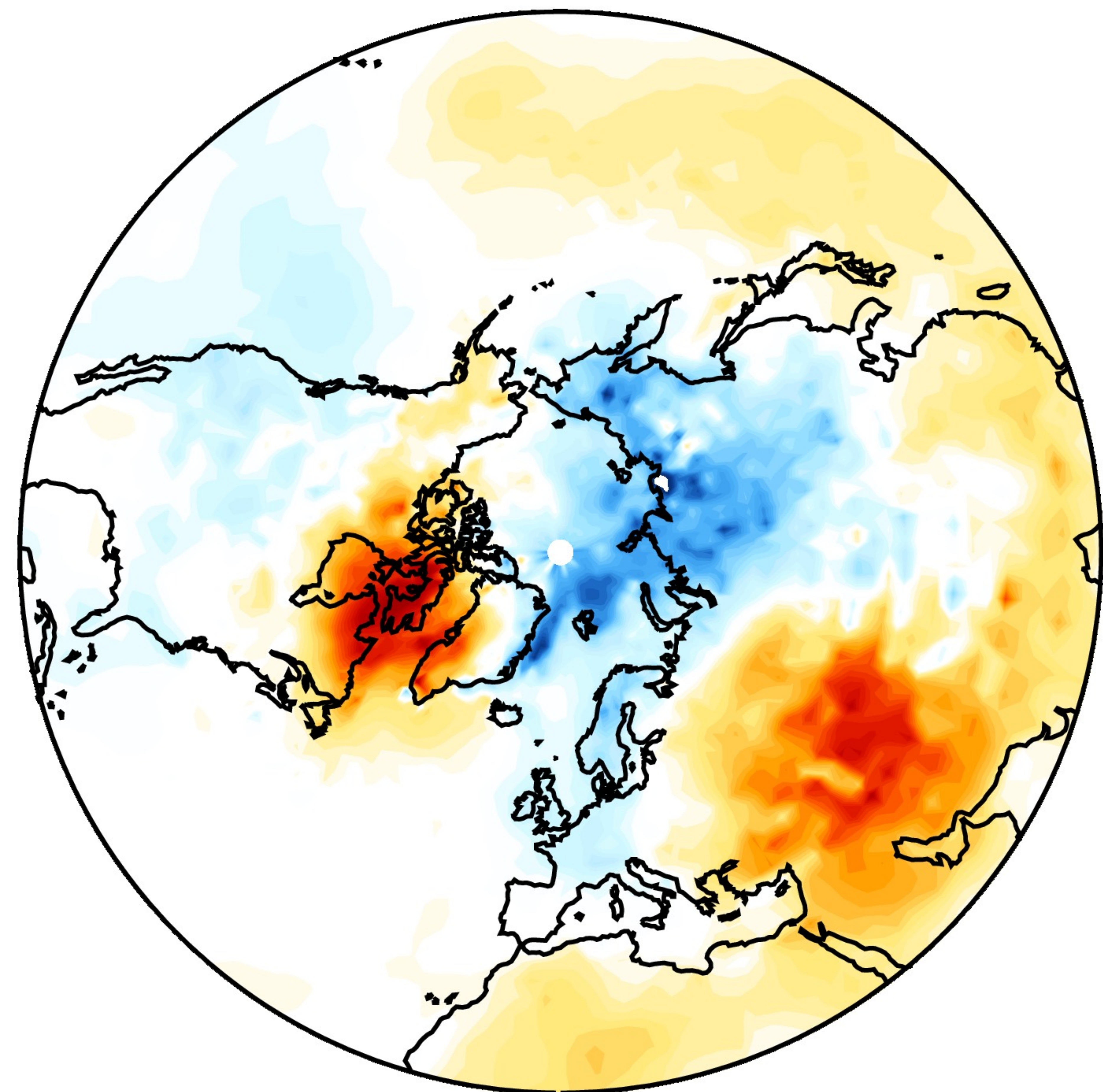


Figure.

Accepted Article

(a) ERA-Interim (weak)

(b) Multi-model mean (weak)



(c) ERA-Interim (strong)

(d) Multi-model mean (strong)

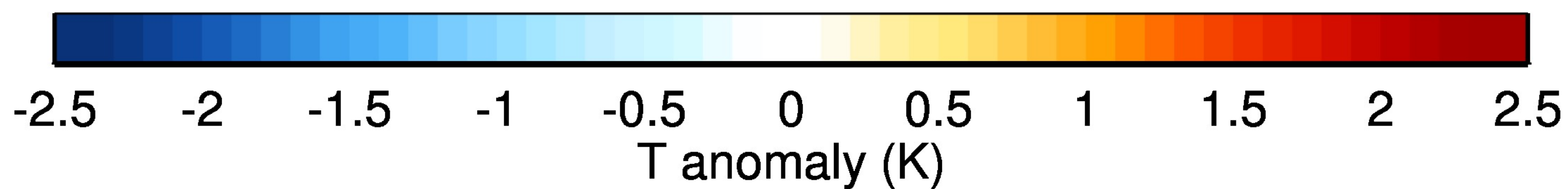
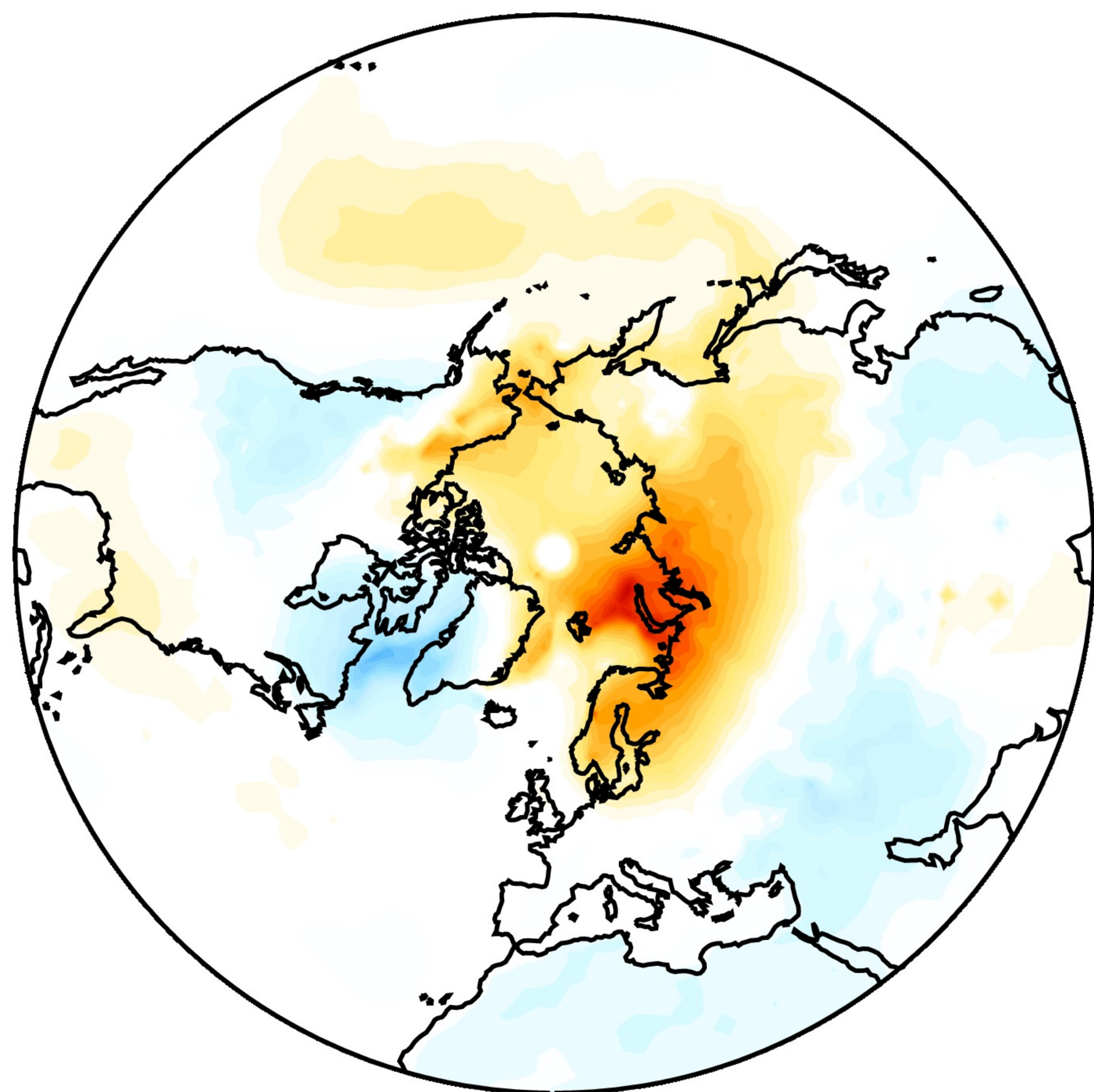
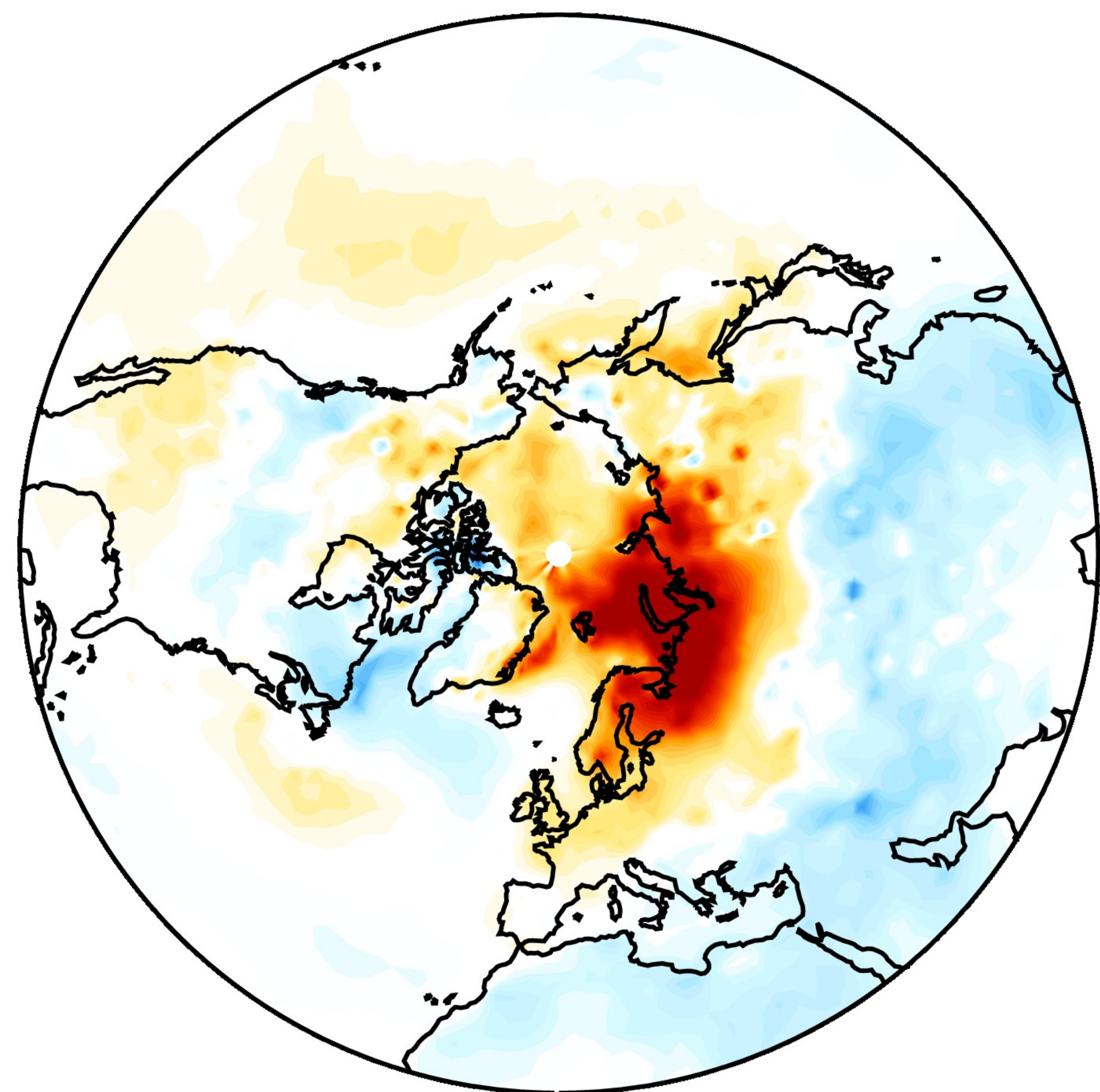
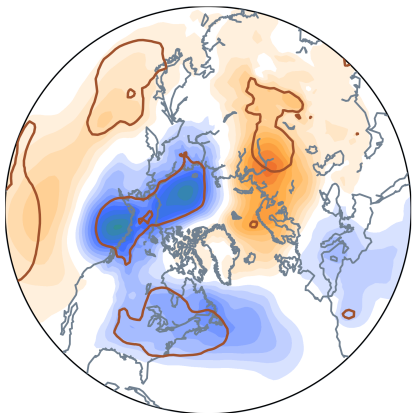


Figure.

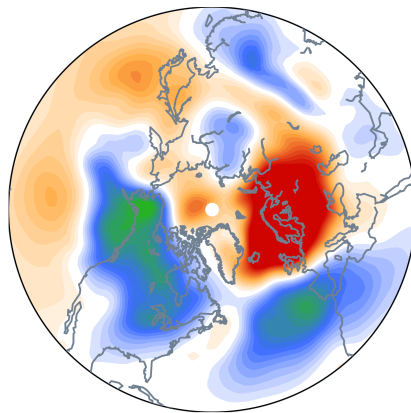
Accepted Article

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

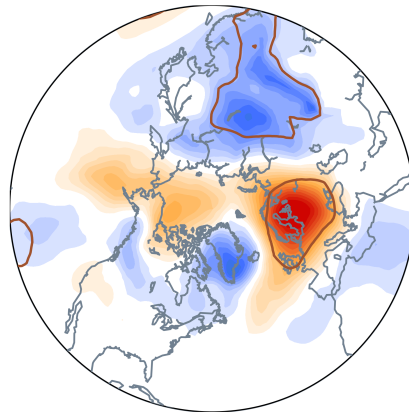
(a) ERA-INTERIM ($N = 11$)



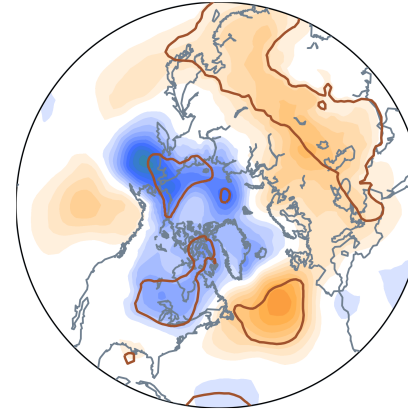
(b) BOM ($N = 2$)



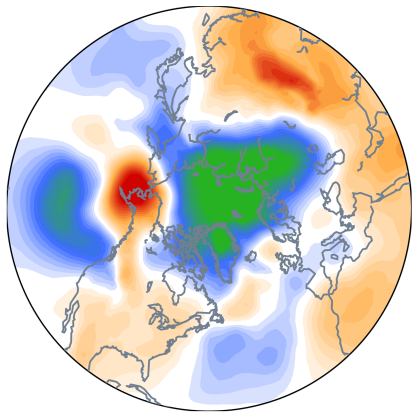
(c) CMA ($N = 13$)



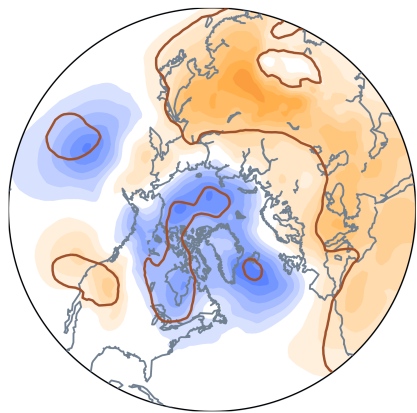
(d) ECMWF^X ($N = 16$)



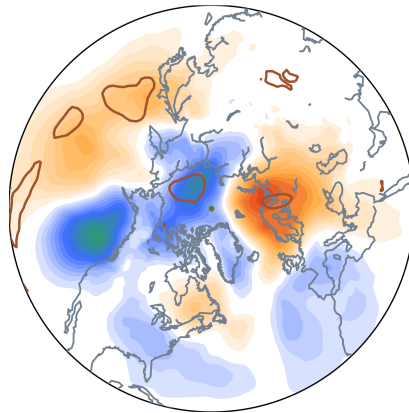
(e) CNRM-Meteo^X ($N = 2$)



(f) NCEP^X ($N = 21$)



(g) UKMO^X ($N = 5$)



(h) Multi-Model Mean ($N = 59$)

