Uncertainty in the response of sudden stratospheric warmings and stratosphere-troposphere coupling to quadrupled CO2 concentrations in CMIP6 models

Article

Accepted Version

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

1   **Uncertainty in the response of sudden stratospheric warmings and stratosphere-**
2   **troposphere coupling to quadrupled CO2 concentrations in CMIP6 models**
3

4   **B. Ayarzagüena[1], A. J. Charlton-Perez[2], A. H. Butler[3], P. Hitchcock[4], I. R. Simpson[5], L. M.**
5   **Polvani[6], N. Butchart[7], E. P. Gerber[8], L. Gray[9], B. Hassler[10], P. Lin[11], F. Lott[12], E.**
6   **Manzini[13], R. Mizuta[14], C. Orbe[15], S. Osprey[9], D. Saint-Martin[16], M. Sigmond[17], M.**
7   **Taguchi[18], E. M. Volodin[19], S. Watanabe[20]**

8   [1] Departamento de Física de la Tierra y Astrofísica, Universidad Complutense de Madrid, Spain

9   [2] Department of Meteorology, Univ. of Reading, UK

10 [3] Cooperative Institute for Environmental Sciences (CIRES)/National Oceanic and Atmospheric
11 Administration (NOAA) Chemical Sciences Division, USA

12 [4] Department of Earth and Atmospheric Sciences, Cornell University, USA

13 [5] Climate and Global Dynamics Laboratory, National Center for Atmospheric Research, USA

14 [6] Columbia University, USA

15 [7] Met Office Hadley Centre, Exeter, UK.

16 [8] Courant Institute of Mathematical Sciences, New York University, USA

17 [9] NCAS-Climate, Department of Physics, University of Oxford, UK

18 [10] Deutsches Zentrum für Luft-und Raumfahrt (DLR), Oberpfaffenhofen, Germany

19 [11] Atmospheric and Oceanic Sciences Program, Princeton University, USA,

20 [12] Laboratoire de Météorologie Dynamique, Ecole Normale Supérieure, France

21 [13] Max-Planck-Institut für Meteorologie, Germany

22 [14] Meteorological Research Institute, Japan

23 [15] NASA Goddard Institute for Space Studies, USA

24 [16] Centre National de Recherches Météorologiques (CNRM), Université de Toulouse, Météo-
25 France, CNRS, Toulouse, France

26 [17] Canadian Centre for Climate Modelling and Analysis Environment and Climate Change,
27 Canada

28 [18] Department of Earth Science, Aichi University of Education, Kariya, Japan

29 [19] Marchuk Institute of Numerical Mathematics, Russia

30 [20] Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Japan

31

32

33 Corresponding author: Blanca Ayarzagüena (bayarzag@ucm.es)

Key Points:

- The tropospheric signal of Sudden Stratospheric Warming (SSWs) in the North Atlantic does not change under $4xCO_2$ forcing.

- There is high uncertainty in changes of SSW frequency under $4xCO_2$ forcing; single models show the rate to be significantly halved or doubled.

- The boreal polar vortex will form earlier and disappear later under increased CO_2, extending the season of stratosphere-troposphere coupling.

Abstract

44

45   Major sudden stratospheric warmings (SSWs), vortex formation and final breakdown
46 dates are key highlight points of the stratospheric polar vortex. These phenomena are relevant for
47 stratosphere-troposphere coupling, which explains the interest in understanding their future
48 changes. However, up to now, there is not a clear consensus on which projected changes to the
49 polar vortex are robust, particularly in the Northern Hemisphere, possibly due to short data
50 record or relatively moderate CO_2 forcing. The new simulations performed under the Coupled
51 Model Intercomparison Project, Phase 6, together with the long daily data requirements of the
52 DynVarMIP project in preindustrial and quadrupled CO_2 ($4xCO_2$) forcing simulations provide a
53 new opportunity to revisit this topic by overcoming the limitations mentioned above.

54   In this study, we analyze this new model output to document the change, if any, in the
55 frequency of SSWs under $4xCO_2$ forcing. Our analysis reveals a large disagreement across the
56 models as to the sign of this change, even though most models show a statistically significant
57 change. As for the near-surface response to SSWs, the models, however, are in good agreement
58 as to this signal over the North Atlantic: there is no indication of a change under $4xCO_2$ forcing.
59 Over the Pacific, however, the change is more uncertain, with some indication that there will be
60 a larger mean response. Finally, the models show robust changes to the seasonal cycle in the
61 stratosphere. Specifically, we find a longer duration of the stratospheric polar vortex, and thus a
62 longer season of stratosphere-troposphere coupling.

1 Introduction

63

64   The stratospheric polar vortex is a strong wintertime circumpolar cyclonic circulation that
65 isolates the polar air masses from air in the lower latitudes [Andrews et al. 1987]. The
66 stratospheric polar vortex forms in Autumn as solar heating vanishes at the pole, establishing
67 strong meridional temperature gradients. The vortex intensifies during winter and then decays in
68 spring as sunlight returns to high latitudes. The springtime breakdown of the vortex, when the
69 zonal winds revert to easterlies, is also known as the stratospheric final warming (SFW).

70   Interest in the polar vortex has increased in the last decades for two different reasons.
71 First, the magnitude of the Antarctic ozone hole is dependent on the state of the polar vortex, as a
72 strong polar vortex is associated with colder temperatures (crucial for heterogeneous ozone
73 chemistry) and reduced mixing with ozone-rich mid-latitude air [Schoeberl and Hartmann,
74 1991]. Secondly, polar stratospheric variability is known to affect not only the stratosphere but
75 also the troposphere, typically projecting onto Annular Mode patterns [e.g.: Baldwin and
76 Dunkerton, 2001; Kidston et al., 2015]. Polar stratospheric variability peaks in the winter
77 hemisphere when the polar vortex is present, as a major source of stratospheric variability is
78 upward propagating, planetary-scale Rossby waves from the troposphere below [Charney and
79 Drazin, 1961]. Under linear theory, the vertical propagation of Rossby waves is limited to
80 regions with westerly winds [Andrews et al. 1987]. Furthermore, because wave activity is greater
81 in the Northern Hemisphere (NH) than in the Southern Hemisphere (SH) so is the polar
82 stratospheric variability. In the SH, stratospheric variability, and thus the coupling to the
83 troposphere, is mainly associated with SFW [Black and McDaniel 2007]. In the NH apart from
84 SFWs [Black et al. 2006; Ayarzagüena and Serrano, 2009; Hardiman et al., 2011], this coupling
85 is primarily associated with polar vortex extremes, in particular, major sudden stratospheric
86 warmings (SSWs). SSWs happen in midwinter and consist in a reversal of wintertime polar
87 stratospheric circulation with a subsequent recovery of the polar vortex after the event. The

88 tropospheric signal of SSWs can persist for up to two months after the occurrence of each event
89 [Charlton and Polvani, 2007]. Although the exact mechanism for this downward influence is still
90 unclear, different hypothesis have been presented in the literature such as wave reflection,
91 downward control or responses to stratospheric redistributions of potential vorticity among
92 others [Song and Robinson, 2004 and references herein]. In most of these theories the role of the
93 circulation anomalies of the lower stratosphere was found to be extremely important to define the
94 impact on the troposphere. Indeed, recently Hitckcock et al. [2013] defined a subset of SSWs,
95 called Polar-night Jet Oscillation events (PJOs), which are characterized by a very persistent
96 warm polar lower stratosphere and whose signal in the troposphere is particularly strong and
97 persistent too.

98   The importance of polar vortex variability for both atmospheric dynamics and ozone
99 chemistry has spurred considerable efforts in identifying if and how the stratospheric polar
100 vortex might respond to increasing greenhouse gases (GHGs). While several studies have been
101 devoted to this question, there is not consensus at this time on which projected changes to the
102 polar vortex are robust. Here, and throughout the paper, we use the word robust to mean a strong
103 agreement across many models as to the size and amplitude of the changes to the stratospheric
104 polar vortex under increased GHG. To offer a trivial example: a two-model ensemble in which
105 one model predicted a halving of SSW frequency and the other model predicted a doubling of
106 SSW frequency would not represent a robust prediction of future changes, although both these
107 changes might be statistically significant in each model. On the contrary, if one model predicted
108 a significant increase of SSW frequency by a factor of 2.5 and the other by a factor of 2, we
109 would regard this as a robust prediction.

110   Early studies using simple models demonstrated polar stratospheric cooling under
111 increased GHG forcing [Manabe and Wetherland, 1967; Fels et al. 1980]. Global atmospheric
112 modeling work in the 1990s (with prescribed changes in sea surface temperatures) projected a
113 boreal polar warming in winter, but no consensus on the changes in the number of SSWs [Rind
114 et al. 1990; Mahfouf et al 1994; Rind et al. 1998; Butchart et al 2000]. Moreover, after decades
115 of improvement in modeling the stratosphere, a clear consensus about future changes to the polar
116 vortex is still missing. For instance, one can find in the literature a number of single-model
117 studies that report a significant increase in the frequency of SSWs in the future [Charlton-Perez
118 et al., 2008; Bell et al., 2010], while other studies report a non-statistically significant increase
119 [e.g. Mitchell et al., 2012a; Ayarzagüena et al. 2013], and others no significant change in SSW
120 frequency at all [McLandress and Shepherd, 2009; Scaife et al., 2012; Karpechko and Manzini,
121 2012]. Multi-model intercomparisons of Chemistry Climate Model Validation (CCMVal) and
122 Coupled Model Intercomparison Project 5 (CMIP5) models have reported large discrepancies in
123 the sign of change among models [Mitchell et al. 2012b; Kim et al., 2017].

124   Recently, Ayarzagüena et al. [2018] revisited this topic, trying to overcome some of the
125 issues suggested in the literature as potential reasons for this disagreement, such as the use of one
126 single model in the analysis or the dependence of results on the SSW identification criterion.
127 They analyzed 12 different models participating in the Chemistry Climate Model Initiative
128 (CCMI) and applied several different (absolute and relative) criteria for the identification of
129 SSWs. The outcome was again a lack of a significant change in SSWs frequency in the future,
130 although most of the models predicted a slight increase in the frequency of these, regardless of
131 the SSW identification algorithm. One might argue, however, that the limited data record
132 available (40 years in each period of study), and the relatively moderate GHG forcing used in the

133 central CCMI scenario (Representative Concentration Pathway 6.0, RCP6.0), might be
134 insufficient to detect significant changes in SSWs in those simulations.

135   The new CMIP6 model generation together with the special data requirements of the
136 DynVarMIP project [Gerber and Manzini, 2016] provide a new opportunity to revisit the
137 question of the effects of increasing CO_2 on the interannual variability of the stratospheric polar
138 vortex. The very long daily data record at stratospheric levels of the Diagnostic, Evaluation and
139 Characterization of Klima (DECK) experiments allows us, for the first time, to try to isolate
140 forced changes in stratospheric variability in a larger ensemble of high-top models than possible
141 previously [Eyring et al. 2016]. Specifically, one of these DECK simulations consists of a very
142 high CO_2 forcing (abrupt4xCO_2) enabling the exploration of changes in the vortex variability
143 under an extreme future scenario. Furthermore, the daily output of the 1pctCO_2 simulation with a
144 gradual increase of CO_2 allows us to investigate the time of emergence of SSW changes.

145   The goal of this study is to analyze the potential changes in the interannual variability of
146 the polar vortex due to increasing CO_2 concentrations, as simulated by CMIP6 models. Apart
147 from the mentioned new possibilities opened up by the availability of CMIP6 data, we have also
148 examined other characteristics that are relevant for the stratosphere-troposphere coupling such as
149 the seasonal cycle of the polar vortex, i.e. formation and final breakdown, in both hemispheres,
150 as well as changes in stratosphere-troposphere coupling during SSWs, given the importance of
151 these aspects for tropospheric impacts and predictability. However, we do not aim here to fully
152 diagnose stratospheric variability in the CMIP6 models, nor to explain in detail why models
153 differ in their estimates of the sensitivity of the stratospheric polar vortex to CO_2 forcing.
154 Instead, we simply aim to provide a timely, quantitative estimate of how stratospheric variability
155 might change under CO_2 forcing since this information is of critical importance to the upcoming
156 Intergovernmental Panel on Climate Change (IPCC) AR6 report, and for future work on the
157 stratosphere in CMIP6 models.

158

159 **2 Data and methodology**

160 2.1 Data

161   In this study we analyze the daily output of DECK simulations by 12 CMIP6 models
162 participating in the DynVarMIP initiative (Table 1). All the models are coupled to an ocean and
163 sea ice model, and most (8 out of 12) are "high-top" models, defined by having a model top at or
164 above 0.1hPa as in Domeisen et al. [2019]. A priori, we expect the high-top models to have more
165 realistic polar stratospheric variability and, consequently, to better simulate SSWs, and their
166 frequency and surface impacts, than low-top models [Charlton-Pérez et al. 2013]. For the CMIP6
167 ensemble, there is a much larger number of models that have a high model top than in the
168 previous CMIP5 ensemble. In order to make sure our model sample is unbiased, only a single
169 member of each model ensemble is analyzed here; details are shown in Table 1.

170   We focus on four DECK experiments [Eyring et al., 2016], each of them used for
171 different purposes. The historical run is employed for model validation: we compare the
172 simulated SSW frequency, intensity and seasonality to the values obtained from the JRA-55
173 reanalysis [Kobayashi et al., 2015]. In fact, we have specifically restricted the analysis period to
174 1958-2014 to perform a rigorous quantitative comparison with JRA-55. This reanalysis shows a
175 very good performance in representing SSW [Ayarzagüena et al. 2019] and is the most modern

176 reanalyses of the three that extend longer than the satellite era and assimilate more than surface
177 data (ERA-40, NCEP/NCAR reanalysis and JRA-55).

178   The pre-industrial Control (piControl) experiment is used for two purposes. Since it
179 contains a very long data record (more than 450 years for most of the models, Table 1), it is used
180 to characterize both the baseline estimates of SSW frequency and intensity, and to characterize
181 internal atmospheric variability in SSW frequency and trends.

182   The abrupt4xCO$_2$ and 1pctCO$_2$ runs are used to examine the impact of CO$_2$ forcing on
183 SSW properties. Both simulations extend 150 years (except for the abrupt4xCO$_2$ in IPSL-
184 CM6A-LR, which is 900 years long, and GISS-E2.2AP, which is 81 years long). All forcings in
185 the abrupt4xCO$_2$ simulations are identical to those in the piControl run, except for the CO$_2$
186 concentrations, which are abruptly quadrupled from piControl levels, and then are held constant
187 throughout the entire length of the simulation [Eyring et al. 2016]. The large and constant forcing
188 in the abrupt4xCO$_2$ makes it possible to isolate robust changes, if any, to the size and nature of
189 changes to SSW properties. In the 1pctCO$_2$ simulation, the CO$_2$ concentration starts at pre-
190 industrial levels and is increased at the rate of 1% per year. This simulation is used to estimate
191 the rate at which SSW frequency might change in the future [one aspect of the so-called
192 'dynamical sensitivity' of the stratosphere, Grise and Polvani 2016].

193   Anomalies are defined as the departure from the daily evolving annual cycle of each
194 respective model. In the piControl run, the climatology is based on the whole period, while in the
195 historical run only the 1979-2014 is considered for calculating the climatology. In the
196 abrupt4xCO$_2$ runs, a trend is identified in some variables during the first 50 years following the
197 switch-on of the forcing. To avoid this trend, the climatologies are computed after omitting the
198 first 75 years except for IPSL-CM6A-LR where we omit the first 300 years but we keep the
199 following 600years. A similar omission of data is performed for the analysis of SFW or vortex
200 formation dates. In contrast, the full abrupt4xCO$_2$ is considered when looking at SSW frequency
201 as no trend is detectable in the occurrence of these phenomena.

202 2.2 Methods

203   There has recently been a considerable discussion in the literature as to which metrics
204 best characterize the variability of the stratospheric polar vortex, in particular, extreme vortex
205 weakening events [Butler et al. 2015; Butler and Gerber, 2018]. However, in a recent study,
206 Ayarzagüena et al. [2018] found little dependence on the choice of metrics in terms of
207 documenting future changes in SSWs. Thus, we here focus only on a few, widely-used and easily
208 implementing metrics of stratospheric variability. Future work will likely be able to explore
209 stratospheric variability in more detail, and possibly reveal subtleties in changes to stratospheric
210 circulation not apparent in our initial analysis. Furthermore, focusing on commonly used
211 diagnostics allows us to place our work in the context of previously published studies on changes
212 in, for example, SSW frequency.

213   Several aspects of the stratospheric polar vortex (formation, final breakdown and
214 variability) are analyzed using the zonal mean zonal wind at 60ºN and 10hPa ($u_{60N10hPa}$) for the
215 NH, and 60ºS and 10hPa for the SH.

216 • <u>SSWs</u> are identified following the criterion proposed in Charlton and Polvani [2007], which
217   is based on the reversal in the sign of $u_{60N10hPa}$ from November to March. Their criterion
218   includes two additional restrictions: (1) winds must return to westerly for at least 20

219   consecutive days between events and (2) winds must return to westerly for at least 10
220   consecutive days before April 30 of each year. Recall that this definition only identifies so-
221   called "major" SSWs. Here we do not examine other aspects of polar vortex variability, such
222   as vortex intensification events, wave reflection events, or minor stratospheric warmings.

223 • SFWs is defined as the last date in the spring on which $u_{60N10hPa}$ reverses and does not return
224   to westerly for more than 10 consecutive days [Butler and Gerber, 2018].

225 • The polar vortex formation date is identified as the first time that $u_{60N10hPa}$ turns westerly after
226   1 July, in the NH, and stays westerly for at least 10 days.

227 • PJOs are identified by applying a slight variation of criteria established by Hitchcock et al.
228   [2013], as the original required finer vertical resolution than available. The new metric has
229   been validated in reanalysis to ensure that similar results are obtained in this case as to those
230   obtained by applying the original one (not shown). Here, the identification is based on two
231   time series PC_1 = T'(5 hPa) - T'(100 hPa) and PC_2 = T' (50 hPa), where T' indicates the
232   polar-cap averaged temperature anomaly (from climatology) at the specified pressure level.
233   These time series are transformed into polar coordinates r(t) and phi(t), and the central dates
234   of events are defined by when the phase phi(t) passes counter-clockwise through $3\pi/2$, so
235   long as the amplitude r(t) is greater than 2.5σ. Once a central date is defined, the starting date
236   of the event is defined by the most recent date prior to the central date when r(t) is below
237   1.5σ, and similarly the ending date of the event is defined by the earliest date following the
238   central date when the r(t) is below 1.5σ.

239 2.2 Statistical methods

240   Two methods to calculate the statistical significance of changes to the SSW frequency are
241 used: a parametric method based on an assumption that the SSW frequency can be estimated
242 using a Poisson point process, and a non-parametric bootstrapping technique based on
243 resampling the piControl run of each model. Trends in SSW frequency and the time of
244 emergence of these trends are estimated by fitting a Generalized Linear Model (GLM) to the
245 decadal SSW frequency estimates from each model. All three statistical methods are described in
246 detail in Appendix 1.

247 **3 Model simulation of SSWs during the Historical Period: Mean frequency and seasonal**
248 **distribution**

249   Prior to reporting changes in SSWs caused by increased CO_2 concentrations, it is
250 important to document the models' ability to simulate SSW events during the period of overlap
251 with re-analysis data: we do so by analyzing the historical simulations. Figure 1a shows the
252 average frequency of SSWs during the period 1958-2014 in JRA-55 reanalysis (horizontal
253 dashed line) and the corresponding value for the CMIP6 models (bars; the numerical values are
254 given in Table S1). In agreement with prior studies [e.g., Charlton-Perez et al. 2013;
255 Ayarzagüena et al. 2018] we find a large spread across the models in the mean frequency of
256 SSW over that period. This spread is likely due, in part, to the large internal variability of the
257 polar wintertime stratosphere; even with an identical climate model the frequency of SSWs can
258 vary greatly across different realizations, as demonstrated by Polvani et al. [2017].

259   Mindful of this large internal variability, it appears that only four of the models are
260 significantly different from JRA-55, at the 95% confidence level. Three of these are the models

261 with the lowest model tops (CESM2, CanESM5 and GFDL-CM4) that simulate fewer SSW
262 events than JRA-55 re-analysis. When comparing the seasonal distribution of SSW activity in
263 these models with JRA-55 (Fig. 2) it is clear that for two of them (GFDL-CM4 and CESM2), the
264 SSW activity is significantly shifted towards March, with few SSWs observed in December and
265 January. This is another common bias in low-top models [Charlton-Perez et al., 2013], and more
266 generally, in models with an overly strong polar vortex. It is also worth noting that the three low-
267 top models mentioned above are the only ones lacking a simulated Quasi-Biennial Oscillation
268 (QBO). The fourth model with an unrealistic SSW frequency (IPSL-CM6A-LR), in contrast,
269 simulates a very high number of SSWs, on average one per year during the historical period
270 (instead of one every other year). As detailed below, this model also stands out for its high
271 frequency of warmings in the piControl run. While we retain these four models in our analysis,
272 the simulated changes produced by these models should be treated with caution given these
273 biases.

274 Finally, considering the surprising occurrence of an SSW in the SH in 2002 [Krüger et al.
275 2005], we extended the analysis to that hemisphere. Not a single SSW event was identified in the
276 SH over the historical period in the models analyzed here. One may be tempted to claim that the
277 CMIP6 models are underestimating the stratospheric variability in the SH, as spontaneous SSWs
278 in the absence of stationary waves have been reported in simple models [Kushner and Polvani,
279 2005]. However, it remains to be demonstrated whether five or six decades of observations are
280 sufficient to make that claim.

281

4 Future changes in polar stratospheric variability

283 4.1 Future changes in sudden stratospheric warmings

284 Figure 1b displays the mean frequency of SSWs in both the piControl and abrupt4xCO$_2$
285 simulations (numerical values in Table S2). As discussed in Section 2, all SSWs identified in the
286 entire abrupt4xCO$_2$ simulation have been considered. We stress, however, that the main results
287 presented below do not change significantly if only the second 75 years of each abrupt4xCO2
288 simulation are used (not shown). Two different tests of the statistical significance of the changes
289 are conducted, providing a consistent indication of the statistical significance of changes,
290 although the precise p-values vary due to difference in the underlying assumptions.

291 Of the 12 models in our study, four models indicate a statistically significant *decrease* in
292 SSW frequency, while four indicate a statistically significant *increase* in SSW frequency. Thus,
293 no consensus in the sign of the change exists in the CMIP6 models, in agreement with the
294 diversity of claims reported in the earlier literature. The lack of a robust change across the
295 models is not due to a lack of sensitivity of SSW frequency to increasing CO$_2$: in fact, 8 of the 12
296 models indicate significant changes. Rather, the CMIP6 models suggest that there is a great deal
297 of uncertainty in the sign of the change, which varies between a near doubling in the frequency
298 of SSWs in some models, to a near halving in others. These divergent responses of the models
299 may now be clearer in the CMIP6, where we can consider a stronger forcing (4xCO$_2$) and have
300 access to longer records of daily data, compared to previous studies.

301 We also note that the lack of consensus in the CMIP6 models agrees with the recent study
302 of Ayarzagüena et al. [2018], who analyzed the chemistry climate model projections of the
303 CCMI models, which were forced with RCP6.0 scenario. While reporting a general tendency

304 towards an increased frequency of SSWs by the end of the current century, they also emphasized
305 that most changes were *not* statistically significant.

306   We do not attempt to further analyze the causes of differences in the model responses
307 here, other than to note that within our set of models, one of the models indicating a significant
308 reduction of SSW frequency (CanESM5) and one of the models indicating a significant increase
309 of SSW frequency (CESM2) have anomalously low SSW frequency and (in the case of CESM2)
310 a biased seasonal distribution of SSW in the historical simulations (Fig. 2). Additionally, two
311 models which show significant decreases in SSW frequency (HadGEM3-GC31-LL and IPSL-
312 CM6A-LR) have the highest frequency of SSW events in the piControl and historical
313 simulations. The IPSL-CM6A-LR has a significant bias in SSW frequency and presents some
314 strong biases in the representation of QBO in the abrupt4xCO$_2$ simulation. Nevertheless, even if
315 we did not consider the four models with biases in the representation of SSWs in the historical
316 period (CanESM5, CESM2, IPSL-CM6A-LR and GFDL-CM4), the main conclusion on the
317 uncertainty in the sign of SSW changes would remain the same.

318   We also briefly examined the relationship between the change in SSW frequency and
319 possible predictors of the change, including the frequency of SSWs in the piControl and
320 historical simulations and the Effective Climate Sensitivity (ECS, Gregory et al. [2004]) (Fig. 3).
321 Recall that ECS gives a measure of the equilibrium change of the global surface temperature
322 after a doubling of CO$_2$. As can be seen from Fig. 3, models that have a larger frequency of
323 SSWs in the piControl run and models that have a larger ECS seem to produce large reductions
324 in SSW frequency under large CO$_2$ forcing. A notable outlier from the main relationship here is
325 the GISS-E2.2AP model but note that shorter simulations are available for this model than for
326 others in the ensemble which also means that the uncertainty on the estimate of the piControl
327 SSW frequency for this model is large.

328   Excluding GISS-E2.2AP, the correlation between SSW frequency changes and ECS is -
329 0.52 with a probability value of obtaining results at least as extreme as the computed correlation
330 (p-value) of 0.12. However, with GISS-E2.2AP included in the ensemble, the correlation drops
331 to -0.33 and is not significant. The correlation between piControl frequency and SSW frequency
332 changes is -0.50 with a p-value of 0.10 with all models included. Further analysis of a larger
333 ensemble would be required to determine the robustness of these relationships.

334   Although not addressed in the literature, a relationship between ECS and SSW frequency
335 changes might be possible given some previous results connected to this topic. Shepherd and
336 McLandress, [2011] and Grise and Polvani [2016] documented a link between the strengthening
337 of the sub-tropical jet and stratospheric wave driving. Moreover, Li et al. [2007] have argued that
338 the subtropical jet, and tropospheric state in general, might control the upward planetary wave
339 propagation. In this sense, the meridional gradient of the upper tropospheric temperature in the
340 piControl simulation (computed as in Harvey et al [2014]) was found to be linked to the SSW
341 frequency changes under high CO$_2$ concentrations. The correlation between both variables is -
342 0.61 (p-value 0.04). Thus, a model bias in the tropospheric state affects the stratospheric
343 response to increasing CO$_2$, probably due to its effects on wave propagation. In addition, an
344 intriguing examination of the relationship between changes in the tropospheric state and SSW
345 frequency is shown in the bottom row of panels of Fig. 3. Again, GISS-E2.2AP is an outlier in
346 Fig, 3(c)-(f). Excluding, GISS-E2.2AP, there is a significant correlation between changes in
347 SSWs and changes in the polar lower tropospheric temperature (-0.89, p-value < 0.01) and the
348 lower tropospheric temperature gradient (0.79, p-value < 0.01). In contrast, correlations between

349 the upper tropospheric temperature changes and SSW frequency are generally smaller, with the
350 highest correlation between the tropical upper tropospheric temperature change and SSW
351 frequency change (-0.62, p-value 0.06). With GISS-E2.2AP included, the lower tropospheric
352 correlations are reduced but have p-values smaller than 0.05, while the correlation with tropical
353 upper tropospheric temperature does not (-0.49, p-value 0.12).

354   Of these three critical temperature parameters, temperatures in the upper tropical
355 troposphere and polar lower troposphere are correlated with the ECS. As more dynamical
356 diagnostics suitable for detailed examination of the wave generation and propagation in the
357 models become available, it will be very interesting to try to understand the robustness and
358 causes of these relationships. We also note the interesting recent result of Zelinka et al. (2020),
359 that models with higher climate sensitivity in CMIP6 generally have reduced low cloud cover in
360 mid-latitude and polar regions.

361   To further examine the changes in SSW frequency under $4xCO_2$ forcing, we have
362 analyzed the entire distribution of daily $u_{60N10hPa}$ in December-January-February in the piControl
363 and abrupt$4xCO_2$ simulations (Fig. 4). The four models with a significant decrease in SSWs
364 frequency in Fig. 1b (HadGEM3-GC31-LL, CanESM5, IPSL-CM6A-LR, and INM-CM5-0) are
365 also those that show the largest shift of the $u_{60N10hPa}$ distribution towards stronger vortex speeds
366 in the abrupt$4xCO_2$ experiment. Interestingly, the opposite does not always apply to models with
367 a significant increase in SSWs. The models with the largest changes in SSW frequency,
368 MIROC6 and CESM2-WACCM show small changes to either the median or standard deviation
369 of the $u_{60N10hPa}$ (Table S3). This would agree with the results of Taguchi [2017] who pointed out
370 SSW frequency does not only correlate with vortex strength but also wave activity.

371   A similar analysis was repeated for the zonal-mean zonal wind at 10hPa averaged
372 between 70° and 80°N (not shown). That latitude band was found by Manzini et al. [2014] to
373 display significant future changes in wind in most models, unlike the 60°N latitude where no
374 robust future changes were found in CMIP5 models because the opposed effects of subtropical
375 jet and stratospheric polar vortex changes might combine at that latitude. However, in our case,
376 the main conclusions remain the same. Those models that show a shift of the $u_{60N10hPa}$
377 distribution towards stronger vortex speeds under 4xCO2 forcing also display a sharper peak of
378 high values u at 70-80°N suggesting lower variability in that region, consistent with a stronger
379 and larger vortex.

380   We have also examined potential changes in SSW seasonality. However, despite the
381 already mentioned changes detected in SSW frequency in some models, the drastic increase in
382 CO_2 concentrations does not appear to substantially affect the seasonal distribution of SSWs (not
383 shown).

384   Finally, motivated by the recent occurrence of a minor but highly publicized SSW event
385 in the SH in September 2019 [Hendon et al., 2019], together with the occurrence of a major SSW
386 in September 2002, we also examined the CMIP6 models to determine the extent to which the
387 likelihood of similar events might change under the extreme climate forcing in the abrupt$4xCO_2$
388 runs. Only one of our twelve models (MRI-ESM2-0) simulates an SSW in both the piControl and
389 the abrupt$4xCO_2$ simulations. Thus, these runs provide no evidence for the claim of possible
390 trends in the frequency of SSWs in the SH that would be caused by increased CO_2
391 concentrations.

392   4.2 Trends in SSW frequency and time of emergence

393   For the model integrations which show a statistically significant increase or decrease in
394   SSW frequency between the piControl and the abrupt4xCO$_2$ runs, it is useful to consider when
395   and whether the trend in SSW frequency might be detected in a simulation with continuously
396   increasing CO$_2$ forcing. A useful way to frame climate trends is in terms of the time of
397   emergence of the signal from the unforced climate noise [Hawkins and Sutton, 2012]. This
398   question is examined by studying the occurrence of SSWs in the 1pctCO$_2$ runs, an idealized
399   scenario.

400   Trend estimates for each model are shown in the Fig. 5a (numerical values in Table S4).
401   Results reveal that there are six models (light gray bars) for which the null hypothesis of no trend
402   in SSW frequency can be rejected, but consistent with the results of the previous section, the sign
403   of this trend is not robust across models. While CanESM5 and HadGEM3-GC31-LL show a
404   significant decrease, CNRM-ESM2-1, CESM2-WACCM, GFDL-CM4 and MRI-ESM2-0 show
405   a significant increase. Recall that for the abrupt4xCO$_2$ runs (Fig. 1b), CNRM-ESM2-1 and
406   CESM2-WACCM also indicated a statistically significant increase in SSW frequency compared
407   to the piControl runs, while GFDL-CM4 and MRI-ESM2-0 did not (although they did indicate
408   an increased frequency). CanESM5 and HadGEM3-GC31-LL both showed a statistically
409   significant decrease.

410   One can also estimate a time of emergence of the trend by comparing the trend in the
411   1pctCO2 runs with the natural variability in SSW frequency from the piControl run (see
412   Appendix for details in the procedure). For the models with a significant trend, the decade of
413   emergence is shown in Fig. 5b. There is a wide spread in the projected time of emergence for the
414   models with a significant trend, varying from the 5th decade to 14th decade. This result reflects
415   both the variation in the trend across the models and the spread in the estimated variability in
416   SSW frequency (the noise) in the piControl simulations. Since the time of CO$_2$ doubling occurs
417   between the 6th and 7th decade in the 1pctCO$_2$ run and approximately by 2060-70 in the RCP8.5
418   scenario [Meinshausen et al., 2017], these results indicate that the emergence of a detectable
419   change in SSW frequency is extremely unlikely prior to the end of the 21st century.

420

421   **5 Future changes in the seasonal cycle of the polar stratosphere**

422   Since, according to linear theory, the vertical propagation of stationary Rossby waves is
423   restricted to periods with westerly winds, stratospheric variability is largely confined to the
424   winter season [e.g., Charney and Drazin 1961]. When considering how stratospheric variability
425   might change in future climates it is therefore also important to consider the extent to which the
426   timing and length of the winter season in the stratosphere might also change.

427   Figure 6a and b show the distribution of dates of formation and final breakdown of the
428   boreal stratospheric polar vortex, respectively, in the piControl, historical and abrupt4xCO$_2$
429   CMIP6 simulations. In these plots the first years of the abrupt4xCO$_2$ simulations (75 or 300
430   years) have been omitted similar to the procedure followed to calculate the climatology.
431   Nevertheless, conclusions do not change when considering the whole data record for
432   abrupt4xCO$_2$ runs.

433   First, let us consider the historical model simulations, and contrast them to the reanalysis.
434   Over the period 1958-2014, the polar vortex forms earlier in all models than it does in the

435 reanalysis, with the exception of IPSL-CM6A-LR. In contrast, the SFW date is well reproduced
436 by models. The latter implies an improvement with respect to previous generations of climate
437 models, such as those contributing to CCMVal and CMIP5, which simulated a delayed SFW
438 [Butchart et al., 2011; Kelleher et al., 2019]. CMIP6 models are also good at simulating the
439 different range of interannual variability in the dates of vortex formation and SFW, the latter
440 being considerably larger than the former.

441   Second, we consider the changes caused by increased CO_2, both for the formation and the
442 final breakdown of the boreal polar vortex: these display robust changes across models. The
443 polar vortex forms earlier and persists for longer in the abrupt4xCO_2 scenario than in the
444 piControl runs (Fig. 6a and b). This signal is particularly clear, and is significant in most of the
445 models in the case of the vortex formation. Although half of the models do not show a significant
446 change, there is a clear consensus in the sign of the SFW change across these models.

447   Interestingly, the models with the largest delay of SFW in the abrupt4xCO_2 simulation
448 (CanESM5, HadGEM3.GC31-LL and IPSL-CM6A-LR) are also those with the largest reduction
449 in the frequency of SSWS. This indicates that the long persistence of the vortex is related to a
450 stronger and colder vortex during the extended winter, rather than to the effect of SSWs on the
451 SFWs timing suggested by Hu et al. [2014]. The year-round radiative effect of CO_2, which is
452 associated with a warming tropical upper troposphere and a cooling stratosphere, increases the
453 upper-level meridional temperature gradient and leads to a longer-lived polar vortex. Indeed, a
454 positive and significant correlation (~0.65) has been found between the degree of change in the
455 duration of the polar vortex per winter and the warming of the tropical upper troposphere in
456 models between piControl and abrupt4xCO_2 simulations. Why this influence occurs primarily in
457 early fall and spring may be tied to the seasonality of the upper tropospheric warming [Harvey et
458 al. 2014], and the dynamical driving of the polar vortex. Indeed, the wave activity is typically
459 weaker during the transition season (particularly in Autumn) than in mid-winter [Kodera et al.,
460 2003], and so, the radiative effect of increased CO_2 on the stratosphere dominates. In sum,
461 models predict an increase of around 30 days of westerly winds in the abrupt4xCO_2 simulations,
462 a substantial increase in the time of the year over which stratospheric variability is active and can
463 couple with the troposphere.

464   A similar analysis has been performed for the SH. Because planetary wave activity is
465 much weaker in the SH than in the NH [Andrews et al. 1987], radiative CO_2 forcing dominates
466 the SH polar vortex response to increasing CO_2 concentrations and so, causes a robust
467 strengthening. In many models the extreme CO_2 concentrations prevent the polar vortex from
468 disappearing at all during austral summer, leading to perpetual westerly conditions in the
469 stratosphere, so we do not show the results for the abrupt4xCO_2 simulation. The distribution of
470 SFW dates for piControl and historical simulations are displayed in Figure 6c. Unlike in the NH,
471 the distribution of SFWs in the SH already shifts towards a later date in the historical period with
472 respect to the piControl conditions. Although the attribution of changes in the length of the
473 winter season to CO_2 is complicated, ozone depletion in austral spring over the historical period
474 might be responsible, based on previous literature [e.g. McLandress et al. 2010; Oberländer-
475 Hayn et al. 2015].

476

477   **6 Future changes in the surface impact of SSW events**

478   6.1 Surface response to SSW events

479   In addition to changes in SSW frequency, amplitude and seasonality, it is also
480   conceivable that the surface impact of SSW events might change as a consequence of increased
481   CO_2. While detailed quantitative description of the mechanism for coupling between SSW events
482   and surface remains elusive, there is now a large body of evidence quantifying the amplitude and
483   spatial structure of the surface pressure and temperature responses following SSW events [e.g.,
484   Baldwin and Dunkerton, 2001; Polvani et al. 2017; Butler et al. 2017]. A number of studies point
485   to the importance of eddy-jet feedbacks in determining this surface response [e.g.: Kushner and
486   Polvani, 2004; Song and Robinson, 2004; Garfinkel et al. 2013;]. It is therefore plausible that
487   together with changes in the position and variability of the extra-tropical jet caused by CO_2
488   increases, one might be able to detect changes in the surface response following SSW events.

489   To test this idea, we analyze first composite maps of anomalous surface temperature and
490   sea-level pressure (SLP) for the period 15-60 days after SSWs in the piControl simulation (Fig.
491   7). In nearly all models we obtain the typical SLP and surface temperature patterns following
492   SSWs that are also detected in reanalysis (although CO_2 forcing is different), i.e., negative
493   Northern Annular Mode pattern (particularly over the pole), and Eurasian cooling and
494   Northeastern American warming. None of the models produce a positive SLP anomaly in the
495   Pacific basin that can be found in the JRA55 composite though. Despite the relatively structural
496   similarities across models, the amplitude of the response can vary by a factor of two or three
497   between them. The amplitudes of SLP anomalies in five of the eleven models (CESM2-
498   WACCM, GFDL-CM4, HadGEM3-GC31-LL, IPSL-CM6A-LR, and MIROC6) are too weak.
499   Moreover, even the rest of the models that do a reasonable job of the polar cap SLP signal
500   significantly underestimate the surface temperature response over the Labrador Sea and to the
501   east of Greenland. This is consistent with Hitchcock and Simpson [2014] that argued the near-
502   surface temperature response to SSW was underestimated in specific regions in CMIP5 models.
503   The amplitude of the signal in the troposphere does not correlate with the SSW frequency. It is
504   also not a problem of model biases in the simulation of SSWs mentioned in Section 3 either. The
505   large SSW sample size from the piControl simulations means that the estimates of surface impact
506   are very robust.

507   Secondly, we compare the SLP pattern after SSWs in the abrupt4xCO2 and piControl
508   simulations (Fig. 8, differences in SLP between both runs are shown in shading). The overall
509   SSW signal in SLP appears unchanged between the piControl and abrupt4xCO2 simulations,
510   except in three models (CESM2, HadGEM3-GC31-LL and IPSL-CM6A-LR) that produce a
511   significantly stronger Northern Annular Mode-like response. However, in the Pacific basin there
512   are some indications about a potential more general change due to a higher CO_2 loading. Indeed,
513   six of the ten models exhibit a statistically stronger negative SLP anomaly in that area under
514   abrupt4xCO_2 forcing than in the piControl runs. This could be related to some changes in the
515   tropospheric precursors of SSWs because these anomalies have been identified as the remainder
516   of the deepening of the Aleutian low preceding SSWs in observations [Charlton and Polvani,
517   2007; Ayarzagüena et al., 2019]. Nevertheless, more work is required to understand all the
518   details.

519   Please note that when restricting the analysis to the years 75-150 in IPSL-CM6A-LR,
520 similar results are found but with a reduction in the areas with statistical significance due to a
521 lower number of events considered.

522

523   6.2 Polar-night Jet Oscillation events

524   In this subsection we focus on specific events (PJO events) that are closely related to
525 SSWs and the stratosphere-troposphere coupling [Hitchcock et al. 2013]. As indicated in the
526 Introduction section, their strong and persistent tropospheric response explains the interest in
527 investigating possible changes in the occurrence of these events for increasing CO_2
528 concentrations.

529   First, examining the surface response to PJOs in the piControl experiment (Fig. S1)
530 confirms that these events in models have a stronger signal in the troposphere than all SSWs too.
531 In JRA-55 roughly half of all SSWs are associated with a PJO event (PJO SSW) (solid line in
532 Fig. 9). Six models include the JRA-55 value of the ratio of PJO SSW events in their confidence
533 interval in the piControl simulations (MRI-ESM2-0, UKESM1-0-LL, CanESM5, HadGEM3-
534 GC31-LL, INM-CM5-0 and GFDL-CM4). The other models underestimate this fraction.
535 However, we do not find a clear relationship between this fraction and the amplitude of SLP
536 pattern following SSWs. For instance, HadGEM3-GC31-LL and GFDL-CM4 simulate a very
537 weak SLP pattern (Fig. 7), but the ratio of PJO SSWs is close to observations or even larger.

538   In the future, similar to changes in SSW frequency, there is no robust response of PJO
539 SSWs across models to increasing CO_2 (Fig. 9). Roughly half of the models show a decrease and
540 half of them an increase in PJO SSW events between the piControl and abrupt4xCO2
541 simulations. More interestingly, two of the three models with a stronger Northern Annular
542 Mode-response to SSWs in the abrupt4xCO_2 run (IPSL-CM6A-LR and HadGEM3-GC31-LL)
543 display an increase in this subset of SSWs too. The other one (CESM2) does not show a
544 significant change in the fraction of SSWs that are PJOs. Nevertheless, given the low number of
545 models, it is difficult to make a direct link between changes in the number of PJO SSWs and
546 stronger SSW coupling to the surface under increased CO_2 loading.

547

548 **7 Conclusions**

549   SSWs are the primary dynamical event in the wintertime polar stratosphere and have
550 clear impacts on the tropospheric circulation on sub-seasonal to seasonal timescales. This study
551 takes advantage of the new sets of simulations available through the DynVarMIP sub-project of
552 CMIP6 to revisit a number of questions about how SSW events and the stratospheric seasonal
553 cycle might respond to quadrupled CO_2 concentrations. In comparison with previous rounds of
554 CMIP and comparisons made as part of the CCMVal and CCMI projects, the new simulations
555 provide significant advances in our ability to study SSWs. In particular, the long piControl runs
556 and the availability of daily data of abrupt4xCO_2 simulations from a large number of high-top
557 models is unprecedented.

558   From our analysis of the twelve models for which sufficient daily time resolution
559 stratospheric data was available, these conclusions can be drawn about the impact of extreme
560 CO_2 concentrations on SSW events:

561 • There is no consensus among models on the sign of changes in SSW frequency to
562   increase in CO_2 forcing.

563 • It is, however, possible to say with confidence that many models predict that SSW
564   frequency is sensitive to increase in CO_2 forcing.

565 • There is no change to the impact of SSW events in the N. Atlantic between the
566   abrupt4xCO_2 and piControl simulations. In the N. Pacific, there is some indication that
567   under large CO_2 forcing there will be a larger mean response to SSW events.

568 • With the exception of MRI-ESM-2-0, predicted trends in SSW frequency are small
569   relative to natural variability (as characterized by the piControl simulations of each
570   model). This is not to say that SSW changes are themselves small (three models predict
571   frequency changes of more than a factor of two compared to piControl conditions) but
572   more a reflection of the large, natural decadal variability in SSW occurrence. As such,
573   changes in SSW frequency are unlikely to be observed until the end of the 21st century.

574 • Robust changes to the seasonal cycle in the stratosphere are predicted by all models. The
575   stratospheric polar vortex is likely to form earlier and decay later in the future. This
576   extends the season in which the stratosphere can actively couple to the troposphere and
577   influence surface weather.

578 • There is no evidence of an increased likelihood of major SSWs in the SH in the future.

579   These results underscore the conclusions of a number of previous studies of SSW events
580 and also motivate the need for more detailed understanding of the stratospheric momentum
581 budget in models as advocated by, for example, Wu et al. [2019], which is now possible with the
582 simulations available through DynVarMIP. Similarly, developing an understanding of how both
583 model formulation and resolution and ECS might influence dynamical sensitivity in the
584 stratosphere remains an important but unsolved challenge for the stratospheric dynamics
585 community.

Acknowledgments, Samples, and Data

609

610   **Appendix 1: Statistical framework**

611   <u>Statistical methodology for comparing SSW frequency</u>

612   *Parametric method*

613   To compare the frequency of SSW events in two models or between a model and
614   observations, it can be assumed that each data sample is a Poisson process with an annual rate λ_i.
615   The difference between the intensity of the two processes Δ_λ is given in equation (1)

616

$$\Delta_\lambda = \frac{(\lambda_0 - \lambda_1)}{\sqrt{\frac{\lambda_0}{N_0} + \frac{\lambda_1}{N_1}}} \qquad (1)$$

617   This can be modeled with a normal distribution providing the frequency of observed events is
618   greater than 30 [Charlton et al. 2007]. This approach has been widely used in the literature.

619   An alternative approach that compares the ratio of the rate of the two Poisson processes
620   has been studied by Gu et al. [2008].

621 $$H_0: \lambda_0/\lambda_1 = 1 \qquad against \qquad H_A: \lambda_0/\lambda_1 \neq 1 \qquad (2)$$

622   Gu et al. [2008] suggest that a conservative test statistic with high power is the one
623   suggested by Huffman [1984] (here X_i is the number of SSWs in each dataset and $\rho = t_0/t_1$ the
624   ratio of the length of observation of the two processes):

625 $$W(X_0, X_1) = \frac{2[\sqrt{X_0 + 3/8} - \sqrt{\rho(X_1 + 3/8)}]}{\sqrt{1+\rho}} \qquad (3)$$

626   The p-value for this statistic is estimated as in equation (4), where ϕ is the cumulative
627   distribution function of the standard normal and the observed value of the test statistic
628     $W(X_0, X_1) = w(x_0, x_1)$:

629 $$p = 1 - 2 * \Phi(w_j(x_0, x_1)) \qquad (4)$$

630   This is the parametric test statistic used to compare SSW frequency. In addition to
631   calculating the p-value of any test statistic it is also useful, a priori, to estimate the statistical
632   power of any testing framework. Tests with high statistical power minimize the likelihood of
633   Type-II errors (i.e. that the null hypothesis is not rejected when it is, indeed, false). For the test
634   statistic described above, we estimated the statistical power for a comparison with observations
635   of 60 winters with an SSW frequency of 0.6. Assuming a p-value of 0.05, the statistical power of
636   the test is high (above 0.8) for model integrations of more than 100 winters (the null hypothesis
637   will be rejected with a probability above 0.8) which is the case for all comparisons in this study
638   apart from the comparison between the historical simulations and the JRA-55 re-analysis. In this
639   later case, the power of the test is low only for cases in which the observed and modelled SSW

640 frequency is very similar (i.e. for model SSW frequencies of 0.2 and 1 SSW per year, the power
641 is greater than 0.8).

642

643 *Bootstrapping method*

644 　　　As an alternative to the parametric test, we can also construct a bootstrapping test as
645 outlined by Boos [2003]. We assume that there are two sets of independent samples of the
646 number of SSW events in each season $\{X_1,…,X_m\}$ and $\{Y_1,…,Y_n\}$. To determine the confidence
647 interval for the difference of mean frequency of the two sets $\mu_x - \mu_y$ two samples (of equal size
648 to the original samples) are drawn from the pooled observation set $\{X_1,…,X_m, Y_1,…,Y_n\}$, with
649 replacement. The p-value of the true observation is calculated as the number of bootstrap
650 samples with an absolute difference greater than the true value. In all cases, 10,000 boostrap
651 sample are drawn.

652 　　　This bootstrapping technique was also applied to determine the confidence intervals on
653 the seasonal distribution of SSW frequency. We choose to perform the bootstrapping on
654 individual winters over a block bootstrapping approach to increase the sample size available for
655 models that have a limited length of piControl simulation available. We have, therefore, assumed
656 that there is no autocorrelation from one winter to the next, but comparison with a block-
657 bootstrapping approach for the models that have long piControl simulations produced similar
658 uncertainty ranges (not shown), indicating that this assumption is reasonable. For Figure 2, the
659 uncertainty range is derived from the piControl simulation. Since there are 57 years in the JRA55
660 record, we resample 57 years from the piControl simulation, with replacement and recalculate
661 the SSW distribution, normalized by the number of SSWs in that sample. This is repeated 1000
662 times and the uncertainty range shows the 2.5th to 97.5th percentile range of these 1000 samples
663 (95% confidence interval) i.e., this is the uncertainty range from the model with an equivalent
664 number of years to that of the observations.

665

666 <u>Trend in SSW frequency and Time of Emergence</u>

667 　　　Analogously to the method of Hawkins and Sutton [2012] the time of emergence of a
668 'signal' in the frequency of SSW events is estimated by comparing the size of the trend in SSW
669 frequency in the 1pctCO2 simulations with the 'noise' determined from the piControl simulation
670 of the same model.

671 　　　To calculate the signal term in each integration, a Generalized Linear Model fit to the
672 data with a logarithmic link function, implemented in R is used. Trend estimates for decadal
673 SSW frequency in the 1pctCO2 simulations. Modification to the method following
674 (https://stats.idre.ucla.edu/r/dae/poisson-regression/) to account for cases with mild violation of
675 the Poisson distribution in the models is included. The resulting regression equation is of the
676 form:

$$F_{ssw}(t) = e^{\beta_0 + \beta_t t} \quad (5)$$

678 Trend terms are expressed as a fractional multiplier of the count per decade. Due to the low mean
679 annual frequency of SSW events, the noise on annual mean frequency estimates is large,
680 therefore when estimating trends in SSW frequency and time of emergence we consider the

681 decadal mean SSW frequency. This means that time of emergence calculations are limited to the
682 decade of emergence.

References

684 Andrews, D. G., J. R. Holton, and C. B. Leovy, (1987) Middle Atmosphere Dynamics,
685   Academic, Orlando, Fla.

686 Ayarzagüena B. and E. Serrano (2009) Monthly characterization of the tropospheric circulation
687   over the Euro-Atlantic area in relation with the timing of stratospheric final warmings,
688   J.Clim. 22, 6313-6324.

689 Ayarzagüena, B., Langematz, U., Meul, S., Oberländer, S., Abalichin, J., and Kubin, A. (2013)
690   The role of climate change and ozone recovery for the future timing of major
691   stratospheric warmings, Geophys. Res. Lett., 40, 2460–2465.

692 Ayarzagüena, B., Polvani, L.M., Langematz, U., Akiyoshi, H., Bekki, S., Butchart, N., Dameris,
693   M., Deushi, M., Hardiman, S.C., Jöckel, P. and Klekociuk, A., (2018) No robust evidence
694   of future changes in major stratospheric sudden warmings: a multi-model assessment
695   from CCMI. Atmos. Chem. Phys., 18(15), 11277-11287, https://doi.org/10.5194/acp-18-
696   11277-2018.

697 Ayarzagüena, B., Palmeiro, F. M., Barriopedro, D., Calvo, N., Langematz, U. and K. Shibata,
698   (2019) On the representation of major stratospheric warmings in reanalyses, Atmos.
699   Chem. Phys., 19, 9469-9484, https://doi.org/10.5194/acp-19-9469-2019.

700 Baldwin, M. P and T. J. Dunkerton (2001) Stratospheric harbingers of anomalous weather
701   regimes, Science, 294, 581-584.

702 Bell, C. J., Gray, L. J., and Kettleborough, J. (2010) Changes in Northern Hemisphere
703   stratospheric variability under increased CO2 concentrations, Q. J. Roy. Meteor. Soc.,
704   136, 1181–1190.

705 Black, R. X., B. A. McDaniel, and W. A. Robinson (2006) Stratosphere-troposphere coupling
706   during spring onset, J. Clim. 19, 4891-49001.

707 Black, R. X., and B. A. McDaniel (2007) Interannual variability in the Southern Hemisphere
708   circulation organized by stratospheric final warming events, J. Atmos. Sci. 64, 2968-
709   2974.

710 Boos D. D. (2003) Introduction to the Boostrap World. Statistical Science, 18 (2), 168-174.

711 Boucher, O., S. Denvil, A. Caubel, M.A. Foujols (2018). IPSL IPSL-CM6A-LR model output
712   prepared for CMIP6 CMIP. Version 20190722.Earth System Grid Federation.
713   https://doi.org/10.22033/ESGF/CMIP6.1534

714 Butchart, N., J. Austin, J. R. Knight, A. A. Scaife, and M. L. Gallani (2000) The response of the
715   stratospheric climate to projected changes in the concentration of well-mixed greenhouse
716   gases from 1992 to 2051, J Clim, 13, 2141-2159.

717 Butchart, N., and Coauthors (2011) Multimodel climate and variability of the stratosphere. J.
718   Geophys. Res., 116 (D5), doi:10.1029/2010jd014995.

719 Butler, A.H., J. Sjoberg, D. Seidel., and K.H. Rosenlof (2017) A sudden stratospheric warming
720   compendium, Earth Syst. Sci. Data, 9, 63-76.

721 Butler, A. H., and E. P. Gerber (2018), Optimizing the definition of a sudden stratospheric
722   warming, J. Clim., 31, 2337-2344, doi: 10.1175/JCLI-D-17-0648.1

723 Charlton, A. J., and L. M. Polvani (2007), A new look at stratospheric sudden warmings. Part I:
724   Climatology and modeling benchmarks, J. Clim., 20, 449–469.

725 Charlton, A. J, L. M. Polvani, J. Perlwitz, F. Sassi, E. Manzini, K. Shibata, S. Pawson, J. E.
726   Nielsen, and D. Rind (2007), A New Look at Stratospheric Sudden Warmings. Part II:
727   Evaluation of Numerical Model Simulations, J. Clim., 20, 470–88.

728 Charlton-Pérez, A. J., L. M. Polvani, J. Austin, and F. Li (2008) The frequency and dynamics of
729   stratospheric sudden warmings in the 21st century, J. Geophys. Res. 113, D16116, doi:
730   10.1029/2007JD009571.

731 Charlton-Pérez, A. J. and coauthors (2013) On the lack of stratospheric dynamical variability in
732   low-top versions of the CMIP5 models, J. Geophys. Res. Atmos., 118, 2494–2505,
733   doi:10.1002/jgrd.50125.

734 Charney, J. G., and P. G. Drazin (1961), Propagation of planetary-scale disturbances from the
735   lower into the upper atmosphere, J. Geophys. Res., 66, 83–109,
736   doi:10.1029/JZ066i001p00083.

737 Danabasoglu, Gokhan (2019). NCAR CESM2-WACCM model output prepared for CMIP6
738   CMIP. Version 20190730. Earth System Grid Federation.
739   https://doi.org/10.22033/ESGF/CMIP6.10024

740 Danabasoglu, G., D. Lawrence, K. Lindsay, W. Lipscomb, G. Strand (2019). NCAR CESM2
741   model output prepared for CMIP6 CMIP historical. Version 20190730.Earth System Grid
742   Federation. https://doi.org/10.22033/ESGF/CMIP6.7627.

743 Danabasoglu, G. and coauthors (2020) The Community Earth System Model version2 (CESM2),
744   J. Adv. Model. Earth System, doi: 10.1029/2019MS001916.

745 Domeisen D.I.V, and coauthors (2019) The role of the stratosphere in subseasonal to seasonal
746   prediction. Part I: Predictability of the stratosphere, J. Geophys. Res. Atmos.,
747   https://doi.org/10.1029/2019JD030920.

748 Eyring, V., Bony, S., G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor
749   (2016), Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6)
750   experimental design and organization, Geosci. Model Dev., 9, 1937-1958,
751   doi:10.5194/gmd-9-1937-2016.

752 Fels, S. B., J. D. Mahlman, M. D. Schwarzkopf and R. W. Sinclair (1980) Stratospheric
753   sensitivity to perturbations in ozone and carbon dioxide: radiative and dynamical
754   response, J. Atmos. Sci, 37, 2265-2297.

755 Garfinkel, C. I., D. W. Waugh, and E. P. Gerber (2013) The effect of tropospheric jet latitude on
756   coupling between the stratospheric polar vortex and the troposphere, J. Atmos. Sci., 26,
757   2077-2095.

758 Gerber, E. and E. Manzini (2016) The dynamics and variability model intercomparison project
759   (DynVarMIP) for CMIP6: assessing the stratosphere-troposphere system, Geosci. Model
760   Dev., 9, 3413-3425.

761 Gettelman, A. and coauthors (2019) The Whole Atmosphere Community Climate Model Version
762   6 (WACCM6), J. Geophys. Res. Atmos., https://doi.org/10.1029/2019JD030943

763 Gregory J. M., W. J. Ingram, M. A. Palmer, G. S. Jones, P. A. Stott, R. B. Thorpe, J. A. Lowe, T.
764   C. Johns and K. D. Williams (2004) A new method for diagnosing radiative forcing and
765   climate sensitivity. Geophys. Res. Lett., 31, L03205, doi: 10.1029/2003GL018747.

766 Grise, K.M. and Polvani, L.M., (2017). Understanding the time scales of the tropospheric
767   circulation response to abrupt CO2 forcing in the southern hemisphere: seasonality and
768   the role of the stratosphere. J. Clim., 30(21), 8497-8515, doi: 10.1175/JCLI-D-16-0849.1

769 Gu, K., H. K. Tony Ng, M. L. Tang, and W. R Schucany (2008) Testing the Ratio of Two
770   Poisson Rates. Biometrical Journal: Journal of Mathematical Methods in Biosciences 50,
771   283–98.

772 Guo, H. and coauthors (2018) NOAA-GFDL GFDL-CM4 model output. Version 20190718.
773   Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.1402

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

777 Harvey, B. J., L. C. Shaffrey, and T. J. Woollings (2014), Equator-to-pole temperature
778   differences and the extra-tropical storm track responses of the CMIP5 climate models,
779   Clim. Dyn., 43, 1171-1182, doi:10.1007/s00382-013-1883-9.

780 Hawkins, E. and R. Sutton (2012) Time of Emergence of Climate Signals, Geophys. Res. Lett.,
781   39, L01702, doi:10.1029/2011GL050087.

782 Held, I.M., and coauthors (2019) Structure and Performance of GFDL's CM4.0 Climate Model,
783   J. Adv. Model. Earth System, https://doi.org/10.1029/2019MS001829.

784 Hendon H., D. W. J. Thompson, E.-P. Lim, A. H. Butler, P. A. Newman, L. Coy, A. Scaife, I.
785   Polichtchouk, R. S. Gerrard, T. G. Shepherd and H. Nakamura (2019), Nature, 573, 496,
786   doi: 10.1038/d41586-019-02858-0

787 Hitchcock, P., T. G. Shepherd, and G. L. Manney (2013) Statistical characterization of Arctic
788   polar-night jet oscillation events, J. Clim., 26, 2096-2116.

789 Hitchcock, P., and I. R. Simpson (2014) The downward influence of stratospheric sudden
790   warmings, J. Atmos. Sci., 71, 3856-3876, doi: 10.1175/JAS-D-

791 Hu, J., R. Ren and H. Xu (2014) Occurrence of winter stratospheric sudden warming events and
792   the seasonal timing of spring stratospheric final warming, J. Clim., 71, 2319-2334.

793 Huffman, M. (1984), An Improved Approximate Two-Sample Poisson Test. J. Roy Stat. Soc.:
794   Series C (Applied Statistics) 33 (2), 224–26.

795 Karpechko, A. Y. and E. Manzini (2012) Stratospheric influence on tropospheric climate change
796   in the Northern Hemisphere. J. Geophys. Res. 117, D05133, DOI
797   10.1029/2011JD017036.

798 Kelleher, M., B. Ayarzagüena and J. Screen (2019): Inter-seasonal connections between the
799   timing of the stratospheric final warming and Arctic sea ice, J. Climate,
800   https://doi.org/10.1175/JCLI-D-19-0064.1.

Kidston, J., A. A. Scaife, S. C. Hardiman, D. M. Mitchell, N. Butchart, M. P. Baldwin and L- J. Gray (2015) Stratospheric influence on tropospheric jet streams, storm tracks and surface weather, Nat. Geos., 8, 433-440.

Kim, J., Son, S.-W., Gerber, E. P., and Park, H.-S. (2017) Defining sudden stratospheric warming in climate models: Accounting for biases in model climatologies, J. Clim., 30, 5529–5546.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K., and Takahashi, K. (2015): The JRA-55 reanalysis: General specifications and basic characteristics, J. Meteor. Soc. Jpn., 93, 5–48.

Kodera, K., K. Matthes, K. Shibata, U. Langematz, and Y. Kuroda (2003), Solar impact on the lower mesospheric subtropical jet: A comparative study with general circulation model simulations, Geophys. Res. Lett., 30(6), 1315, doi:10.1029/ 2002GL016124.

Krüger, K., B. Naujokat, and K. Labitzke (2005) The Unusual Midwinter Warming in the Southern Hemisphere Stratosphere 2002: A Comparison to Northern Hemisphere Phenomena, J. Clim., 62, 603-613.

Kuhlbrodt, T., C. G. Jones, A. Sellar, D. Storkey, E. Blockley, M. Stringer, et al. (2018). The low-resolution version of HadGEM3 GC3.1: Development and evaluation for global climate. J. Adv. Model. Earth Systems, 10, 2865–2888, https://doi.org/10.1029/2018MS001370.

Kushner, P.J., and L.M. Polvani (2004). Stratosphere-troposphere coupling in a relatively simple AGCM: The role of eddies. J. Clim., 17, 629-639.

Kushner, P.J., and L.M. Polvani (2005). A very large, spontaneous stratospheric sudden warming in a simple AGCM: A prototype for the Southern Hemisphere warming of 2002? J. Atmos. Sci., 62, 890-897.

Li, Q., H.-F. Graf and M. A. Giorgetta (2007) Stationary planetary wave propagation in the Northern Hemisphere winter – climatological analysis of the refractive index, Atm. Chem. Phys., 7, 183-200.

Mahfouf, J. F., D. Cariolle, J.-F. Royer, J.-F Geleyn and B. Timbal (1994) Response of the Météo-France climate model to changes in CO2 and sea surface temperature, Clim. Dyn. 9, 345-362.

Mailier, P. J, D. B Stephenson, C. AT Ferro, and K. I Hodges (2006) Serial Clustering of Extratropical Cyclones, Mon. Wea. Rev. 134 (8): 2224–40.

Manabe, S. and R. T. Wetherland (1967) Thermal equilibrium of the atmosphere with a given distribution of relative humidity, J. Atmos. Sci, 24, 241-259.

Manzini, E., and coauthors (2014) Northern winter climate change: Assessment of uncertainty in CMIP5 projections related to stratosphere-troposphere coupling, J. Geophys. Res. Atmos., 119, 7979-7998, doi:10.1002/2013JD021403.

McLandress, C., and T. G. Shepherd (2009), Impact of climate change on stratospheric sudden warmings as simulated by the Canadian Middle Atmosphere Model, J. Clim., 22, 5449–5463, doi:10.1175/ 2009JCLI3069.1

841 McLandress, C., A. I Jonsson, D. A. Plummer, M. C. Reader, J. F. Scinocca, T. G. Shepherd
842   (2010), Separating the dynamical effects of climate change and ozone depletion. Part I:
843   Southern Hemisphere stratosphere, J. Clim., 23, 5002-5020.

844 Meinshausen, M. and coauthors (2017) Historical greenhouse gas concentrations for climate
845   modelling (CMIP6), Geosci. Model Dev., 10, 2057-2116, doi:10.5194/gmd-10-2057-
846   2017.

847 Mitchell, D. M., Osprey, S. M., Gray, L. J., Butchart, N., Hardiman, S. C., Charlton-Perez, A. J.,
848   and Watson, P. (2012a) The effect of climate change on the variability of the Northern
849   Hemisphere stratospheric polar vortex, J. Atmos. Sci., 69, 2608–2618, 2012a.

850 Mitchell, D. M., Charlton-Perez, A. J., Gray, L. J., Akiyoshi, H., Butchart, N., Hardiman, S. C.,
851   Morgenstern, O., Nakamura, T., Rozanov, E., Shibata, K., Smale, D., and Yamashita, Y.
852   (2012b) The nature of Arctic polar vortices in chemistry-climate models, Q. J. Roy.
853   Meteor. Soc., 138, 1681–1691.

854 NASA Goddard Institute for Space Studies (NASA/GISS) (2018). NASA-GISS GISS-E2.1G
855   model output prepared for CMIP6 CMIP piControl. Version 20191120.Earth System
856   Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.7380

857 Oberländer-Hayn, S., S. Meul, U. Langematz, J. Abalichin and F. Haenel (2015) A chemistry-
858   climate model study of past changes in the Brewer-Dobson circulation, J. Geophys. Res.
859   Atmos., 120, 6742-6757.

860 Polvani, L.M., Sun, L., A.H. Butler, J.H. Richter, and C. Deser (2017), Distinguishing
861   stratospheric sudden warmings from ENSO as key drivers of wintertime climate
862   variability over the North Atlantic and Eurasia, J. Clim., 30, 1959-1969, doi:
863   10.1175/JCLI-D-16-0277.1.

864 Rind, D., R. Suozzo, N. K. Balachandran, and M. J. Prather (1990) Climate change and the
865   middle atmosphere. Part I: The doubled CO2 climate, J. Atmos. Sci., 47, 475-494.

866 Rind D., D. Shindell, P. Lonergan, and N. K. Balachandran (1998) Climate change and the
867   middle atmosphere. Part III: The doubled CO2 climate revisited, J. Clim., 11, 876-894.

868 Roberts, M. (2017). MOHC HadGEM3-GC31-LL model output prepared for CMIP6. Version
869   20190723.Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.1901

870 Scaife, A., and coauthors (2012), Climate change projections and stratosphere–troposphere
871   interaction, Clim. Dyn., 38, 2089-2097, doi:10.1007/s00382-011-1080-7.

872 Schoeberl, M. R., and D. L. Hartmann (1991) The dynamics of the stratospheric polar vortex and
873   its relation to springtime ozone depletions, Science, 251, 46-52.

874 Sellers, K. F, and D. S Morris (2017) Underdispersion Models: Models That Are 'Under the
875   Radar', Communications in Statistics-Theory and Methods 46 (24), 12075–12086.

876 Séférian, R. (2018) CNRM-CERFACS CNRM-ESM2-1 model output prepared for CMIP6
877   CMIP for experiment piControl-spinup. Version 20180423. Earth System Grid
878   Federation. https://doi.org/10.22033/ESGF/CMIP6.4169

Séférian, R. and coauthors (2019) Evaluation of CNRM Earth-System model, CNRM-ESM2-1: role of Earth system processes in present-day and future climate, J. Adv. Model. Earth Systems, https://doi.org/10.1029/2019MS001791.

Shepherd T.G. and C. S. McLandress (2011) A robust mechanism for strengthening of the Brewer–Dobson circulation in response to climate change: Critical-layer control of subtropical wave breaking, J. Atmos. Sci., 69, 784-797, doi: 10.1175/2010JAS3608.1

Song, Y. and W. A. Robinson (2004) Dynamical mechanisms for stratospheric influences on the troposphere, J. Atmos. Sci., 61,1711-1725.

Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., Anstey, J., Arora, V., Christian, J. R., Hanna, S., Jiao, Y., Lee, W. G., Majaess, F., Saenko, O. A., Seiler, C., Seinen, C., Shao, A., Sigmond, M., Solheim, L., von Salzen, K., Yang, D., and Winter, B. (2019a) The Canadian Earth System Model version 5 (CanESM5.0.3), Geosci. Model Dev., 12, 4823–4873, https://doi.org/10.5194/gmd-12-4823-2019.

Swart, Neil Cameron; Cole, Jason N.S.; Kharin, Viatcheslav V.; Lazare, Mike; Scinocca, John F.; Gillett, Nathan P.; Anstey, James; Arora, Vivek; Christian, James R.; Jiao, Yanjun; Lee, Warren G.; Majaess, Fouad; Saenko, Oleg A.; Seiler, Christian; Seinen, Clint; Shao, Andrew; Solheim, Larry; von Salzen, Knut; Yang, Duo; Winter, Barbara; Sigmond, Michael (2019). CCCma CanESM5 model output prepared for CMIP6 CMIP piControl. Version 20190730. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.3673

Taguchi, M. (2017) A study of different frequencies of major stratospheric sudden warmings in CMIP5 historical simulations, J. Geophys. Res. Atmos., 122, 5144-5156, doi: 10.1002/2016JD025826.

Tang, Y. and coauthors (2019). MOHC UKESM1.0-LL model output prepared for CMIP6 CMIP piControl. Version 20190904. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.6298

Tatebe, H. and Watanabe, M. (2018). MIROC MIROC6 model output prepared for CMIP6 CMIP piControl. Version 20190903.Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.5711

Tatebe, H., and coauthors (2019) Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6, Geosci. Model Dev., 12, 2727–2765, https://doi.org/10.5194/gmd-12-2727-2019.

Volodin, E. M., Mortikov, E. V., Kostrykin, S. V., Galin, V. Y., Lykossov, V. N., Gritsun, A. S., Diansky, N. A., Gusev, A. V., and Iakovlev, N. G. (2017) Simulation of the present day climate with the climate model INMCM5, Clim. Dyn., 49, 3715, https://doi.org/10.1007/s00382-017-3539-7

Wilks, Daniel S. (2011). *Statistical Methods in the Atmospheric Sciences*. Vol. 100. Academic press.

Williams, K., Copsey, D., Blockley, E., Bodas-Salcedo, A., Calvert, D., Comer, R., Davis, P., Graham, T., Hewitt, H., & Hill, R. (2018). The Met Office global coupled model 3.0 and 3.1 (GC3. 0 and GC3. 1) configurations. J. Adv. Model. Earth Systems, 10(2), 357–380.

920   Wu, Y., Simpson, I. R. and Seager, R. (2019) Inter-model spread in the Northern Hemisphere
921   stratospheric polar vortex response to climate change in the CMIP5 models, Geophys.
922   Res. Lett., 46, 13290-13298, https://doi.org/10.1029/2019GL085545.

923   Yukimoto, S. and coauthors (2019). MRI MRI-ESM2.0 model output prepared for CMIP6
924   CMIP. Version 20190726.Earth System Grid Federation.
925   https://doi.org/10.22033/ESGF/CMIP6.621

926   Yukimoto, S. and coauthors (2019) The Meteorological Research Institute Earth System Model
927   Version 2.0, MRI-ESM2.0: Description and Basic Evaluation of the Physical Component,
928   J. Meteor. Soc. Jpn, 97(5),931−965, doi:10.2151/jmsj.2019-051

929   Zelinka M. D., T. A. Myers, D. T. McCoy, S. Po-Chedley, P. M. Caldwell, P. Ceppi, S. A. Klein
930   and K. E. Taylor (2020) Causes of Higher Climate Sensitivity in CMIP6 Models,
931   Geophys. Res. Lett., doi:10.1029/2019GL085782
932
933

934 **Table 1.** List of models included in the analysis indicating their resolution and the ensemble
935 members considered in simulations (rXiXpXfX: where r corresponds to realization, i to
936 initialization, p to physics and f to forcing). Effective climate sensitivity for CO_2 doubling is
937 taken from analysis by A. G. Pendergrass using Gregory et al. [2004] method
938 (https://github.com/apendergrass/cmip6-ecs) apart from the estimate for GISS-E2.2AP which
939 was provided by a reviewer. We use the term 'Effective Climate Sensitivity' here following the
940 discussion in and recommendation of Zelinka et al. (2020)

941

Models	Model resolution	Ensemble members	Internally generated QBO	Nr. of years piControl run	Effective Climate Sensitivity / K
CanESM5 [Swart et al. 2019a,b]	T63L49, top 1hPa	r1i1p2f1	No	450	5.59
CESM2 [Danabasoglu et al. 2019, 2020]	1ºx1º L32, top 40km	r1i1p1f1	No	1200	5.12
CESM2-WACCM [Danabasoglu, 2019; Gettelman et al. 2019]	1ºx1º L70, top 150km	r1i1p1f1	Yes	500	4.61
CNRM-ESM2-1 [Séférian, 2018; Séférian et al. 2019]	Tl127L91, top 0.01hPa	r1i1p1f2	Yes	500	4.66
GFDL-CM4 [Guo et al. 2018; Held et al. 2019]	C96L33, top 1hPa	r1i1p1f1	No	140	3.84
GISS-E2.2AP [NASA-GISS et al. 2018]	2ºx2.5º, top 0.002hPa	r1i1p1f1	Yes	81	2.1
HadGEM3-GC31-LL [Roberts, 2017 Williamson et al. 2018]	N261L85, top 85km	r1i1p1f3 except for piControl run: r1i1p1f1	Yes	500	5.41
INM-CM5-0 [Volodin et al. 2017]	2x1.5L73, top 0.2hPa	r1i1p1f1	Yes	154	2.1
IPSL-CM6A-LR [Boucher et al. 2018]	N96, top 80km	r1i1p1f1	Yes	1200	4.49
MIROC6 [Tatebe et al. 2018; 2019]	T85L81, top 0.004hPa	r1i1p1f1	Yes	800	2.54
MRI-ESM2-0 [Yukimoto et al. 2019a b]	TL159L80, top 0.01hPa	r1i1p1f1	Yes	200	3.30
UKESM1-0-LL [Tang et al., 2019; Kuhlbrodt et al. 2018]	N96L85, top 85 km	r1i1p1f2	Yes	1100	5.27

942

943   **Figures captions**
944
945   **Figure 1**. **(a)** Average annual SSW frequency in the historical simulations (1958-2014) of the 11
946   models. Black lines show 95% confidence estimates for the annual frequency. Dashed black line
947   corresponds to SSW frequency in the JRA-55 reanalysis, with its 95% confidence interval in the
948   light gray shading. **(b)** Same as (a) but for SSW occurrence in the piControl (light gray bars) and
949   abrupt4xCO$_2$ simulations (dark gray bars). Black lines show 95% confidence intervals for each
950   estimate. Bars are ordered by the size of the difference between the two simulations.
951
952   **Figure 2.** SSW frequency distribution in the historical simulation of each model (blue line) and
953   JRA-55 reanalysis period (orange dashed line). The distribution has smoothed by a kernel
954   smoother of a bandwidth of 10 days. Shading corresponds to 2.5th-97.5th percentile range of the
955   bootstrap samples i.e., the 95% confidence interval on the mean of the piControl simulation. (See
956   more details about the determination of this interval in Appendix).
957
958   **Figure 3:** Scatter plots of the change of SSW frequency between the piControl and abrupt4xCO$_2$
959   simulations vs. (a) the frequency in the piControl simulations, (b) the frequency in the historical
960   simulations, (c) the ECS, (d) the change in tropical temperature at 250hPa, (e) the change in
961   polar temperature at 850hPa and (f) the difference in polar-tropical temperature difference at
962   850hPa. In (a) and (b) the grey dashed line shows the observed SSW frequency in the JRA-55 re-
963   analysis (0.64 SSW yr^{-1}). The temperature regions in (d)-(f) are defined as in Harvey et al.
964   (2014).
965
966   **Figure 4.** Probability distribution of daily zonal mean zonal wind at 60ºN and 10hPa (m/s) for
967   the piControl (blue) and abrupt4xCO$_2$ (orange) experiments. Dashed lines represent the median
968   value of the distribution in each integration.
969
970   **Figure 5. (a)** Estimated fractional change in SSW frequency by the seventh decade of the
971   1pctCO$_2$ simulations. Light gray shaded lines indicate that the trend of SSW frequency in the
972   model is significantly different from zero at a p-value of 0.05. Dashed line indicates trend equal
973   to 1, i.e. no trend in the SSWs frequency. **(b)** Decade of emergence of SSW frequency trend for
974   those models in which the trend term is significantly different from zero, calculated as described
975   in the main text.
976
977   **Figure 6**. Box plots showing the distribution of dates of **(a)** polar vortex formation and **(b)**
978   stratospheric final warming in the Northern Hemisphere for the piControl (blue), historical
979   (green) and abrupt-4xCO$_2$ (red) simulations for all models and JRA-55 reanalysis. **(c)** Same as
980   (b) but for the Southern Hemisphere and only in piControl and historical runs. The interquartile
981   range is represented by the size of the box and the inside line (black cross) corresponds to the
982   median (mean). Whiskers indicate the maximum and minimum points in the distribution that are
983   not outliers. Outliers (red crosses) are defined as points with values greater than 3/2 times the
984   interquartile range from the ends of the box.
985
986   **Figure 7.** Composite maps of anomalous SLP (contour interval 1hPa) and 2m temperature
987   (shading) for 15/60 days after SSWs in piControl simulation and JRA-55 reanalysis (bottom
988   left). Green stippling indicates stat. significant differences in SLP from JRA-55 reanalysis at the
989   95% confidence level. Numbers in titles indicate the number of events considered.

990
991   **Figure 8**. Abrupt4xCO$_2$-minus-piControl composite maps of anomalous SLP (shading, hPa) for
992   15/60 days after SSWs. Anomalous SLP after SSWs in piControl run is shown in contours
993   (interval: 1hPa). Green stippling indicates stat. significant differences from piControl run at the
994   95% confidence level. Numbers in titles indicate the number of events considered in the
995   piControl simulation (piC) and abrupt4xCO2 (4x).
996
997   **Figure 9.** Fraction (%) of SSWs that are also PJO events in piControl (solid bars) and
998   abrupt4xCO$_2$ (open bar) runs. Horizontal black solid and dashed line correspond to the mean
999   value and the 2.5^{th}-97.5^{th} percentile range in JRA-55 reanalysis, respectively. Error bars are
1000   based on bootstrapping.
1001

Figure 1.

a **Historical**

b **PiControl vs Abrupt4xCO2**
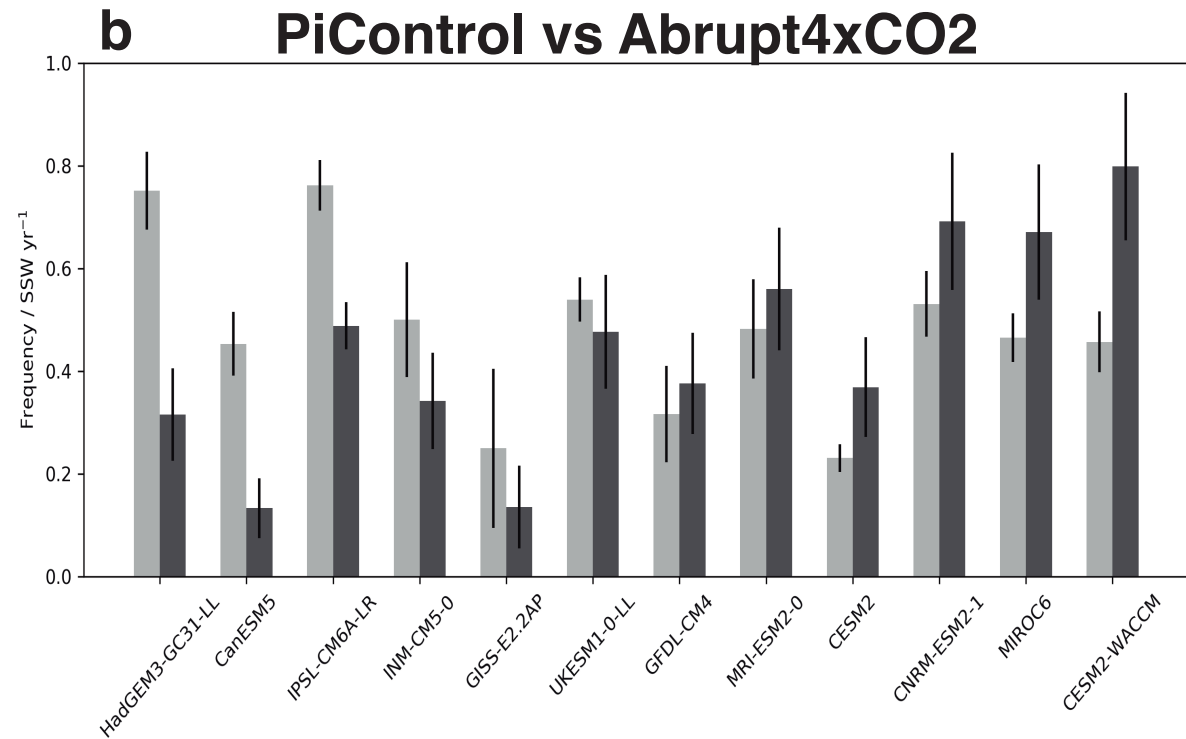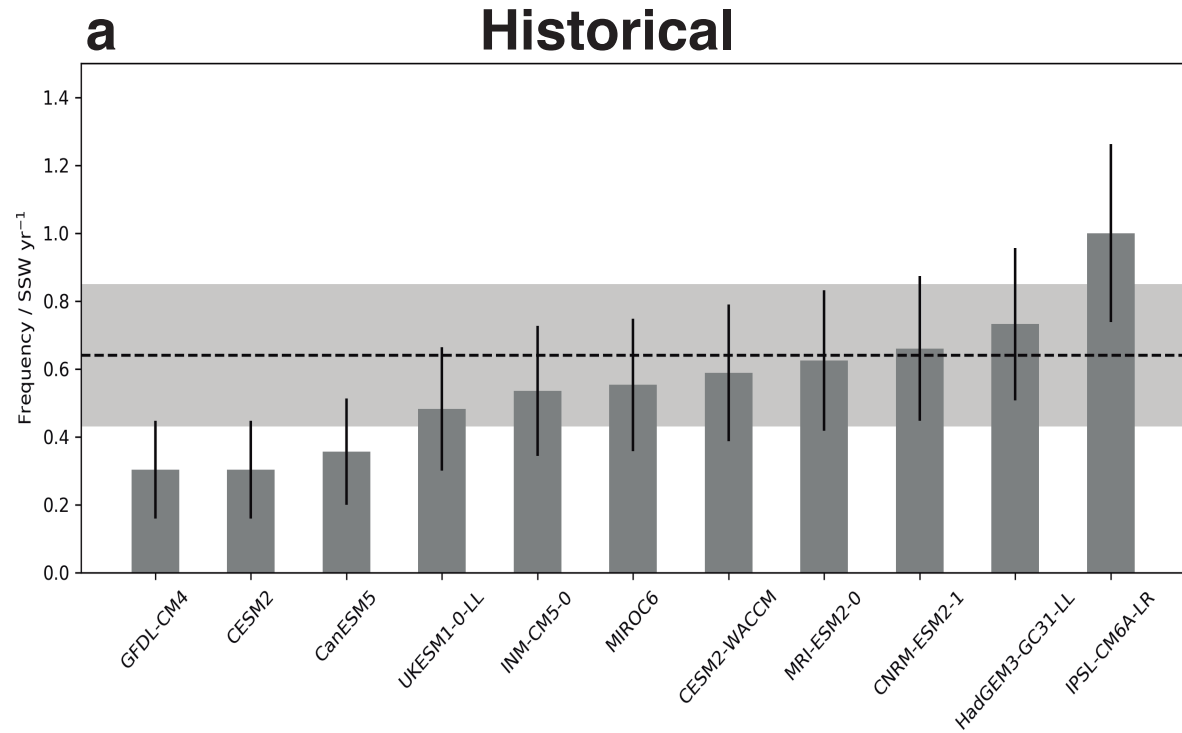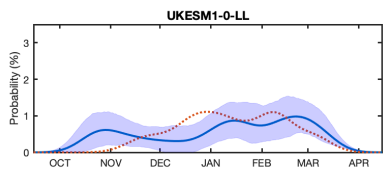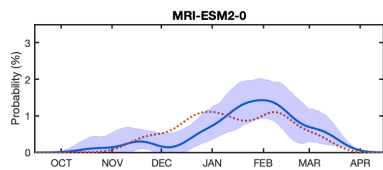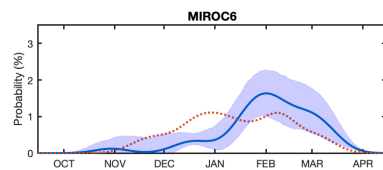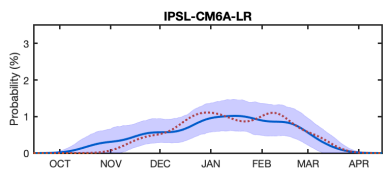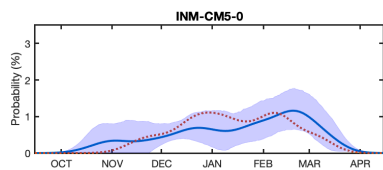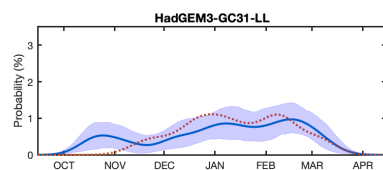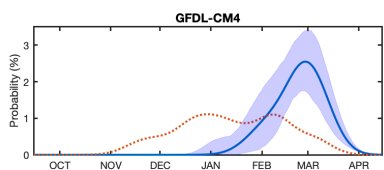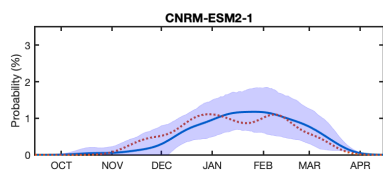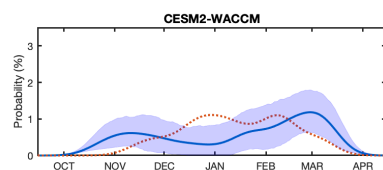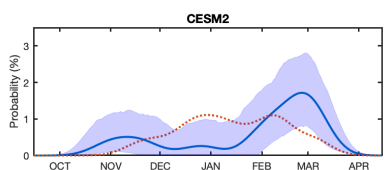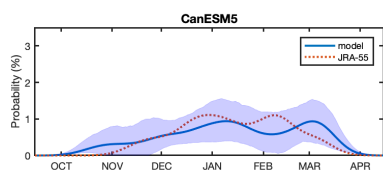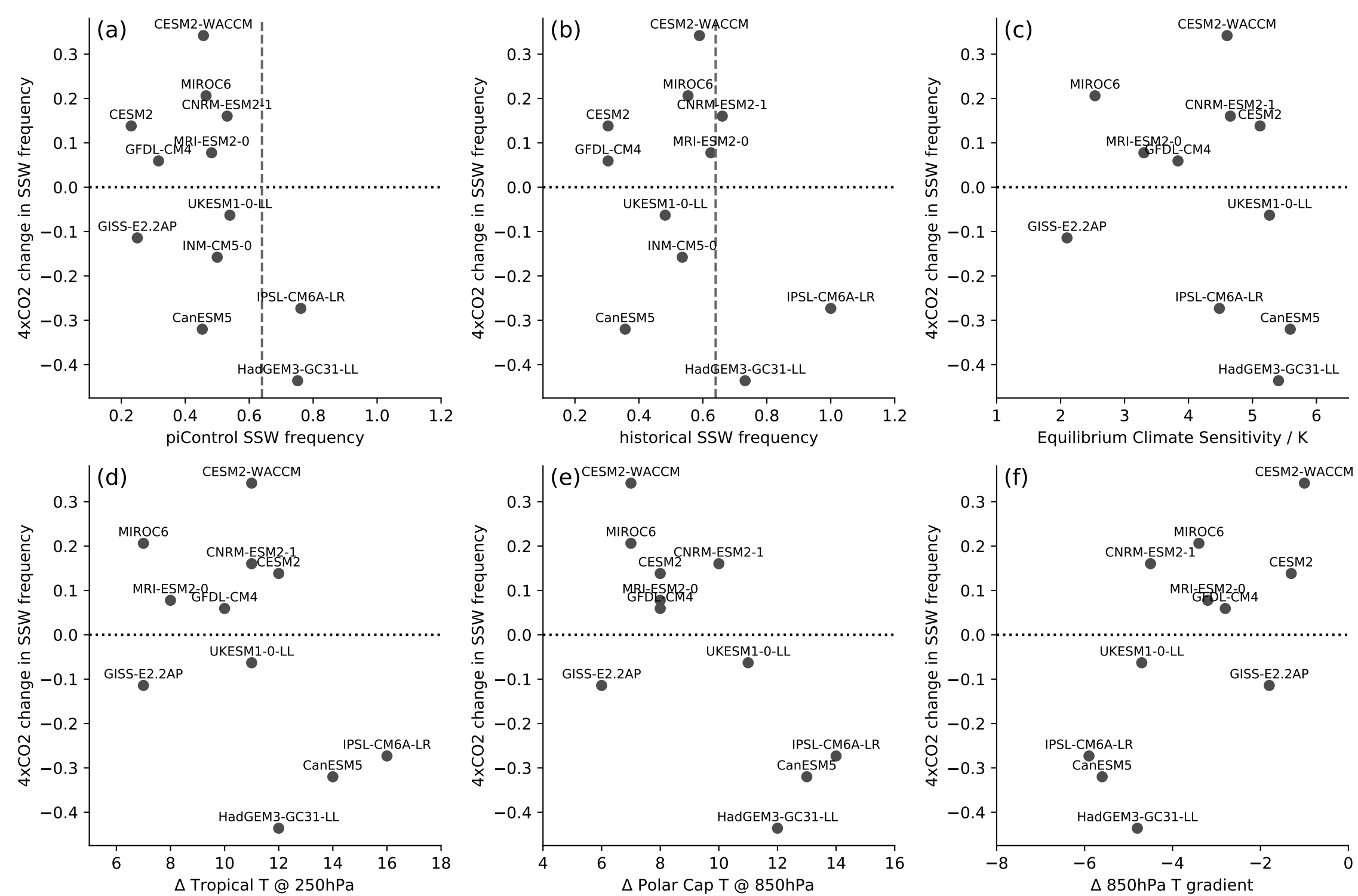
Figure 2.

Figure 3.

Figure 4.
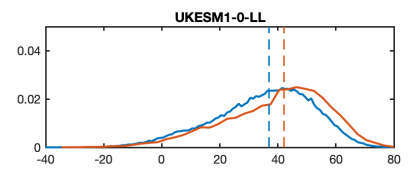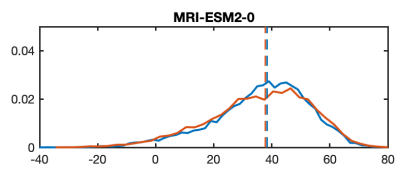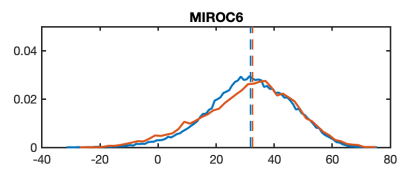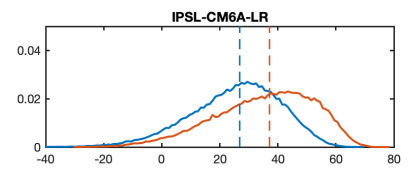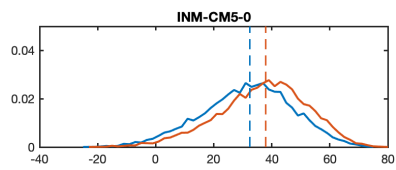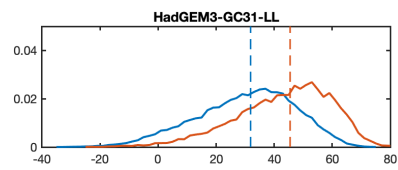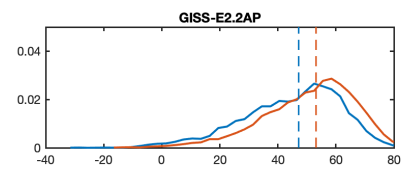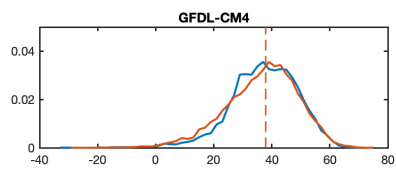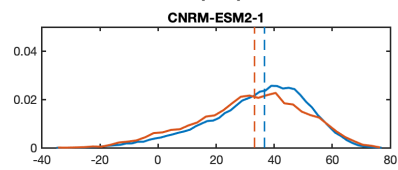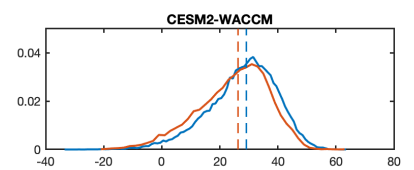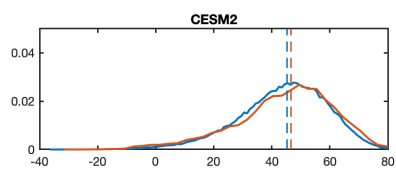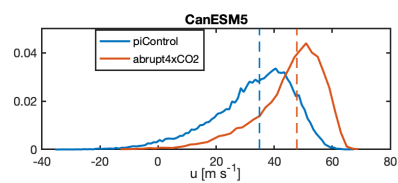
Figure 5.

Fractional change in SSW freq.
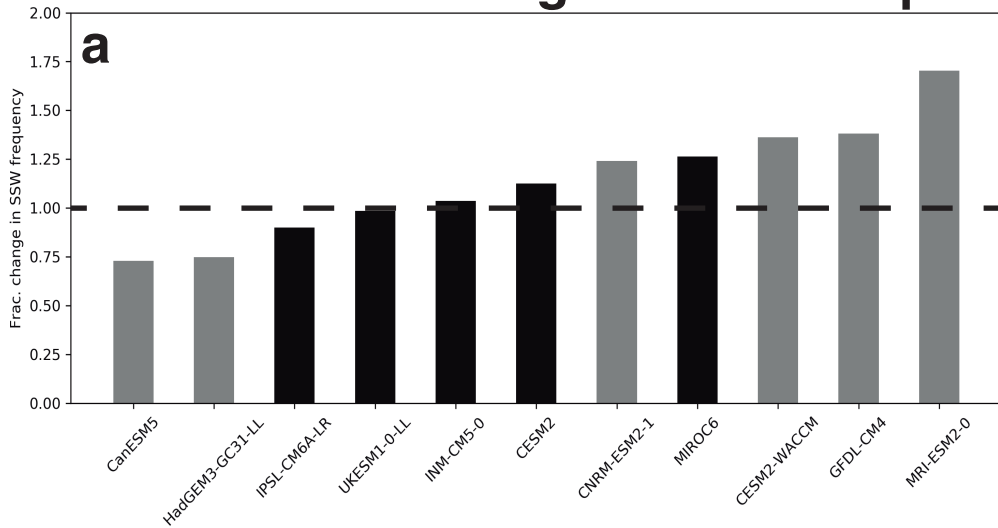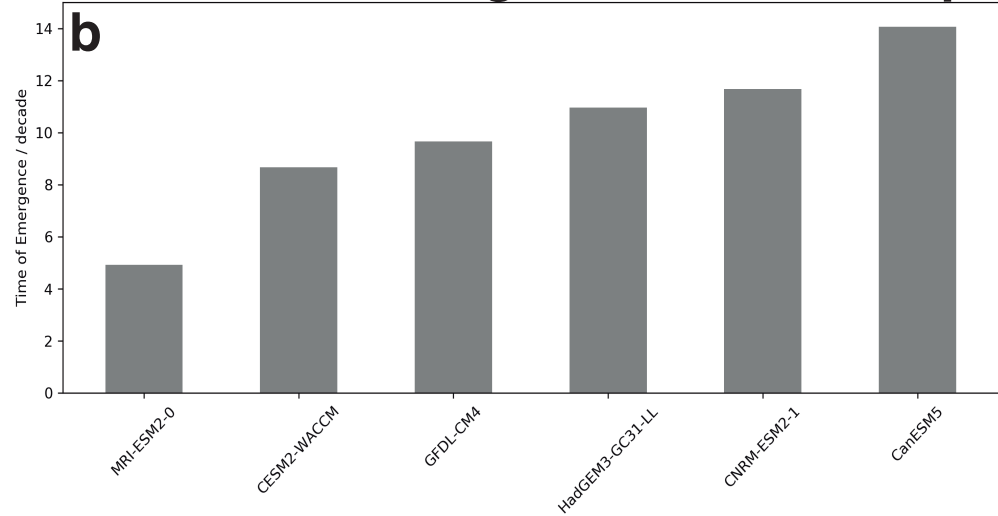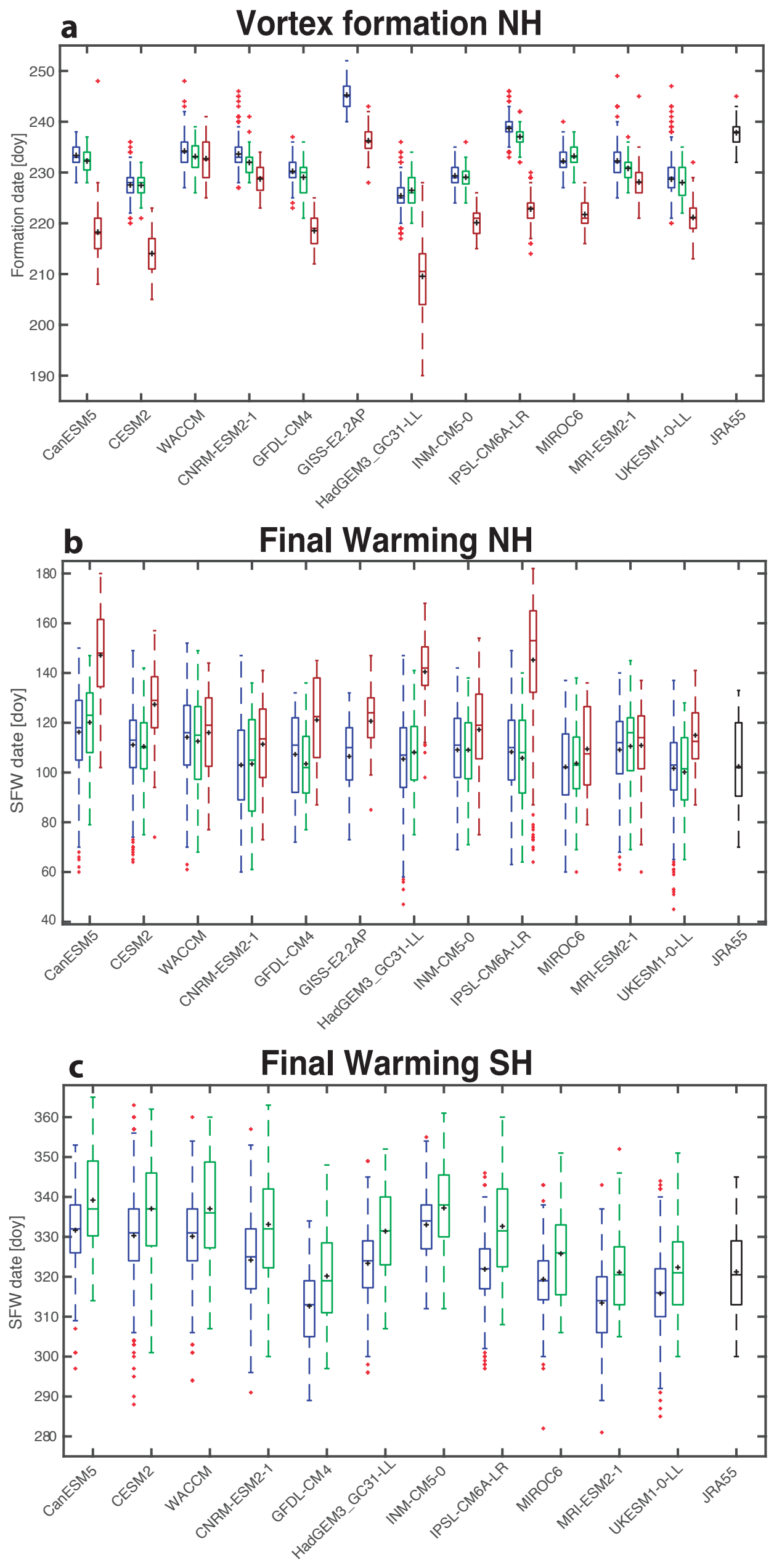
a

Decade of emergence of SSW freq.

b

Figure 6.

a **Vortex formation NH**

b **Final Warming NH**

c **Final Warming SH**
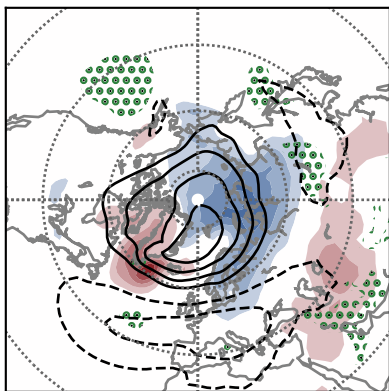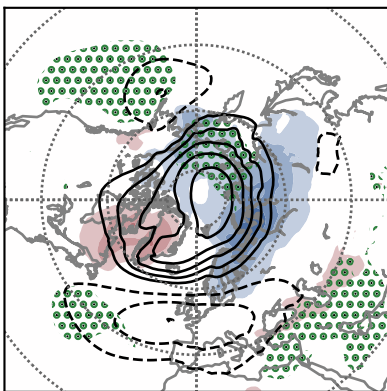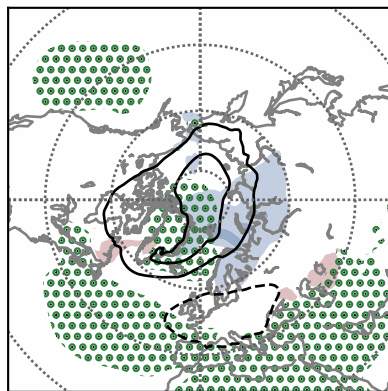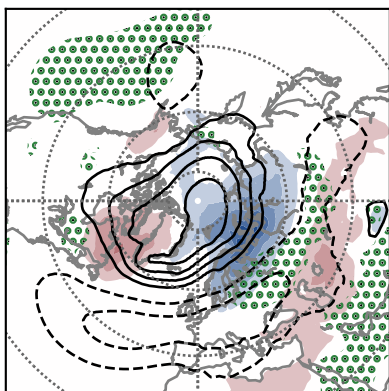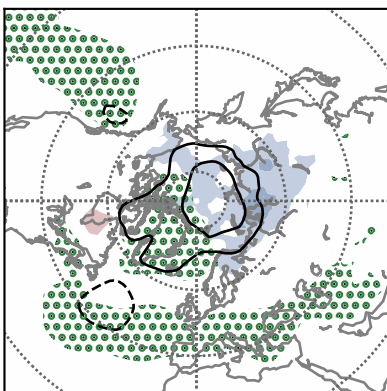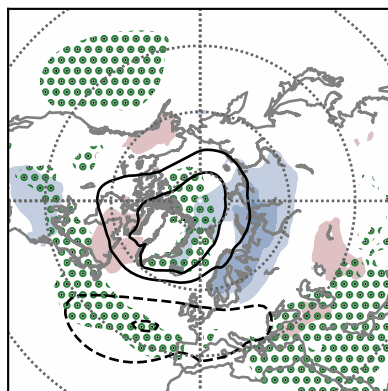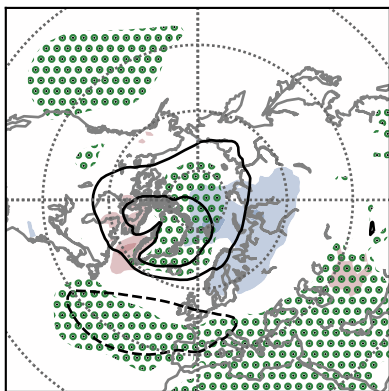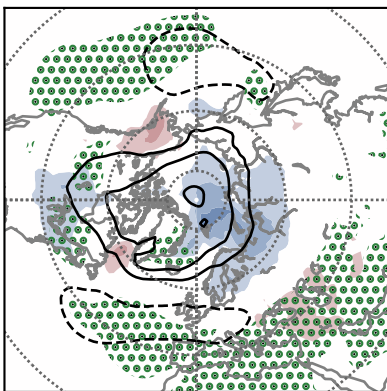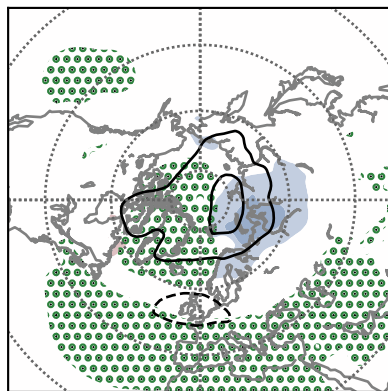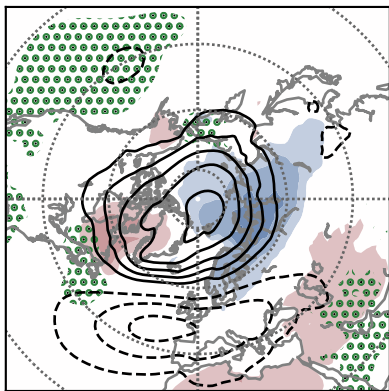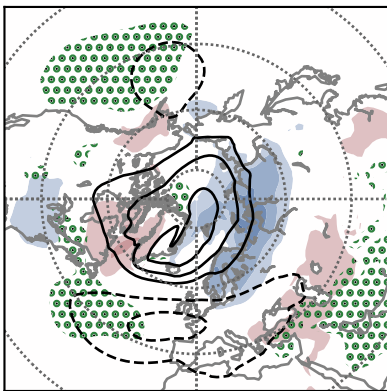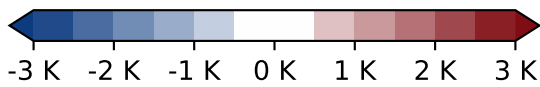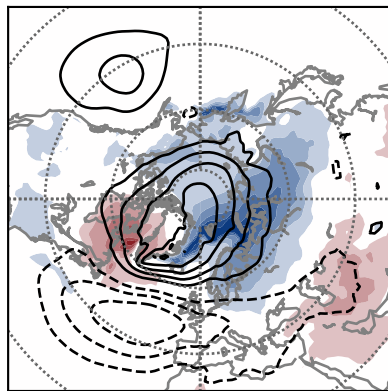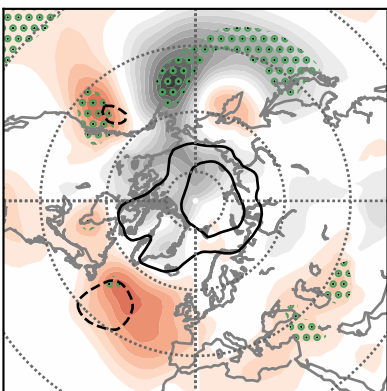
Figure 7.

Figure 8.

CANESM5
4x:7 piC:205
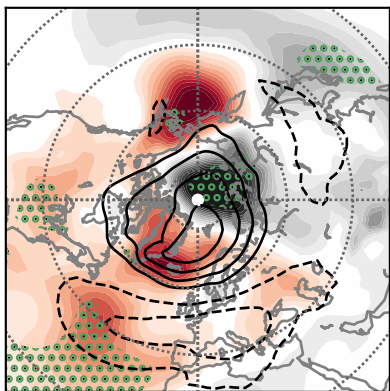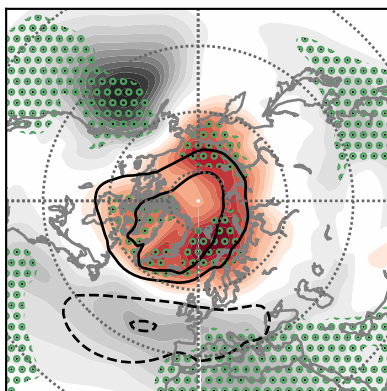
CESM2
4x:30 piC:278
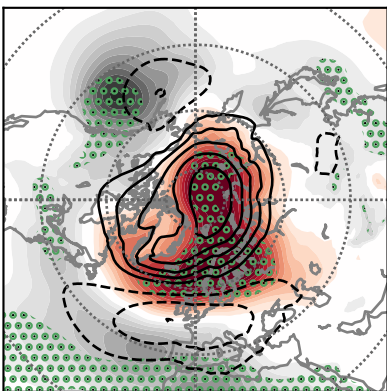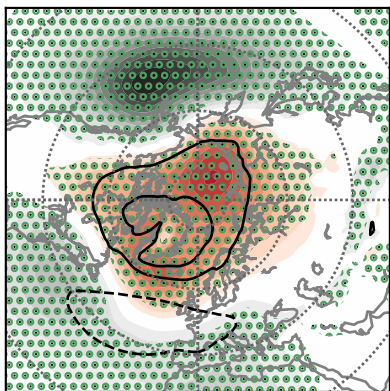
CESM2-WACCM
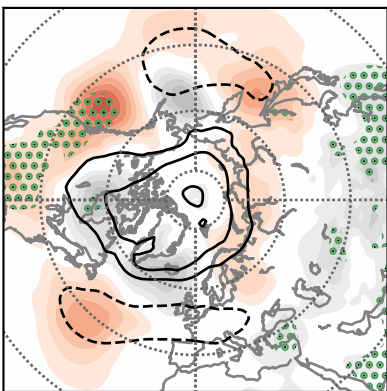4x:68 piC:229
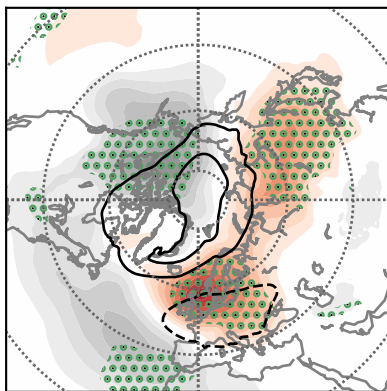
CNRM-ESM2-1
4x:55 piC:267
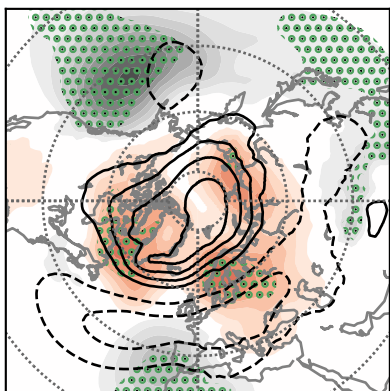
GFDL-CM4
4x:25 piC:44

HADGEM3-GC31-LL
4x:23 piC:375
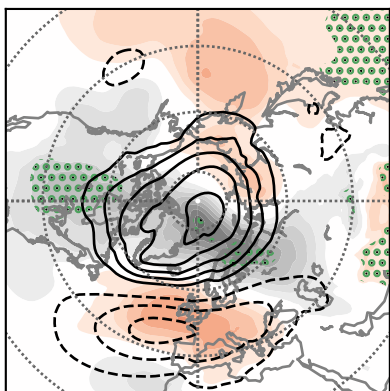
IPSL-CM6A-LR
4x:269 piC:915

INM-CM5-0
4x:25 piC:77

MIROC6
4x:56 piC:234
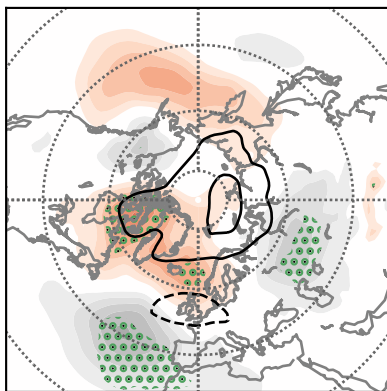
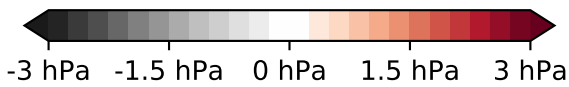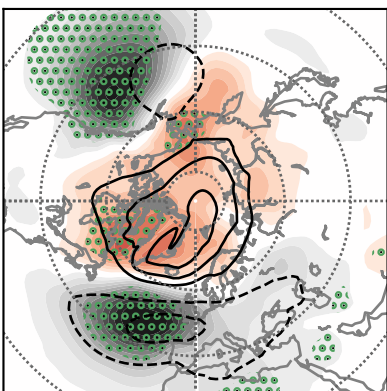MRI-ESM2-0
4x:41 piC:96

UKESM1-0-LL
4x:28 piC:595

-3 hPa   -1.5 hPa   0 hPa   1.5 hPa   3 hPa

Figure 9.

% of SSWs that are PJOs