JGR Atmospheres

RESEARCH ARTICLE
10.1029/2019JD032345

Uncertainty in the Response of Sudden Stratospheric Warmings and Stratosphere-Troposphere Coupling to Quadrupled CO_2 Concentrations in CMIP6 Models

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

[1]Departamento de Física de la Tierra y Astrofísica, Universidad Complutense de Madrid, Madrid, Spain, [2]Department of Meteorology, University of Reading, Reading, UK, [3]Cooperative Institute for Environmental Sciences (CIRES)/National Oceanic and Atmospheric Administration (NOAA) Chemical Sciences Division, Boulder, CO, USA, [4]Department of Earth and Atmospheric Sciences, Cornell University, Ithaca, NY, USA, [5]Climate and Global Dynamics Laboratory, National Center for Atmospheric Research, Boulder, CO, USA, [6]Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY, USA, [7]Met Office Hadley Centre, Exeter, UK, [8]Courant Institute of Mathematical Sciences, New York University, New York, NY, USA, [9]NCAS-Climate, Department of Physics, University of Oxford, Oxford, UK, [10]Deutsches Zentrum für Luft-und Raumfahrt (DLR) Oberpfaffenhofen, Weßling, Germany, [11]Atmospheric and Oceanic Sciences Program, Princeton University, Princeton, NJ, USA, [12]Laboratoire de Météorologie Dynamique, Ecole Normale Supérieure, Paris, France, [13]Max-Planck-Institut für Meteorologie, Hamburg, Germany, [14]Meteorological Research Institute, Tsukuba, Japan, [15]NASA Goddard Institute for Space Studies, New York, NY, USA, [16]Centre National de Recherches Météorologiques (CNRM), Université de Toulouse, Météo-France, CNRS, Toulouse, France, [17]Canadian Centre for Climate Modelling and Analysis, Environment and Climate Change Canada, Victoria, BC, Canada, [18]Department of Earth Science, Aichi University of Education, Kariya, Japan, [19]Marchuk Institute of Numerical Mathematics, Moscow, Russia, [20]Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Yokohama, Japan

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

Supporting Information:
- Supporting Information S1

Correspondence to:
B. Ayarzagüena,
bayarzag@ucm.es

Citation:
Ayarzagüena, B., Charlton-Perez, A. J., Butler, A. H., Hitchcock, P., Simpson, I. R., Polvani, L. M., et al. (2020). Uncertainty in the response of sudden stratospheric warmings and stratosphere-troposphere coupling to quadrupled CO_2 concentrations in CMIP6 models. *Journal of Geophysical Research: Atmospheres*, *125*, e2019JD032345. https://doi.org/10.1029/2019JD032345

Received 30 DEC 2019
Accepted 26 FEB 2020
Accepted article online 28 FEB 2020

Abstract Major sudden stratospheric warmings (SSWs), vortex formation, and final breakdown dates are key highlight points of the stratospheric polar vortex. These phenomena are relevant for stratosphere-troposphere coupling, which explains the interest in understanding their future changes. However, up to now, there is not a clear consensus on which projected changes to the polar vortex are robust, particularly in the Northern Hemisphere, possibly due to short data record or relatively moderate CO_2 forcing. The new simulations performed under the Coupled Model Intercomparison Project, Phase 6, together with the long daily data requirements of the DynVarMIP project in preindustrial and quadrupled CO_2 ($4xCO_2$) forcing simulations provide a new opportunity to revisit this topic by overcoming the limitations mentioned above. In this study, we analyze this new model output to document the change, if any, in the frequency of SSWs under $4xCO_2$ forcing. Our analysis reveals a large disagreement across the models as to the sign of this change, even though most models show a statistically significant change. As for the near-surface response to SSWs, the models, however, are in good agreement as to this signal over the North Atlantic: There is no indication of a change under $4xCO_2$ forcing. Over the Pacific, however, the change is more uncertain, with some indication that there will be a larger mean response. Finally, the models show robust changes to the seasonal cycle in the stratosphere. Specifically, we find a longer duration of the stratospheric polar vortex and thus a longer season of stratosphere-troposphere coupling.

1. Introduction

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

Interest in the polar vortex has increased in the last decades for two different reasons. First, the magnitude of the Antarctic ozone hole is dependent on the state of the polar vortex, as a strong polar vortex is associated with colder temperatures (crucial for heterogeneous ozone chemistry) and reduced mixing with ozone-rich midlatitude air (Schoeberl & Hartmann, 1991). Second, polar stratospheric variability is known to affect not only the stratosphere but also the troposphere, typically projecting onto Annular Mode patterns (e.g., Baldwin & Dunkerton, 2001 ; Kidston et al., 2015). Polar stratospheric variability peaks in the winter hemisphere when the polar vortex is present, as a major source of stratospheric variability is upward propagating, planetary-scale Rossby waves from the troposphere below (Charney & Drazin, 1961). Under linear theory, the vertical propagation of Rossby waves is limited to regions with westerly winds (Andrews et al., 1987). Furthermore, because wave activity is greater in the Northern Hemisphere (NH) than in the Southern Hemisphere (SH), so is the polar stratospheric variability. In the SH, stratospheric variability, and thus the coupling to the troposphere, is mainly associated with SFW (Black & McDaniel, 2007). In the NH apart from SFWs (Ayarzagüena & Serrano, 2009; Black et al., 2006; Hardiman et al., 2011), this coupling is primarily associated with polar vortex extremes, in particular, major sudden stratospheric warmings (SSWs). SSWs happen in midwinter and consist in a reversal of wintertime polar stratospheric circulation with a subsequent recovery of the polar vortex after the event. The tropospheric signal of SSWs can persist for up to 2 months after the occurrence of each event (Charlton & Polvani, 2007). Although the exact mechanism for this downward influence is still unclear, different hypothesis have been presented in the literature such as wave reflection, downward control, or responses to stratospheric redistributions of potential vorticity, among others (Song & Robinson, 2004, and references herein). In most of these theories, the role of the circulation anomalies of the lower stratosphere was found to be extremely important to define the impact on the troposphere. Indeed, recently, Hitchcock et al. (2013) defined a subset of SSWs, called Polar-night Jet Oscillation events (PJOs), which are characterized by a very persistent warm polar lower stratosphere and whose signal in the troposphere is particularly strong and persistent too.

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

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

Recently, Ayarzagüena et al. (2018) revisited this topic, trying to overcome some of the issues suggested in the literature as potential reasons for this disagreement, such as the use of one single model in the analysis or the dependence of results on the SSW identification criterion. They analyzed 12 different models participating in the Chemistry Climate Model Initiative (CCMI) and applied several different (absolute and relative) criteria for the identification of SSWs. The outcome was again a lack of a significant change in SSWs frequency in the future, although most of the models predicted a slight increase in the frequency of these, regardless of the SSW identification algorithm. One might argue, however, that the limited data record available (40 years in each period of study), and the relatively moderate GHG forcing used in the central CCMI scenario (Representative Concentration Pathway 6.0, RCP6.0), might be insufficient to detect significant changes in SSWs in those simulations.

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

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

2. Data and Methodology

2.1. Data

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

We focus on four DECK experiments (Eyring et al., 2016), each of them used for different purposes. The historical run is employed for model validation: We compare the simulated SSW frequency, intensity, and seasonality to the values obtained from the JRA-55 reanalysis (Kobayashi et al., 2015). In fact, we have specifically restricted the analysis period to 1958–2014 to perform a rigorous quantitative comparison with JRA-55. This reanalysis shows a very good performance in representing SSWs (Ayarzagüena et al., 2019) and is the most modern reanalyses of the three that extend longer than the satellite era and assimilate more than surface data (ERA-40, NCEP/NCAR reanalysis, and JRA-55).

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

The abrupt4xCO_2 and 1pctCO_2 runs are used to examine the impact of CO_2 forcing on SSW properties. Both simulations extend 150 years (except for the abrupt4xCO_2 in IPSL-CM6A-LR, which is 900 years long, and GISS-E2.2AP, which contains 81 years). All forcings in the abrupt4xCO_2 simulations are identical to those

Table 1
List of Models Included in the Analysis Indicating Their Resolution and the Ensemble Members Considered in Simulations (rXiXpXfX: Where r Corresponds to Realization, i to Initialization, p to Physics, and f to Forcing)

Models	Model resolution	Ensemble members	Internally generated QBO	Nr. of years piControl run	Effective climate sensitivity (K)
CanESM5 (Swart et al., 2019a, 2019b)	T63L49, top 1 hPa	r1i1p2f1	No	450	5.59
CESM2 (Danabasoglu et al., 2019, 2020)	1° × 1° L32, top 40 km	r1i1p1f1	No	1,200	5.12
CESM2-WACCM (Danabasoglu, 2019; Gettelman et al., 2019)	1° × 1° L70, top 150km	r1i1p1f1	Yes	500	4.61
CNRM-ESM 2-1 (Séférian, 2018; Séférian et al., 2019)	Tl127L91, top 0.01 hPa	r1i1p1f2	Yes	500	4.66
GFDL-CM4 (Guo et al., 2018; Held et al., 2019)	C96L33, top 1 hPa	r1i1p1f1	No	140	3.84
GISS-E2.2AP (NASA Goddard Institute for Space Studies (NASA/GISS), 2018)	2° × 2.5°, top 0.002 hPa	r1i1p1f1	Yes	81	2.1
HadGEM3-GC31-LL (Roberts, 2017; Williams et al., 2018)	N261L85, top 85 km	r1i1p1f3 except for piControl run: r1i1p1f1	Yes	500	5.41
INM-CM5-0 (Volodin et al., 2017)	2° × 1.5º L73, top 0.2 hPa	r1i1p1f1	Yes	154	2.1
IPSL-CM6A-LR (Boucher et al., 2018)	N96, top 80 km	r1i1p1f1	Yes	1,200	4.49
MIROC6 (Tatebe et al., 2019; Tatebe & Watanabe, 2018)	T85L81, top 0.004 hPa	r1i1p1f1	Yes	800	2.54
MRI-ESM 2-0 (Yukimoto, Koshiro, et al., 2019 Yukimoto, Kawai, et al., 2019)	TL159L80, top 0.01 hPa	r1i1p1f1	Yes	200	3.30
UKESM1-0-LL (Kuhlbrodt et al., 2018; Tang et al., 2019)	N96L85, top 85 km	r1i1p1f2	Yes	1,100	5.27

Note. Effective climate sensitivity for CO_2 doubling is taken from analysis by A. G. Pendergrass using Gregory et al. (2004) method (https://github.com/apendergrass/cmip6-ecs) apart from the estimate for GISS-E2.2AP which was provided by a reviewer. We use the term “Effective Climate Sensitivity” here following the discussion in and recommendation of Zelinka et al. (2020).

in the piControl run, except for the CO_2 concentrations, which are abruptly quadrupled from piControl levels, and then are held constant throughout the entire length of the simulation (Eyring et al., 2016). The large and constant forcing in the abrupt4xCO_2 makes it possible to isolate robust changes, if any, to the size and nature of changes to SSW properties. In the 1pctCO_2 simulation, the CO_2 concentration starts at preindustrial levels and is increased at the rate of 1% per year. This simulation is used to estimate the rate at which SSW frequency might change in the future (one aspect of the so-called “dynamical sensitivity” of the stratosphere; Grise & Polvani, 2016).

Anomalies are defined as the departure from the daily evolving annual cycle of each respective model. In the piControl run, the climatology is based on the whole period, while in the historical run, only the 1979–2014 is considered for calculating the climatology. In the abrupt4xCO_2 runs, a trend is identified in some variables during the first 50 years following the switch-on of the forcing. To avoid this trend, the climatologies are computed after omitting the first 75 years except for IPSL-CM6A-LR where we omit the first 300 years, but we keep the following 600 years. A similar omission of data is performed for the analysis of SFW or vortex formation dates. In contrast, the full abrupt4xCO_2 is considered when looking at SSW frequency as no trend is detectable in the occurrence of these phenomena.

2.2. Methods

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

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

1. *SSWs* are identified following the criterion proposed in Charlton and Polvani (2007), which is based on the reversal in the sign of $u_{60N10hPa}$ from November to March. Their criterion includes two additional restrictions: (1) Winds must return to westerly for at least 20 consecutive days between events and (2) winds must return to westerly for at least 10 consecutive days before 30 April of each year. Recall that this definition only identifies so-called "major" SSWs. Here we do not examine other aspects of polar vortex variability, such as vortex intensification events, wave reflection events, or minor stratospheric warmings.
2. *SFWs* are defined as the last date in the spring on which $u_{60N10hPa}$ reverses and does not return to westerly for more than 10 consecutive days (Butler & Gerber, 2018).
3. *The polar vortex formation date* is identified as the first time that $u_{60N10hPa}$ turns westerly after 1 July, in the NH, and stays westerly for at least 10 days.
4. *PJOs* are identified by applying a slight variation of criteria established by Hitchcock et al. (2013), as the original required finer vertical resolution than available. The new metric has been validated in reanalysis to ensure that similar results are obtained in this case as to those obtained by applying the original one (not shown). Here, the identification is based on two time series $PC_1 = T'(5\ \text{hPa}) - T'(100\ \text{hPa})$ and $PC_2 = T'(50\ \text{hPa})$, where T' indicates the polar-cap-averaged temperature anomaly (from climatology) at the specified pressure level. These time series are transformed into polar coordinates r(t) and phi(t), and the central dates of events are defined by when the phase phi(t) passes counterclockwise through $3\pi/2$, so long as the amplitude r(t) is greater than 2.5σ. Once a central date is defined, the starting date of the event is defined by the most recent date prior to the central date when r(t) is below 1.5σ, and similarly, the ending date of the event is defined by the earliest date following the central date when the r(t) is below 1.5σ.

2.3. Statistical Methods

Two methods to calculate the statistical significance of changes to the SSW frequency are used: a parametric method based on an assumption that the SSW frequency can be estimated using a Poisson point process and a nonparametric bootstrapping technique based on resampling the piControl run of each model. Trends in SSW frequency and the time of emergence of these trends are estimated by fitting a Generalized Linear Model to the decadal SSW frequency estimates from each model. All three statistical methods are described in detail in Appendix A.

3. Model Simulation of SSWs During the Historical Period: Mean Frequency and Seasonal Distribution

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

Mindful of this large internal variability, it appears that only four of the models are significantly different from JRA-55, at the 95% confidence level. Three of these are the models with the lowest model tops (CESM2, CanESM5, and GFDL-CM4) that simulate fewer SSW events than JRA-55 reanalysis. When comparing the seasonal distribution of SSW activity in these models with JRA-55 (Figure 2), it is clear that for two of them (GFDL-CM4 and CESM2), the SSW activity is significantly shifted toward March, with few SSWs observed in December and January. This is another common bias in low-top models (Charlton-Pérez et al., 2013) and, more generally, in models with an overly strong polar vortex. It is also worth noting that the three low-top models mentioned above are the only ones lacking a simulated Quasi-Biennial Oscillation (QBO). The fourth model with an unrealistic SSW frequency (IPSL-CM6A-LR), in contrast, simulates a very high number of SSWs, on average one per year during the historical period (instead of one every other year). As detailed below, this model also stands out for its high frequency of warmings in the piControl

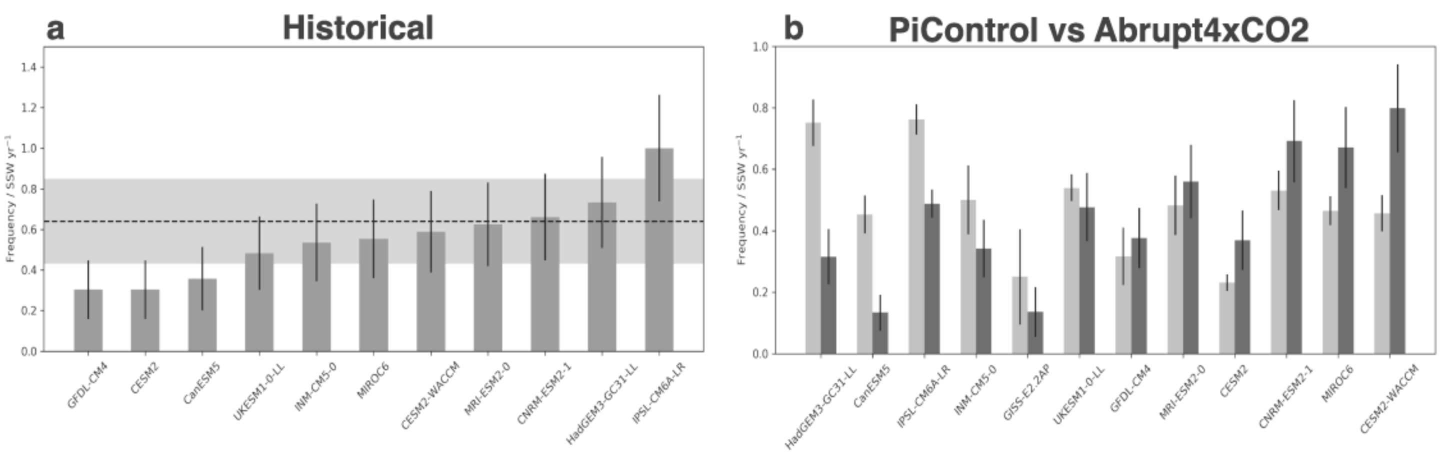

Figure 1. (a) Average annual SSW frequency in the historical simulations (1958–2014) of the CMIP6 models. Black lines show 95% confidence estimates for the annual frequency. Dashed black line corresponds to SSW frequency in the JRA-55 reanalysis, with its 95% confidence interval in the light gray shading. (b) Same as (a) but for SSW occurrence in the piControl (light gray bars) and abrupt4xCO_2 simulations (dark gray bars). Black lines show 95% confidence intervals for each estimate. Bars are ordered by the size of the difference between the two simulations.

run. While we retain these four models in our analysis, the simulated changes produced by these models should be treated with caution given these biases.

Finally, considering the surprising occurrence of an SSW in the SH in 2002 (Krüger et al., 2005), we extended the analysis to that hemisphere. Not a single SSW event was identified in the SH over the historical period in

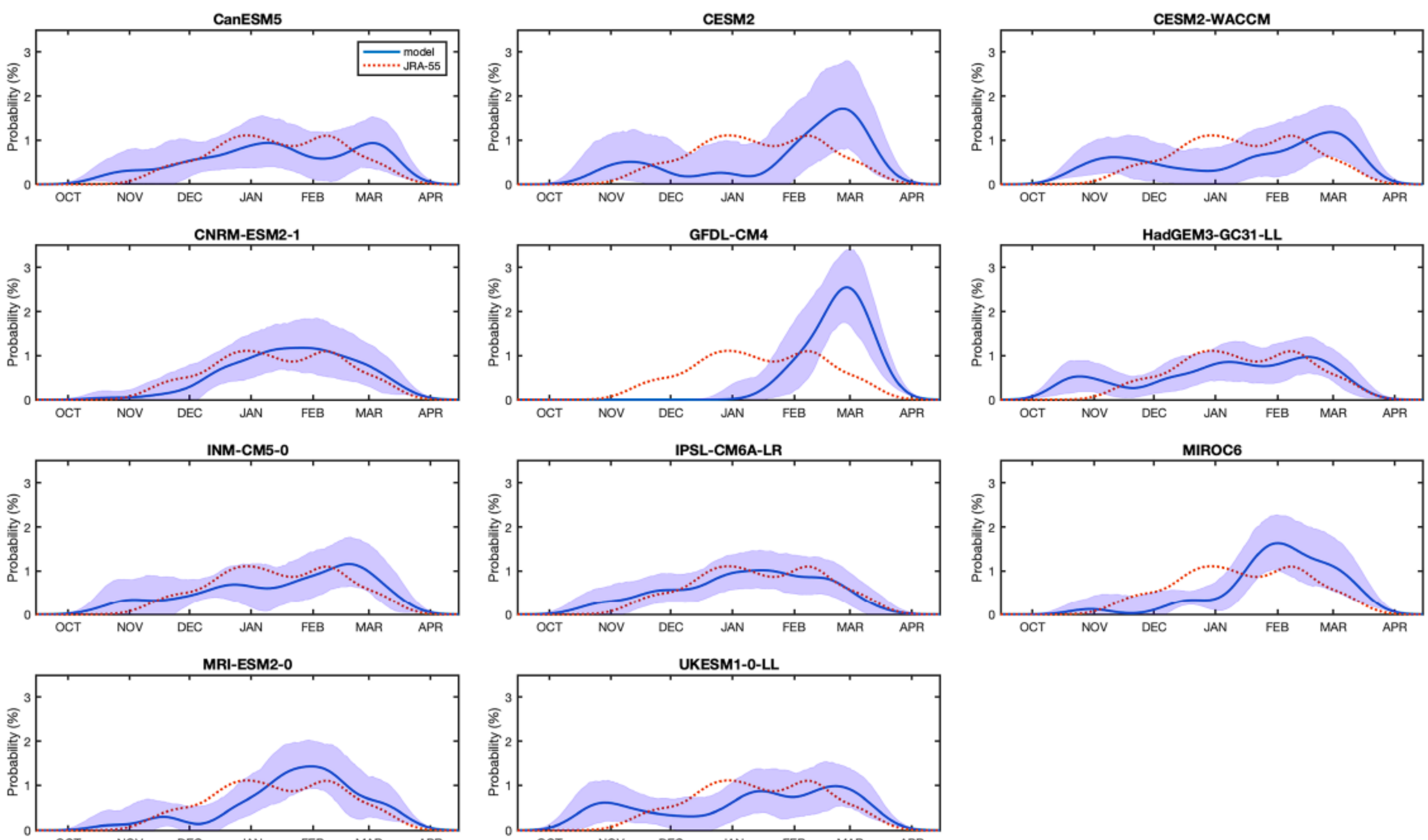

Figure 2. SSW frequency distribution in the historical simulation of each model (blue line) and JRA-55 reanalysis period (orange dashed line). The distribution has been smoothed by a kernel smoother of a bandwidth of 10 days. Shading corresponds to 2.5th–97.5th percentile range of the bootstrap samples, that is, the 95% confidence interval on the mean of the piControl simulation. (See more details about the determination of this interval in Appendix A1.2).

the models analyzed here. One may be tempted to claim that the CMIP6 models are underestimating the stratospheric variability in the SH, as spontaneous SSWs in the absence of stationary waves have been reported in simple models (Kushner & Polvani, 2005). However, it remains to be demonstrated whether five or six decades of observations are sufficient to make that claim.

4. Future Changes in Polar Stratospheric Variability

4.1. Future Changes in SSWs

Figure 1b displays the mean frequency of SSWs in both the piControl and abrupt4xCO_2 simulations (numerical values in Table S2). As discussed in section 2, all SSWs identified in the entire abrupt4xCO_2 simulation have been considered. We stress, however, that the main results presented below do not change significantly if only the second 75 years of each abrupt4xCO2 simulation are used (not shown). Two different tests of the statistical significance of the changes are conducted, providing a consistent indication of the statistical significance of changes, although the precise *p*- values vary due to difference in the underlying assumptions.

Of the 12 models in our study, four models indicate a statistically significant *decrease* in SSW frequency, while four indicate a statistically significant *increase* in SSW frequency. Thus, no consensus in the sign of the change exists in the CMIP6 models, in agreement with the diversity of claims reported in the earlier literature. The lack of a robust change across the models is not due to a lack of sensitivity of SSW frequency to increasing CO_2: In fact, 8 of the 12 models indicate significant changes. Rather, the CMIP6 models suggest that there is a great deal of uncertainty in the sign of the change, which varies between a near doubling in the frequency of SSWs in some models and a near halving in others. These divergent responses of the models may now be clearer in the CMIP6, where we can consider a stronger forcing (4xCO_2) and have access to longer records of daily data, compared to previous studies.

We also note that the lack of consensus in the CMIP6 models agrees with the recent study of Ayarzagüena et al. (2018), who analyzed the chemistry climate model projections of the CCMI models, which were forced with RCP6.0 scenario. While reporting a general tendency toward an increased frequency of SSWs by the end of the current century, they also emphasized that most changes were *not* statistically significant.

We do not attempt to further analyze the causes of differences in the model responses here, other than to note that within our set of models, one of the models indicating a significant reduction of SSW frequency (CanESM5) and one of the models indicating a significant increase of SSW frequency (CESM2) have anomalously low SSW frequency and (in the case of CESM2) a biased seasonal distribution of SSW in the historical simulations (Figure 2). Additionally, two models which show significant decreases in SSW frequency (HadGEM3-GC31-LL and IPSL-CM6A-LR) have the highest frequency of SSW events in the piControl and historical simulations. The IPSL-CM6A-LR has a significant bias in SSW frequency and presents some strong biases in the representation of QBO in the abrupt4xCO_2 simulation. Nevertheless, even if we did not consider the four models with biases in the representation of SSWs in the historical period (CanESM5, CESM2, IPSL-CM6A-LR, and GFDL-CM4), the main conclusion on the uncertainty in the sign of SSW changes would remain the same.

We also briefly examined the relationship between the change in SSW frequency and possible predictors of the change, including the frequency of SSWs in the piControl and historical simulations and the Effective Climate Sensitivity (ECS; Gregory et al., 2004) (Figure 3). Recall that ECS gives a measure of the equilibrium change of the global surface temperature after a doubling of CO_2. As can be seen from Figure 3, models that have a larger frequency of SSWs in the piControl run and models that have a larger ECS seem to produce large reductions in SSW frequency under large CO_2 forcing. A notable outlier from the main relationship here is the GISS-E2.2AP model but note that shorter simulations are available for this model than for others in the ensemble which also means that the uncertainty on the estimate of the piControl SSW frequency for this model is large.

Excluding GISS-E2.2AP, the correlation between SSW frequency changes and ECS is −0.52 with a probability value of obtaining results at least as extreme as the computed correlation (*p*value) of 0.12. However, with GISS-E2.2AP included in the ensemble, the correlation drops to −0.33 and is not significant. The correlation between piControl frequency and SSW frequency changes is −0.50 with a *p*value of 0.10 with all models

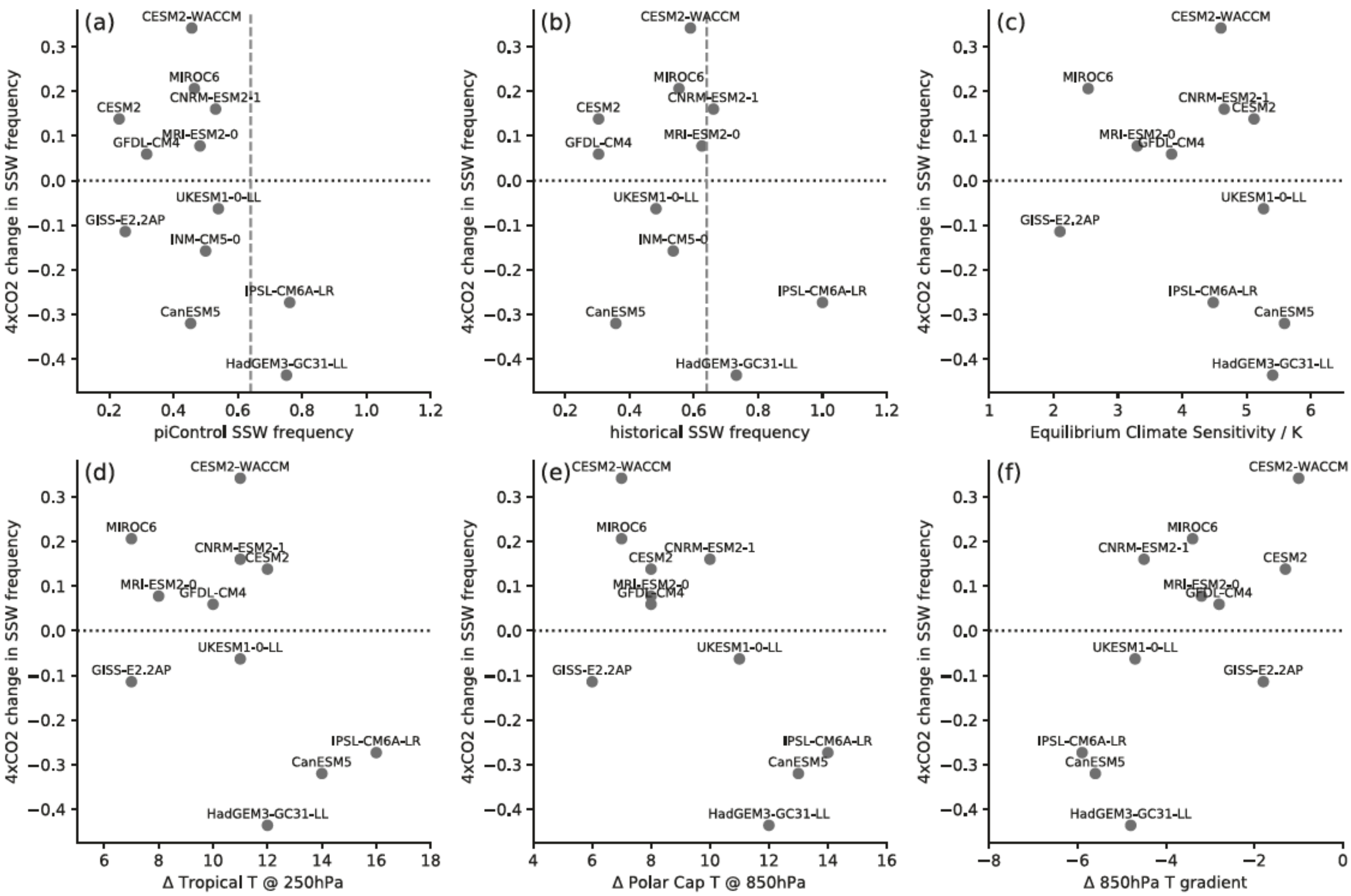

Figure 3. Scatter plots of the change of SSW frequency between the piControl and abrupt4xCO_2 simulations versus (a) the frequency in the piControl simulations, (b) the frequency in the historical simulations, (c) the ECS, (d) the change in tropical temperature at 250 hPa, (e) the change in polar temperature at 850 hPa, and (f) the difference in tropical-polar temperature difference at 850 hPa. In (a) and (b), the gray dashed line shows the observed SSW frequency in the JRA-55 reanalysis (0.64 SSW per year). The temperature regions in (d)–(f) are defined as in Harvey et al. (2014).

included. Further analysis of a larger ensemble would be required to determine the robustness of these relationships.

Although not addressed in the literature, a relationship between ECS and SSW frequency changes might be possible given some previous results connected to this topic. Shepherd and McLandress (2011) and Grise and Polvani (2016) documented a link between the strengthening of the subtropical jet and stratospheric wave driving and between ECS and dynamical changes, respectively. Moreover, Li et al. (2007) have argued that the subtropical jet, and tropospheric state in general, might control the upward planetary wave propagation. In this sense, the meridional gradient of the upper tropospheric temperature in the piControl simulation (computed as in Harvey et al., 2014) was found to be linked to the SSW frequency changes under high CO_2 concentrations. The correlation between both variables is −0.61 (*p*value 0.04). Thus, a model bias in the tropospheric state affects the stratospheric response to increasing CO_2, probably due to its effects on wave propagation. In addition, an intriguing examination of the relationship between changes in the tropospheric state and SSW frequency is shown in the bottom row of panels of Figure 3. Again, GISS-E2.2AP is an outlier in Figures 3c–3f. Excluding, GISS-E2.2AP, there is a significant correlation between changes in SSWs and changes in the polar lower tropospheric temperature (−0.89, *p*value < 0.01) and the lower tropospheric temperature gradient (0.79, *p*value < 0.01). In contrast, correlations between the upper tropospheric temperature changes and SSW frequency are generally smaller, with the highest correlation between the tropical upper tropospheric temperature change and SSW frequency change (−0.62, *p*value 0.06). With GISS-E2.2AP

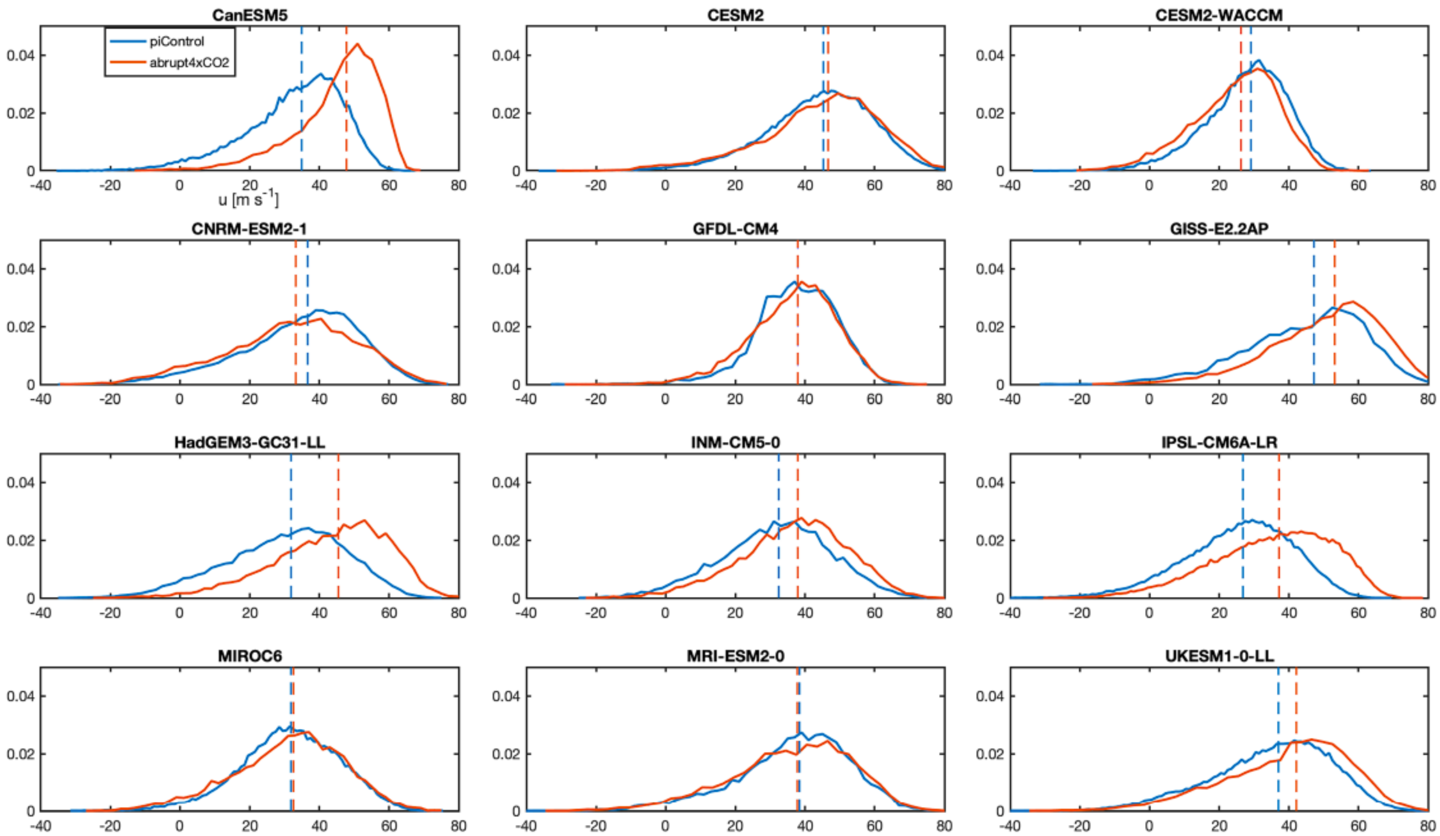

Figure 4. Probability distribution of daily zonal mean zonal wind at 60°N and 10 hPa (m/s) for the piControl (blue) and abrupt4xCO_2 (orange) experiments. Dashed lines represent the median value of the distribution in each integration.

included, the lower tropospheric correlations are reduced but have *p*values smaller than 0.05, while the correlation with tropical upper tropospheric temperature does not (−0.49, *p*value 0.12).

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

To further examine the changes in SSW frequency under 4xCO_2 forcing, we have analyzed the entire distribution of daily $u_{60N10hPa}$ in December-January-February in the piControl and abrupt4xCO_2 simulations (Figure 4). The four models with a significant decrease in SSWs frequency in Figure 1b (HadGEM3-GC31-LL, CanESM5, IPSL-CM6A-LR, and INM-CM5-0) are also those that show the largest shift of the $u_{60N10hPa}$ distribution toward stronger vortex speeds in the abrupt4xCO_2 experiment. Interestingly, the opposite does not always apply to models with a significant increase in SSWs. The models with the largest changes in SSW frequency, MIROC6 and CESM2-WACCM, show small changes to either the median or standard deviation of the $u_{60N10hPa}$ (Table S3). This would agree with the results of Taguchi (2017) who pointed out SSW frequency does not only correlate with vortex strength but also wave activity.

A similar analysis was repeated for the zonal-mean zonal wind at 10 hPa averaged between 70° and 80°N (not shown). That latitude band was found by Manzini et al. (2014) to display significant future changes in wind in most models, unlike the 60°N latitude where no robust future changes were found in CMIP5 models because the opposed effects of subtropical jet and stratospheric polar vortex changes might combine at that latitude. However, in our case, the main conclusions remain the same. Those models that show a shift of the $u_{60N10hPa}$ distribution toward stronger vortex speeds under 4xCO2 forcing also display a sharper peak

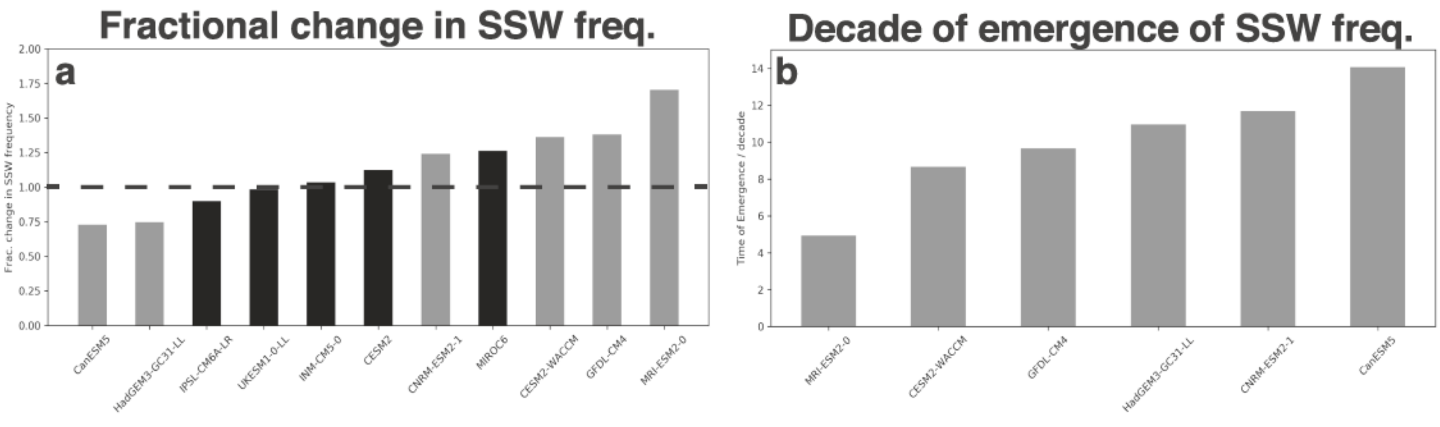

Figure 5. (a) Estimated fractional change in SSW frequency by the seventh decade of the 1pctCO_2 simulations. Light gray shaded lines indicate that the trend of SSW frequency in the model is significantly different from zero at a *p*value of 0.05. Dashed line indicates trend equal to 1, that is, no trend in the SSWs frequency. (b) Decade of emergence of SSW frequency trend for those models in which the trend term is significantly different from zero, calculated as described in the main text.

of high values of u at 70°–80°N suggesting lower variability in that region, consistent with a stronger and larger vortex.

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

Finally, motivated by the recent occurrence of a minor but highly publicized SSW event in the SH in September 2019 (Hendon et al., 2019), together with the occurrence of a major SSW in September 2002, we also examined the CMIP6 models to determine the extent to which the likelihood of similar events might change under the extreme climate forcing in the abrupt4xCO_2 runs. Only 1 of our 12 models (MRI-ESM 2-0) simulates an SSW in both the piControl and the abrupt4xCO_2 simulations. Thus, these runs provide no evidence for the claim of possible trends in the frequency of SSWs in the SH that would be caused by increased CO_2 concentrations.

4.2. Trends in SSW Frequency and Time of Emergence

For the model integrations which show a statistically significant increase or decrease in SSW frequency between the piControl and the abrupt4xCO_2 runs, it is useful to consider when and whether the trend in SSW frequency might be detected in a simulation with continuously increasing CO_2 forcing. A useful way to frame climate trends is in terms of the time of emergence of the signal from the unforced climate noise (Hawkins & Sutton, 2012). This question is examined by studying the occurrence of SSWs in the 1pctCO_2 runs, an idealized scenario.

Trend estimates for each model are shown in Figure 5a (numerical values in Table S4). Results reveal that there are six models (light gray bars) for which the null hypothesis of no trend in SSW frequency can be rejected, but consistent with the results of the previous section, the sign of this trend is not robust across models. While CanESM5 and HadGEM3-GC31-LL show a significant decrease, CNRM-ESM 2-1, CESM2-WACCM, GFDL-CM4, and MRI-ESM 2-0 show a significant increase. Recall that for the abrupt4xCO_2 runs (Figure 1b), CNRM-ESM 2-1 and CESM2-WACCM also indicated a statistically significant increase in SSW frequency compared to the piControl runs, while GFDL-CM4 and MRI-ESM 2-0 did not (although they did indicate an increased frequency). CanESM5 and HadGEM3-GC31-LL both showed a statistically significant decrease.

One can also estimate a time of emergence of the trend by comparing the trend in the 1pctCO2 runs with the natural variability in SSW frequency from the piControl run (see Appendix A.2 for details in the procedure). For the models with a significant trend, the decade of emergence is shown in Figure 5b. There is a widespread in the projected time of emergence for the models with a significant trend, varying from the fifth decade to fourteenth decade. This result reflects both the variation in the trend across the models and the spread

10.1029/2019JD032345

in the estimated variability in SSW frequency (the noise) in the piControl simulations. Since the time of CO_2 doubling occurs between the sixth and seventh decade in the 1pctCO_2 run and approximately by 2060–2070 in the RCP8.5 scenario (Meinshausen et al., 2017), these results indicate that the emergence of a detectable change in SSW frequency is extremely unlikely prior to the end of the 21st century.

5. Future Changes in the Seasonal Cycle of the Polar Stratosphere

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

Figures 6a and 6b show the distribution of dates of formation and final breakdown of the boreal stratospheric polar vortex, respectively, in the piControl, historical, and abrupt4xCO_2 CMIP6 simulations. In these plots the first years of the abrupt4xCO_2 simulations (75 or 300 years) have been omitted similar to the procedure followed to calculate the climatology. Nevertheless, conclusions do not change when considering the whole data record for abrupt4xCO_2 runs.

First, let us consider the historical model simulations and contrast them to the reanalysis. Over the period 1958–2014, the polar vortex forms earlier in all models than it does in the reanalysis, with the exception of IPSL-CM6A-LR. In contrast, the SFW date is well reproduced by models. The latter implies an improvement with respect to previous generations of climate models, such as those contributing to CCMVal and CMIP5, which simulated a delayed SFW (Butchart et al., 2011; Kelleher et al., 2019). CMIP6 models are also good at simulating the different range of interannual variability in the dates of vortex formation and SFW, the latter being considerably larger than the former.

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

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

A similar analysis has been performed for the SH. Because planetary wave activity is much weaker in the SH than in the NH (Andrews et al., 1987), radiative CO_2 forcing dominates the SH polar vortex response to increasing CO_2 concentrations and so causes a robust strengthening. In many models, the extreme CO_2 concentrations prevent the polar vortex from disappearing at all during austral summer, leading to perpetual westerly conditions in the stratosphere, so we do not show the results for the abrupt4xCO_2 simulation. The distribution of SFW dates for piControl and historical simulations is displayed in Figure 6c. Unlike in the NH, the distribution of SFWs in the SH already shifts toward a later date in the historical period with

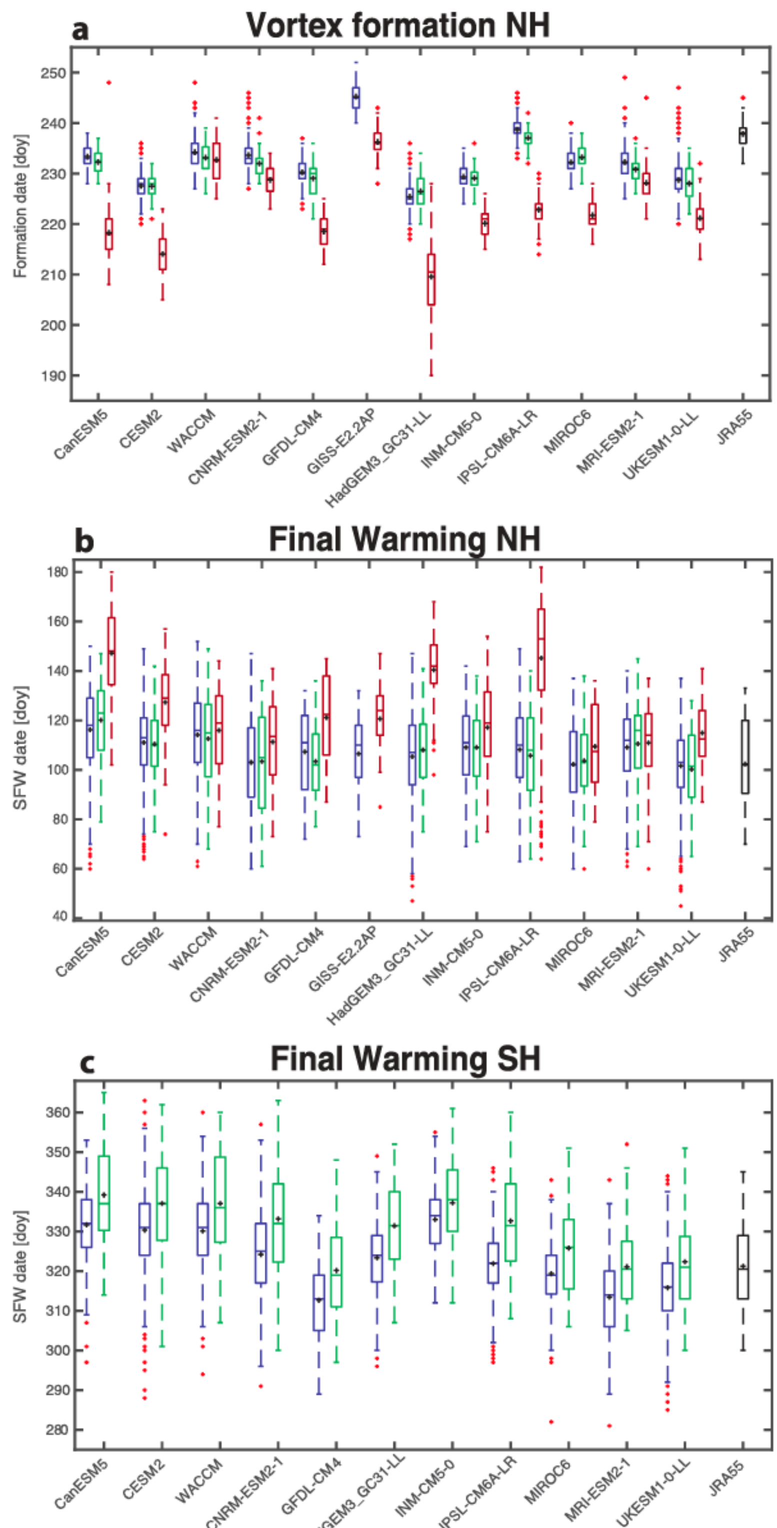

Figure 6. Box plots showing the distribution of dates of (a) polar vortex formation and (b) stratospheric final warming in the Northern Hemisphere for the piControl (blue), historical (green), and abrupt-4xCO_2 (red) simulations for all models and JRA-55 reanalysis. (c) Same as (b) but for the Southern Hemisphere and only in piControl and historical runs. The interquartile range is represented by the size of the box, and the inside line (black cross) corresponds to the median (mean). Whiskers indicate the maximum and minimum points in the distribution that are not outliers. Outliers (red crosses) are defined as points with values greater than 3/2 times the interquartile range from the ends of the box.

respect to the piControl conditions. Although the attribution of changes in the length of the winter season to CO_2 is complicated, ozone depletion in austral spring over the historical period might be responsible, based on previous literature (e.g., McLandress et al., 2010; Oberländer-Hayn et al., 2015).

6. Future Changes in the Surface Impact of SSW Events

6.1. Surface Response to SSW Events

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

To test this idea, we analyze first composite maps of anomalous surface temperature and sea-level pressure (SLP) for the period 15–60 days after SSWs in the piControl simulation (Figure 7). In nearly all models, we obtain the typical SLP and surface temperature patterns following SSWs that are also detected in reanalysis (although CO_2 forcing is different), that is, negative Northern Annular Mode pattern (particularly over the pole) and Eurasian cooling and Northeastern American warming. None of the models produce a positive SLP anomaly in the Pacific basin that can be found in the JRA55 composite though. Despite the relatively structural similarities across models, the amplitude of the response can vary by a factor of 2 or 3 between them. The amplitudes of SLP anomalies in 5 of the 11 models (CESM2-WACCM, GFDL-CM4, HadGEM3-GC31-LL, IPSL-CM6A-LR, and MIROC6) are too weak. Moreover, even the rest of the models that do a reasonable job of the polar cap SLP signal significantly underestimate the surface temperature response over the Labrador Sea and to the east of Greenland. This is consistent with Hitchcock and Simpson (2014) that argued the near-surface temperature response to SSW was underestimated in specific regions in CMIP5 models. The amplitude of the signal in the troposphere does not correlate with the SSW frequency. It is also not a problem of model biases in the simulation of SSWs mentioned in section 3 either. The large SSW sample size from the piControl simulations means that the estimates of surface impact are very robust.

Second, we compare the SLP pattern after SSWs in the abrupt4xCO2 and piControl simulations (Figure 8, differences in SLP between both runs are shown in shading). The overall SSW signal in SLP appears unchanged between the piControl and abrupt4xCO2 simulations, except in three models (CESM2, HadGEM3-GC31-LL, and IPSL-CM6A-LR) that produce a significantly stronger Northern Annular Mode-like response. However, in the Pacific basin, there are some indications about a potential more general change due to a higher CO_2 loading. Indeed, 6 of the 11 models exhibit a statistically stronger negative SLP anomaly in that area under abrupt4xCO_2 forcing than in the piControl runs. This could be related to some changes in the tropospheric precursors of SSWs because these anomalies have been identified as the remainder of the deepening of the Aleutian low preceding SSWs in observations (Ayarzagüena et al., 2019; Charlton & Polvani, 2007). Nevertheless, more work is required to understand all the details.

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

6.2. PJOs

In this subsection we focus on specific events (PJOs) that are closely related to SSWs and the stratosphere-troposphere coupling (Hitchcock et al., 2013). As indicated in section 1, their strong and persistent tropospheric response explains the interest in investigating possible changes in the occurrence of these events for increasing CO_2 concentrations.

First, examining the surface response to PJOs in the piControl experiment (Figure S1) confirms that these events in models have a stronger signal in the troposphere than all SSWs too. In JRA-55, roughly half of all SSWs are associated with a PJO event (PJO SSW) (solid line in Figure 9). Six models include the

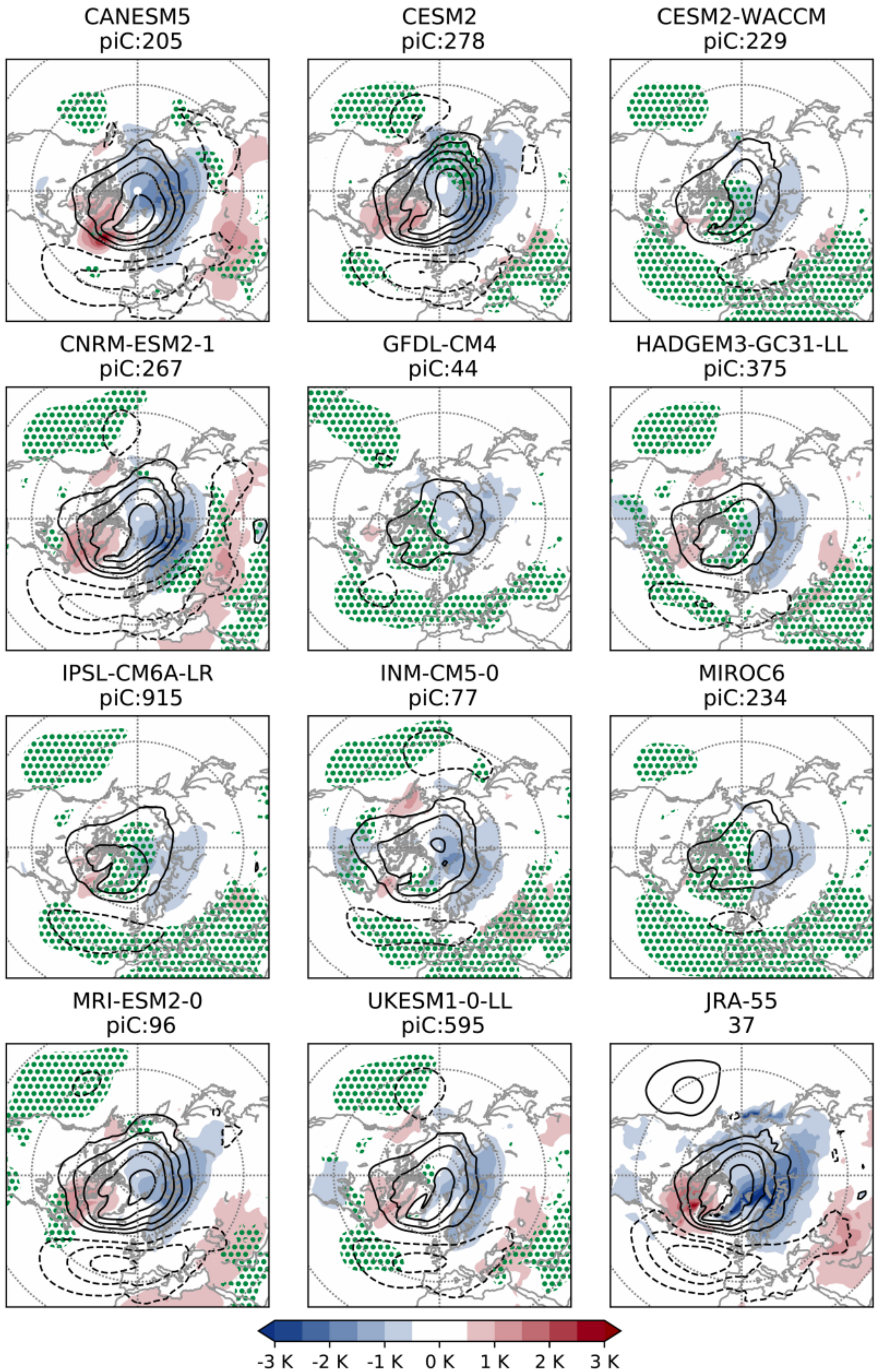

Figure 7. Composite maps of anomalous SLP (contour interval 1 hPa) and 2 m temperature (shading) for 15/60 days after SSWs in piControl simulation and JRA-55 reanalysis (bottom right). Green stippling indicates statistically significant differences in SLP from JRA-55 reanalysis at the 95% confidence level. Numbers in titles indicate the number of events considered.

10.1029/2019JD032345

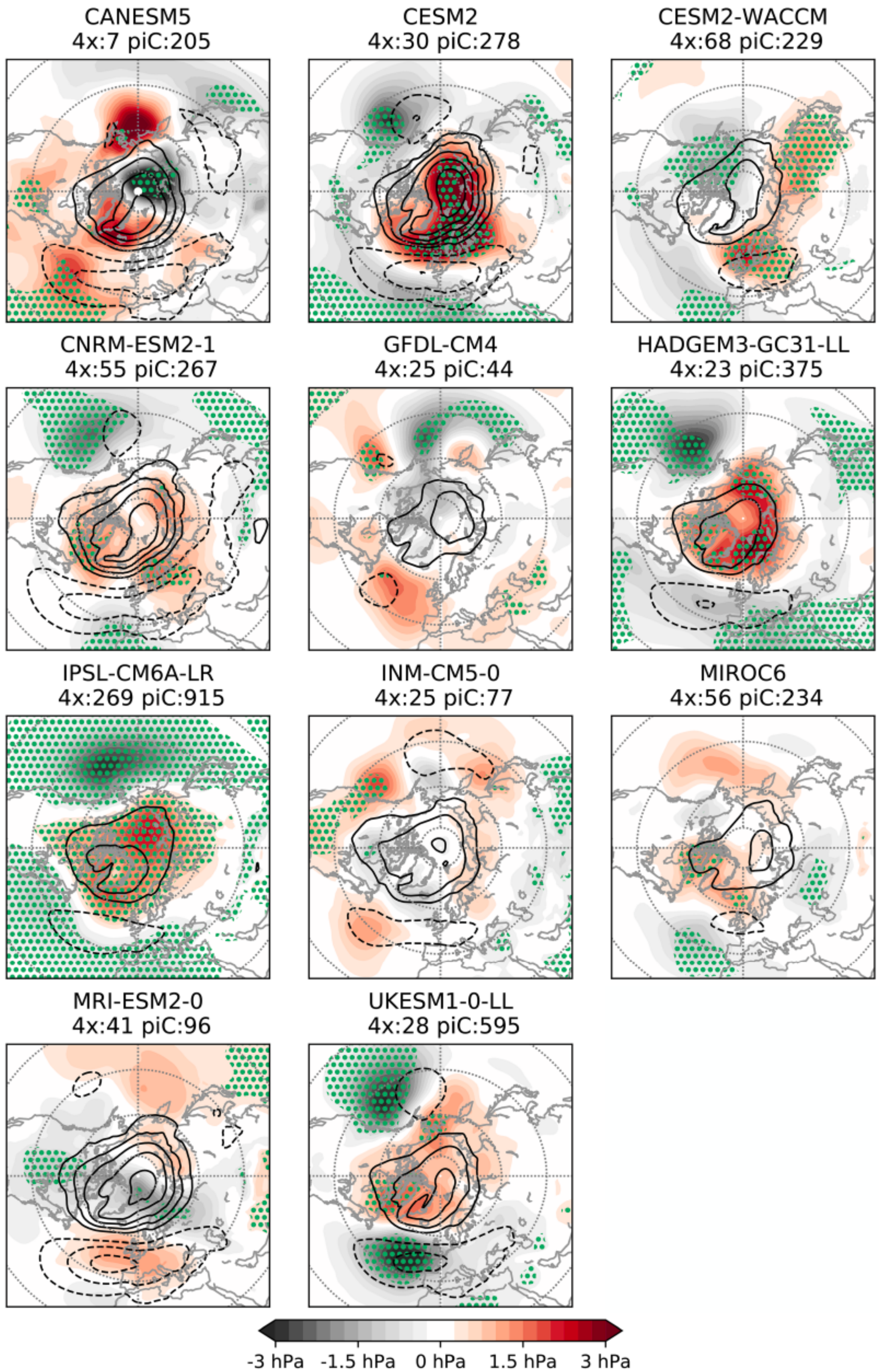

Figure 8. Abrupt4xCO_2-minus-piControl composite maps of anomalous SLP (shading, hPa) for 15/60 days after SSWs. Anomalous SLP after SSWs in piControl run is shown in contours (interval: 1 hPa). Green stippling indicates statistically significant differences from piControl run at the 95% confidence level. Numbers in titles indicate the number of events considered in the piControl (piC) and abrupt4xCO2 (4x) simulations.

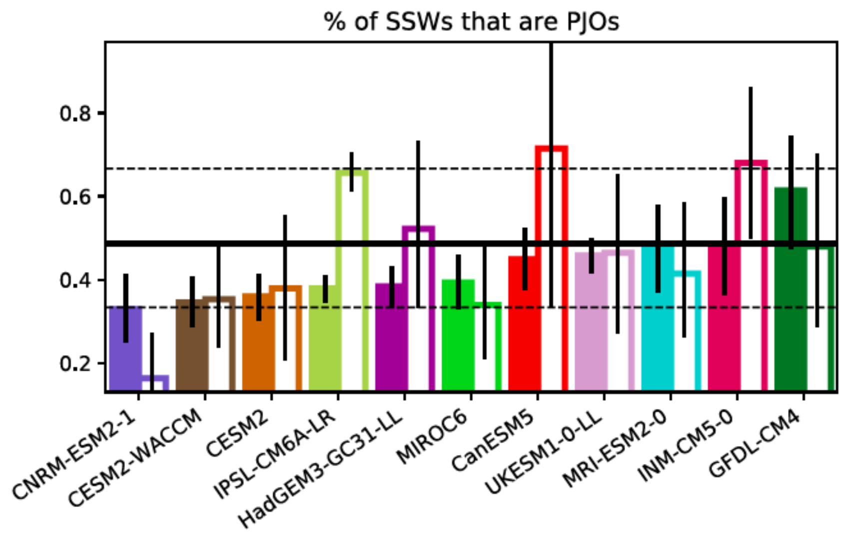

Figure 9. Fraction (%) of SSWs that are also PJOs in piControl (solid bars) and abrupt4xCO_2 (open bar) runs. Horizontal black solid and dashed line correspond to the mean value and the 2.5th–97.5th percentile range in JRA-55 reanalysis, respectively. Error bars are based on bootstrapping.

JRA-55 value of the ratio of PJO SSW events in their confidence interval in the piControl simulations (MRI-ESM 2-0, UKESM1-0-LL, CanESM5, HadGEM3-GC31-LL, INM-CM5-0, and GFDL-CM4). The other models underestimate this fraction. However, we do not find a clear relationship between this fraction and the amplitude of SLP pattern following SSWs. For instance, HadGEM3-GC31-LL and GFDL-CM4 simulate a very weak SLP pattern (Figure 7), but the ratio of PJO SSWs is close to observations or even larger.

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

7. Conclusions

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

From our analysis of the 12 models for which sufficient daily time resolution stratospheric data were available, these conclusions can be drawn about the impact of extreme CO_2 concentrations on SSW events:

1. There is no consensus among models on the sign of changes in SSW frequency to increase in CO_2 forcing.
2. It is, however, possible to say with confidence that many models predict that SSW frequency is sensitive to increase in CO_2 forcing.
3. There is no change to the impact of SSW events in the North Atlantic between the abrupt4xCO_2 and piControl simulations. In the North Pacific, there is some indication that under large CO_2 forcing, there will be a larger mean response to SSW events.

4. With the exception of MRI-ESM-2-0, predicted trends in SSW frequency are small relative to natural variability (as characterized by the piControl simulations of each model). This is not to say that SSW changes are themselves small (three models predict frequency changes of more than a factor of 2 compared to piControl conditions) but more a reflection of the large, natural decadal variability in SSW occurrence. As such, changes in SSW frequency are unlikely to be observed until the end of the 21st century.
5. Robust changes to the seasonal cycle in the stratosphere are predicted by all models. The stratospheric polar vortex is likely to form earlier and decay later in the future. This extends the season in which the stratosphere can actively couple to the troposphere and influence surface weather.
6. There is no evidence of an increased likelihood of major SSWs in the SH in the future.

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

Appendix A: Statistical Framework

A.1. Statistical Methodology for Comparing SSW Frequency

A.1.1. Parametric Method

To compare the frequency of SSW events in two models or between a model and observations, it can be assumed that each data sample is a Poisson process with an annual rate λ_i. The difference between the intensity of the two processes Δ_λ is given in equation (A1)

$$\Delta_\lambda = \frac{(\lambda_0 - \lambda_1)}{\sqrt{\frac{\lambda_0}{N_0} + \frac{\lambda_1}{N_1}}}. \tag{A1}$$

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

An alternative approach that compares the ratio of the rate of the two Poisson processes has been studied by Gu et al. (2008).

$$H_0{:}\lambda_0/\lambda_1 = 1 \qquad \text{against} \qquad H_A{:}\lambda_0/\lambda_1 \neq 1. \tag{A2}$$

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

$$W(X_0, X_1) = \frac{2\left[\sqrt{X_0 + \frac{3}{8}} - \sqrt{\rho\left(X_1 + \frac{3}{8}\right)}\right]}{\sqrt{1+\rho}}. \tag{A3}$$

The *p*value for this statistic is estimated as in equation (A4), where ϕ is the cumulative distribution function of the standard normal and the observed value of the test statistic $W(X_0, X_1) = w(x_0, x_1)$:

$$p = 1 - 2^* \Phi\left(w_j(x_0, x_1)\right). \tag{A4}$$

This is the parametric test statistic used to compare SSW frequency. In addition to calculating the *p*value of any test statistic, it is also useful, a priori, to estimate the statistical power of any testing framework. Tests with high statistical power minimize the likelihood of Type-II errors (i.e., that the null hypothesis is not rejected when it is, indeed, false). For the test statistic described above, we estimated the statistical power for a comparison with observations of 60 winters with an SSW frequency of 0.6. Assuming a *p*value of

0.05, the statistical power of the test is high (above 0.8) for model integrations of more than 100 winters (the null hypothesis will be rejected with a probability above 0.8) which is the case for all comparisons in this study apart from the comparison between the historical simulations and the JRA-55 reanalysis. In this later case, the power of the test is low only for cases in which the observed and modeled SSW frequency is very similar (i.e., for model SSW frequencies of 0.2 and 1 SSW per year, the power is greater than 0.8).

A.1.2. Bootstrapping Method

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

This bootstrapping technique was also applied to determine the confidence intervals on the seasonal distribution of SSW frequency. We choose to perform the bootstrapping on individual winters over a block bootstrapping approach to increase the sample size available for models that have a limited length of piControl simulation available. We have, therefore, assumed that there is no autocorrelation from one winter to the next, but comparison with a block-bootstrapping approach for the models that have long piControl simulations produced similar uncertainty ranges (not shown), indicating that this assumption is reasonable. For Figure 2, the uncertainty range is derived from the piControl simulation. Since there are 57 years in the JRA55 record, we resample 57 years from the piControl simulation, with replacement and recalculate the SSW distribution, normalized by the number of SSWs in that sample. This is repeated 1,000 times, and the uncertainty range shows the 2.5th to 97.5th percentile range of these 1,000 samples (95% confidence interval), that is, this is the uncertainty range from the model with an equivalent number of years to that of the observations.

A.2. Trend in SSW Frequency and Time of Emergence

Analogously to the method of Hawkins and Sutton (2012), the time of emergence of a "signal" in the frequency of SSW events is estimated by comparing the size of the trend in SSW frequency in the 1pctCO2 simulations with the "noise" determined from the piControl simulation of the same model.

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

$$F_{\mathrm{ssw}}(t) = e^{\beta_0+\beta_1 t}. \tag{A5}$$

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

Acknowledgments
B. A. was supported by the Spanish Ministry of Science, Innovation and Universities through the JeDiS (RTI2018-096402-B-I00) project. This research is part of POLARCSIC activities. The work of L. M. P. is funded, in part, by a grant from the U.S. National Science Foundation to Columbia University. N. B. was supported by the Met Office Hadley Centre Climate Programme funded by BEIS and Defra. E. P. G. acknowledges support from the NSF through Award AGS-1852727. E. M. acknowledges support from the Blue-Action project, Grant Agreement 727852, European Union's Horizon 2020 research and innovation programme. C. O. thanks the support and resources provided by the NASA Modeling, Analysis and Prediction program and the NASA High-End Computing (HEC) Program through the NASA Center for Climate Simulation (NCCS) at Goddard Space Flight Center. S. W. was supported by the "Integrated Research Program for Advancing Climate Models (TOUGOU Program)" from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT), Japan. We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP6. We thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP6 and ESGF. The authors thank the WCRP SPARC for support of the DynVar Activity, which spurred the formation of the DynVarMIP. CMIP6 data are allocated at the ESGF archive (https://esgf-node.llnl.gov/projects/cmip6/). The Japanese 55-year Reanalysis (JRA-55) project was carried out by the Japan Meteorological Agency (JMA). JRA-55 data were accessed through NCAR- UCAR Research Data Archive (https://rda.ucar.edu).

References

Andrews, D. G., Holton, J. R., & Leovy, C. B. (1987). *Middle atmosphere dynamics*. Orlando, FL: Academic.

Ayarzagüena, B., Langematz, U., Meul, S., Oberländer, S., Abalichin, J., & Kubin, A. (2013). The role of climate change and ozone recovery for the future timing of major stratospheric warmings. *Geophysical Research Letters*, *40*(10), 2460–2465. https://doi.org/10.1002/grl.50477

Ayarzagüena, B., Palmeiro, F. M., Barriopedro, D., Calvo, N., Langematz, U., & Shibata, K. (2019). On the representation of major stratospheric warmings in reanalyses. *Atmospheric Chemistry and Physics*, *19*, 9469–9484. https://doi.org/10.5194/acp-19-9469-2019

Ayarzagüena, B., Polvani, L. M., Langematz, U., Akiyoshi, H., Bekki, S., Butchart, N., et al. (2018). No robust evidence of future changes in major stratospheric sudden warmings: A multi-model assessment from CCMI. *Atmospheric Chemistry and Physics*, *18*(15), 11,277–11,287. https://doi.org/10.5194/acp-18-11277-2018

Ayarzagüena, B., & Serrano, E. (2009). Monthly characterization of the tropospheric circulation over the Euro-Atlantic area in relation with the timing of stratospheric final warmings. *Journal of Climate*, *22*, 6313–6324.

Baldwin, M. P., & Dunkerton, T. J. (2001). Stratospheric harbingers of anomalous weather regimes. *Science*, *294*, 581–584.

Bell, C. J., Gray, L. J., & Kettleborough, J. (2010). Changes in Northern Hemisphere stratospheric variability under increased CO_2 concentrations. *Quarterly Journal of the Royal Meteorological Society*, *136*, 1181–1190.

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

Black, R. X., McDaniel, B. A., & Robinson, W. A. (2006). Stratosphere-troposphere coupling during spring onset. *Journal of Climate*, *19*, 4891–49001.

Boos, D. D. (2003). Introduction to the bootstrap world. *Statistical Science*, *18*(2), 168–174.

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

Butchart, N., Austin, J., Knight, J. R., Scaife, A. A., & Gallani, M. L. (2000). The response of the stratospheric climate to projected changes in the concentration of well-mixed greenhouse gases from 1992 to 2051. *Journal of Climate*, *13*, 2141–2159.

Butchart, N., Charlton-Perez, A. J., Cionni, I., Hardiman, S. C., Haynes, P. H., Krüger, K., et al. (2011). Multimodel climate and variability of the stratosphere. *Journal of Geophysical Research*, *116*, D05102. https://doi.org/10.1029/2010JD014995

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

Butler, A. H., Seidel, D. J., Hardiman, S. C., Butchart, N., Birner, T., & Match, A. (2015). Defining sudden stratospheric warmings. *Bulletin of the American Meteorological Society*, *96*, 1913–1928. Butler, A. H., Sjoberg, J., Seidel, D., & Rosenlof, K. H. (2017). A sudden stratospheric warming compendium. *Earth System Science Data*, *9*, 63–76.

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

Charlton, A. J., Polvani, L. M., Perlwitz, J., Sassi, F., Manzini, E., Shibata, K., et al. (2007). A new look at stratospheric sudden warmings. Part II: Evaluation of numerical model simulations. *Journal of Climate*, *20*, 470–488.

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

Charlton-Pérez, A. J., Polvani, L. M., Austin, J., & Li, F. (2008). The frequency and dynamics of stratospheric sudden warmings in the 21st century. *Journal of Geophysical Research*, *113*, D16116. https://doi.org/10.1029/2007JD009571

Charney, J. G., & Drazin, P. G. (1961). Propagation of planetary-scale disturbances from the lower into the upper atmosphere. *Journal of Geophysical Research*, *66*, 83–109. https://doi.org/10.1029/JZ066i001p00083

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

Danabasoglu, G., Lamarque, J. F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., et al. (2020). The Community Earth System Model version2 (CESM2). *Journal of Advances in Modeling Earth Systems*, *12*(2). https://doi.org/10.1029/2019MS001916

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

Domeisen, D. I. V., Butler, A. H., Charlton-Perez, A. J., Ayarzagüena, B., Baldwin, M. P., Dunn-Sigouin, E., et al. (2019). The role of the stratosphere in subseasonal to seasonal prediction. Part I: Predictability of the stratosphere. *Journal of Geophysical Research: Atmospheres*, *125*. https://doi.org/10.1029/2019JD030920

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*, 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Fels, S. B., Mahlman, J. D., Schwarzkopf, M. D., & Sinclair, R. W. (1980). Stratospheric sensitivity to perturbations in ozone and carbon dioxide: Radiative and dynamical response. *Journal of the Atmospheric Sciences*, *37*, 2265–2297.

Garfinkel, C. I., Waugh, D. W., & Gerber, E. P. (2013). The effect of tropospheric jet latitude on coupling between the stratospheric polar vortex and the troposphere. *Journal of the Atmospheric Sciences*, *26*, 2077–2095.

Gerber, E., & Manzini, E. (2016). The Dynamics and Variability Model Intercomparison Project (DynVarMIP) for CMIP6: Assessing the stratosphere-troposphere system. *Geoscientific Model Development*, *9*, 3413–3425.

Gettelman, A., Mills, M. J., Kinnison, D. E., Garcia, R. R., Smith, A. K., Marsh, D. R., et al. (2019). The Whole Atmosphere Community Climate Model Version 6 (WACCM6). *Journal of Geophysical Research: Atmospheres*, *124*, 12,380–12,403. https://doi.org/10.1029/2019JD030943

Gregory, J. M., Ingram, W. J., Palmer, M. A., Jones, G. S., Stott, P. A., Thorpe, R. B., et al. (2004). A new method for diagnosing radiative forcing and climate sensitivity. *Geophysical Research Letters*, *31*, L03205. https://doi.org/10.1029/2003GL018747

Grise, K. M., & Polvani, L. M. (2016). Is climate sensitivity related to dynamical sensitivity? *Journal of Geophysical Research*, *121*, 5159–5176. https://doi.org/10.1002/2015JD024687

Gu, K., Tony Ng, H. K., Tang, M. L., & Schucany, W. R. (2008). Testing the ratio of two Poisson rates. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, *50*, 283–298.

Guo, H., John, J. G., Blanton, C., McHugh, C., Nikonov, S., Radhakrishnan, A., et al. (2018). NOAA-GFDL GFDL-CM4 model output. Version 20190718. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.1402

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

Harvey, B. J., Shaffrey, L. C., & Woollings, T. J. (2014). Equator-to-pole temperature differences and the extra-tropical storm track responses of the CMIP5 climate models. *Climate Dynamics*, *43*, 1171–1182. https://doi.org/10.1007/s00382-013-1883-9

Hawkins, E., & Sutton, R. (2012). Time of emergence of climate signals. *Geophysical Research Letters*, *39*, L01702. https://doi.org/10.1029/2011GL050087

Held, I. M., Guo, H., Adcroft, A., Dunne, J. P., Horowitz, L. W., Krasting, J., et al. (2019). Structure and Performance of GFDL's CM4.0 Climate Model. *Journal of Advances in Modeling Earth Systems*, *11*, 3691–3727. https://doi.org/10.1029/2019MS001829

Hendon, H., Thompson, D. W. J., Lim, E.-P., Butler, A. H., Newman, P. A., Coy, L., et al. (2019). Rare forecasted climate event under way in the Southern Hemisphere. *Nature*, *573*(7775), 496. https://doi.org/10.1038/d41586-019-02858-0

Hitchcock, P., Shepherd, T. G., & Manney, G. L. (2013). Statistical characterization of Arctic polar-night jet oscillation events. *Journal of Climate*, *26*, 2096–2116.

Hitchcock, P., & Simpson, I. R. (2014). The downward influence of stratospheric sudden warmings. *Journal of the Atmospheric Sciences*, *71*, 3856–3876. https://doi.org/10.1175/JAS-D-14-0012.1

Hu, J., Ren, R., & Xu, H. (2014). Occurrence of winter stratospheric sudden warming events and the seasonal timing of spring stratospheric final warming. *Journal of Climate*, *71*, 2319–2334.

Huffman, M. (1984). An improved approximate two-sample Poisson test. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, *33*(2), 224–226.

Karpechko, A. Y., & Manzini, E. (2012). Stratospheric influence on tropospheric climate change in the Northern Hemisphere. *Journal of Geophysical Research*, *117*, D05133. https://doi.org/10.1029/2011JD017036

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

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

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

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., et al. (2015). The JRA-55 reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan*, *93*, 5–48.

Kodera, K., Matthes, K., Shibata, K., Langematz, U., & Kuroda, Y. (2003). Solar impact on the lower mesospheric subtropical jet: A comparative study with general circulation model simulations. *Geophysical Research Letters*, *30*(6), 1315. https://doi.org/10.1029/2002GL016124

Krüger, K., Naujokat, B., & Labitzke, K. (2005). The unusual midwinter warming in the Southern Hemisphere stratosphere 2002: A comparison to Northern Hemisphere phenomena. *Journal of Climate*, *62*, 603–613.

Kuhlbrodt, T., Jones, C. G., Sellar, A., Storkey, D., Blockley, E., Stringer, M., et al. (2018). The low-resolution version of HadGEM3 GC3.1: Development and evaluation for global climate. *Journal of Advances in Modeling Earth Systems*, *10*(11), 2865–2888. https://doi.org/10.1029/2018MS001370

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

Kushner, P. J., & Polvani, L. M. (2005). A very large, spontaneous stratospheric sudden warming in a simple AGCM: A prototype for the Southern Hemisphere warming of 2002? *Journal of the Atmospheric Sciences*, *62*, 890–897.

Li, Q., Graf, H.-F., & Giorgetta, M. A. (2007). Stationary planetary wave propagation in the Northern Hemisphere winter—Climatological analysis of the refractive index. *Atmospheric Chemistry and Physics*, *7*, 183–200.

Mahfouf, J. F., Cariolle, D., Royer, J.-F., Geleyn, J.-F., & Timbal, B. (1994). Response of the Météo-France climate model to changes in CO_2 and sea surface temperature. *Climate Dynamics*, *9*, 345–362.

Manabe, S., & Wetherland, R. T. (1967). Thermal equilibrium of the atmosphere with a given distribution of relative humidity. *Journal of the Atmospheric Sciences*, *24*, 241–259.

Manzini, E., Karpechko, A. Y., Anstey, J., Baldwin, M. P., Black, R. X., Cagnazzo, C., et al. (2014). Northern winter climate change: Assessment of uncertainty in CMIP5 projections related to stratosphere-troposphere coupling. *Journal of Geophysical Research: Atmospheres*, *119*, 7979–7998. https://doi.org/10.1002/2013JD021403

McLandress, C., Jonsson, A. I., Plummer, D. A., Reader, M. C., Scinocca, J. F., & Shepherd, T. G. (2010). Separating the dynamical effects of climate change and ozone depletion. Part I: Southern Hemisphere stratosphere. *Journal of Climate*, *23*, 5002–5020.

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

Meinshausen, M., Vogel, E., Nauels, A., Lorbacher, K., Meinshausen, N., Etheridge, D. M., et al. (2017). Historical greenhouse gas concentrations for climate modelling (CMIP6). *Geoscientific Model Development*, *10*(5), 2057–2116. https://doi.org/10.5194/gmd-10-2057-2017

Mitchell, D. M., Charlton-Perez, A. J., Gray, L. J., Akiyoshi, H., Butchart, N., Hardiman, S. C., et al. (2012). The nature of Arctic polar vortices in chemistry-climate models. *Quarterly Journal of the Royal Meteorological Society*, *138*, 1681–1691.

Mitchell, D. M., Osprey, S. M., Gray, L. J., Butchart, N., Hardiman, S. C., Charlton-Perez, A. J., & Watson, P. (2012). The effect of climate change on the variability of the Northern Hemisphere stratospheric polar vortex. *Journal of the Atmospheric Sciences*, *69*, 2608–2618.

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

Oberländer-Hayn, S., Meul, S., Langematz, U., Abalichin, J., & Haenel, F. (2015). A chemistry-climate model study of past changes in the Brewer-Dobson circulation. *Journal of Geophysical Research: Atmospheres*, *120*, 6742–6757. https://doi.org/10.1002/2014JD022843

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

Rind, D., Shindell, D., Lonergan, P., & Balachandran, N. K. (1998). Climate change and the middle atmosphere. Part III: The doubled CO_2 climate revisited. *Journal of Climate*, *11*, 876–894.

Rind, D., Suozzo, R., Balachandran, N. K., & Prather, M. J. (1990). Climate change and the middle atmosphere. Part I: The doubled CO_2 climate. *Journal of the Atmospheric Sciences*, *47*, 475–494.

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

Scaife, A., Spangehl, T., Fereday, D. R., Cubasch, U., Langematz, U., Akiyoshi, H., et al. (2012). Climate change projections and stratosphere–troposphere interaction. *Climate Dynamics*, *38*(9-10), 2089–2097. https://doi.org/10.1007/s00382-011-1080-7

Schoeberl, M. R., & Hartmann, D. L. (1991). The dynamics of the stratospheric polar vortex and its relation to springtime ozone depletions. *Science*, *251*(4989), 46–52. https://doi.org/10.1126/science.251.4989.46

Séférian, R. (2018). CNRM-CERFACS CNRM-ESM 2-1 model output prepared for CMIP6 CMIP for experiment piControl-spinup. Version 20180423. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.4169

Séférian, R., Nabat, P., Michou, M., Saint-Martin, D., Voldoire, A., Colin, J., et al. (2019). Evaluation of CNRM Earth-System model, CNRM-ESM 2-1: Role of Earth system processes in present-day and future climate. *Journal of Advances in Modeling Earth Systems*, *11*, 4182–4227. https://doi.org/10.1029/2019MS001791

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

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

Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., et al. (2019a). The Canadian Earth System Model version 5 (CanESM5.0.3). *Geoscientific Model Development*, *12*, 4823–4873. https://doi.org/10.5194/gmd-12-4823-2019

Swart, N. C., Cole, J. N.S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., et al. (2019b). 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. *Journal of Geophysical Research: Atmospheres*, *122*, 5144–5156. https://doi.org/10.1002/2016JD025826

Tang, Y. et al. (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., Ogura, T., Nitta, T., Komuro, Y., Ogochi, K., Takemura, T., et al. (2019). Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. *Geoscientific Model Development*, *12*(7), 2727–2765. https://doi.org/10.5194/gmd-12-2727-2019

Tatebe, H., & 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

Volodin, E. M., Mortikov, E. V., Kostrykin, S. V., Galin, V. Y., Lykossov, V. N., Gritsun, A. S., et al. (2017). Simulation of the present day climate with the climate model INMCM5. *Climate Dynamics*, *49*(11-12), 3715–3734. https://doi.org/10.1007/s00382-017-3539-7

Williams, K., Copsey, D., Blockley, E., Bodas-Salcedo, A., Calvert, D., Comer, R., et al. (2018). The Met Office global coupled model 3.0 and 3.1 (GC3. 0 and GC3. 1) configurations. *Journal of Advances in Modeling Earth Systems*, *10*(2), 357–380. https://doi.org/10.1002/2017MS001115

Wu, Y., Simpson, I. R., & Seager, R. (2019). Inter-model spread in the Northern Hemisphere stratospheric polar vortex response to climate change in the CMIP5 models. *Geophysical Research Letters*, *46*, 13,290–13,298. https://doi.org/10.1029/2019GL085545

Yukimoto, S., Kawai, H., Koshiro, T., Oshima, N., Yoshida, K., Urakawa, S., et al. (2019). The Meteorological Research Institute Earth System Model Version 2.0, MRI-ESM 2.0: Description and Basic Evaluation of the Physical Component. *Journal of the Meteorological Society of Japan*, *97*(5), 931–965. https://doi.org/10.2151/jmsj.2019-051

Yukimoto, S. Koshiro, T., Kawai, H., Oshima, N., Yoshida, K., Urakawa, S., et al. (2019). MRI MRI-ESM 2.0 model output prepared for CMIP6 CMIP. Version 20190726.Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.621

Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., et al. (2020). Causes of higher climate sensitivity in CMIP6 models. *Geophysical Research Letters*, *47*(1). https://doi.org/10.1029/2019GL085782