JGR Atmospheres

RESEARCH ARTICLE
10.1029/2022JD037123

Comparison of Arctic and Antarctic Stratospheric Climates in Chemistry Versus No-Chemistry Climate Models

Olaf Morgenstern[1], Douglas E. Kinnison[2], Michael Mills[2], Martine Michou[3], Larry W. Horowitz[4], Pu Lin[5], Makoto Deushi[6], Kohei Yoshida[6], Fiona M. O'Connor[7], Yongming Tang[7], N. Luke Abraham[8,9], James Keeble[8,9], Fraser Dennison[10], Eugene Rozanov[11,12,13], Tatiana Egorova[11], Timofei Sukhodolov[11,13,14], and Guang Zeng[1]

[1]National Institute of Water and Atmospheric Research, Wellington, New Zealand, [2]National Center for Atmospheric Research, Boulder, CO, USA, [3]Centre National des Recherches Météorologiques (CNRM), Toulouse, France, [4]Geophysical Fluid Dynamics Laboratory (GFDL), National Oceanic and Atmospheric Administration, Princeton, NJ, USA, [5]Program in Atmospheric and Oceanic Sciences, Princeton University, Princeton, NJ, USA, [6]Meteorological Research Institute (MRI), Japan Meteorological Agency, Tsukuba, Japan, [7]Hadley Centre, MetOffice, Exeter, UK, [8]Department of Chemistry, University of Cambridge, Cambridge, UK, [9]National Centre for Atmospheric Science, University of Cambridge, Cambridge, UK, [10]Commonwealth Scientific and Industrial Research Organization, Aspendale, VIC, Australia, [11]Physikalisch-Meteorologisches Observatorium, World Radiation Centre, Davos, Switzerland, [12]Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland, [13]St. Petersburg State University, St. Petersburg, Russia, [14]Institute of Meteorology and Climatology, University of Natural Resources and Life Sciences, Vienna, Austria

Key Points:

- Coupling in ozone chemistry causes an increase in persistence of low temperature anomalies over both poles
- In the Antarctic, coupling in chemistry amplifies pre-existing stratospheric cold biases
- These effects can be masked by other dynamical differences present in some models

Correspondence to:

O. Morgenstern,
olaf.morgenstern@niwa.co.nz

Citation:

Morgenstern, O., Kinnison, D. E., Mills, M., Michou, M., Horowitz, L. W., Lin, P., et al. (2022). Comparison of Arctic and Antarctic stratospheric climates in chemistry versus no-chemistry climate models. *Journal of Geophysical Research: Atmospheres*, *127*, e2022JD037123. https://doi.org/10.1029/2022JD037123

Received 15 MAY 2022
Accepted 16 SEP 2022

Author Contributions:

Conceptualization: Olaf Morgenstern
Data curation: Olaf Morgenstern
Formal analysis: Olaf Morgenstern
Funding acquisition: Olaf Morgenstern, Douglas E. Kinnison, Michael Mills, Martine Michou, Larry W. Horowitz, Pu Lin, Makoto Deushi, Kohei Yoshida, Fiona M. O'Connor, Yongming Tang, N. Luke Abraham, James Keeble, Fraser Dennison, Eugene Rozanov, Tatiana Egorova, Timofei Sukhodolov, Guang Zeng
Investigation: Olaf Morgenstern
Methodology: Olaf Morgenstern

© 2022 Commonwealth of Australia and National Institute of Water and Atmospheric Research. This article is published with the permission of the Controller of HMSO and the King's Printer for Scotland. This article is a U.S. Government work and is in the public domain in the USA.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Abstract Using nine chemistry-climate and eight associated no-chemistry models, we investigate the persistence and timing of cold episodes occurring in the Arctic and Antarctic stratosphere during the period 1980–2014. We find systematic differences in behavior between members of these model pairs. In a first group of chemistry models whose dynamical configurations mirror their no-chemistry counterparts, we find an increased persistence of such cold polar vortices, such that these cold episodes often start earlier and last longer, relative to the times of occurrence of the lowest temperatures. Also the date of occurrence of the lowest temperatures, both in the Arctic and the Antarctic, is often delayed by 1–3 weeks in chemistry models, versus their no-chemistry counterparts. This behavior exacerbates a widespread problem occurring in most or all models, a delayed occurrence, in the median, of the most anomalously cold day during such cold winters. In a second group of model pairs there are differences beyond just ozone chemistry. In particular, here the chemistry models feature more levels in the stratosphere, a raised model top, and differences in non-orographic gravity wave drag versus their no-chemistry counterparts. Such additional dynamical differences can completely mask the above influence of ozone chemistry. The results point toward a need to retune chemistry-climate models versus their no-chemistry counterparts.

Plain Language Summary Ozone is a chemical constituent of the atmosphere acting as an absorber of both solar ultraviolet light and infrared radiation emitted by the Earth. It therefore needs to be considered in climate models. Explicit ozone chemistry is a computationally challenging addition to a climate model; hence in most cases ozone is simply prescribed. Especially during relatively cold stratospheric winter/spring seasons, Antarctic and Arctic ozone depletion can be considerable. Such anomalous ozone loss is not reflected in the imposed ozone field, and hence differences in behavior are expected for such situations between chemistry- and no-chemistry models. Indeed for such cold winters/springs, we find an enhanced persistence of such cold spells in a set of chemistry-climate models, versus their no-chemistry counterparts; such enhanced persistence generally makes the chemistry model less realistic than its no-chemistry counterpart. However, if there are substantial further differences between the members of these model pairs, such as regarding their grid configuration or physical processes beyond chemistry, these can obscure the effect of ozone chemistry. We thus claim that adding stratospheric ozone chemistry to a climate model necessitates retuning to counteract a deterioration of the simulated stratospheric climate that can otherwise occur.

1. Introduction

Climate feedbacks involving ozone have long been known to be important in large-scale climate change. Most notably, stratospheric ozone depletion has been linked to a strengthening of the Southern Annular Mode (SAM)

Software: Olaf Morgenstern, Douglas E. Kinnison, Michael Mills, Martine Michou, Larry W. Horowitz, Pu Lin, Makoto Deushi, Kohei Yoshida, Fiona M. O'Connor, Yongming Tang, N. Luke Abraham, James Keeble, Fraser Dennison, Eugene Rozanov, Tatiana Egorova, Timofei Sukhodolov, Guang Zeng
Visualization: Olaf Morgenstern
Writing – original draft: Olaf Morgenstern, Douglas E. Kinnison, Michael Mills, Martine Michou, Larry W. Horowitz, Pu Lin, Makoto Deushi, Kohei Yoshida, Fiona M. O'Connor, Yongming Tang, N. Luke Abraham, James Keeble, Fraser Dennison, Eugene Rozanov, Tatiana Egorova, Timofei Sukhodolov, Guang Zeng
Writing – review & editing: Olaf Morgenstern

since roughly the 1970s (Son et al., 2010; Fogt & Marshall, 2020, and references therein). Ozone depletion of the Antarctic polar vortex in spring drives a cooling of this airmass, stabilizing the vortex, delaying the transition to summertime easterlies, and via deep coupling causing a strengthening of the Southern Annular Mode (SAM) during southern summer (Morgenstern, 2021; Thompson et al., 2011). In the Arctic, ozone depletion is usually less pronounced than in the Antarctic (although recent years have seen two Arctic "ozone holes"; Kuttippurath et al., 2021), residual ozone is larger, and consequently ozone depletion has not been implicated in a long-term strengthening of the Northern Annular Mode (NAM; Eyring et al., 2021). However, large ozone depletion does tend to be followed by anomalous tropospheric weather, that is, an anomalously strong NAM (Friedel et al., 2022; Ivy et al., 2017). The pertinent observed long-term strengthening of the NAM however remains unexplained (Eyring et al., 2021).

Climate models regularly simulate a delayed breakdown of the polar vortex. This behavior leads to exaggerated stratospheric cooling following ozone depletion, driven by biases in the dynamical responses to ozone depletion (Lin et al., 2017). Also in some single-model studies, ozone chemistry has been found to impact timescales of variability of the polar vortices (Haase & Matthes, 2019; Oehrlein et al., 2020; Rieder et al., 2019). We will investigate whether these findings apply to present-generation climate models as a group as these models transition from almost all excluding to in the future increasingly including explicit ozone chemistry.

At the time of writing, the portal of the sixth Coupled Model Intercomparison Project (CMIP6) lists 120 models and model variants. Morgenstern (2021) uses 29 different models in his assessment of the SAM in CMIP6, essentially sidelining many model variants to reduce redundancy. Of these 29 models, only six have explicit interactive ozone chemistry. A feature of CMIP6 is that pairs of models have participated with interactive ozone chemistry constituting the main or only point of difference between them. Simulations performed by these model pairs thus offer an opportunity to assess what the impact is of interactive chemistry versus the alternative approaches, that is, usually prescribing the pre-computed CMIP6 ozone climatology (Checa-Garcia et al., 2018). A comparison of such model pairs will of course not only find impacts due to interactive ozone – or lack thereof – but would also be sensitive to any peculiarities of the precomputed ozone field itself, its implementation (Hardiman et al., 2019), and any differences versus the interactive ozone. For example, Morgenstern et al. (2020, 2021) have shown that the recommended CMIP6 ozone climatology (Checa-Garcia et al., 2018) greatly underestimates Northern-Hemisphere mean ozone loss over the period 1979–2000. Also in a few cases there are other differences between these pairs beyond ozone chemistry which can complicate this comparison. In a recent study Lin and Ming (2021) find substantially enhanced cooling in a model variant with interactive ozone versus the same model using prescribed ozone, even though the simulated and prescribed ozone are quite similar. The authors explain this as the effect of co-variance of ozone and temperature anomalies that does not exist in the no-chemistry model.

In the below we will compare simulations of pairs of CMIP6 models (supplemented with three non-CMIP6 models) with and without interactive ozone, and will assess differences between the two members of the pair regarding polar stratospheric dynamics and associated stratosphere-troposphere coupling. Where significant, such differences will be indicative of the role of climate-ozone coupling. We will assess both hemispheres, noting that Morgenstern (2021) has already made the case, using CMIP6 simulations, for why interactive ozone is important for simulating climate trends of the Southern Hemisphere. Here we will complement his analysis with a focus on timescales of variability and on anomalously cold stratospheric winters when polar ozone chemistry is particularly impactful.

2. Models and Observational Reference Data

Models used here are listed in Table 1.

We use all chemistry-climate models from CMIP6 for which daily and zonal-mean temperature and geopotential height (GPH) fields are available for "historical" simulations, and their no-chemistry CMIP6 equivalents where such models exist. Furthermore we use the SOCOL (Sukhodolov et al., 2021), ACCESS-CM2-Chem, and UKESM1-StratTrop models from the Chemistry-Climate Model Initiative Phase 2 (CCMI2) set of models (Plummer et al., 2021), and their no-chemistry CMIP6 equivalents. UKESM1-StratTrop is a further development of the UKESM1-0-LL model (Sellar et al., 2019), based on the same no-chemistry background model (HadGEM3-GC31-LL; Williams et al., 2018; Kuhlbrodt et al., 2018) but with some updates to photolysis and other reaction rates which reduce a general overestimation of ozone in the extrapolar stratosphere. (Other CCMI2

Table 1
Sixth Coupled Model Intercomparison Project/Chemistry-Climate Model Initiative Phase 2 Chemistry and Corresponding CMIP6 No-Chemistry Models Considered Here

CCMs		No-chemistry models		Differences	References
CESM2-WACCM	3	CESM2	11	higher top, NGWD	G19, DA20
CESM2-WACCM-FV2	3	CESM2-FV2	3	higher top, NGWD	G19, DA20
CNRM-ESM2-1	9	CNRM-CM6-1	28	same settings	S19, V19, M20
GFDL-ESM4	3	GFDL-CM4	1	higher top, NGWD	D20, H19
MRI-ESM2-0	5				Y19
UKESM1-0-LL	13	HadGEM3-GC31-LL	3	same settings	SE19, K18, W18
UKESM1-StratTrop	*3*	*HadGEM3-GC31-LL*	*5*	same settings	SE19, K18, W18
ACCESS-CM2-Chem	*3*	*ACCESS-CM2*	*3*	same settings	B20, BO20
SOCOL4	*3*	*MPI-ESM1-2-LR*	*3*	same settings	S21, M19

Note. The second and fourth columns denote the number of "historical," REF-D1, or AMIP simulations used in the analysis. For the purposes of this paper, models listed in italics are atmosphere-only; we use their CCMI2 REF-D1 and CMIP6 AMIP simulations, respectively. The "differences" pertain only to the model dynamics shared between the model pairs. Other differences exist because of the more comprehensive functionalities of the Earth System models versus their "physics only" counterparts. References: B20 = Bi et al. (2020), BO20 = Bodman et al. (2020), D20 = Dunne et al. (2020), DA20 = Danabasoglu et al. (2020), G19 = Gettelman et al. (2019), H19 = Held et al. (2019), K18 = Kuhlbrodt et al. (2018), M19 = Mauritsen et al. (2019), M20 = Michou et al. (2020), S19 = Séférian et al. (2019), S21 = Sukhodolov et al. (2021), SE19 = Sellar et al. (2019), V19 = Voldoire et al. (2019), W18 = Williams et al. (2018), Y19 = Yukimoto, Kawai, et al. (2019).

models are not used here because they do not have no-chemistry equivalents in the CMIP6 group of models.) References in table 1 are for the chemistry models (Morgenstern, 2021). In the CCMI2 "REF-D1" simulations used here the three CCMI2 models are not coupled to an interactive ocean; rather they use prescribe observational (HadISST) sea-surface conditions (Rayner et al., 2003). The simulations are therefore more comparable to the Atmosphere Model Intercomparison Project (AMIP) simulations of CMIP6 (although these use a different observational climatology for sea surface conditions vs. the REF-D1 simulations; Taylor et al., 2015). ACCESS-CM2-Chem and ACCESS-CM2 share an atmosphere model with UKESM1-0-LL and HadGEM3-GC31-LL but use a different land model. MRI-ESM2-0 does not have a no-chemistry equivalent amongst the CMIP6 models.

CESM2-FV2 and CESM2-WACCM-FV2 are identical to CESM2 and CESM2-WACCM but with the atmospheric resolution degraded from about ∼1° to ∼2°. Apart from chemistry, the leading differences between CESM2 and CESM2-WACCM (at both resolutions) are the higher top of the latter, raised from 2.26 hPa in CESM2 to $4.5 \cdot 10^{-6}$ hPa in CESM2-WACCM, with an associated increase in the number of levels from 32 to 70. With the higher top also comes some additional middle- and upper-atmosphere physics absent in CESM2 (Danabasoglu et al., 2020). Importantly, the non-orogrophic gravity wave (NGWD) parameterization after Richter et al. (2010) used by CESM2-WACCM is absent in CESM2 (Gettelman et al., 2019). Between GFDL-CM4 and GFDL-ESM4, similarly the model top has been raised from 1 hPa to 0.01 hPa, the number of levels has been increased from 33 to 49, and the tunings of the NGWD scheme differ between the two model versions. These and several other differences between these models are listed in Table 1 of Dunne et al. (2020). The other models not discussed in more detail (the CNRM models, UKESM1–both versions/HadGEM3-GC31-LL, the ACCESS-CM2 versions, SOCOL/MPI-ESM1-2-LR) share the same code, dynamics settings, and grids between the two model versions, with differences apart from atmospheric chemistry restricted to "Earth System" components such as aerosols and ocean biogeochemistry considered not relevant for the purposes of this paper.

The no-chemistry model CNRM-CM6-1 uses a simplified ozone scheme (Voldoire et al., 2019, and references therein). In this regard the model differs from the other no-chemistry models considered here. It is therefore included in Figures 1–3.

Previous evaluations have shown that the UKESM1-0-LL and CNRM-ESM2-1 models well simulate 1979–2000 Arctic ozone trends, GFDL-ESM4, CESM2-WACCM, and MRI-ESM2-0 underestimate Arctic ozone depletion (Morgenstern et al., 2020), and SOCOL quite faithfully reproduces extrapolar ozone (Sukhodolov et al., 2021). The ozone field used to drive the no-chemistry models HadGEM3-GC31-LL, MPI-ESM1-2-LR, GFDL-CM4,

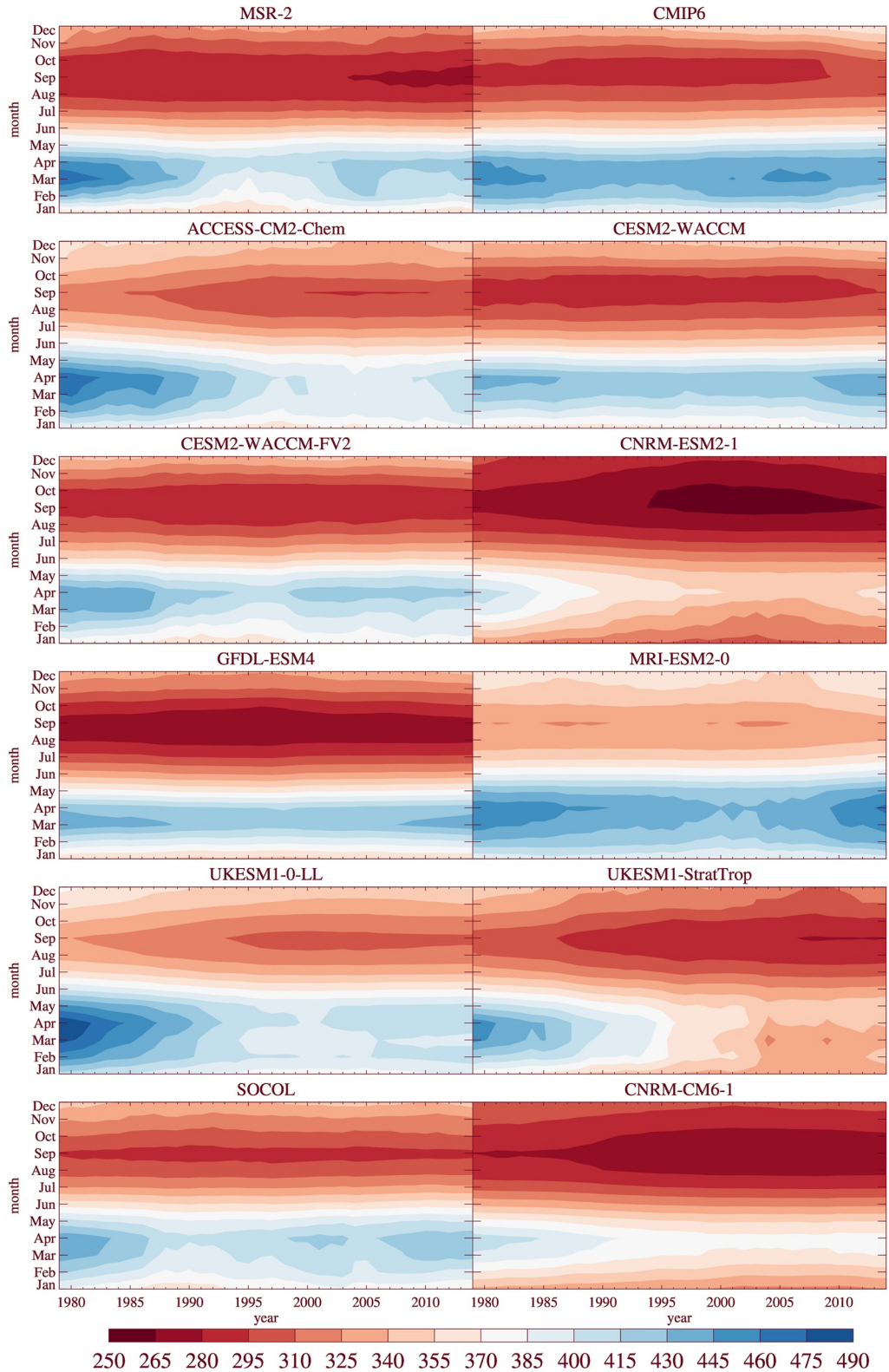

Figure 1. 1979–2014 monthly mean TCO (DU) averaged over the Arctic polar cap (north of 75°N), expressed as functions of the year and month of the year and smoothed with an 11-year boxcar filter, for the MSR-2 observational reference (van der A et al., 2015a), the CMIP6 ozone forcing dataset (Checa-Garcia et al., 2018), and the single-model ensemble-means of the "historical" and REF-D1 simulations, respectively, by the nine chemistry-climate models and CNRM-CM6-1.

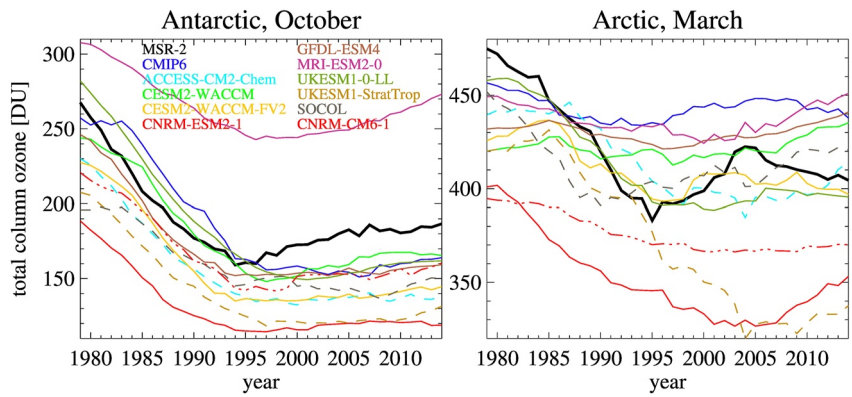

Figure 2. Polar-cap (polewards of 75°N/S) mean, ensemble-average TCO (DU) for the Antarctic and Arctic in October and March, respectively, smoothed with an 11-year boxcar filter. Thick black: Observations (van der A et al., 2015a). Dark blue: CMIP6 climatology (Checa-Garcia et al., 2018). Other colors: models. Solid: CMIP6 "historical" ensemble means. Dashed: CCMI2 REF-D1 atmosphere-only ensemble means. Red dash-dot-dot-dot: CNRM-CM6-1.

CESM1, CESM2-FV2, and ACCESS-CM2 however much underestimates these Northern-Hemisphere and especially Arctic ozone trends (Morgenstern et al., 2020), with hemispheric- and annual-mean TCO trends for 1979–2000 in the CMIP6 climatology (Checa-Garcia et al., 2018) only reaching approximately a third of observed trends (Morgenstern, 2021). In the Antarctic, UKESM1-0-LL and MRI-ESM2-0 under- and overestimate, respectively, Antarctic ozone during spring, whereas the other CMIP6 chemistry-climate models simulate more realistic Antarctic ozone depletion (Morgenstern et al., 2020).

The results will be compared to version 2 of the National Center for Environmental Prediction (NCEP)/Department of Energy (DOE)/NCEP-DOE2 reanalysis (Kanamitsu et al., 2002), the fifth European Centre for Medium-Range Weather Forecasts Reanalysis (ERA5, Hersbach et al., 2020), and the Multi-Sensor Reanalysis 2 (MSR-2) total-column ozone climatology (van der A et al., 2015a).

3. Method

In a seminal paper Baldwin and Dunkerton (2001) showed how stratospheric circulation anomalies in the Arctic propagate to low altitudes and affect tropospheric circulation for the approximately two months that such features may last. For example, impacts include anomalous states of the NAM, the positions of the northern storm tracks, and mid-latitude storms. Equivalent influences of the stratosphere on the weather of the Southern Hemisphere have also been demonstrated (Thompson et al., 2005). Baldwin and Dunkerton (2001)'s method also lends itself to a comparison of chemistry versus equivalent no-chemistry models presented here. While Baldwin and Dunkerton (2001) present composites of stratospheric NAM indices, here we modify their method to using polar-cap mean stratospheric temperature as our key metric. The reason for this is that (a) this diagnostic is available for both chemistry- and no-chemistry models, unlike e. g. ozone, and (b) wintertime low temperatures are associated with heterogeneous chlorine activation on polar stratospheric clouds followed by ozone depletion in models with interactive chemistry. Much of the rest of our analysis is inspired by Baldwin and Dunkerton (2001), namely:

1. We use all available "historical", REF-D1, or AMIP daily and zonal-mean temperature and GPH fields on pressure levels for 1980–2014, for the chemistry and no-chemistry models listed in table 1. Ensemble sizes for these vary greatly between 1 and 28.
2. For each individual model ensemble of simulations separately, we calculate the polar-cap (75°N–90°N and 90°S–75°S, respectively) average temperature and GPH fields.
3. We smoothen both fields using 15-day boxcar filters, to reduce the impact of outliers, and subtract the mean annual cycles of polar-cap temperature and GPH, creating for each model ensemble temperature and GPH polar-cap mean anomaly timeseries.

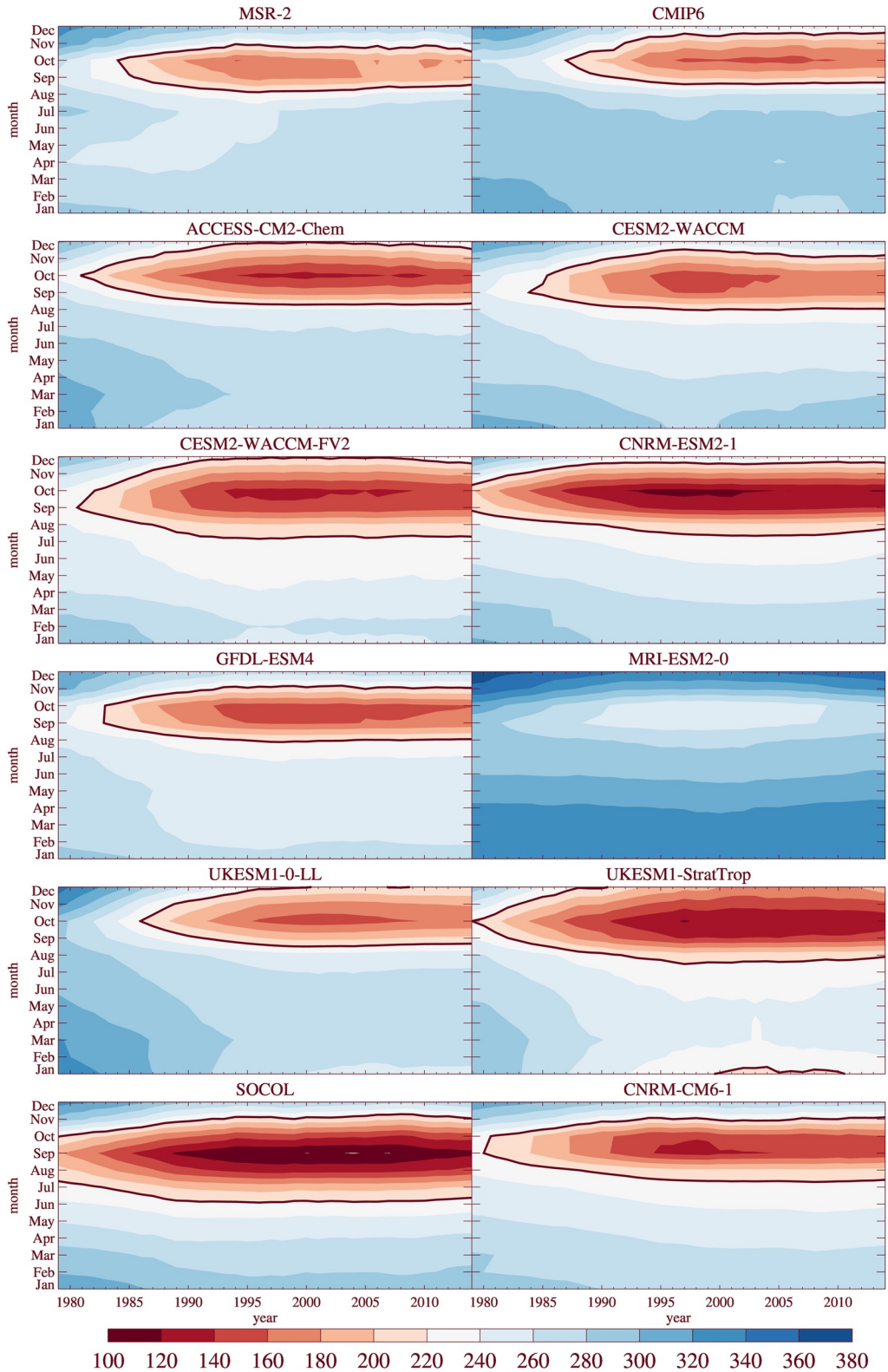

Figure 3. Same as Figure 1, but for the Antarctic polar cap (south of 75°S). The thick contour marks TCO = 220 DU, the traditional threshold defining the ozone hole.

4. We determine, for every year starting on 1 September (for the Arctic) and 1 May (for the Antarctic) and for every ensemble member, the lowest value at 70 hPa of the polar-cap average temperature anomaly, and the day of its occurrence. This temperature is then used to rank the years by stratospheric temperature.
5. Using this ranking, we select only the 20% coldest winters. Defining time 0 to be the day of the occurrence of the largest cold anomaly in these years, we average the temperature during these cold seasons from 130 days before to 100 days after the coldest day. The number of winter/spring seasons entering these averages is thus $0.2 \times 34 \times n$, rounded to the nearest integer, where n is the ensemble size of every individual model, and 34 complete winters occur in the study period 1980–2014 (Table 1). For the reanalyses, this amounts to 7 winter/spring seasons.

4. Results

4.1. General Model Performance for Monthly Mean Ozone and Temperature

Arctic total-column ozone, in the decades before ∼1995, experienced a decline of nearly 100 DU in March since 1979 but recovered slightly thereafter, see the MSR-2 panel of Figures 1 and 2. The relatively fast partial "recovery" between 1995 and 2005 indicates that a portion of the earlier loss might reflect internal variability (Figure 4–4 of WMO, 2018). We estimate that sustained March Arctic losses in the MSR-2 climatology are roughly 50 to 70 DU. Losses in other seasons were much smaller. The loss was mainly driven by increasing halogens in a well-understood mechanism involving chlorine activation on polar stratospheric clouds (WMO, 2018); hence the much smaller trends outside the spring season. In the nine chemistry-climate models and the CMIP6 climatology (itself derived from model results, Checa-Garcia et al., 2018), this springtime loss is captured but with varying degrees of realism. March trends come close to MSR-2 in UKESM1-0-LL, UKESM1-StratTrop, and ACCESS-CM2-Chem, but in these models, unrealistically, the ozone loss is bigger in April than in March. SOCOL and CNRM-ESM2-1 also both simulate substantial though underestimated ozone loss. CESM2-WACCM, CESM2-WACCM-FV2, GFDL-ESM4, and MRI-ESM2-0 all substantially underestimate the amount of ozone loss, as does the CMIP6 ozone climatology used to force no-chemistry CMIP6 models. A failure to simulate a realistic impact of halogen increases on Arctic ozone can indicate that chlorine activation in these models is not realistic, for example, because of a stratospheric warm bias reducing the occurrence of polar stratospheric clouds (PSCs), or for other reasons such as an incorrect representation of the physics of PSC formation or of heterogeneous chemistry leading to chlorine activation. CNRM-CM6-1, with its simplified ozone scheme, underestimates trends in TCO during spring; the simulated trends are also substantially smaller than those simulated by CNRM-ESM2-1.

Similarly, the Antarctic has experienced substantial ozone loss in spring, manifesting as the Antarctic "ozone hole" (Figures 2 and 3). The models capture this, but again with various biases. Several models have severe ozone loss persisting for too long into summer (the UKESM1 models, ACCESS-CM2-Chem, the CESM2-WACCM versions, and CNRM-ESM2-1). The MRI-ESM2-0 model substantially underestimates ozone loss, whereas the SOCOL model strongly overestimates it, partly due to an early onset of the ozone hole, with lowest polar ozone occurring in September not October. The GFDL-ESM4 model overall has the most realistic timing and small biases of Antarctic ozone–we note however the much underestimated ozone loss in the Arctic in this model. CNRM-CM6-1 slightly underestimates ozone over the Antarctic during the ozone hole season, but less so than CNRM-ESM2-1, and simulates quite realistic trends.

Next we assess the simulation of temperature in these models.

An inspection of the mean 1980–2014 bias and standard deviation for the 70 hPa polar-cap mean temperature (Figure 4) indicates that for both polar regions, there is excellent agreement between the NCEP-DOE2 (Kanamitsu et al., 2002) and the newer ERA5 reanalyses (Hersbach et al., 2020), with essentially identical standard deviations and absolute biases between the two reanalyses of mostly less than 1 K, much smaller than typical model biases. Over both poles, a majority of models (chemistry and no-chemistry alike) exhibits cold biases during spring. In the Antarctic, the cold bias reaches −15 to −20K in November in ACCESS-CM2-Chem, CNRM-ESM2-1, UKESM1-0-LL, and UKESM1-StratTrop. These biases are all worsened versus their no-chemistry counterparts. The cold biases are reflected in an increase in stratospheric variability during December and January (as evidenced by the anomalously large standard deviation of temperature for these models), indicating an extension of the lifetime of the Antarctic polar vortex versus their no-chemistry counterparts. The CESM2-WACCM

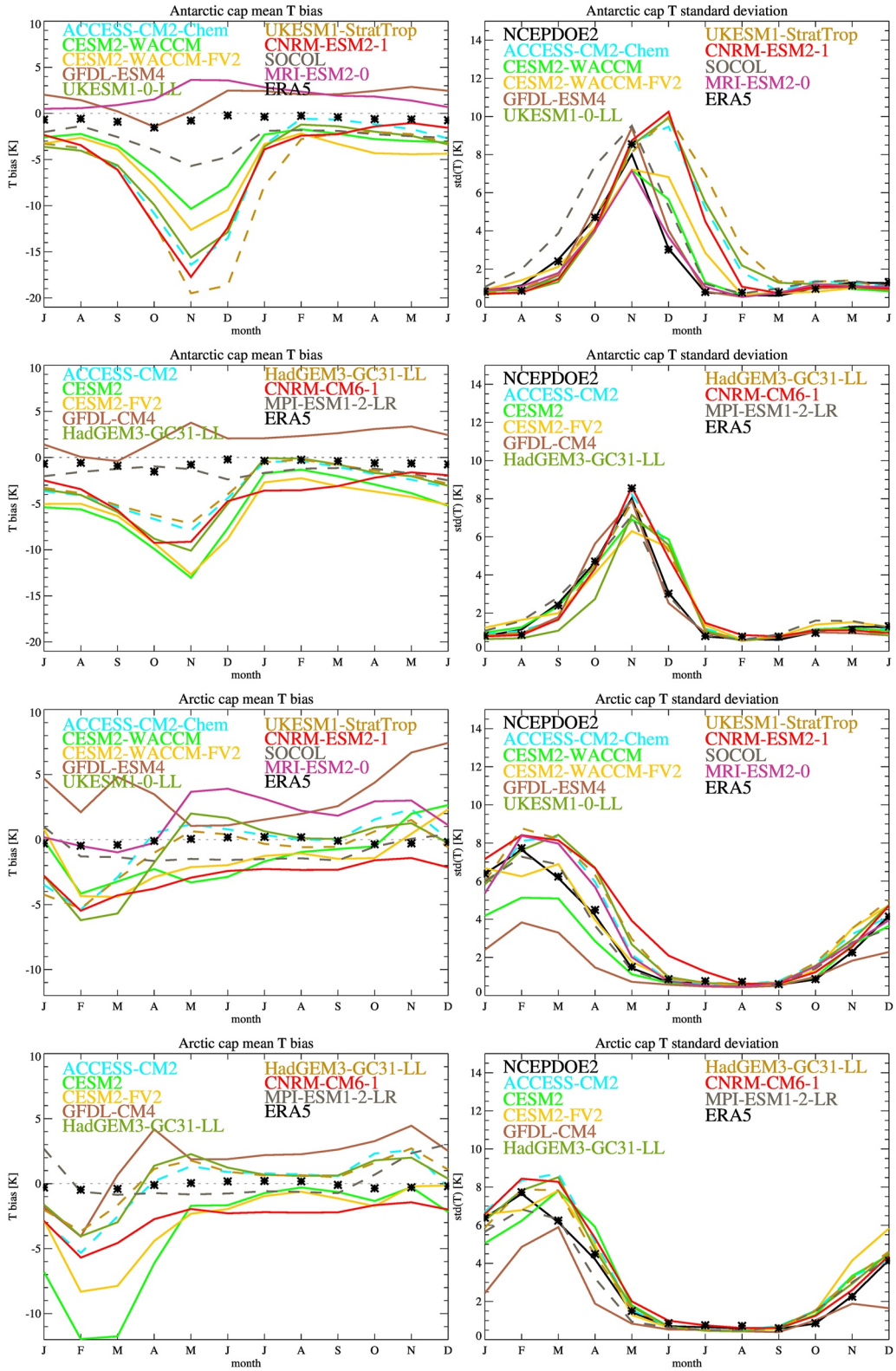
Antarctic cap mean T bias
ACCESS-CM2-Chem
CESM2-WACCM
CESM2-WACCM-FV2
GFDL-ESM4
UKESM1-0-LL
UKESM1-StratTrop
CNRM-ESM2-1
SOCOL
MRI-ESM2-0
ERA5
T bias [K]
month
Antarctic cap T standard deviation
NCEPDOE2
std(T) [K]
Antarctic cap mean T bias
ACCESS-CM2
CESM2
CESM2-FV2
GFDL-CM4
HadGEM3-GC31-LL
CNRM-CM6-1
MPI-ESM1-2-LR
Antarctic cap T standard deviation
Arctic cap mean T bias
Arctic cap T standard deviation

Figure 4.

models exhibit largely unchanged biases and variability in the Antarctic versus the no-chemistry equivalents, but a decreased cold bias in the Arctic in the chemistry versions. The GFDL-ESM4 model exhibits only a small warm bias in the Antarctic, which is consistent with its good simulation of Antarctic ozone depletion noted above, but a substantial warm bias (∼4K) in the Arctic in spring. Here, even this seemingly moderate warm bias can substantially suppress PSC formation, explaining the weak Arctic ozone depletion simulated by this model. SOCOL simulates relatively small biases in both polar regions but exaggerated variability in the Antarctic in September and October, reflecting the early onset of ozone depletion in this model noted above.

Comparing the two CNRM models, in the Antarctic CNRM-ESM2-1 simulates an increased cold bias and large variability extending into summer. This is consistent with its larger ozone depletion and a longer lifetime of the ozone hole found above, compared to CNRM-CM6-1 which has a smaller bias and a better representation of Antarctic stratospheric temperature variability.

4.2. Temperature Variablity and Cold Episodes in Chemistry- and No-Chemistry Models

Figures 5 and 6 confirm that practically all variability in polar-cap 70 hPa temperature occurs during the cold season–during summer this variability is no more than a few K but in the daily polar-cap average can reach and exceed ±20 K during winter and spring. For both polar regions there are asymmetries between cold and warm winters: For warm winters, the anomalies occur nearly symmetrically around the middle of the cold season (in the Arctic, approximately day 30, i.e., 31 January; in the Antarctic, approximately day −40, i.e., 22 November), whereas during extremely cold winters the temperature anomaly builds until the wintertime circulation collapses and temperatures rapidly return to the average, with the largest cold anomalies occurring in spring or even summer. Also models with larger ensembles (e.g., CNRM-CM6-1, MPI-ESM1-2-LR) show that for warm anomalies there is no sharp upper bound for the largest warm anomalies that can occur, whereas the cold anomalies, until well into spring, are sharply bounded by a lower envelope function which decreases during the course of the winter. This is more evident for the Antarctic (Figure 6) than the Arctic (Figure 5). During spring some rare extremely cold events occur, that is, long-lasting cold polar vortices, for example, in the UKESM1, CNRM-ESM2-1, and ACCESS-CM2-Chem models. This asymmetric nature of variability reflects coupling with mid-latitudes, or lack thereof. During warm winters, the Arctic and Antarctic receive their heat from mid-latitudes in dynamical disturbances. This mechanism is different from the radiative cooling that dominates during cold, dynamically relatively unperturbed winter seasons and causes temperatures to gradually drop throughout the season, until the final warming marks the end of the polar vortex.

A remarkable warm outlier is seen in the NCEP-DOE2 and ERA5 reanalyses around day −100 (i.e., 23 September; Figure 6). This is the vortex breakup and major stratospheric warming of 2002 which at the time was considered very unusual as it had never before been seen in the observational record (Newman & Nash, 2003). Both chemistry (UKESM1, SOCOL) and no-chemistry (CESM2) models exhibit similar extremely warm episodes around this time of the year, meaning that some CMIP6/CCMI2 models can qualitatively simulate such events (Jucker et al., 2021).

CNRM-CM6-1 has a tendency, perhaps more so than the other no-chemistry models, to simulate long-lasting extremely cold polar vortices. This may be the effect of having an interactive albeit simplified ozone scheme, which sets this model apart from the other no-chemistry models.

Restricting our attention to the 20% of years with the lowest 15-day mean temperature anomalies in the Arctic and Antarctic at 70 hPa, Figure 7 indicates that the median of the date of occurrence of the most anomalously cold day, in the reanalyses, is around day 51 or 52 (21 or 22 February) in the Arctic. In the Antarctic, the two reanalyses exhibit a somewhat larger disagreement regarding the timing of the coldest day, with ERA5 indicating a later occurrence (day −33, i.e., 29 November) than NCEP-DOE2 (day −40, i.e., 22 November). It is noteworthy that all 17 models considered here simulate a later median date for the thus defined coldest day in the Arctic at 70 hPa

Figure 4. (left) Monthly mean 70 hPa bias of polar-cap (poleward of 75°N/S) mean temperature (K) relative to National Center for Environmental Prediction (NCEP)-DOE2 for the period 1980–2014. (right) Standard deviation of monthly mean 70 hPa polar-cap mean temperature (K). (first row) Chemistry-climate models, Antarctic polar cap mean. (second row) Same for the associated no-chemistry models. (third row) Chemistry-climate models, Arctic polar cap mean. (fourth row) Same for the associated no-chemistry models. Solid lines represent Sixth Coupled Model Intercomparison Project (CMIP6) "historical" ensembles, dashed lines are CMIP6 AMIP (ACCESS-CM2, HadGEM3-GC31-LL, MPI-ESM1-2-LR) and CCMI2 REF-D1 (ACCESS-CM2-Chem, UKESM1-StratTrop, SOCOL) ensembles. Respectively. Black "*" symbols denote the ERA5 reanalysis.

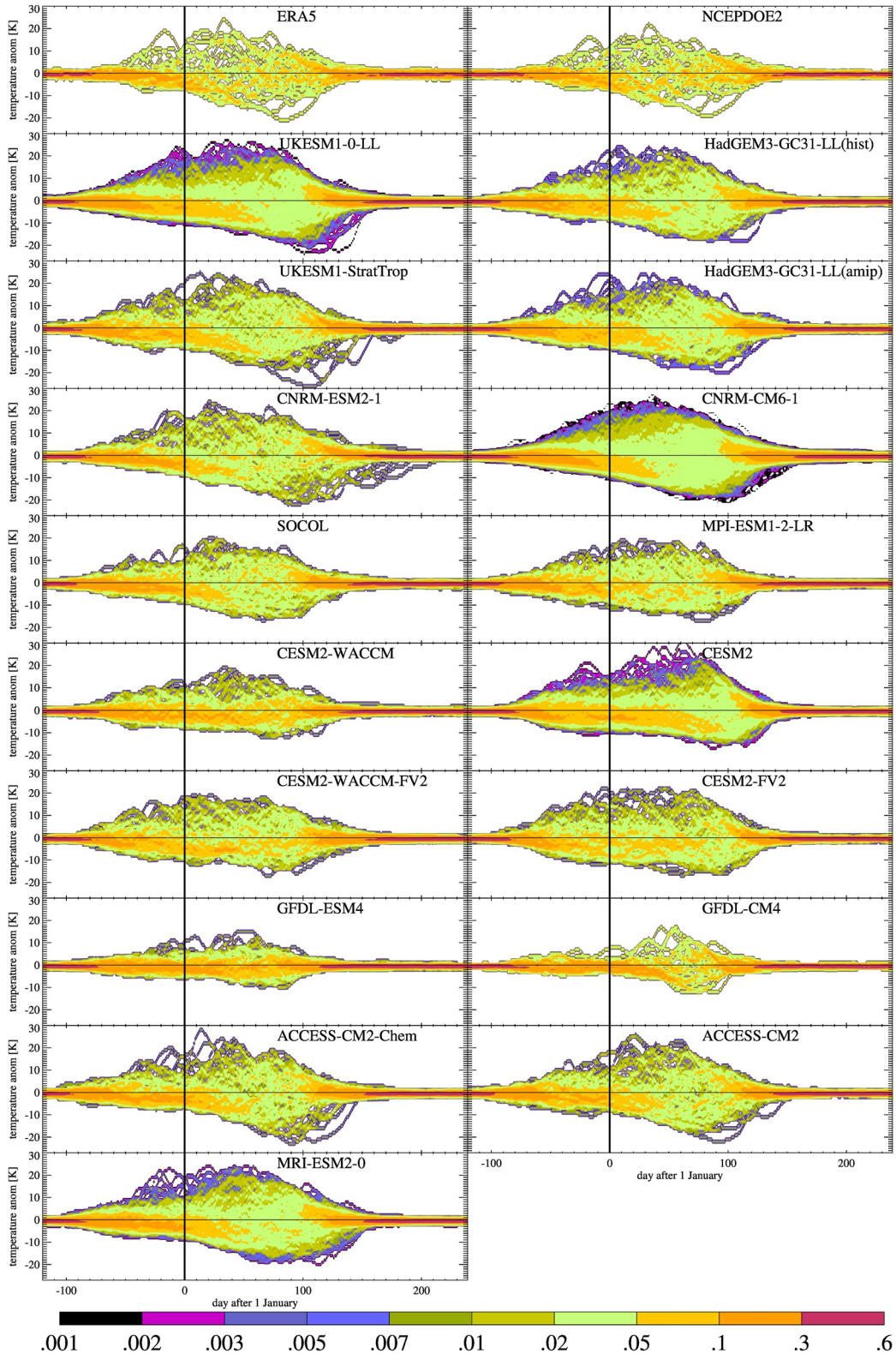

Figure 5. Probability density plot (K^{-1}) of the Arctic-mean (75°N–90°N) temperature anomaly relative to the 1980–2014 mean seasonal cycle at 70 hPa, in the NCEP-DOE2 and ERA5 reanalyses and the climate models as a function of the day of the year, for September 1980 to August 2014. Models with larger ensembles allow for better sampling of low-probability temperature anomalies (colored in blue and violet); these colors are therefore absent for small-ensemble models and the reanalyses.

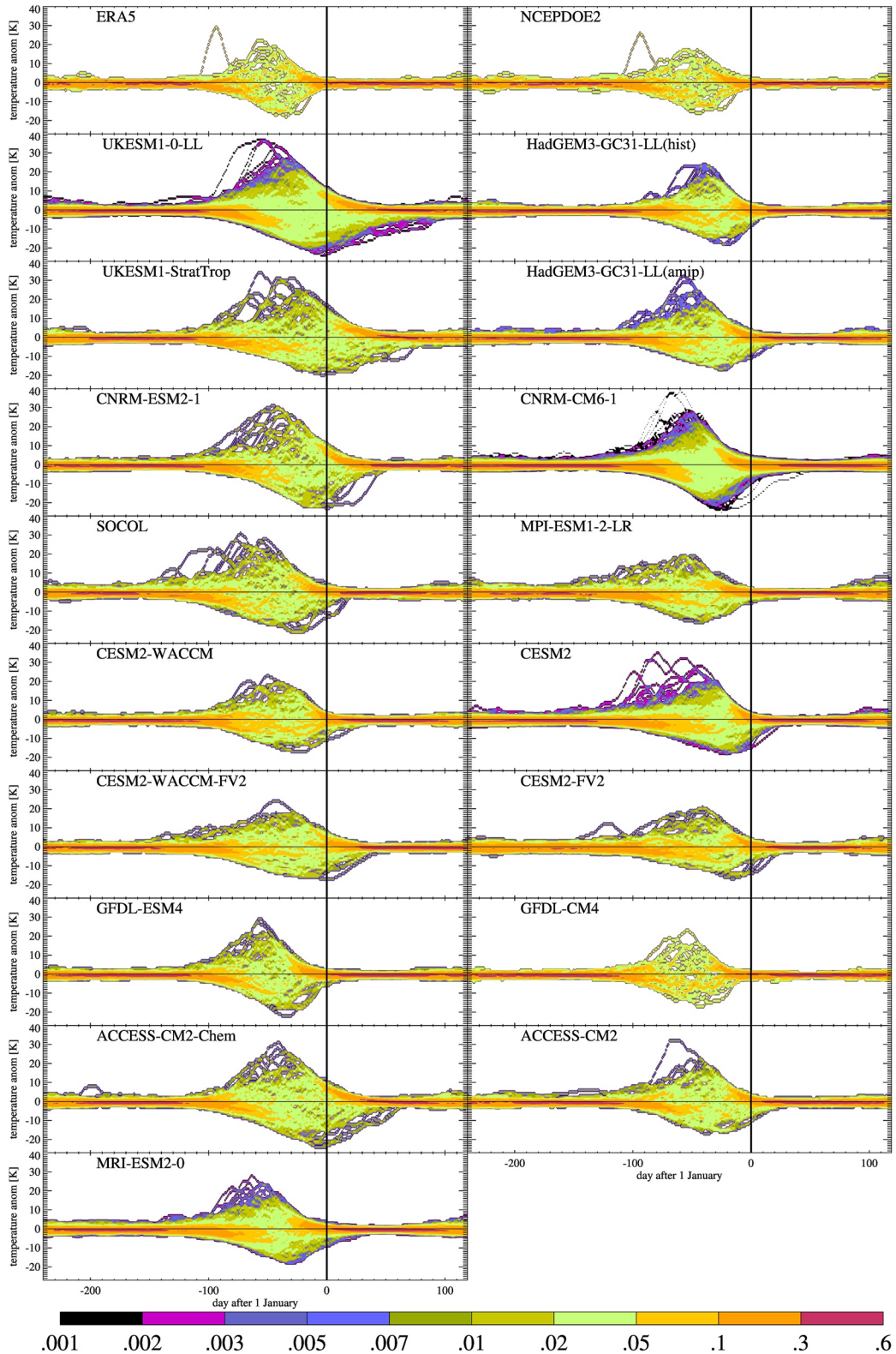

Figure 6. Same as Figure 5, but for Antarctic 70 hPa polar-cap mean temperatures.

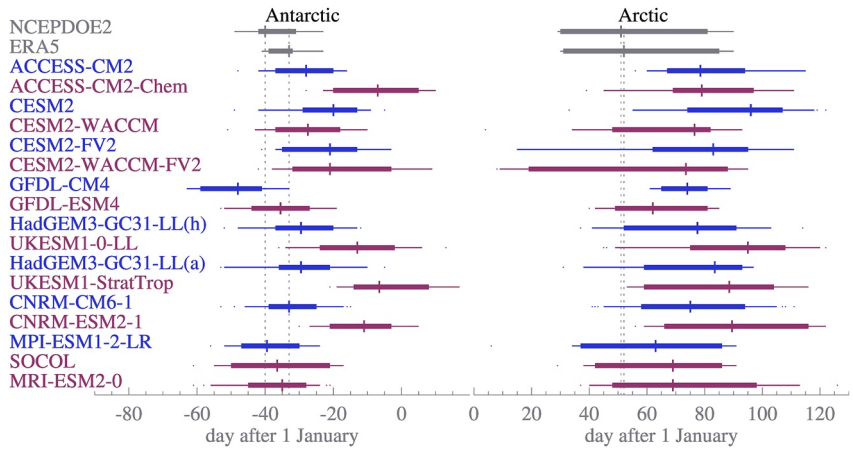

Figure 7. Date of occurrence of lowest temperatures in years with the minimum of the 15 day filtered 70 hPa polar-mean temperature anomaly in the lowest 20%. Left: Antarctica (75–90°S). Right: Arctic (75–90°N). Vertical bars: Median day. Thick lines: 16 to 84-percentile range. Thin lines: 2.5 to 97.5-percentile range. Dots: Outliers outside the 2.5 to 97.5 percentiles. Note that the NCEP-DOE2 and fifth European Centre for Medium-Range Weather Forecasts Reanalysis (ERA5) reanalyses and the GFDL-CM4 simulations only have 7 such cold winters each in this 20% category (out of a total of 34, for the period of September 1980–August 2014). The long dashed vertical lines mark the medians of the coldest days in the two reanalyses. Chemistry models are represented in red, no-chemistry models in blue. "(h)" stands for the "historical" ensemble of HadGEM3-GC31-LL, '(a)' for AMIP.

by 10 days or more. Also in the Antarctic most models simulate a delay in the coldest day relative to both reanalyses. Both findings may illustrate that climate models struggle with correctly capturing stratospheric dynamics in the polar regions, although it is impossible to be sure given that only the seven coldest winters are considered in the reanalyses. In the Arctic, all models have some degree of overlap of the 16 to 84-percentile interval for this coldest date with the reanalyses, whereas in the Antarctic, where the reanalyses show relatively little variation in the date of the coldest day, the models simulating the most severe ozone depletion (ACCESS-CM2-Chem, UKESM1, and CNRM-ESM2-1) all have 16 to 84-percentile ranges for this diagnostic that do not overlap with those of the reanalyses.

For these four model pairs (ACCESS, HadGEM3/UKESM1–both versions, CNRM) for the Antarctic the chemistry variants simulate a delay in the median occurrence of the coldest day by around 20 days versus their no-chemistry counterparts. With the exception of the ACCESS pair, this also holds in the Arctic but with smaller shifts. However, for other model pairs this is not the case: The MPI-ESM1-2-LR/SOCOL pair exhibits quite similar behavior for both polar regions, and the GFDL pair simulates shifts in the coldest day of different signs in the two polar regions. The GFDL model pair is the only example where for both polar regions these shifts are in the right direction and the observed dates of the coldest day are within the 16 to 84 percentile ranges of the dates of the modeled coldest days. For the Antarctic, where the timing of the coldest day in GFDL-ESM4 is excellent, this coincides with the good simulation of Antarctic ozone depletion shown in Figure 3. The CESM2 pairs, in most cases, produce an earlier coldest day if interactive chemistry is used. We will discuss these findings more in Section 5.

4.3. Composite Analysis of Cold Stratospheric Winters

Next we produce composites for temperature and GPH for cold stratospheric winter/spring seasons, similar to Baldwin and Dunkerton (2001)'s method. We express these fields relative to the time of occurrence of the largest absolute temperature anomaly (deviation from the mean) at 70 hPa. Baldwin and Dunkerton (2001) and Thompson et al. (2005) had used NAM and SAM indices instead, respectively.

Figures 8–10 show that the 20% coldest winters, at the time of the lowest temperature, in the reanalyses are generally between 10 and 15 K colder at 70 hPa than the average winter, in agreement with Figures 5 and 6. ERA5 suggest cold anomalies that are systematically colder than NCEP-DOE2 by a few K. Substantial cold anomalies often start at least two months before the largest temperature anomalies occur, and last typically 30–40 days

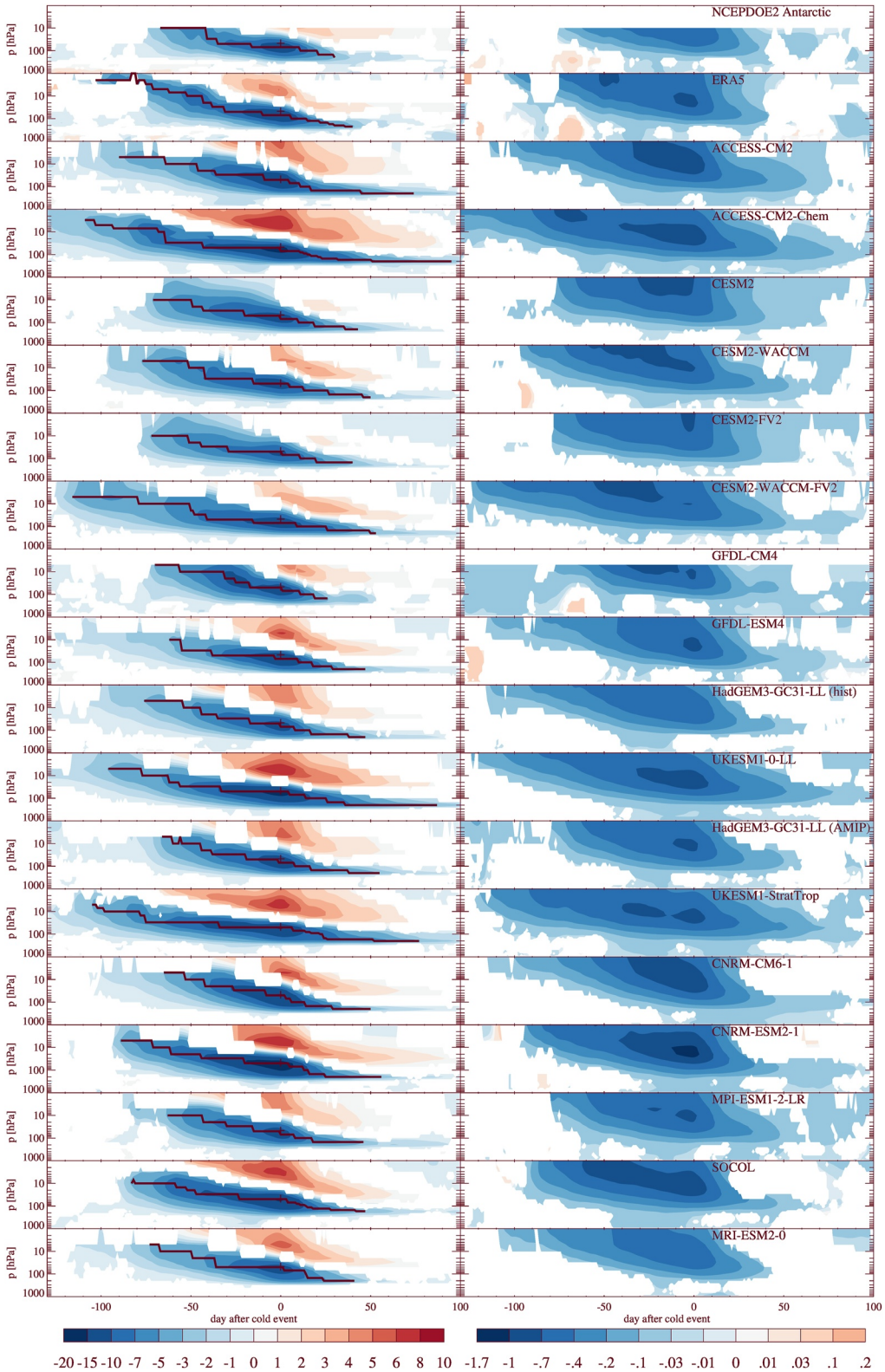

Figure 8. Antarctic polar cap (90°S–75°S) mean temperature (left; in K) and GPH (right; in km) anomalies (relative to their 1980–2014 mean seasonal cycles) for the 20% coldest winters in the chemistry-climate and no-chemistry models. Time is relative to the day of occurrence of the coldest day at 70 hPa, marked by a small "+" symbol. The pressure where the temperature anomaly minimizes, where this minimum is less than −3 K, is marked by bold lines. White pixels indicate that fewer than two thirds of the data composited here are of the same sign, that is, these composites are not significantly different from 0.

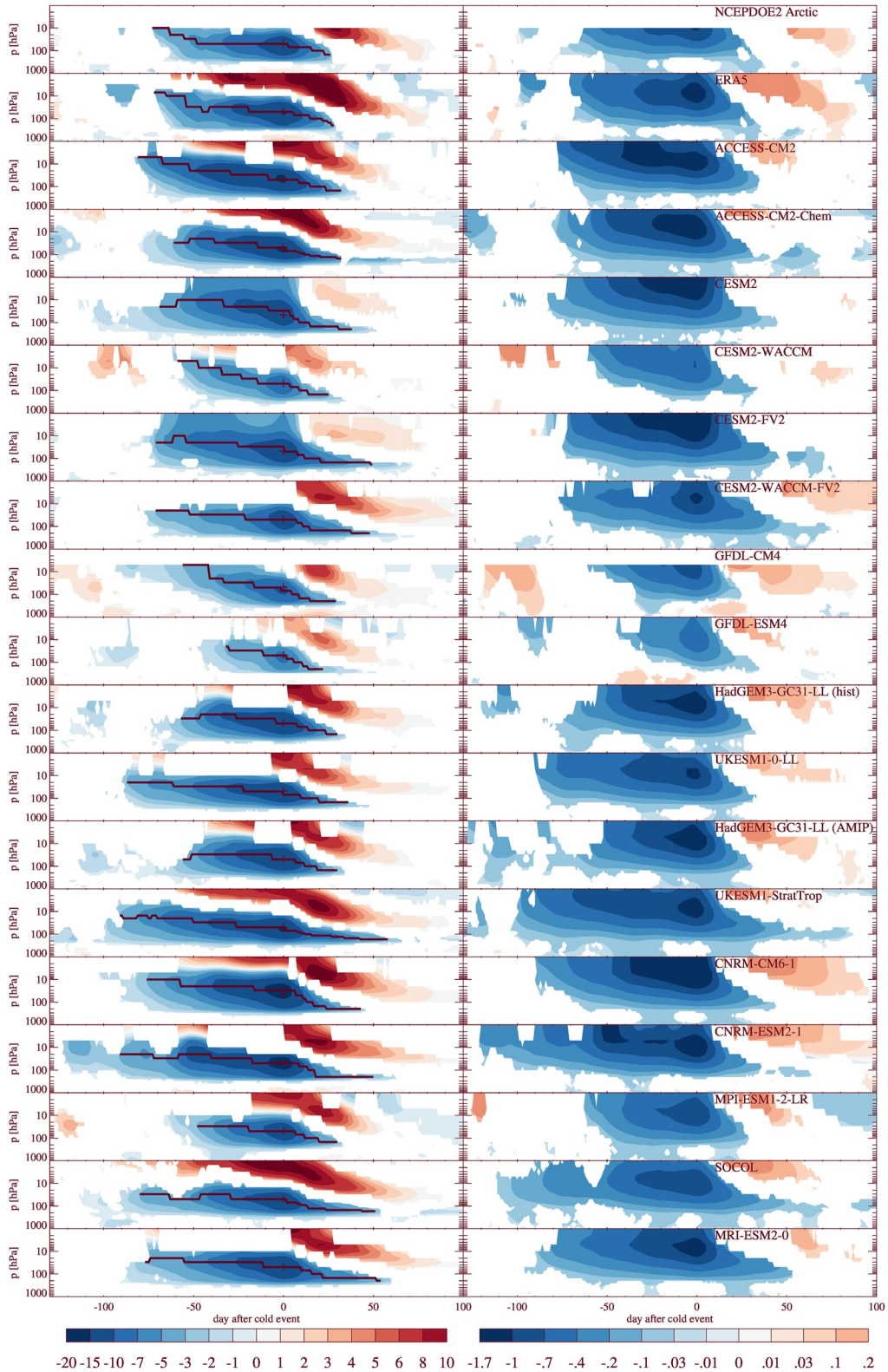

Figure 9. Same a Figure 8, but for the Arctic polar cap (75°N–90°N).

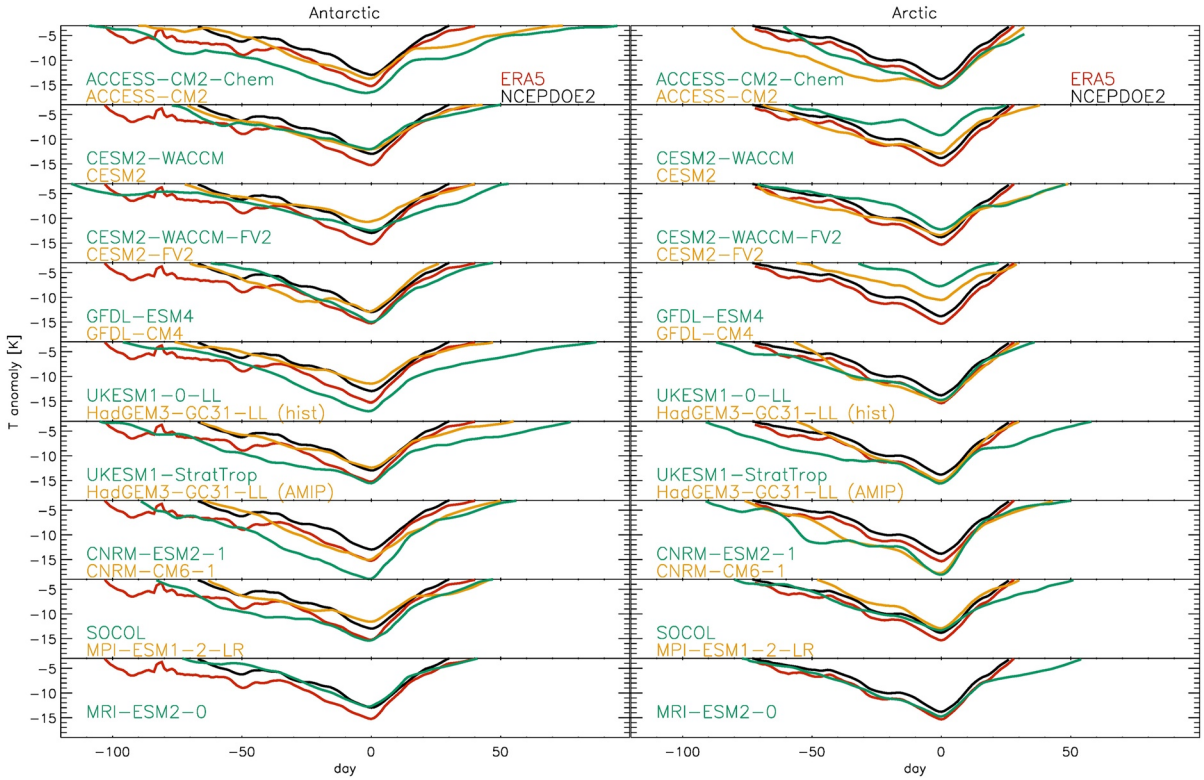

Figure 10. Temperature of the composite cold anomalies, evaluated along the bold lines in Figures 8 and 9. Black: NCEP-DOE2. Red: ERA5. Orange: no-chemistry models. Green: chemistry models. Left: Antarctica. Right: Arctic.

beyond this date. They are accompanied by corresponding negative GPH anomalies that sometimes extend into the troposphere, in agreement with Baldwin and Dunkerton (2001)'s and Thompson et al. (2005)'s findings. In the Antarctic, the cold anomalies start earlier in ERA5 than NCEP-DOE2, mostly because unlike NCEP-DOE2, ERA5 covers the upper stratosphere where these features originate before descending through the stratosphere over the coarse of the season. For the Arctic, the reanalyses agree on the start and end dates of the cold anomalies, when the cold anomaly reduces to less than −3 K, about 28 days after the coldest day. For the Antarctic, ERA5 has the cold anomalies persist for about 10 days longer into spring than NCEP-DOE2.

For both polar regions, qualitative agreement between the CMIP6 models is generally remarkably good. In the Antarctic, in several model pairs temperature and GPH anomalies are systematically more persistent, usually both at the start and end, and also of larger-amplitude in the chemistry-climate models (UKESM1-0-LL, UKESM1-StratTrop, CNRM-ESM2-1, SOCOL, ACCESS-CM2-Chem) than in the corresponding no-chemistry models (HadGEM3-GC31-LL, CNRM-CM6-1, MPI-ESM1-2-LR, ACCESS-CM2, Figure 10). In all cases, the chemistry models produce more long-lasting cold anomalies than their corresponding no-chemistry models, with extensions ranging from a marginal one day for SOCOL/MPI-ESM1-2-LR to 40 days for UKES1-0-LL/HadGEM3-GC31-LL. However most models (both chemistry and no-chemistry, except for CESM2-FV2, GFDL-CM4, and MRI-ESM2-0) have the cold anomalies lasting longer than day 40, the date of disappearance in ERA5. The cold anomalies lasting well into spring are reflected in GPH anomalies also lasting longer and spawning low-GPH anomalies in the troposphere (Figure 8), signaling impacts of this behavior on simulated tropospheric weather. With the exceptions of CESM2/CESM-WACCM and the GFDL models, the chemistry models generally produce bigger-amplitude cold anomalies and larger negative temperature biases versus the reanalyses than their corresponding no-chemistry models with cold anomalies typically about 3–5K colder in chemistry versus no-chemistry models (Figure 10). We note that ERA5 places the genesis of the cold anomalies at 1 hPa about 110 to 100 days before the occurrence of the maximum anomaly at 70 hPa (Figure 8). In reality, these anomalies might even originate in the mesosphere; 1 hPa is only the highest pressure level which the ERA5 and CMIP6 fields are interpolated to. None of the models has the cold anomalies starting this high

up, meaning if these models were used in seasonal prediction, there would be concern that the models do not correctly represent the upper-level genesis of such anomalies, hence missing out on some potential predictability.

For the Arctic, where the influence of ozone depletion is smaller than in the Antarctic, there is also a tendency for some chemistry models to simulate a delayed end of the cold anomalies relative to the reanalyses and their corresponding no-chemistry model (ACCESS-CM2-Chem, UKESM1-StratTrop, CNRM-ESM2-1, SOCOL; Figure 10). However, the picture is more complex than in the Antarctic. There are counterexamples: For example, CESM2 produces longer-duration cold spells than CESM2-WACCM, GFDL-ESM4 simulates shorter-duration, more warm-biased, less realistic cold anomalies than GFDL-CM4, and in no case do the chemistry models simulate consistently colder cold anomalies (as measured by the minimum of the temperature anomaly) than their no-chemistry counterparts. Coinciding with the underestimated amplitudes of cold anomalies, CESM2-WACCM and GFDL-ESM4 are characterized by much underestimated Arctic ozone loss (Figure 1) and in the case of GFDL-ESM4, underestimated variability of polar-cap mean temperature (Figure 5. In all cases, geopotential height anomalies sometimes reaching into the troposphere is in agreement with Baldwin and Dunkerton (2001).

The CNRM model pair exhibits remarkably similar behavior in the Arctic. Both model configurations have cold anomalies lasting longer into spring than in the reanalyses by about the same amount of time. Also the depths of the cold anomalies are quite similar (Figure 10). The fact that these two models behave so similarly may be related to similarities in ozone between both configurations. As noted above, CNRM-CM6-1 uses a simplified ozone scheme, that is, like in the chemistry models, in this model ozone is a predicted variable coupled to dynamics (Voldoire et al., 2019).

Comparing now the climate models with the NCEP-DOE2 and ERA5 reanalyses (Hersbach et al., 2020; Kanamitsu et al., 2002), in the cases where the chemistry models exhibit increased persistence, the no-chemistry counterparts are often in better agreement with observations than the chemistry models, both in the Arctic and the Antarctic. This means the persistence of cold anomalies long into spring occurring in most CCMs is not reflected in the reanalyses.

5. Discussion

We have analyzed the dynamics of stratospheric cold winters in 13 CMIP6 and three CCMI2 climate and chemistry-climate models and compared them to reanalyses. The behavior of the chemistry models depends crucially on whether substantial additional differences, extending beyond interactive ozone chemistry, exist regarding their dynamics configurations between the chemistry models and their no-chemistry equivalents. In three model pairs considered here, in particular there are differences in the model top, the number of levels representing the stratosphere, and NGWD. In four cases where the dynamics configurations are essentially unchanged versus the no-chemistry configuration (ACCESS-CM2-Chem, CNRM-ESM2-1, UKESM1-0-LL/UKESM1-StratTrop), coupling in chemistry results in a delay in the occurrence of the coldest day the Antarctic lower stratosphere. In the Arctic, two models with large Arctic ozone depletion (UKESM1-0-LL, CNRM-ESM2-1) also exhibit a delayed occurrence of the coldest day relative to their no-chemistry configurations, but this behavior is not consistent across the group of models. The SOCOL model does not exhibit any substantial shift in the timing of the coldest day versus its reference model MPI-ESM1-2-LR for both polar regions; SOCOL is also characterized by a generally good representation of ozone trends (Sukhodolov et al., 2021), albeit with an early onset of ozone loss in the Antarctic, and a good representation of Arctic temperature and variability (Figure 4). Several of these chemistry models (ACCESS-CM2-Chem, UKESM1-0-LL, UKESM1-StratTrop) exhibit timescales of persistence of stratospheric cold anomalies over both poles that are longer than in their no-chemistry counterparts, reflecting extensions of the lifetimes of both polar vortices. Possibly in the SOCOL model, this increased persistence is counteracted by the early onset of ozone depletion in the Antarctic, resulting in no shift of the occurrence of the coldest day relative to the background model, MPI-ESM1-2-LR, over the South Pole. The MRI-ESM2-0 model also behaves similarly to these chemistry models in that over the Arctic it exhibits too large a persistence of cold anomalies compared to the reanalyzes. Over the Antarctic its climatology compares relatively well to the reanalyzes.

The behavior of this group of models contrasts with the GFDL and two CESM2 pairs of models. In these pairs, the chemistry models differ more substantially in their dynamics configurations from their no-chemistry counterparts, namely the chemistry versions operate on a vertically extended grid with more levels in the stratosphere,

compared to their no-chemistry counterparts. CESM2-WACCM and CESM-WACCM-FV2 include an additional NGWD parameterization (Gettelman et al., 2019) completely absent in CESM2 and CESM2-FV2. GFDL-ESM4 also differs in terms of NGWD and a few other aspects (Dunne et al., 2020). NGWD drives the Brewer-Dobson Circulation and influences the stability of the polar vortices (for a recent review see Eichinger et al., 2020), so may well explain the differences in behavior between the CCMs and their no-chemistry equivalents. GFDL-ESM4 is the only chemistry model studied here with a substantial warm bias in the Arctic stratosphere in winter. Together with the much underestimated variability (Figure 4) this indicates this model does not realistically simulate Arctic ozone depletion (Morgenstern et al., 2020), but ranks amongst the top-performing models for Antarctic ozone depletion. CESM2-WACCM, like most other models studied here, has a cold bias in the Arctic winter stratosphere. Together also with the underestimated variability this suggests that the model simulates too many "cold" polar vortices with too regular ozone depletion. Both in the CESM2 and the GFDL models, however, the timings of the coldest days, for both polar regions, are either unchanged or more realistic in the chemistry models. The timescales of persistence are not appreciably different between the chemistry and no-chemistry configurations of these models.

The extended lifetimes of the polar vortices occurring in some chemistry-climate models compare worse to reanalyses than the shorter lifetimes of the polar vortices characterizing the corresponding no-chemistry models. This impact of interactive chemistry is consistent with earlier studies based on fewer models (Haase & Matthes, 2019; Lin & Ming, 2021; Oehrlein et al., 2020). An exception to this is CNRM-CM6-1 which uses a simplified ozone scheme. The timing of the end of cold spells in this model, about 50 days after the most anomalously cold day at 70 hPa (Figure 2), is almost the same as for CNRM-ESM2-1 which uses fully interactive chemistry. In both models, the cold spells last longer into spring than in the reanalyses.

The delayed occurrence of the most anomalously cold day, combined with the extended duration of the stratospheric cold spells in several chemistry models, suggests that also the final warming dates and the dates of wind reversal from westerlies to easterlies, both indicating the switching from winter to summer in the stratosphere, are delayed in these models.

In our analysis we have used "historical" coupled and atmosphere-only, REF-D1 and AMIP, simulations. The one model appearing both coupled and uncoupled in our analysis (HadGEM3-GC31-LL) does not exhibit any properties (within the limits of what has been investigated here) that could be unambiguously attributed to it being coupled or not. If the AMIP experiments were replaced with the corresponding "historical" coupled simulations, largely the same results would be found (not shown). Nonetheless, the influence of ocean coupling on the results could be investigated further.

The findings illustrate that in the cases where ozone chemistry is the only significant difference between two model configurations, ozone chemistry, and even simplified ozone chemistry as in CNRM-CM6-1, introduce additional "memory" into the atmosphere. Feedbacks of ozone chemistry onto radiation, for a cold winter, enhance radiative cooling and stabilize the vortex to last longer into spring; similar results were found in earlier single-model studies (Lin & Ming, 2021; Oehrlein et al., 2020). These effects can however be counterbalanced by retuning and/or additional physics, especially the non-orographic gravity wave schemes added or modified in GFDL-ESM4 and CESM2-WACCM (both versions).

Based on first principles, additional "physics" such as ozone chemistry can be expected to better capture Earth system feedbacks. However, this will only lead to a better reproduction of atmospheric dynamics and climate if other processes are tuned to account for its presence in a climate model. In particular, NGWD schemes are often adjusted to improve the simulation of stratospheric dynamics. In the absence of such tuning, adding in interactive ozone chemistry may degrade performance, which might erroneously be understood to count against including this process in a climate model. Philosophically, this situation is of course dissatisfying. It exemplifies that model behavior reflects the necessary presence of parameterizations that substitute actual climate physics. Changing model physics generally upsets the balance of these parameterizations achieved in a previous tuning step, explaining some increased biases. We assert that interactive ozone chemistry, a costly addition to a climate model, does require retuning of stratospheric climate, or else the benefit of coupling in chemistry may not be obvious.

Data Availability Statement

CMIP6 data are available at https://esgf-node.llnl.gov/search/cmip6/. Specifically, the following datasets are used: Dix et al. (2019); Danabasoglu (2019b, 2019a, 2019d, 2019c); Séférian (2018); Voldoire (2018); Guo et al. (2018); Krasting et al. (2018); Yukimoto, Koshiro, et al. (2019); Wieners et al. (2019); Tang et al. (2019); Byun (2020); Ridley et al. (2019a, 2019b). CCMI2 data are downloaded from ftp://ftp.ceda.ac.uk/badc/ccmi/data/post-cmip6/ccmi-2022. Specifically, the following datasets have been used: Dennison and Woodhouse (2021); Rozanov et al. (2021); Abraham and Keeble (2021). NCEP-DOE2 data were provided by the NOAA/OAR/ESRL PSL, Boulder, Colorado, USA, from their web site at https://psl.noaa.gov/data/gridded/data.ncep.reanalysis2.pressure.html. Hersbach et al. (2018, 2019) was downloaded from the Copernicus Climate Change Service (C3S) Climate Data Store. The results contain modified Copernicus Climate Change Service information. Neither the European Commission nor ECMWF are responsible for any use that may be made of the Copernicus information or data it contains. MSR-2 data are available at https://www.temis.nl/protocols/O3global.php (van der A et al., 2015b). Scripts and intermediate data used in the generation of the figures of this paper may be downloaded at Morgenstern et al. (2022).

Acknowledgments
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. We also acknowledge IGAC, SPARC, and CEDA for their support of CCMI. We acknowledge NOAA/OAR/ESRL PSL, Boulder, Colorado, USA, for providing the NCEP/DOE Reanalysis 2 data, ECMWF for providing the ERA5 data, and KNMI for providing the MSR-2 total-column ozone climatology. OM and GZ were supported by the NZ Government's Strategic Science Investment Fund (SSIF) through the NIWA programme CACV. MD was supported by the Japan Society for the Promotion of Science KAKENHI (grant number: JP20K04070). FMO'C and YT were supported by the Met Office Hadley Centre Climate Programme funded by BEIS. PL is under award NA18OAR4320123 from the National Oceanic and Atmospheric Administration, U.S. Department of Commerce. The statements, findings, conclusions, and recommendations are those of the author(s) and do not necessarily reflect the views of the National Oceanic and Atmospheric Administration, or the U.S. Department of Commerce. All calculations with the SOCOLv4.0 model were supported by the Swiss National Supercomputing Centre (CSCS) under projects S-901 (ID 154), S-1029 (ID 249), and S-903. ER, TE, and TS were supported by the Swiss National Science Foundation (SNSF) project POLE (Grant No. 200020-182239). TS and ER have also been partly supported by the Ministry of Science and Higher Education of the Russian Federation (Grant No. 075-15-2021-583). Open access publishing facilitated by National Institute of Water and Atmospheric Research, as part of the Wiley - National Institute of Water and Atmospheric Research agreement via the Council of Australian University Librarians.

References

Abraham, N. L., & Keeble, J. (2021). CCMI-2022: REF-D1 data produced by the UKESM1-StratTrop model at NCAS Cambridge. NERC EDS Centre for Environmental Data Analysis. Retrieved from https://data.ceda.ac.uk/badc/ccmi/data/post-cmip6/ccmi-2022/NCAS-CAMBRIDGE/UKESM1-StratTrop/refD1

Baldwin, M. P., & Dunkerton, T. J. (2001). Stratospheric harbingers of anomalous weather regimes. *Science*, *294*(5542), 581–584. https://doi.org/10.1126/science.1063315

Bi, D., Dix, M., Marsland, S., O'Farrell, S., Sullivan, A., Bodman, R., et al. (2020). Configuration and spin-up of ACCESS-CM2, the new generation Australian Community Climate and Earth System Simulator Coupled Model. *Journal of Southern Hemisphere Earth Systems Science*, *70*(1), 225–251. https://doi.org/10.1071/es19040

Bodman, R. W., Karoly, D. J., Dix, M. R., Harman, I. N., Srbinovsky, J., Dobrohotoff, P. B., & Mackallah, C. (2020). Evaluation of CMIP6 AMIP climate simulations with the ACCESS-AM2 model. *Journal of Southern Hemisphere Earth Systems Science*, *70*(1), 166–179. https://doi.org/10.1071/ES19033

Byun, Y.-H. (2020). NIMS-KMA UKESM1.0-LL model output prepared for CMIP6 CMIP historical. *Earth System Grid Federation*. https://doi.org/10.22033/ESGF/CMIP6.8379

Checa-Garcia, R., Hegglin, M. I., Kinnison, D., Plummer, D. A., & Shine, K. P. (2018). Historical tropospheric and stratospheric ozone radiative forcing using the CMIP6 database. *Geophysical Research Letters*, *45*(7), 3264–3273. https://doi.org/10.1002/2017GL076770

Danabasoglu, G. (2019a). NCAR CESM2 model output prepared for CMIP6 CMIP historical. *Earth System Grid Federation*. https://doi.org/10.22033/ESGF/CMIP6.7627

Danabasoglu, G. (2019b). NCAR CESM2-FV2 model output prepared for CMIP6 CMIP historical. *Earth System Grid Federation*. https://doi.org/10.22033/ESGF/CMIP6.11297

Danabasoglu, G. (2019c). NCAR CESM2-WACCM model output prepared for CMIP6 CMIP historical. *Earth System Grid Federation*. https://doi.org/10.22033/ESGF/CMIP6.10071

Danabasoglu, G. (2019d). NCAR CESM2-WACCM-FV2 model output prepared for CMIP6 CMIP historical. *Earth System Grid Federation*. https://doi.org/10.22033/ESGF/CMIP6.11298

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., et al. (2020). The Community Earth System Model version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, *12*(2), e2019MS001916. https://doi.org/10.1029/2019ms001916

Dennison, F., & Woodhouse, M. (2021). CCMI-2022: REF-D1 data produced by the ACCESS-CM2-Chem model at CSIRO-ARCCSS. *NERC EDS Centre for Environmental Data Analysis*. Retrieved from https://data.ceda.ac.uk/badc/ccmi/data/post-cmip6/ccmi-2022/CSIRO/ACCESS-CM2-Chem/refD1

Dix, M., Bi, D., Dobrohotoff, P., Fiedler, R., Harman, I., Law, R., et al. (2019). CSIRO-ARCCSS ACCESS-CM2 model output prepared for CMIP6 CMIP AMIP. *Earth System Grid Federation*. https://doi.org/10.22033/ESGF/CMIP6.4239

Dunne, J. P., Horowitz, L. W., Adcroft, A. J., Ginoux, P., Held, I. M., John, J. G., et al. (2020). The GFDL Earth System Model version 4.1 (GFDL-ESM4.1): Overall coupled model description and simulation characteristics. *Journal of Advances in Modeling Earth Systems*, *12*(11), e2019MS002015. https://doi.org/10.1029/2019MS002015

Eichinger, R., Garny, H., Šácha, P., Dietmuller, S., & Oberlander-Hayn, S. (2020). Effects of missing gravity waves on stratospheric dynamics; part 1: Climatology. *Climate Dynamics*, *54*(5–6), 3165–3183. https://doi.org/10.1007/s00382-020-05166-w

Eyring, V., Gillett, N., AchutaRao, K., Barimalala, R., Parrillo, M. B., Bellouin, N., et al. (2021). Human Influence on the Climate System. In V. Masson-Delmotte et al. (Eds.), Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press.

Fogt, R. L., & Marshall, G. J. (2020). The Southern Annular Mode: Variability, trends, and climate impacts across the Southern Hemisphere. *WIREs Climate Change*, *11*(4), e652. https://doi.org/10.1002/wcc.652

Friedel, M., Chiodo, G., Stenke, A., Domeisen, D., Fueglistaler, S., Anet, J., & Peter, T. (2022). Springtime arctic ozone depletion forces Northern Hemisphere climate anomalies. *Nature Geoscience*, *15*(7), 541–547. https://doi.org/10.1038/s41561-022-00974-7

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*(23), 12380–12403. https://doi.org/10.1029/2019jd030943

Guo, H., John, J. G., Blanton, C., McHugh, C., Nikonov, S., Radhakrishnan, A., et al. (2018). NOAA-GFDL GFDL-CM4 model output historical. *Earth System Grid Federation*. https://doi.org/10.22033/ESGF/CMIP6.8594

Haase, S., & Matthes, K. (2019). The importance of interactive chemistry for stratosphere–troposphere coupling. *Atmospheric Chemistry and Physics*, *19*(5), 3417–3432. https://doi.org/10.5194/acp-19-3417-2019

Hardiman, S. C., Andrews, M. B., Andrews, T., Bushell, A. C., Dunstone, N. J., Dyson, H., et al. (2019). The impact of prescribed ozone in climate projections run with HadGEM3-GC3.1. *Journal of Advances in Modeling Earth Systems*, *11*(11), 3443–3453. https://doi.org/10.1029/2019ms001714

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*(11), 3691–3727. https://doi.org/10.1029/2019ms001829

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., et al. (2018). ERA5 hourly data on pressure levels from 1959 to present. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*. https://doi.org/10.24381/cds.bd0915c6

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., et al. (2019). ERA5 monthly averaged data on pressure levels from 1979 to present. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*. https://doi.org/10.24381/cds.6860a573

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Ivy, D. J., Solomon, S., Calvo, N., & Thompson, D. W. J. (2017). Observed connections of Arctic stratospheric ozone extremes to Northern Hemisphere surface climate. *Environmental Research Letters*, *12*(2), 024004. https://doi.org/10.1088/1748-9326/aa57a4

Jucker, M., Reichler, T., & Waugh, D. W. (2021). How frequent are Antarctic sudden stratospheric warmings in present and future climate? *Geophysical Research Letters*, *48*(11), e2021GL093215. https://doi.org/10.1029/2021gl093215

Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S.-K., Hnilo, J. J., Fiorino, M., & Potter, G. L. (2002). *NCEP-DOE AMIP-II reanalysis (R-2)* (pp. 1631–1643). Bulletin of the American Meteorological Society. https://doi.org/10.1175/bams83111631(2002)083<1631nar>2.3.co;2

Krasting, J. P., John, J. G., Blanton, C., McHugh, C., Nikonov, S., Radhakrishnan, A., et al. (2018). NOAA-GFDL GFDL-ESM4 model output prepared for CMIP6 CMIP historical. *Earth System Grid Federation*. https://doi.org/10.22033/ESGF/CMIP6.8597

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

Kuttippurath, J., Feng, W., Müller, R., Kumar, P., Raj, S., Gopikrishnan, G. P., & Roy, R. (2021). Exceptional loss in ozone in the Arctic winter/spring of 2019/2020. *Atmospheric Chemistry and Physics*, *21*(18), 14019–14037. https://doi.org/10.5194/acp-21-14019-2021

Lin, P., & Ming, Y. (2021). Enhanced climate response to ozone depletion from ozone-circulation coupling. *Journal of Geophysical Research: Atmospheres*, *126*(7), e2020JD034286. https://doi.org/10.1029/2020jd034286

Lin, P., Paynter, D., Polvani, L., Correa, G. J. P., Ming, Y., & Ramaswamy, V. (2017). Dependence of model-simulated response to ozone depletion on stratospheric polar vortex climatology. *Geophysical Research Letters*, *44*(12), 6391–6398. https://doi.org/10.1002/2017gl073862

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., et al. (2019). Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and its response to increasing CO_2. *Journal of Advances in Modeling Earth Systems*, *11*(4), 998–1038. https://doi.org/10.1029/2018ms001400

Michou, M., Nabat, P., Saint-Martin, D., Bock, J., Decharme, B., Mallet, M., et al. (2020). Present-day and historical aerosol and ozone characteristics in CNRM CMIP6 simulations. *Journal of Advances in Modeling Earth Systems*, *12*(1), e2019MS001816. https://doi.org/10.1029/2019ms001816

Morgenstern, O. (2021). The Southern Annular Mode in 6th Coupled Model Intercomparison Project Models. *Journal of Geophysical Research: Atmospheres*, *126*(5), e2020JD034161. https://doi.org/10.1029/2020jd034161

Morgenstern, O., Frith, S. M., Bodeker, G. E., Fioletov, V., & van der A, R. J. (2021). Reevaluation of total-column ozone trends and of the effective radiative forcing of ozone-depleting substances. *Geophysical Research Letters*, *48*(21), e2021GL095376. https://doi.org/10.1029/2021gl095376

Morgenstern, O., Kinnison, D., Mills, M., Michou, M., Horowitz, L., Lin, P., et al. (2022). Comparison of Arctic and Antarctic stratospheric climates in chemistry versus no-chemistry climate models. *Zenodo*. https://doi.org/10.5281/zenodo.6972644

Morgenstern, O., O'Connor, F. M., Johnson, B. T., Zeng, G., Mulcahy, J. P., Williams, J., et al. (2020). Reappraisal of the climate impacts of ozone-depleting substances. *Geophysical Research Letters*, *47*(20), e2020GL088295. https://doi.org/10.1029/2020gl088295

Newman, P., & Nash, E. (2003). The unusual Southern Hemisphere stratosphere winter of 2002. *Journal of the Atmospheric Sciences*, *62*(3), 614–628. https://doi.org/10.1175/JAS-3323.1

Oehrlein, J., Chiodo, G., & Polvani, L. M. (2020). The effect of interactive ozone chemistry on weak and strong stratospheric polar vortex events. *Atmospheric Chemistry and Physics*, *20*(17), 10531–10544. https://doi.org/10.5194/acp-20-10531-2020

Plummer, D., Nagashima, T., Tilmes, S., Archibald, A., Chiodo, G., Fadnavis, S., et al. (2021). CCMI-2022: A new set of Chemistry-Climate Model Initiative (CCMI) community simulations to update the assessment of models and support upcoming Ozone Assessment activities. *SPARC Newsletter*, *57*, 22–30.

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., & Kaplan, A. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research*, *108*(D14), 4407. https://doi.org/10.1029/2002jd002670

Richter, J. H., Sassi, F., & Garcia, R. R. (2010). Toward a physically based gravity wave source parameterization in a general circulation model. *Journal of the Atmospheric Sciences*, *67*(1), 136–156. https://doi.org/10.1175/2009jas3112.1

Ridley, J., Menary, M., Kuhlbrodt, T., Andrews, M., & Andrews, T. (2019a). MOHC HadGEM3-GC31-LL model output prepared for CMIP6 CMIP AMIP. *Earth System Grid Federation*. https://doi.org/10.22033/ESGF/CMIP6.5853

Ridley, J., Menary, M., Kuhlbrodt, T., Andrews, M., & Andrews, T. (2019b). MOHC HadGEM3-GC31-LL model output prepared for CMIP6 CMIP historical. *Earth System Grid Federation*. https://doi.org/10.22033/ESGF/CMIP6.6109

Rieder, H. E., Chiodo, G., Fritzer, J., Wienerroither, C., & Polvani, L. M. (2019). Is interactive ozone chemistry important to represent polar cap stratospheric temperature variability in Earth-system models? *Environmental Research Letters*, *14*(4), 044026. https://doi.org/10.1088/1748-9326/ab07ff

Rozanov, E., Egorova, T., & Sukhodolov, T. (2021). *CCMI-2022: REF-D1 data produced by the SOCOL model at ETH-PMOD*. NERC EDS Centre for Environmental Data Analysis. Retrieved from https://catalogue.ceda.ac.uk/uuid/f088e7a33a25409197ea9b6aa3b90864

Séférian, R. (2018). *CNRM-CERFACS CNRM-ESM2-1 model output prepared for CMIP6 CMIP historical*. Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.4068

Séférian, R., Nabat, P., Michou, M., Saint-Martin, D., Voldoire, A., Colin, J., et al. (2019). Evaluation of CNRM Earth System Model, CNRM-ESM2-1: Role of Earth system processes in present-day and future climate. *Journal of Advances in Modeling Earth Systems*, *11*(12), 4182–4227. https://doi.org/10.1029/2019ms001791

Sellar, A. A., Jones, C. G., Mulcahy, J. P., Tang, Y., Yool, A., Wiltshire, A., et al. (2019). UKESM1: Description and evaluation of the U.K. Earth System Model. *Journal of Advances in Modeling Earth Systems*, *11*(12), 4513–4558. https://doi.org/10.1029/2019MS001739

Son, S.-W., Gerber, E. P., Perlwitz, J., Polvani, L. M., Gillett, N. P., Seo, K.-H., et al. (2010). Impact of stratospheric ozone on Southern Hemisphere circulation change: A multimodel assessment. *Journal of Geophysical Research*, *115*(D3), D00M07. https://doi.org/10.1029/2010jd014271

Sukhodolov, T., Egorova, T., Stenke, A., Ball, W. T., Brodowsky, C., Chiodo, G., et al. (2021). Atmosphere–ocean–aerosol–chemistry–climate model SOCOLv4.0: Description and evaluation. *Geoscientific Model Development*, *14*(9), 5525–5560. https://doi.org/10.5194/gmd-14-5525-2021

Tang, Y., Rumbold, S., Ellis, R., Kelley, D., Mulcahy, J., Sellar, A., et al. (2019). MOHC UKESM1.0-LL model output prepared for CMIP6 CMIP historical. *Earth System Grid Federation*. https://doi.org/10.22033/ESGF/CMIP6.6113

Taylor, K. E., Williamson, D., & Zwiers, F. (2015). AMIP sea surface temperature and sea ice concentration boundary conditions. Retrieved from https://pcmdi.llnl.gov/mips/amip/details/index.html

Thompson, D., Baldwin, M., & Solomon, S. (2005). Stratosphere–troposphere coupling in the Southern Hemisphere. *Journal of the Atmospheric Sciences*, *3*(62), 708–715. https://doi.org/10.1175/jas-3321.1

Thompson, D., Solomon, S., Kushner, P., England, M. H., Grise, K. M., & Karoly, D. J. (2011). Signatures of the Antarctic ozone hole in Southern Hemisphere surface climate change. *Nature Geoscience*, *4*(11), 741–749. https://doi.org/10.1038/ngeo1296

van der A, R. J., Allaart, M. A. F., & Eskes, H. J. (2015a). Extended and refined Multi Sensor Reanalysis of total ozone for the period 1970–2012. *Atmospheric Measurement Techniques*, *8*(7), 3021–3035. https://doi.org/10.5194/amt-8-3021-2015

van der A, R. J., Allaart, M. A. F., & Eskes, H. J. (2015b). *Multi-Sensor Reanalysis (MSR) of total ozone, version 2 dataset*. Royal Netherlands Meteorological Institute (KNMI). https://doi.org/10.21944/temis-ozone-msr2

Voldoire, A. (2018). CMIP6 simulations of the CNRM-CERFACS based on CNRM-CM6-1 model for CMIP experiment historical. *Earth System Grid Federation*. https://doi.org/10.22033/ESGF/CMIP6.4066

Voldoire, A., Saint-Martin, D., Sénési, S., Decharme, B., Alias, A., Chevallier, M., et al. (2019). Evaluation of CMIP6 DECK experiments with CNRM-CM6-1. *Journal of Advances in Modeling Earth Systems*, *11*(7), 2177–2213. https://doi.org/10.1029/2019ms001683

Wieners, K.-H., Giorgetta, M., Jungclaus, J., Reick, C., Esch, M., Bittner, M., et al. (2019). MPI-M MPI-ESM1.2-LR model output prepared for CMIP6 CMIP AMIP. *Earth System Grid Federation*. https://doi.org/10.22033/ESGF/CMIP6.6464

Williams, K. D., Copsey, D., Blockley, E. W., 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

WMO. (2018). *Scientific Assessment of Ozone Depletion: 2018*. World Meteorological Organization. Retrieved from https://www.esrl.noaa.gov/csl/assessments/ozone/2018/

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-ESM2.0: Description and basic evaluation of the physical component. *Journal of the Meteorological Society of Japan Series II*, *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-ESM2.0 model output prepared for CMIP6 CMIP historical. *Earth System Grid Federation*. https://doi.org/10.22033/ESGF/CMIP6.6842