The Southern Annular Mode in 6th Coupled Model Intercomparison Project models

O. Morgenstern[1]

[1]National Institute of Water and Atmospheric Research (NIWA), Wellington, New Zealand

Key Points:

- The influences of ozone changes and greenhouses gases on the Southern Annular Mode are analyzed in CMIP6 simulations.
- Ozone depletion exerts a stronger and GHGs a weaker influence on the SAM in chemistry versus no-chemistry models.
- Three sensitivity experiments are explained considering an impact of GHGs onto the SAM resulting from ozone changes.

Corresponding author: Olaf Morgenstern, olaf.morgenstern@niwa.co.nz

Abstract

We analyze trends in the Southern Annular Mode in CMIP6 simulations. For the period 1957-2014, simulated linear trends are generally consistent with two observational references but seasonally in disagreement with two other reconstructions of the SAM. Using a regression analysis applied to model simulations with interactive ozone chemistry, a strengthening of the SAM in summer is attributed completely to ozone depletion because a further strengthening influence due to long-lived greenhouse gases is fully counterbalanced by a weakening influence due to stratospheric ozone increases associated with these greenhouse gas increases. Ignoring such ozone feedbacks would yield comparable contributions from these two influences, an incorrect result. In winter, trends are smaller but an influence of greenhouse gas-mediated ozone feedbacks is also identified. The regression analysis furthermore yields significant differences in the attribution of SAM changes to the two influences between models with and without interactive ozone chemistry, with ozone depletion and GHG increases playing seasonally a stronger and weaker, respectively, role in the chemistry models versus the no-chemistry ones.

1 Introduction

The Southern Annular Mode (SAM) is the leading mode of variability of the Southern Hemisphere (Gong & Wang, 1999). In essence, it constitutes a see-saw of atmospheric mass between Southern high and middle latitudes, with corresponding expressions in surface pressure, the position and strength of the westerly jet, temperature, and precipitation. Its development under global climate change is a leading concern to countries in the Southern Hemisphere.

The SAM is subject to a variety of external influences, the leading of which are increasing anthropogenic greenhouse gases (GHGs) and ozone depletion due to chlorofluorocarbons and other anthopogenic ozone-depleting substances (ODSs, e.g. Arblaster & Meehl, 2006; Kang et al., 2011; Thompson et al., 2011). Previous studies have identified that ozone depletion is the leading cause of of a strengthening of the SAM during austral summer (DJF,e.g. Son et al., 2009, 2010; Eyring et al., 2013; Gillett & Fyfe, 2013; Son et al., 2018). In other seasons, however, trends are smaller and other influences play a role (Marshall, 2003; Arblaster & Meehl, 2006). Difficulties here include that the SAM exhibits considerable internal variability obscuring to some extent external influences. Observational references for the period before the onset of the satellite era are based on ground-based measurements which are sparse or even nonexistent, leaving large areas of the footprint of the mode without any long-term observations extending into the 19^{th} century. This is particularly so for the Antarctic continent and the remote Southern Ocean

There is a considerable range of responses of climate models to ODS and GHG changes. Despite the role of ozone depletion in driving trends in the summertime SAM (Thompson et al., 2011), some previous studies have found that modelling ozone interactively in global climate models does not lead to a significantly different trend in the SAM versus prescribing ozone from a precomputed climatology (Son et al., 2018), the method pursued in most contemporary climate models. However, aspects of variability of the SAM, and particularly its extreme states (Dennison et al., 2015), are affected by the type of representation of ozone (Haase & Matthes, 2019; Haase et al., 2020), and also there are interactions between greenhouse gases, ozone, and the SAM that would not be represented in models that use prescribed ozone (Morgenstern et al., 2014). The few models participating in the 5^{th} Coupled Model Intercomparison Project (CMIP5) which had interactive stratospheric chemistry were characterized by various deficiencies in their simulations of ozone and hence did not produce obviously higher-quality projections of other climate variables versus models that used simpler representations of ozone (Eyring et al., 2013).

The 6th Coupled Model Intercomparison Project (CMIP6) brings together the latest generation of climate models to produce simulations informing the upcoming 6th Assessment Report of IPCC (Eyring et al., 2016). These models have generally undergone further development since CMIP5. Also there is a much more diverse range of sensitivity experiments available than under CMIP5, targeting a large variety of forcing types and processes aimed at better characterizing and improving understanding of how models respond to forcings. Using three of these experiments and many "historical" all-forcings simulations, here we will conduct a seasonally resolved attribution of trends in the SAM to the two leading influences, ODSs and GHGs.

2 Definition of the SAM

While not perfectly "annular", the SAM is characterized by a large zonally symmetric component. Hence for simplicity we here follow Gong and Wang (1999) and define the SAM index to be the difference in monthly- and zonal-mean sea-level pressure (psl), or for above-surface features geopotential height on pressure levels (zg), between 40°S and 65°S. For every simulation the SAM index is smoothed with a three-months boxcar filter (such that the SAM index represents the seasonal mean centred on a given month). The first and last months of each dataset are invalidated.

The results are, for every model, month of the year, and simulation covered by that model, a timeline of seasonal SAM indices covering 1850-2014 (1851-2014 for DJF, 1850-2013 for NDJ).

3 Method of data analysis

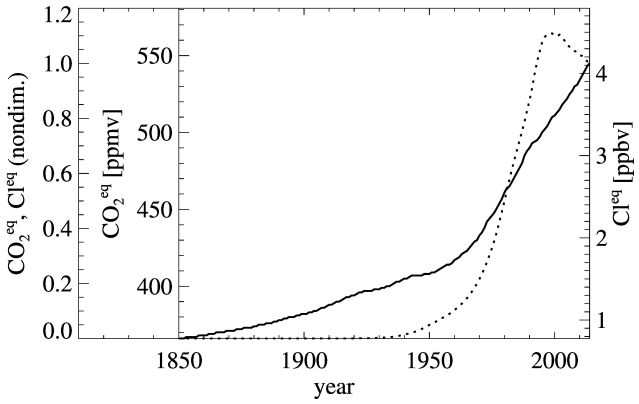

Figure 1. (solid) Equivalent CO_2 (CO_2^{eq}, in ppmv) and (dotted) equivalent chlorine (Cl^{eq}, in ppbv) in the historical simulations of CMIP6. See text for definitions of both quantities. Nondimensional scaling (far left axis) is used in the regression analysis.

For the purposes of attributing any long-term trends in these data using a multiple-linear regression analysis, we form two regressor functions (Morgenstern et al., 2018), both using forcing data provided by Meinshausen et al. (2017):

- Equivalent chlorine (Cl^{eq}) is the sum of the abundances of all chlorinated or brominated ODS gases weighted with the numbers of chlorine and bromine atoms per molecule, and additionally multiplying any bromine source gas by a factor of 60 to account for the larger per-atom depletion of ozone caused by bromine than by

chlorine (Newman et al., 2007). Cl^{eq} is shifted by four years to account for the time it takes for the ODSs to be delivered into the stratosphere.

- Equivaent CO_2 (CO_2^{eq}) is the sum of the surface abundances of all long-lived GHGs weighted by their specific radiative efficiencies divided by that of CO_2, with specific radiative efficiency coefficients taken from table 8.A.1 of AR5 (Myhre et al., 2013).

Both regression functions are normalized such that the functions equal 0 at the start of the record (1850) and increase to 1 in 2014 (figure 1). $Cl^{eq} = 0$ until about 1950 followed by a ramp-up in the 1970s to 1990s to values exceeding 1 and a slow decay in the 21st century, reflecting the removal of ODSs from the atmosphere after emissions of ODSs have mostly ceased. By contrast, CO_2^{eq} is increasing throughout the "historical" period, with substantial increases already between 1850 and 1950 and a speed-up in the latter decades. The shapes of these two functions mean that any trend in the SAM index during 1850-1950 will be reflected in a projection onto CO_2^{eq}.

We thus determine, using least-squares linear regression, coefficients S_0, S_1, and S_2 in the function

$$S(m,y) = S_0(m) + S_1(m)\mathrm{CO}_2^{eq}(y) + S_2(m)\mathrm{Cl}^{eq}(y) + \epsilon(m,y) \tag{1}$$

such that the residual ϵ is minimized in the root-mean-square metric. Here S is an observed or modelled SAM index, a function of month m and calendar year y. We interpret the terms as S_0 standing for the baseline mean seasonal cycle of the SAM index in preindustrial times, S_1 the change in the SAM index driven by long-lived greenhouse gases, reflecting in particular any trends in the first 100 years when $Cl^{eq} = 0$, S_2 the change in the SAM index driven by ODSs which have been elevated from the 1960s onwards, and $S_1 + S_2$ the total strengthening due to both anthropogenic causes in 2014. Both S_1 and S_2 are seasonally resolved for the 12 overlapping seasons of the year.

We furthermore model the residual ϵ using an analogous least-squares regression approach:

$$\epsilon^2(m,y) = E_0(m) + E_1(m)\mathrm{CO}_2^{eq}(y) + E_2(m)\mathrm{Cl}^{eq}(y) + \mu(m,y) \tag{2}$$

In this approach, ϵ^2 is assumed to also vary with the two regressor functions. Where the regression fit produces negative numbers for any model or season (which would indicate the regression fit is poor) we replace equation 2 with $E_1(m) = E_2(m) = 0$, $E_0(m) = \overline{\epsilon^2(m,.)}$ (i.e. the regression becomes a constant in time). For the ensemble-mean SAM indices derived from "historical" simulations detailed below, this is the case only for one month each in two models (FIO-ESM-2-0 in November, NorCPM1 in March; see below).

If the regression model (equation 1) was perfect and ϵ was normally distributed, the variance terms would scale with the inverse of the ensemble size n where single-model ensemble means are considered: $E_i \sim 1/n$.

The regression model developed above is complemented with a simpler linear trend analysis conducted on the period since 1957 when permanent meteorological observations started in Antarctica. If T_{ij} is the linear trend in ensemble member j of model i, and σ_{ij} its uncertainty at 68% confidence, then

$$p_{ij}(T) = G\left(\frac{T - T_{ij}}{\sigma_{ij}}\right) \tag{3}$$

is the probability that a given trend T is larger than the best-estimate trend, T_{ij}. Here, G is the Gaussian integral,

$$G(x) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{x} \exp\left(-\xi^2\right) \mathrm{d}\xi. \tag{4}$$

We thus form the weighted mean of this distribution over all CMIP6 historical simulations

$$P(T) = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{n_i} \sum_{j=1}^{n_i} p_{ij}(T) \tag{5}$$

Here $P(T)$ is the cumulative probability distribution function that any given trend T exceeds the trend derived from a random CMIP6 model simulation, thus accounting for model as well as statistical uncertainties in these trends. Here m is the number of models in the ensemble and n_i is the number of historical simulations provided by model i. The trends T for which $P(T)$ evaluates to 0.5, 0.16, 0.84, 0.025 and 0.975 mark the multi-model mean and the 68 and 95% uncertainty ranges of the combined distribution of trends in the CMIP6 ensemble. Unlike other forms of averaging, this method does not inherently reduce the statistical uncertainty as ensemble sizes or model numbers increase, meaning that the thus obtained uncertainty range of trends remain comparable to those derived from observations. We also note that in equation 5 every model enters the averaging with equal weight, irrespective of n_i.

$P(T)$ is evaluated separately for all 12 overlapping seasons of the year.

4 Data and models

Models used in the below analysis are listed in table 1. We use almost all such simulations available at the time of download for the "historical", hist-GHG, hist-stratO3, and hist-1950HC experiments (for definitions of the experiments see below). To limit model redundancy, some model variants that are nearly identical to one of the models used here are not included, for example high-resolution versions of some models. Also models for which only one historical simulation is available are generally not used, with the exception of GFDL-CM4 which is retained because it is the basis of the chemistry-model GFDL-ESM4. In total, we consider 282 historical, 61 hist-GHG, 28 hist-stratO3, and 8 hist-1950HC simulations.

Briefly, the experiments are characterized as follows:

- "historical": This all-forcings experiment covering 1850-2014 is conducted by all CMIP6 models. Models and simulations used here are listed in table 1.
- "hist-1950HC": This experiment is identical to "historical" except it covers only 1950-2014 and ODSs are kept at their 1950 abundances. The five models participating in this experiment all have interactive stratospheric ozone chemistry, thus consistently representing the impact of ODSs.
- "hist-GHG": In this experiments all forcings are kept at their 1850 values except for greenhouse gases. Ozone is also kept invariant in this experiment. Ten models have participated in this experiment.
- "hist-stratO3": In this experiment all forcings are kept at their 1850 status except for stratospheric ozone which follows the CMIP6 ozone climatology. This experiment covers 1850-2020.

In addition to CMIP6 simulations, we also consider various observational references for the SAM (table 2).

Figure 2 shows that even in the decades since 1957 some disagreements occur between the observational references. Apart from some offsets between the datasets, in particular HadSLP2 displays stronger increases in all seasons in the 21st century than the other datasets. In summer (DJF), NOAA-20CR appears to agree well with CERA-20C, but less so in winter (JJA). In the 19th century, some substantial discrepancies appear between the two datasets covering this period (HadSLP2 and NOAA-20CR), reflecting

Model	reference	historical	hist-GHG	hist-stratO3
ACCESS-ESM1-5	Ziehn et al. (2020)	1-10	1-3	
AWI-CM-1-1-MR	Semmler et al. (2020)	1-5		
BCC-CSM2-MR	Wu et al. (2019)	1-3	1-3	
BCC-ESM1	Wu et al. (2020)	1-3		
CAMS-CSM1-0	Chen et al. (2019)	1-2		
CanESM5	Swart et al. (2019)	1-25	1-25	1-10
CESM2	Danabasoglu et al. (2020)	1-11	1-3	
CIESM	Lin et al. (2019)	1-3		
CNRM-CM6-1	Voldoire et al. (2019)	1-30	1,3-10	
E3SM-1-0	Golaz et al. (2019)	1-5		
EC-Earth3	Wyser and et al (2020)	2-4,7,10,12,14,16-25		
FIO-ESM-2-0	Bao et al. (2020)	1-3		
GFDL-CM4	Held et al. (2019)	1		
GISS-E2-1-G	Kelley et al. (2020)	1-10, 101-102	1-5	1-5
HadGEM3-GC31-LL	Williams et al. (2017)	1-4	1-4	
INM-CM5-0	Volodin and Gritsun (2018)	1-10		
IPSL-CM6A-LR	Boucher et al. (2020)	1-32	1,4,6,7,9,10	1-10
KACE-1-0-G	Lee et al. (2020)	1-3		
MIROC6	Tatebe et al. (2019)	1-50	1-3	1-3
MIROC-ES2L	Hajima et al. (2020)	1-10		
MPI-ESM1-2-LR	Müller et al. (2018)	1-10		
NorCPM1	Bethke et al. (2019)	1-30		
NorESM2-LM	Seland et al. (2020)	1-3		
			hist-1950HC	
CESM2-WACCM	Gettelman et al. (2019)	1-3	1	
CNRM-ESM2-1	Séférian et al. (2019)	1-5,7-8	1-2	
GFDL-ESM4	Dunne et al. (2020)	1-3	1	
GISS-E2-1-Gchem	Kelley et al. (2020)	1-6,8-10	1	
MRI-ESM2-0	Yukimoto et al. (2019)	1-5	1,3,5	
UKESM1-0-LL	Sellar et al. (2019)	1-4,8-14,16-19		

Table 1. CMIP6 models considered here, key references, and historical, hist-GHG, hist-stratO3, and hist-1950HC run numbers used in the below analysis. The first 23 models (ACCESS-ESM1-5 to NorESM2-LM) do not use fully interactive ozone; most use CMIP6 prescribed ozone. The bottom six models (CESM2-WACCM to UKESM1-0-LL) use fully interactive formulations of ozone involving transporting and calculating the chemistry of stratospheric ozone. We rename here GISS-E2-1-G to "GISS-E2-1-Gchem" for simulations using interactive ozone chemistry.

the paucity of data coverage during this period. A substantial strengthening of the SAM index is evident from the 1950s onwards.

5 Results

5.1 Synoptic analysis of the SAM in CMIP6 simulations

In a first step we consider the seasonally resolved linear trend in the CMIP6 simulations and the observations for the period of 1957-2014. We stipulate that before 1957, there were no continuous meteorological observations made in Antarctica and therefore SAM reconstructions become more uncertain. Figure 3 shows that the CMIP6 simula-

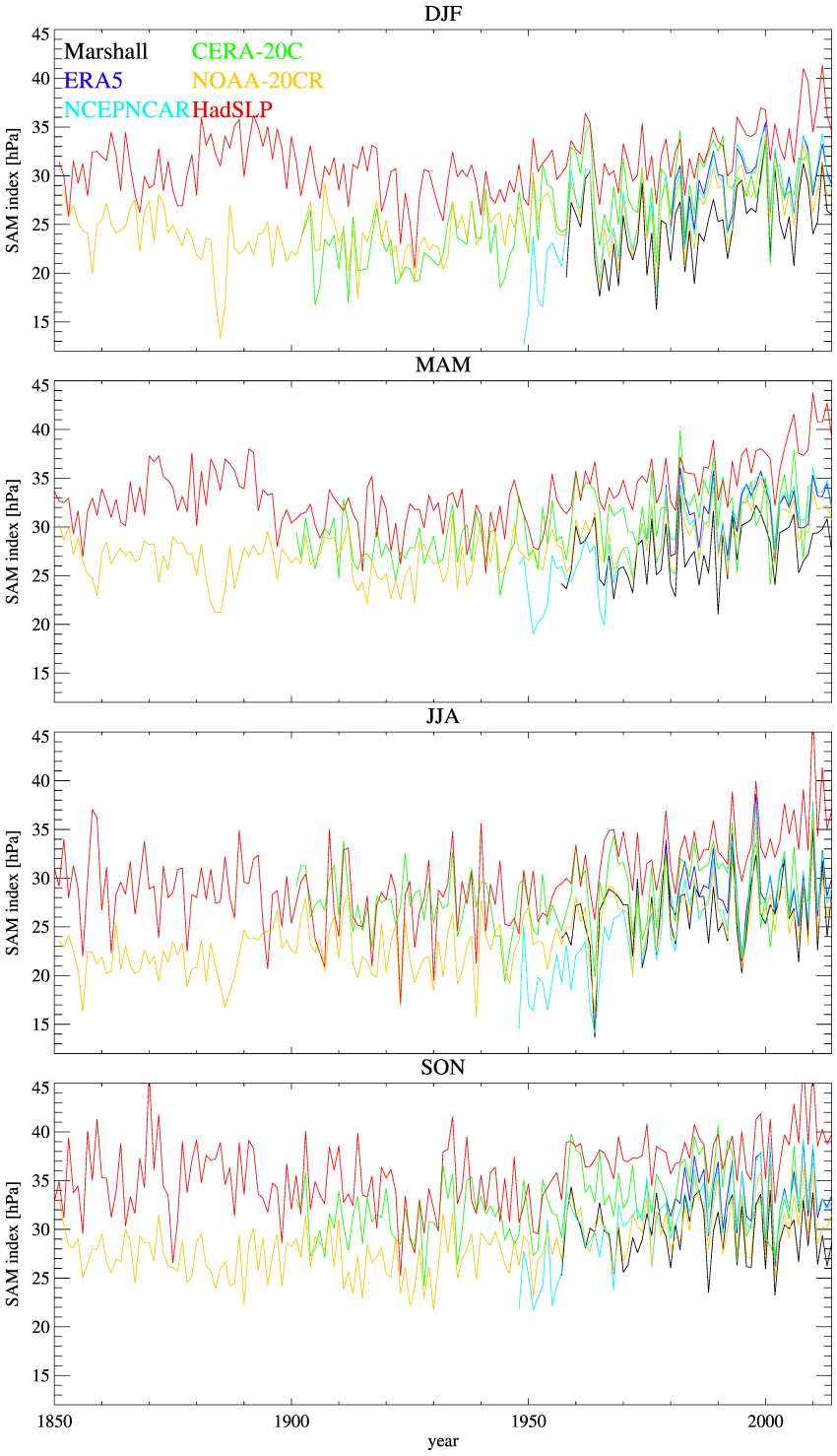

Figure 2. SAM index (hPa) as derived from Marshall (2003), ERA5, NCEP/NCAR, CERA-20C (ensemble mean), NOAA-20CR, and HadSLP2 observational references, for the four seasons.

Reference	literature	Ens. size	Coverage
ERA5	Simmons et al. (2020)	1	1979-2014
NCEP/NCAR	Kalnay et al. (1996)	1	1948-2014
CERA-20C	Laloyaux et al. (2018)	10	1901-2010
NOAA-20CRv2c	Compo et al. (2011)	1	1851-2014
Marshall index	Marshall (2003)	1	1957-2014
HadSLP2	Allan and Ansell (2006)	1	1850-2014

Table 2. Global gridded SLP reconstructions used here. The first four are reanalysis products. ERA5 and NCEP/NCAR are reanalyses using a wide variety of observations, whereas CERA-20C and NOAA-20CRv2c are using surface observations only. HadSLP2 is a gridded interpolation of SLP observations using station and ship data. The Marshall index is a reconstruction of the SAM index using observations from twelve stations only.

tions for 1957-2014 exhibit a linear strengthening trend of around 0.08±0.05 hPa a^{-1} in summer but only around 0.03 ± 0.04 hPa a^{-1} in winter. (Uncertainties are for the 68% confidence level.) In summer, these trends are in general agreement with the observational references (although all four references show stronger best-estimate trends than the CMIP6 mean), but in winter, only the Marshall index and NOAA-20CR are consistent with this ensemble, whereas HadSLP2 and particularly NCEP/NCAR (Marshall, 2003) have trends that are very unlikely to be consistent with the CMIP6 distribution of trends. Given the simplicity of construction of the Marshall (2003) index, and the fact that HadSLP2 shows anomalously high values also for the most recent decades characterized by comparatively excellent data coverage, in the following we will only consider the NOAA-20CR reanalysis for the regression analysis outlined above which uses the full 165-year historical simulations. We note however persistent reservations also about the quality of this dataset for the first century (1851-1956) for which few actual southern high-latitude observations have entered this reanalysis.

Moving now to analyzing the SAM index for the whole "historical" period in the CMIP6 ensemble for the austral summer and winter seasons (figure 4), it is clear that the regression model (equation 1) well approximates the multi-model mean "historical" evolution of the SAM index for all subsets of models displayed. In summer, all four model subsets in the multi-model mean (MMM) exhibit about the same strengthening in 2014 (∼4 hPa) which is larger than in winter (figure 4). In summer, the groups are in general agreement with NOAA-20CR which also shows strengthening, albeit of slightly larger magnitude than the multi-model mean. The regression analysis as well as the hist-1950HC experiment (for the chemistry models) indicate that the strengthening in summer is caused by the growth in ODSs from about 1960 onwards. However, the hist-GHG experiment indicates substantial strengthening (exceeding ∼2 hPa in the MMM) also when all forcings other than the GHGs and particularly ozone are held invariant, and the hist-stratO3 experiment yields a strengthening in summer which is smaller than in the equivalent historical ensemble. We will discuss all of these findings in more detail in section 6.

For the chemistry group, the amount of strengthening found here aligns broadly with the amount of Southern-Hemisphere ozone depletion, with models simulating strong ozone loss (UKESM1, CNRM-ESM2-1) showing a larger strengthening than those with comparatively weak ozone loss (MRI-ESM2, CESM2-WACCM; Morgenstern et al., 2020).

In winter, all model groups indicate much weaker strengthening than in summer which however is in disagreement with an ∼6 hPa strengthening in NOAA-20CR (figure 4). Given the good agreement between the trends in NOAA-20CR and the CMIP6 models for the period post-1957 during which much of the anthropogenic forcing was established (figure 3), it is plausible that the disagreement found here is an artefact of the

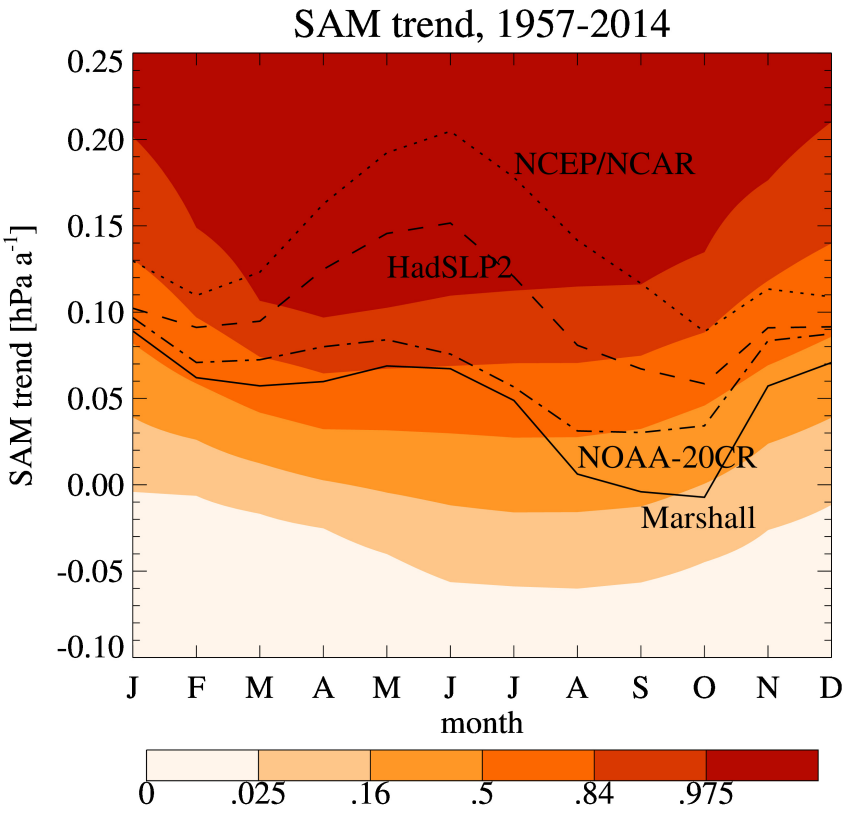

Figure 3. Contours: Weighted mean cumulative probability distribution function P for the SAM trend (equation 5). The 0.5 contour of P is the best-estimate linear trend in the SAM index derived from CMIP6 historical simulations. The other contours mark its 68 and 95% confidence intervals. Lines: trends in observational datasets. Solid: Marshall (2003) index. Dash-dot: NOAA-20CR. Dashed: HadSLP2. Dotted. NCEP/NCAR.

NOAA-20CR data.The regression analysis suggests only a small to no role for ozone depletion. Both the strong influence of ozone depletion in summer and the weak one in winter are in agreement with previous studies (Son et al., 2010; Morgenstern et al., 2014, 2018).

In the following sections we will address three questions, motivated by this analysis:

1. Are the CMIP6 model simulations, as an ensemble, consistent with NOAA-20CR under the regression analysis laid out above?
2. Are there any statistically robust differences between the chemistry and no-chemistry groups of models?
3. How can we reconcile the apparently contradictory findings regarding what is driving the summertime SAM derived from the hist-1950HC, hist-GHG, and hist-stratO3 experiments?

5.2 Validation of the CMIP6 historical simulations versus NOAA-20CR

A fundamental problem here is that there are hundreds of CMIP6 "historical" simulations used here, but only one realization produced by nature, approximated by NOAA-

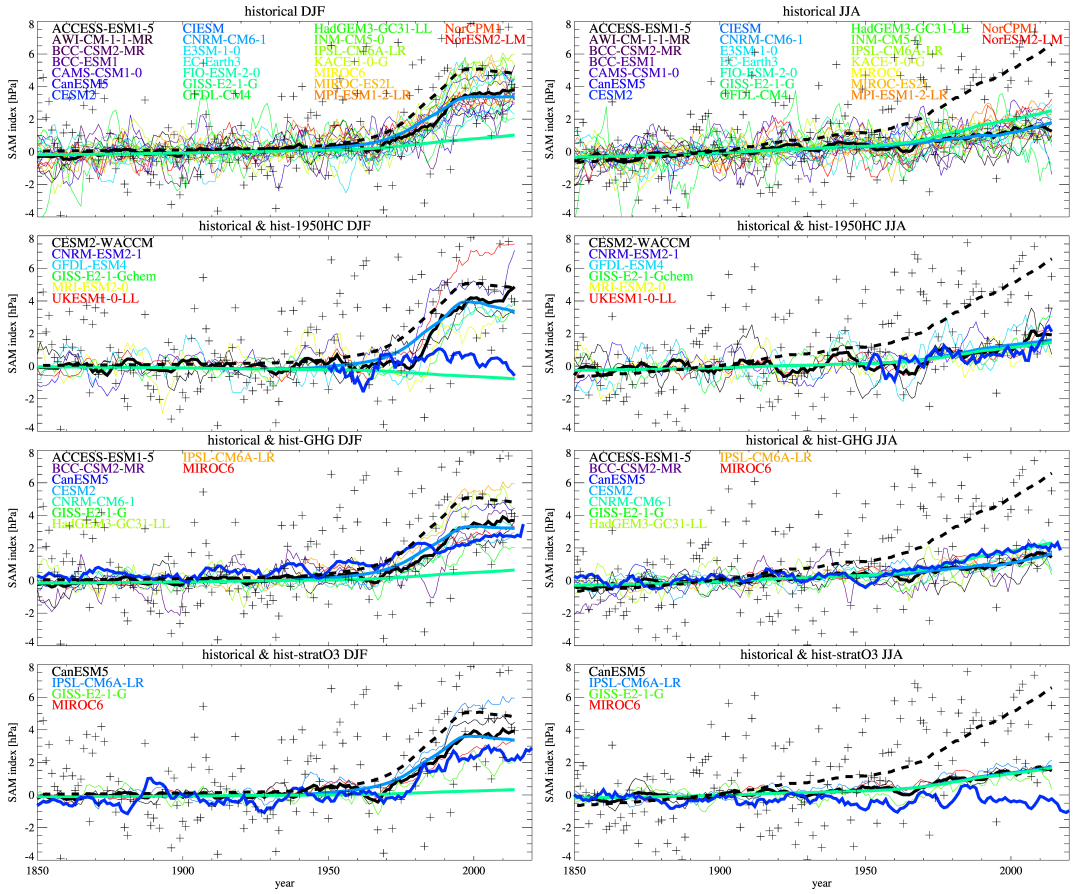

Figure 4. Sea-level pressure based SAM index for (left) austral summer (DJF) and (right) winter (JJA) for (top) all no-chemistry models, baseline-adjusted so the index averages to 0 over 1851-1950. Thin coloured lines: Single-model ensembles means. Thick solid black: Multi-model mean. Thick light blue: Regression fit to multi-model mean. Thick green: Regression component S_1 just due GHGs. '+': NOAA-20CR. Black dashed: Regression fit to NOAA-20CR. 2$^{\text{nd}}$ panel: Same for the six chemistry models. Thick dark blue: Multi-model mean for the hist-1950HC experiment. (2$^{\text{nd}}$ row) Same, for the chemistry models. Thick blue: hist-1950HC multi-model mean (MMM). (3$^{\text{rd}}$ row) Same, for the nine no-chemistry models that have participated in hist-GHG. Thick blue: hist-GHG MMM. (bottom): Same, for four no-chemistry models that have participated in hist-stratO3. Thick blue: hist-stratO3 MMM.

20CR. We thus need to adequately consider natural variability when comparing observations to CMIP6 simulations. Figure 5 shows density plots for the three regression coefficients and their sum S_1+S_2 as derived from individual historical CMIP6 simulations, with every simulation weighted with n_i^{-1}, with n_i being the number of simulations in the historical ensemble of model i. An inspection of the results for individual models (not shown) indicates that the spread in S_0 evident in figure 5 is not the result of natural variability (i.e. spread within single-model ensembles) but rather reflects the differences in the base states of individual models that are seen in all ensemble members with little random variability. However, the models generally well reflect the two maxima in spring and autumn (known in New Zealand as the "windy seasons"), in agreement with NOAA-20CR.

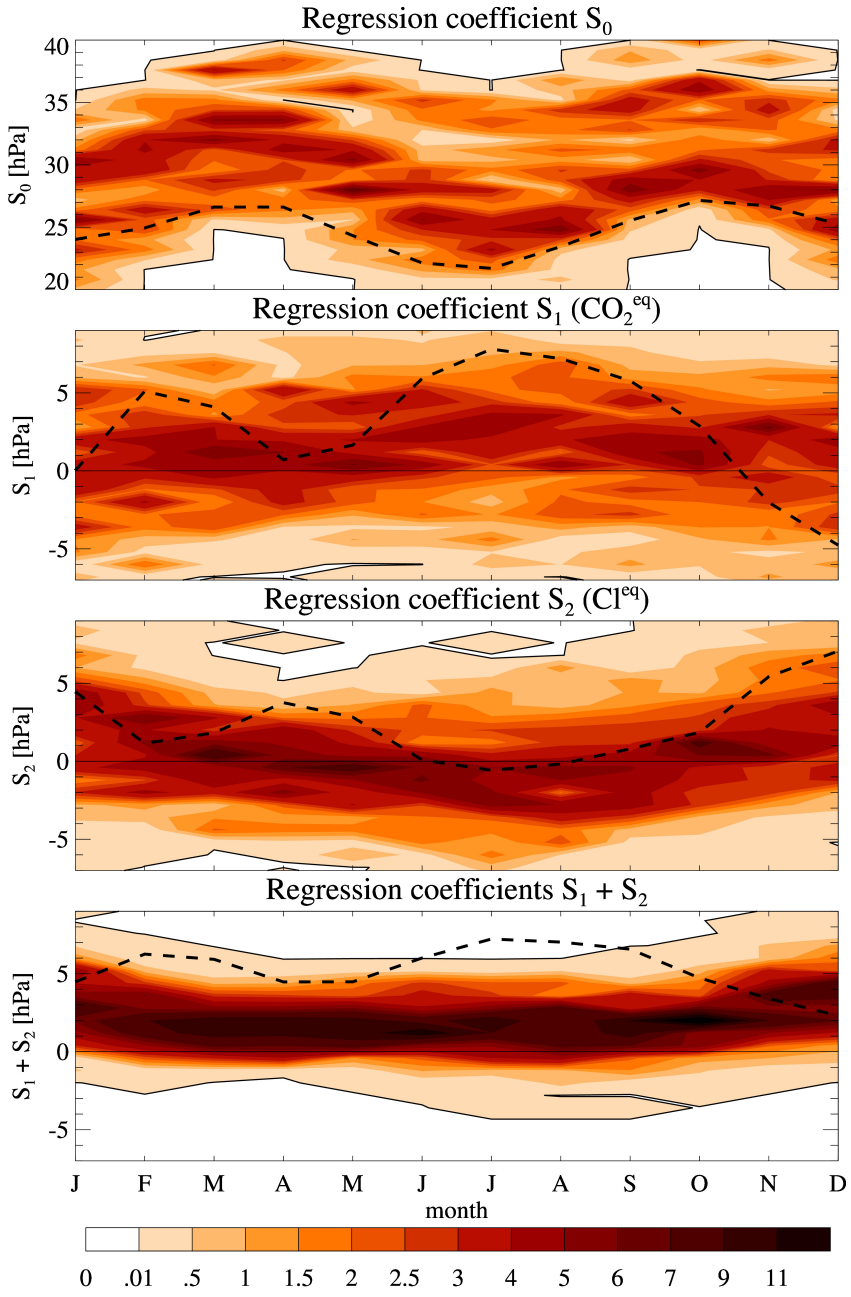

Figure 5. Density plot (number of models/hPa) of the distribution of regression coefficients S_0, S_1, S_2, and $S_1 + S_2$ derived from the CMIP6 historical ensemble. Solid lines: Boundary of the CMIP6 range. Dashed: NOAA-20CR. For every month of the year, we count the number of regression coefficients derived from the CMIP6 ensemble, weighted by the inverse of ensemble size, n^{-1}, that fall into bins sized 0.8 hPa each.

For the component S_1 of the SAM that goes with CO_2^{eq}, during austral summer (DJF) the CMIP6 simulations are roughly evenly divided between positive and negative values, but in austral winter, they favour a positive contribution to the SAM (i.e. increasing GHGs drive a strengthening of the SAM). An inspection of individual model results shows that the spread in S_1 has contributions both due to natural variability and inter-model disagreements. NOAA-20CR is generally within the CMIP6 range for all months but in different seasons (JJA, NDJ) tracks close to both ends of the range spanned by the CMIP6 models.

For the component S_2 reflecting the influence of ODSs, the CMIP6 historical ensemble shows predominantly positive influences during summer (i.e. ODSs driving a summertime strengthening of the SAM), in agreement with NOAA-20CR. During winter, CMIP6 favours a negative influence (weakening) of ODSs on the SAM, but a substantial fraction of simulations also show a strengthening. NOAA-20CR is generally in agreement with this behaviour.S_2 exhibits a generally better agreement between NOAA-20CR and the CMIP6 models than S_1, likely because ODSs only started to increase around the time of the onset of measurements in Antarctica, and thus trends driven by ODS increases are better captured by observations than earlier, possibly spurious trends projecting onto GHG increases in NOAA-20CR which occurred before these observations started.

The sum of the regression coefficients S_1+S_2 shows less variability than the two coefficients individually. This is the result of non-zero correlation between the two indices. $S_1 + S_2$ is therefore more robustly diagnosed from both the references and the models than both indices individually. The much larger wintertime strengthening in NOAA-20CR, which is outside the range spanned by the CMIP6 ensemble, is evident here as well.

We conclude from this analysis that (a) during summer the CMIP6 simulations are broadly in agreement with NOAA-20CR regarding the influences of ODSs and GHGs and their combination, whereas (b) during winter the total best-estimate strengthening of the SAM evident in NOAA-20CR is irreconcilable with the CMIP6 ensemble. Given the better agreement between CMIP6 and NOAA-20CR, in all seasons, for the SAM strengthening in 1957-2014 (figure 5), we conjecture that this may well reflect spurious variations in the NOAA-20CR reanalysis for the period before 1957.

Next we perform a formal analysis of differences between the behaviours of chemistry- and no-chemistry models, with a particular emphasis on an uncertainty calculation.

5.3 Chemistry- versus no-chemistry models

In comparing the two model groups, we consider two sources of statistical uncertainty: The first is the random-noise uncertainty associated with the regression approach itself. To account for this, for every model (chemistry and no-chemistry alike), we produce 1000 synthetic variant realizations of each single-model ensemble-mean SAM index using a Monte-Carlo simulation approach outlined in more detail in the appendix. From these synthetic realizations, for each model 1000 synthetic regression coefficients $s_0^{i,j}(m)$, $s_1^{i,j}(m)$, and $s_2^{i,j}(m)$ are derived which now reflect the statistical uncertainties in the regression coefficients of the SAM for each model. Here i denotes the model, j the Monte-Carlo realization, and m as before the 12 overlapping seasons of the year.

We then form the six-model mean and statistical uncertainty range of this data from the six chemistry models. Here the uncertainty only refers to the statistical component reflecting natural variability, which due to the 6-model averaging is smaller than for individual models.

For the no-chemistry group, we note that there are 100,947 distinct 6-model subsets that can be formed from the 23 no-chemistry models. We thus form those 100,947 six-model averages of the s_0, s_1, and s_2 coefficients. Multiplied by the 1000 Monte-Carlo realizations each, this yields a total of 100,947,000 regression realizations. The distribution functions of s_0, s_1, and s_2 define the uncertainties in 6-model-mean regression coefficients derived from a randomly chosen 6-model subset of the no-chemistry models. This analysis is accounting for statistical/random noise as well as model-selection uncertainties. The presence of both these types of uncertainties in the no-chemistry group means that the no-chemistry 6-model-mean coefficients are subject to larger total uncertainties than those of the chemistry group (see below).

We then, for both ensembles, reduce the 1000 and 100 million realizations, respectively, to their means and 2.5, 16, 84, and 97.5 percentiles, representing the one- and two-standard deviation uncertainty bounds of the 6-model-mean regression parameters (figures 6 and 7). For the baseline SAM index S_0, there are relatively small differences between the two model sets although the large inter-model differences noted before produce a substantial spread in the uncertainty range for the no-chemistry ensemble which is absent in the chemistry group (as there is only one such ensemble possible, characterized by a very small random uncertainty). Except during austral spring, the SAM in the chemistry group tends to be slightly stronger than in the no-chemistry group with a probability exceeding 60% in February and September and a tendency for chemistry models to have a weaker S_0 in May-June than the no-chemistry models (figure 7).

The GHG influence S_1 is generally stronger in the no-chemistry group (with probabilities of 90% or larger during summer and autumn (DJFMAM)). During summer (DJF) there is a disagreement in sign, with no-chemistry models favouring a strengthening influence of GHGs but chemistry models favouring a weakening. The ODS influence S_2 is stronger throughout most of the year in the chemistry versus no-chemistry group, with particularly large differences occurring in summer and autumn when the probability exceeds 90%. In winter, chemistry models show a zero influence but no-chemistry models display a weakening influence of ODSs. Both are much weaker than the GHG influence and insignificant at the 68% confidence level.

As for the combination of both influences $S_1 + S_2$, no-chemistry models (with a probability of 60 to 90%, depending on season) favour a larger strengthening than the chemistry models, although the difference is always less than 1 hPa.

5.4 Deep coupling of the SAM

In the preceding sections we have only studied the surface expression of the SAM. Here we discuss briefly what the trends and the regression analysis yield when applied to the SAM index derived from geopotential height (zg) fields. In particular we will contrast the chemistry models with their near-equivalent no-chemistry variants (CESM2, CNRM-CM6-1, GFDL-CM4, GISS-E2-1-G, HadGEM3-GC31-LL; table 1).

The 1957-2014 linear trends in the SAM (figure 8) show in all models (both chemistry and equivalent no-chemistry models) a maximum in the strengthening of the SAM in late spring between 100 and 10 hPa. In the chemistry models, the amplitude of this maximum depends on the amount of ozone depletion simulated, with UKESM1-1-LL and MRI-ESM2-0 simulating the largest and smallest trends in zg for relatively large and small amounts of Southern-Hemisphere ozone loss, respectively (Morgenstern et al., 2020). The strengthening disappears when halocarbons are suppressed in the hist-1950HC experiment; they are replaced with insignificant trends of inconsistent sign. In four out of five cases where this comparison is possible, spring/summer trends in zg are larger in the chemistry models than in their no-chemistry equivalents. Only the CESM2 / CESM2-WACCM pair exhibits about the same strengthening. The hist-GHG experiment (where ozone is prescribed and GHGs provide the only forcing) in some cases (CESM2/CESM2-WACCM,

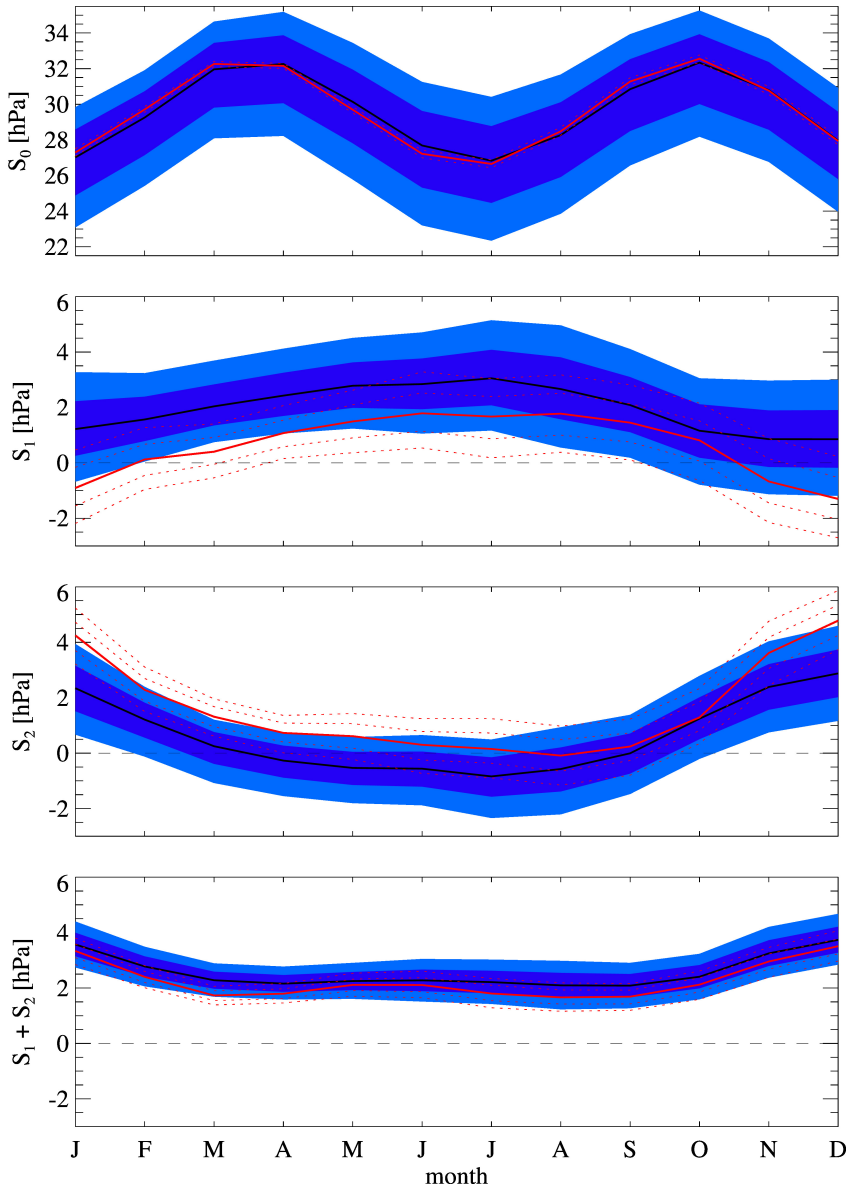

Figure 6. Thick black: Mean parameters S_0, S_1, S_2, and $S_1 + S_2$ from the no-chemistry model ensemble. Dark and light blue: Their 68 and 95% confidence intervals. Solid red: Mean parameters S_0, S_1, S_2, and $S_1 + S_2$ from the chemistry model ensemble. Dotted red: Their 68 and 95% confidence intervals.

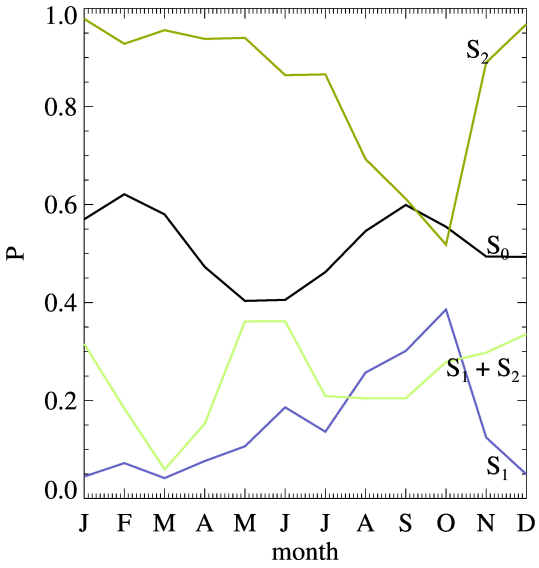

Figure 7. Probabilities P that the multi-model mean of (black) S_0 (blue) S_1 (olive) S_2 and (light green) $S_1 + S_2$ is larger for the chemistry group of models than for a randomly chosen 6-model no-chemistry group.

GFDL-CM4/GFDL-ESM4-1, GISS-E2-1-G/GISS-E2-1-Gchem) the trends are larger than in hist-1950HC (where ozone is interactive), but this is not the case for all models.

Extending now the regression analysis (covering 1850-2014) to zg, for five of the six chemistry models (except for MRI-ESM2-0) the regression analysis (figure 9) shows that ozone depletion is driving a strengthening of the SAM during spring and summer, but there is a sizeable offset due to increasing GHGs. In MRI-ESM2-0 a weak strengthening of the SAM associated with the small ozone depletion characterizing this model is not partially offset by a weakening influence of GHGs. There is substantial anticorrelation between the influences of both forcings also in other seasons. The no-chemistry equivalents largely show similar behaviour although in two of the models (GISS-E2-1-G, HadGEM3-GC31-LL) the offsetting effect maximizes earlier in the year and is not perfectly aligned with the ozone depletion season, unlike in most chemistry models. However, in all cases, deep coupling is evident whereby relatively large zg trends in the stratosphere, with a delay of a few months, drive corresponding zg trends in the troposphere in summer which then manifest as the trends in psl discussed above.

6 Discussion

Validations of multi-model climate experiments versus observations, and comparisons of different groups of models, require a careful consideration of statistical uncertainty bounds. This is particularly the case for the SAM which for the period before the onset of routine meteorological measurements in Antarctica is subject to substantial uncertainties. For this reason evaluations of trends in the SAM are often restricted to recent decades only, and often only involve calculating linear trends. In the face of large random variability and considerable differences in model behaviour, such approaches can result in substantial uncertainties in the resultant trends and consequently only weak conclusions about any drivers of such trends.

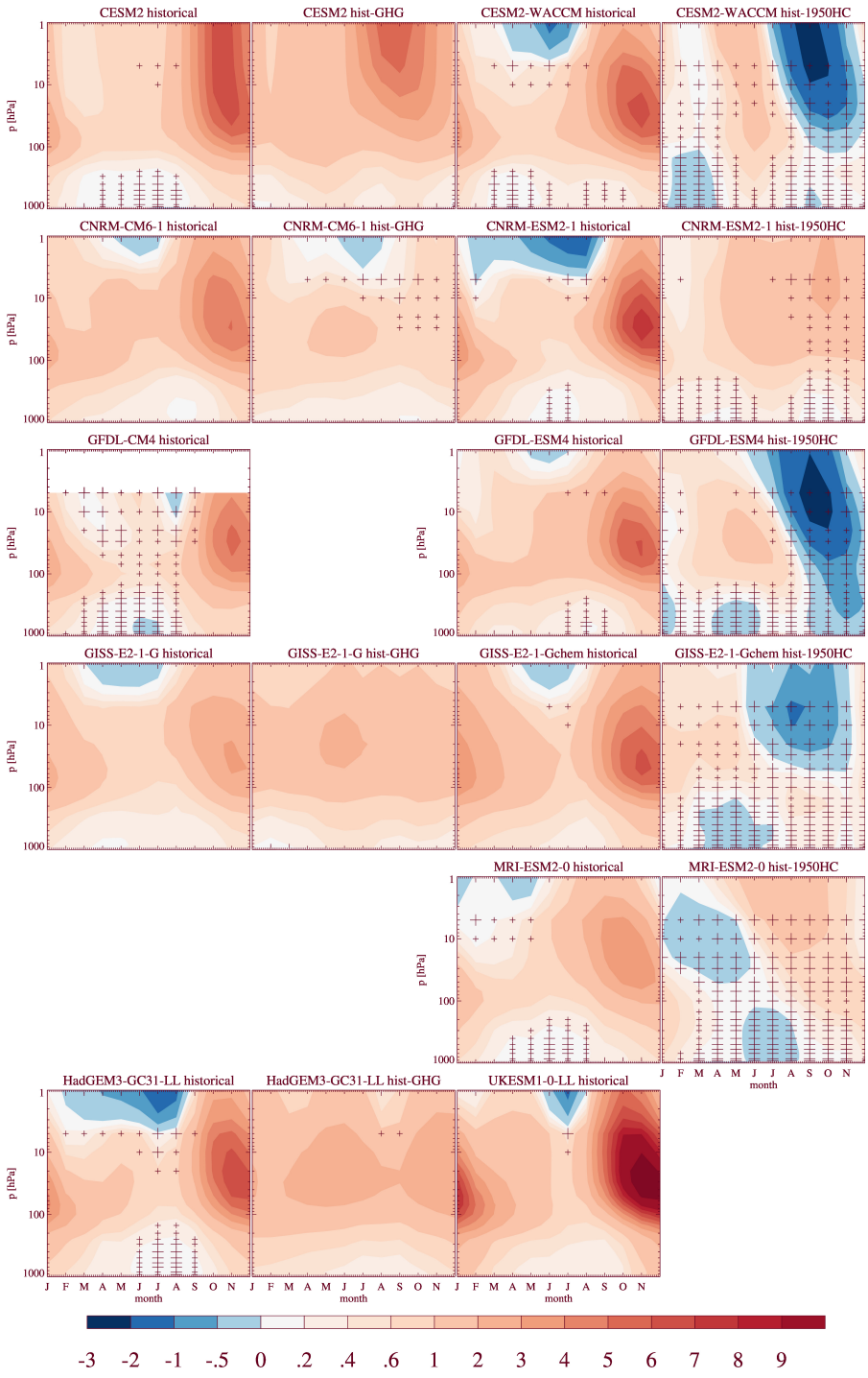

Figure 8. 1957-2014 trends in the ensemble-mean geopotential height fields (m a^{-}1) in the chemistry models and their no-chemistry equivalents (where available). '+' symbols denote insignificant trends (small '+': at the 68% confidence levels; large '+': 95% significance). (left) No-chemistry, historical. (2nd colum) No-chemistry, hist-GHG. (3rd column) Chemistry, historical. (right) Chemistry, hist-1950HC.

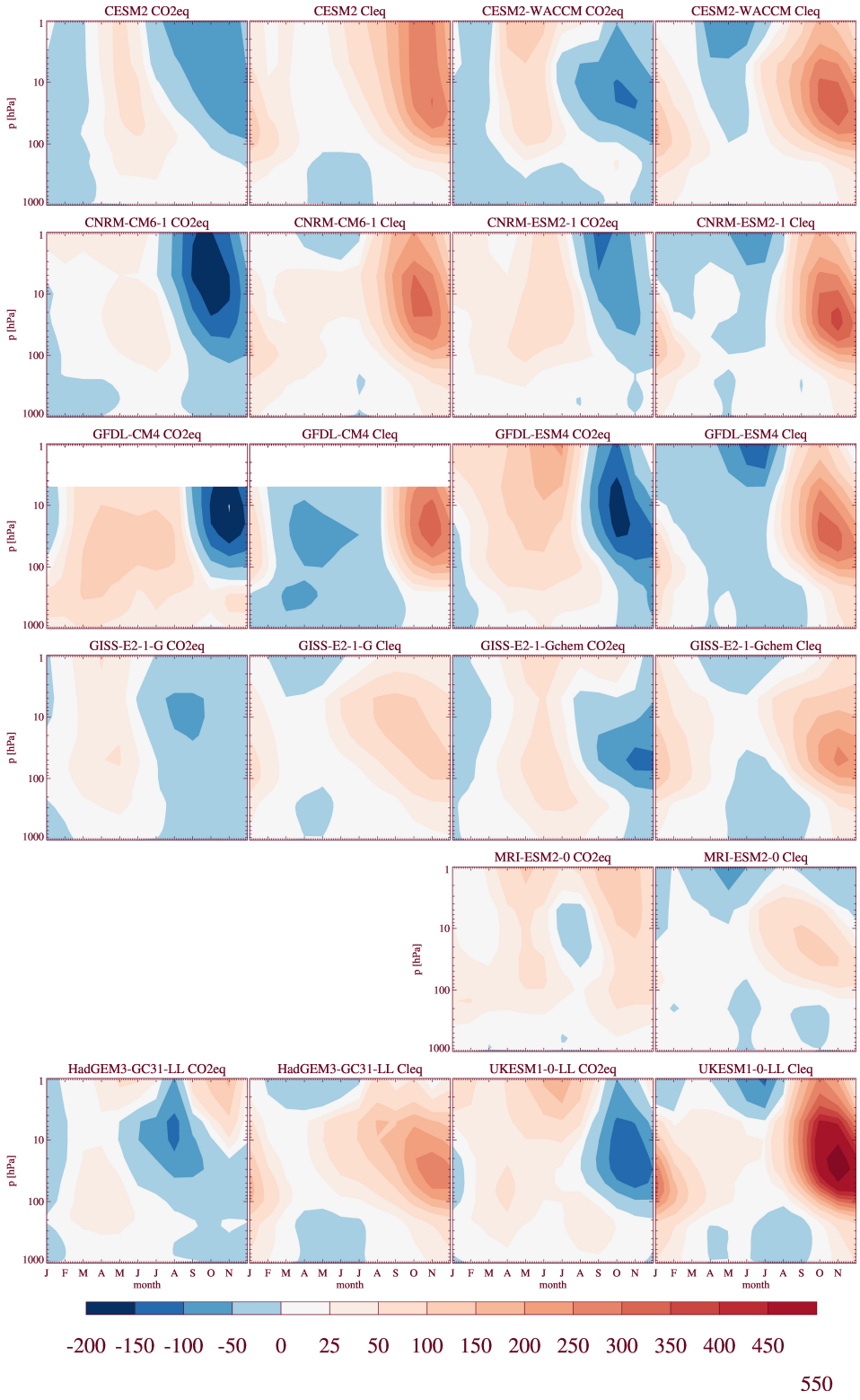

Figure 9. Regression coefficients S_1 and S_2 derived from the ensemble-mean geopotential height fields (m) in the chemistry models and their no-chemistry equivalents (where available). (left column) S_1, no-chemistry models. (2^{nd} column) S_2, no-chemistry. (3^{rd} column) S_1, chemistry. (right column) S_2, chemistry.

Table 3. Approximate MMM SAM strengthening (hPa) in the four experiments in 2014, discerned from figure 4, and factors present in those experiments. Magnitudes of the factors as derived from the experiments.

Experiment	factors	DJF	JJA
historical	ODS + dGHG + iGHG	3.7	1.7
hist-1950HC	dGHG + iGHG	0	1.7
hist-GHG	dGHG	2.6	2
hist-stratO3	ODS + iGHG	2.5	−0.5
	ODS	3.7 to 5.1	0 to 0.2
	dGHG	1.2 to 2.6	2 to 2.2
	iGHG	−2.6 to −1.2	−0.3 to −0.5

To advance in the face of these issues, we here pursue an approach that (a) maximizes the usage of available simulations (i.e., we use almost all available CMIP6 historical simulations). (b) Using a regression model, we consider the whole "historical" period (1850-2014). The regression model accounts for the leading anthropogenic influences that modulate the SAM on decadal-to-century timescales. (c) In comparing the mean behaviour of two different groups of models (with and without interactive ozone) we form all possible subsets of the larger group that are of equal size to the smaller group. This ensures strict comparability.

For the period 1957-2014, during which continuous Antarctic meteorological observations exist, we find a mean strengthening trend in the CMIP6 ensemble which in summer is consistent with four observational datasets. In winter however two of these datasets (HadSLP2, NCEP/NCAR) exhibit spuriously large trends and are thus removed from further analysis. The remaining two datasets (Marshall (2003) and NOAA-20CR) are consistent with each other and with the distribution of trends in the CMIP6 ensemble. This makes NOAA-20CR our primary observational reference dataset as it extends back to 1851.

Forced multidecadal SAM variations in the four experiments analyzed here cannot be reconciled using the traditional approach of only accounting for ODS and GHG influences. The hist-1950HC experiment clearly shows that in the absence of ozone depletion, the models in the mean exhibit no strengthening in DJF. In this two-factor framework this would be in contradiction with the hist-GHG experiment which does produce a significant MMM growth of the SAM index. Following Morgenstern et al. (2014), we therefore stipulate that a three-factor approach is needed to explain this, comprising the factors ozone depletion (ODS), greenhouse-gas induced warming (the "direct GHG" effect, dGHG), and the impact on dynamics of greenhouse-gas induced ozone changes (the "indirect GHG" effect, iGHG). These factors are variously taking effect in the four experiments considered here (table 3).

Table 3 represents an overdetermined system of four linear equations for the three factors ODS, dGHG, and iGHG. There are four three-equation subsets, one of which has no solution. The other three yield the ranges as indicated in the lower three lines of table 3.

The three-factor qualitative model better explains the findings than the traditional two-factor approach. In summer ozone depletion dominates the other influences in the CMIP6 ensemble. However, the GHG influence is negligible because of a cancellation of two non-negligible terms, dGHG and iGHG, which each are about a third to half the size of the ODS influence. In winter, the ODS influence is small, and there is some offset of

dGHG by iGHG. These findings are corroborated by our upper-level analysis (figure 8) which indicates for the hist-1950HC experiment essentially zero trends in the SAM index during the ozone-depletion season but persistent positive trends in the index for the hist-GHG experiment. The results corroborate Morgenstern et al. (2014) however now using a multi-model analysis.

The analysis presented here fundamentally relies on the presence of a subset of models with interactive ozone chemistry having completed the hist-1950HC experiment. In the hypothetical absence of this experiment, the "contradiction" mentioned above would be reduced to some nonlinearity between the historical, hist-GHG and hist-stratO3 experiments (i.e. the summertime strengthening in hist-GHG and hist-stratO3 not adding up to that in the historical experiment; table 3), which however would not be clear indication of a missing factor in the attribution analysis.

We also find some statistically robust differences in behaviour between the 6-model chemistry and the 23-model no-chemistry ensembles. Corroborating previous literature on the differences between chemistry- and no-chemistry models (Haase & Matthes, 2019; Haase et al., 2020), we find stronger influences of ozone depletion offset by weaker influences of increasing GHGs in the chemistry group, in most seasons. A comparison of the chemistry models with their no-chemistry equivalents suggests that indeed the SAM strengthening in the stratosphere is mostly weaker in the no-chemistry counterparts; this behaviour has previously been attributed to misalignment of the ozone hole with the dynamical polar vortex in no-chemistry models, causing a systematically weaker polar vortex and a weaker influence of ozone depletion also on the tropospheric SAM (Haase & Matthes, 2019).

In summary, our study illustrates that using only no-chemistry models for attribution of trends in the SAM carries the risk of erroneous conclusions. CMIP6 marks the first time that a sizeable set of fully coupled climate models, with and without ozone chemistry, has been applied to a range of sensitivity scenarios required to identify and quantify the factors that drive the SAM. The results of our analysis align well but also advance on previous analyses of the SAM.

Appendix A The Monte-Carlo based comparison of chemistry and no-chemistry models

The method used to compute the data in figure 6 is laid in more detail here.

- For all models i, we form the single-model ensemble-means of the SAM index.
- The ensemble-mean SAM indices are decomposed into their regression components S_0^i to S_2^i, and the variance ϵ^2 is decomposed into its components E_0^i to E_2^i following equations 1 and 2.
- Based on these regression functions, for every model 1000 synthetic random realizations $z^{ij}(m, y)$ are produced. (j stands for one of the 1000 random realizations, and m and y are the month and year as before.) This process is illustrated in figure A1 for the HadGEM3-GC31-LL model. Note that the spread in figure A1 reduced with increasing ensemble size.
- The random realizations $z^{ij}(m, y)$ are again expanded according to equation 1 to derive regressions coefficients $s_0^{ij}(m)$, $s_1^{ij}(m)$, and s_2^{ij}.
- We now group the models into 6 chemistry and 23 no-chemistry models. For the chemistry group, we form the 6-model means of s_0, s_1, and s_2, resulting for every calendar month in 1000 random realizations for those 6-model means. From these, we derive their mean across the 1000 realizations and their 2.5, 16, 84, and 97.5 percentiles, marking the 68 and 95% confidence intervals as displayed in figure 6.

- For the no-chemistry subset of models, we form all $23!/(6! \cdot 17!) = 100,947$ six-model subsets. For every one of the subsets we follow the same process as above for the chemistry models, resulting in 100,947,000 six-model mean realizations of the regression coefficients for the no-chemistry group. We again reduce these to their means and 2.5, 16, 84, and 97.5 percentiles as above.

For the Monte-Carlo analysis of SAM indices to be valid, we require an autocorrelation length of less than 1 year for the residual ϵ. This would ensure that the uncorrelated random noise produced by the random-number generator is a good reflection of what is produced by the models. We have checked that for all individual models this is indeed the case. We furthermore require ϵ to be normally distributed. Following Morgenstern et al. (2014), figure A2 indicates that the assumption of a normal distribution of the residual ϵ is good but not perfect. Relative to a perfect Gaussian distribution, large deviations from the mean in both directions (where $G < 0.15$ or $G > 0.85$) are slightly overrepresented and small deviations (where $0.2 < G < 0.8$) are underrepresented, across practically all models. Morgenstern et al. (2014) found similar deviations from a normal distribution in their analysis.

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. Model and observational data used in this study can be downloaded at `https://esgf-node.llnl.gov/search/cmip6/`, `https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5`, `https://psl.noaa.gov/data/gridded/data.ncep.reanalysis.html`, `https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/cera-20c`, `https://psl.noaa.gov/data/gridded/data.20thC_ReanV2c.html`, `https://www.metoffice.gov.uk/hadobs/hadslp2/data/download.html`, and `https://legacy.bas.ac.uk/met/gjma/sam.html`. The author acknowledges ECMWF, NOAA, the UK MetOffice, and the British Antarctic Survey for provision of these data. The author was supported by the NZ Government's Strategic Science Investment Fund (SSIF) through the NIWA programme CACV. He acknowledges the contribution of NeSI high-performance computing facilities to the results of this research. New Zealand's national facilities are provided by the New Zealand eScience Infrastructure (NeSI) and funded jointly by NeSI's collaborator institutions and through MBIE's Research Infrastructure programme. Scripts and data need to reproduce the analyses and figures in this paper can be found at `https://doi.org/10.5281/zenodo.4145770`.

References

Allan, R., & Ansell, T. (2006). A new globally complete monthly historical gridded mean sea level pressure dataset (HadSLP2): 1850–2004. *Journal of Climate*, *19*, 5816–5842. Retrieved from `https://doi.org/10.1175/JCLI3937.1`

Arblaster, J. M., & Meehl, G. A. (2006). Contributions of external forcings to Southern Annular Mode trends. *Journal of Climate*, *19*(12), 2896-2905. Retrieved from `https://doi.org/10.1175/JCLI3774.1`

Bao, Y., Song, Z., & Qiao, F. (2020). FIO-ESM version 2.0: Model description and evaluation. *Journal of Geophysical Research: Oceans*, *125*, e2019JC016036. Retrieved from `https://doi.org/10.1029/2019JC016036`

Bethke, I., Wang, Y., Counillon, F., Kimmritz, M., Fransner, F., Samuelsen, A., ... Keenlyside, N. (2019). *NCC NorCPM1 model output prepared for CMIP6 CMIP historical.* Earth System Grid Federation. Retrieved from `https://doi.org/10.22033/ESGF/CMIP6.10894` doi: 10.22033/ESGF/CMIP6.10894

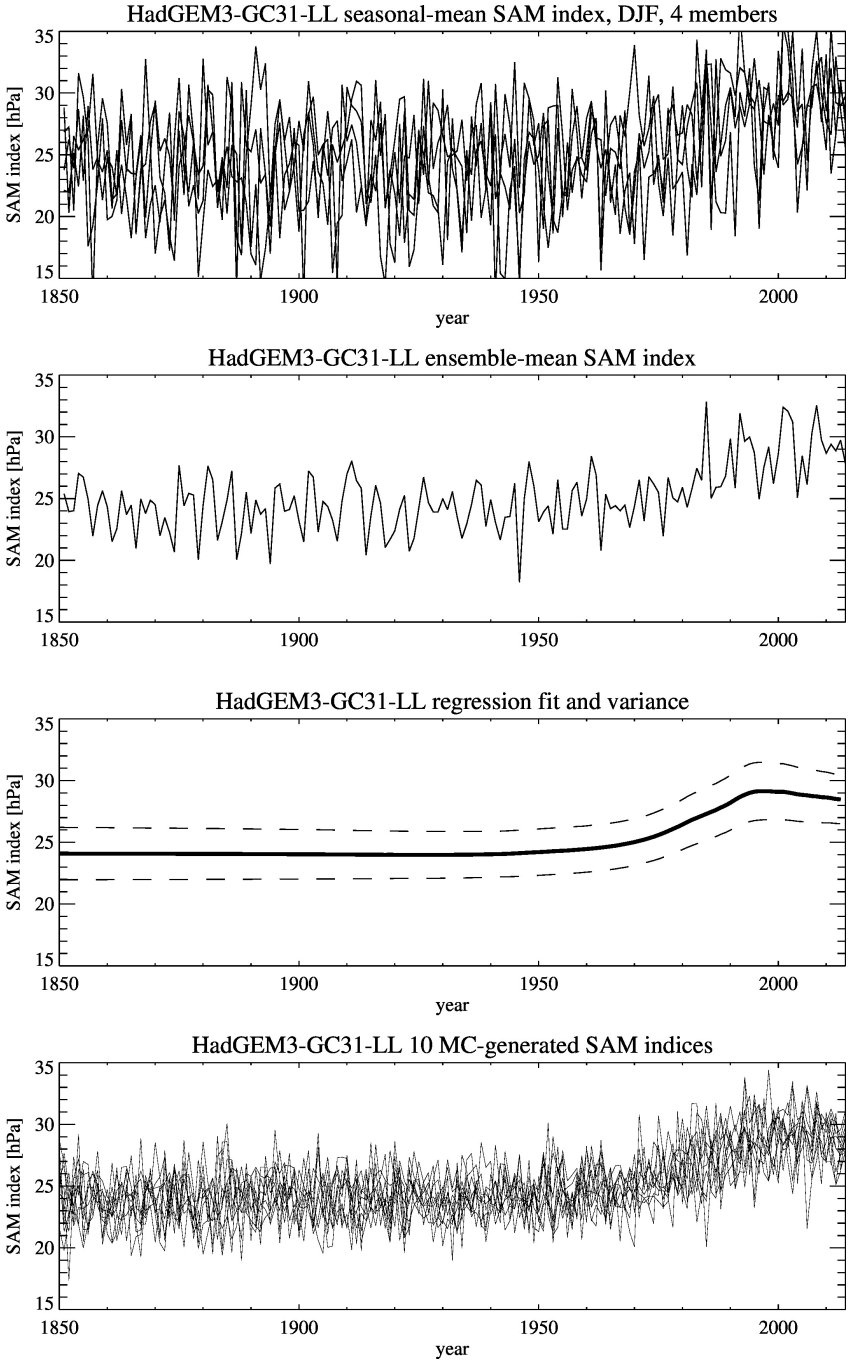

Figure A1. Example of Monte-Carlo based analysis. (top) Seasonal-mean SAM indices in DJF for the four "historical" simulations of HadGEM3-GC31-LL (hPa). (2nd panel) Ensemble mean. Note the reduced variance. (3rd panel) Regression expansion (solid) with one-standard deviation variance range (dashed). (bottom) Monte-Carlo based perturbations z^{ij} around the regression expansion. Only the first 10 realizations are shown here.

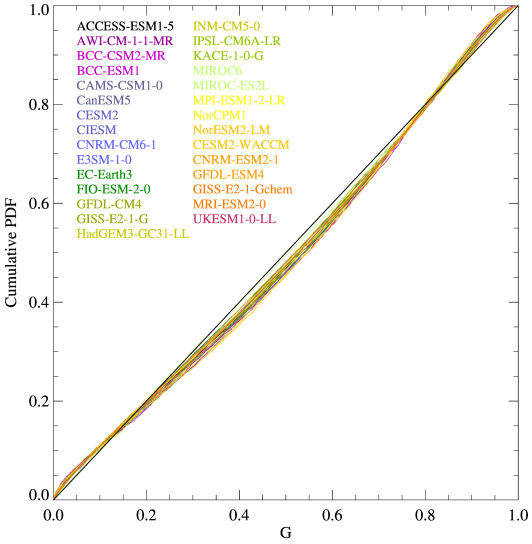

Figure A2. Cumulative probability distribution function (PDF) of the normalized regression residual $\epsilon/\sqrt{E_0 + E_1 \mathrm{CO}_2^{eq} + E_2 \mathrm{Cl}^{eq}}$ against a perfect Gaussian integral G. The PDF is evaluated individually for every model but across all seasons and across all historical simulations.

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., ... Vuichard, N. (2020). Presentation and evaluation of the IPSL-CM6A-LR climate model. *Journal of Advances in Modeling Earth Systems*, *12*(7), e2019MS002010. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS002010` (e2019MS002010 10.1029/2019MS002010) doi: 10.1029/2019MS002010

Chen, X. L., Guo, Z., Zhou, T. J., & et al. (2019). Climate sensitivity and feedbacks of a new coupled model CAMS-CSM to idealized CO_2 forcing: A comparison with CMIP5 models. *Journal of Meteorological Research*, *33*, 31-45. doi: 10.1007/s13351-019-8074-5

Compo, G. P., Whitaker, P. D., J. S. Sardeshmukh, Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., ... Worley, S. J. (2011). The Twentieth Century Reanalysis Project. *Quarterly Journal of the Royal Meteorological Society*, *137*, 1-28. doi: 10.1002/qj.776

Danabasoglu, G., Lamarque, J., 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*, e2019MS001916. Retrieved from `https://doi.org/10.1029/2019MS001916`

Dennison, F. W., McDonald, A. J., & Morgenstern, O. (2015). The effect of ozone depletion on the Southern Annular Mode and stratosphere-troposphere coupling. *Journal of Geophysical Research: Atmospheres*, *120*(13), 6305-6312. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014JD023009` doi: 10.1002/2014JD023009

Dunne, J. P., Horowitz, L. W., Adcroft, A. J., Ginoux, P., Held, I. M., John, J. G., ... Zhao, M. (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*, e2019MS002015. Retrieved from `https://doi.org/10.1029/2019MS002015`

Eyring, V., Arblaster, J. M., Cionni, I., Sedláček, J., Perlwitz, J., Young, P. J., ... Watanabe, S. (2013). Long-term ozone changes and associated climate impacts in CMIP5 simulations. *Journal of Geophysical Research: Atmospheres*,

118(10), 5029-5060. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/jgrd.50316 doi: 10.1002/jgrd.50316

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*(5), 1937–1958. Retrieved from https://www.geosci-model-dev.net/9/1937/2016/ doi: 10.5194/gmd-9-1937-2016

Gettelman, A., Mills, M. J., Kinnison, D. E., Garcia, R. R., Smith, A. K., Marsh, D. R., ... Randel, W. J. (2019). The Whole Atmosphere Community Climate Model version 6 (WACCM6). *Journal of Geophysical Research: Atmospheres*, *124*(23), 12380-12403. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019JD030943 doi: 10.1029/2019JD030943

Gillett, N. P., & Fyfe, J. C. (2013). Annular mode changes in the CMIP5 simulations. *Geophysical Research Letters*, *40*(6), 1189-1193. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/grl.50249 doi: 10.1002/grl.50249

Golaz, J., Caldwell, P. M., Van Roekel, L. P., Petersen, M. R., Tang, Q., Wolfe, J. D., & et al. (2019). The DOE E3SM coupled model version 1: Overview and evaluation at standard resolution. *Journal of Advances in Modeling Earth Systems*, *11*, 2089–2129. Retrieved from https://doi.org/10.1029/2018MS001603

Gong, D., & Wang, S. (1999). Definition of Antarctic Oscillation index. *Geophysical Research Letters*, *26*(4), 459-462. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/1999GL900003 doi: 10.1029/1999GL900003

Haase, S., Fricke, J., Kruschke, T., Wahl, S., & Matthes, K. (2020). Sensitivity of the Southern Hemisphere tropospheric jet response to Antarctic ozone depletion: prescribed versus interactive chemistry. *Atmospheric Chemistry and Physics Discussions*, *2020*, 1–33. Retrieved from https://acp.copernicus.org/preprints/acp-2020-441/ doi: 10.5194/acp-2020-441

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

Hajima, T., Watanabe, M., Yamamoto, A., Tatebe, H., Noguchi, M. A., Abe, M., ... Kawamiya, M. (2020). Development of the MIROC-ES2L Earth system model and the evaluation of biogeochemical processes and feedbacks. *Geoscientific Model Development*, *13*, 2197–2244. Retrieved from https://doi.org/10.5194/gmd-13-2197-2020

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. Retrieved from https://doi.org/10.1029/2019MS001829

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., ... Joseph, D. (1996, 03). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, *77*(3), 437-472. Retrieved from https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2 doi: 10.1175/1520-0477(1996)077⟨0437:TNYRP⟩2.0.CO;2

Kang, S. M., Polvani, L. M., Fyfe, J. C., & Sigmond, M. (2011). Impact of polar ozone depletion on subtropical precipitation. *Science*, *332*, (6032):951-4. doi: 10.1126/science.1202131

Kelley, M., Schmidt, G. A., Nazarenko, L., Bauer, S. E., Ruedy, R., Russell, G. L., ... Yao, M. (2020). GISS-E2.1: Configurations and climatology. *Journal of Advances in Modelling Earth Systems*, *12*, e2019MS002025. Retrieved from

https://doi.org/10.1029/2019MS002025

Laloyaux, P., de Boisseson, E., Balmaseda, M., Bidlot, J.-R., Brönnimann, S., Buizza, R., ... Schepers, D. (2018). CERA-20C: A coupled reanalysis of the twentieth century. *Journal of Advances in Modeling Earth Systems*, *10*(5), 1172-1195. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018MS001273 doi: 10.1029/2018MS001273

Lee, J., Kim, J., Sun, M., Kim, B.-H., Moon, H., Sung, H. M., ... Byun, Y.-H. (2020). Evaluation of the Korea Meteorological Administration Advanced Community Earth-System model (K-ACE). *Asia-Pacific Journal of Atmospheric Science*, *56*, 381–395. Retrieved from https://doi.org/10.1007/s13143-019-00144-7

Lin, Y.-L., Huang, X.-M., Liang, Y.-S., Qin, Y., Xu, S.-M., Huang, W.-Y., ... Gong, P. (2019). The Community Integrated Earth System Model (CIESM) from Tsinghua University and its plan for CMIP6 experiments. *Climate Change Research*, *15*, 545-550. doi: 10.12006/j.issn.1673-1719.2019.166

Marshall, G. J. (2003, 12). Trends in the Southern Annular Mode from observations and reanalyses. *Journal of Climate*, *16*(24), 4134-4143. Retrieved from https://doi.org/10.1175/1520-0442(2003)016<4134:TITSAM>2.0.CO;2 doi: 10.1175/1520-0442(2003)016⟨4134:TITSAM⟩2.0.CO;2

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

Morgenstern, O., O'Connor, F. M., Johnson, B. T., Zeng, G., Mulcahy, J. P., Williams, J., ... Kinnison, D. E. (2020). Reappraisal of the climate impacts of ozone-depleting substances. *Geophysical Research Letters*, *47*(20), e2020GL088295. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020GL088295 (e2020GL088295 10.1029/2020GL088295) doi: 10.1029/2020GL088295

Morgenstern, O., Stone, K. A., Schofield, R., Akiyoshi, H., Yamashita, Y., Kinnison, D. E., ... Chipperfield, M. P. (2018). Ozone sensitivity to varying greenhouse gases and ozone-depleting substances in CCMI-1 simulations. *Atmospheric Chemistry and Physics*, *18*, 1091–1114. Retrieved from https://doi.org/10.5194/acp-18-1091-2018

Morgenstern, O., Zeng, G., Dean, S. M., Joshi, M., Abraham, N. L., & Osprey, A. (2014). Direct and ozone-mediated forcing of the Southern Annular Mode by greenhouse gases. *Geophysical Research Letters*, *41*(24), 9050-9057. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014GL062140 doi: 10.1002/2014GL062140

Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., ... Zhang, H. (2013). Anthropogenic and Natural Radiative Forcing. In *Climate Change 2013 - The Physical Science Basis* (chap. 8). Geneva, Switzerland: Intergovernmental Panel on Climate Change (IPCC). Retrieved from https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5_Chapter08_FINAL.pdf

Müller, W. A., Jungclaus, J. H., Mauritsen, T., Baehr, J., Bittner, M., Budich, R., ... Marotzke, J. (2018). A higher-resolution version of the Max Planck Institute Earth System Model (MPI-ESM1.2-HR). *Journal of Advances in Modeling Earth Systems*, *10*(7), 1383-1413. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2017MS001217 doi: 10.1029/2017MS001217

Newman, P. A., Daniel, J. S., Waugh, D. W., & Nash, E. R. (2007). A new formulation of equivalent effective stratospheric chlorine (EESC). *Atmospheric Chemistry and Physics*, *7*(17), 4537–4552. Retrieved from https://acp.copernicus.org/articles/7/4537/2007/ doi: 10.5194/acp-7-4537-2007

Seland, O., Bentsen, M., Seland Graff, L., Olivié, D., Toniazzo, T., Gjermundsen, A., ... Schulz, M. (2020). The Norwegian Earth System Model, NorESM2 – Evaluation of the CMIP6 DECK and historical simulations. *Geoscientific Model Development Discussions*, in review. Retrieved from https://doi.org/10.5194/gmd-2019-378

Sellar, A. A., Jones, C. G., Mulcahy, J. P., Tang, Y., Yool, A., Wiltshire, A., ... Zerroukat, M. (2019). UKESM1: Description and evaluation of the U.K. Earth System Model. *Journal of Advances in Modeling Earth Systems*, *11*(12), 4513-4558. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001739 doi: 10.1029/2019MS001739

Semmler, T., Danilov, S., Gierz, P., Goessling, H. F., Hegewald, J., Hinrichs, C., ... Jung, T. (2020). Simulations for CMIP6 with the AWI climate model AWI-CM-1-1. *Journal of Advances in Modeling Earth Systems*, *12*(9), e2019MS002009. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS002009 (e2019MS002009 2019MS002009) doi: 10.1029/2019MS002009

Simmons, A., Soci, C., Nicolas, J., Bell, B., Berrisford, P., Dragani, R., ... Schepers, D. (2020). Global stratospheric temperature bias and other stratospheric aspects of ERA5 and ERA5.1. *ECMWF Technical Memo*, *859*. Retrieved from https://www.ecmwf.int/sites/default/files/elibrary/2020/19362-global-stratospheric-temperature-bias-and-other-stratospheric-aspects-era5-and-era51.pdf

Son, S.-W., Gerber, E. P., Perlwitz, J., Polvani, L. M., Gillett, N. P., Seo, K.-H., ... Yamashita, Y. (2010). Impact of stratospheric ozone on Southern Hemisphere circulation change: A multimodel assessment. *Journal of Geophysical Research: Atmospheres*, *115*(D3). Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010JD014271 doi: 10.1029/2010JD014271

Son, S.-W., Han, B.-R., Garfinkel, C. I., Kim, S.-Y., Park, R., Abraham, N. L., ... Zeng, G. (2018). Tropospheric jet response to Antarctic ozone depletion: An update with Chemistry-Climate Model Initiative (CCMI) models. *Environmental Research Letters*, *13*(5), 054024. Retrieved from https://doi.org/10.1088%2F1748-9326%2Faabf21 doi: 10.1088/1748-9326/aabf21

Son, S.-W., Tandon, N. F., Polvani, L. M., & Waugh, D. W. (2009). Ozone hole and Southern Hemisphere climate change. *Geophysical Research Letters*, *36*(15). Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2009GL038671 doi: 10.1029/2009GL038671

Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., ... Winter, B. (2019). The Canadian Earth System Model version 5 (CanESM5.0.3). *Geoscientific Model Development*, *12*(11), 4823–4873. Retrieved from https://gmd.copernicus.org/articles/12/4823/2019/ doi: 10.5194/gmd-12-4823-2019

Séférian, R., Nabat, P., Michou, M., Saint-Martin, D., Voldoire, A., Colin, J., ... Madec, G. (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. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001791 doi: 10.1029/2019MS001791

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

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*, 741–749. Retrieved from https://doi.org/10.1038/ngeo1296

Voldoire, A., Saint-Martin, D., Sénési, S., D., B., Alias, A., & Chevallier, e. a., M. (2019). Evaluation of CMIP6 DECK experiments with CNRM-CM6-1. *Journal of Advances in Modeling Earth Systems*, *11*, 2177–2213. Retrieved from https://doi.org/10.1029/2019MS001683

Volodin, E., & Gritsun, A. (2018). Simulation of observed climate changes in 1850–2014 with climate model INM-CM5. *Earth System Dynamics*, *9*, 1235–1242. Retrieved from https://doi.org/10.5194/esd-9-1235-2018

Williams, K. D., Copsey, D., Blockley, E. W., Bodas-Salcedo, A., Calvert, D., Comer, R., & et al.. (2017). 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*, 357–380. Retrieved from https://doi.org/10.1002/2017MS001115

Wu, T., Lu, Y., Fang, Y., Xin, X., Li, L., Li, W., ... Liu, X. (2019). The Beijing Climate Center Climate System Model (BCC-CSM): the main progress from CMIP5 to CMIP6. *Geoscientific Model Development*, *12*, 1573–1600. Retrieved from https://doi.org/10.5194/gmd-12-1573-2019

Wu, T., Zhang, F., Zhang, J., Jie, W., Zhang, Y., Wu, F., ... Hu, A. (2020). Beijing Climate Center Earth System Model version 1 (BCC-ESM1): Model description and evaluation of aerosol simulations. *Geoscientific Model Development*, *13*, 977–1005. Retrieved from https://doi.org/10.5194/gmd-13-977-2020

Wyser, K., & et al. (2020). Warmer climate projections in EC-Earth3-Veg: the role of changes in the greenhouse gas concentrations from CMIP5 to CMIP6. *Environmental Research Letters*, *15*, 054020.

Yukimoto, S., Kawai, H., Koshiro, T., Oshima, N., Yoshida, K., Urakawa, S., ... ISHII, M. (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 Ser. II*. doi: 10.2151/jmsj.2019-051

Ziehn, T., Chamberlain, M. A., Law, R. M., Lenton, A., Bodman, R. W., Dix, M., ... Srbinovsky, J. (2020). The Australian Earth System Model: ACCESS-ESM1.5. *Journal of Southern Hemisphere Earth Systems Science*. Retrieved from https://doi.org/10.1071/ES19035