Reply to reviews

Nora Bergner, Marina Friedel, Daniela I.V. Domeisen, Darryn Waugh, Gabriel Chiodo

August 5, 2022

RC = Reviewer Comment
AR = Author Reply

Reviewer 1

RC 1.1 The paper aims to explore the robustness of the tropospheric response to major stratospheric 'extreme' events in a reanalysis and two chemistry-climate models 'on a wide range of timescales'. The topic is important and outcomes of such an analysis are potentially valuable for the communities in the Southern Hemisphere. The paper is well structured and clearly written; conclusions are supported by the analysis.

However, I think that the aim of the analysis has not been achieved and, perhaps, could not have been achieved using the selected models. The reason must be that models don't capture the intensity of polar vortex anomalies and tropospheric SAM impacts well enough to asses robustness of the stratospheric-stratospheric coupling and surface impacts. Hence, I think that the title and parts of introduction need to be re-written to be consistent with the skills of the models.

AR 1.1 Thank you for this comment. We would like to point out that the bulk of the polar vortex anomalies is represented by the two models and that the magnitude of polar vortex anomalies come reasonably close to the 2019 event, although no event in the magnitude of 2002 occurs in the models (see Fig. 1 in the manuscript). Assessing robustness cannot be achieved using observations alone, and models are the best, albeit imperfect, tools that are currently available. It is also one point of the paper to document the model shortcomings and biases in the stratosphere-troposphere coupling and comparing two different models allows us to draw conclusions despite their biases. Nevertheless, we would agree that using the term 'robustness' could be misinterpreted and we would therefore suggest to adapt the title to: *Exploring the linkage between austral stratospheric polar vortex anomalies and surface climate* and adapt our introduction to: *In this study, we aim to explore linkages between polar vortex anomalies and their surface response in the*

SH in reanalysis data and chemistry-climate models (CCMs). In particular, we investigate the sensitivity of stratosphere-troposphere coupling and linkages to surface climate to the short observational record, internal variability, and model biases.

RC 1.2 Furthermore, I am wondering why all simulations are forced with the same GHGs, ODS, boundary conditions of the year 2000. Byrne et al. (2019) have shown that ENSO affects the eddy-driven jet via the stratosphere, with no evidence of a direct tropospheric pathway. Therefore, how results of the experiments are sensitive to the use of different initial conditions? Moreover, MERRA composites include various ENSO (and other large-scale drivers) phases, hence, how fair is to compare model and reanalysis composites?

AR 1.2 The primary objective of this study is to investigate the impact of interannual variability of the polar vortex on surface climate. Simulations forced with constant boundary conditions eliminate the effects of climate change and ozone trends due to long-term changes in ODSs, aiding the purpose of this paper. When comparing our time-slice simulations with historical simulations (1980-2020) of the chemistry climate model intercomparison (CCMI-1) (Morgenstern et al., 2017) using CESM1 WACCM and SOCOL, the stratospheric SAM anomalies are reasonably similar to reanalysis. In fact, the time-slice simulations (referred to as "const. BCs" in Figure 1) get closer to the more extreme events, such as 2019 (Figure 1). For the tropospheric SAM signal in Figure 2, the simulations show slightly different results: the CCMI-1 historical simulations of SOCOL show a weaker signal following weak polar vortex anomalies than our time-slice simulations, whereas the opposite is the case for WACCM. The differences between our runs and CCMI simulations are most likely due to the different set-ups: CCMI-1 runs are much shorter (40 years vs our own 200 years time-slice experiments), they include ozone and climate-change trends, and in the case of the CCMI-1 runs from SOCOL, SSTs are prescribed.

We wish to clarify that the time-slice simulations of both models that we are using are coupled to an ocean model and therefore include various ENSO states. Hence, these experiments also capture any ENSO effects on the tropospheric and stratospheric SAM, and are directly comparable to MERRA-2. Based on the Nino 3.4 index which we show in Figure 3, we see that the stratospheric extremes occur during all ENSO phases, which renders models and reanalysis comparable. There is a weak negative correlation between ENSO and the stratospheric SAM in reanalysis and one model (SOCOL) captures it (Spearman correlation coefficients based on monthly averaged indexes of Nino 3.4 and 10 hPa SAM: MERRA-2: -0.11, p=0.01; SOCOL: -0.16, p=3.18×10^{-15}; WACCM: -0.01, p=0.44). While it would be interesting to further investigate the impact of ENSO teleconnections and other initial conditions on the stratosphere, it is outside the aim and the scope of this study.

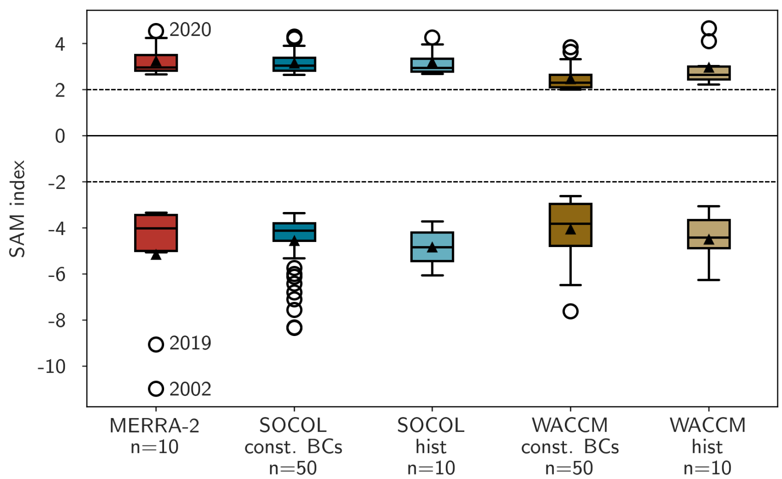

Figure 1: The 10 hPa SAM index values of the 25 % lowest and highest springtime SAM indices in MERRA-2 (10 events per weak/strong category), and in CCMs WACCM and SOCOL for simulations using constant boundary conditions (50 events per weak/strong category), and the reference "ref-C2" simulations from CCMI-1 (Morgenstern et al., 2017), which are referred to as "hist" (solely focusing on 1980-2020 for comparability with MERRA-2 reanalysis, i.e. 10 events per weak/strong category). The most extreme events in the reanalysis data are annotated with the year of occurrence.

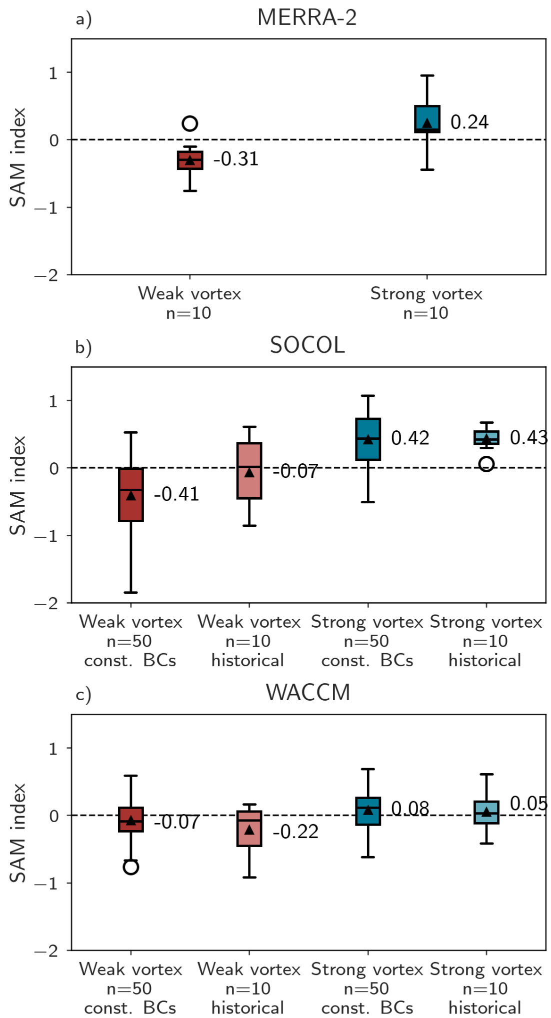

Figure 2: Distribution of the mean 500 hPa SAM index averaged over the October-January period following weak and strong polar vortex anomalies for MERRA-2 (a), and CCMs SOCOL (b), and WACCM (c) with each constant BC simulations and historical simulations (1980-2020). The box extends from the lower to upper quartile of the data, the whiskers extend from the lower quartile −1.5 IQR to the upper quartile +1.5 IQR. Data points outside of the whiskers are shown as circles. The horizontal line marks the median value and the triangle indicates the mean of the distribution, which is annotated next to the box.

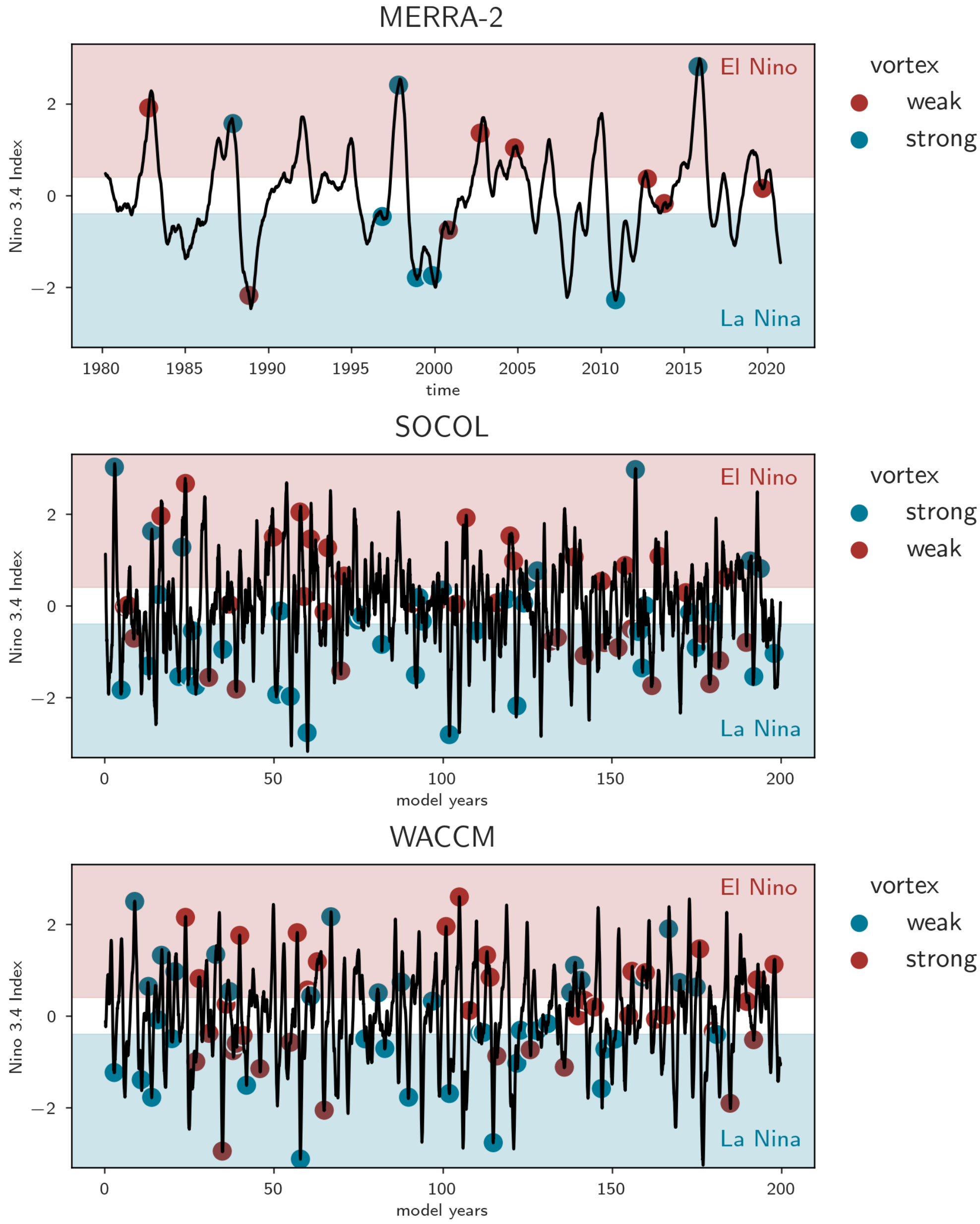

Figure 3: The figure shows the evolution of the Nino3.4 index, estimated using the 5-month average of surface temperatures in the Nino3.4 region. The detected polar vortex anomalies are marked with red circles for weak events, and blue circles for strong events.

RC 1.3 Therefore, while the paper is well written, (1) more work is needed to bring the goals of the research in line with actual skills of SOCOL and WACCM; (2) other experiments may be needed that produce polar vortex anomalies of magnitudes that are similar to the observed values (particularly, for WACCM) and that account for other boundary and initial conditions.

AR 1.3 We are in favour of the referee's suggestions, but are of the opinion that the suggested additional experiments are outside of the aim and the scope of this study. Additionally, experiments that produce polar vortex anomalies of similar magnitudes to the observed values do come with new challenges and questions, e.g. the nudging timescales, parameters, height range etc, that could determine the effect of the nudging. Nudging has not been systematically evaluated for CCMI models and e.g. Chrysanthou et al. (2019) shows that nudging does not necessarily lead to improvements in the stratospheric residual circulation. Therefore, while the suggested nudged experiments would potentially be beneficial for an improved understanding of linkages between stratosphere and troposphere in our models, further work would be needed to evaluate and test such experiments. Thus, experiments with a nudged stratospheric circulation are interesting material for an extensive follow-up study.

Moreover, we wish to note that while 2002 and 2019 show the strongest surface responses, there is still a surface effect, particularly over Antarctica, when leaving these two most extreme years out (Fig. 4). Leaving out 2002 and 2019 makes the polar vortex anomalies comparable between models and observations.

Other comments:

RC 1.4 By extreme events, the authors mean top/bottom 25% SAM index of all years. Those events may be called anomalous, but not extreme. Probably +/- 1 std would be a better threshold.

AR 1.4 We agree that using the term "extreme" for the 25% top and bottom years is misleading. We chose this percentile as a trade off between a sufficiently large sample size and events that are still anomalous enough to produce a surface result. Furthermore, using this percentile includes the same events in the reanalysis as in previous studies (Thompson et al., 2005; Byrne and Shepherd, 2018). For the above reasons, we have not adapted the threshold, but we now refer to these events as "anomalous" in the manuscript, instead of "extreme".

RC 1.5 l. 296, Fig. A5: 'Model biases in the jet position are consistent with the CCMs' tropospheric SAM surface patterns' - what do you mean? I agree that there are biases in the jet location (Fig. 7), but for the temperature response I don't see biases, I see

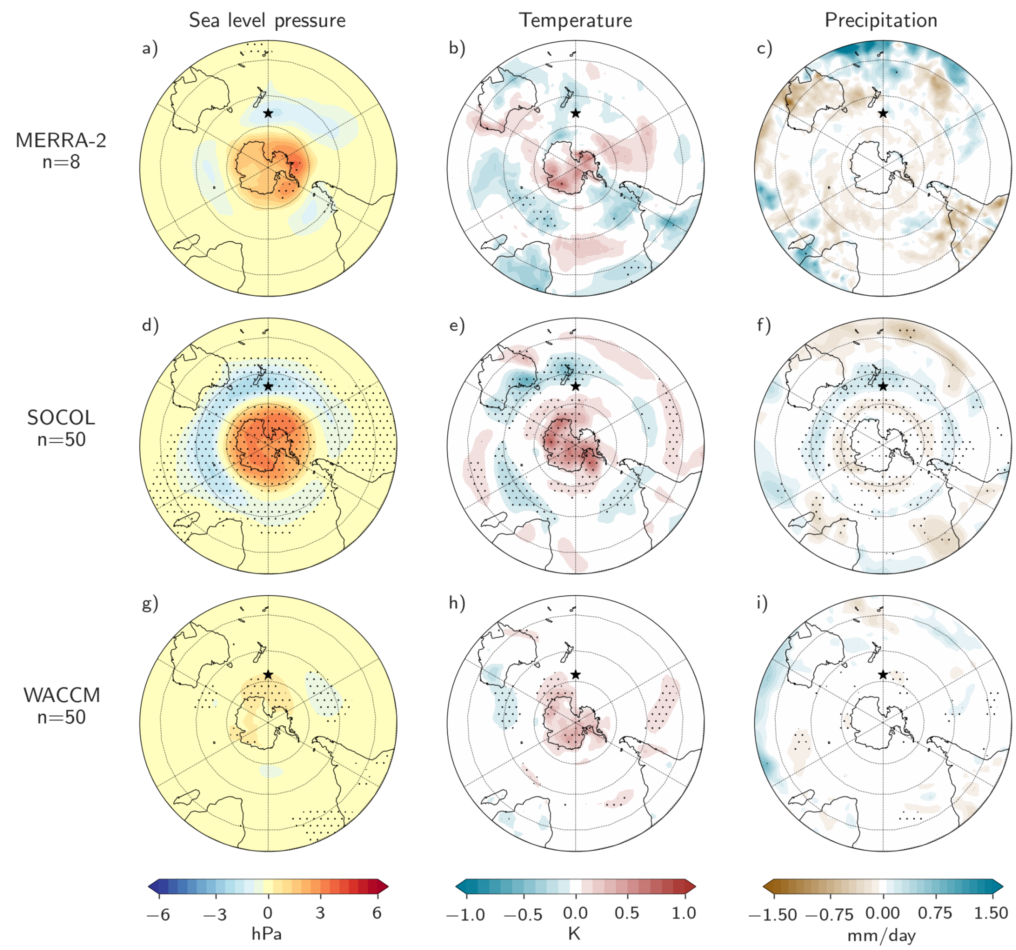

Figure 4: Surface composites following weak polar vortex anomalies without the 2002 and 2019 events in MERRA-2.

lack of skills outside of polar regions and high midlatitudes. This is particularly true for Australia. This plot illustrates why, as I stated earlier, these models could not be used to study surface impacts of polar vortex events outside of Antarctica. Therefore, I'd suggest limiting the analysis to the processes that are reasonably well simulated by the models (e.g., jets).

AR 1.5 We mean that the tropospheric SAM in models projects differently on surface climate (e.g., temperature and precipitation) than in reanalysis, as is visible in Figure A5. Furthermore, these biases in the SAM surface pattern (of the tropospheric SAM in general, not necessarily related to the downward coupling of the stratosphere) are likely related to the model biases. For instance, as the tropospheric jet is typically more equatorward in SOCOL, the corresponding temperature patterns are slightly shifted equatorward as well. To clarify, we therefore reformulate to: *Model biases in the jet position are consistent with the CCMs' tropospheric SAM surface pattern, as a northward shift implies a northward shift of the temperature patterns, and vice versa.*

RC 1.6 Fig. 8: For WACCM, instead of saying that 'tropospheric SAM timescales are shorter than in the reanalysis', it would be good to say that the stratospheric SAM does not quite reach the surface. Also, weak stratospheric SAM in WACCM becomes very obvious.

AR 1.6 We would like to emphasise that the SAM timescale is a constructed value and does not directly represent the SAM. From Figure 2 in the manuscript, we see that the SAM anomaly in WACCM reaches the surface for intermittent time periods. This means that although the stratospheric SAM reaches the surface in WACCM, the surface SAM response decays rapidly due to the small time scale in the troposphere (i.e. indicating too little persistence of the tropospheric circulation) in WACCM.

RC 1.7 Considering the strength of modelled polar vortices, perhaps most extreme modelled vortices can be compared with vortex anomalies in the reanalysis of the same magnitude? Alternatively (or additionally), other experiments can be conducted that amplify the polar vortex anomalies to observed values to assess their impact.

AR 1.7 In our bootstrap analysis and Figure 6 in the manuscript, we look at subsampled composites (so the number of samples is the same for the models and reanalysis data) and here we can compare similar magnitudes of polar vortex anomalies to the average surface temperature over Antarctica and Australia:

For Antarctica, when using similar stratospheric SAM magnitudes (red box in Fig. 5), the Antarctic temperature anomaly between models and reanalysis becomes comparable, although models show a larger spread (likely due to the larger number of years available).

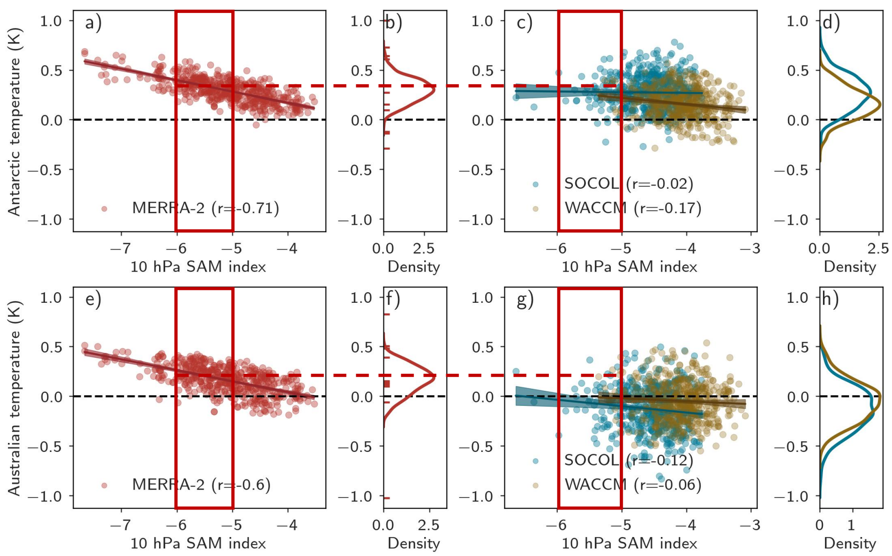

Figure 5: Scatterplots of 500 synthetic bootstrapped composites (n=10) for Antarctic (a,c) and Australian (e,g) 2-meter temperature anomalies in weak polar vortex regimes vs. the composite stratospheric 10 hPa SAM peak anomaly. The r-value refers to the Pearson correlation coefficient between the two quantities, the line refers to the fitted linear regression with the 95 % confidence interval of the slope in shading. The kernel density estimation (KDE) of the temperature composites is shown on the right of the scatterplots (b,d,f,h), and the mean 2 m temperature values for the 10 weak polar vortex regimes are marked with red dashes for the reanalysis datasets.

For the Australian temperature anomaly, only very few composites are comparable to the reanalysis but the bulk shows lower temperatures. This leads to the conclusion that the lack of extreme stratospheric weak events in models is not responsible for the missing surface response.

RC 1.8 l. 331: if the observed warming signal over Antarctica and Australia is robust, then do you need to use models at all?

AR 1.8 Even though the bootstrapped analysis suggests that these signals are robust, they are nevertheless constrained to a small number of observations (that can be resampled with replacement). We compared our results to another reanalysis dataset (JRA-55) (Figure 6) and see surprisingly large differences between JRA-55 and MERRA-2, which is likely also influenced by the fact that over polar regions, less data is available to be assimilated and the reanalysis product is more strongly influenced by the model. Climate models aid in the understanding of the underlying processes and causalities. Therefore, it is also relevant to evaluate how they simulate certain processes and document biases that could motivate future improvements.

Minor:

RC 1.9 Fig. 3: Three subplots can be merged into one plot, similar to Fig. 1. There is enough space alone X axis to show six box-whiskers (colours clearly separate strong/weak events).

AR 1.9 We have adjusted the figure accordingly (Fig. 7).

RC 1.10 Fig. A2: please add labels for Y axis

AR 1.10 We thank the reviewer for pointing this out. We have added the labels (Fig. 8).

RC 1.11 l. 333, 'we find that the Australian temperature signal is even more uncertain than the warming over Antarctica': The fact that Australian signal is more uncertain than warming over Antarctica would be expected, please re-word.

AR 1.11 We reformulated to *In the model experiments, we find that the Australian temperature signal is very uncertain, with equal likelihoods of warming and cooling.*

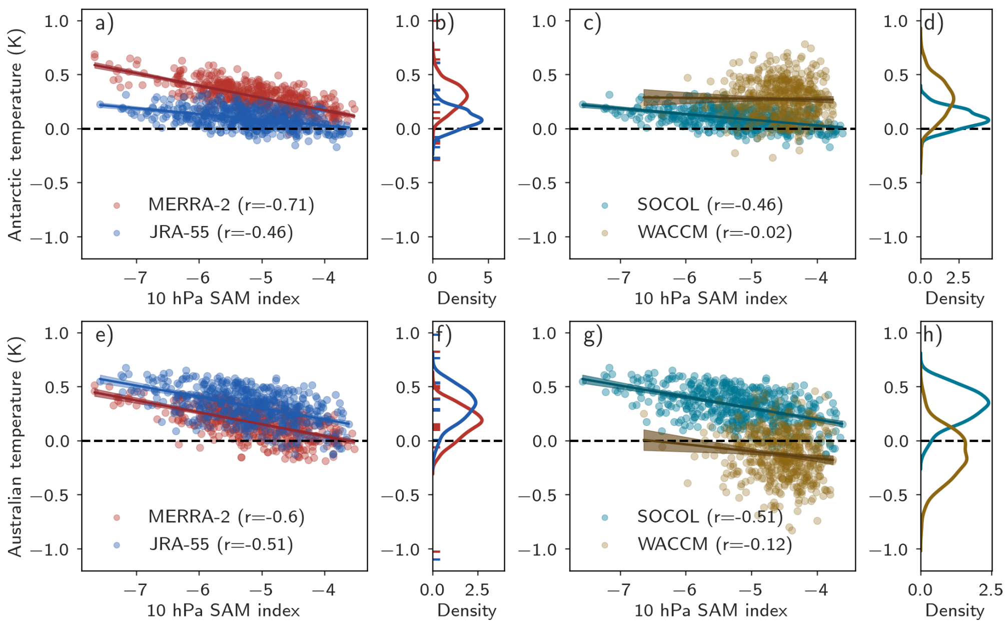

Figure 6: Scatterplots of 500 synthetic bootstrapped composites (n=10) including JRA-55 reanaylisis data for Antarctic (a,c) and Australian (e,g) 2-meter temperature anomalies in weak polar vortex regimes vs. the composite stratospheric 10 hPa SAM peak anomaly. The r-value refers to the Pearson correlation coefficient between the two quantities, the line refers to the fitted linear regression with the 95 % confidence interval of the slope in shading. The kernel density estimation (KDE) of the temperature composites is shown on the right of the scatterplots (b,d,f,h), and the mean 2 m temperature values for the 10 weak polar vortex regimes are marked with red dashes for the reanalysis datasets.

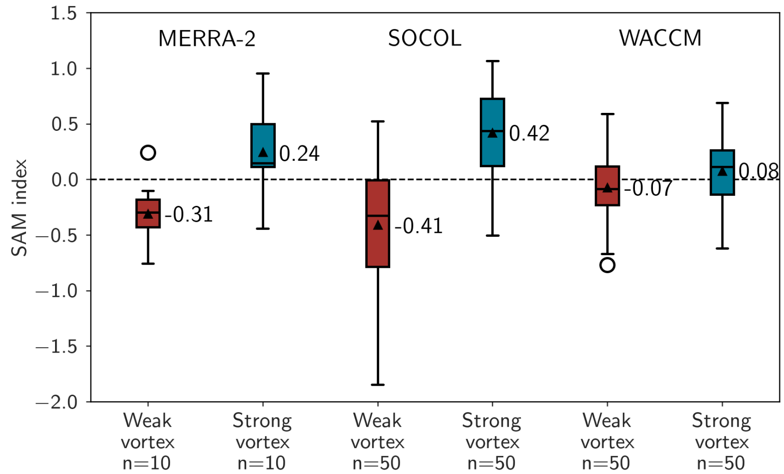

Figure 7: Distribution of the mean 500 hPa SAM index averaged over the October-January period following weak and strong polar vortex anomalies for MERRA-2 (a), SOCOL (b), and WACCM (c). The box extends from the lower to upper quartile of the data, the whiskers extend from the lower quartile −1.5 IQR to the upper quartile +1.5 IQR. Data points outside of the whiskers are shown as circles. The horizontal line marks the median value and the triangle indicates the mean of the distribution, which is annotated next to the box.

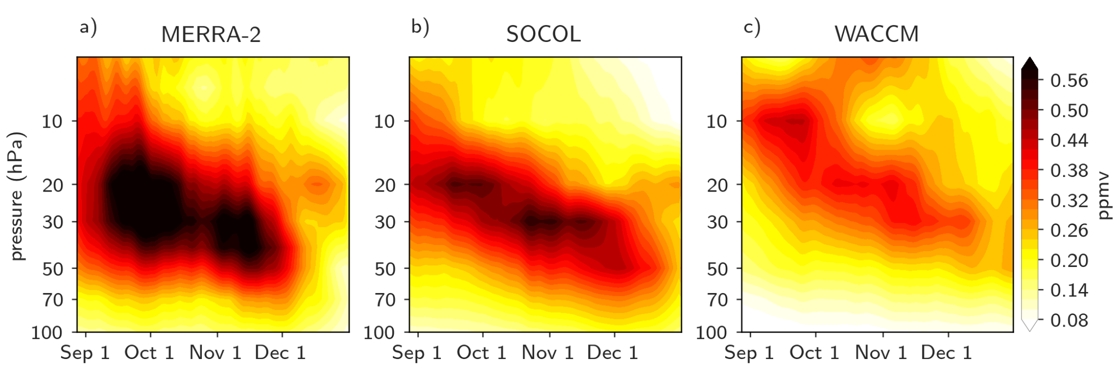

Figure 8: Ozone standard deviation in austral spring in MERRA-2 (a), SOCOL (b) and WACCM (c).

References

Byrne, N. J. and T. G. Shepherd (May 2018). "Seasonal Persistence of Circulation Anomalies in the Southern Hemisphere Stratosphere and Its Implications for the Troposphere". In: *Journal of Climate* 31.9, pp. 3467–3483. DOI: 10.1175/JCLI-D-17-0557.1.

Chrysanthou, A., A. C. Maycock, M. P. Chipperfield, S. Dhomse, H. Garny, D. Kinnison, H. Akiyoshi, M. Deushi, R. R. Garcia, P. Jöckel, O. Kirner, G. Pitari, D. A. Plummer, L. Revell, E. Rozanov, A. Stenke, T. Y. Tanaka, D. Visioni, and Y. Yamashita (2019). "The effect of atmospheric nudging on the stratospheric residual circulation in chemistry–climate models". In: *Atmospheric Chemistry and Physics* 19.17, pp. 11559–11586. DOI: 10.5194/acp-19-11559-2019.

Morgenstern, O., M. I. Hegglin, E. Rozanov, F. M. O'Connor, N. L. Abraham, H. Akiyoshi, A. T. Archibald, S. Bekki, N. Butchart, M. P. Chipperfield, M. Deushi, S. S. Dhomse, R. R. Garcia, S. C. Hardiman, L. W. Horowitz, P. Jöckel, B. Josse, D. Kinnison, M. Lin, E. Mancini, M. E. Manyin, M. Marchand, V. Marécal, M. Michou, L. D. Oman, G. Pitari, D. A. Plummer, L. E. Revell, D. Saint-Martin, R. Schofield, A. Stenke, K. Stone, K. Sudo, T. Y. Tanaka, S. Tilmes, Y. Yamashita, K. Yoshida, and G. Zeng (Feb. 2017). "Review of the global models used within phase 1 of the Chemistry–Climate Model Initiative (CCMI)". In: *Geoscientific Model Development* 10.2, pp. 639–671. DOI: 10.5194/gmd-10-639-2017.

Thompson, D. W. J., M. P. Baldwin, and S. Solomon (Mar. 2005). "Stratosphere–Troposphere Coupling in the Southern Hemisphere". In: *Journal of Atmospheric Sciences* 62.3, pp. 708–715. DOI: 10.1175/JAS-3321.1.