Comment on acp-2022-229
Anonymous Referee #1

Referee comment on "On the Robustness of the Surface Response to Austral Stratospheric Polar Vortex Extremes" by Nora Bergner et al., Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2022-229-RC1, 2022

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.

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?

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.
Other comments:

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.

I. 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 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).

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.

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.

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

Minor:

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).
Fig. A2: please add labels for Y axis

I. 33, ‘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.