“**Significant additional Antarctic warming in atmospheric bias-corrected ARPEGE projections”** by J. Beaumet, M. Déqué, G. Krinner, C. Agosta, A. Alias & V. Favier

Author response to referee’s comment.
Anonymous Referee #1
Received and published: 19 November 2020

General comment:
This study consists of two parts: implementation of simulated bias correction (with respect to a reanalysis dataset) in an atmospheric general circulation model (AGCM) and the effect of bias correction on assessing the future climate change in the Antarctica. The use of the AGCM with refined resolution near the Antarctica allows one to acquire detailed, useful information with only prescribed lower (oceanic) boundary conditions and without lateral (regional) boundary conditions. It can be seen as a type of dynamical downscaling. The authors implemented a method of correcting biases in the model, and their approach is, I think, unique for this context, and potentially very practical. The description of the methodology (Sect. 2) is clearly given and it is easy to follow the text throughout. The authors demonstrated that the result differs significantly with and C1 without the bias correction. I found that the first part of the study (implementation of the bias correction) (Sects. 3.1 and 4.1) is nicely done but the second part (Sect 4.2) needs significant improvement as argued below.

Authors : We thank the referee for their encouraging review and constructive comments on the manuscript. A point by point response to each comment is given below.

Major comment:
1. As in the title, abstract, and conclusion, the most important scientific finding in this study is that the climate change signal is assessed differently with and without the bias correction. The heart of the discussion should then be placed on whether the bias-corrected climate change simulations are more reliable or not compared to the uncorrected climate change simulations. It is not trivial because the addition of extra terms to the tendency equations violates the conservation laws of physics and may distort the processes operating in the model. There are at least two ways to check the validity of the approach. The first method is a perfect model study in which the bias in a model with respect to the simulated present-day climate by another model is to be corrected and one can investigate whether the climate change signal simulated by the second model is better reproduced by the bias-corrected first model. The second method is to provide a physically persuasive mechanism/rationale why the bias-corrected simulations are more reliable. Now, the

first approach requires two models and too much extra work, and still the result may depend on the reference models used. The second approach is, on the other hand, feasible and essential.

Authors : We agree with the reviewer on this relevant remark. Regarding the first point, we will investigate the typical values of the added/removed energy, moisture and momentum associated with the values of the correction terms. Unfortunately, physical tendencies associated with the radiative scheme or the dynamics in our ARPEGE simulations were not saved in the output, but we will take typical values in inter-model studies from the literature (e.g., Cesana et al., 2019, https://doi.org/10.1175/JCLI-D-17-0136.1). Results of this comparison will be shown in the final answer, and added to the supplementary materials of the paper. In the meantime, we can already inform the reviewer that in the paper by Krinner et al., (2019) using the same method with the atmospheric model LMDZ at lower horizontal resolution, the correction terms associated with the tendency errors were found to be much smaller that the tendency associated with the other processes of the model physics (radiative transfer, dynamics…).

Besides, we want to stress the fact that in AMIP-style experiments as those presented in this study, the fact that the oceans are considered as a surface boundary condition and that it can act as an infinite source or sink of energy already distort the laws of energy conservation in the experiment.

On the second point regarding the possibility to demonstrate the validity of the approach for climate change projections using a perfect model study, this is what has been done in Krinner et al., 2020 (https://doi.org/10.1038/s43247-020-00035-0). In this study, three global atmospheric models, one of them being ARPEGE, emulates one of the two other model corresponding coupled climate model (CGCM) considered as the reference (pseudo-reality). Then future projections were performed using the same corrections term and the "bias-corrected" future projections (RCP8.5 scenario) are compared to the original future projections of the AOGCM used as reference in the beginning. Results show that about 70% of the added-value of the bias-correction remains in this perfect model study for RCP8.5 projections at the end of the 21st century. This study demonstrates the validity of the approach even for climates that are significantly warmer than the one in which the bias-correction terms have been built. These results are consistent with the results from Krinner and Flanner., 2019 (https://doi.org/10.1073/pnas.1807912115) who showed that a climate model can be easily and automatically identified by its departure from the ensemble mean in both present historical simulation and late 21st century projection (abrupt 4xCO_2 scenario). These results together suggest that climate model biases are mostly stationary within (and even beyond) the range of climate changes projected during this century and allow for the use of empirical bias-correction terms derived in present-day climate using observations or reanalysis as references for future climate projections.

Regarding the last point, some physical explanations on why climate models with more poleward/broader jet structure (that agree better with present-day climate) show less poleward shift in a warmer climate have been proposed in Bracegirdle et al., 2013 and other previous works. The following can be found in Bracegirdle et al., 2013 : "*These mechanisms, put forward by Barnes and Hartmann [2010] and Simpson et al. [2012], both relate shifts in the jet to tropospheric eddy feedbacks that depend on the time mean jet structure. Barnes and Hartmann [2010] suggest that differences in eddy feedback could originate from differences in wave breaking on the poleward side of the sub-polar jet. According to this mechanism a poleward shift occurs when poleward breaking is suppressed and the resulting wider jet extends to higher latitudes. The implication is that models with jets that already exhibit weak poleward wave breaking (and a wider, more poleward, structure) show weak shifts under global warming, since wave breaking is already suppressed in those models. An alternative mechanism relating to eddy feedbacks on the equatorward side of the jet was suggested by Simpson et al. [2012], who show that the tropospheric eddy feedback (and poleward shift) is stronger when the distance between the sub-polar eddy-driven jet and the sub-tropical critical line is smaller. Higher latitude jets with a larger distance exhibit a weaker poleward jet shift due to a weaker latitudinal coherence of eddy momentum flux convergence across the phase speed spectrum. Their results are for a specific case of tropical stratospheric heating in a simplified GCM (sGCM), but may be relevant more generally.*"

We propose to briefly investigate in our corrected and non corrected simulation the upper-level meridional temperature gradient and/or the distance between the sub-polar jet and the sub-tropical critical line such as done in Bracegirdle et al., 2013 for the CMIP5 ensemble. If relevant, the results of these analysis will be added to the discussion and/or the supplementary material. These processes are widely influenced by the interdecadal climate variability and no robust conclusion can be drawn from only two pairs of 30-years simulations. However, when we put them in the context of previous CMIP5 large ensemble analyses (Bracegirdle et al. 2013), we confirm that these processes are likely at play in our simulations (Fig R1).

Besides, the findings Barnes and Hartmann, 2012 (https://doi.org/10.1029/2012JD017469) suggest that the reduced poleward shift of the eddy-driven jet and the lesser deepening of circum-antarctic low pressure systems found in the empirically bias-corrected simulations is indeed likely to be more realistic. Barnes and Hartmann, (2012) showed that the poleward shift of cyclonic wave breaking reaches a poleward limit around 60°S and that wave breaking on the poleward side of the jet will become less frequent for any further poleward shift of the jet. Their results suggest that there is a theoretical limit to how far South the location of the maximum cyclonic wave breaking can move poleward, most likely controlled by the absolute vorticity gradient and that the observed Southern Hemisphere circulation may already be close to this limit. Therefore, we will refer to this paper in our discussion, as well as the previous arguments in order to highlight more in the discussion the reliability of projected circulation changes in bias-corrected simulation.

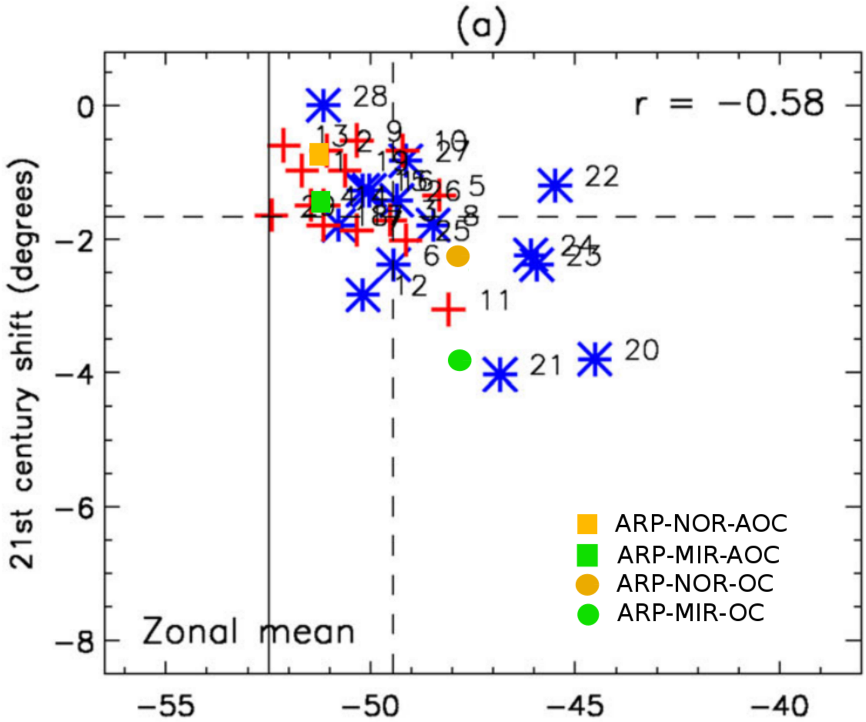

Fig R1 : Projected twenty first century surface westerly wind maximum poleward shift as function of the latitudinal position in the historical simulation for the CMIP5 ensemble (red and blue symbols) and for the ARPEGE corrected (squares) and uncorrected (circles) projections presented in this study. Modified from Bracegirdle et al., (2013).

In Sect. 4.2.1, the authors describe how the changes in atmospheric circulation and sea level pressure are different with and without the bias correction, but they do not discuss the reason and mechanism. It is, thus, difficult to assess which result is more reliable. Before discussing the difference from previous studies, they should investigate and explain the mechanism in their own model.

Authors : We think that the arguments brought in the previous question concerning the reference between previous studies that established a link between a poleward position of the eddy-driven mid-latitude jet and a reduced poleward shift (Bracegirdle et al., 2013, Barnes and Hartmann, 2012) together with the stability of the added-value of the bias-correction in future projection realized within the framework of a perfect model test in Krinner et al., (2020) are good arguments in favour of a higher reliability of the projected circulation changes in the bias-corrected projections.

Moreover, we want to stress the fact that for future climate projections not only the projected climate change signal (relative difference with present-day climate) matters but also (maybe even more) the absolute mean state of the climate at the end of the 21st century, especially within the framework of impact assessment studies. In this regard, the biases in the non corrected

simulations being, especially for the atmospheric circulation in the high southern latitude, of the same order of magnitude as the projected changes, its is by construction not possible that the climate mean state at the end of the 21st century is more reliable in uncorrected future projections.
Investigating the causes of changes in circulation between corrected and uncorrected projections cannot be done with simple diagnostics and would require complex diagnostics such as done in Barnes and Hartmann, (2012) or run-time diagnostics such as radiative kernels (e.g., Soden et al., 2008, https://doi.org/10.1175/2007JCLI2110.1) that are beyond the scope of this study.

In Sect. 4.2.2, it is stated that the difference in summer SAM change has large impact on the surface warming, but again they do not investigate why the summer SAM is different and which dynamical mechanism responsible for it is affected by the imposed bias correction. Moreover, they cite PSA1, PSA2, and Amundsen Sea Low differences as potentially important to understand the temperature difference with and without the bias correction, but they do not explain the link of these differences to the background climate state.

Authors : The projected evolution towards a more recurrent positive phase of the SAM at the end of the current century results from the projected increase in meridional pressure gradient (deepening lows and increasing mid-latitude high pressure systems) associated with the poleward shift of these structures and of the westerly jet caused by the increase in greenhouse gas concentration. Here again, this poleward shift is likely constrained by the absolute vorticity gradient such as suggested by the results of Barnes and Hartmann (2012) and models that are strongly underestimating the meridional pressure gradient and show an equatorward bias in the position of the westerly jet such in the historical climate as in the uncorrected simulation with ARPEGE are prone to overestimate this increase towards a positive phase of the SAM.

The variability of the Southern Annular Mode was found to have larger influence on the temperature anomalies over the East Antarctic Plateau (Marshall, 2007) and over the Antarctic Peninsula (Clem et al., 2016) in summer and autumn.

Therefore, we will modify the text to explain more explicitly why the bias-corrected simulations evolve towards a less pronounced positive phase anomaly in future projections, why it is likely to be more realistic than the uncorrected projections and why this has a large impact on summer temperatures over East Antarctica.

We agree with the reviewer that introducing the PSA1 and PSA2 mode of variability as possible explanations for the differences in climate change between corrected and uncorrected projection without properly introducing these mode of variability and their link to background climate, and without investigating their representation in our climate simulation is scientifically questionable. We think that the impact of these processes are most likely of second order of importance compared to the impact of the correction on the main pattern of the atmospheric general circulation but these processes would nevertheless deserve to be investigated in a separate study. Therefore, we will delete any reference to the PSA1 and PSA2 mode of variability to interpret our results. The impact of the correction on the position and depth of the Amundsen Sea Low and its projected displacement in future projection being more straightforward and easier to interpret, it will still be discussed but we will introduce more details and reference to previous work in this part of the discussion.

In Sect. 4.2.3, the difference in precipitation change is attributed to the atmospheric circulation difference, but the link between the unperturbed climate state and the circulation difference is not explained. To my view, these are the most important scientific points of the paper, and the necessary data to explore are all at the authors' hands. Without these explanations, one may see the paper as simply demonstrating how the present-day simulation becomes close to the reanalysis dataset after the correcting terms are added (as implemented), and the simulated future climate change signal are affected by the bias correction for unknown reasons. As it is unclear which one (bias-corrected or uncorrected) is more realistic in the climate change simulations, one could argue that the main conclusion is not convincingly established.

Authors : Regarding the reliability of projected precipitation change, we think a good example can be given in the western West Antarctica region (Maria Byrd Land). In this region, the position and depth of the Amundsen Sea Low, currently located at the fringe of the Amundsen and Ross Sea in winter and spring, has a large influence on the advection of moist air and therefore precipitation on these coastal regions. In the uncorrected historical simulation, the position and depth of this climatological pressure minimum is widely biased. Unsurprisingly, total precipitation in this region is better simulated (better agreement with MAR and RACMO and also with CloudSat snowfall, see response to the second reviewer below) in the bias-corrected simulation. Similarly to the example given by Maraun et al., 2017 (https://doi.org/10.1038/nclimate3418) for projected changes in precipitation across western Europe, since the position and depth of this main low pressure system is largely biased in the uncorrected historical simulation, the absolute deepening and displacement of the centre of the climatological low cannot be correct in the uncorrected future projection and so will be its impact on regional precipitation change in coastal areas.

2. As the tendency terms are corrected at each time step so that it should reproduce the reanalysis datasets, it is not surprising that the simulated result shows better agreement with the targeted dataset. I would not describe it as "improvement" as the authors claim. It only demonstrates that the implementation worked as designed. I do not at all mean that it is easy to achieve it (indeed I appreciate the effort and see that the approach has a great potential), but this part of the study is more a technical advancement, rather than a scientific one. Please highlight which results are unexpected (or surprising) in simulating the present-day climate with the bias correction and implication of those unexpected results. The climate modelling community had taken the similar

approach of so-called flux adjustments decades ago and virtually abandoned it by now due to various side effect. I am a bit surprised that the authors do not touch upon such a historical path in the model development and not discuss why the authors consider it worth to be revived.

Authors : We agree with the reviewer that the term “improvement” has been misused here and we will avoid using it in this context, and use the terms “bias reduction” instead.

We think that obtaining a better representation of the daily variability of the large-scale atmospheric circulation, such as evidenced in the application of self-organizing maps on sea-level pressure fields, was a non expected (by construction, the bias-correction is only expected to improve mean state) and interesting result. Krinner et al. (2020) found similar results for both inter-annual and synoptic scale variability in their application until 2100 using the perfect model test framework and so did Kharin and Scinocca (2012) in their application of the method for seasonal prediction (doi:10.1029/2012GL052815).

We will add some historical perspectives about the flux adjustments method in the climate modelling community in the introduction. Interestingly, we note that Guldberg et al., 2005 in their first application of flux correction with ARPEGE model for seasonal forecasting found improved skills mostly in the Southern Hemisphere. More recently, Dommenget and Rezny (2018) (https://doi.org/10.1002/2017MS000947) argued in a pilot study that a transparent, well documented flux correction (which is what we achieve in this study) is more desirable and cheaper in computational cost than model tuning which is involved in the development of many climate model. Unlike model tuning, flux correction does not introduce artificial errors between the submodels of the CGCMs.

Besides, as long as climate models have biases of the order of magnitude as the projected changes, *a posteriori* statistical bias correction will be applied to climate model output in the framework of climate change impact assessment. However, these methods are not able to correct for errors associated with large biases on atmospĥeric general circulation, particularly projected changes in precipitation (Maraun et al., 2017). In this regard, empirical bias correction (or “flux” correction) such as implemented in this study have large potential benefits. Even though these methods are not perfect and distort slightly the tendency associated with the model physics, they do so to a reduced extent compared to model parameter tuning which is widely implemented in the development of climate models. Moreover, empirical bias correction or “flux” correction can be easily switched off so that its impact on the model performance and on projected changes are well known. This way, using both bias corrected and uncorrected projections, we can assess remaining uncertainties on projected climate change and emphasize what should be the priorities in climate model development in order to reduce these uncertainties.

We will add the reference of Dommenguet and Rezny (2018) in the introduction and refer to it in the discussion (section 4.3). In this later section, we will insist more clearly on the points mentioned above, precisely the potential assets of flux correction methods with respect to model parameter tuning in climate model development and climate change assessment.

Minor comment:

1. In Fig. 1-(b)-left, is there explanation why the SLP bias in the vicinity of Antarctica does not disappear after the bias correction?

Authors response: This positive bias in the seas surrounding Antarctica, even though substantially reduced, especially in the Amundsen Sea sector, remains mostly in winter and spring. During these seasons these areas witness the formation of rapidly developing and evolving meso-cyclons (polar lows). It is therefore likely that the model, even in the bias-corrected simulation, fails to fully capture the formation of these polar lows. Here are two possible explanation :

- The characteristic time of formation of these cyclones is much smaller than the characteristic time of other larger scale cyclones. Therefore, the relaxation time of 72h used in the first nudged simulation towards climate reanalyses that is used to derive correction terms might be too wide to retrieve the right values of correction terms that should be applied to correct for the model deficiencies in simulating these phenomena.
- Katabatic winds flowing from the ice-sheet towards the coast play a key role in the formation of these meso-cyclones. Besides, the formation of a very stable, cold boundary layer at the surface of the ice-sheet plays a key role in the formation of the katabatic winds. In this study and in the previous one (Beaumet et al., 2019b), we have seen that the version of ARPEGE used in these studies has some deficiencies in capturing the formation of very stable boundary layer at the surface of the ice-sheet in winter (similarly to many climate models), which likely impacts the capacity of the model to reproduce correctly the katabatic winds regime around Antarctica and latter the formation of meso-cyclones over near-by seas. We remind that variables in the boundary layer (<100 m) are not corrected at all in the bias-corrected simulations.

We will briefly mention these hypotheses in our discussion of the remaining biases in the corrected historical simulation.