This paper evaluates the variability in diagnostics critical to ice sheet mass balance from simulations over Antarctica in three state-of-the-art regional climate models and two reanalysis products. This is a substantial piece of work that furthers our understanding of uncertainty in the atmospheric component of coupled climate models over Antarctica and informs the configuration of such models. The structure of the paper is well thought out, and the quality of the writing and figures are excellent. I recommend that the paper is accepted for publication following minor revisions addressing my comments, below.
From your plots, it looks to me as though model differences are generally greater over steep, mountainous terrain (e.g. along the coastline and over the Transantarctic mountains). This highlights differences in the representation of orography and meteorological conditions over orography. Previous studies have demonstrated that a major source of such differences is model resolution, e.g. there are significant differences in the reproduction of influential mountains winds between simulations with grid spacings of 12-km and km-scale (e.g. Orr et al., 2014, 10.1002/qj.2296 for katabatic winds in E Antarctica; Heinemann et al., 2021, 10.3390/atmos12121635 for foehn winds over the Antarctic Peninsula). These differences are particularly pertinent for surface mass balance over ice shelves, since the Antarctic coastline is generally found at the foot of the steep slopes of the Antarctic plateau, and/or is in the vicinity of mountainous terrain. So I’m intrigued as to whether in your results you can see larger systematic differences in the vicinity of steep terrain that can be attributed to resolution (using the two MetUM models), and whether these differences are of sufficient magnitude and spatial scale to be pertinent for ice shelf mass balance. Also, as alluded to above, it could be that even the 12 km MetUM simulations are insufficient to reproduce climatically important influences of terrain-induced airflows. You do mention katabatic winds as a potential source of model differences, specifically on the Amery IS, but I wonder if it might be worth commenting further on the influence of orography on model differences and the implications of this.

**Specific comments**

Line 8: “suggested” here is too weak. Suggest instead “Our results imply that...”

Line 26: “The primary method of ice shelf retreat is through oceanic basal melting”. I think this statement requires further qualification. Specifically, adding the word "currently", and something like "with the notable exception of some of the ice shelves on the Antarctic Peninsula (Pritchard et al., 2012)". Recent climate/ice-sheet modelling studies indicate that atmosphere-driven hydrofracture has in the distant past been, and will in the future be, the principal cause of Antarctic ice-shelf collapse (e.g. DeConto et al., 2021, 10.1038/s41558-021-03427-0; Pollard et al., 2015, 10.1016/j.epsl.2014.12.035).
Figure 3c: Why are the interiors of the Ross and Filchner-Ronne ice shelves masked out?

Figures 4-6: From neither the text nor the figures is it totally clear to me whether the difference is model - ensemble, or ensemble - model. I'd expect it to be the former, and that is indeed my impression from the text. However, "different to ensemble average" implies to me the opposite. Please make this clear, in the text where these figures are first referenced, and in the figure captions.

Line 446: “The primary sources of large-scale, systematic differences between the simulations, for all variables and components, are identified as deriving from differences in: the model dynamical core; the surface scheme; parametrisation and tuning.” In the discussion, the sensitivity of melt to the subsurface scheme is highlighted, and justification is given for the "secondary" importance of the factors listed in the subsequent sentence (driving data, resolution, domains etc.). We may then assume by way of elimination that the model dynamics and physics are the primary sources of systematic differences. I'm not sure though that this reasoning is actually stated, and I think it should be, somewhere in the Discussion section.

Line 456: “Therefore, as concluded in Mottram et al. (2021), there is an importance on observational campaigns to correct for biases.” Do you mean there is demand for new field observations with which to constrain model physics parameterisations? Or for (post-processing) model bias correction? This statement needs expanding on.

Line 458: “2100 SLR” A bit specific. Suggest simply "future SLR". The same applies in abstract, line 3.
Technical corrections / suggestions

Line 5: Suggest italicising “Seasonal and trend decomposition using Loess”, and also perhaps capitalising the T in trend, to make it clear that this is what STL stands for.

Line 168: Suggest italicising “Seasonal and trend decomposition using Loess”

Line 201: RMSD: This abbreviation is defined in the abstract, but should be defined when first used in the main text also

Line 220 and other instances: “mmWEqm\textsuperscript{-1}”... I think there should be spaces between the units, so perhaps “mm WEq m\textsuperscript{-1}”. Or “mm w.e. m\textsuperscript{-1}” as I’ve seen this notation used before.

Line 299: Remove repeated number 1 before “mm”