Carter et al. presents a comparison of key variables (Snowfall, temperature, melt) for a set models including MetUM, MAR and RACMO as well as reanalysis (ERAI and ERA5). They use statistical methods which compare with regard to mean, trend, seasonal cycle and the residual (which should capture physical differences). They examine the potential impact of several differences in the models, including the different use of DEMs, forcing at the boundaries, the size of the boundary itself and differences in the underlying physics, concluding largely that the major differences occur in the differences in the physics.

In general, figures are clear and helpful, the analysis is sound and this will be a valuable addition to the canon of model intercomparisons (especially with regard to how methods were applied systematically here). The main critiques listed below are about presentation, language, or which parts of the discussion should be highlighted:

Main Critique:

As surface melt occurs primarily during the summer, any differences in the seasonal cycle require a bit more explanation. For example, in Figure 6, the differences in the seasonal standard deviation for MAR are quite large over East Antarctic ice shelves. What does this mean? Does this mean that seasons are shorter in MAR? It is also worth highlighting immediately that (here in plain language): the only real estimates of surface melt that matter here are from MAR and RACMO. The simpler surface schemes in ERAI/ERA5/MetUM simply don’t capture these processes. I note, however, that the authors have provided a thorough explanation of the underlying physics (and similarly did a great job explaining
Agosta’s work on how the difference in how precipitation is treated in MAR and may lead to more precip in the interior). It is just worth noting these basic differences at the start (potentially within table 1)

Minor revisions:

Table 1: Include a column for time-period at which forced at the boundaries. This is mentioned in the text, but is a very important difference.

Potentially my misunderstanding, but how does the value and the magnitude of correlation differ from one another?

Fig. 3: Use different color scales for (a) and (b) to make explicit that they are different scales (alternatively, just add a note in the caption). This is stated in text, but easy to misread.

Line 205: The impact of systematic differences in snowfall/snow melt on estimates...

Line 208: For near-surface air temperature, differences...
Lne 245: ...each component of the timeseries and that for temperature and melt ...

Line 258-263: This is partially mentioned in methods, and should perhaps just be moved there

Line 264: “significant systematic” implies a more quantitative assumption. Perhaps use “substantial” or “spatially-coherent”

Line 268: Possibly worth mentioning the length scale in the Antarctic Peninsula as well.

Line 270: mean of the time series is highly correlated : as “correlated” is a specific quantitative term here, I think that “has a similar spatial signature” might be more accurate

Line 281: Similarly (to above) I wouldn’t say “weak, negative correlation”, but rather something like “contrasting spatial patterns” and specify where.
Line 284: Figure 5: Point to location of “example grid-cell” in the figure

Line 294: How meaningful is the seasonal standard deviation if the majority of melt happens in summer? (What does this mean physically? Does this imply that the seasons shift?)

Line 295: The physical meaning of this metric isn’t entirely clear to me: I understand this as a measure of the bias correcting for the effect of season and residual, but it might be helpful to make this explicit and expand a little more.

Line 298: The masking for 1mmW Eq was mentioned in Methods, I don’t think it needs to be mentioned again.

Line 310: Again, I’m having a hard time parsing what “adjusting for equal means and seasonal/residual standard deviations” means here. I think it’s worth explaining a bit more what this metric means physically (e.g. when correcting for the seasonal effect, the mean trend, the residual is a metric of the physics).

Line 321: The large-scale differences in snowfall (ERAII vs ERA5) are attributed here to model physics, but we’re also seeing very large differences in the DEMs. Similarly, Met models show quite similar DEMs (Figure C1)
Line 342-344. This sentence is a little long and could be broken up.

Line 354: It is found that for the MeUM(044) run, the buffer...

Line 355: to the buffer zone boundary, and that even...

Line 386 – 396: This paragraph is a little convoluted generally.

Line 418: this seems to be reversed. (positive over ocean, negative over large region of East Antarctica near the Transantarctic mts.

Line 424: The use of the term “systematic differences” needs to be more clearly differentiated from “correlation” here. I think what you mean is that “this correlation occurs over large portions of the ice sheet”

In general, I think this paragraph needs a summary sentence for melt, e.g. that the melt bias is consistent, and wide-spread, even accounting for the seasonal and trend components

443: Despite this, there exists (add comma)
Sincerely,

R. Tri Datta