Reviewer 2

General Comments:

This paper presents results from a first intercomparison of polar regional climate models (RCM) applied in the Antarctic Ice Sheet (AIS). The model performances were compared and assessed in terms of surface pressure, near-surface air temperature, near-surface wind speed, surface temperature, and surface mass balance (SMB) of the AIS. The models that participate in this intercomparison project are COSMO-CLM2, HIRHAM5, MetUM, MAR, and RACMO. For some models, results from different versions are provided additionally.

We thank the reviewer for their thoughtful comments and have addressed in detail the points they raise below.

My first honest impression after reading through this manuscript is that the current title “What is the Surface Mass Balance of Antarctica?” is a bit misleading, because meltwater runoff is not considered in the most participating models except for MAR and RACMO. It is true that a contribution by runoff to the changes in the present- day AIS SMB is relatively small than contributions from precipitation and sublimation/evaporation. But, runoff in the present-day AIS already cannot be neglected as presented by several studies cited in this manuscript. In the future in a warming world, the contribution by runoff to the changes in AIS SMB will become much higher almost certainly as pointed out by the authors (P. 2, L. 18 ~ 19). Therefore, this reviewer expected that all the models calculated runoff in the present study, and as a result, I was a bit disappointed when I found the relevant description in Sect. 2.2.1.

Related to the point indicated above, the intercomparison procedure for SMB sounds a bit inadequate to me, because the authors employ different definitions of SMB (Sect. 2.2.1). If the authors focus intercomparisons only for precipitation and sublimation/evaporation (in addition to the three surface meteorological properties as well as the surface temperature), it makes sense and highlights key differences in model physics employed by these participating models more clearly. This reviewer recommends the authors to reconsider the title of this manuscript: maybe something like “intercomparison of Antarctic ice sheet surface meteorological conditions simulated by five different regional climate models” would be appropriate. However, the intercomparison of RCMs performed in the AIS itself is a considerable new challenge, so provides the latest comprehensive information related to these RCMs, which is very informative for readers certainly, so deserved to be published

In this paper we focus on the Surface Mass Budget of Antarctica and the uncertainties introduced by using different regional climate models to calculate it, even when those models are forced by the same global model. Overall, precipitation dominates SMB to such an extent at the present day that even subtracting all runoff from the models that calculate melt and refreezing leaves the overall SMB virtually unchanged. The difference is negligible even on a basin scale. We agree that melt is likely to become more important in the future, but at the present day melt and runoff are only observed at a few very specific locations.

However, while all of the models simulate melt, they have varying degrees of complexity to calculate refreezing so purely to simplify the comparison here we focus on precipitation, evaporation and sublimation terms. Two of the models include sublimation from blowing snow subroutines, and as the physical parameterisations have been developed with these processes in the model, we have also used the sublimation from snow schemes in the results.

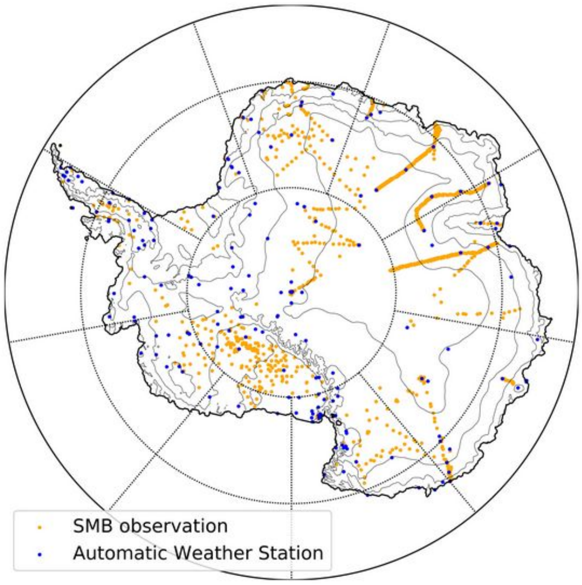
We absolutely agree that there are important questions around melt extent that have important implications for future SMB projections and we plan to extend this study with a detailed look at how the models simulate melt and runoff in a paper currently in preparation. For now, we have however added in the introduction and discussion sections more detail on how SMB is computed (see response to reviewer 1 also) and the processes that can be and are included and address the issue the reviewer raises in more details.

We note also that we do compare modelled SMB with measured SMB from stakes and from other studies such as the IMBIE study (see also reply to reviewer 1) and we think therefore it is justified to keep the title as it is.

Specific comments (major)

P. 2, L. 7 ~ 8: What kind of measurements do the authors think here (observational campaigns)? Maybe it is not necessary to indicate explicitly here; however, please suggest something at least in the discussion and/or conclusion sections.

Our results suggest that in particular stake measurements of SMB are crucial. These need to cover locations where there are very few recent observations and where there are large disagreements between the models. We propose adding a new figure (see below) to the supplementary materials that show the locations of SMB observations as well as locations of weather stations in order to demonstrate the significant data gaps. We have made this clearer in the conclusions.



P. 12, L. 2 ~ 3: What kind of physical mechanisms do the authors think here? Please detail more.

In locations with melt we expect that the lack of refreezing will affect the latent heat release in the snowpack which will in turn affect observed 10m depth temperatures. Furthermore, the diffusion of and conductivity of the surface snow layers is affected by the presence or absence of ice layers and by density which is dependent on densification schemes that this version of HIRHAM and COSMO-CLM2 do not have. We have clarified this to read:

"However, biases in cloud cover and long-wave radiation reaching the surface are likely the main explanation for divergence from observations and should be investigated for all RCMs run for Antarctica as shown by van Wessem et al. (2014). In their study, significant improvements in the RACMO2.3p2 model were obtained by adjustments to the cloud microphysics. Furthermore, the lack of detailed subsurface snow pack schemes including processes such as refreezing (and subsequent latent heat release) and densification also likely has an impact on the temperature bias in HIRHAM5 and MetUM (see also figure 2)"

P. 12, L. 5: How large do the authors think the uncertainties are?

It is very difficult to quantify uncertainties on these observations, particularly wind, as they are made at mostly automatic unstaffed stations and are subject to different biases depending on location from effects such as burial by snow, changes in orientation due to wind and breakdown of sensors among others.

Modified to:
"This is likely in part due to large uncertainties in the observations especially at unattended stations where burial by snow, changes in orientation and sensor breakdown are more likely. However, the effects of different resolution and differences in turbulent schemes between the models may also be important. In particular the extremely stable boundary layer over most of Antarctica is hard to represent in models particularly at lower resolutions"

P. 12, L. 5 ~ 6: Readers cannot know the difference in turbulent heat schemes, because they are not described in this manuscript.

A detailed description of all of the models turbulent energy schemes is beyond the scope of this paper, we have however added relevant references to these for each model to support our interpretations of model biases in an expanded section 3.1.

P. 24, L. 13 ~ 15: It is interesting to see the model-simulated precipitation integrated over the common ice sheet mask by COSMO-CLM2 tends to be lower than that by the parent data ERA. It is because precipitation in a dynamically downscaled data is higher than precipitation in its parent data in general. Please discuss.

Our analysis suggests that the COSMO-CLM2 model used in this simulation has indeed a dry bias compared to the other models and the reasons behind this are the subject of ongoing work. The bias was first identified by Souverijns et al., 2019 and seems in part to be a consequence of a particularly dry bias on the coast, especially in the peninsular and west Antarctica but there is conversely an overestimate of precipitation in the interior. We have added these details to the paper.

P. 29, L. 27: Please suggest what kind of measurements do the authors think necessary in the "observational campaigns"?

Given the importance of precipitation and snow processes to the SMB in Antarctica, stake and radar measurements, supplemented with firn cores are a clear priority. New observations should focus where possible in regions where there is a lack of measurements, but also where there is strong disagreement between models, as identified in the Results section. We have added these points explicitly to the conclusions.

Specific comments (minor)

P. 4, L. 13 ~ 18: Is it OK to understand MAR 3.6 is older than MAR v3.10? If yes, it is a bit confusing isn't it?

MAR v.3.10 is the more recent version of MAR

P. 8, L. 6: What do the authors mean by "cloud physics"? To me, it is difficult to understand why "cloud physics" is resolved better in nudged models.

As nudged models better represent cyclones when compared to observations, the presence or absence of clouds is more likely to be closer to observed, however we agree saying "cloud physics" is not quite technically correct so we have modified to "the presence of clouds"

P. 12, L. 24 ~ 25: It is not clear why the authors think so. Please explain more.

The explanation here is derived from Van Wessem et al., 2014 as noted above, we have made this more clear by adding the reference explicitly

P. 12, L. 31: "For the warmer coastal regions": From which data can we see this argument?

We have deleted the word "coastal" as this was a little misleading. See also previous answer.

Figure 2: This figure is a bit difficult to see. Please provide a table indicating ME, RMSE, and R2.

We have added this table as requested. See image:

	a. Shelves (N=112, L=112) Mean obs: 199±132				b. 0 - 1200 m (N=130, L=128) Mean obs: 223±224			
	MB	RMSE	r	rlog	MB	RMSE	r	rlog
COSMO-CLM^2	-85	125	0.75	0.84	-79	174	0.73	0.81
HIRHAM5 0.44°	-37	89	0.79	0.75	-22	143	0.77	0.82
HIRHAM5 0.11°	-59	122	0.60	0.67	-26	194	0.68	0.76
MAR$_{v3.10}$	-12	98	0.69	0.79	-5	159	0.74	0.79
MetUM	-32	83	0.82	0.82	-41	142	0.79	0.84
RACMO2.3p2	-25	90	0.78	0.78	-29	147	0.78	0.87
	c. 1200 - 2200 m (N=158, L=154) Mean obs: 225±240				d. 2200 - 2800 m (N=259, L=258) Mean obs: 89±55			
	MB	RMSE	r	rlog	MB	RMSE	r	rlog
COSMO-CLM^2	-22	187	0.63	0.75	-9	42	0.67	0.61
HIRHAM5 0.44°	33	143	0.89	0.78	-18	45	0.65	0.59
HIRHAM5 0.11°	-19	119	0.89	0.68	-16	46	0.64	0.56
MAR$_{v3.10}$	20	115	0.90	0.79	-14	42	0.70	0.63
MetUM	-16	119	0.87	0.80	-22	46	0.68	0.63
RACMO2.3p2	12	95	0.94	0.77	-13	41	0.68	0.66
	e. 2800 - 3400 m (N=161, L=161) Mean obs: 58±27				f. 3400 m - top (N=103, L=103) Mean obs: 36±12			
	MB	RMSE	r	rlog	MB	RMSE	r	rlog
COSMO-CLM^2	-1	23	0.59	0.61	-1	9	0.70	0.72
HIRHAM5 0.44°	-6	40	0.35	0.53	-12	15	0.72	0.72
HIRHAM5 0.11°	-5	26	0.55	0.62	-9	12	0.72	0.72
MAR$_{v3.10}$	-2	32	0.41	0.54	-1	9	0.67	0.69
MetUM	-10	25	0.59	0.61	-10	14	0.73	0.73
RACMO2.3p2	-2	27	0.46	0.56	0	9	0.70	0.72

	g. All (N=923, L=916) Mean obs: 133±160			
	MB	RMSE	r	rlog
COSMO-CLM^2	-28	113	0.74	0.79
HIRHAM5 0.44°	-9	91	0.85	0.82
HIRHAM5 0.11°	-20	101	0.81	0.79
MAR$_{v3.10}$	-3	88	0.85	0.83
MetUM	-22	82	0.87	0.84
RACMO2.3p2	-9	79	0.88	0.85

Table 2. Comparison of the modelled SMB to the SMB observations over the ice shelves (A), by elevation bins (B-F) and over the whole Antarctic ice sheet (G). Unit of Mean Biases (MB), Root Mean Square error (RMSE), and Mean of the observation is kg/m²yr. N denotes the number of comparison used for each bin while L represents the number of comparison used the log distribution (See the supplementary materials for more details)

18

P. 23, L. 4 ~ 7: What is an interesting point here? I don't think the lower panel of Fig. 8 is necessary; however, if the authors think it is necessary, please discuss more about the figure. Maybe, inter-annual variations of these model results should be discussed more.

We have added more discussion in the section following Figure 8 to make explicit the finding that while all the models have the same anomaly when compared to their own mean, the sign of the anomaly compared to the ERA-Interim value can be different. Since the most highly constrained models show the lowest anomaly compare to ERA-Interim, we suggest that most of the variation is related to internal variability (weather) within the domain.

Technical corrections

P. 2, L. 1: “compar”: typo

P. 2, L. 11: “a potentially important potential contributor” -> “an important potential contributor”?

Fixed

P. 2, L. 22: surface mass balance -> SMB; Note this term is already defined.

Fixed

P. 3, L. 22 ~ 25: This sentence especially after “and better understand drive sea level rise . . .” is a bit difficult to understand. Please reformulate it.

Fixed

P. 5, L. 11: “regional mesoscale model e.g.” -> “regional mesoscale model as presented by e.g.”?

Fixed

P. 7, L. 14: Define SU_{ds} and ER_{ds}.

Fixed

P.9, L. 2: “GrIS”: typo, right?

Fixed

C4P. 10, L. 27: “assessing” -> “assess”

Fixed

P. 12, L. 7: Indicate the publication year for Zentek and Heinemann.

Fixed

P. 12, L. 25: “downwelling longwave and surface albedo” -> “downwelling longwave radiation flux and surface albedo”

Fixed

P. 14, L. 18: “HIRHAM5.011”: typo

Fixed