Referee Comments 1

The authors have used the RACMO2.3 model and a downscaled, long-term dataset to investigate temporal and spatial differences in precipitation and extreme precipitation events (EPEs), and their role on Surface Mass Balance (SMB) at three different ice core sites in Dronning Maud Land, Antarctica.

The paper is well structured and flows logically from each section. The introduction and motivation outline the gap in our knowledge and the importance of investigating precipitation and extreme events. The discussion is also comprehensive and provides a different perspective on the results. However, moving some of the discussion points from the discussion to the results or introduction would provide the reader more trust in the results and the use of the model for assessing the precipitation. In addition, the results could be separated into spatial and temporal distribution to better aid the understanding. I do think major revision is required, as some of the results are not clearly explained or presented. I am also unsure about the use of the multiple datasets and how they compare or validate the observations and each other. I believe the study sheds light on an important topic, and showcases the difficulties of using relatively lower-resolution models to investigate SMB at site-specific observations, especially in a geographically complex region. I think the conclusions are valid and important, but the results need more clarification and justification before it can be published.

We thank the reviewer for taking the time to review the manuscript and for providing constructive and precise feedback. The reviewer's comments and suggestions have been addressed, please see the responses below.

Comments:

Your introduction is very comprehensive, and your motivation is clearly outlined. No comments for the introduction.

Major revision:

Section 2.3: *(1)* More information is required for this product. Add the time period to line 140 - 'Extending the time period covered by RACMO' is not enough for the reader to assess the length of time. Even though you later say the ESM was used from 1850-2014, it is not clear if the downscaled final product also uses this same timeframe. Which version of high-resolution RACMO is used? 2km or 5.5km? What resolution is the downscaled dataset? The reader shouldn't have to read Ghilain et al. 2022 for this information, given that it is important to the study.
(2) The downscaled product uses RACMO data, but there are large differences in the outcome, especially for the EPEs, but this isn't reflected in results or discussion – you should discuss this.

(1) We agree that the downscaling section could be more detailed. We will modify the manuscript to provide a more detailed description of the dataset: "Considering the strong internal variability of the Antarctic climate system (Jones et al., 2016), the satellite period might be too short to study the variability of the SMB. An opportunity to analyze longer

periods rises from statistical downscaling which extends the time period covered by the RACMO simulation. Here, we use the dataset from Ghilain et al. (2022a) that employed a statistical downscaling method combining the daily snowfall from RCM of high spatial resolution to specific weather patterns observed in state-of-the-art atmospheric reanalyses. Based on these dynamical relationships, Ghilain et al. (2022a) applied this relationship on ensemble of simulations performed with a Earth System Model covering the 1850–2014 period to downscale precipitation over the coastal region of DML at high resolution (5.5km; see for details and specifications). Briefly, this dataset provides daily snowfall by using the RACMO2.3 simulation at 5.5 km horizontal resolution, ERA5 as the atmospheric reanalysis and an ensemble of 10 simulations performed with the Community Earth System Model version 2 (CESM2) at low horizontal resolution (1 degree; about 70 km). The resulting daily snowfall simulations combine the advantages of RCM and GCM since they are characterized by the high spatial resolution of RACMO2.3 and the longer record (1850-2014) of CESM2."

(2) Thank you for highlighting this missing point. As detailed in one of the following comments, we will reformat the manuscript to better analyze the differences between the datasets (from models but also the ice core records) and focus more on the downscaling outputs in the results and discussion. We will also address the potential limitations of the downscaled dataset.

Which variables are you using? Snowfall or precipitation? Are these synonymous or comparable between the downscaled dataset and RACMO? Does total precipitation = snowfall in this region, or are their times of rainfall in summer?

We use precipitation (snowfall + rain) from RACMO2.3 and snowfall from the downscaled dataset. Differences between precipitation and snowfall are negligible - of the order of 0.1 mm yr^{-1} (0.14, 0.13, and 0.07 mm yr^{-1} for IC, FK, and TIR, respectively, while the average precipitation per year are 431.3, 497.0, and 360.19 mm yr^{-1}) in RACMO2.3. To make it simpler and avoid confusion, we will mention this negligible difference and will then use a generic term (precipitation) throughout the manuscript.

Section 3: There are major differences between the simulated SMB by RACMO and the ice core records. This is the first thing reported and then makes it very difficult for the reader to trust that RACMO is going to be used for the rest of the study. The justification for using it comes in the discussion, but you should consider moving this earlier, and perhaps bolstering this justification further. Almost 50% of the ice core SMB is not represented in the model – if precipitation is the main component, are you convinced that the model is representing the precipitation properly? Whilst models are always wrong, there is additional model justification and testing which is presented in the discussion which could perhaps come early in the results to bolster the reason for continuing to use RACMO despite the consistent, large underestimation. It would be ideal to see more comparison of the key variables such as precipitation with other observations.

The discussion section regarding complexity of SMB in ice cores should perhaps be moved to the introduction or results. In my case, I am very familiar with RACMO and the SMB analysis in the polar regions, but haven't used ice cores as observations before. Therefore, a straight comparison of the ice core and the model seems like a bad idea, given how poorly

RACMO captures the ice core observations. However, I do see value in continuing the study to analyse the RACMO data and assume that it can be a tool for showing SMB differences in time/space.

We understand that the differences between the ice core records and the model's outputs can be confusing and lower confidence in the results. This should indeed be addressed earlier in the manuscript. We will start the Results with a section dedicated to the comparison of the SMB from the ice core observations and from the models. From this point, we will reformat the manuscript to better highlight the logic we have followed. Here is a proposition of structure:

3 Results

3.1 SMB: comparison between the ice cores and the models

→ this section will show the time-series of precipitation from RACMO2.3 and the downscaled dataset, as well as SMB from ice core records (including adapted Figures) and explain the possible reasons and uncertainties behind the differences (both for the ice core observations and for the models). Note that we don't have the SMB data for the downscaling results (only the snowfall), this will be addressed.

3.2. Contribution of the SMB components in RACMO2.3

→ this section will show that, according to RACMO2.3, precipitation is the main driver of SMB. By relying on RACMO2.3, we can thus make the hypothesis that the SMB variability observed in ice cores is explained by the precipitation. RACMO, recognized as an excellent tool for modelling precipitation in polar regions, will therefore be used to study/evaluate the spatial variability of precipitation over Dronning Maud Land, in order to shed light on the large spatial variability observed from ice core records. In this section, we thus make the hypothesis that RACMO2.3 performs well enough to investigate the potential processes explaining the large spatial variability of precipitation. However, since RACMO spans a short period of time, we might under-sample the impact of internal variability on precipitation over Dronning Maud land. Therefore, we need to obtain longer time-series. We introduce the downscaled dataset from Ghilain et al. (2022).

3.3 Precipitation: spatial variability and temporal trends

→ here, we will merge the two sub-sections (3.2.1 and 3.2.2) and will rework the text to better show the main results (distribution of the precipitation days, trends, and correlation between the time series), including those of the downscaled dataset.

3.4 Extreme Precipitation Events

→ in this section, we will detail how we calculated EPE thresholds for both RACMO and downscaling datasets. More emphasis will be given on the downscaling results, and how they differ from RACMO outputs, as this downscaled dataset can be used to "increase the sample size" of EPEs. Potential bias towards higher precipitation of the downscaling technique, potential lack of spatial variability(see the reply to next comment) as well as

issues in the downscaling method of Ghilain et al. (2022a) will also be mentioned. Synoptic patterns during EPEs will remain in this section, with details on the use of ERA5.

3.5 Frequency distributions of precipitation anomalies

→ we understand this section needs to be clarified. More explanations will be given on the method and purpose of looking at the frequency distributions.

The discussion and conclusion will be adapted accordingly.

With the downscaled data, it isn't even capturing the long-term trends found in ice cores (section 3.2.2), which makes it even more challenging to justify using. If it isn't capturing long-term trends, which typically models can capture, how do you know it is capturing any spatial variability?

Ghilain et al. (2022) observed a general good agreement between the downscaled snowfall (precipitation) and eight ice cores located in the DML coastal region. This will be mentioned, as well as the clear indication that the downscaled dataset is not capturing the long-term trends of our records (in the section comparing SMB from ice cores and models).

Looking at the differences between EPEs from RACMO and the downscaling, we made additional calculations to obtain the average number of EPEs per year in both datasets and it appears that there is no clear spatial variability (in terms of number of events per year), as shown on the following table:

	Downscaling (1850-2014)			RACMO (1979-2016)		
EPE 95	IC	FK	TIR	IC	FK	TIR
# EPEs	2656	2599	2590	426	374	370
avg # EPEs/year	16.09	15.75	15.70	11.21	9.84	9.74
EPE 98						
# EPEs	1062	1040	1036	171	149	148
avg # EPEs/year	6.44	6.30	6.28	4.50	3.92	3.89

This table shows that the downscaled product might be biased towards higher precipitations, thus overexpressing those, because of the way the downscaling is built - i.e., by focusing on the precipitation events. However, this does not mean that downscaling is significantly less effective than RACMO at capturing spatial variability.These important observations will be added to the manuscript.

Ghilain et al. (2022) highlight an important bias reduction in comparison to CESM2 without downscaling. This shows that, despite the existing room for improvements, the downscaled dataset is the best tool available to increase the sample size and extend the record.

The authors give a short analysis of the blowing snow contribution – which is significant, but then it is not investigated further or mentioned in relation to the spatial differences in SMB between the three sites. Perhaps more emphasis should be given to this investigation, especially as you conclude that precipitation/extreme precipitation is not responsible for the spatial differences between the sites. If there is a blowing snow modified version of RACMO, could you investigate (briefly) the differences in the model output with and without this modification?

We agree that the blowing snow contribution is a solid hypothesis to explain the models-observations discrepancies. It is presented along other hypotheses (small-scale processes, influence of the ice rise dynamics), which will be further detailed in the revised version of the manuscript. We would prefer to keep the focus of this paper on the precipitation/EPE and present the hypotheses not linked to precipitation as perspective for future work.

Figure 3: It would be useful to change this to better allow the reader to compare the products and the locations. Figure 3 is busy and it is too hard to compare SMB from RACMO and observations, this could be a separate figure to the other components. In addition, it is hard to compare the locations when there's a lot happening in one figure. Similarly, it would be useful to get one figure where ice cores, RACMO and downscaled data are presented together. Statements like 'in contract with ice core records' (Line 203) are hard to check in the figures, when long-term SMB from ice cores is on figure 2, satellite-era RACMO is on figure 3 and long-term downscaled data is on figure 4.

As detailed in a previous comment, the structure of the Results will be modified and the Figure 3 will be changed to a figure with the SMB from observations and models and a figure presenting the SMB components in RACMO. We will keep your suggestions in mind when doing the new figures.

Section 3.3: More information is needed on how you calculate the thresholds – does each location have its own 95% threshold – e.g each grid cell which represents the ice core location has a value? Or are you using an area average for all locations? Are the thresholds re-calculated for the downscaled data? Later on, you use ERA5 too – are you calculating the thresholds again for ERA5 data, or simply using ERA5 to extra data on specific dates, which have been above the threshold from RACMO?

Thank you for raising these questions, we understand that we need to elaborate more on these, we will give more details on how each EPE threshold is calculated. Briefly, each location and each dataset has its own threshold calculation. Regarding using ERA5, we use it as an additional data input, keeping the threshold from RACMO. This is discussed more in the next comment.

Section 3.3.2: I think it is a good idea to look at the synoptic situation during these events, but I question the addition of ERA5 data (also because it is not listed in your data section). This is another dataset, which is not compared to RACMO, the downscaled data or ice cores. Whilst synoptic conditions are generally well captured in ERA5 and most models, you are looking at specific dates of these events. How do you know that ERA5 also captures the EPEs which RACMO is seeing? Precipitation is a difficult variable, even in higher resolution reanalysis products. Perhaps it is better to look at the synoptic conditions in RACMO, rather than introduce an additional dataset and therefore additional uncertainties in the conclusions. If you do stick with ERA5 – there should be some discussion of its useability in this region and how it compares to RACMO. In addition, do you select the data from ERA5 based on dates in RACMO, or do you re-calculate the 95% threshold with ERA5 data?

Thank you for identifying the missing information regarding ERA5 data in the data section, we will add it.
We indeed use ERA5 as an additional dataset allowing us to investigate the synoptic conditions during EPEs because the domain of RACMO2.3 used in the study is too small. We assume that ERA5 capturessimilar EPEs as RACMO, since RACMO uses the previous version of the ECMWF reanalysis (ERA-Interim) for the boundary conditions and the domain is not very large, so the atmospheric dynamics should be relatively similar. We will add this to the description of how the synoptic patterns are obtained.

Section 3.4: The geographic and synoptic set up doesn't particularly align with the characteristics of foehn winds. With the Peninsula, a long, high ridge prevents the air from flowing around the obstacle and therefore forces it over 2000m– this is what creates the foehn winds. However, in the case of ice rises, the airflow could flow around the obstacles, given their size, and likely not create the warm and dry leeside conditions. The definition of a foehn wind is also the warm and dry lee slope winds, and not the reduction of precipitation down wind. Instead, you're perhaps referring to orographic precipitation characteristics, such as rain shadow. I wouldn't introduce the foehn effect here, as it doesn't really apply.

Thank you for raising this point out, we will modify the text to orographic precipitation.

Section 3.4.1: Can you find a different name for the EPEs in this section? Up until now, you have defined EPEs as extreme precipitation events with a 95% or 98% threshold for the value of extreme. However, in this section, EPEs now mean 'percentage of cases where the other sites receive more precipitation than their site-specific EPE threshold'. It is then confusing to try and interpret the results – especially the relative wet/dry of the locations. However, this definition does answer a question I had earlier about whether thresholds were site-specific. With the current definition, I am left confused about whether IC is drier or wetter than other sites during EPEs.

We understand that this is confusing, so we will replace the term "EPEs" in the table and text with "Above EPE threshold". We will also better explain what is done in this section so that the results are easier to interpret. We will rework the text to better highlight how we obtain the frequency distributions of precipitation anomalies and how the classification is made (i.e., that the "above EPE threshold" events correspond to EPEs too, meaning that EPEs are simultaneously occurring at the two sites).

Whilst the differences in EPEs and negative anomalies per location from RACMO (table 4) seem significant, the differences between the sites in the downscaled data seems negligible or insignificant. Have you run any statistical tests on these results? The neg.anom for EPEs at IC and EPEs at TIR are very similar, and the two data sets do not agree with each other. The results here focus on the RACMO set, but you don't discuss the lack of consensus among the datasets. This could be because the downscaled data includes a longer time period, but as stated in your earlier results, there is no long-term trend in the data for precipitation, SMB or EPEs, so this perhaps doesn't answer it. I am really not convinced with section 3.4.1 I understand your hypothesis and perhaps the method of trying to look at it, but the results are quite confusing.

We understand this section is confusing, we will rewrite it to make it easier to understand. As stated previously, the downscaled data will be more discussed in the reworked version of the manuscript. We will include a discussion on the similarities and differences between the RACMO and downscaling outputs.

Section 4: This first paragraph about ice core complexity should go in the introduction – throughout the results, I am concerned with how RACMO is representing observations, but this section gives me pause about the ice cores as observations. This level of complexity regarding SMB from ice cores should come earlier, especially for readers who are not experts in ice core interpretation.

As said previously, we agree that this should come earlier in the manuscript. A brief overview of the SMB variability in ice core records is already given in the introduction (lines 70-74), but this will be further detailed in the revised version of the manuscript with the section dedicated to the comparison of the different SMB.

If 2.2km RACMO was found to be more representative of the ice cores than 5.5km RACMO, why not use the higher resolution one? Or is the 2.2km RACMO the downscaled product you have used?

Regarding the comparison with the ice core SMBs, the 2 km scale RACMO is closer than the one of the 5.5 km scale for IC and TIR, but not for FK. However, as the downscaling dataset is based on the 5.5 km scale, we made the choice to use this version throughout the paper, to analyse "comparable items". Moreover, the 5.5 km scale RACMO has been well evaluated in our region of interest (i.e., Lenaerts et al., 2014; Kausch et al., 2020; Ghilain et al., 2022) and we had easy access to the daily datasets because of previous projects we were involved in while the 2 km RACMO dataset has been only publicly archived at annual resolution. In addition, the 2 km RACMO results from the statistical downscaling of the 27 km scale RACMO with a high resolution surface topography at 2 km. RACMO 5.5 is thus expected to better represent the atmospheric and surface dynamics.The comparison we made between the 5.5 and 2 km scales aimed to recognize that the resolution used to analyze RACMO has a potential impact on the results. However, we realize that mentioning the 2 km RACMO dataset adds confusion in the manuscript. Instead of conducting the comparison in the discussion, we will propose the use of models with a finer resolution (both for RACMO and downscaling) as well as enhanced representation of processes, in particular processes related to the blowing snow, as a perspective to develop in the future.

Minor:

Section 2.1: Can you provide the elevation of the ice rises – this becomes fairly important for your discussion on foehn winds and the loss of moisture across the trajectory.

This is a good point, it will be added.
The IC12 core was drilled in December 2012 at the crest of DIR (−70.24218 °S, 26.34162 °E, ~429 m ASL*) and is 120 m long. The FK17 core was drilled during the 2017/2018 austral summer at the crest of LIR (-70.53648° S, 24.07036° E,* ~333 m ASL*) and is 208 m long.*

The TIR18 core was drilled during the 2018/2019 austral summer at the crest of HIR (-70.49960° S, 21.88017° E, ~348 m ASL*) and is 262 m long.*

Line 165 and 168 say the same thing.

The first sentence of line 165 will be deleted.

Section 3.2.1 – is this really interannual variability section, or is it more spatial variability? Apart from the first line, the rest of this section is about the different locations.

Thank you for pointing that out. We will merge the two sections and keep the section 3.2 title.

Line 182: What do you mean by opposing signals in ice core records? Is this figure 2? Apart from TIR which has a decreasing trend, they don't seem to have opposing signals. This is hard to tell in figure 3 too.

We will reference the text to Fig. 2. Opposite is indeed an inappropriate word. We will replace it by contrasting or different, as the three ice core records indicate different behaviors/trends.

Table 3: caption says it is contribution to the total annual precipitation, which you also confirm in line 252, however in line 259 you say that EPE variance accounts for 2/3 of the SMB variance. So is the variance SMB and the average contribution annual precipitation? Different variables are used between RACMO (annual precipitation) and downscaled data (snowfall) – are they comparable?

The caption of Table 3 mentions both contribution and variance, which are two distinct concepts. However, we will define variance, as a proper definition is lacking in the text. Regarding the different variables between RACMO and downscaled data, we addressed the question above.

Line 327: I don't understand this sentence – where are the observed global atmospheric pathways observed?

We meant that from Figure 6, it seems that the synoptic conditions are similar during EPEs at each of the three sites. But it also appears that some 95th EPEs at IC result in negative anomalies at the two other sites, thus we want to see if these correspond to particular synoptic patterns or not. We will change the text to make it clearer.

Figure 8: ERA5 data?

Indeed, these are ERA5 data, we will mention it.

Line 356: change 'excludes' to 'rejects'.

We agree.

Line 438-439: Perhaps include that this is a conclusion from a model, not from observations.

Thank you for pointing that out, it is important to remind readers that this comes from a model.