We thank the reviewers for their constructive critic. We have responded to all *comments* (displayed in *italic*) and adjusted our manuscript accordingly.

In summary we:

- run more simulations and added another section that investigates the importance of resolving extremes.

- We discuss more our dynamics runs without the surface elevation feedback.

-moved some figures from the SI to the main part to improve readability

-identify more clearly the difference of our experiment in comparison to the one of Delhasse et al. and compare to them in the discussion part as they have included extremes as well,

- discuss in more detail our results and add the role of the surface elevation feedback

- rephrase our conclusion and hope to that the discussion and conclusion are readable to a broader audience.

Reviewer1

> *Review of "Effects of extreme melt events on ice flow and sea level rise of the Greenland Ice Sheet" by Beckmann and Winkelmann*
>
> *The authors present a set of ice-sheet model simulations to 2300 that explore the impact of extreme events of varying frequency and intensity relative to a simulation with a baseline climate forcing. They construct the baseline temperature forcing using regional climate model estimates of Greenland surface temperature and an emulated global mean temperature time series. They then calibrate a positive degree day model for this temperature time series in order to attain surface mass balance forcing to 2300. They construct nine scenarios for combinations of periods of 5, 10, and 20 years and relative intensities of 1.25, 1.5 and 2, relative to the running decadal mean temperature in the baseline forcing. Running the ice sheet model under these forcing scenarios, they find that including extreme warm events can have a significant impact on long-term mass loss for events with high frequency and intensity. They find a 14% increase relative to the baseline scenario for the most extreme scenario that includes ice dynamics and SMB-elevation feedback. For a case that considers only surface mass balance (i.e., neglecting ice dynamics changes), they find a 16% increase in mass loss.*

Overall, I think the study is interesting and the paper is well written. I have two primary concerns with the paper, however, which cause me to recommend major revisions.

The first concern is that the initial condition is not a good representation of the present day ice sheet, with many major outlet glaciers over- or under-estimating observed velocities by a wide margin. It is not possible to discern from the figures just how far from the modern state the initial condition is, but some major outlets seem to differ from observed velocities by >100% (estimating from Fig S2 by eye). Some misfit statistics are reported in the text, but these are skewed by the very large slow-flowing part of the ice sheet and so the reported average misfit of 9 m/yr is not terribly relevant. It is unclear why this initial condition was used, when there are other PISM initial conditions that look much closer to observations. The present initial condition makes it difficult to interpret the results, as I would not expect this model configuration to respond to external forcing in the same way as a configuration that is closer to observations. I recommend improving the initial condition (possibly just using one that has been published) and repeating these simulations and analysis, or at least somehow demonstrating that the initial condition does not significantly bias the results relative to a more realistic initial condition.

Our project aims to investigate a first order approximation on the effects of extremes on the overall mass balance of the Greenland Ice Sheet. Therefore, our model calibration was focused on achieving an overall sea level rise potential close to observation rather than simulating each glacier correctly. While glacier speed can be very well represented by ice sheet models, these models usually use inversion as their initialization method, which is not a feature of PISM. The initialization method of PISM is done via a spin-up over a glacier cycle and temperature anomalies of a scalar temperature field. This procedure makes it more challenging to simulate each glacier correctly, especially when using the positive degree day model as the surface model. From this initialization method and surface model, biases in the ice thickness (Fig. S1,Fig.5c) arise, that in turn influence the surface velocity though differences in thickness and surface slope (driving stress). However, another influence on the velocity patterns of the GrIS stems from the resolution applied in PISM. As the reviewer correctly pointed out, it is possible to simulate glacier velocities close to present-day with PISM, demonstrated by Aschwanden et al 2016. However, there the authors use a flux correction that artificially adds or reduces the ice thickness to match closely the observed surface topology. They also show that the good match to observed velocity is only achieved when using resolution of 600m. Using this resolution would not allow us to run an ensemble of projections until 2300 due to high computational costs. Even when using 600m Aschwanden et al. 2016 show, that especially Humboldt glacier shows the poorest fit in their study and reasons for that would have to be investigated in another study. In their paper and supporting information the authors depict observed and modelled velocity profiles for different resolutions. For a resolution of 4.5 km, velocity difference in the order of several 100 m/a are common for almost all glacier profiles even with flux correction. Similar published initial states with PISM to ours are published under Zeitz et al, 2021 and 2022 ,We therefore do not think that another spin-up is necessary or could lead to a more realistic velocity representation with this model configuration (glacial spinup, scalar temperature anomalies, pdd surface model ,resolution).

As a second constraint we show that the overall mass balance historical change form 1979-2017 is reasonably well captured with our model simulations (Fig. S3)

> *My second concern is that the experimental design is to essentially add extra temperature forcing, which leads to the conclusion that extreme events are important. But applying these extreme events to the baseline temperature forcing time series results in a stronger average temperature forcing than the baseline. Thus, there is no way to determine how much of the excess mass loss is really due to the extreme events and how much is due to this increase in average temperature forcing. It seems that the proper methodology would be to ensure that the baseline and extreme scenarios have the same mean temperature forcing over some long-term average (probably a few decades to a century). This would much more clearly show the impact of variability vs mean forcing.*

Yes absolutely. We thank the reviewer for pointing this out and added a whole section in the result part to investigate this.

> *The question arises whether the additional temperature excess inserted by the extreme is important to resolve on this monthly time step or whether the evenly distribution of this excess temperature would lead to the same impact in terms of SLR.*

We therefore averaged the original temperature forcing of a monthly resolution over different time frames and asses their contribution to SLR (Tab 4). The average temperatures were recalculated to their monthly equivalent to produce a data set with a monthly output. This allows us to keep the same experimental setting of a monthly time step for PISM as in our original experiment and exclude any numerical biases that might be introduced by e.g. changing the time step of PISM. We concentrated on investigating the most severe scenario (I_2,f_5). All results can be found in section 3.5 Resolving extremes

I also found the Discussion section to be rather limited in scope. I have added a few suggestions below of topics to enhance the Discussion. A number of more specific edits, comments, and questions that are also listed below. There are a number of mis-referenced figures, especially in the very long supplement, so that should be checked carefully during revision.

Agreed, we added several discussion points and carefully checked the references to the SI.

Specific comments:

L 73: "Consider changes in ocean melt or sliding due subglacial or subglacial processes"

We explain and discuss the setup more in the methods part. We therefore write now: "We do not consider changes in submarine melt rates or subglacial processes as e.g. glacial channel building that could in turn influence the basal sliding (Methods). As of yet, these processes are not well understood and require different experimental setup (Methods) that would go beyond of the scope of this study (Discussion)."

L 88: submarine melting is kept constant, but how is it calculated, or what dataset is used? Is the melt rate constant for each glacier for all time, or does each cell have an associated melt rate that is applied when the glacier terminus is in that cell? Are there different treatments for floating and grounded ice? Please provide more information about this.

Is there any calving law or criterion used here?

We thank the reviewer for spotting this lack of clarity. We now add calving front of glaciers are not allowed to advance. We also do not allow for floating ice thinner than 50 m at the calving front and use the von Mise calving law, appropriate for glaciers in Greenland" … …."submarine melting is kept constant in time and space with a melt rate of 0.051914 m/year (PISM default setting).

L 123–124: By this logic, Humboldt Glacier should have a fairly good match to observations because its width is large compared with the model resolution. However, the fit is very poor there.

Indeed, Humboldt Glacier has be shown do achieve only poor results when modelled with PISM even for very high resolution of 600m that leads to a good match between observation and simulation for other glaciers as (Aschwanden et al. 2016) demonstrate.

L 125: There should be similar statistics reported for just the fast-flowing part of the ice-sheet (e.g., where speed > 100 m/yr or some other reasonable threshold), where the velocity and thickness are much more relevant than over the ice-sheet interior.

Agreed, the Fig. S2 caption now states as RMSE "of 413 m/yr for regions flowing faster than 100m/yr." which is also now mentioned in the text.

Specify which version of BedMachine is being used. Presumably v3? Citing the paper rather than the dataset (ie., the NSIDC page) is a bit ambiguous because the Morlighem et al (2017) should be cited when using v4 or v5 as well.

Agreed, we now cite both, the dataset, and the paper, as the title of the paper specifies version 3.

Fig 1B: Subplot title missing a letter?

Changed.

Figure S1: please add a panel showing thickness misfit as a fraction of observed (BedMachine) ice thickness

Done.

Figure S2: Color bars on all plots are too narrow, resulting in very large areas of saturated color that make it impossible to judge the fit to the observations. Please use wider color bars; 10^4 m/yr would be a more reasonable upper limit for panels (a) and (b). Also consider using a signed log-scale (e.g., -10^3, -10^2, ..., 10^2, 10^3) for panel c to aid with visualization. There should also be a plot that shows the misfit as a percentage of observed velocity. Some of the velocities at these large outlets (notably Humboldt, NEGIS, most of the NW sector, and potentially others, but hard to tell on this color scale) are very far from the observed velocities, which will significantly bias model results in these regions. This makes interpreting the results rather difficult, as the modeled ice-sheet state is quite far from the true modern state.

We adjusted the figure and added a panel showing the relative error. We agree that errors are large for some glaciers, but again looking at the best fits with 600m resolution and flux correction in Aschwanden et al. (2016) shows that the mentioned regions by the reviewer are the typical ones that are rather challenging to model.

L 138–140 and Fig S3: The agreement between the modeled and observed mass balance from 1972–2017 seems overstated to me, given that the slope of the observed mass balance is almost twice the slope of the modeled mass balance from 2000–2017.

We added ",however not capturing the strong slope from 2000-2017."

L 157–162: Difficult to understand. I don't understand how the anomaly years contain the monthly anomalies, for instance. Please revise these lines for clarity.

We rewrote this section and hope it is more clearly now:

To this end, from the annual MIROC5 results until 2100 we first derived a quadratic trend function ($T_{GrIS,trend} = 1280.16 ◦C − 1.31 ◦C/year · years + 20 ◦C/year^{−2} · years^2$) to exclude the inter annual variability (Fig. S4a). Together with the GMT until 2100 we determined a fitting function (Fig. S4b) $T_{GrIS,emulated\ trend} = 0.1 ◦C + 0.96 ◦C^{−1} · GMT + 0.15 ◦C^{−2} · GMT^2$,in order to emulate $T_{GrIS,trend}$ beyond the year 2100 in dependence of the GMT. Thus, with the GMT until 2300 and the fitting function, we established the GrIS trend function until 2300 ($T_{GrIS,emulated\ trend}$, red dashed line). To this, we added the inter- and intra- annual variability to receive a more realistic monthly temperature projection. This was done by first calculating the yearly anomalies of the 2050-2100 from the fitting function to the actual annual values: $ΔT_{yr}(2050−2100) = T_{GrIS}(2050−2100) − T_{GrIS}(2050−2100),trend$ and then randomly picking out of $ΔT_{yr}(2050−2100)$ and adding on to the emulated trend $T_{GrIS,emulated\ trend}$ until the year 2300. This gave us the new annual temperature curve until the year 2300 with inter-annual variability ($T_{GrIS,emulated}$ red solid line, Fig.S 4a). However, as we need monthly temperatures we recalculated the monthly temperature values for our newly created $T_{GrIS,emulated}$ by adding the monthly temperature anomalies as well. Thus, for each annual anomaly $ΔT_{yr}(x)$ we picked, we calculated its monthly anomalies by collecting the full 12 months of that year and subtracting them by the mean of the annual trend (and not the annual mean) ($ΔT_x(1,...,12) = T_{GrIS}(x(1,..,12)) − T_{GrIS}(x),trend$). Thus each annual anomaly contains now monthly anomalies as well $ΔT_x = ΔT_x(1,...,12)$ that were added to the emulated trend. These values served as our baseline scenario until the year 2300 (Fig. 2, dark grey lines).

L 164: Is Figure S9 the correct figure to reference here? I don't see how it relates to the text here. Seems like it should be Figure S5

Yes correct, we thank the reviewer for spotting this mistake and changed the Figure number.

L 181: Should be I1.5f5?

No, we are here referring to the intensity factor in general.

Figure S10: There is only one tick on the vertical axis here, which makes it impossible to determine the vertical scale.

We added a y axis.

Section S2.1: " Figures S5 and S6 show that the extremes would increase…" These don't look like the correct figures. Should be S8 and S9? Also, the brown curve is not defined in S8 and S9.

Correct, we changed the Figure number again. There is no brown curve on its own. This brown colour stems from overlaying the black and the yellow curve, which both are defined.

Figure S12: Why are the two MAR curves here so different over most of the century? I don't fully understand what is meant by: "SLR from the original MAR data set (Miroc5) of 1km resolution was derived from the â SMB", so perhaps that can be phrased more clearly, with a reference to another figure if relevant.

When calculating the SLR from SMB loss only the mass loss from volumes above floatation must be calculated. Our experiments were run on the 4.5km gird on which we interpolated the MAR SMB original data set that had a resolution of 1km. Thus, for the 4.5km SMB changes from MAR we can estimate the true loss of volume of floatation with our bedrock data set. This was not done for the 1km data set. But to give an estimate of SLR from the original data set we estimated it by calculating the SMB loss over the entire ice volume (also including floating ice cells.)

We write now in the caption:

"SLR from the original MAR data set (Miroc5) of 1km resolution was estimated by the cumulative changes in SMB over the entire ice sheet (also floating cells). The original data set was interpolated to our 4.5km grid as MAR (4.5km). where SLR was calculated by the cumulative changes of surface mass loss over over the volume above flotation in order to compare it correctly to the PISM PDD simulations."

L 205: It would be helpful to remind the reader in this sentence of what the scenarios are.

Agreed, we add now a sentence before:" We derived a set of 10 different temperature forcing scenarios, that include a MIROC5 baseline scenario and its 9 versions of extremes. These extremes differ by three intensities (I1.25, I1.5, I2) with each having 3 different frequencies (5, 10 and 20 years).

Figure 2 and in general: It seems strange that only extreme warm events are included in these scenarios, rather than including both extreme cold and extreme warm events. By including only warm events, you've essentially just increased the decadal (or multidecadal to centennial) average temperature by a few degrees, which will of course lead to correspondingly more mass loss. It seems that the proper comparison would be to make temperature time-series that have the same multidecadal average, so that the impact of variability is actually quantified, rather than to add extra temperature forcing to a baseline temperature time series as is done here.

The AMAP of 2021 showed, that extremes have increased with a bias towards the higher temperatures whereas cold temperature extremes have not increased. We therefore concentrate on looking into increases in temperatures. However, to demonstrate the necessary of resolving these extremes we added a new section that includes runs with equal mean temperatures but disregarding the climate variability. (Section 3.5)

> *Figure S13: The vertical axis label should be dST/dz, correct?*

Correct! We adjusted the axis.

> *L 210: In the SMB-only experiments, does ice thickness change due to SMB? Or is ice thickness held constant in time? Or is advection active, but velocity is held constant? Please add a bit more detail about this set of experiments.*

We added " and thickness was held constant. Thus, SMB changes were calculated with PISM's internal PDD model for a constant surface topography but with changing temperatures."

> *Figure S14: Text seems to reference something that isn't present in the figure: "the corresponding ice sheet extent in 1971 (i) and the emerging ice retreat in years 2100 (ii), 2200 (iii) and 2300 (iv) are given in light blue and red shading, respectively."*

True, we deleted that part.

> *L ~245: Is this shown in a figure or table anywhere?*

We added a new Figure now in the SI and refer to it in the text.

> *L267: typo: Mirco5*

Changed.

L 300: Could it also be that CW is the only one that continues to accelerate because Jakobshavn remains a marine-terminating outlet, and that's not the case for most other large outlets? From Fig1, it looks like the only other outlets that remain in contact with the ocean are Petermann, maybe NEGIS, and maybe Humboldt.

We thank the reviewer for this insightful idea. Indeed, most of the other sectors become land terminating and thus velocities decrease. Remaining in contact with the ocean clearly hinders a slowdown for this sector, We added this to the part"

The CW-sector is the only sector in which average surface velocities continue to increase until 2300 (Fig.~S23) with a large part of the glaciers remaining in contact to the ocean, keeping a reduced basal friction and therefore high basal velocities (Fig SX) while other sectors become land terminating.

What basal friction law is used here? I see that you use an exponent of 0.6, but what is the form of the law? That could have an effect on the slow-down you observe while driving stress decreases.

We rewrote this passage now as the following:" The basal sliding velocities are related to basal shear stress via a pseudo-plastic power-law with a power of q and the yield
stress. The yield stress in turn follows the Mohr–Coulomb criterion, and is determined by models of till material property (the till friction angle) and by the effective pressure on the saturated till. We linearly altered the friction angle between 5◦ and 40◦between -700m and 700m of bedrock elevation after Aschwanden et al. (2016). The resulting lower friction for lower altitudes
and below sea level leads to an additional increase in surface velocities at the ice sheet margins, resulting in an improved match
of flow structure for the glaciers."

The Discussion section is very short and the Conclusions section reads like it should be in the Discussion. Consider expanding the Discussion and including more of a summary of your findings in the Conclusions. Particularly, the discussion should touch on the impact of the initial ice sheet state on these results, as the spun-up initial condition is quite far from the observed modern ice sheet state (Fig S2). This initial condition should be compared with other model initial conditions for Greenland, at least with the initial condition from Aschwanden et al. (2019).

We agree with the reviewer here and added a summary of our findings to the conclusion. Now the Conclusion reads:"

Previous studies did not consider extreme melt events when projecting sea-level rise (SLR), and predictions of weather and climate extremes are generally accompanied by high uncertainty (Otto, 2016, 2019). Hence, our idealized experiments offer an initial assessment of how future extreme events may impact the Greenland Ice Sheet, highlighting the significance of incorporating extremes in future SLR projections. It is essential to take into account both the intensity and frequency of extreme events in future projections. Comparatively, our most severe scenario, in contrast to a scenario without additional extremes, could result in an additional SLR of approximately half a meter or 14% within a fully dynamic ice sheet experiment. To accurately capture the effects of extremes, it is crucial to account for monthly temperature extremes. Our experiments demonstrated that mass loss is primarily driven by surface melting, which is significantly amplified by the surface elevation feedback. Incorporating this feedback is critical for precise projections of future sea-level rise. Building upon an SLR estimate derived from a non-dynamic surface mass balance-only model, our experiments indicate that accounting for the dynamic surface elevation feedback would contribute roughly 35% to the sea-level rise, and the most severe extreme scenarios would add an additional 20%.

Another topic to touch on in the Discussion is that temperature extremes will in reality increase the flux of meltwater to the bed and thus affect ice dynamics through subglacial hydrology, which is not accounted for in these simulations

Yes we added this to the discussion:

"The increased surface runoff during the next centuries can clearly influence the subglacial hydrology via the formation of subglacial channels which in turn can influence basal sliding. Our experiments due not consider any changes of this nature, many of these processes are not fully understood and their long term effect is still unclear (Shannon et al., 2013; Tedstone et al., 2015)"

Finally, some discussion of the full dynamics runs vs the runs without SMB-elevation feedback would be good.

There is no equivalent of Fig 4 given for the dynamic case without SMB-elevation feedback. Overall, it seems like those runs were ignored compared with the SMB-only and full dynamics cases. There should be another subsection analogous to 3.2 in which the full dynamics and no-feedback runs are compared in more detail.

We wanted the reader to concentrate more on the SMB only and the full dynamic runs because these are the experiments typically run in the community. As the SMB only scenario is close to a setting in which the regional model MAR would simulate SMB loss and common ice sheet models include lapse rate correction when calculating with a pdd model. However, we now add a few more lines in the result part that discuss more the dynamic runs without surface elevation feedback and also added Figure visualizing our analysis.

We added in the results:

"The dynamic experiments that do not consider the surface elevation feedback lie in between our SLR projections for the SMB-only and full dynamic experiment but are only slightly higher than the SMB-only runs (about 1 % higher). For this constellation, the dynamic mass loss adds only a little more SLR than the pure SMB loss. This clearly shows the importance of the surface elevation feedback, i.e. it is important to include the effects of both the surface-mass balance and the ice dynamics in sea-level projections in an interactive manner"

....

"The role of the surface elevation feedback for these simulations can be estimated when comparing the full dynamics runs with the dynamic runs without surface elevation feedback in Figure S25 for the baseline scenario. In the year 2100, thinning is more prominent along the margins and the southwest of the GrIS for the full dynamic runs. The thinned ice cells decrease in surface velocities, but further inland surface velocities speed up due to the steepening gradient. This effect is even further amplified in 2300: Thinning is more and more amplified, reaching into the ice interior of the GrIS leading to further retreat. Steepening the interior leads to speedup while thinning at the margins reduces surface velocities. Dynamic runs without the surface elevation loose only about a quarter of the present day ice sheet area (440.3 $10^3 km^2$ Tab S3). Additional retreat due to extremes is about 4.5,9 and 18 ·103km2 for I1.5 with frequencies of 20,10 and 5 years respectively (Tab S4). Here as well, we see roughly a doubling and halving of the additional mass loss for the other intensities (I2 and I1.25 respectively). Likewise, the surface elevation feedback shows a bigger effect on ice sheet retreat than the additional extremes"

I have rated Presentation Quality as "Fair" because there I think the manuscript relies too heavily on the numerous figures in the Supplement, while there are only a few figures in the main paper.

We moved some Figures from the SI to the main text.