Response to reviews

We would like to thank both reviewers for their thoughtful comments on our paper. We were glad to see that both thought a revised manuscript would make a timely and important contribution to the field as a full GMD paper. We agree that additional analysis and content along the lines they suggest will improve our model description - these revisions will take some time to complete, and in the meantime, this point-by-point response will address the comments they made and outline how we intend to revise the paper.

As written, our paper was intended primarily as a technical description of how our novel FAMOUS-to-ice-sheet coupling has been implemented, with a focus on the downscaling techniques and enough illustrative detail in the results to show that our coupled model simulations could be scientifically plausible even with the low native resolution of the FAMOUS atmosphere. We are pleased to see that Reviewer 1 thinks we have succeeded in this aim, at least for a modern climate state (first blue sentence below). Given this intention, we are not keen on having this paper include extensive analysis and evaluation of any individual climate simulation where that would not speak to our primary technical focus. Certain issues that significantly affect the setup of a simulation with a coupled climate-ice model (for instance spinup/initialisation techniques, which are still the subject of much ongoing research) we considered to be out of scope for this paper. It's clearly not sensible to suggest that a modelling technique should be described without giving some evidence that its results are likely to be fit for some stated purpose if applied appropriately, and the results and analysis we have done - and will expand on in a revised manuscript, guided by the review comments - are framed with this in mind. We wish to demonstrate the capabilities of the coupling techniques in the model, rather than evaluate the large-scale climate of the FAMOUS atmosphere or the general behaviour of the Glimmer ice sheet (both of which have been done in previous studies) under a particular set of boundary conditions. Our responses below will note where we agree with an additional analysis suggested by the reviewers, and also where we do not think that what they suggest will usefully serve the purpose we have in mind for this paper.

Below, the reviewers' comments are in blue italics and our responses to each follow in black.

Reviewer 1

The presented analysis demonstrates that FAMOUS-ice in combination with an ice sheet model is capable to simulate a reasonably realistic GrIS under present day climate conditions. The surface mass balance, however, appears to be extended to too high elevations and albedo seems to be overestimated in the ice sheet's interior. Given that other ESMs of similar design and resolution exhibit a qualitatively better skill (e.g. Kapsch et al., 2020; Fettweis et al., 2020), the analysis of the model should answer the question, whether these shortcomings are the result of a biased climate or of a misrepresentation of specific processes at the ice sheet's surface.

It is true of course that the simulation of surface mass balance components in FAMOUS-ice contains biases. We would note that the comparison with the ESMs contributing to GrSMBMIP (Fettweis et al., 2020) is perhaps too stringent a test for FAMOUS-ice - both CESM2 and MPI-ESM are much more physically complex and computationally expensive CMIP6-class models run at significantly higher resolution (CESM2's GrSMBIP submission was from a 1 degree model (37 gridboxes for every 1 in FAMOUS) and 32 vertical levels (almost 3x as many as FAMOUS) and MPI-ESM submitted results from an even higher resolution model - T127, 95 vertical levels.

Regardless of what is used as a comparison or its computational cost, FAMOUS-ice must be shown to stand on its own merits as suitable for its intended purpose, which we think it does - it seems Reviewer one agrees (their first sentence). Their question of whether the SMB, albedo and other GrIS biases that we do have are rooted in the large-scale climate or our downscaled surface modelling is a good one though, and very relevant to what we want to address. We will discuss this topic further in the revised paper. It's unlikely to be a straightforward either/or question, and shortcomings in both will play different roles. It's also unlikely to be just a question of misrepresented processes, as the resolution in FAMOUS is going to have a significant part on its own.

A very similar approach to ice sheet surface mass balance - similar down to the code level - is taken in UKESM1 (Sellar et al. 2019, a paper with significant additional description and evaluation of the ice surface schemes is in preparation) at a much higher resolution, equivalent to the CESM2 configuration noted above. GrIS surface fields in UKESM1 are generally much more similar to those in RCMs (eg MAR) than the FAMOUS fields we use to illustrate this paper, which suggests that the process parameterisations in our surface scheme are fundamentally capable of doing better, but are being held back by other factors in FAMOUS-ice.

Accumulation is not downscaled to elevation tiles in FAMOUS-ice, and is a pure output of the 5x7.5 degree atmosphere. Global GCM atmospheres are well-known for producing widespread, persistent drizzle where they should be dry even in modern, higher resolution models, and the smoothed, relatively low orography of GrIS and relatively high levels of moisture diffusivity inherent to the low resolution FAMOUS grid exacerbate this tendency. The significant bias in downwelling shortwave radiation over GrIS shown in our Figure 2 is also a product of the FAMOUS atmosphere, and can't be directly attributed to failings in the surface modelling. One might assume that the low albedo biases in Figure 2 and 3 would be a clear case of shortcomings in the surface scheme, but in fact these are partly result from tuning to compensate for the downwelling shortwave bias, so even the albedo bias is not simple to attribute.

Ultimately, to produce a "good" set of surface fields from a seriously biased background climate would require significant, deliberately unphysical interventions, such as using flux adjustments. That would not be an appropriate strategy for a model such as ours, intended for use in climate very different from modern, observed ones, and a balance must be struck between accepting the biases that are present and some model tuning that makes allowances for them.

Fig. 2 indicates that the downward shortwave radiation is strongly biased towards lower values while downward longwave radiation does not exhibit any clear bias. As argued by the authors, this may indicate an overestimated cloud cover, which should be substantiated by decomposing long wave radiation into atmospheric emissivity and near surface temperature. A potentially biased cloud cover should be reflected in atmospheric emissivity. Near surface temperature influences both, downward longwave radiation and turbulent heat flux and is an important driver of the SMB, which should be analysed as well. Given that the spatial distribution of surface mass balance and albedo from FAMOUS-ice differ strongly from respective fields simulated by MAR, I would propose to additionally show selected surface exchange characteristics (Fig. 2) as 2-D fields in comparison to MAR. It is important to not only analyse the spatial characteristics of the simulated SMB but to also evaluate the model's response to climate variations. The model is explicitly designed to simulate fundamental climate change and consequently this paper should allow to assess the model's skill to respond to climate change. This could be accomplished by using a transient ocean forcing from a 21st century climate projection.

Reviewer 2 also queries how the cloud is affecting both short- and longwave components of the radiation forcing, and we will look further to explain this for the revision and see how the downscaling that is applied to downwelling longwave in each gridbox features in this. These radiation components are also relevant to our scheme's modelling of albedo, so they are relevant from the point of view of our technical focus. As suggested, we will produce additional plots of the spatial variation of surface characteristics to complement the line plots in figure 2, and disentangle the effects of the elevation downscaling from simple variation with height along the mean ice sheet surface as suggested.

The sensitivity of the coupling scheme to climate change in FAMOUS is a matter raised by both reviewers. This is clearly an important topic, given the intended use of the coupled model, and we will include some analysis of it in the revised paper. However, given the caveats towards general simulation evaluation given in our preamble we are wary of going too far down this path in this paper since the SMB response will be controlled to a large degree by the climatic of FAMOUS to a specific forcing. We note that Gregory et al (2020, currently under revision for TC (tc-2020-89), first version available in TCD, GGS20 hereafter) already explores in some depth the response of an ensemble of the coupled FAMOUS-ice Glimmer system to a range of climate change forcings, so our additions will take GGS20 as context and provide complementary detail. GGS20 shows that GrIS SMB climate sensitivity of FAMOUS-ice is a reasonable match to the polynomial scaling developed by Fettweis et al. 2013 and used in AR4.

Rather than use a transient ocean forcing as suggested by the reviews, for our revision we propose to compare detail from one of the GGS20 ensemble, forced continuously by end of 21st century MIROC RCP4.5 SSTs, with MAR results diagnosed at the end of the 21st century from their transient forcing with the MIROC RCP4.5 projection. We note that this will be presented as a comparison between model projections, rather than assessment of "skill" as suggested by Reviewer 1. We don't consider that it is possible to assess a model's skill in the context of a climate projection, for which there is no verifiable truth (yet).

Furthermore, I found the manuscript at times hard to read,as essential information with respect to the snow pack schemes are only found in referenced literature:Page 10: It would be helpful to briefly summarize which processes influence the grain-size of snow and to generally characterize the relationship between snow albedo and climate forcing, and to give typical grain *sizes of new snow and old/wet snow. (Alternatively, a figure illustrating albedo as a function of influencing factors could be beneficial.) This is particularly desirable as Marshall (1989) is not easily available and as Gardner and Sharp (2010) indicates a sensitivity to grain size (0.05 /mm) towards greater grain sizes, which seems to be considerably smaller than the values used here*

We will provide more background, and summarise factors that influence snow albedo in the revised paper. Snow grain size distributions and seasonal evolution will be included in the set of fields analysed to characterise the capability of the model and explain the resulting albedo fields.

To address the particular comment on the sensitivity of the albedo to snow grain size: improving our match to observations of this relationship was the primary tuning criteria used in changing the parameters in our scheme. Compared to the original parameters in this version of the snow scheme in JULES, our model has an increased sensitivity to grain size (figure R1). The changes were made for the visible part of the two-stream radiation scheme in FAMOUS. Snow grain size is limited to 2000μm, which leads to an associated minimum in albedo. These adjustments were made to provide a more realistic representation of seasonal snow albedo evolution (e.g., Stroeve (2013) https://doi.org/10.1016/j.rse.2013.07.023). Additional observational studies were referenced, as was work on albedo optimisation in two-stream radiation code in ECHAM4 (Roesch(2012), 10.1029/2001JD000809).

The tuning parameter for which values are given in the main text, $\Delta a_{\mathrm{snow,visible}}$, is not directly used as the sensitivity of albedo to the grain size, but as part of a larger formulation. It is thus not directly comparable to the 0.05 /mm inferred from Gardner and Sharp (2010) by the reviewer. As can be seen from the orange line in figure R1, the albedo sensitivity to grain size that results from our formulation is in fact similar to this 0.05 /mm estimate, although the gradient is not constant across the full range of grainsize.

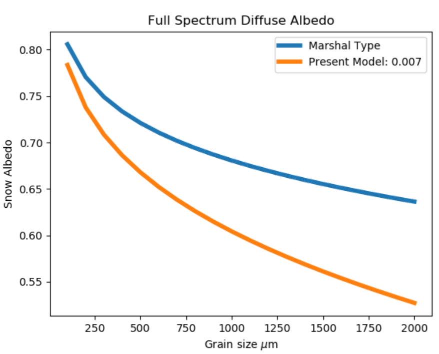

Figure R1. Snow albedo changes with grain size. Blue line shows the original tuning of the scheme in FAMOUS-ice, the orange line is the result of our tuning to observational studies

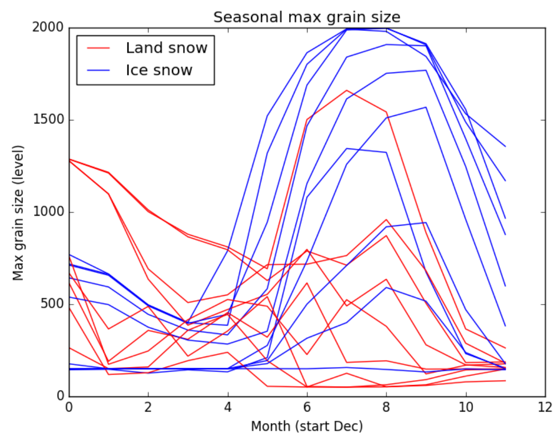

Figure R2. Seasonal maximum grain size occurring somewhere in FAMOUS-ice. Each line represents a different sample year. Red lines are for snow on non-ice surfaces, blue lines are for snow on the surface of the main ice sheet

As can be seen in figure R2, grain size for surface snow on the main body of the ice sheet often reaches the maximum value possible in one or more places during the summer melt season. This will lead to the lowest possible snow albedos in our scheme for these surfaces. This figure does not show influence of bare ice or near-surface air temperatures high enough to cause melt that will lower albedo further. These low albedos do not persist year round, and by the end of the winter fresh snow has covered over the higher grain-size snow. The low albedos shown in figure 3 of the main text are a summer phenomenon only - summer is when the albedo is most important, of course!

I also missed a separate short paragraph on how the snow pack evolves over time:How does density change if snow accumulates on top, what happens if the first layer melts, how and when are densities of the snow column prognostically calculated, what is the typical depth of compacted snow?

As requested, the revision will include more description of the snow relayering process, to save the need to refer further back in the literature

I also recommend to improve the structure of the paper by a further break down of sub-sections. Separate subsections might cover initial conditions, upper and lower boundary conditions of the snow pack as seen by the land surface model, and boundary conditions at the ice surface as seen from the atmosphere. Finally the analyzed experiments, together with the MIROC forcing and the MAR and RACMO simulation should be introduced in a separate section, possibly before section 4

We have no objection to reconsidering the subdivision of material in the paper. However, as noted already, we would like the focus to remain on a technical description of the methods used and the general capabilities of the model rather than a detailed evaluation of one particular climate realisation. Some of the suggestions made in this comment would not fit with our technical focus - for instance, description of initial conditions for the simulation. The choice of appropriate initial conditions can depend on the purpose of a specific experiment, so is not clearly part of a description of general model capabilities. GGS20 describe and justify the spinup strategy used for the experiments presented there. Once the additional analysis described above is complete, we will reconsider the structure of the paper

p. 5, l. 8: Is it possible to quantify the computational cost of the new model version in comparison to the old version?

yes, this will be noted. The additional tiles do have a non-negligible impact on the cost of FAMOUS, although appropriately configured on 10 processors throughput can still be measured in 100s of simulated years per wallclock day.

p. 5, l. 17: "We will describe later..." Please refer to the respective section.

will do

p. 7, l. 9: high-latitude -> high-altitude?

either would fit, I think, for this example. By "high-latitude" we meant latitudes near the poles

p. 7, l. 11: Is the sub-surface lapse rate in line with the coarse resolution subsurface temperature distribution?

That's a good question. In places is looks like it is, on this metric other transects across the ice sub-surface would suggest a higher lapse rate might be more appropriate. It's not, in fact a value we have tried to tune, so we don't know what the sensitivity of our results to it is. Other experimentation with the ice sub-surface formulation during development suggests that this lapse rate would not be expected to have a major effect on our results.

p. 11, l. 12: (fig:4) -> (Fig. 2) ?

yes, that's a typo.

Fig. 3, 5,6, : As the colorbar does not cover the full range of values, maybe include the range of values in the figure caption.

good idea

p. 12-14: does the spatial representation of SMB, albedo improve qualitatively if the number of elevation classes is increased?

The number of elevation classes used in a full simulation has not been changed since initial tests at the start of model development, and further simulations to assess how sensitive the SMB simulation is to this factor alone would not be straightforward. In an early version of FAMOUS-ice we used 25 evenly-spaced levels, and found that results with the same 10 class level placement as CESM1 used produced equivalent large-scale results at less computational cost. All further snow model developments and tuning were then done with these 10 classes as a base.

On the evidence we have I don't think it's likely that additional classes, with potential retuning of the surface parameters, would make a major difference to our results or the biases present in the surface fields - see discussion above of whether surface processes or large-scale climate are more responsible for the biases we see. Those background climate factors would not change with increased numbers of elevation classes at the surface.

p. 12, l. 9: It should be stressed that FAMOUS and MAR are consistently forced with the climate model output while RACMO is forced by reanalysis data. Also please specify the ERA forcing used and the period which was analysed.

will do

p. 14, Tab. 1: Please specify which ERA forcing is used here, Fettweis 2013 only has RACMO(ECMWF), and the RACMO(ERA) experiment that was used in this paper should be added to this table

will do

Reviewer 2

You argue that Helsen et al., 2012 developed an empirical parameterization to translate global climate fields to usable boundary conditions. However, they state that their method is not usable for models where ablation areas are not resolved.

OK, this will be rephrased.

How does FAMOUS-ice prevent snow from growing infinitely thick? Does it include a firn-to-ice densification scheme, or do you cap the snow above a certain thickness?

The snow scheme does include snowpack compaction and densification processes, although over time this alone would not prevent infinite columns of ice-density "snow" from forming where local accumulation is greater than ablation. Indeed, when not coupled to an ice sheet model, all UM configurations old and modern that we are aware of do slowly accumulate very large snow masses in such regions until someone thinks to reset the initial conditions - this is a numerical quirk that has very little effect on the climate simulation due to the way that the snow surface is implemented in these configurations. In FAMOUS-ice we use the coupling scheme as outlined in section 8.1, and the snowpack mass is reset to a default value at the start of every year. The amount of mass added/subtracted in the course of this resetting is what we pass to the ice sheet as the SMB term. In the revised paper this will be clarified as part of the new paragraph requested by reviewer 1 that will describe the relayering of the evolving state of the snowpack.

For downscaling of the temperature and (incoming?) longwave radiation from the atmosphere to the elevation classes you use an empirically tuned method, where you assume that the near-surface climate gradients are similar to the climate gradients in the lower atmosphere. Is this a valid assumption? For example, the temperature decreases with elevation along the GrIS near-surface, while the lower atmosphere has a temperature increase with height due to inversion.

The surface temperature and downwelling longwave are downscaled with elevation using a fixed lapse rate. This is commonly done in this sort of downscaling (see references in the original paper), although it is undeniably a rather simplistic method and cannot reflect local conditions such as inversions that deviate from its cooling-with-height paradigm. The value used for this lapse rate is sometimes treated as a tunable parameter in order to achieve a realistic surface temperature distribution on the ice sheet. Used, as we do, to reflect subgrid-scale surface variation in a single land model gridbox, this near-surface-like scaling is more justified than trying to use the state of the free atmosphere column, especially in a lower-resolution model like FAMOUS where the vertical resolution in the atmosphere is rather low. As noted in the text, we did try this latter approach initially but found it did not produce good results in FAMOUS. Some of this discussion will be included in the revised text.

I am not convinced you improve on the elevation class implementation analysed by Sellevold et al., 2019, as you are not comparing the same metric - that is, you do not subtract the grid-cell average from the same grid-box elevation class simulated value.

My impression of the CESM1 classes shown in Sellevold et al. (2019) was that several quantities were found, per gridbox, to increase with height where they should decrease and vice versa (their Figures 1,2) whereas our Figures 2 and 4 imply that overall we reproduce relationships with height in the correct sense for all variables. I do accept that their metric is more detailed, and perhaps more of a stringent test than what we have shown here. We will attempt to reproduce their metric more precisely and amend the revised manuscript accordingly.

Below 1000 m, the latent heat flux is positive in FAMOUS (Fig. 2), while negative in MAR and RACMO. First, it should be indicated whether the positive values means energy loss or gain at the surface. I am assuming the latter, as MAR and RACMO simulate climatological mean sublimation at lower elevations. Can you explain why FAMOUS simulates climatological mean deposition at these elevations?

Thanks for this - there was a sign error in converting the FAMOUS-ice sublimation diagnostic into latent heat in this plot. There is indeed sublimation at lower elevations, and low levels of deposition higher up. Latent heat fluxes on Greenland in FAMOUS-ice are generally very small: looking at the spatial patterns of the surface fluxes in the revision, as requested by Reviewer 1, will help clarify what is going on.

You argue that the shortwave down is too low because of a high biased cloud cover. I am surprised to see that this is not compensated for by an overestimation of longwave down. Is this due to the atmosphere being warmer in FAMOUS?

Reviewer 1 also requested more analysis of how the cloud and radiation components produce the surface energy balance we see: this will be addressed in the revised paper.

The simulation of albedo shows clear differences, particularly at higher elevations,when compared to MAR. You argue that the trigger for the lowered albedo at the higher elevations is a warm temperature bias at higher altitudes. Including a (supplementary) figure of April/May near-surface temperature or snow grain size compared to MAR/observations would likely strengthen this argument. Further, is it possible to assess whether the albedo is triggered too much and/or too early in the melt season?

As requested by reviewer 1, a number of additional fields comparing the 2d surface state of FAMOUS-ice with MAR. The low simulation of albedo is a melt season phenomenon - winter albedo, dominated by the low grain size of freshly fallen snow, is generally high everywhere. Figure R2 shows that grain size for surface snow often reaches maximum values on the ice sheet during the summer. There is a feedback between higher melt, lower albedo and faster-evolving grainsize, so it is not trivial to clearly identify a single initial trigger. We will show more analysis of the causes and timing of the albedo we simulate in the revised paper.

Also, if the atmospheric temperatures are too high at high altitudes, the melt energy contribution from the sensible heat flux should also be higher. As you don't show the sensible heat flux, I am left to wonder how the simulation of the sensible heat flux is.

this will be included as part of the additional comparison of the surface state compared with MAR

How well is rain simulated over the ice sheet? As the ice sheet is warm at higher altitudes, is there an overestimation of rainfall at higher altitudes? Could this also contribute to a lower albedo?

There is an over-estimation of low-intensity rain. In common with most GCMs, especially lower resolution ones, FAMOUS tends to drizzle too much everywhere. However, rain is ignored by the snow pack model in FAMOUS-ice - it isn't allowed to pool on top or percolate through the snowpack, but instead runs off straight away to the ocean. It has no direct effect on the albedo calculation or snow pack properties in general

The presented surface mass balance shows a realistic distribution with ablation at lower elevations and accumulation at higher elevations. However, the ablation areas are biased large and the high accumulation areas in the Southeast and Northwest show too little accumulation. As you don't present a transient climate simulation, I think the possible effect of these biases in transient simulations deserves a paragraph in the discussion section.

Reviewer 1 also requested some consideration of how the model responds to climate change, this is discussed above. We note again that GGS20 shows that the sensitivity of GrIS SMB to climate change forcing in FAMOUS-ice is similar to the polynomial scaling estimated by Fettweis et al. (2013) using MAR simulations. Some analysis of the SMB and surface exchange fluxes in FAMOUS-ice under climate change will be considered in the revised manuscript.

It would be easier to understand some of the thickness anomalies in Fig. 8 if you showed the ice velocities. Consider adding it to supplementary materials.

this will be done

I found many occurrences of double parentheses with citations. If latex is used, use(text before citation) or to avoid.

we'll fix this

P. 11, l. 12: replace fig:4 with figure 4.

will do

P. 19, l. 2: "a model"

will fix

P. 19, l 2: "allow for coupled"

I think this sentence is actually grammatically OK as it stands?

Fig. 7, caption: replace x in: 10x km

will do