Review 1

The authors analyse how differences in three climate models contribute to difference in the modelled sea ice. They apply the ISF method to break out contributions of individual components to drivers of melt and freeze throughout the year. Overall the manuscript is well written, clear, and has sound methodology. Well done! My concerns are primarily related to clarification and minor in scope.

We thank the reviewer for their kind comments about our study. We apologise for the time taken to produce this response. Some of the issues raised by Reviewer 2 required detailed consideration of the timescales of aspects of ice thermodynamics, and of the atmospheric boundary layer, before we felt an adequate reply could be submitted. In addition, three weeks were lost due to COVID in the family. We wished to ensure maximum consistency between our responses, and therefore submit them at the same time.

Specific comments

- *Line 110, remove "and" after Wang et al.*

This will be done.

- *Line 197: I would recommend rewording "thicker ice in HadGEM3-GC3.1-LL and UKESM1.0-LL causing a colder surface temperature, and less heat loss to space, than is the case in HadGEM2-ES" to: "compared to HadGEM2-ES, the two CMIP6 models have thicker ice which leads to a colder surface due to reduced heat conduction through the ice, and the colder surface results in less longwave radiative loss to space."*

Yes, this is much clearer and we will amend this as suggested.

- *Line 216: You haven't cited or mentioned CICE yet. You may want to expand a bit about why or how the ice models are different.*

We have added information in section 2 to clarify that the two CMIP6 models use CICE, and specified the configuration.

- *Figure 3:*
    - *Could you plot surface albedo differences to show the net spatial impacts of all components?*

This step would be quite similar to the main ISF analysis – quantifying the impacts on surface flux of all the components, the main difference being that it involves quantifying the impact on surface albedo, i.e. going back one step in the causal chain. We will consider how best to work this in, whether in a separate figure, additional panels, or in a supplementary figure.

Taking any sort of average of surface albedo values involves a subjective judgment: take a simple arithmetic mean, or first weight by downwelling SW? We would probably adopt the first approach as it will show the immediate effect of the various differences more clearly, and the second approach would produce an answer identical (up to scaling by a constant) to the main ISF analysis.

- *It might be nice to compare with observations here, where appropriate. The different models have hugely different ice fractions and melt pond fractions. Which are most reasonable given observations?*

This is a good idea and we will try to work this in, whether in additional panels or by means of text description. Some of the variables plotted here are quite poorly observed (particularly snow fraction) but it should be possible to say something about the melt pond and ice fractions.

- *Equation 2: instead of MODEL it says MODE. Same two lines above.*

In fact these should read MODEL1, MODEL2, MODEL1 and MODEL2, and seem to have been truncated somehow before publication, which we should have noticed. We will make sure that these are correct in any future published revisions.

- *Equation 5: At first it is nearly impossible to tell the difference between a for area and alpha for albedo. Could Area be changed to A_i or bolded to make this clearer? Same for lines below.*

This is also a good idea, which Reviewer 2 mentioned. Thank you.

- *Line 282: Please clarify how bare ice fraction is found.*

Bare ice fraction is one minus the sum of the other terms. This will be stated in the text.

- *Line 285: If this equation is relevant, may want to number it. Also, please define the albedo of the ocean. Is the albedo over different surface types output directly?*

This equation will be moved out of the paragraph and numbered. The albedo of the ocean will be defined.

- *Line 304-306: Did you verify that the answer is similar by using either/both CMIP6 models and comparing ensemble mean to the individual ensemble members?*

No, and this would be a good idea. We will carry out this comparison and report on this in a revision.

- *Figures 4,5,6:*
    - *These figures are really nice and clear, but I had a lot of trouble with the colors. Please modify the colors to improve readability, change dash patterns, or bold particular lines of relevance.*

Thank you, this is useful feedback. We will try to use a mixture of linestyles and colours to distinguish the variables.

    - *I think monthly mean figures of the spatial difference contributions might be useful as a supplementary figure to see which components dominate in different regions.*

This is a good idea and we will look at spatial patterns of the ISF differences next. The spatial patterns observed will probably be described in the main text, though as you say any associated figure can be supplementary (unless there is anything really striking).

- *Line 328: Why is the snow thickness so different between these models?*

We are not sure. The baseline climate in UKESM1.0-LL is considerably colder in the period in question, so snow accumulation during the freezing season may be greater (though this association is obviously by no means automatic, as one would also expect total precipitation to be lower in the colder model). We will look into this further by examining seasonal and spatial patterns of snowfall, but a more in-depth analysis (e.g. by analysing the role of atmospheric circulation differences) would probably be beyond the scope of this study.

- *Line 379: I would recommend rewording "differences in the thicker categories contribute much more towards the total" to: "contributions across all thickness categories are similar rather than being dominated by very thin ice as was found in the previous comparison."*

Yes, this is much clearer and will be amended as suggested.