Reply to Reviewer #1:

We thank David Bailey for the time and efforts he spent reading our manuscript and providing valuable advices. Please find below a discussion of the reviewer's comments (italic). Changes/additions made to the text are underlined and given in quotes.

This is a very nice study that assesses surface albedo from the MOASAiC campaign versus the HIRHAM-NAOSIM model. Maybe a lot of it is my misunderstanding of what was done here. The pieces are mostly here and I do think this is a worthwhile study, but I have some fairly significant concerns here.

1. In terms of originality, Light et al. have recently published a similar study in Elementa. While there is more emphasis on the model results here, I still think some additional contrast to what they found would be useful here.

Following the Reviewer's comment we refer to this study in the Section 5 when summarizing the results of the comparison with MOSAiC data:

"Simulations and ground-based measurements of the seasonal evolution of surface albedo during MOSAiC were previously presented by \cite{Light_2022}. The authors used an Earth system model (1° spatial resolution) for comparison with surface albedo measurements manually made along three survey lines. These measurements could not be performed with the same high temporal frequency during the complete campaign for logistical reasons. Therefore, the transition from dry to wet snow during the onset of melting was less well observed than in our study, which relied on autonomous measurements from a radiation station. Similar to our results, \cite{Light_2022} showed that in particular the representation of melt pond albedo in the model needs to be improved, while the general surface albedo values and properties of the different ice types were captured quite well."

In the context of the study by Light et al., we did not explicitly mention the difficulty of evaluating the modeled surface type fraction from ground-based observations. Even using observations along a 200 m survey line instead of single point measurements will hardly represent the variability within a modeled grid cell with 1° spatial resolution. At the end of Section 5, we write:

"We invite the modeling community to use this airborne data set to evaluate other surface albedo schemes, as it provides decoupling of surface type fraction and surface albedo parametrization for larger spatial scales than covered by ground-based observations."

2. My biggest concern is the bias in absorbed shortwave (irradiance), Figure 8. Did the authors compare the incoming shortwave between the model and observations? The albedo could be perfectly correct, but if the incoming shortwave is biased, then the absorbed will be similarly biased. I am not an expert in atmospheric radiation, but I think it would be helpful to see a comparison of incoming and outgoing shortwave. Perhaps this was mentioned, but I think this could be expanded upon.

Indeed, the net irradiance is highly dependent from the representation of the incoming (here called downward) irradiance. An underestimated modeled downward irradiance directly leads to an underestimation of the modeled net irradiance assuming the same surface albedo. Therefore, we included the information on the bias of modeled downward irradiance in Figure 8. The size of the symbols directly corresponds to the difference of the modeled and measured downward irradiance. To make this clear, we adjusted the figure caption:

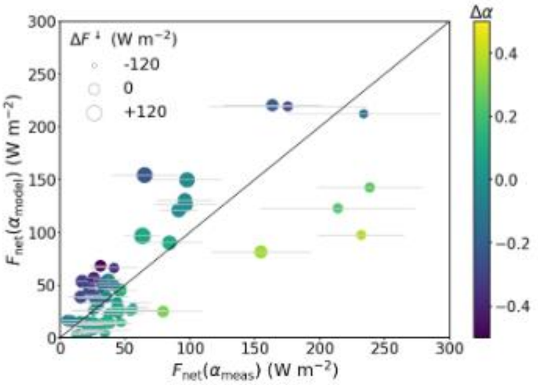
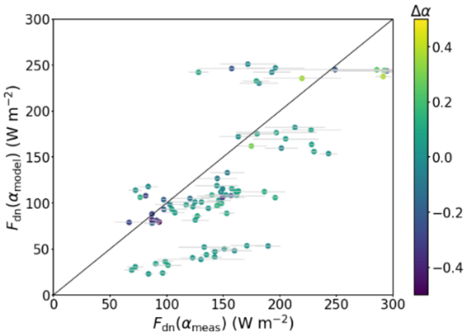


Figure 8. Scatterplot of net irradiance based on measured and modeled surface albedo covering the flights performed during PAMARCMiP and MOSAiC-ACA. The horizontal bars indicate the standard deviation of the averaged measured F_{net}. Color code gives the surface albedo difference ($\Delta\alpha = \alpha_{\text{model}} - \alpha_{\text{meas}}$) and the symbol size the difference of the modeled and measured downward irradiance.



Not included in the manuscript: a figure showing the measured and modeled downward irradiance. The mostly negative bias of the modeled $F\downarrow$ is visible. In the manuscript, we report on the mean deviation.

We rephrased parts of this section:
"The F_{net}-differences between measurement and model depend not only on $\Delta\alpha$, but we must also take into account the difference in the downward irradiance ($\Delta F\downarrow$). A negative $\Delta F\downarrow$ (smaller symbols in Fig. 8) may occur when the modeled extinction of $F\downarrow$ caused by modeled clouds is higher than an observation would show. This is especially the case when cloudless situations were observed but not modeled. It would lead to an underestimation of the modeled net irradiance, assuming the same surface albedo. In fact, a mean negative bias of the modeled $F\downarrow$ (mean $\Delta F\downarrow = -31\ \mathrm{W\ m^{-2}}$) was found, which can be related to an overestimation of the modeled cloud cover. However, the downward irradiance itself also depends on the surface albedo. In particular, below clouds $F\downarrow$ is enhanced for brighter surfaces due to multiple-scattering between surface and cloud base. A positive surface albedo bias would lead to a positive bias in $F\downarrow$, assuming a similar cloud representation. On average $\Delta\alpha$ was 0, indicating a small effect of surface albedo on the modeled $F\downarrow$.
Overall, both cloud properties and surface albedo must be well represented for modeling net irradiance correctly. To estimate whether the representation of clouds or the surface albedo potentially contribute more to the uncertainty of F_{net}, we calculated …"

3. On a similar note, the authors talk about the importance of albedo for climate model simulations. However, related to point 2, we often have to adjust the snow albedo to compensate for biases in the incoming shortwave. So, it is possible to have the "correct" albedo, but for the wrong reasons.

The surface albedo is a crucial parameter for modeling radiative transfer in the atmosphere, especially for calculating the upward irradiance. The downward irradiance at the surface is very sensitive to the properties of the atmospheric components (aerosol and cloud particles, trace gases, ...). Therefore, in atmospheric applications, these components must be properly reproduced by the model to obtain a correct model output. Adjusting the surface albedo to get a correct downward irradiance is less effective. However, if you are interested in modeling radiative transfer within a snow layer, for example, you need the incident irradiance, which can be adjusted by changing the surface albedo. Perhaps the reviewer is aiming in this direction. Since we are interested here in the atmospheric solar irradiance effect of surface albedo, we would rather suggest that the atmospheric parameters be adjusted so that the downward irradiance is well reproduced by the model.

4. What is the temporal resolution here? It wasn't obvious to be if these are instantaneous, hourly, etc. I assume the model is saving the fields at the same temporal resolution? How is albedo defined when there is no sun?

The temporal resolution of the model output for MOSAiC (2020) was three hours, and one hour for PAMARCMiP (2018). It is mentioned in Section 2.3:

"The model output was given with a spatial resolution of about 27 km distributed over 200 x 218 grid points on a circum-Arctic domain. [...] The HIRHAM-NAOSIM model was run for 2018 covering the time frame of the PAMARCMiP campaign (temporal resolution of 1 hour), and for the entire MOSAiC period (temporal resolution of 3 hours) that includes the time frame of the ground-based measurements from spring to autumn 2020 and the period of the aircraft observations during MOSAiC-ACA."

For the comparison of measured and model data, the spatial and temporal overlap between the two data sets was taken into account when filtering the data.

How is albedo defined when there is no sun?

Since the surface albedo is not determined by a dependence of the solar zenith angle, there is no difference for the case when there is no sun. All data considered in this study were taken during the presence of the sun (polar day).

5. I'm very confused about the use of "online" and "offline" models here. Is the difference that one has prognostic radiation and the other has specified radiation? I would like the authors to expand upon the description of these. I think this is where you are trying to get at the question raised earlier about whether the incoming shortwave is biased, or the albedo is biased. I think a bit more could added to section 4.2 to help alleviate these concerns.

We have tried to explain the differences between online and offline simulations in section 2.4. The "offline" mode applies only the two parameterizations of subtype albedo and subtype fraction as they are implemented in HIRHAM-NAOSIM. It uses measured parameters that were derived from the observations along the flight tracks. In contrast, for the "online" simulations, the HIRHAM-NAOSIM model package was run completely independently of the measurements. So, the results discussed in Section 4.1 refer to an "online" application of the model, whereas Section 4.2 takes only the parametrizations into account.
We have adapted the beginning of Section 4.2 to introduce the independent radiative transfer simulations that allow a sensitivity study of the F_{net} dependence on surface albedo.

"In contrast to the study of the HIRHAM-NAOSIM results, the application of the offline evaluation allows a reduction in the dependencies of the Fnet bias for the comparison of the parameterization with the airborne measurements. The measured subtype fractions were used to identify only the influence of the bias of the parameterized surface albedo on Fnet, without having to consider the uncertainties of the subtype fraction parametrization. The net irradiance was determined along the flight path for seven selected days during all five flight campaigns, covering cloudy and cloudless conditions. Radiative transfer simulations were performed for these cases using the measured and parameterized surface albedo. In this way, the sensitivity of net irradiance to surface albedo was quantified under the same predefined atmospheric condition. These conditions were matched to the measurements made during the selected flights (see Appendix A).”

Minor points.

1. In figure 3, the panels that show the surface type are hard to see (a, g, c, i). Maybe just lines instead of filled contours. The red of melt ponds in particular is hard to see.

It is true that the proportions of each type are difficult to read when their contributions are small, as in the case of melt ponds. However, we deliberately chose to use a stack plot so that we could immediately identify the dominant surface types. Individual lines, as suggested by the reviewer, would not be helpful because the temporal variation is quite high.
The proportions of surface types are presented as a stacked area plot to identify the predominant subtypes. We have improved the figure caption to better indicate the surface types, and added the following to the figure caption:

“Figure 3. (a) - (j) Temporal development of surface types, surface albedo (blue lines; left y-axis) and surface skin temperature (grey lines; right y-axis) for all five flight campaigns. The proportions of surface types are presented as a stacked area plot to identify the predominant subtypes. Vertical green lines separate the individual flight days. Dates given in the panels are explicitly mentioned in the text. (k) Averaged surface albedo as a function of sea ice fraction (bin size of 10 %), separately for cloudless and cloudy conditions. The standard deviation of the averages is represented by thin vertical bars.”

At the beginning of Section 3.1, we mentioned the type of plot directly:

“An overview of the proportions of classified subtypes along the flight tracks of the five campaigns is shown in Fig. 3 as a stacked area plot. The temporal development …”

2. In figure 4, I prefer you not use the description of "violin" plot. While this might describe the shape it doesn't say anything about what you are showing. Just a description of what you are showing is sufficient. Also, you could refine the Y-axis. Everything below 0.6 is not interesting in spring and summer.

We have replaced the term “violin” plot by “distribution” in the figure caption. A deeper description of the figure is given in the main text.

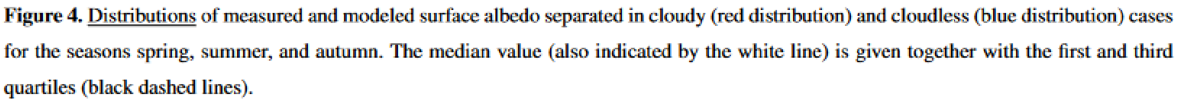
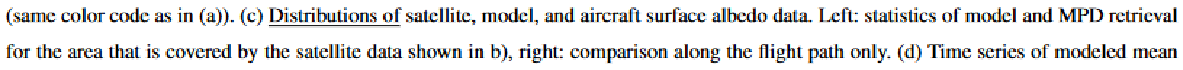


Figure 4. Distributions of measured and modeled surface albedo separated in cloudy (red distribution) and cloudless (blue distribution) cases for the seasons spring, summer, and autumn. The median value (also indicated by the white line) is given together with the first and third quartiles (black dashed lines).

We prefer to keep the y-axis in order to have a uniform scale for all seasons. This facilitates comparability between the individual distributions.

3. Similarly in Figure 5c. Are you simply reflecting the same information on both sides of the line?

The four individual distributions are symmetrical because they are not divided into cloudy and non-cloudy cases as in Fig. 4. Similar to Figure 4, we have changed the figure caption:


(same color code as in (a)). (c) Distributions of satellite, model, and aircraft surface albedo data. Left: statistics of model and MPD retrieval for the area that is covered by the satellite data shown in b), right: comparison along the flight path only. (d) Time series of modeled mean

4. Figure 7b is a similar issue to point 1. I find that these "stacked" plots are kind of tricky to interpret. Maybe line plots are better here.

For consistency with Figure 3, we would like to keep the stacked area plots.