# Reviewer 1

Overall, the authors have addressed the review comments to my satisfaction. The changes made by the authors clarify the text and better justify the results compared to the previous revision. I just have a few specific remaining questions and technical corrections detailed below. So long as these points are addressed, I would be happy to recommend the article for publication.
Regarding the expanded discussion of the snow density measurements, I can agree with the authors that detailed examination of snow density is beyond the scope of the paper. I very much agree that further analysis of snow density is required, but this is difficult given the high variability in snow density as is mentioned. As such, I think the amount of discussion in the revised article is sufficient.

I appreciate that the authors have now included simulation data and code, since that helps with data availability. Regarding the analysis for the other locations (Sloop and Runway), having now seen the figures I can agree they are similar, and since the output data is now provided in case others are interested in these other results, I can grant that it suffices to simply show the location with the most data (Nloop) as is done in the revised manuscript. The section restructuring for the article (e.g. Section 5) also helps the article flow better.
Remaining questions:
1. I appreciate that the authors now including the correlation breakdown by melt period in Fig 7 per my previous suggestion. However, I am confused as to why the breakdown is not shown for snow density (Fig 7 b), since there do appear to be points available during the melt period (the grey dots around 2020-07 in Fig 7 b). Is this omitted because these densities are indirectly calculated from SWE? I am somewhat confused as to what the grey dots in Fig 7b represent here and would appreciate some further clarification; perhaps specify more clearly in the text where the grey dot values are obtained from.
2. Although the authors acknowledge my suggestions in their review response, Figure 6 (previously figure 5; the MOSAiC bulk snow density time series plot) appears to not have been updated accordingly. I assume this was just an accidental omission but would appreciate for the figure to be updated for the final manuscript as the authors describe.
3. The rearranged sections are clearer, but now I think it would be helpful if the authors would briefly specify in Section 3.2 which specific precipitation type provides the forcing for MicroMet (i.e. is it total precipitation or snowfall rate and rain rate from MERRA-2). Likewise please briefly specify here what precipitation type is used from the KAZR for this study.

Technical corrections:
Backward opening single quotes e.g. at L72 ‘rotten ice’ (same issue at L101, 112, 280, 425, 435)
L 272: “These periods and marked” -> “These periods are marked”
L 540: Katlein et al. citation should be parenthetical, not in-line
L 590: “Simulations results” -> “simulation results”

**Author’s Response**

Thank you very much for your time and expertise. Our manuscript was improved greatly by your efforts. See below the additional modifications we made in this revision.

Question 1:
This is a very good point. There were no observations of snow density after the snow melt onset. The value used here was at constant 550 kg/m3 that was estimated. Since this is not a real measured value nor it is used in the simulations, we have now modified Figure 7 and omitted these data points.

Question 2:
This is true – the figure was updated, but not uploaded into the revision. This is now updated as promised. Apologies!

Question 3:
KAZR data for MOSAiC (published by Matrosov et al, 2021) has only type – total precipitation. From MERRA-2 we also used total precipitation, and we let MicroMet distinguish between snowfall and rain using the air temperature formulation of Dai (2008) as described in Liston et al. (2020). This information is now briefly clarified by:

All of the MetCity, KAZR, and MERRA-2 data were then aggregated to 3-hourly values used in the model assimilations (averages for air temperature, relative humidity, and wind speed and direction, and sums for **total** precipitation). **MicroMet was used to distinguish between snowfall and rainfall using the air temperature formulation of Dai (2008) as described in Liston et al. (2020).**

Dai, A. (2008). Temperature and pressure dependence of the rain-snow phase transition over land and ocean. Geophysical Research Letters, 35, L12802. https://doi.org/10.1029/2008GL033295

We thank the reviewer for the technical corrections – they were implemented! In addition, we did some minor technical (language) corrections in the text and figures. All of them are tracked in the 'difference' PDF file.

## Reviewer 2

First of all, I would like to express that I appreciate that the authors have worked on their manuscript! I think the structure and thus readability have really improved, and it is easier to understand what has been done here. The motivation behind and main idea of the paper are of course still very relevant and would serve the scientific community. And I think that the continuous year-long snow and ice time series produced here would be really useful. However, I think that the authors should either focus on producing the most realistic estimation of a continuous snow and ice time series assimilating the model to the observations and "do the best they can" to achieve this OR focus on "praising" their model and show how well the simulations agree with observations (see my comments 1 and 3) . Here, it looks a bit like they are trying to do both… In addition, there are still inconsistencies (see comment 2), and it is still unclear to me which of the described observations have been used (and which not) and why (comment 4). As I think that most of these concerns can mainly be addressed by modifying the description instead of the actual work, I would consider these minor revisions.

1. The authors (still) claim that the SnowModel LG reproduces the snow evolution accurately (p. 21, l. 309-401 and l. 420-421*1 and p. 28, l. 588-589) during the time periods where there are no observations. The difference D (in terms of SWE) between the simulations and the observations has been subtracted from the simulations to match the observations (especially important for the two SYI cases), so of course the time series agrees well with the observations but we do not have information for the time periods in between, so how can the authors judge on this...? An exception is the good agreement between simulations and observations in June/July, which were not assimilated, which is why - here - the model can be "evaluated" (as is separately mentioned on p.21, 402-405). But note also my comment 3 as the starting point of the model might have been adjusted to potentially match these observations (or the ice thickness observations) better (?).
*1Additional remark: I had already mentioned this in the first review round, pointing at the sentence that is now found on l. 420-21 (before: l. 394) and the authors replied "Will be corrected", which was obviously not done as the exact same sentence is still in the manuscript.

2. The authors write that D (= the difference between observations and simulations, assumed to be related to deformation) is small throughout the simulations (p.22, l. 443) and that "its accumulated effect by 7 May (the last winter observation) was 10% in Nloop". I guess they mean 10% of the precipitated SWE value (which still means that the simulated SWE at this point in time has to be reduced by ~50% to match the observations). However, based on Fig. 8 it should be ~15% (visual inspection)... The number of 10% is also given in the abstract, should be checked.

3. The authors did not explain (although I asked for it in the first review) why the surface freeze-up date (= start of snow accumulation) for the fresh ice ("deformed SYI") was determined using the (3 hourly) air temperature time series, while for the saline ice ("ponded SYI") it was determined using the 3 day running mean of that time series. (In their reply they explained why different thresholds were used, which was clear (different freezing temperatures of fresh and saline ice) but not why the 3 day running mean was used in one case but not the other). If this cannot be explained, I (have to) guess it was because the results (e.g. regarding the match of simulations and not assimilated observations in June/July and/or the agreement between simulated and observed ice thicknesses) fit better when doing it like this...? I think it is ok to use a more "subjectively" determined starting date for the simulations (as compared to using identically - and thus more objectively – derived dates for both) but this has to be communicated. Especially, as it is the authors themselves who stress that "The correct initialization of our simulations proved to be the most critical aspect of our work." (p. 28, l. 603).

4. In Section 3.4, the MagnaProbe snow depth measurements along the transects during MOSAiC are described but I cannot see where they are used in this study. In Section 5.1, the authors write that they use SWE observations (described in the section on snow density 3.5) to assimilate theSnow Model LG model. And in a second step they use the snow density evolution over time (eq. 1, snow density is a function of days since 25 Oct 2019) to assimilate snow density. The simulated snow depths are then compared to the snow depths*2 that result from the SWE observations and this snow density evolution (snow_depth = SWE * density_water/density_snow) and NOT to any (directly) observed snow depths. Thus, I guess snow depth observations have not been used for SnowModel LG simulations or assimilations (?). Later on, the authors write that "Our simulations used observed MOSAiC snow depths ... to drive our ice growth simulations" (p. 19, l. 370-371). Here, it sounds like (directly) observed snow depths have been used, but this would contradict the statement a few lines earlier, where it says that "we performed ... SnowModel-LG simulations ... to create the snow forcing for the HIGHTSI sea ice simulations… using SWE observations" (p. 19, l. 359-362). Only in Fig. 9c (directly) observed snow depths are shown, but only the ones from coring and the "Ridge Ranch". Are the snow depths measured along the transects with MagnaProbes shown or used anywhere?
Another aspect that is unclear to me: The SWE observations are introduced in section 3.5 (‘Snow density’) as being measured in the snow pits. Figures 7, 8 and 9 show time series for the three ‘transects’ (Nloop, Sloop, Runway). On page 6 the locations of the snow pits are given (l. 128 ff). Nloop and Runway are among these. Did you use only SWE measurements located at these transects or all of them? For Sloop, did you take the closest ones (temporally and/or spatially) or all the measurements at any time on level and ponded SYI? Wouldn’t it be interesting to see how the snow depths measured along the transects compare with the simulations and the SWE-based snow depth estimations? Or maybe the snow depth observations along the transects ARE used somewhere but I cannot find it…
*2 It is confusing that the snow depths used in this context are called ‘snow depth observations’, as they are rather ‘snow depth estimates based on SWE observations’, especially because snow depth observations have actually been collected, too.

Other minor comments:

p. 8, l. 161: analysis of (Boisvert et al., 2018) --> remove parentheses
p. 14, l. 272: So far, mainly "Central Observatory" was used, here you use "CO" + "These periods and marked...": and -> are?
p. 15, eq. (3): In the equation "M" is used, while in the text "S_M" is used.
p.25, l. 511-512: "... led to younger and thinner sea ice thicknesses ... equal to that of the oldest and thickness ice types" -> still does not make sense to me, did you mean "oldest and thickest ice type"?

**Author's Response**

Here we would first like to state our gratitude for the work of this reviewer. Her/his comments were again very detailed and valuable for improvement of this manuscript. The first revision required a lot of text rewriting and some details have slipped our attention there. It is remarkable how this reviewer found many (hopefully all!) of them and pointed them out to us. It was our pleasure to be able to prepare the second revision.

Issue 1:
We agree with the reviewer of why the model is giving good results – in addition to the sound physics in the model, the initial conditions were carefully assessed and afterwards observations were assimilated. We carefully examined the wording in the listed statements and made small adjustments that keep the statements sufficiently short to be understandable. Here are the disputed statements one-by-one:

OLD:
However, the bias-corrected atmospheric reanalyses is of sufficient quality that SnowModel-LG **(apparently) accurately reproduced** these missing periods.

NEW:
However, the bias-corrected atmospheric reanalyses **data are** of sufficient quality that SnowModel-LG **simulated physically-credible values during** these missing periods.

OLD:
Using observed and estimated atmospheric forcing data, and periodic SWE and snow density observations, SnowModel-LG **reproduced the observed** snow evolution on the three sea ice types with different age and sea ice deformation characteristics found at MOSAiC (Figure 8).

NEW:
Using observed and estimated atmospheric forcing data, and periodic SWE and snow density observations, SnowModel-LG **simulated physically-credible** snow evolution on the three sea ice types with different age and sea ice deformation characteristics found at MOSAiC (Figure 8).

OLD:
Here, we have combined physics-based modeling tools, with temporally incomplete measurements, to create a full annual time series of 3-hourly snow and ice property values that match the observations when and where they occurred. Finally, the time series data contain **realistic** values when observations were not available.

NEW:
Here, we have combined physics-based modeling tools, with temporally incomplete measurements, to create a full annual time series of 3-hourly snow and ice property values that match the observations when and where they occurred. Finally, the time series data contain **physically-credible** values when observations were not available.

Issue 2:
We are very grateful for the meticulous work of the reviewer on this issue!
We have recalculated the fractions of SWE removed by different sinks in comparison to the total precipitation. The recalculated results are listed below and the numbers in the text are adjusted to correspond to them. The changes are noticeable, but still small and do not alter the interpretation of the findings in this paper.

```
Nloop
May 7 cummulative effects:
Cumulative snowfall:  0.5396220300000004
model mean no D:  0.17634
model mean with D:  0.09262
fraction of SWE removed by S 0.6732157135986463
fraction of SWE removed by D 0.1551456303590866

Sloop
May 7 cummulative effects:
Cumulative snowfall:  0.42294580000000026
model mean no D:  0.13291
model mean with D:  0.09871
fraction of SWE removed by S 0.685751696789518
fraction of SWE removed by D 0.08086142479721982

Runway
May 7 cummulative effects:
Cumulative snowfall:  0.31302074
model mean no D:  0.09034
model mean with D:  0.09095
fraction of SWE removed by S 0.7113929255933649
fraction of SWE removed by D -0.0019487526609259162
```

In addition we get an impression that the ability of the model to simulate the SWE sinks was not explained sufficiently. The precipitated SWE is not removed to match the observed SWE. 68-70% of the SWE is simulated to be removed by a $S_{SS}$ and $S_{BS}$ model sink terms. These sinks are operating at the synoptic length scales (approx 100km). Only afterwards the remaining 10-1% of the excess comparing to the observed SWE is removed by our parametrization that works at LOCAL scale. To clarify also this the text has been modified as presented below. Also the value in the abstract was modified to 15%:

OLD:
On all three ice types, the strongest winter season sinks were static and blowing snow sublimation (\$S_{SS}\$ and\$S_{BS}\$) which by 7 May cumulatively removed 68, **68**, and **70** \% of \$SWE\$ from snowfall (\$P_S\$) in Nloop, Sloop, and Runway, respectively. This is represented by the difference between 'precipitation' and 'model no D' on Figure \ref{fig8}. Note that, in this environment, if the blowing snow is not captured by a ice-topographic drift trap, or blown into an open lead, it blows perpetually and, in air that has a humidity deficit, it eventually sublimates completely away \citep{tabler1975,liston_sturm2004}. These \$S_{SS}\$ and\$S_{BS}\$ values were about three times as large as in \cite{liston2020}. This is likely due to the specific weather during MOSAiC winter and location during the drift, including, generally low snowfall (\$P_S\$) after freeze-up, frequent storms with high winds \citep{rinke2021}, and relatively high sea ice concentration \citep{krumpen2021} with low near-surface relative humidity during winter. \$P_S\$,

$S_{SS}$, and $S_{BS}$ operate at synoptic temporal and length scales, and were the same (or very similar for $S_{SS}$ and $S_{BS}$**, which depend on grain bounding**) for all ice types.

The differences in $SWE$ evolution on the three ice types were largely controlled by the ice (and snow) onset date, and the differences in the remaining wintertime snow sink or source - the ice dynamics term $D$. This is represented by the difference between 'model no D' and 'model with D' on Figure \ref{fig8}. $D$ is the only **simulated local** source or sink; in the natural system, $D$ produces ice roughness features such as rubble ice and pressure ridges, and lead timing, size, and distribution. Following any sea ice deformation, a certain amount of airborne snow will be removed to open water in the leads \citep{clemens2022} or stored in snowdrifts at the deformed ice roughness features \citep{liston2018,itkin2023}. During winter, the wind velocity is frequently above the blowing threshold value (7.7 m s$^{-1}$) following \cite{li_pomeroy1997}, which provides the justification for our parametrization of $D$ as a sea ice deformation snow sink or source (see Section \ref{swe_calc}). After melt onset, the snow grains are wet and no drifting snow is observed \citep[\textit{sensu}][]{pomeroy1997}. Following this principle, $D$ was set to zero in May after the last transect measurements.

$D$ remained small throughout the simulations, but its accumulated effect by 7 May (the last winter observation) was **10**, 8, and **>1** \% in Nloop, Sloop, and Runway, respectively. $D$ is likely large right after freeze-up; this is a period of thin ice with **lots of** deformation. More studies of this fast-changing period with thin ice are needed to understand what exactly is happening when the ice first forms. The importance of erosion for $SWE$ at MOSAiC was explored by \cite{wagner2022}, who gave estimates of erosion based on uncalibrated snowfall rates, wind speeds, and $SWE$ in Sloop and Nloop. While the magnitude of the combined snow sink by \cite{wagner2022} is similar to ours (53-68\%), their study could not differentiate between erosion and sublimation. Our study shows that $D$ was predominantly a sink in the case of Nloop and Runway and, after ridge formation in November, was occasionally a source in Sloop.

NEW:
On all three ice types, the strongest winter season sinks in our simulations were static and blowing snow sublimation ($S_{SS}$ and$S_{BS}$) which by 7 May cumulatively removed 67, 68, and 71 \% of $SWE$ from snowfall ($P_S$) in Nloop, Sloop, and Runway, respectively. This is represented by the difference between 'precipitation' and 'model no D' on Figure \ref{fig8}. **The magnitude of $S_{SS}$ and $S_{BS}$ depends on grain bonding, which is for $S_{SS}$ determined by latent heat flux, while for $S_{BS}$ wind speed, humidity and solar radiation are the main factors \citep{liston2020}.** Note that, in this environment, if the blowing snow is not captured by **an** ice-topographic drift trap, or blown into an open lead, it blows perpetually and, in air that has a humidity deficit, it eventually sublimates completely away \citep{tabler1975,liston_sturm2004}. These $S_{SS}$ and$S_{BS}$ values were about three times as large as in \cite{liston2020}. This is likely due to the specific weather during MOSAiC winter and location during the drift, including, generally low snowfall ($P_S$) after freeze-up, frequent storms with high winds \citep{rinke2021}, and relatively high sea ice concentration \citep{krumpen2021} with low near-surface relative humidity during winter. $P_S$, $S_{SS}$, and $S_{BS}$ operate at synoptic temporal and length scales **(on scales comparable to, e.g., 3 hours and 100 km)**, and were the same (or very similar for $S_{SS}$ and $S_{BS}$) for all ice types.

The differences in $SWE$ evolution on the three ice types were largely controlled by the ice (and snow) onset date, and the differences in the remaining wintertime snow sink or source - the ice dynamics term $D$. **These factors operate at much shorter local spatial scales, e.g., 10 m.** This is represented by the difference between 'model no D' and 'model with D' on Figure \ref{fig8}. $D$ is the only **local simulated** source or sink; in the natural system, $D$ produces ice roughness features such as rubble ice and pressure ridges, and lead timing, size, and distribution. Following any sea ice deformation, a certain amount of airborne snow will be removed to open water in the leads \citep{clemens2022} or stored in snowdrifts at the deformed ice roughness features \citep{liston2018,itkin2023}. During winter, the wind velocity is frequently above the blowing threshold value (7.7 m s$^{-1}$) following \cite{li_pomeroy1997}, which provides the justification for our parametrization of $D$ as a sea ice deformation snow sink or source (see Section \ref{swe_calc}). After melt onset, the snow grains are wet and no drifting snow is observed \citep[\textit{sensu}][]{pomeroy1997}. Following this principle, $D$ was set to zero in May after the last transect measurements.

$D$ remained small throughout the simulations, but its accumulated effect by 7 May (the last winter observation) was **15**, 8, and **<1** \% in Nloop, Sloop, and Runway, respectively. $D$ is likely large right after freeze-up; this is a period of thin ice with **frequent** deformation. More studies of this fast-changing period with thin ice are needed to understand what exactly is happening when the ice first forms. The importance of erosion for $SWE$ at MOSAiC was explored by \cite{wagner2022}, who gave estimates of erosion based on uncalibrated snowfall rates, wind speeds, and $SWE$ in Sloop and Nloop. While the magnitude of the combined snow sink by \cite{wagner2022} is similar to ours (53-68\%), their study could not differentiate between erosion-**deposition** and sublimation. Our study shows that $D$ was predominantly a sink **(erosion)** in the case of Nloop and Runway and, after ridge formation in November, was occasionally a source **(deposition)** in Sloop.

Issue 3
This section was modified significantly in the revision and the reviewer is right that the use of the 3-day running mean was still not explicitly addressed. For clarity we now add the statement:
'**The 3-day running mean is used here to take into account the heat capacity of the freezing water column, opposed to freezing the ice surface.**'

Issue 4
The Magnaprobe and snow stake snow depth observations are used in combination with snow densities from snow pits to calculate observed SWE. This is now explained at the end of Section 4.1, right after the conversion equation is introduced:

**Because SnowModel-LG operates in $SWE$, while the MOSAiC observations provide $h_s$ (Section \ref{sec_snowdepth}) and $\rho_s$ (Section \ref{sec_snowdens} and specifically Equation \ref{eq_rho} with continuous values for $\rho_s$), any comparison can be made by using Equation \ref{eq_hs} which determines the relationship between the three variables.**

We realize that the mentioning of 'SWE cylinder' and 'SWE observations' can be confusing for the reader. We decided to replace the word '**SWE cylinder**' with '**snow core sampler**' in this manuscript.

Also all minor comments by this reviewer were addressed. In addition, we did some minor technical (language) corrections in the text and figures. All of them are tracked in the 'difference' PDF file.