MOSAiC SnowModel

Reviewer 2

In this paper, the authors combine the MOSAiC observational time series with two one-dimensional models, one for the snow and one for the ice, to bridge gaps in the time series and to complete one full annual cycle. The produced data set of SWE, snow density, snow and ice thickness can be interesting for other applications and from the analysis some (first) conclusions can be drawn about the impact of snow deformation on the snow depth evolution and on how snow depth heterogeneity influences heat transfer through the ice.

While the paper is clearly written in the sense that you can easily read and follow the sentences, I had often difficulties to find out what EXACTLY has been done and why. This is reflected in the numerous comments you will find below, which also might suggest that a final consistency check/internal review would have been useful and should not be task of the reviewers...

My main concerns are listed below, while individual comments follow after that:

1. The authors stress how important the initial conditions (for freeze up/snow accumulation start date and ice thickness) are for the simulations but the freeze up date is determined inconsistently. For the first ice type the 10m (why not 2m?) 3 hourly air temperature is used, while for the second ice type the 3-day running mean is used. However, this is not stated or discussed anywhere, I inferred it from the figure (Fig. 4). Where the assumed initial ice thicknesses, especially for the second ice type (10cm), comes from is not clear. This combined with the uncertainty from using reanalysis precipitation multiplied by 2.13 (?, see 3. point below) for the first 2.5 months of the simulations should be more appreciated as a bigger uncertainty of the produced time series.

2. A linear fit to the observed snow density data is assumed, which I am not convinced of. The implications and resulting uncertainties of this assumption should at least be discussed more extensively.

3. To drive the snow model with atmospheric data especially in the initial and final phase of the simulations, the authors use MERRA-2 reanalysis data. The precipitation from MERRA-2 is multiplied by 2.13 to do so and the authors claim that this is the result from comparing in situ precipitation observations with the reanalysis data were available, but the comparison is not presented, the high value of 2.13 is not discussed anywhere, and the precipitation in the time frames when only the reanalysis data is used is considerably higher than during the observation times (see Fig. 6c and my comments for Sect. 3.7 and Tab.1)! This raises some questions which are not addressed at all.

4. I suspect that something has been mixed up in Sect. 5.1 and 5.2. My guess would be that 5.2 should be given before 5.1, otherwise I would have even more difficulties to understand what has been done here (see the specific comments below). This caused lots of confusion. Because my comments would be different if or if not this is the case, I prefer this to be checked first. Anyways, I would prefer to see what the results for the simulations are with and also without the assimilations. If I understand it correctly, this is only shown for SWE in Fig. 8, and then the accordingly assimilated SWE is shown in Fig. 7a (?). I assume that in Fig. 7b and c only the assimilated snow density and snow depth are shown? It may be useful to also discuss what happens in the 'model world' when the model is dragged to match the observations (maybe strange things happen when the model is forced to use values that are not consistent within the model physics?)

5. Different numbers of observations are used at different locations of the paper, and I cannot always follow where these come from.

6. The authors claim that their analysis confirms the good quality of the precipitation data because the difference between the simulated and observed snow evolution is (relatively) small and can partly be explained by deformation. I would suggest to state more clearly that this assumption (if I understood this correctly) is mainly drawn because a correlation between the difference between simulated and observed snow evolution/SWE (which is hypothesized to be attributed to deformation) and total deformation obtained from buoy position measurements is found ($r2=0.58$). This chain of arguments should be stated more clearly (if this is what the authors meant).

7. The authors should differentiate more on what are outcomes from the MOSAiC observational timeseries (published already earlier) and the NEW insights due to combining the observations with models. I suspect that parts of the conclusions contain information related to the first and not the latter one (see specific comments below).

(Specific comments are listed just in the response.)

Author’s Comment 2

Dear referee,

Thank you very much for your review, comments and suggestions. All your comments will be addressed, references will be improved, and the simulation data and code will be provided / published accordingly in the revision.

Response to the main concerns:

1. Uncertainty of the initial conditions (freeze-up date and thickness) and the atmospheric conditions: All the temperature observational data in the paper are 10-m temperatures. MicroMet model is used to simulate the 2-m temperatures that are used further in the observations. We will check that this is consistently explained in the paper. The reason the different temperature criteria choice for the deformed and ponded SYI is the nature of the freezing point. The first has a fresh surface without any contact with the sea ice water and we assume it freezes at zero. The second is salty and connected to the sea water; we tracked that with the 3-day running mean. This will be better explained in the revision. The importance of the atmospheric forcing bias will also be better explained in the revision (see point 3 here).

2. Linear trend in snow density: using a linear trend is the most simple approximation. We believe this is sufficient for this study. Potential biases will be discussed briefly. For a more complex fit a separate analysis (paper) could be done, but we believe that a spatially-distributed approach would be necessary. It is also questionable if the spatial distribution of the MOSAiC snow pits is sufficient for such a study based solely on observations. Potentially a distributed run of a data/model fusion tool would be helpful, but this also is beyond the scope of our initial study here.

3. Atmospheric precipitation bias correction: we will discuss our~~this~~ bias correction and provide references to published literature on this problem. First, we will discuss how the 2.13 precipitation correction was obtained. There were two steps to the precipitation correction we implemented. 1) the “drizzle clipping” described by Liston et al. (2020) was performed, but without the mass-

balance correction. Then 2) the bias correction described in the manuscript was implemented. This totaled a correction of 2.13. If the mas-balance correction is removed from the bias correction, then the precipitation bias correction is 1.01. This will be clarified in the manuscript. Then, we will discuss this relative to the first assumption (line 62) about the high quality of the precipitation data collected in-situ at MOSAiC (Matrosov et al, 2022) relative to the MERRA-2 precipitation forcing. In particular with respect to the bias correction of reanalysis in early winter prior to MOSAiC in-situ observations (August and September): it is true that the difference between snowfall and SWE on the sea ice are forming about 1/3 to 1/2 of accumulated difference (see Figure 8, specially for the Nloop case), but 1) our bias correction results in realistic snow depth, snow density and ice types for all ice types, 2) even larger difference in snowfall and SWE occurs from February to May when the atmospheric forcing is exclusively from MOSAiC weather observations. Based on the early winter difference between snowfall and SWE on sea ice we give recommendations on earlier onset of seasonal observations in future experiments.

4. Sequence of SWE calculations: The sequence of sections of 5.1 and 5.2 corresponds to the sequence in the simulation. What will be pointed out explicitly in the revision is that procedure described under 5.1. is done again after the introduction of the sink described under 5.2. We will consider making Section 5.2 a subsection 5.1. All simulations results that the reader requires are already shown on Figure 8. There is the snowfall (where no SnowModel mass balance sinks have been used), then there is ‘model without D’ (simulation after procedure described in 5.1 is applied for the first time) and finally ‘model with D’ (simulation after procedure described in 5.1 is applied for the second time).

5. Problematic observation numbers: Yes, the observation number (for example Nloop) are not the same in all cases. We will check that the observation numbers are correct and explain better why they are not identical in all the cases. The other referee also had concerned about this, so the paper obviously needs improvements here.

6. Justification of good quality precipitation data: The explanation here is in line with our message, which will be improved correspondingly.

7. Differentiation between what is based on the old observational paper and what is based on the simulations here: We will check this and make a clear difference. This paper also brings some new insights about the observations.

Response to the specific comments (in blue)**:**

l. 10: 'D ... and was at times as high as 10% of all winter snowfall' -> When only reading the abstract, this sentence is hard to understand correctly. Maybe something like 'deformation appears to contribute/explain 10% of ...'.

This is good suggestion and it will be applied.

l. 28: roles of snow... has

Will be corrected.

l. 45: I suggest: ... collected once a week, the following caused interruptions of up to 2.5 months: list of things -> would be much easier to read. What do you mean with 'early fall MOSAiC ship arrival' and 'summer ship departure'? Sounds like the starting and ending point of the time series? Do you mean that there is a gap because the time series does not cover the full 12 months?

This part will be rewritten for clarity.

l. 72: 'This implies that no snow had accumulated on it during summer.' -> Do you mean: we do not know whether snow has accumulated on the ice DURING summer but at the end of summer all precipitated snow had been transformed to ice/frozen melt ponds?

Yes, this will be rephrased.

l. 84: that -> which (also elsewhere)? and 'are not generally' -> 'are generally not'?

This will be rephrased.

l. 139: '...was 0.5 m thick on 1 August': Ok, you argue that October ice thickness is a good estimate for end-of-summer ice thickness, does this mean that 0.5m ice thickness was the mean/modal/? ice thickness as measured on this ice type in October? (I had to speculate as this is not explicitly stated here)

All ice thickness are modal, this will be explained.

Fig. 3, legend: 'Runwy'

Will be corrected.

l. 140: survided

Will be corrected.

l. 143: Why did you use 10 m air temperature from reanalysis? (and not 2m as it would be available for reanalysis data) I would state already here what reanalysis you used (and not only refer to section 3.7)

We prefer MicroMet to simulate 2-m air temperature. MicroMet is also used to simulate other boundary layer variables that are consistent to each other. We will double-proof that this is obvious in the paper.

Caption of Fig. 4:

1. the caption should include what temperature time series is shown (biased corrected MERRA2 reanalysis)

These are observed MOSAiC temperatures. Only a fraction of the data is from reanalysis (black lines).

2. sea ice water -> sea water?

Will be corrected.

3. 'The dates when air temperature AND its running mean depart continuously from the freezing temperature...' -> inconsistent/confusing choice of the two dates (blue and purple vertical bars): if you choose to set the dates according to the running mean, the second date agrees with this definition but the first date is not detectable from the shown time series (running mean is below the

threshold for the whole tine series shown here). However, when making the choice based on the air temperature series (instead of the running mean), the first date agrees with this definition, but the second date would be shifted to a later day (and from the figure it would not be obvious whether it should be even later or not). Or if you meant to refer to this difference by using 'respectively' in the end, this is not clear and not distinguishable for the reader and would in addition require further explanation why these different definitions are used.

This part will be rewritten according to the explanation under major concern 1.

l. 146/150: Where do these assumptions (0.1 m/0.05m thickness) come from, especially the 10cm assumption?

These are the minimal sea ice thickness that HIGTSI can simulate. This will be explained.

l. 147: bellow

Will be corrected.

Section 3.2 + 3.3: Maybe it would be better to combine these sections as these paragraphs are a bit confusing because they jump back and forth from the different measuring methods (transects, drilling sites, stake sites) without clear transitions.

We will modify this section to better describe what was done.

l. 156: 'no Magnaprobe snow depth measurements were used after mid-July' -> more precisely (I guess): after mid-July 2020 (i.e. the final phase of MOSAiC)

Will be corrected. Similar will be also done for other snow depth measurements (coring site).

Section 3.4, l. 174-192.: the jumping back and forth between the different methods to determine snow density makes this section harder to read 1. you mention all three methods (ok), then you write something about the cutters, then about the SMP, then about the SWE cylinders, then again the cutters, the SMP, the SWE (I suggest to go through this cycle once, this also avoids unnecessary repetitions).

We will modify this section to better describe what was done.

l. 183: why 'still'?

Will be removed.

l. 219: ' In addition, the final sampling on Snow 1 provided deeper and denser snow values.' -> meaning what?

Explanation will be added.

Fig. 5:

1. you give N=591 but maybe you could add in the caption the numbers for the three different methods such that the reader does not have to search them from the text.

We will modify this to better describe these details.

2. fit line for cylinder measurements: in spring 2020 there seem to be enough measurements (marked by crosses) but all of them are far below the fit line. It is hard to judge from this figure (with all the others points for the other two measurement methods) whether the fit is mathematically correct (which it probably is) but it also shows very clearly that a linear fit as calculated here is not a very feasible assumption for these data points... (which should at least be mentioned somewhere)

We discuss the spatial representation problem of these measurements (due to restricted access during break up) and we will make this information more clear.

3. equation rho_s = 0.22 x is given without units (also in the text description)

The units will be added (kg/m^3).

Section 3.5: for the other sections you already show some numbers/results, here not. Are the shaded areas, e.g. in Fig. 6, determined here?

We will move the shared area definitions (and potentially some deformation description) to this section. This is a good idea!

Section 3.6 + Fig. 1: Is the 'buoy' given in Fig. 1 the buoy 2019I3 mentioned in the text? For me, from the description it is not clear why there are different drift trajectories and what the different trajectories are: So, in the figure, the trajectories marked as MOSAiC

CO1/CO2 are from the ship positions? and they are almost identical to the buoy, right? Reading the text, I would assume that the back trajectory model is only used to extend the ship's or the buoy's trajectory at the beginning and the end, but the figure suggests that this is a separately derived drift trajectory for the whole time series. Wouldn't it be better to use the buoy's trajectory where available and only to add the beginning and the end? Also, the beginning of the trajectory looks strange...

This figure shows how very similar is the trajectory of the coarse passive microwave satellite-based (PMW) sea ice drift that was used to extract the atmospheric reanalysis data from August 2019 to August 2020.The distance is never far enough to justify atmospheric forcing from a different MERRA-2 model grid. The beginning has no buoy as none was deployed and only 'strange PWM drift is shown'. This will be better explained.

Section 3.7 and Tab. 1: It would be interesting to see the MERRA-2 data compared to the observations. The authors multiplied the water equivalent precipitation from MERRA-2 with 2.13 and (as a result?) in the time sections where the reanalysis data is used, the precipitation is considerably higher than in the remaining time: increasing from 0 to 5m within less than 2.5 months and later by approx. 3m in 2.5 months (of reanalysis data), while the cumulative precipitation increased by only approx. 4m in the remaining more than 7 months (of observation data). The reader should a least have a possibility to retrace whether this is realistic and where this comes from and what uncertainties are involved.

We will address this as suggested under major concern 3.

l. 260: I suggest to write: 'In this paper, only ocean surface heat fluxes derived from SIMBA buoys by Lei et al. (2022) were used.

We will modify this to better communicate the information.

l. 275: SWE is calculated -> the change in SWE (with time) is calculated

We will modify this to better communicate the information.
l. 280: the units ... is -> the unit ... is / units ... are

We will modify this to better communicate the information.

l. 304: 'any and all'?

We will modify this to better communicate the information.

Section 5.1: The description in this section could be clearer. I am also not convinced that a linear fit to the snow density values (Fig. 5) is the best choice to describe the observed snow densities and whether this fit is something you would want to use to correct/assimilate the model. Again, I would like to see how the model and the observations (e.g. of snow density) differ when the model is run without assimilations? And if there is large differences the reasons for this should be discussed somewhere. And what happens in the model when snow density and thickness are just changed (assimilated) without physical reasons (in the model's world).

Implementing more complex methods to describe and apply the observed snow densities is not justified in light of the general goals and purpose of this paper. A detailed analysis of the differences between the modeled and observed density values could be done in another paper focusing specifically on snow density observations and simulations. In general, SnowModel's data assimilation and data-model fusion codes never "just change" or ingest state variables like snow depth and density; the model's codes have been developed so they do not compromise the physics that is driving the systems of interest.

Fig. 7: Do I understand correctly that in the scatter plots the assimilated model values are compared with the observations (and thus are located (almost exactly) on the 1:1 line and only because the model is not assimilated for the melt season, there is some scatter points where the model and the observations differ (which then lead to r2 values not equal to 1.0). I am not sure how much sense these scatter plots make under these circumstances...?

Yes, you are correct. This figure clearly shows two things: that our assimilation methodology works and was done appropriately, and that the model physics realistically describes what happened during the melt period when the observations were not assimilated. This understanding is critical for the reader to accept the general findings of the paper.

I guess the circles are not only 'assimilated values' (purple) and 'melt period values' (grey) but could be marked as 'observed values' (which are used for assimilating the model (purple) or not (grey, melt season)).

Yes, these circles are observed values. We will modify this to better communicate the information.

l. 310: 'the final SWE evolution defined above was further modified' -> what is the 'final' evolution and were 'above' is it defined and what do you mean with 'further' modified, what is the first modification?

This will be resolved as described under major concern 4.

l. 320: Snow depths were created using eq. 2 (or 3?) and the assimilated SWE and snow density values, but they were also (on top) assimilated with observed snow depth values?

Both equations (finally 3, of course). We will add the information about assimilation.

Section 5.2.: The modifications to SWE described here are they meant in l. 310?

This will be resolved as explained under the major concern 4.

Fig. 8: Caption does not explain all markers used in the figure. It is confusing that the SWE observations are (blue and gray) crosses (I am guessing, it is not explicitly stated in the caption) while the other crosses are the derivatives. Why are the absolute precipitation values so different for the three ice types? Are they only shown qualitatively, do they refer to the values on the y-axis?

The markers are explained on the legend. The accumulated precipitation differs by the start of the accumulation date.

l. 339: strange sentence: snow depth variability ... snow-depth variability?

Will be corrected.

l. 340 f.: Did you come up with the idea to use mean snow depth minus 1 std or has it been used elsewhere earlier? Is this related to the fact that the effective thermal conductivity of snow (partly due to snow depth being variable instead of one mean value) would be much higher than the thermal conductivity (point) measurements that you are using here? (edit: you make this connection much later in the paper)

However, if only 16% of the data show this or a lower value, it sounds like a quite small value to be used for the simulations? Maybe this is more like 'tuning' the ice thickness model (for Nloop and Sloop) and there are other reasons why the simulated ice is otherwise thinner than the observed one?

We will make an obvious statement this is a 1-D idealized study and that the distributed study with uneven snow depth is required to understand this further.

l. 378: than than

Will be corrected.

For me, what is written in Section 6.1 is a summary with some conclusions but not a discussion...

We will shorten this section (some facts from the introduction are repeated here) and we will refer to alternative solutions (e.g. new expeditions after MOSAiC).

l. 394: 'SnowModel-LG reproduced the observed snow evolution' because it was assimilated

Will be corrected.

l. 400: 'These SSS and SBS values were about three times as large as in Liston et al. (2020)' -> or maybe this is related to the precipitation in the reanalysis being multiplied by 2.13?

The reasons for this are discussed right after the statement until line 405. Again, we would like to point out that the atmospheric forcing was for most of the period from the MOSAiC weather observations and not form the reanalysis. Reviewer is right at the point that about 1/3 of the difference between precipitation and model without D originates from the period in September (with large snowfall), but ever larger difference appears during spring snowfalls (from February on) when no reanalysis forcing is used.

l.414/5: This could be mentioned already on l. 327 to explain why you set D to zero for these time periods.

We will clarify this assumption on lines 326 and 327 where we first discuss setting D to zero. For example:
‘
Before and after these dates *D*was set to zero, under the assumption that no blowing snow occurs when air temperatures are above freezing, and the snow may be melting (e.g., Li and Pomeroy 1997).
’

l.416 affect -> effect + I suggest to phrase this more precisely: deformation contributed to ... % of ...

This will be rewritten.

l.426: between ... to ... -> between ... and ... + Why is there only n=16 data points for Nloop and Sloop and only n=2 for Runway? (in Fig. 7 there are more...)

This will be explained as in major concern 5.

l.432/3: 'snow redistribution ... was able to be detected' -> 'we were able to ...' or 'it was possible to...' or '...could be detected'?

This will be rewritten.

l.432/3: 'Our findings confirm that the precipitation observations have sufficiently low errors that a minor signal such as snow redistribution due to sea ice deformation was able to be detected' -> With 'our findings confirm' do you mean the high correlation between the time derivative of D and the total deformation? If so, please write it like that. If not, please elaborate more on why you think the findings (which?) confirm the accuracy of the precipitation data. Because otherwise I would think that the higher the uncertainty/error in the precipitation data, the higher the fraction of snow that would have to be (erroneously) attributed to deformation in order to make the simulations and the observations match.

This will be addressed as listed in major concern 6.

l. 461: 'a known phenomena' -> phenomenon

Will be corrected.

l. 480: check sentence: for example, what is 'younger ice thickness' + 'the oldest and thickness ice'

This will be rewritten.

Fig. 11: Why are there only 3 data points used for Runway? Why are the ice thicknesses from the coring site and the Ridge Ranch representative enough to be included in the comparison in Fig.9 but not here? If they were included, the correlation would be smaller because these observations show clearly thinner ice than the simulations with snow depth = level ice mean minus 1 std.

This is discussed elsewhere in the paper: There are 3 different ages of FYI, each resulting in different SWE and snow depth. Here the simulation is only done for the Runway (that has enough measurements to provide standard deviation). Other measurements (from stakes and coring) are provided to give a supporting information, but can not be used for quantitative analysis. Again, one of the weaknesses of the MOSAiC dataset is poor sampling of the FYI (due to logistic restrictions). After re-reading of the paper we will decide of this information needs to be more highlighted to the reader.

l. 499: Why not write 'Our bedform parametrization (using the level ice mean snow depth minus one standard deviation),...' instead of 'Our bedform parametrization, as described in Section 5.3,...'? Easier for the reader, only a few letters more.

This will be rewritten.

l. 499-505: Interesting idea, but I think a more detailed and more comprehensive analysis focusing on this aspect should be conducted before any recommendations can be given of what works better (using a different thermal conductivity value vs. a different snow thickness value (and especially which one))

This is a comment about the 3-d heat fluxed through the ice. On this point we disagree with the reviewer as this is the fourth publication the idea is raised. If reviewer insists, we can remove this part.

l. 500: that -> than

Will be corrected.

l. 507: citation not correctly embedded

Katlein et al, 2020 is correctly embedded in the PDF we submitted, be will pay attention to this and other citations.

l. 509: I assume you want to refer to Section 3.1 also for the initial thicknesses.

Correct, reference to section 2.1 will be removed.

Section 6.4: I would suggest to use consequently 'the simulated snow/ snow density/SWE/whatever' to make it clearer when you write about the simulated results in contrast to observations

Will be rewritten.

l. 531 f: 'All [simulated, see comment above] snow melted by 8 July, which fits well with the transect observations ... On level ice with snow reduced by one standard deviation, [simulated] snow was fully melted even earlier - about 3 weeks prior to the average snow cover. This coincides well with the estimates from eight thermistor chains deployed on level ice (Lei et al., 2022).' -> What does this mean? According to transect observations (i.e. real people on the ice) the ice melted on 8 July, according to (eight) thermistor chains this happened already three weeks later? Are eight thermistor chains only a subset of all (how many?) thermistor chains? Is this related to spatial heterogeneity/a different ice type?

This section will be rewritten with clarifications. Here we use thermistor chains used by other publications.

l. 544/5: What do you mean? There is more melting around thermistor chains?

Yes, this will be rewritten.

l. 556: Again, I guess you mean Sect. 3.1?

Yes, will be corrected.

l. 559-567: Are these results obtained from combining the model and the observations or are/can they (be) obtained from the observations? (and have been published before) The newly added information through your approach of combining the models with the observational time series should be made visible here.

The list of conclusions here is based on the simulations in this paper. Some findings co-incides with the findings from other MOSAiC publications. This will be explained on top of the list (before the conclusions are listed). What is important here that the simulations unite all these findings that are otherwise disconnected and scattered in several papers.

l. 579: is the parametrization really found in 5.1?

As of now, this is in section 5.2.

l. 594: allow -> allow for?

Will be corrected.

Thank you very much for your work and time. Your comments will greatly improve this manuscript.

On behalf of both co-authors,

Polona Itkin