# **Response to Referee Comments**

Referee comments are shown in black, our response is in <u>blue</u>, and changes to the manuscript are shown in red.

# Referee 1

The authors have made substantial changes with considerable improvements in the manuscript. I just have a few minor comments that should be considered.

Firstly, we would like to thank the referee for reviewing the manuscript again and for the thorough and thoughtful comments on our manuscript that they have provided during the peer review process. We appreciate the recognition of the improvement in quality of the manuscript. We have addressed the remaining comments as detailed below.

# Minor comments.

Abstract, line 20. "shifts 14 days earlier"

I'd suggest that you explicitly state here what time period this is relative to.

# Edit made as suggested.

Line 291. "no clear change over time"

Later in the paragraph, you mention that there is a clear change over time (with an increase in growth in the "always pack" region from the 1980s to the 2010s). Please either remove or revise this "no clear change over time" statement.

Thank you for identifying the above oversight. The comment has been removed, as suggested.

Line 304. "Figure 8 shows a large reduction in congelation growth"

I believe that this is referring to the change in 2040. This change looks quite small, especially in the "becomes MIZ" region. Perhaps remove the word "large" here or revise in some other way.

Agreed, the word 'large' has been removed as suggested.

Line 455-456 "it does not account for any ocean warming that we might expect to see in the 2040s"

This statement could be confusing since ocean warming is possible in your simulations from changes in surface fluxes. This is stated further in the paragraph. To clarify for the reader, I'd suggest revising this statement to better clarify that it is specific to ocean warming associated with changing ocean heat flux convergence.

A caveat to address the above point has been added to the text:

'it does not account for any ocean warming that we might expect to see in the 2040s beyond that captured in the model via changes to atmospheric surface fluxes'

Line 485-486. "There is no significant difference over time or between the pack ice and MIZ in the processes of sea ice growth."

This statement seems misleading since you show that there is a difference between 1980 and 2010 in the congelation growth in the "always pack ice" region. Please revise.

Thank you for identifying the above oversight. The sentence referred to here has been replaced with the following:

'The main difference in the growth terms was a general increase in these terms for the 'always pack' region from the 1980s to the 2010s, likely resulting from the higher growth rates associated with thinner ice.'

# Referee 2

# Summary:

This is a modeling study striving to shed light on changes in sea ice volume melt as the Arctic sea ice cover shifts from a more compact state to a state with a more open ice cover for which processes such as lateral melt or advection into warmer water can play an enhanced role. The authors use an improved version of CICE forced by two different atmospheric models (one for present day and one for a seamless investigation of present day conditions and future projections) for their study. They do a consistency check of the model forcing by means of discussing the forcing itself and by comparing resulting model sea ice quantities such as timeseries of the total sea ice extent and volume against independent data. They investigate and discuss the temporal changes of ice volume melt for three regions defined by means of different temporal development of their characteristic sea ice concentration. They investigate processes such as ice growth, ice melt (top, basal and lateral), and dynamic processes and provide a quite sound discussion of their results.

I have mixed feelings with this manuscript. On the one hand it seems to be well written and many aspects of the study are laid out very well, illustrations of results are quite comprehensive as is the discussion. However, I have a few question about why certains things were done as they were done; I am not sure whether dynamic processes have been investigated as thorough as it might have been needed (also in terms how good the atmospheric forcing is in this regard); I am also not sure whether the findings are relevant in view of the substantial differences in the forcings of the two models used for the atmospheric forcing and I hence do not have a clear feeling about the uncertainty of the results shown. Finally, I find some expressions and definitions as written could lead to mis-understandings and mis-interpretations of the results shown and discussed.

Firstly, we would like to thank the referee for providing thorough and thoughtful comments on our manuscript. They have been invaluable in improving the manuscript. We have addressed the concerns raised above in our response to specific comments.

# **General Comments**

GC1: I can understand that you applied the easier-to-use definition of the MIZ. My question is, however: How close is this definition to the one based on ocean waves and what would be the difference you'd expect in case you could use the wave-based definition? I am concerned about that. I thought that, in the meantime, the community has moved away from defining the MIZ as the area with sea-ice concentrations below 80%. It has been shown that this sea-ice concentration threshold is not adequate to approximate the part of the sea ice cover that is

influenced by waves. I was wondering whether the strength of the sea ice cover, i.e. discriminating between compact pack ice and freely drifting ice wouldn't be a more natural way to define the two different sea ice areas used in your study. I would avoid to term this MIZ or marginal ice zone given the vast extent the MIZ as defined by you has.

Firstly, both the concentration-based and wave-based definition of the MIZ continue to be used in the literature. It is not true to say that the community has moved away from using the concentration-based definition (see e.g. Concetta et al., 2024; Strong et al., 2024; and our own previous papers on this topic). We think it would be misleading to the community to avoid the use of the term MIZ when considering a concentration-based definition, as this would be going against several decades of accepted usage. We apply the concentration-based definition in this study since waves are not the critical factor in determining sea ice mass balance in the Arctic (Bateson et al., 2020; Bateson, 2021). We agree that the wave-based and concentration-based definitions of the MIZ do likely refer to different regions of the sea ice (Horvat et al., 2020). It is therefore important to be clear in the definition of the MIZ being used in this study, and this is indeed something that we have stated clearly in the opening few sentences of the introduction.

We have updated the text with the more recent references included above. We have included an additional sentence in the first paragraph of the introduction to address some of the above concerns:

'Whilst the two definitions of the MIZ likely refer to different regions of the sea ice (Horvat et al., 2020), we apply the concentration-based definition in this study since waves are not the critical factor in determining sea ice mass balance in the Arctic (Bateson et al., 2020; Bateson, 2021).'

GC2: Closely connected to GC1 is the question: Did you check how your results would change if you would change the threshold from 80% to 60% or to 90% - independent of the ice strength definition by Hibler?

A threshold of 80% (corresponding to the upper limit of the concentration-defined MIZ) is both a well-established threshold for distinguishing between regions of sea ice in the literature (as noted above and in the manuscript) and is physically motivated since it approximately corresponds to the transition from ice in free drift to pack ice. We have previously considered applying different thresholds, however there are also practical reasons why we selected a threshold of 80%, in addition to the physical motivation and widely-accepted usage. If we increase the threshold to 90%, much of the central pack region will be identified as MIZ due to the opening of leads. If we reduce the threshold to 60% or even 70%, the 'always MIZ' region over the 1980s to 2010s period will be too small to allow any useful analysis. Figure 6 already shows the limited extent of this region even for a threshold of 80% (see the orange areas in panels c and d).

We have included the following comments in section 2.3 of the manuscript to provide more details of our choice of threshold:

'We selected a threshold of 80% for our analysis because it is both a well-established threshold for distinguishing between regions of sea ice in the literature (i.e. the upper limit of the concentration-defined MIZ) and is physically motivated since it approximately corresponds to the transition from ice in free drift to pack ice. In additional, alternative choices for this threshold have practical limitations e.g. for a higher threshold of 90%, much of the central pack region will be identified as MIZ due to the opening of leads. For a lower threshold e.g. 70%, too few grid cells are identified as part of the MIZ in the 1980s to allow useful analysis.' GC3: I am aware that you have expanded the work put into the initial version of this manuscript substantially. But given the importance of surface turbulent heat fluxes and the potential increasing role of sea ice transport from one region to another region, I was wondering whether you should not comment and/or justify why you did not also check NCEP2 and HadGEM2-ES model data with respect to wind speed and direction



**Figure A:** Average wind speed and vectors for NCEP in the 1980s (a), 2010s (b) and for HadGEM2-ES in the 1980s (d), 2010s (e), and 2040s (f). Averages are taken over June to August for ocean north of 66.5°N over the 5-year study periods as outlined in Section 2.3.

We agree that wind speed is an additional important component of the atmospheric forcing to consider in understanding the changes to the sea ice mass balance budget that we find in this study. In an updated Fig. 2 (see the revised manuscript), we have included an additional panel (e) showing the mean wind speed for the atmospheric forcings for each of the time periods considered. This figure shows high daily variability in each case, with no clear trends in wind speed from the 1980s to the 2010s/2040s for either of the forcing data sets. The magnitude of the wind speed is, however, consistently higher for the NCEP forcing compared to the HadGEM2-ES forcing. Since it is not possible to produce an analogous, meaningful comparison for wind direction, we have instead included Fig. A (above) showing map plots of average wind speed and vectors for both forcing data sets for each of the time periods considered in the study averaged over the summer months. The changes in wind speed are consistent with the updated Fig. 2, i.e. consistently higher wind speeds can be seen for NCEP compared to HadGEM2-ES, with no clear trends over time for either dataset. Whilst there are more distinct changes in the

pattern of wind vectors, the high daily variability illustrated in Fig. 2 will also apply to wind direction. As such, interpreting the impacts of these changes requires a much more involved analysis, as the precise impact of any change in wind speed or direction will depend on the seasonality of these changes and whether the change only acts to redistribute sea ice within the pack ice or MIZ, or between these two regions (and of course export from the MIZ to open ocean). It is worth noting, however, that despite the different patterns of, and changes in, wind vectors, Figs 8 and 9 show that both products produce a comparable reduction in the contribution of dynamics to sea ice volume loss. In addition, whilst the NCEP forcing does have higher wind speeds, a larger dynamics contribution to sea ice volume loss for the NCEP forced simulation compared to the Hadgem2-ES forced simulation can only be found in the always pack region, particularly in the 1980s, with only small differences found for the other regions. As discussed above, we want to maintain focus in this study on understanding the changes in thermodynamics due to the greater magnitude of these terms (as shown in Figs 8 and 9) compared to dynamics, not least since the changes in dynamics can at least partly be understood through the role of thermodynamics in producing thinner ice (as discussed in the manuscript). Whilst wind speeds are important for latent and sensible heat fluxes and therefore can also impact the thermodynamics terms, the lack of change in average wind speed over time shown by Fig. 2 suggests wind speeds are not a major driver for the changes in the thermodynamics terms over time identified in this study.

We have updated Fig. 2 to include a panel comparing average wind speed for the different forcing products over the different time periods considered. We have updated the relevant section to discuss the additional panel comparing wind speed in Fig. 2, and have also made references to this comparison of average wind speed where we discuss the dynamics term in section 3.2 (paragraph 4).

GC4: My impression is that a number of the "main" conclusions of the paper are based on comparably small changes in ice volume melt by the different processes investigated. What I am missing in this context is a critical discussion of what the error bar of the obtained results is. I don't get a clear picture of whether a change in, e.g., the contribution by lateral melt by 10% from the 1980s to the 2010s does not simply fall into the uncertainty range of the model results. One could argue that the differences you show between NCEP2 forced and HadGEM2-ES forced ice model runs are an indication for the uncertainty of your "physics-rich" model representing the processes shown. Emphasizing the uncertainty of the model results could solidify the credibility of your results.

We agree that different atmospheric forcing products will produce different results, indeed it would be surprising if these differences were not typically significant compared to those associated with sea ice physical process dependent on sea ice state (whether associated with the MIZ or not). A major motivation of producing hindcasts using two different products is to estimate how sensitive our results are to the forcing product used. This is why we account for both hindcasts when discussing our results and then use the understanding gained in analysing the results from the hindcasts when discussing the projection. Our results show systematic differences in the changes that we find from the 1980s to the 2010s and between the MIZ and pack ice, which suggests a significant result that is not peculiar to the choice of forcing product. In some cases, these differences are larger than the sensitivity to the choice of forcing product, particularly when comparing across different regions (e.g. the differences in basal melt volume between the 'always pack' region and 'always MIZ' region in both the 1980s and 2010s are much larger than the differences between the basal melt volume for the same region and time period

for the two different forcing products). However, even where the magnitude of these differences is comparable or smaller than the sensitivity to the choice of forcing product (particularly when considering changes over time rather than between regions), the consistency of these changes across both hindcasts provides confidence that they are robust results.

At the end of paragraph 4 of section 2.1, where we describe the two atmospheric forcings datasets, we have included the following motivation:

'In addition, the use of two different atmospheric forcing datasets allows us to estimate the sensitivity of the results to the forcing used.'

We have also provided some additional discussion in the final paragraph of section 4, where we discuss the advantages and disadvantages of using a forced sea ice model:

'Whilst individual atmospheric forcing products will still have an associated uncertainty, our approach of evaluating hindcasts using two different atmospheric forcing products allows us to determine the sensitivity of our results to the forcing used. We find that, despite the differences between the two atmospheric forcing products, there are systematic differences in the changes that we find from the 1980s to the 2010s and between the MIZ and pack ice. In some cases, these differences are larger than the sensitivity to the choice of forcing product, particularly when comparing across different regions. However, even where the magnitude of these differences is comparable or smaller than the sensitivity to the choice of forcing product (particularly when considering changes over time rather than between regions), the consistency of these changes across both hindcasts provides confidence that they are robust results.'

#### **Specific Comments**

L9/10: "We validate ..." --> After having read the paper I would say that you did not do a validation but you carried out a consistency check of the model. I suggest to change the writing accordingly throughout the manuscript.

#### We agree validate is too strong.

We have replaced validate with 'demonstrate the model is realistic' in the abstract, and have also replaced validate / validation elsewhere in the manuscript with suitable alternative phrasing.

L18-20: "As more ... days earlier" --> These lines can be misunderstood. I recommend to clarify that by "melts earlier" you mean that the sea ice is completely gone (melted) earlier. I am pretty sure that it is not the transition from > 80% SIC to < 80% SIC that causes an earlier melt-onset. By the same token I recommend to clearly state what you mean by "peak melt"; you are not referring to the largest forcing but you are referring to the largest change in ice volume due to a combination of different melt processes.

By 'melts earlier', we are referring to the results presented in Fig. 10 that show that for the 'becomes MIZ' region, there is overall shift earlier in the melt season in the later period compared to the earlier period for all three panels. This applies not just to the end of melting but also to melt onset and the peak melt fluxes, hence the behaviour being referred to here is not just a result of earlier melt out of the sea ice.

We have modified the relevant sentence to avoid any ambiguity in interpretation:

'For areas of sea ice that transition to being MIZ in summer, we find an earlier melt season: in the region that was pack ice in the 1980s and became MIZ in the 2010s, the peak in the total melt volume flux occurs 20(12) days earlier.'

L56: "location of ..." --> I can agree that the atmospheric forcing plays a large role here. But the way how melt ponds form, their depth and their size distribution depends largely on the sea ice topography and the amount of snow. Melt ponds on level ice can spread over large areas and might remain shallow for quite a while. In contrast, melt ponds on rough sea ice form (first) in local depressions of the surface. There have been quite a few studies about how the formation of melt ponds is governed by the snow depth and the sea ice surface topography - even hypothesizing that it is possible to use the sea ice topography and related snow depth distribution in spring as a means to predict the melt pond fraction during summer. I suggest you take this into account in your text.

# We agree with the above comments.

We have modified the relevant text to address the above comments:

'In the Arctic, the snow thickness is generally modest compared to that on Antarctic sea ice, and the location of top melting and the initial formation of surface melt ponds is primarily driven by atmospheric conditions with the subsequent evolution controlled by sea ice topography.'

#### L126: Why 2010? Why not 2017?

We selected 2010 as we assessed it to be a typical year for wave forcing. Also, as we discussed above (and also in previous rounds of reviews), there is a low model sensitivity to the wave forcing.

# We have noted in the manuscript that we assessed 2010 to be a typical year for wave forcing.

L132/133: I was wondering whether you would like to comment on your choice of forcing data sets. An obvious alternative for NCEP2 would have been ERA5 to be more in line with the wave climatology used.

We used NCEP2 for atmospheric forcing for the hindcast since NCEP models generally perform better for near-surface variables compared to other forcing datasets, including ERA-Interim (Jakobson et al., 2012). Studies comparing ERA5 and ERA-Interim concluded that both reanalyses had significant near-surface warm biases, with ERA5 demonstrating particularly strong biases in winter, substantially reducing sea ice growth (Wang et al., 2019). We used ERA-Interim for the wave forcing since NCEP2 does not include wave output and ERA-Interim performs favourably in comparison to other reanalyses in simulating wind speeds in the Arctic (Jakobson et al., 2012; deBoer et al., 2014).

# We have provided justification in our choice of forcing data sets in the relevant paragraphs of section 2.1 of the manuscript.

And I was wondering what kind of a model HadGEM2-ES is. It does not belong to the CMIP6 model suite it seems? A bit more detail would be great. In addition: What is your motivation to use RCP8.5?

HadGEM2-ES is one of the Met Office Unified Model climate configurations produced for CMIP5 (Martin et al., 2011). We selected the RCP8.5 pathway since this will produce the maximum possible change in the sea ice state by the end of our study period, best enabling us to explore

the implications of a higher MIZ fraction in the future Arctic for the different processes that contribute to sea ice mass balance.

# We have provided a reference for HadGEM2-ES within the manuscript, and also provided a justification for using RCP8.5.

And, furthermore: The time period you cover here is 1980 through 2050. This differs from what I would expect from CMIP5 or CMIP6 model runs where there are historical runs that end in 2014 for CMIP6 with predictions following thereafter. To my understanding there is no seamless transition between the historical runs and the future predictions, or - in other words - one should not use them together. Therefore I am a bit puzzled here.

We are particularly interested in changes in the sea ice state from 1980 to 2050 since simulations suggest this period is critical in terms of Arctic and MIZ change. To achieve this, we have to include projected forcing to drive our forced ice-ocean model and to do this we use CMIP forcing (note that we are using a CMIP5 model where the transition is earlier in the 2010s). Note that we are not trying to replicate a CMIP simulation in this study. The initial conditions for the projection are taken from the historical run, so whilst there is a transition in the early 2010s in terms of how the CMIP forcing is constructed, it is nevertheless a continuous dataset. Since we only include 2015-2019 in our average, we avoid this transition.

L153/154: "A full discussion ... 2017)." --> I am not sure how much that publication also looks into differences of these two SIC products during summer - which appears to be the main focus of your study. I invite you to look into papers published in 2020 in the journal "The Cryosphere" about sea ice concentration product evaluation during Arctic summer conditions to obtain additional information about the performance and potential biases of these products during summer. Doing so will also help with what you call "Model validation" and the discussion of the results later in the paper.

# We have provided an additional reference, Kern et al. (2019), within the manuscript to a study that compared the two SIC produces during summer. As discussed above, we have also changed the terminology to avoid the term 'validation'.

L161-164: Please provide just a little bit of more context here so that the reader understands what you mean by "radar returns". I note that sea ice thickness has been estimated from satellite radar or laser altimeters. I note further that melt ponds on sea ice are one issue with these instruments but that already the wet snow conditions encountered during late spring inhibit an adequate sea ice thickness retrieval beyond, currently, April and before October.

Another issue I find important to mention in the context of your work is that PIOMAS provides consistent estimates of the sea ice thickness and volume but that it is well known that PIOMAS tends to overestimate the thickness of thin ice and underestimate the thickness of thick ice. I am pretty sure you will find respective literature about this. I invite you to clearly state this deficit of the PIOMAS model data and bring this issue back into your paper when you are talking about the "Model validation" and when you are discussing your results.

As a general response to the above comment (and also the previous comment for L153-154), our aim in using the SIC products and PIOMAS is to ensure we are simulating a realistic sea ice state rather than an accurate sea ice state. This is sufficient for our purposes in assessing the contribution of different processes represented within the sea ice model to the total sea ice mass balance. As discussed in previous responses, we have changed the term 'model

validation' throughout the manuscript to better reflect this intent. We have also provided additional references in section 2.2 highlighting potential biases in these products.

In the relevant paragraph, we have noted the deficit of the PIOMAS model identified above and provided a reference for this (Schweiger et al., 2011). We have also included a comment about the impact of wet snow conditions in the relevant sentence in the manuscript and clarified what we mean by 'radar returns':

'In addition, satellite thickness products have not historically been produced during summer since both wet snow conditions that emerge during late spring and melt ponds interfere with satellite altimeter retrievals.'

L162: Finally, there is this "summmer" typo in L162.

This typo has been corrected.

L170: "In each case we use the last 5 years" --> Why? Why don't you use the full 10-year period? This is not clear and should be motivated.

Our choice in time period was motivated by balancing the need to capture the transient climatology of a changing system (better achieved using a smaller number of years) whilst not overrepresenting extreme events (better achieved by using more years). In addition, as noted above, there is a transition in the early 2010s in how the HadGEM2-ES is constructed. Using only the last 5 years means we avoid including this period of transition.

At the end of the relevant paragraph, we have included the following motivation for using the last 5 years of each decade:

'We use the final five years of each decade considered for our analysis to balance the need to capture the transient climatology of a changing system whilst not over-representing extreme events. In addition, this avoids including in the analysis a transition in the early 2010s for the HadGEM2-ES simulation where the model changes from a hindcast to a projection.'

L179: "annual volume fluxes" --> Not sure I understand what you mean by volume fluxes in this context. My understanding of a volume flux would be the flux of a certain volume of ice across a boundary, for instance the Fram Strait or from the Pacific to the Atlantic sector of the Arctic Ocean. Would you mind to specify a bit better what you mean?

In this context, volume flux refers to the rate of ice volume loss or gain from a given process.

#### We have included the above definition within the manuscript.

L182++: I note that here you explicitly mention the seasons during which this part of ice (and snow) volume change happens. You do not specify the seasons for most of the other processes. Why?

To ensure consistency in definitions and given this information can easily be established via Fig. 10, we no longer describe the seasonality for specific terms.

I note further that you list sinks for snow (snow ice formation, melting, sublimation) but you do not include the source terms. Why? How about - in this context - snow blown off the sea ice into the openings?

As stated when introducing the referred to list in section 2.3, these are terms that directly impact the sea ice volume budget. We are specifically interested in processes that impact the sea ice volume, not snow volume. Hence, we do not include source terms for snow since they do not directly impact sea ice volume. Similarly, snow that is blown off from the sea ice does not directly impact sea ice volume. Instead, we capture the contribution of snow to sea ice volume via the 'snowice' term. The three melt terms (top, basal, and lateral) only include ice melt, since snow melt is calculated separately in CICE. The sublimation term does include contributions from both ice and snow because CICE model output does not separate these contributions. Fig. 8 shows that this term is very small and it does not form part of our later analysis and discussion.

We have modified the description of the sublimation term and included the following note at the end of the list:

'These terms collectively account for all sources and sinks of sea ice volume captured by our model setup within a given region.'

L189/190: "sum of advection, convergence and ridging" --> So, what you take into account here is whether and how much sea ice is advected (aka transported) from one region to another region and/or outside the MIZ region into open water. While you explicitly mention convergence you do not mention divergence and its linkage to new ice formation during the freezing season on the one hand and its influence on lateral melt during the melting season on the other hand. Why?

Since you only look at the sum of the three initially-mentioned processed you refrain from looking a their individual contributions to the budget. Why?

Ridging I would see as a consequence of convergent ice motion in case the sea ice is mechanically weak enough to fail. I note further that ridging does not change the ice volume in a region; it only changes the thickness distribution.

In summary, I am not sure whether the brevity of the description of the processes you want to consider here in the bullet point list is not leading to misunderstandings and I invite you to consider expanding on these a bit.

A final thing I would like to mention in this regard is that I would find it helpful to see some motivation about why you have been choosing these processes - why you, for instance, look at the ice volume that is associated with congelation growth or frazil growth.

As discussed above, our aim here is to include in our analysis all terms included within our model setup that directly impact sea ice volume. This is why the growth terms are included within our analysis. We have modified our description of the dynamics term to now reference divergence, although it should be noted that convergence and divergence are the same process with a different sign. We agree with the point made above that ridging does not directly change the ice volume within a given region and is instead a consequence of processes that do, and have therefore removed the reference to ridging. We only include the sum of the dynamic terms since convergence / divergence is a specific case of advection and not a distinct process.

As noted above, we have included a comment at the end of the description of processes noting that we are aiming to account for all sources and sinks of sea ice volume captured by our model setup. We have also updated the description of each process to ensure we are only describing

# the direct mechanism of how the sea ice volume is modified, rather than any indirect mechanisms. Finally, we have updated the description of dynamics as described above.

L193-205: In this paragraph you illustrate the atmospheric forcing. There are quite some differences between NCEP2 and HadGEM2-ES. While you list these differences you are refraining from commenting whether these differences are not eventually jeopardizing your research goal. I was wondering for instance, where the difference between NCEP2 and HadGEM2-ES air temperatures during winter has its origin. It cannot be the longwave radiation. I was wondering furthermore, whether the massive differences in the longwave radiation but also in the shortwave radiation between NCEP2 and HadGEM2-ES data for the same 5-year periods do not have a considerable influence on the ice model performance, i.e. how the processes are realized.

It is not in the scope of our study to analyse the differences between the two atmospheric forcing products. As discussed in our response to comment GC4, a major motivation for choosing these two atmospheric forcing products is to evaluate the sensitivity of our results to the forcing used. We find that, despite the differences between these two products, there are systematic differences in the changes that we find from the 1980s to the 2010s and between the MIZ and pack ice, strengthening the conclusions that we reach based on these changes.

As noted in our response to comment GC4, we have included additional details in both section 2.1 and section 4 of the manuscript to more clearly explain our motivation in using the two atmospheric forcing products and how this can strengthen our conclusions.

Finally, I note that you did not include surface wind data in your "model validation" - even though the dynamics and any surface processes that depend on the near surface turbulent fluxes of sensible and latent heat are dependent on the near surface wind. Isn't the near surface wind speed essential for a correct simulation of how the area of the different regions of pack ice, MIZ and "open water" changes? Both, the wind speed and the wind direction might be of importance here (see GC3)

We have addressed this comment through our response to GC3.

We have modified Fig. 2 in the manuscript to include a comparison of average wind speed and have discussed this additional comparison within the text of the manuscript. See our response to GC3 for details.

Fig. 6: (panels a&b) It is clear that both forcings lead to model runs that show a too fast SIE decrease at the beginning of summer and a too fast freeze-up in fall - when compared to the observations. (panels c&d) Very interesting to see that the HadGEM2-ES forced model runs follow the Bootstrap SIE only until July (in the 1980s), after that they "switch" to be closer to NASA-Team and even reach the same MIZ extent as NASA-Team, only a month later. Also in the 2010s, the HadGEM2-ES forced model runs are first closer to Bootstrap.

# We agree this is interesting, but not of central relevance to our manuscript.

L225-236: I would like to remind you of the potential mis-estimation of the SIC by the NASA-Team and Bootstrap algorithms during summer while melt ponds are present (see one of my previous comments).

As mentioned previously, we have now provided additional references where we first introduce the two SIC products (Kern et al., 2019, 2020) that provide a more complete discussion of the

differences between these two products, particularly during summer. We have also noted the contribution of melt ponds to the uncertainty in these SIC products.

L237-243: I was wondering whether it is sufficient for the main focus of your study to look into these annual time series of the total Arctic sea ice volume. How do the sea ice volumina time series look like if you separate pack ice and MIZ?

In this context I would like to make you aware one more time of the issue of the PIOMAS I mentioned earlier in my review.

As noted above, PIOMAS is known to overestimate the thickness of thinner ice and vice versa for thicker ice. Whilst the impact of these biases should be reduced over a pan-Arctic scale product due to averaging out, the impact will be larger when only considering regions that overrepresent thinner or thicker ice compared to the pan-Arctic average, which will be the case when considering the pack ice and MIZ separately.

L274-282: In the context of these results I would like to ask you whether you did consider at some point to have a look at the floe-size distribution and at the actual sea ice concentrations in the regions you are using. As you rightly say, top and basal melt primarily depend on the actual area of the sea ice that is subject to melt (here: top and basal) while lateral melt is primarily driven by the floe size distribution and only secondarily driven by the sea ice concentration.

For the 'becomes MIZ' region discussed in this paragraph, a reduction in sea ice concentration is implicit in how the region has been defined (i.e. it generally forms part of the pack ice in July in the 1980s, but is part of the MIZ in July in the 2010s). Whilst the above description of the dependencies of the different melt components is broadly true, more specifically lateral melt depends on the total floe perimeter, which is a function of the floe size distribution (FSD) and sea ice concentration. Noting this, and using simple geometry, it can be shown that the ratio of basal melt to lateral melt scales with the linearly weighted average floe size i.e. as floe size decreases, the relative volume of lateral melt compared to basal melt increases. The ratio of lateral to basal melt can therefore be used to indirectly infer changes in the FSD, which is the approach we have taken in the manuscript. Whilst we did look at maps of the perimeter weighted average floe size (which is the most useful metric when considering lateral melt rate), we decided not to include these in the manuscript due to the small size of the lateral melt term and, as discussed above, we are already able to infer the relevant changes in the FSD through the analysis presented.

L283-290: Ok, great. This is the first time that I learn in your paper that by "advection" you indeed mean the import or export of sea ice into or out of a specific region. Now I understand why the contribution of the dynamics is negative. I invite you to clarify this considerably earlier in your paper.

The definition of dynamics in the list in section 2.3 has been modified to provide additional clarity:

'the net import or export of sea ice within a given domain resulting from advection and convergence/divergence'.

L305-307: "The former ... 2040s" --> so you say that the warming (during winter) sets off the fact that continued thinning of the existing ice would allow for more rapid congelation ice growth and hence more ice volume to be formed?

#### Yes. See the response to the later comment for L375/376.

"whereas the latter ... sea ice formation" --> Here I was wondering how "good" CICE is when it comes to forming frazil ice. Isn't this a notoriously difficult ice type to grow? Can you perhaps refer to publications where it has been shown that the model is capable for this ice type?

CICE is routinely used to simulate frazil ice for both the Arctic and Antarctic, and produces frazil ice where water is identified as supercooled (Hunke et al., 2015). Keen et al. (2021) provides discussion of the differences in frazil ice production between different CMIP6 model and the reasons for these differences.

L350/351: "whilst the end of summer ... simulation." --> These are drastic shifts in the end of summer melt towards summer - this is per se in contradiction to studies looking at the melt onset, freeze-up and length of the melting season from observations. In this context I encourage you to back-up your results with existing literature about melt season length in the Arctic Ocean. Could it be that the cause for this shift simply is the lack of sea ice - aka the sea ice has melted out earlier? Otherwise it is not really clear why with elevated air temperatures and longwave heat fluxes the melt is supposed to stop earlier.

The referred to lines are discussing the timings of the melt season specifically for the 'becomes MIZ' region only, not the whole Arctic Ocean. Since this region has been specifically defined to capture locations that transition from being pack ice to MIZ in July from the 1980s to the 2010s, we expect this region to behave differently to the Arctic as a whole in terms of melt season timings. Indeed, we see very different results for these timings in the 'always pack' region. We agree that the earlier onset of the end of summer melting is a result of melt out of sea ice in this location in the 2010s.

We have modified the relevant paragraph to note that the earlier onset of the end of summer melting in the 'becomes MIZ' region results from melt out of sea ice in this region during summer in the 2010s.

L353: "will be driven by reduced sea ice mass balance" --> I suggest to reformulate. I guess the earlier timing "is driven" or "is associated" with partial complete melt-out of the sea ice, right? You mention this fo the "always MIZ" region further below in your text but it might also be the case here.

The point being made here by 'reduced sea ice mass balance' is that the total melt volume flux is related to the sea ice mass balance (e.g. lateral melt volume is proportional to sea ice thickness, and top melt volume is proportional to sea ice area), hence a reduction in sea ice mass balance can lead to reduced melt volume flux, even without there being complete melt-out of sea ice.

The relevant text has been modified to better explain the relationship between sea ice mass balance and melt fluxes. We have also added a comment about the specific role of melt out of sea ice in determining the end of summer melting (as noted in the previous response).

For my understanding: Peak melt is when there is a maximum of melt water flux, right? This date is not necessarily related to the peak in solar and/or longwave radiation?

The full phrasing used in the sentence referred to is 'peak melt rate', and it indeed refers to the maximum of the melt water flux as plotted in Fig. 10.

We have modified the third paragraph in section 3.3 to clarify the definition of peak melt rate.

L371/372: "The start of melting shifts earlier by 7 days ..." --> I don't agree for the always MIZ region where Fig. 10 clearly shows that melt begins mid May in 1980s, a bit later in the 2010s but as late as the beginning of June in the 2040s. I guess this disagreement has to do with you definition of the threshold (see L330)? If not, then there is some rewriting / explanation required here.

Subplot (h) in Fig. 10 shows that the melt season has shifted earlier for the 'always MIZ' region. Please note, the results presented in subplot (c) cannot directly be compared to the result in subplot (h) since the 'always MIZ' region is different for the two panels, as discussed in section 2.3 and shown in Fig. 1.

We have added an additional comment to the captions for Figs 8-10 to clarify the different region used for each panel in these figures.

Figure 8: I guess I would appreciate an explanation about why, if the dynamics do not include transport from one region to another you have a change in the sea ice volume (see my previous comment on this). Convergence and the potentially associated ridging just redistributes the ice mass but does not lead to a change in ice mass or volume.

The dynamic term includes transport (as indicated in the caption for Fig. 8).

As discussed in earlier responses, the description of the dynamics term in section 2.3 has been updated for clarity.

I find it surprizing that the fraction of the volume decrease by lateral melt is not increasing over time (from the 1980s to 2010s to 2040s), i.e. is not increasing with the increasing area covered by MIZ-type ice, hence more isolated floes.

The change in lateral melt volume over time (particularly relative to the other melt components) is something we also found worthy of note and have already addressed in the manuscript (for example, see paragraph 3 in section 3.2 and paragraph 4 of section 4).

I find it also surprizing that the fraction of frazil ice volume is not increasing with time - at least for the "becomes MIZ" and the "always MIZ" regions.

Figure 8 shows that the frazil ice volume is increasing relative to congelation growth (assuming this is what you mean here by fraction) in both the 'becomes MIZ' and 'always MIZ' regions. In 5/6 subplots for these two regions, the volume flux for congelation growth decreases whereas the volume flux for frazil ice either increases or shows negligible change. In subplot (b), whilst we do see an increase in congelation growth, we see a larger percentage increase in the frazil ice volume.

L375/376: "followed by slower ... December." --> Since this is counter-intuitive when taking into account that the sea ice underneath which the congelation growth occurs, is possibly considerably thinner, hence allowing faster growth. It might be good to mention that there is the tie between thinner ice but at the same time warmer temperatures?

We have added an additional sentence to end of the relevant paragraph to address the above point.

'For the becomes MIZ region, it might be expected that the average rate of congelation growth would increase from the 2010s to 2040s over October to December due to the reduced mean ice thickness, however in this case the changes that act to reduce congelation growth, e.g. the

increase in longwave radiation and surface air temperature over the autumn and winter months shown by Fig. 2, have a larger impact.'

L470++ Given the fact that the discussion section already partly repeats parts of the results section I find the concluding remarks a bit long. I was wondering whether you could condense it in view of what can be found in the results and discussion sections.

We have expanded the conclusion in response to previous reviewer comments.

I was wondering into which direction future research should go

We have edited the final section of the penultimate paragraph to better highlight the implications of this study for future research.

#### **Editorial Comments / Typos:**

L45-47: "Models studying ... 2021)." --> Please check this sentence and eventually split it into two. I find it difficult to understand the way written.

#### Sentence has been split into two, as suggested.

L58-60: "The SIC budget ... seasonal cycle." --> I don't understand this sentence. I don't understand in particular what you mean by "SIC budget" ... as far as I know sea ice concentrations have been observed by satellite sensors since the 1970s .... so why are SIC budgets constructed?

The text has been updated to provide a clearer definition of what a SIC budget is:

'The purpose of a SIC budget is to evaluate the relative contribution of thermodynamic and dynamic processes to the seasonal cycle of is concentration. The SIC budget from observations has been constructed for the Arctic by Holland and Kimura (2016) using AMSR-E satellite observations spanning 2003-2010.'

There is a typo in the sentence: It needs to read: "AMSR-E"

#### Typo has been corrected.

L71: "growth reducing in autumn" --> Could it be that you wanted to say "delayed freeze-up" or the like? This would fit better to the "melting happening earlier"

The phrasing here is intentional. The full wording used here is 'growth reducing in autumn and increasing in winter' i.e. a contrast is being made between the behaviour in autumn and the behaviour in winter.

L98: "calibration to Cryosat-2 data" --> please mention the physical parameter provided by CryoSat-2 against which the model has been calibrated.

Text has been updated to note that it is Cryosat-2 thickness data.

L116: Please check whether it would not enhance the clarity of this sentence if you would write "ERA-Interim".

#### Corrected as suggested.

L184: "when the snow layer on top of the sea ice is pushed below water" --> This could be misunderstood. What is actually pushed under the water by the weight of the snow is the ice-

snow interface, then the basal snow layers may become flooded and, in case it is cold enough, refreeze. As written one might think that the entire snow cover needs to be pushed into the water.

The text has been edited to note that it is the ice-snow interface that needs to be pushed below the surface of the water.

L196: "2010s and 1980s" --> I suggest to use the same order as you used before in L194.

### Corrected as suggested.

L201: "The trend over time in both data sets is increasing humidity in all months," --> I suggest to write: "In both data sets the humidity is increasing in all months,"

The relevant sentence has been edited to improve clarity.

L204: "but particularly summer values" --> I suggest to write "particularly during summer"

#### Corrected as suggested.

L208/209: "around the Fram Strait region" --> Revisiting Fig. 1 suggests that it might be better to speak about "North of Svalbard" or even "North of the Barents Sea".

#### Agreed. Corrected as suggested.

L229/231: "The simulations show ..." --> It might make sense to also let the reader know that the increase of the July MIZ extent in the models runs kicks in around the year 2000.

The following additional sentence has been included in the text: 'For July, this transition primarily emerges from 2000 onwards'.

L326: "by more than" --> please check ... "more than"?

We think the original phrasing is fine here.

L357: "will primarily be driven" or "are primarily driven"?

I note that you are using the future tense in this and also the following paragraph even though you are referring to changes that happened already, i.e. between the 1980s and the 2010s. You might want to change this.

#### Corrected as suggested.

Fig. 9: It appears to me that "total melt" and "dynamics" add to 1 and that "total melt" is actually the sum of "top melt", "dasal melt" and "lateral melt". If you can confirm this then I think you need to change at least the caption of this figure. It is not sufficiently clear as written.

The above description of Fig. 9 is correct. We have edited the caption to make this point clearer.

#### **References**

\*We only include additional references here that are not already provided in the manuscript.

Wang, Caixin, Robert M. Graham, Keguang Wang, Sebastian Gerland, and Mats A. Granskog. "Comparison of ERA5 and ERA-Interim near-surface air temperature, snowfall and precipitation over Arctic sea ice: effects on sea ice thermodynamics and evolution." The Cryosphere 13, no. 6 (2019): 1661-1679.