SECOND RESPONSE TO REVIEWERS

Manuscript title: “Improved teleconnection between Arctic sea ice and the North Atlantic Oscillation through stochastic process representation”

GENERAL RESPONSE TO ALL REVIEWERS

We thank the reviewers, and also the co-editor, for their comments. In our attempt to assimilate the new ensemble members, incorporate the new LIM analysis and address the reviewer comments, the manuscript had clearly become quite bloated and hard to follow. The co-editor has explicitly urged us to simplify and streamline, and we have now made an effort to do so by making several notable changes, which we now summarise.

1. **Barents-Kara is now used for all data sets**. We had intended to do this anyway purely to simplify/streamline, but the co-editor’s comment on colorbars made us realise we had made a really stupid oversight when updating the analysis to use 6 members. In Figure 4, showing ice-NAO correlations at gridpoints, the colorbar had been manually set so that the zero-contour of +-0.2 roughly matched the 95% CI estimate. This estimate was based on 3 members: with 6 members, the 95% CI is approximately 0.13, but we failed to update the colorbar accordingly. The authors were therefore led to conclude that with 6 members there were still no significant gridpoint correlations in the Kara sea in OCE. In fact, while the BKS correlations in OCE are dominated by those in the Barents sea (as our sensitivity analysis had shown), there are in fact statistically significant correlations in the Kara sea with 6 members: see the new Figure 4. The reason for wanting to use the Barents sea for CTRL and OCE was, ultimately, because that appeared to be the only place with statistically significant gridpoint correlations in OCE. Statistically significant correlations are of course higher in general than non-significant ones (to address a point made by Reviewer #1), but it was the significance that was guiding us, not the magnitude.

   In any case, we now use BKS for everything. This immediately allowed for several simplifications throughout the paper. We apologise for the trouble this oversight, and our stubbornness on the point, have caused the reviewers and co-editors.

2. **Section 6 has, at the suggestion of the co-editor, been cut entirely**. The lack of changes to direct ocean-atmosphere coupling is still mentioned elsewhere, and the Figure suggesting an overly dominant ENSO in CTRL is included in Appendix B and briefly discussed in Section 5. The Blackport et al 2019 methodology has been cut entirely, based on feedback from Reviewer #1.

3. **The LIM analysis section has been cleaned up, simplified, and split in two**. As part of this, the LIM `forecast correlations' are removed from Table 1, as these seem to be mostly adding confusion rather than clarity. The figure of lag-correlations between ice and heatfluxes is now discussed explicitly and is included in the main text rather than the appendix.

4. **The focus of the paper is now more squarely on the period 1980-2015**, with the more speculative discussion on decadal variability kept to its own subsection (with a figure in the appendix). As part of this, the former Figure 5 has been removed, and we now rather display the correlation distributions for CMIP6, CTRL and OCE in a much easier to visualise boxplot (new Figure 3). The SPHINX data has been cut from the Figures to keep things simple: we now just state in the Methods section that the SPHINX experiments are consistent with the CTRL ensemble.

With regards to all the above changes, the reviewers and co-editor should find that no actually new information or analysis is included, it's just a case of things being cut, simplified or moved around. With the space that was made available from this, we did however decide to add one new piece of information: a very brief new sub-section 4.3 on tropospheric vs stratospheric pathways. This one-paragraph section contains one new figure showing that the teleconnection signal in OCE extends all the way up to the stratosphere, and that therefore both pathways may be active in OCE. This was included based on feedback from other senior academics in the field, and may help reassure readers that the signal in OCE is not restricted purely to the troposphere.

We now address the co-editor and three reviewers in turn.

CO-EDITOR

Co-editor: *Overall, the result that the ice-NAO correlation is different in OCE versus CTRL is convincing, and as such, an exploration of the reasons is certainly warranted. The authors have performed a number of analyses to get at the reasons. I believe where the third reviewer's concerns lie is in the interpretation of these analyses, which is very difficult to follow in places, and could use more balance/transparency. There are statements appearing in different places in the manuscript that seem contradictory, some statements seem at odds with what I see in the figures, and at times the discussion has an appearance of over-emphasizing certain points while dismissing others. In short, it is clear that the stochastic schemes are doing something to the ice-NAO teleconnection – probably related to the ability of ice to influence the NAO, according to the LIM analysis. However, I'm not sure the conclusion that it is due to differences in local ice-atmosphere coupling is well supported by the manuscript as it stands.*

Response: We thank the co-editor for their extensive feedback. We hope the changes made make the paper easier to follow and address the issues raised.

Co-editor: *- Lagged correlations: I find Fig. 7 and Fig. B7 equally important, and Fig. B7 especially so for understanding mechanisms, and thus am not sure why both should not be in the main text. Furthermore, I don't understand some of the statements made about B7. Fig. 7 shows that, when ice leads the NAO, OCE and ERA5 are behaving in a similar way while CTRL is doing something different (all okay until here). But Fig. B7 shows that the ice-flux lead-lag relationships (i.e., the processes) are operating the same way in both CTRL and OCE. Why do you say that Fig. B7 demonstrates that "the CTRL model is simply failing to propagate the same initial anomaly correctly" (L524) while not mentioning that the same holds for OCE? Also, it doesn't seem to me that this statement is accurate either: "Figure B7 gives some hint that the way the ice in the CTRL model adjusts to heatflux forcing is unrealistic, which may result in biases in the non-local adjustment. The same Figure suggests that these biases are reduced in OCE, suggestive of improved coupling." In fact, doesn't Fig. B7 suggest that it's not the local coupling that is making the difference between OCE and CTRL?*

Response: Our text was not clear here, leading to some confusion about when we were talking about the pressure anomaly (which ultimately is supposed to become an NAO pattern) and the heatflux response. We should also in general have been more explicit in how we were interpreting Figure B7 (now Figure 10). What we were trying to say was:

1. We do not find that the initial heatflux anomaly (in response to a BKS anomaly), on average, is significantly different in CTRL vs OCE. This is basically what the thick blue and red lines of Figure B7 show, as well as Figures B4 and B3 (which show the mean November heatflux and temperature signal).
2. In addition, we saw in Figure 6 that also the initial pressure anomaly in November looks basically the same in CTRL and OCE. So the conclusion of these figures suggests that, on average, the response of CTRL and OCE to a BKS anomaly is roughly the same within the first month or so, whether one considers heatfluxes or pressure anomalies.
3. Because CTRL nevertheless fails to have a teleconnection, this suggests that CTRL is failing to propagate this initial pressure anomaly correctly across the winter season. Note that this conclusion is not based on Figure B7, but on the fact that OCE has a teleconnection while CTRL does not. This interpretation was spelled out in Section 5.2 (see the paragraph starting line 425, as well as the first sentence of the following paragraph).

However, Figure B7 does suggest that there *are* some potentially interesting differences between the local ice-heatflux coupling in OCE vs CTRL. This is because of the difference in the ensemble spread (shown in shaded blue for OCE and stipled red for CTRL). The ensemble spread of CTRL is much bigger than for OCE, which means that, e.g. for small negative lags, all the OCE ensemble members give correlations with the correct sign (i.e. the same sign as ERA5), but in CTRL the sign is correct only on average, with half the members giving the wrong sign. We interpret this to suggest that the ice-heatflux coupling has become more tightly constrained in OCE, in a manner which appears to bring the ensemble as a whole closer to ERA5.

We have now tried to clarify this in our rewrite of this section. We also moved Figure B7 into the main text (now Figure 10) and made our interpretation of it explicit.

Co-editor: *- Related to the above, it seems like one of the main points is that atmosphere-ice coupling (and not atmosphere-ocean coupling) is key to getting the teleconnection, but the discussion seems to circle the point in many places. For example, while much is made of the generally “improved” coupling in OCE compared to CTRL, section 6.3 (especially statements like L617-619) and the concluding point on L657 talk specifically about how the atmosphere-ocean coupling does very little. It seems like it would have been useful to run an experiment with just the stochastic ice scheme (not the ocean schemes), which would allow firmer conclusions and cleaner/more direct writing.*

Response: We agree, it would have been nice if we had such experiments. The reason we do not has to do with the fact that the origin of this experimental protocol had nothing to do with testing the impact on Arctic-midlatitude teleconnections, but was rather concerned with assessing the impact of stochasticity in broad terms (as part of the Horizon 2020 PRIMAVERA project). Given limited computing resources and prior experience we decided at the time to test both schemes at once.

Since it seems clear now that having 6 members is really much better than just 3, running additional experiments to single out the impact of each scheme would require us to run an additional 12 simulations covering 1950-2015, i.e. 780 simulated model years. Even if we had computing resources to do this, it would take many months to carry this out, and even longer to carry out any relevant new analysis. We therefore hope the co-editor will accept that we will not be able to do these experiments for this present paper. We hope to carry out such experiments in the future.

We added some lines in the Conclusion about the lack of such experiments.

Co-editor: *- Fig. 9: Shouldn't the top row be the same as the bottom row of Fig. 6? The "ERA All" pattern is dominated by "ERA Atmos2ice", while for both CTRL and OCE, the All pattern is dominated by Ice2atmos. How does this fit with the LIM or the idea that the ice forcing is somehow stronger/better in OCE?*

Response: We found a minor coding error which meant Figure 9 was using data covering 1979-2015, rather than 1980-2015, which explains the minor differences in the plots. However, we have taken the suggestion to remove the entirety of Section 6, including Figure 9, from the paper.

Co-editor: *I ask the authors to revise the manuscript taking into account all the comments of the referees, as well as some additional comments from me below. This would include addressing the two main comments of referee #1 (removing seasonal cycle and clarifying text) and the various comments of referee #3 about clarifying/streamlining the text and using references appropriately. I leave it to the authors to decide whether or not to change the definition of the Barents-Kara region, but the comments of referee #3 on improving the discussion of similarities/differences in Figures 1 and 2 should be considered regardless. (Related to this, the colour scale of Fig. 4 should probably be adjusted to better show spatial features of the correlation structure, and also to show more of what's happening in CTRL. In c-f, the 95% ci is +/- 0.13 right? So it would be fair to show correlations at least as small as around 0.1.)*

Response: Thank you for the various suggestions. As described earlier, we have undertaken many sweeping changes to try to simplify and streamline the manuscript.

We particularly thank the co-editor for their comment on the colorscale. We realised that, when adjusting the analysis to the doubled ensemble size, we had stupidly not adjusted this colorscale appropriately: +-0.2 was the 95% CI using 3 ensemble members, hence the choice before. We now changed it to go down to +-0.13, the approximate 95% CI using 6 members. This gives a clearer picture of the situation, and had the embarrassing benefit of showing that, with 6 members, the gridpoint correlations in the Kara sea are significant in the OCE ensemble: see the General Response.

Co-editor: *The new text in 4.1 justifying the use of different sea ice regions could be presented in a much clearer and more condensed way. The EOF methods are set up as somewhat of a strawman (text starting L266 "This latter method has the advantage…") only to be taken down later (starting L289). I understand that this text was added to address one of the reviewers concerns, but I think this is better done via (a) the sensitivity tests presented later and (b) discussion/interpretation of the results.*

Response: This whole text has been made redundant by using BKS for everything, so it has all been removed.

Co-editor: *L32-35: These two sentences as written could be interpreted as separate, contradictory findings. I think this text need some editing to reflect the fact that ice-NAO correlations are weak when the averaging period/simulation is long (first sentence), but can be stronger and positive or negative when the averaging period is shorter, i.e., long simulations are subsampled (second sentence). This second part is actually true for many models, not just a single one – e.g., CESM as shown in Kolstad and Screen, and a range of LENS experiments and CMIP5/6 models as shown in Siew et al.*

Response: We added a clarifying comment about the length of the simulation period.

Co-editor: *L161: "in the same way by initial condition perturbation" -> "by initial condition perturbation, as above"*

Response: We made the suggested change.

Co-editor: *L185-187: What about the additional simulations (SPHINX, CMIP6, HighResMIP)?*

Response: We clarified that we use 1980-2015 for these additional simulations as well.

Co-editor: *L196: Can you add one sentence explaining why different statistical tests are used for the std dev of SST and sea ice?*

Response: Done: the Levene test is more appropriate in situations where the distributions might be more non-Gaussian, which is often the case for sea ice.

Co-editor: *L257-259: "to have a bigger impact" compared to what?*

Response: Compared to the impact of the ocean perturbations at other regions of the globe. We rephrased to say "have a notable impact" to avoid this confusion.

Co-editor: *L263:264: Not sure this makes sense - to regress a "region" against gridded pressure data or the NAO index? Also small grammar issue with "regressing".*

Response: This line was cut. Other situations where similar phrasing occurred have been clarified.

Co-editor: *Fig. 2 – EOF1 of SIC for OCE doesn't look more realistic than that for CTRL – in fact, maybe less.*

Response: This Figure has been removed.

Co-editor: *Footnote 2: I understand the first part of the sentence, that the teleconnection might only emerge when the storm track is in a favourable position, but how does it follow that climate model biases would promote shifts in the key region? Perhaps part of the problem is that I'm not clear what biases or shifts we're talking about – storm tracks, ice line, something else?*

Response: This footnote was removed, and in general the discussion of this has been removed from the manuscript in an attempt to streamline the discussion in response to comments by yourself and the reviewers.

For the benefit of the co-editor, we were suggesting that in order for a sea ice region to be capable of generating a teleconnection with the NAO, three conditions may need to be satisfied:

1. It needs to be a region which actually experiences significant interannual variability.
2. The interannual variability needs to be associated with large heatflux anomalies (i.e., variability which still leaves the region fully covered by sea ice wouldn't be expected to be important).
3. The region needs to be such that stationary Rossby waves generated there can be constructively reinforced by the storm track.

If these criteria are correct, then biases in both the position of the storm track and the mean sea ice state (particularly the position of the ice edge) in models mean there is no reason, a priori, for Barents-Kara to be the most important region in a model. E.g. if a model has a huge bias in the Kara sea, then Kara sea ice may not be generating notable interannual variability in heatfluxes (because, e.g., the ocean is mostly always covered up in the model). Or, e.g., Rossby waves generated in the Kara sea don't have a chance to get reinforced by the storm track because the storm track is biased south in the model. Or some more complicated combination of these.

However, we acknowledge that this is essentially speculative, and likely distracts from the main point of the paper, so we have removed it.

Co-editor: *L291: "prescribe Barents-Kara across all models" needs rewording*

Response: This line was cut.

Co-editor: *Fig. 5: In panel b, where is there only one one OCE member (dashed blue line) with correlation greater than ERA5 (dashed black line), while Table 1 suggests there should be 2 members?*

Response: Thank you, this was well spotted. This was regrettably due to an error in our analysis code, which meant there was a difference in the trends being removed in the two cases (one trend was defined using 1950-2015, the other was using 1980-2015). This led to the correlations marked in Figure 5b being slightly different from those in Table 1. We had not noticed because, of course, the picture is basically the same either way (OCE has notably higher correlations than CTRL and CMIP6).

We have opted to remove Figure 5 either way. It has been replaced with the more visually transparent boxplot figure (Figure 3 in the revised). The correlations that go into this boxplot have now been checked to match those of Table 1. We apologise for the mistake and associated confusion.

Co-editor: *Section 5.1: Section 3 shows small improvements in the mean state of OCE compared to CTRL for some features (e.g., sea ice in the Barents-Kara) but certainly not others (L246 mentions that the mean SST bias shows no improvement. This makes the hypothesis stated on L405 a bit confusing/misleading. In addition, I'm not sure Fig. 3 should be referenced on L548, as it more or less contradicts the comment on L246.*

Response: We changed the phrasing in both cases to refer to the possible impact of changes to the mean state, rather than improvements.

Co-editor: *Sections 5.1/5.2: The AMIP experiments certainly suggest that the ocean mean state is not the answer, but it in the interest of understanding the implications, it should be mentioned that if internal atmospheric variability plays a role, then prescribing sea ice would also be expected to destroy the ice-NAO correlation. The Blackport and Screen 2021 JClim paper is very relevant for this interpretation (even though it does not address the ice-NAO teleconnection directly).*

Response: We added this point to the beginning of section 5.2 and cited Blackport and Screen 2021.

Co-editor:*Section 6: The goal should be either to provide a more balanced discussion of the existing results or to really pin down the ice-atmos coupling as critical. Please consider whether this section is necessary.*

Response: We chose to remove the entire Section 6. Instead, we split up the LIM discussion of Section 5 by adding a new subsection where discussion on possible explanations of the LIM results are contained.

Co-editor: *Discussion/summary: In general, I like the idea of bullet points in the summary, but eight is probably a bit too many. Also, please be sure that summary statements about the role of mean state vs coupling in creating the differences are accurate and well supported by what has been presented in the manuscript, i.e., what is actually improved or not improved in OCE versus CTRL should be clear from figures, try to summarize if a certain feature seems to contribute to OCE vs CTRL differences in one analysis but not in another, etc.*

Response: The simplifications made to the paper meant we could reduce the bullet points to 5, which we hope is seen as a more reasonable number. We have also generally tried to rephrase things more cautiously/accurately.

REVIEWER #1

Reviewer #1: *The authors have significantly improved the manuscript in response to my comments, but I don't feel my comment about the use of different regions for the models and observations has been adequately addressed. The authors have changed the regions so that they are not as different as they were in the original manuscript, but they are still different and it is still a problematic. The authors go into detail about why there could be plausible physical arguments for using different regions for the comparisons (L274-288 and much longer discussion in their response). However, they do not actually use any of these physical arguments to justify the specific regions chosen. The authors do state in a few places that the choice is based in the results in Figure 1 and 2 (e.g. L626-628), but I do not see what they are seeing in these figures. There is clearly sea ice variability in both the Barents and Kara seas in both ERA5 and OCE (Figure 2) and if anything, I see bigger differences in the means and stds in the Barents compared to the Kara sea (Figure 1). This is the opposite of what is claimed. In my view, the main reason the authors use different regions is still that there is a stronger correlation in the model when only considering only the Barents Sea.*

The authors have included some analysis showing that it makes only a small quantitative difference if they use the Barents-Kara Sea for all analysis and then use this to justify the use of different regions. But then why not use the same region for all analysis? This would be simpler, avoids the possibility of cherry-picking (or the appearance of cherry-picking), and is more consistent with past work that focuses on Barents-Kara sea. It also would help shorten the paper, which is quite long. It just seems like the authors are over-complicating the analysis for little benefit and plenty of downsides.

Response: Upon further reflection we have decided we agree with the Reviewer, and have now changed the paper to use Barents-Kara for everything. The choice of using just Barents (or Barents-Greenland in the first version) was an attempt to be transparent about where there were significant correlations at actual gridpoints (as opposed to for area-averaged timeseries): statistically significant correlations are of course generally going to be larger than non-significant ones. However, we acknowledge that firstly, this is ultimately distracting from the more central point of the paper, and, secondly, that our attempt to motivate the different region is essentially speculative.

This has allowed for various simplifications elsewhere in the paper as well. We thank the reviewer for their patience with us on this point.

Reviewer #1: *L40-41: This mention of Blackport at al. 2019 is misleading and unnecessary, so can be deleted. It is not about the sea ice-NAO connection (more about this later).*

Response: We deleted this reference.

Reviewer #1: *L79: Similar to Blackport et al 2019, Mori et al. 2019 is about the 'warm Arctic cold Eurasia' pattern and has little to do with the sea ice-NAO connection, so should not be presented as such. If anything, the results presented in Mori et al. 2019 would suggest that there is positive NAO response to reduced Barents-Kara sea ice and that models are underestimating this. This is the opposite of what is suggested here, so does not support this work.*

Response: We clarified to say that Mori et al. made a similar hypothesis in a different Arctic-midlatitude context: "A similar hypothesis has also been emphasised in \citet{Mori2019a} and \citet{Mori2019b} in the context of surface level teleconnections".

Reviewer #1: *L260-273: It seems unnecessary to provide these arguments to use an EOF-based method to identify the critical sea ice regions and then not even use these methods to then identify the region.*

Response: This was removed entirely due to the change to using BKS only.

Reviewer #1: *L274: "Besides the somewhat abstract consideration of modes of variability…" But the modes of variability are not even used to identify the region?*

Response: As above.

Reviewer #1: *L350-353: The lack difference in trends is intriguing and could be highlighted and discussed in more detail. This would seem to contradict the idea that the differences in correlations in CTRL and OCE are because of the CTRL (and models more generally) underestimating the response to sea ice anomalies. Why would the differences in interannual variability not translate into differences in trends in the NAO? This is particularly important because I could see the results presented here regarding the interannual correlations be misinterpreted to have implications for climate change, when it appears to have little effect. It is also interesting that the models show a robust negative NAO trend over 1950-2015, while the observed winter NAO trend is clearly positive and well outside of the model distribution shown Figure 2 of the response. Because this is outside the scope of the paper, the authors do not need to do additional analysis, but some discussion would be useful.*

Response: It is possible that the physical mechanisms that relate sea ice trends with NAO trends are different than those determining interannual variability. As a hypothetical, the climate change signal may be due to changes in the large-scale meridional temperature gradient, while interannual variability may be driven by more localised Rossby wave activity. If this were the case, CTRL may well be underestimating the response to one of these but not the other. The fact that the global warming signal differs from the interannual signal has been noted in other contexts, such as ENSO (https://doi.org/10.1175/2008JCLI2200.1), so it seems hard to rule this out.

For these reasons, we hesitate to draw conclusions on impacts of climate change signals without actual climate change experiments. We have, however, added a new, short subsection of section 4 discussing decadal variability, trends and climate change, including some of these points.

Reviewer #1: *Section 4.2: What are the differences in CTRL and OCE over the entire period (1950-2015)? Looking at Figure 5a, the differences will likely be much smaller than only over the most recent period.*

Response: The correlation of the concatenated CTRL ensemble is 0.067 and for OCE is 0.184. The individual correlations of OCE ensemble members are generally higher as well, as expected.

Though we note the edited discussion on Decadal Variability (new sub-section 4.2): when looking at the spatial pattern of correlations in the first 35 years, it appears as if there is a much larger role being played by the Greenland and Labrador sea at that time, with less coming from Barents and particularly from Kara. The correlations in these regions are still greater in the OCE ensemble members than for CTRL. We included some speculative comments (which are labelled as such) about the possibility that much decadal variability in BKS correlations may arise because of the importance of other regions in the past, implying that a focus on BKS only may give a somewhat misleading picture. This discussion has been kept brief and cautiously worded, because it would require further work to substantiate, but we hope that it may be of interest anyway: we are not aware of this point having been made previously.

Reviewer #1: *Figure 5 caption: 'three' should be 'six'*

Response: This Figure has been cut for the latest revision.

Reviewer #1: *L365: Another difference is the exact months used. A number of studies have looked at the correlation between November-December sea ice and the subsequent February circulation (e.g. Blackport and Screen 2021, De and Wu 2019 doi: 10.1007/s00382-018-4576-6).*

Response: We added this point to the text.

Reviewer #1: *L368-369: While CTRL may not be unusually weak compared to CMIP6 mean, they are more negative than the CMIP6 mean. This would be a lot easier to see if the CTRL values were plotted on Fig 5b like they are with the OCE.*

Response: In our effort to streamline and simplify the presentation, we have removed Figure 5, and replaced it with a boxplot (new Figure 3), which shows more clearly that the CTRL ensemble, while consistent with a random draw from CMIP6, is on the negative side of the distribution.

Reviewer #1: *L369-371: Whether the CMIP6 distribution changes or not seems like an irrelevant point, so should be deleted. This will depend more on the size of the CMIP6 ensemble than how well the CMIP6 and CTRL ensemble agree or not.*

Response: We agree, that Figure really wasn't that helpful. The relationship between CMIP6, CTRL and OCE is now visualised in the new boxplot Figure 3, which we hope is more illuminating. We have removed the SPHINX correlations from these as well, again to help streamline/simplify (a statement has been added about consistency of CTRL and SPHINX instead).

Reviewer #1: *L425-431: Blackport et al. 2019 is not relevant in other places in the manuscript, but it is relevant for this discussion. The pattern seen in the instantaneous November regressions are very similar to the instantaneous DJF regressions found in Blackport et al. 2019 that they concluded were because of the atmosphere forcing the sea ice.*

Response: At the suggestion of the co-editor, the entirety of Section 6 has now been removed, including the discussion of the Blackport et al. (2019) methods.

Reviewer #1: *L566-576: The authors are again misrepresenting Blackport et al. 2019. They show that a positive NAO is associated with a negative sea ice anomaly in DJF. This is clearly not 'nearly identical' and is in fact the opposite to what is found in Figure 9 here. More generally, it is somewhat concerning that the authors are continuing to misrepresent a number of different studies on this topic even after it is pointed out. It makes me wonder whether the authors are misrepresenting other studies that I am not as familiar with and do not have time to carefully check.*

Response: We have removed the Blackport 2019 inspired analysis from the paper. It seems to us that the methods of that paper could still be helpful for understanding the role of atmospheric variability in the ice-NAO teleconnection, but based on your feedback, and that of the co-editor, we have decided to not pursue this here.

Reviewer #1: *L591-592: Are there citations for this statement?*

Response: This section was removed.

Reviewer #1: *L651: It looks like the start of the sentence was cutoff?*

Response: Thank you, the word “These” had been cut.

Reviewer #1: *L653: Again, Mori et al. 2019 is not about the sea ice-NAO connection.*

Response: We clarified here as well that the hypotheses of Mori et al was in a different context.

REVIEWER #2

Reviewer #2: *Line 488: do you mean second "column" not "row"?*

Response: Thanks, we corrected

Reviewer #2: *Table 1: I think it would be helpful to clarify in the table caption what the LIM correlations are. You explain in the text, but it is a bit unclear in the table caption.*

Response: We have added a brief explanation in the caption of what these LIM forecast correlations are measuring, with a pointer to the relevant section.

Reviewer #2: *Line 568: I think there is a Figure reference missing at the beginning of the sentence in this line.*

Response: This entire section was removed.

Reviewer #2: *Line 651: I think there is a word missing at the beginning of this sentence.*

Response: Thanks, the word “These” had been accidentally cut.

REVIEWER #3

Reviewer #3: *First, it looks like the authors missed part of my comment #5 in the previous review, so I’m reiterating this here. For the NAO index calculation, I didn’t* *get why they are using non-deseasonalized data for their EOF analysis, and removing the seasonal cycle only in a second step (L175). This choice would make sense if the seasonal cycle is seen as an important component of the variability, but that's not the case here (since the authors are deseasonalizing anyway). I'd recommend deseasonalizing before the EOF analysis. It's entirely possible it would make very little difference to the results, but would seem cleaner to me.*

Response: Apologies, we must have overlooked this. We have now tested the sensitivity to this choice by recomputing the NAO time series in the suggested manner: removing a seasonal cycle from every gridpoint in the domain, then doing the EOF computation. It turns out that this has a negligible impact on the results. For example, here are the ice-NAO correlations for the 6 OCE ensemble members to 3 decimal places, first as in the paper, then with the recomputed NAO indices:

Member1:	Corr = 0.129 (p=0.46)
Member2:	Corr = 0.166 (p=0.34)
Member3:	Corr = 0.359 (p=0.03)
Member4:	Corr = 0.078 (p=0.66)
Member5:	Corr = 0.542 (p=0.00)
Member6:	Corr = 0.271 (p=0.12)

Member1:	Corr = 0.134 (p=0.44)
Member2:	Corr = 0.162 (p=0.35)
Member3:	Corr = 0.356 (p=0.04)
Member4:	Corr = 0.075 (p=0.67)
Member5:	Corr = 0.543 (p=0.00)
Member6:	Corr = 0.276 (p=0.11)

We have therefore decided to leave the Table and Figures as they are, rather than redo the analysis with these new NAO timeseries. A sentence has been added to the Methodology pointing out that this choice does not matter for the results.

Reviewer #3: *Secondly, section 5.3 was somewhat hard work to read, so I'd encourage the authors to simplify the text if possible. For example, is the paragraph starting L482 useful? The following paragraph states there is a better way to test the LIM hypothesis, so why not jump to that straight away?*

Response: Yes, this section ended up quite heavy going, sorry about that. We have now decided to remove the LIM forecast correlations from Table 1, since these just seem to promote confusion and distract from the main point: for consistency, the perfect LIM reconstruction plot (Figure B6) has been replaced with a similar Figure showing examples of randomly drawn LIM reconstructions. We also now jump straight to the “better way” in Section 5.3 as suggested.

To further aid readability, the entire subsection has been split in half, with the first half just dealing with the validation (i.e. which of the data sets are consistent with the LIM), and the second half handling the discussion of why the LIM hypothesis might be failing in CTRL. This includes a more explicit discussion of the ice-heatflux lag correlation plot. Hopefully this has made the whole thing less burdensome to the reader.

Reviewer #3: *L175: Is the EOF analysis based on NDJF data only? Please clarify.*

Response: The EOF (i.e. the NAO pattern) is computed using DJF data only, and then an NDJF time series is obtained by projecting the NDJF zg500 anomalies onto this pattern. We have now clarified this in the Methodology.

Reviewer #3: *L189: “the period is”*

Response: Fixed.

Reviewer #3: *L235: “suggest that”*

Response: The discussion on EOFs has been cut from the paper.

Reviewer #3: *L281: “heat flux anomalies”*

Response: Fixed.

Reviewer #3: *L337: “concerning both the”*

Response: This line was cut in the revised version.

Reviewer #3: *L391–394: This is unimportant, but I think the logic behind this statement is flawed, or at least I didn’t follow it. If Kara sea ice is highly correlated* *with Barents sea ice, then I'd expect Kara sea ice to correlate with the NAO even in the absence of a direct physical link between the two.*

Response: This is only guaranteed to be true if the correlations involved are close to 1 between everything, but may easily fail when the correlations are lower. This is the case here, where the correlation between sea ice and the NAO is small at each gridpoint (~0.2), implying only a weak linear relationship. In fact though, the new Figure 4 (with a better colorbar) shows that there *are* weak positive correlations between Kara sea ice and the NAO, just much lower than in the Barents sea. But either way, this discussion has been removed due to using BKS for everything now.

Reviewer #3: *L400: Does this mean the ERA5 correlation would be improved if only the Kara sea region were used?*

Response: Not really: we rather find that you can get the full signal (i.e. a correlation of ~0.4) using just Kara sea ice, and adding Barents to the region doesn't change this. This suggests that the BKS signal in ERA5 is primarily coming from the Kara sea, with the signal in Barents emerging mostly from spatial correlations in ice.

Reviewer #3: *L407: To clarify, the AMIP ensemble uses the OCE SSTs and sea ice, not the CTRL values?*

Response: No: the AMIP ensemble use *prescribed observational* SSTs and sea ice, obtained in this case from HadISST. This was stated in the Methodology, but we have also now added a reminder of this in the section your line is from.

Reviewer #3: *L568: Word missing after the period*

Response: This whole Section was removed.

Reviewer #3: *L651: Missing "The"*

Response: This bulletpoint was removed in the latest version.

Reviewer #3: *L699: "The study *of* Juricke et al."*

Response: Fixed.

Reviewer #3: *Caption of Fig. 4: "The ensemble members" (extra x)*

Response: Fixed.

Reviewer #3: *Fig. 8: I'd recommend plotting this as a barplot, or at least expanding the y-axis to encompass 0. That would make it much easier to compare the magnitudes of the coefficients among CTRL, OCE and ERA5.*

Response: Good idea: we have added a horizontal line at y=0. This actually helps make it clearer why the variation in coefficients between CTRL and OCE has only a small impact on the correlations generated by the LIM fit. Thanks!