THIRD RESPONSE TO REVIEWERS

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

GENERAL RESPONSE

We once again thank the reviewers and co-editor for their valuable feedback.

The changes made in the revised manuscript are mostly minor tweaks and clarifications. The two most notable changes are as follows:

- To support our speculation in Section 4.2 on decadal variability of the teleconnection, we added the new Figure B3, showing that there are substantial changes to the sea ice variability in the Labrador, Greenland and Kara seas between 1950-1980 and 1980-2015 in EC-Earth3. We also now cite and discuss the paper Kelleher and Screen (2018), which shows a statistically significant link between Greenland sea ice and polar cap geopotential height anomalies in the CMIP5 pre-industrial simulations (cf. Figure 8c of ibid). Both these inclusions corroborate our speculation that other regions, such as the Greenland sea, may have exerted a stronger influence on the circulation in the past.

- Text has been added to the lag-correlation plots to indicate the direction of causality, and negative lags now always correspond to the ice leading the atmosphere.

The co-editor has also pointed out to us that Siew et al. (2021) corroborates the findings of Koenigk et al. (2017) in a multi-model context, so these papers are now cited together.

We now respond to each reviewer in turn.

REVIEWER #1

Reviewer #1: *Abstract: I think the authors should say something in the abstract about how it is not clear what is causing the stronger correlations in the simulations with stochastic parameterizations.*

Response: We edited the abstract to clarify that the exact mechanisms are still unclear:

“While the exact mechanisms causing this remain unclear, we argue that it can be accounted for by an improved ice-ocean-atmosphere coupling due to the stochastic perturbations, [...]”

Reviewer #1: *L33: I don’t understand this statement. Why would the signal be much smaller over shorter time periods? Wouldn’t it be that the signal is the same over shorter time periods, but that it is harder determine the signal because there is too much noise?*

Response: Yes, this was not well phrased. We edited to just say that the signal is less consistent on short time scales.

Reviewer #1: *L38/footnote 1: I certainly agree with the authors that the sea ice record is more uncertain further back in time, which questions some of the results from Kolstad and Screen 2019, particularly the results from the early 20th century. However I think the authors are still misrepresenting their analysis and results here. They use multiple datasets, not only the HadISST and some of the strongest negative correlations are seen in more recent periods (after 1950) which likely are more certain.*

Response: We are aware they used more than one sea ice data set, which is why we were careful to phrase it as “For example, HadISST [...]”. We did not want to explicitly discuss all the data sets used in that study, because that would make the footnote overly long. The fact that one widely used data set has clear issues already seemed sufficient to make the point that the quality of sea ice data in the past may not be trustworthy. After all, there are only so many sea ice measurements floating around and all the data sets will generally be trying to assimilate the same measurements. Kolstad and Screen do not carefully discuss the quality of any of the data sets they use.

As far as the time period is concerned, as indicated in the footnote (and visualised in a figure included in an earlier response to reviewers comments), inspection of the data suggests the reduced quality of data in the Barents-Kara region persists well beyond 1950. In private communication with John Walsh (of the Walsh and Chapman data set), it was suggested to the lead author that there is a lot of intermittency and unevenness in sea ice measurements in the Barents-Kara region in these decades (50s to 70s), since availability of measurements mostly depended on how often Russian planes were flying over this region. There seems, therefore, to be good reason to question how meaningful BKS-NAO correlations are for *any* period prior to 1979.

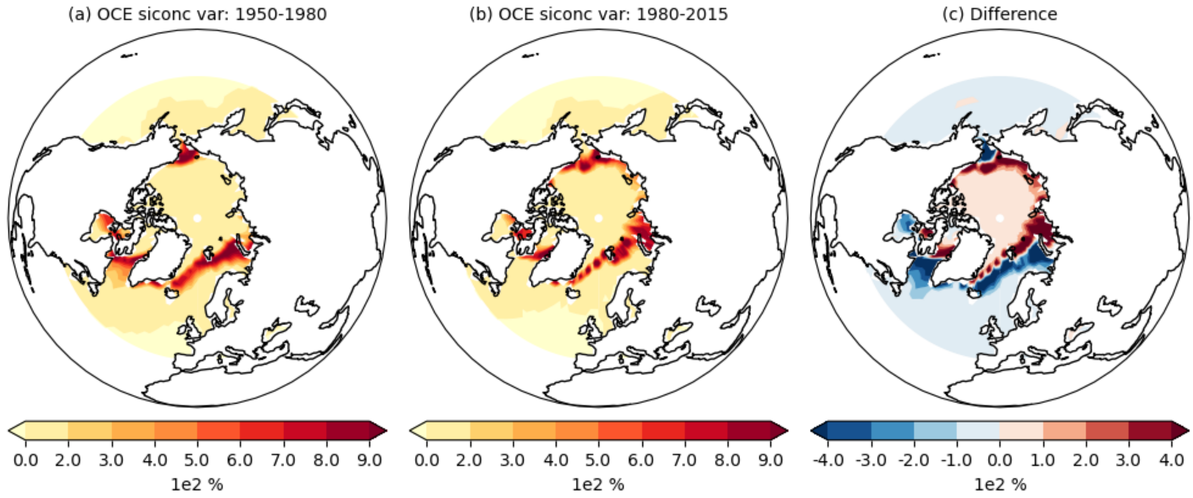
Of course, all this should really be discussed in its own research article, and there is certainly no space in the footnote to get into all these details. But we do not feel we are misrepresenting or overstating concerns about the results of Kolstad and Screen, so we have left the text as is.

Reviewer #1: *L53: This is misrepresenting Blackport and Screen 2021, or maybe I am not understanding what the authors are saying about it. What does "the overall signal is too small to be robust" mean? Blackport and Screen 2021 find that there is a weak signal between sea ice and the NAO in coupled models.*

Response: We agree our text was misleading: "not robust" here was our own very informal interpretation of Blackport and Screen 2021, based on their conclusion that while there is a statistically significant correlation in the coupled multi-model mean, a) the correlation is very small, and b) may overstate the causal ice→NAO link because a lot of the signal might be coming from atmospheric variability. We now rephrased this to focus more clearly on the key point, which is that the findings in Baker et al. 2018 implies that many models will have model errors related to the forced NAO variability, making it hard to rule out that model error is causing the signal to be underestimated in climate models.

Reviewer #1: *L313-330: I am not sure how useful this discussion is. It is very speculative, but it could be interesting if it was backed up by analysis. For example, what does the sea ice variability look like before 1980 vs after in OCE? Is there substantial variability in the Greenland Sea before 1980, but not after? How do these compare to observations (not only the differences like is shown in Fig 1, but the actual variability in each)?*

Response: We made the following figure, showing the change in November sea ice concentration variance between the periods 1950-1980 and 1980-2015, in the OCE ensemble:



This corroborates the discussion in L313-330, showing that there is a substantial reduction in variability in the Labrador and Greenland seas across these periods. There is also almost no variability in the Kara sea in the early period, unlike in the second period, again consistent with the change in the region of significant correlations. We did not make this plot using observational data because of the uncertainties in sea ice data prior to 1979, discussed here earlier.

We have added this figure to the appendix (new Figure B3) and refer to it in the discussion. We also found the paper Kelleher and Screen (2018) (Atmospheric precursors of and response to anomalous Arctic sea ice in CMIP5 models | SpringerLink), which analyses pre-industrial CMIP5 simulations and finds a significant impact of Greenland sea ice area on polar cap geopotential height anomalies in winter, corroborating the potential importance of the Greenland sea in the past. This point, and citation, have also been added to the discussion.

We agree that on the whole this discussion is speculative, but we also believe it raises interesting points that seem plausible (or at least worth considering) based on the preliminary analysis we have done (such as the above figure) and which seem to us to not have been given adequate attention in recent literature. The speculative nature has also been flagged right at the start of this sub-section, and we have now also rephrased the most speculative point (the importance of other regions in the past) to re-emphasise this. We hope the reviewer will not therefore object to us keeping this sub-section in the paper.

Reviewer #1: *L369: Strong et al 2009 is an observational study with no climate model analysis, so shouldn't be included here.*

Response: We removed this citation from this location.

Reviewer #1: *L382-385: What am I supposed to be looking at in Fig B4? OCE and CTRL look very similar around the Barents-Kara Sea. There are differences in the North Atlantic, but these are likely driven by the NAO and are likely not a driver of the NAO.*

Response: The similarity of CTRL and OCE is the point, as we explain in the text:

“Similar plots showing the evolution of heatflux and temperature anomalies (Figures B4 and B5) corroborate this story, with a similar initial anomaly that evolves relatively realistically over time in OCE but simply peters out in CTRL.”

Granted, this exact description is perhaps better in the case of 850hPa temperatures and 500hPa geopotential height, since for heatfluxes, the anomalies over Barents and Kara look basically the same over all three averaging periods (i.e., the `petering out’ of CTRL relative to OCE isn’t as clear). We rephrased therefore to “persists or peters out in CTRL.”

The basic point being made in this paragraph as a whole is that when you compare CTRL and OCE in November, there isn’t a notable difference (in pressure, temperatures or heatfluxes), but as you progress through the season they diverge, with OCE evolving more like ERA5 and CTRL more just persisting or fading out. We think this should be clear when considered in the context of the paragraph as a whole, especially the discussion on zg500 just preceding the reference to these figures.

Reviewer #1: *Figure 8/10: The sign conventions for the lags are confusing for Fig 8 and 10. Figure 8 has negative lags for sea ice leading the NAO, but figure 10 has negative lags corresponding to sea ice lagging the heatflux. Because sea ice is the common variable between the two plots it would make more sense to be consistent and have negative lags in both plots correspond to sea ice leading (or have them both lagging). It would also be helpful to have labels on the plots what is leading and lagging (e.g. "sea ice leading the NAO" on the left hand side of Fig 8).*

Response: Good point: we have changed the order in the ice-heatflux plot to have the ice leading the atmosphere for negative lags in both. We also added text indicating the lead/lag direction as suggested.

Reviewer #1: *L476-484: I don't follow the logic here. Why would the decreased between-realization variability of the sea ice - heat flux relationship lead to a more realistic sea ice - NAO teleconnection? Why would this be more important than the ensemble mean?*

Response: High interannual correlations require a consistent response. It's not enough to just have the correct association when averaged over a period of decades, you need the correct association happening most years. The large spread in the ice-heatflux relationship in CTRL is suggestive of the possibility that there are many years when CTRL does the opposite of what it `should' do, leading to an inconsistent response. Conversely, the more consistent ice-heatflux relationship in OCE may be part of why the ice-NAO link is also more consistent.

We rephrased the text here slightly to emphasise that the more consistent ice-heatflux relationship may plausibly lead to a more consistent ice-NAO link, and hence larger correlations.

Reviewer #1: *L491: What about Figure 10 suggests this?*

Response: The idea that remote adjustments are important is that, potentially, the initial circulation response to a sea-ice anomaly in Barents-Kara affects SSTs and/or ice in nearby regions (e.g. the Greenland sea), and that these changes to SSTs/ice act to promote or strengthen the subsequent NAO response. This mechanism depends therefore on how the atmosphere adjusts the SSTs/ice in these regions. This adjustment at the surface would be mediated by heatfluxes, and Figure 10 shows that the ice-heatflux coupling is frequently very different from what ERA5 says it `should' be. Hence one might plausibly expect that any important remote adjustments are often being done badly in CTRL.

Reviewer #1: *L492: Why would the anomalous upward heat fluxes around the Greenland and Labrador sea lead to a positive NAO? These differences in heat flux can be explained by the NAO forcing the sea ice/ocean. A positive NAO leads to cooling around Greenland (seen in Fig B5), which leads to the anomalous heat flux from the ocean to atmosphere.*

Response: We don't have an explicit mechanism in mind. Our speculation was based mostly on the fact that if remote adjustments are important for generating the correct NAO response, and the adjustments are related to ice-ocean-atmosphere coupling, then the obvious place to look for relevant remote adjustments is in nearby sea ice regions. Greenland and Labrador seas arise naturally when thinking along these lines, and we do see a difference in the heatflux evolution there.

We agree that the signal seen in these regions can be explained by the NAO forcing alone, but because there will also be continuous two-way coupling between the ice and the atmosphere in these regions (as in the Barents-Kara region), it seems hard to rule out that the forcing from the ice to the NAO isn't also playing a role.

Of course, this is entirely speculative, but we did label it as such. In general in this section, we decided it was ultimately better to offer some speculation about the causes of the changes seen, rather than offering no potential explanations at all.

Reviewer #1: *L485-498:Aren't these explanations contradicted by Figure 8 and 9? Figure 8 and 9 suggest that the differences between OCE and CTRL occur because of the initial response to the NAO on daily timescales, but this explanation requires processes that would occur on weekly and monthly timescales.*

Response: We are not sure we understand the reviewer here.

The lag-correlations in Figure 8 are computed using all the daily data over each NDJF season (120 days), as explained in the text (see caption to Figure 8), and are not centred on November 1st. If, for example, the CTRL and OCE ensembles both have identical responses in the first 20 days say, but then the signal in CTRL is blown away by ENSO for the following 100 days, the 1-day lag-correlation between the ice and NAO will be much smaller in CTRL than OCE, because the contribution coming from days 20-120 is much greater than that from days 1-20. This means that processes happening on time-scales much longer than 1 day (such as ENSO) may well be having an impact which would be reflected both in the lag-correlations and, consequently, the LIM coefficients. In other words, the lag-correlations aren't just showing the “initial” response, but the ice-NAO link as estimated across the entire season.

Another point to emphasise is that the LIM model does generate a process which takes place on longer than daily time-scales. Due to the persistence of the ice, the daily time-scale forcing accumulates over time into a forcing acting on monthly to seasonal time-scales. Conversely, the high persistence of the ice (and the NAO) means that the correlations on daily time-scales may be reflecting dynamical processes taking place on longer (e.g. weekly) time-scales. We added two extra lines in Section 5.3 to help emphasise this. The caption to Figure 8 has also been edited slightly to make it more clear that all daily data in NDJF is used.

We hope these points have clarified any potential confusion.

Reviewer #1: *L537-538: I don't think this accurately reflects the results of Blackport and Screen 2021. They show good agreement across models (the models agree that the connections are weak).*

Response: We rephrased this as follows:

"In particular, the inconsistency across the CMIP6 ensemble (Figure 3) and within long integrations of a single model \citep{Koenigk2017}, as well as the weak signals in large ensemble studies such as \citet{Blackport2021}, appears consistent with a hypothesis that most models fail to simulate a consistent and realistic teleconnection due to inadequate coupling."

Though it is interesting to note that Figure 9b of Blackport and Screen 2021 shows that 2 out of the 5 coupled models do not show an overall significant ice-NAO correlation. While 3 out of 5 does constitute a majority, it would be interesting to see how much agreement there would have been for this particular link (ice→NAO) if more models had been available.

Reviewer #1: *L537-539: I don't follow the logic here. Why would models having inadequate coupling lead to larger spread between models or even within a model? If anything, CTRL has less spread across the ensemble (Fig 3), which would contradict this statement.*

Response: We have argued in this paper that the forced component of the winter NAO is driven by BKS sea ice, in a way which is accounted for by ice-NAO coupling. Inadequate coupling would therefore, in our framework, be expected to break this forced component of the NAO, leaving the NAO to be driven entirely by unforced, internal variability. In our framework, this would mean that BKS-NAO correlations in a random 35 year period are just random draws from the null hypothesis, and therefore have mean 0 with an equal likelihood of being positive or negative. By contrast, the OCE ensemble have a clearly positive mean correlation and all the randomly drawn correlations are positive. It is in this sense that we mean that inadequate coupling may produce a weak and inconsistent signal.

We have rephrased slightly to say "consistent and realistic teleconnection", rather than just "realistic", to emphasise that the lack of coupling would, in our framework, produce a model which is overly driven by atmospheric internal variability, and therefore have an inconsistent teleconnection.

REVIEWER #2

Reviewer #2: *Overall, I find the organization and readability of this paper much improved. I think the streamlining the authors have done around the speculation of mechanisms is helpful, although I also appreciated the thorough analysis in the previous iterations. I think the challenge of writing about the influence of stochastic parameterizations in climate models is explaining why they lead to the improvements that they do. I think this paper is a valuable contribution to the literature in the sense that it demonstrates that a more robust connection between Nov. BKS SIC anomalies and the NAO can be achieved, but there is more work to be done to understand why. Further experiments are needed and the authors have added more about proposed new experiments in the Discussion section.*

Main Comment:
Reviewer #2: *I still find the discussion of the LIM somewhat hard to follow. I guess my question is: if the assumptions of the LIM are not appropriate for the CTRL simulation, as the authors conclude, then what does the LIM analysis add to the paper? As a reader, I am left wanting to gain more insights from this section.*

Response: The LIM hypothesis being inappropriate for CTRL is an interesting result because this hypothesis *is* appropriate for both ERA5 and OCE. This immediately suggests that what is missing in CTRL is precisely what is captured by the LIM, namely continuous two-way ice-NAO coupling. This point is made explicit in the first line of Section 5.4, and also again in the 4th bulletpoint of the Conclusion.

The broader point of wanting to gain more insight is of course reasonable either way. We are very conscious of the fact that we have not decisively understood what is going on in these experiments, and we try to be transparent about this in the paper. The puzzle is essentially to come up with an explanation which explains why OCE has a teleconnection, but both CTRL and AMIP do not. In particular, the fact that AMIP does not have a teleconnection more or less immediately invalidates all the most easy and obvious potential explanations (“the sea ice mean state is bad in CTRL”; “the initial heatflux anomaly is too weak in CTRL”; etc.). The speculation we offer in Section 5.4 are the best ideas we could come up with, but they remain speculation.

We do hope to follow up this work with further, detailed analysis, with the goal of clarifying the mechanisms better. But for the present, we do not have any further insight to add here.

Minor Comments:
Reviewer #2: *Line 95: extra "the" before "land surface".*

Response: Thanks, we corrected this.

Reviewer #2: *Lines 324-326: this seems quite speculative. are there no studies that have looked at this in other models for pre-industrial control integrations?*

Response: After scanning the literature again, we found the paper Kelleher and Screen (2018): https://link.springer.com/article/10.1007/s00376-017-7039-9

This looks at pre-industrial CMIP5 simulations, and does find that there is a significant connection between Greenland sea ice area and polar cap geopotential height anomalies in winter in these models. This seems to corroborate our speculation that other regions, such as the Greenland sea, may have had stronger links to the atmospheric circulation in the past. This is now cited and mentioned in the discussion.

We also added the new Figure B3, in response to Reviewer #1, which shows substantial regional changes in sea ice variability between 1950-1980 and 1980-2015, further corroborating our speculation. We hope the addition of both the above citation and this figure makes the speculation appear more reasonable.

Reviewer #2: *Figure 10: I am a bit confused about Figure 10 and what we can conclude from this. If I consider the arguments of Blackport et al. (2019), at negative lags, when the correlation is positive, doesn't this represent a situation where the atmospheric circulation is driving the ice? Should we not be considering the positive lags to better understand the implications for the Nov. BKS-DJF NAO relationship? Can the authors clarify?*

Response: Yes, you're right in your interpretation that at negative lags, it's circulation forcing the ice (via heatfluxes). The main scenario under which this would be expected to play a role is if remote adjustments to ice/SSTs are important for reinforcing/propagating the initial zg500 anomaly. In such a situation, you'd have the following causal chain:

BKS sea ice anomaly → localised pressure anomaly → forced adjustment to nearby ice/SSTs due to this pressure anomaly → continued growth of the pressure anomaly into a full-blown NAO.

Biases at negative lags in Figure 10 would be expected to tamper with the second arrow, though we now realise we had not made it clear that the conclusion from Figure 10 (“CTRL has an inconsistent ice-heatflux link in the BKS region”) was also found to hold in other regions (such as the Greenland sea). This is a necessary observation for this argument to make sense, since the argument depends on the forcing of the atmosphere on other regions besides BKS being unrealistic, so we now stated this clearly in Section 5.4.