General response

We thank the reviewer for carefully reading the revised manuscript and providing additional feedback in order to improve the manuscript's quality.

Below we provide a one-to-one response to all points raised by the reviewer. The reviewers' comments are in red and our replies are in blue. All line numbers refer to the newly revised manuscript.

Second review of the paper by Riebold et al.

I appreciate the authors effort in addressing point by point the different comments I raised. Overall, I find the revised manuscript much clearer and improved. The authors have added an analysis of the different PAMIP model simulations as suggested in my first review. I find the multi-model results interesting but I regret that none of the figure showing multi-model results has been included in the main part of the manuscript. The multi-model PAMIP results are all in the supplementary material even though they are described in the main text. I suggest moving at least Fig. S2 to the main part of the manuscript. In addition, the different regime occurrence simulated by the different models (Fig. S2 and S3) could have been more extensively discussed before moving to the more in-depth analysis of the ECHAM6 results. This is important as it allows to stress which changes in regime occurrence frequency are robust and which ones are not.

Thanks for the remark, we now included Fig. S2 into the main text and describe and compare the regime occurrence frequencies simulated by the different models more extensively (l. 309-336).

Further, in addition to the 4 supplementary figures there are 7 figures in appendix and I do not really understand the value of keeping them separate. Are the figures in appendix supposed to be more essential to the understanding of the paper than those in supplementary? I suggest putting all the additional figures in one place (appendix or supplementary material) or to better justify the choice of keeping them separate.

Thanks for the remark, as also suggested by the co-editor we now moved the Appendix Figures, as well as the former Figure 2 into the Supplementaries.

The authors argue that choosing ECHAM6 for the subsequent analysis is justified by the fact that ECHAM6 is one of the models that compares best with ERA5 in terms of weather regimes patterns. I do not find this argument convincing as Figure S1 shows that other models than ECHAM6 (e.g. CNRM-CM6-1) show as good results and hence the analysis could have been done on a larger model subset. I understand that the authors do not want to conduct the dynamical adjustment analysis on the 9 PAMIP models but I think they should justify their choice differently. Maybe by saying that the temperature daily data were not available for all models (if it is true) or because the interpretation of 9 models is complicated

and having a first analysis on a single model can serve as a basis for a subsequent multi-model analysis. I do not really agree with this choice of using only one model but I think that the authors should at least provide a honest justification before accepting the paper for publication.

Thanks for the remark. As already indicated in the concluding remarks (l.560) and as correctly suggested by the reviewer, the interpretation of nine models is complicated and the single-model analysis of ECHAM6 can serve as a basis for a subsequent multi-model analysis. We now state this more explicitly in the main text in l. 339:

" We will focus on only one model as especially a comprehensive interpretation of the upcoming decompositions for all nine models is very challenging and beyond the scope of this study."

Detailed comments:

-l.10: add that this is true for at least one winter month

Done

-l.12: This is not consistent with Fig S2 which shows that for ECHAM6, MIROC, E3SM, CESM, the only significant changes are less NAO- days, not more. Please clarify.

The term "most models" was indeed not suitably used when summarizing the NAO response of the different models. We now mention this overall inconsistency of the NAO response between models in the abstract in the following way:

"Forced by future Arctic sea ice conditions, most models show more frequent occurrences of a Scandinavian blocking pattern in at least one winter month, whereas there is an overall disagreement between individual models on the sign of frequency changes of two regimes that respectively resemble the negative and positive phase of the North Atlantic Oscillation. "

-l.21-22: Fixed Regime is sometimes in the manuscript written with capital letters and sometimes not. Same for Changed Regime. Please make it consistent everywhere.

Done

-l.136: "over" should be replaced by "of"

Done

-l.138: "a certain number of members" should be replaced by "at least 100 members"

Done

-l.335: Please specify whether you count only the months where these changes are significant or whether you consider all months. Note that even the models that show an increase it is not true for all winter months. It could be worth stating it more clearly.

We now state this more clearly in the text (l.309)

"All nine models indicate a significant increase of SCAN occurrences in futArcSIC in at least one winter month, while in contrast only the NorESM2-LM and CNRM-CM6-1 models show significantly decreased SCAN occurrences."

-l.354-355: It would be good to add here that ECHAM6 is not among those models, which means that ECHAM6 results are not consistent with the robust changes outlined in Smith et al. (2022).

In l. 327 we now explicitly state that ECHAM6 indicates less frequent NAO- occurrences:

"In contrast, also decreased NAO- occurrences in at least one winter month can be detected in five models as well---including ECHAM6."

-l.363-364: I am not convinced by this argument as stated in my general comments.

See reply to major comment above. We now explicitly state as the main reason of chosing one model that the interpretation of the implemented decomposition for all models would be challenging.

-l.372-373: It is quite confusing to describe the results of the reference simulation saying that it shows more frequent occurrence while in the rest of the section it is the sensitivity experiment changes that are described. I suggest keep the same method and describing everywhere the change in the sensitivity experiment (future sea ice) with respect to the reference experiment (present day sea ice). That would make the results description easier to follow.

Thanks for the remark, we now consistently describe the changes with respect to the sensitivity experiment.

-l.374: Several studies in particular by Screen or Blackport have questioned the validity of using observations or reanalysis to detect the influence of sea ice on the atmosphere pointing out the difficulty to identify causal relationship. Hence, I suggest not insisting too much in the paper on the comparison between ERA5 and the model results (except when validating the weather regimes of course) especially given that the models themselves do not show consistent results for all the features that are described here.

As correctly pointed out by the reviewer the complementary ERA5 regime analysis does not provide any insights into the actual causal sea ice- regime relationship, but is meant as some additional statistical evidence. We now explicitly state this in line 345:

"Such an ERA5 analysis does not prove any causal link between recent sea ice loss and circulation regimes, and does not isolate the effect of recent sea ice retreat. Nevertheless, we consider such ERA5 tendencies as additional statistical evidence, especially when deciding which of the significant ECHAM6 regime frequency changes are considered for the decompositions in Sect. 4.4. "

In addition, we slightly reformulated the beginning of Sect. 4.4 (l. 425) in order to avoid the impression of relying too much on the comparison with ERA5.

-l.411: "Section" should be replaced by "section"

Done

-l.432: Add "," after "months"

Done

-l.605: I suggest better explaining here why extending this analysis to more models has not been done in this paper.

In l. 560 we now elaborate a bit more on why we only focused on one model:

"Furthermore, the results in Sects. 4.2–4.4 can differ for other PAMIP models, but conducting the decomposition method as applied in this study for each PAMIP model individually would be difficult: especially a comprehensive summary and interpretation of decomposition results for different models would be very challenging, in particular due to the fact that each model tends to simulate its distinct significant regime frequency changes in different months. Hence, the presented ECHAM6 analysis might be considered as a first step and adapting the employed decomposition methodology for a feasible implementation into a multimodel analysis might provide a prospect for future studies."

-Figure 3 caption: replace "dotted" by "hatched"

Done

-Figure 6 caption is still confusing as blue and red as supposed to both indicate favored occurrence of cold extremes. My understanding is that red indicates favored occurrence of cold extremes and blue a decreased frequency of cold extremes. This is better explained in Fig 9 caption and should be revised here. Also, in general for the different figures showing maps of the frequency of occurrence of cold extremes (Fig3, Fig6 etc.), the color choice is not quite intuitive and very confusing to me. One would expect that blue means colder conditions, and red means warmer conditions. Hence, I strongly suggest switching or changing the colors to make it easier to understand.

Thanks for the comment, we reversed the colorbar in all plots that refer to cold extremes in order to make the plots more intuitive. (Figs.3a–e,4,6,9,S9,S10)

-Figure A2: the name of the experiments on the figure has not been updated.

Corrected.

-Figure A2, A4, A5: Please add in the caption the model name (ECHAM6) otherwise one could think that all the PAMIP models could have been used here.

Done