Dear Editor and Reviewer,
We would like to thank you for your constructive comments and helpful suggestions, which substantially improve the quality of the manuscript. Answers to your comments are given in detail hereafter.

Reviewer comments are in blue, and are followed by our response (in black) that includes changes and/or additions to the text. All authors agree with the modifications made to the manuscript.

Review report, Reviewer 2

Selivanova et al., examine past and future Arctic sea-ice variability and changes in a subset of models participating in HighResMIP. The authors investigate how high-resolution (HR) and low-resolution (LR) versions of a model affect a range of Arctic sea ice variables, including sea ice thickness, volume, area, and concentration both over the historical record and under a future emissions scenario. The authors find that increasing the horizontal model resolution does not lead to a any significant difference in Arctic sea ice.
This manuscript is clear and the figures are quite clear. Despite the main result showing that HR models does not substantially change Arctic sea ice trends (when compared to LR models), it is important to document and has important implications for future modeling efforts with refined grids. However, I think there are some overlooked aspects of this result that might change the key message. Thus, I think this manuscript should be published after some additional analysis and clarifications. Below I describe these concerns and suggestions.

Major

I am concerned that the authors overlooked the role of internal variability on Arctic sea ice trends and variability. Internal variability is known to be highly model dependent (Bonan et al., 2021) and strongly influence sea ice trends (Swart et al., 2015). I think it would be helpful for the authors conduct additional analyses that examine other members of each model. A quick glance at the HighResMIP archive (https://esgf-node.llnl.gov/search/cmip6/) suggests this is possible for at least some models. For instance, CNRM-CM6-1 has 10 members. If not all models have more ensemble members, it could be worthwhile to focus on comparing HR and LR results in a model with 10 ensemble members (e.g., CNRM-CM6.1). My belief is that HR and LR models will have different "forced" responses and this results itself could broaden the study. I also think the HR and LR models will likely have different internal variabilities based on Fig. 3 which shows that HR and LR models have different SIV mean states.

In summary, I strongly suggest the authors conduct additional analyses that essentially repeat this

analysis but with a more robust quantification of the "forced" response and internal variability.

We agree with the reviewer that internal variability might explain part of the differences between low- and high-resolution configurations. However, a clear assessment of internal variability is not feasable in the context of this paper. Only three models within HighResMIP provide multiple ensemble members (ECMWF, EC-Earth3P and CNRM) for the hist-1950 experiments; only two models of those have multiple members for the highres-future experiments (ECMWF does not provide model outputs for the future runs). Additionally, EC-Earth3P and CNRM have only 3 members which might be not sufficient to capture internal variability, yet. Unfortunately, we cannot overcome the limitation due to model data availability shared with most of the recent studies using CMIP6 HighResMIP. For example, Docquier et al. (2019), Docquier et al. (2022), Koenigk et al. (2019) use ensemble-means only for ECMWF model and the first member for other models. Huang et al. (2021), Belucci et al. (2021), Tsartsali et al. (2022) use only the first member for each model.

Minor

L33 and L38: Cite Fetterer et al., 2016 and Stroeve & Notz, 2018 instead of the https links.

Line 170: Remove capital "A" from addition.

L184-185: It is probably worth mentioning that some models have biases as their summer minimum is in August rather than September like in observations.

Most models and the reference products have similar values in August and September but the annual minimum occurs in September.

L407-409: I would suggest changing models to simulations. Only six models were used.

All minor corrections are done.