We thank the reviewers for their helpful feedback on our manuscript. We address both of the reviewers' concerns and suggestions below. We have included clarifications and changes to the manuscript.

Review 1:

The authors have conducted a study which identifies the direct response of the atmosphere to the 1974 Weddell Sea Polynya (WSP) in Atmospheric General Circulation Models (AGCMs), and compares this to an analysis of the response in the ECMWF Reanalysis version 5 (ERA5). By prescribing ERA5 sea surface temperature (SST) and sea ice concentration (SIC) as surface boundary conditions for the AGCMs, the authors explore the range of direct responses across three different dynamical models, and the response across two resolutions for each model.

Previous coupled simulations have shown an importance of model resolution in their representation of Southern Ocean polynyas (e.g. Lockwood et al. 2021), however there is limited research isolating the model response (as opposed to formation combined with response), and understanding how this varies with model resolution. Therefore understanding the effect of model resolution on the atmospheric response to polynyas is a valuable field of study. Comparing the direct atmospheric response to a reanalysis response can also provide some insight on the pathways for a climate response to the WSP. Finally, the study of the reanalysis response itself, with the state-of-the-art reanalysis model in ERA5 is also valuable in understanding the real-world response to WSPs.

Summarising the above paragraph, in my opinion the three main valuable contributions of this research paper are as follows:

- Exploring how model resolution impacts the atmospheric response to the WSP.
- Understanding the atmospheric-only response, and how this compares to the reanalysis response to provide hints on the relative importance of the direct atmospheric response with respect to responses which include coupled feedbacks.
- Analysing the ERA5 reanalysis response to the WSP.

I also would like to recognise that although the results of this study are primarily not unexpected or conceptually new, I still consider it a valuable contribution to repeat methodologies with state-of-the art models. Additionally, this research extends previous studies with a multi-model, multi-resolution framework to provide a comprehensive study of the direct atmospheric response to the WSP.

We thank the reviewer for their summary, which highlights our contributions, while acknowledging they are not unexpected. As we emphasize in our response to reviewer #2, our findings remain speculation until they are actually computed and disseminated to the community. Also, they are not expected to all researchers, especially those who have suggested large atmospheric responses and coupled feedbacks in the polynya dynamics.

Indeed, our results are somewhat unexpected “negative” results, but they set a robust basis and a novel perspective on how to interpret the dynamics of the Weddell Sea polynya in the coupled system.

Suggested revisions:

In my opinion, the authors need to take care that the wording of conclusory remarks remains within the scope of this direct atmosphere-only response study.

For example, the authors' statement "...our results suggest that a 1974-size polynya (the largest observed) cannot generate a response that would project on the scale of the Weddell gyre." is too much of an extrapolation, the study is isolated to atmosphere-only dynamics, and so can not alone rule out wider responses which could arise from coupled feedbacks.

We have edited this statement as follows: “our results suggest that a 1974-size polynya (the largest observed) cannot generate a *first order atmospheric* response that would project on the scale of the Weddell gyre."

Similarly, I think the following statement should be reworded, its wording implies that the results of this study prohibit a large-scale interaction of the WSP with the climate, whereas in reality, the results only suggest that because the reanalysis results and atmosphere-only results are similar, that the influence of the coupled feedbacks on the atmosphere are small. The WSP may influence wider ocean circulation, which could in turn have climate impacts elsewhere, and this possibility cannot be ruled out by this study alone: "...the WSP may not interact with the climate on a large scale and may be too far south and within sea ice edge for coupled feedbacks to have a substantial impact on results".

We have edited this statement as follows: "...the WSP may not interact, *to first order through the atmosphere* with the climate on a large scale and may be too far south and within sea ice edge for coupled feedbacks to have a substantial impact on results".

Again in the statement "In addition, our results show that the response has little memory (i.e., local in time too) and vanishes rapidly with the polynya", the authors need to clarify that this means that there is little memory in the atmospheric-only system. Memory in a climate system is often established via ocean-atmosphere-sea ice feedbacks, which by design are not present in this study.

We have edited this statement as follows: "In addition, our results show that the response has little memory *in the atmospheric-only system* (i.e., local in time too) and vanishes rapidly with the polynya".

The statement on line 94, "The interpretation of coupled models is made difficult by the potential impact of ocean-atmosphere-sea ice feedbacks", needs to be justified. A reader might argue that a coupled model is easier to interpret, because it includes feedbacks which an atmosphere-only model does not, and the processes which drive these feedbacks exist in the real-world. Comments such as this, and as those quoted above, present the idea that these direct atmosphere-only simulations are a better alternative to coupled simulations, when in fact these different methodologies answer different scientific questions.

We have edited this statement as follows: “The interpretation of coupled models *and of the direct response to the WSP* is made difficult by the potential impact of ocean-atmosphere-sea ice feedbacks ".

The authors have discussed the fact that they are using ERA5 before the assimilation of satellite data, and commented on potential data quality issues. They have also cautioned the reader that the ERA5 data is the result of one realisation. I would therefore strongly recommend that the authors utilise all ten members of the ERA5 reanalysis to improve their analysis of the response to the polynya in ERA5, and to provide uncertainty estimates on this response. ERA5 is likely weakly observationally constrained in the Southern Ocean, particularly in the pre-satellite era, which means that the underlying model may play a significant role in determining the reanalysis conditions in the area of interest.

We have edited our analysis and the manuscript as suggested (figure 8), using all 10 members. We find that this does not change our results, but makes them more reliable.

With regards to the statistical methodology, I would suggest that the authors consider employing multiple-hypothesis testing when applying the students t-test to potentially spatially correlated data (e.g. surface temperature and precipitation response). As described by Wilks (2006), this is a mistake that is made by standard in the field. In my opinion, at least acknowledging this potential statistical caveat is an important step in improving the statistical robustness in our community.

We thank the reviewer for bringing this caveat to our attention. We have redone the statistical analysis to include False Detection Rate to our t-test p-values for figures 2,3,4,5 and 8. This new analysis shows that the dynamic response (and that of the response aloft) is less significant than suggested by our original statistical analysis.

However, in our previous draft, we had already downplayed these responses on the basis that they were inconsistent across models and had suggested they were due to internal variability. The new analysis reinforces and formalizes further our key message.

The paper is generally well written and presented, but could also benefit from a proofread. I have included below a list of some of the minor errors which I noticed, but I did not have the time to go through the entire article from a spelling and grammar point of view:

Spelling mistake in line 16 "isolate" should be "isolates".

Spelling mistake in line 68 "impact" should be "impacts".

Spelling mistake in line 112 "your" should be "our".

Line 121 "...thus, has limitations" the reason for limitations needs to be clarified a little.

Line 180 "limiting".

Authors should decide to use "sea ice" or "sea-ice" but not both.

We thank the reviewer for pointing these mistakes out, we have made corrections and further proofed the manuscript.

Finally, I would like to thank the authors for their contribution, which is a very interesting research topic.

Kind regards, Tarkan A Bilge

Lockwood, J.W., Dufour, C.O., Griffies, S.M., and Winton, M.: On the role of the Antarctic Slope Front on the occurrence of the Weddell Sea polynya under climate change, J. Climate, 34, 1–56, 2021.

Wilks, D. S., 2006: On “Field Significance” and the False Discovery Rate. _J. Appl. Meteor. Climatol._, 45, 1181–1189, https://doi.org/10.1175/JAM2404.1

Review 2:

I appreciate the authors' effort to address some of the issues mentioned in the earlier manuscript review. Most of the authors' rewriting efforts greatly help better understand the results and discussion section of the paper. However, that alone does not recommend publication, considering that, as the authors themselves state, "the results are somewhat expected." The authors have presented no new analysis other than what was in the original submission.

In our earlier response to reviewers, we did point out that our “results are somewhat expected”. However, the reviewer did not account for other points made in our response, where we justify the study’s contribution to the field, both as part of the review and in our manuscript edits. Additionally, we hope and strongly believe that publication should not be based on results being expected or surprising. Many researchers start out a research project with expectations about their results, which is often a good practice. We expected that the direct atmospheric response to the polynya was small, and we did find this. It does not make our work irrelevant or uninteresting to the community and our work still very much delivers a novel result.

Prior to our study, the global three-dimensional atmospheric response to the polynya had not been assessed in detail. As we have highlighted in our introduction, previous studies used a two-dimensional domain, regional scale domain, or coupled models where interpretations are ambiguous due to the effect of secondary feedbacks with the ocean.

We have, for the first time, established the direct atmospheric response to the polynya. We have done it in a robust manner by considering 6 set-ups (3 models at 2 different resolution each) to avoid the usual uncertainties in our field of results relying on one single model. We have analysed the response and compared it to previous simulations and inferences from reanalysis to the extent it was possible, which we have extended upon in this latest version of the manuscript, after suggestions from reviewer #1.

The response is highly localized in space and time, where the small magnitudes of the responses do not lend themselves to generate a complex story. Additionally, we find that the limited dynamic response to the polynya that we downplayed because it was not robust across our set of models is not statistically significant according to the False Detection Rate analysis suggested by reviewer #1, reinforcing our conclusions. We nonetheless explored the link with the southern jet stream and SAM which have been suggested in past studies, and only found a tenuous link. Finally, our results suggest that inferences about the polynya dynamics from coupled models with overestimated polynya size should be taken with caution.

This also emphasizes that, while the results were expected to both us and reviewer #2, they would be unexpected to researchers who have argued that the polynya has a large hemispheric scale impact on the atmospheric circulation and interpreted the polynya as part of a coupled mode (e.g., Diao et al., 2022; Weijer et al., 2017). We could have adopted their perspective and labelled our results as unexpected based upon these studies.

Our results should be published so future studies of the polynya dynamics can account for coupled interactions and feedbacks and the size of the polynya. Hence the importance of 'negative results' such as those presented in our study. Expected or not, the direct response to the polynya remains speculative until it is actually computed and disseminated to the community.

The manuscript does not have enough analysis to present a well-rounded story that would do justice to the goal of the manuscript. The authors do not save any of the outputs related to cloud characteristics, which is strange considering they wanted to study the atmospheric response. Previous studies that look at the atmospheric response to a Weddell Sea Polynya have found that most of the atmospheric response is limited to the vicinity of the open ocean polynya in terms of cloud and precipitation changes. Thus, the authors would have wanted to save the atmospheric variables that quantify the cloud characteristics and analyse the difference in cloud and precipitation characteristics with changing resolution.

We understand that the reviewer is particularly interested in clouds and radiations in the vicinity of the polynya, but when we started the project, our concerns were much wider. We had questions about the atmospheric circulation on a global scale, justly based upon the previous literature. Unlike the suggestion made by the reviewer, there wasn't a consensus about a localized response to the WSP. In fact, previous studies have suggested impact as far as the tropics (e.g., Diao et al., 2022, Chang et al., 2020, Kaufman et al., 2020). Again, this may reflect that reviewer #2 have their view on what the atmospheric response to the polynya is, but this view is not what appears in the literature (including recent papers).

Upon setting up our models, we had to make choice about which variables to save, at which frequency, under the constraints of limited storage (we nonetheless reached over 20 TB of data across all six models). A posteriori, with a different perspective one might want to rerun all the simulations; however, this is not possible within the resource and time limitations of our project. Our manuscript already presents changes in surface fluxes including precipitation from the submission and also provided further radiative fluxes changes in response to the earlier reviewer's comment that support said conclusions.

Furthermore, it is justifiable that at least one of the simulations should have a higher spatial resolution of ~0.1°.

We argue below that the request from the reviewer is unreasonable and unjustified.

It may be "justifiable that at least one of the simulations should have a higher spatial resolution of ~0.1°", but the reviewer does not provide such justification, here nor in the previous review.

We considered what could justify doing so under the constraints of limited resources. The most obvious is that the science question dictates simulations at 0.1° to resolve the process we investigate (for example, simulation of mesoscale front embedded in a synoptic eddy). We carried out simulations with a range of resolution from 0.22° (25 km) to 2.7° (300 km), which show, for all purposes of the study, the same features, except for an increase in the magnitude of the response, which we explore. There are no indications that something drastically different would happen at 0.1°, until perhaps the convection scale is reached but this is much higher than 0.1°, and so the reviewer's focus on 0.1° remains unclear).

Our range of resolution covers the overwhelming majority of the resolution range currently used in global three-dimensional simulations in the atmospheric and coupled modelling (our range covers the entirety of the CMIP6 atmospheric models and a vast majority of high-res simulations such as those of PRIMAVERA project).

It is always possible to ask for more resolution (until the Kolmogorov scale is reached) and one could always speculate that something unexpected would happen at higher resolution. We do not think this is the right approach to planning research. The reviewer's request is a rather blunt statement that could apply to 100% of the published atmospheric modelling literature. Such reasoning could equally apply to observations, one could always ask for a denser array of instruments.

Simulations at 0.1° resolution on the global scale are extremely expensive, and we simply do not have the means to carry such an experiment. Such resolutions begin to bridge the gap between climate simulations and weather forecasts, such as the MPI ICON model, which is not the standard for the nature of our, or any related literatures study.

Working with limited resources we had to make choices to use optimally our resources. We choose to invest in a range of models and resolution that would establish the robustness of our results and cover the range of resolutions that the climate community uses (note that our 25 km runs correspond to the highest atmospheric resolution submitted to CMIP6).

Finally, to emphasize, computer resources roughly scale as the cube of the resolution (doubling the horizontal resolution quadruples the number of grid points, and halves the time steps, i.e. 8 times more resources). A 0.1° resolution runs would use (2.5)^3=15.6 time more resources than our 0.25° runs. In other words, a 0.1° run would have used about *5 times the entirety of the resources* invested in our 6 simulations.