We thank the reviewer and our editor for their helpful comments. We hope that through our efforts the manuscript has been improved and the reader will find it clearer.

In the following we answer to the comments individually:

Second review of " On the drivers of regime shifts in the Antarctic marginal seas, exemplified by the Weddell Sea" by Verena Haid, Ralph Timmermann, Özgür Gürses, and Hartmut H. Hellmer

The manuscript presents a set of model experiments used to explore a potential regime shift in the Antarctic marginal seas, with a strong focus on the Weddell Sea. The authors use various modifications of atmospheric forcing data to study the development of the HSSW and WDW density and find that the density difference between these water masses is a deciding element for shifting into a warm regime. The results agree with other model studies using various forcing combinations, but in this study, the regime shift is reversible, provided that the forcing returns to present-day conditions.
The authors have considered most of my major concerns in the revised manuscript, which is improved significantly. The storytelling works better, and it is much easier to follow the structure. However, I would still encourage the authors to look into some of the concerns raised in major issues 1 and 2 from the previous review round. My remaining major concerns are, therefore, closely linked to these.
Major issues:
1) There exist many studies on a possible regime shift in the Weddell Sea. Your main result appears to be that the regime shift strongly depends on the density differences between the off-shelf warm water and the on-shelf dense water, similar to several previous studies. Can you make a stronger introduction to show more clearly how this study differs from the others and what we can learn from it?

We added further information toward this goal in the introduction. We hope that the reader will now have a clearer idea of the goal of our study and what sets it apart from previous studies.

2) The description of the experiments has improved from the revised manuscript. However, it would be helpful to include explanatory text to the acronyms and the similarities/differences of each experiment. For instance, at the beginning of chapter 2.3.3, you could say something like: Three experiments look into seasonal alterations, denoted by SA in their acronym/name, and two experiments focus on extending the summer season (and making the winter shorter and milder), denoted by SUMMER in their acronym/name: Could you, perhaps, also make hypotheses you hope to answer based on sub-sets of the experiments to show the purpose of each experiment more clearly?

Thanks for the nice suggestion. We added the sentence (slightly adapted) to the beginning of Section 2.3.3.

3) The forcing data needs some more explanation. I appreciate the figure in the supporting information, although the caption needs some more explanation (see comment in annotated pdf). I would like to see an explanation of why you added the difference between the seasonal signal in ERA Interrim and HadCM3 21C-A1B. I guess that the regular model forcing already contains the seasonal

signal in ERA Interrim and that you subtract this from the HadCM3 21C-A1B to make sure it is not doubling up – i.e., you only add a seasonal variation that represents the increase from ERA Interrim. But this is not described in plain language.

We first only added the difference between HadCM3 20C and HadCM3 21C A1B as a seasonal signal to the ERA Interim data. This was not enough to trigger a regime shift. Then we went further and practically forced the atmospheric forcing to follow the HadCM3 21C A1B seasonal cycle. The seasonal cycle from ERA Interim is indeed used as the reference for computing anomalies, so the anomaly only represents the difference and does not double up the seasonal cycle. This is also explained in the beginning of Section 2.4.

Our goal was to force a regime shift on the continental shelf. As we mention in the discussion, we only successfully triggered a regime shift with forcing in some way related to HadCM3 data (especially the wind field). Other attempts remained cold on the continental shelf. There may be something specific to the HadCM3 data that renders the southern Weddell Sea continental shelf particularly vulnerable for a regime shift and it may in reality be more stable. We are also convinced that there exist other (more rigorous) manipulations of the forcing data that would result in a regime shift on the Weddell Sea shelf, but we tried to avoid applying unreasonably large perturbations.

All this, however, does not touch on the main result of our findings (that are based on more than just the featured experiments) that we found only the density evolution as a reliable decisive factor for a regime shift.

4) In the letter to reviewers, you explain that, in the SUMMER simulations, you are also providing shortwave radiation for summer conditions in this extended summer. This information should be provided clearly in the manuscript. Exaggerating a season response (e.g., wind/temperature) is one thing, but changing solar radiation beyond what is natural is not something I would recommend. I understand if this choice made it easier to design the experiments, but it needs to be mentioned so that the reader understands what the experiment represents - and a discussion of the caveats should be included in the manuscript.

This fact is not hidden. The text does not exempt any variable from the alteration. It is also plain in Table 1 and Fig. S2. However, we have now added a specific statement that this includes shortwave radiation despite it being a physically questionable choice.

A series of minor issues are included in the annotated manuscript pdf

From the annotated manuscript:

l. 9 this sentence is not complete. Please revise.

To increase readability, we now repeat the verb.

l. 10 This is a too strong claim! I would suggest to say that you could not identify a universal recipe.

We changed our statement to '… a unique and universal recipe … **was not found** …'.

As we described, we found several different ways to trigger a regime shift, admittedly related, but different. Reduced to the fundamental basics there are the two extreme ways to reach the trigger point: 1) decrease the density of the dense on-shelf waters below the density of the off-shelf waters

at sill depth, or 2) raise the density of the off-shelf waters at sill depth above the density of the dense on-shelf waters; and any variation of density evolutions that result in the off-shelf water getting denser than the on-shelf water. Therefore, we do think it an adequate statement that there is no 'unique and universal recipe' for a regime shift.

We now also slightly adapted the wording in our conclusion for clarity (l.411): '… that **there is not one** universal recipe for a regime shift in the Antarctic marginal seas, but that such a shift can occur under diverse circumstances and depends on local influences ranging from bottom topography to atmospheric conditions.'

caption Fig. 1 Could it be an idea to include the acronyms for the simulation in the figure. This would support the text in the discussion.

We provide such a figure in the supplements (Fig. S7).

caption Fig. 2 It would be good to include the circulation of cold water in the FRIS cavity. The HSSW that enters along the Ronne and the Berkner island part of the ice front is cold (it is produced by cooling and sea ice formation). The yellow arrow in the cavity which is said to be warm in the caption could be misleading.

We understand that our interpretation of 'warm' here, as transporting heat toward the ice shelf, is in conflict to our general differentiation of warm vs. cold water on the continental shelf. We changed the colour of the arrow of HSSW penetrating into the cavity to purple (cold) and have added a comment in the figure caption mentioning it still being a heat source despite its low temperatures.

l. 89 could you explain this acronym? (SRES-A1B)

We now write the full meaning of SRES and added this reference for the interested reader:

*IPCC SRES (2000), Nakićenović, N.; Swart, R. (eds.),* Special Report on Emissions Scenarios: A special report of Working Group III of the Intergovernmental Panel on Climate Change (book)*, Cambridge University Press,* ISBN 0-521-80081-1, https://www.ipcc.ch/site/assets/uploads/2018/03/sres-en.pdf

l. 98 I would suggest to rewrite/reformulate so that it does not sound like a surprise or a problem. Something like: None of our experiments with varying forcing alterations based on ERA Interrim, resulted in a regime-shift." And perhaps include a comment if this was surprising or as expected.

We reworked the text here.

l. 101 inspiration?? I would rather say that you wanted to recreate a scenario that was shown to give a regime-shift in Timmermann/Hellmer, and then explore how extreme the forcing need to be...

We did not want to recreate the same scenario in SA_G, we deliberately introduced measures to ensure e.g. that short term events still have the full variability of the ERA Interim data. Experiment 46 in Table S1 came close to reproducing a very similar scenario to the previous studies, although we skipped a few years in the lead-up. We did so as a first step, to test whether further efforts are worthwhile. For SA_G, our goal was to step away from the HadCM3 data (while still triggering a regime shift). Unfortunately, we found that if we stepped too far from the 'original' the regime shift didn't occur (within four to eight decades), see e.g. experiments 47-49 in Table S1 or SUMMER_S. That is not to say that there do not exist other atmospheric forcings triggering a regime shift unrelated to the HadCM3 A1B results.

l. 102 Therefore, we have focused this study on the seasonal variation to gain a better understanding...

We altered the text to read '…, therefore we focused on modifications of the seasonal signal in our forcing perturbations.'

l. 115 I would recopmmend to add an explanatory sentence here. E.g. Three experiments look into seasonal alterations, denoted by SA in their acronym/name, and two experiments focus on extending the summer season (and making the winter shorter and milder), denoted by SUMMER in their acronym/name. It would also be helpful if you explain what you want to study in these experiments - perhaps include hypotheses.

As previously stated, we thank the reviewer for his suggestion and added a slightly adapted introductory sentence.

We also tried to enhance memorability of the acronyms by printing selected letters in bold font. We now enhanced this by using all letters of the acronyms in the order of their appearance and in capital letters where applicable.

l. 116 Is this done for each grid-cell in ERA-I, or for a region-dependent seasonal anomaly?

It is a locally (individually for each grid cell) varying anomaly and not specifically region-dependent. (Of course, the global mean of the atmospheric variable depends on the choice of the regions in which the anomaly is added. Within the Weddell region, however, SA_G, SA_S and SA_W all experience the same forcing.)

l. 117 It is not straightforward from the figure caption which forcing fits to which experiment.

We cite the relevant experiment names in the figure's caption.

l. 123 (Fig S2b) This figure is hard to read. Could you, perhaps, show only one year and make it visible where the january month repeats - and what happens after this repeat? perhaps this could easily be fixed by adding more text to the caption. After spending a lot of time figuring out what is happening, the green and red portions of the graph make sense, but please add to the caption.

This figure is intended as a visual aid to understand the description in the main text. We intentionally chose to show several years in order to also show how short-term and interannual variability is treated in the process. To help the reader along, we now added references to the subsections, where the method is described in the manuscript.

l. 129 It would be helpful with more text explaining what you want to accomplish with these runs. What are you looking for?

We added an explanation here.

caption Fig. 4 it is still very hard to see the details of the different responses in the experiments. Could the figure fill the whole page width, or could you plot the experiments as anomalies from REF?

Unfortunately, the journal's template does not encourage wider figures. We rearranged the panels instead to increase panel size.

caption Fig. 4 This should be moved to the main text and explained a bit more. Observations of warm inflow typically show that these occur at mid-depth. However, when averaging over three years, such intermittent episodes are smoothed out, and would may not be dominating for the temperture maximum. I would, nevertheless, welcome some comments on this matter. Did you, e.g., do anything in your code to ensure you did not plot surface water?

The plot shows the maximum in the water column of the temporal mean value. Thus, generally one can avoid picking up (summer) surface values. On the continental shelf, the warmest water is usually found at the bottom (as the caption now states), while in the Weddell Gyre the maximum temperature is found in the WDW layer. (As a check: Had the algorithm picked up surface values locally, one would see this as patchiness in the plot.)

caption Fig. 5 Could this area be marked in figure 4?

It is now marked in Fig. 4a.

l. 215 I would suggest to first describe the stable conditions (what measures are relevant for this), and let the reader know which experiments fall within the description of stable conditions.

You could then describe the features that are common for experiments that leads to a regime shift, and go into details of these.

We now added a short explanation on what a 'stable' condition means here.

l. 215 and what is this? It would be great to comment here as an overarching description of the findings from this experiment.

Or is SUMMER_S in fact demonstrating the opposite - factors that do not keep the shelf stable...

The explanations follow in the next paragraphs.

l. 223 Please, also provide numbers for SUMMER_S

We have not provided numbers for the ACC for any of the experiments, but Fig. 7 gives a good overview on the transport magnitude and how it changes. (For the Weddell Gyre, we provide transport numbers for all the experiments.)

l. 226 Could you describe the conditions for all experiments with a regime shift, and then say what (why) is different/interesting with SUMMER_S?

We tried to arrange the material in a way that (as much as possible) naturally leads to the introduction of the density balance as the decisive criterion. In combination with SUMMER_S, SUMMER_S+SAw_W makes a very nice case for the importance of the density balance of HSSW and WDW, also illustrating that each of the two players contributes to the decision.

l. 235 Could you please explain in the text why you compare the bottom Ronne densities with the WDW at the Filchner Trough (FT) sill depth. When the inflowing warm water enters from the FT, it

must be denser than the water residing in the FT (usually a mix of HSSW and ISW), which does not need to have the same density as the HSSW that enters the cavity in front of Ronne. See e.g., Hellmer et.al., (2017). When you compare with a different water mass than other papers, I think you should comment on this here? Could it, e.g., be that the warm water enters the western part of the shelf, and not in the FT in your models?

As can be seen in Fig. 4, the warm water clearly enters the cavity through FT, not through the western part of the shelf.

We are comparing the potential candidates for filling the deep cavity, which are the densest 'warm' and the densest 'cold' water. The densest 'warm' water that could potentially flow into the Filchner Trough is the WDW at sill depth. The densest 'cold' water is found along the Ronne Ice Shelf front, not only because it is the site of the most active polynya in the region (e.g., Haid and Timmermann, 2013), but it is also downstream from other active polynyas, leading to the highest salinities. This is why we pick this water mass to compare with. As we state in our paper, the fact that the inflow paths are spatially separated makes it possible to get 'crossing' density evolutions; e.g. in the Amery case a separation of the water masses is not so easy and the density curves follow each other once they 'meet'.

We explain which water masses we are comparing in the paper (l. 249-257) and added a few more words on why. It now reads: 'The crucial criterion for the warm water to cross the continental shelf and fill the ice shelf cavity turns out to be the balance between the density off-shelf at sill depth and the density of the densest water produced by sea ice formation on the continental shelf. **These two water masses are in direct competition for filling up the depths of the ice shelf cavity. In the stable 'cold' situation, HSSW fills the cavity, while after the regime shift, the warmer WDW/MWDW replaces it.** In the Weddell Sea, the most active polynya exists in front of Ronne Ice Shelf (Haid and Timmermann, 2013) with a seasonal fluctuation of the location of the densest water that is either found on the eastern or the western side of the Ronne Ice Shelf front. We therefore compare the maximum bottom density found along the Ronne Ice Shelf front as a measure of HSSW properties with the bottom density at the shelf break at a position east of (i.e. upstream from) the Filchner Trough sill at 670 m depth (approximate sill depth, location L1 in Fig. 2) as a measure for the properties of the WDW **that is a candidate to enter into the Filchner Trough**. '

l. 295 what do you mean by this? Are you thinking about thermocline depth?

Right, we now speak of 'ASF depth'.

l. 300 or is it the density of the water present in the FT?

Essentially, it has to outweigh both. But the HSSW formed by sea ice production at the Ronne Ice Shelf front (very active polynya site) is denser than the water found in the Filchner Trough (which – in the stable case – is cooled down HSSW with an admixture of melt water).

We also explain this in l. 253-255.

l. 316 I would combine this with the previous paragraph

We did so.

l. 340 I would expect a discussion of the summer experiments. These have shortwave radiation for summer conditions for a longer time of the year than is natural, and it is necessary to comment on the efffect of this here.

We added a comment on this.

l. 373 Thi paragraph fits metter in the method section. It is not dicussing your result from this study.

The paragraph was shortened and now sheds some light on the limits of our study. This is to put in in the right perspective in the larger context and therefore we think it is well placed at the end of the Discussion section.

l. 393 Can you say this with certainty? I am aware that you ran a lot of experiments, but this is a very strong claim.

We slightly altered our wording here to make the meaning of our statement clearer.

Editor decision: Publish subject to minor revisions (review by editor)

by Karen J. Heywood

Public justification (visible to the public if the article is accepted and published):

Your paper was returned to one of the original reviewers, and I have also assessed it myself. The reviewer appreciates the revisions that you have undertaken, but has further suggestions for strengthening the paper. See their report, and also their annotated pdf. If I consider that your responses and revisions sufficiently address the reviewer's concerns, it should not be necessary to send the paper out again to review.

I also have a few minor comments myself:

L3 I don't think multiplication is the right word? enhancement?

We agree and replaced 'multiplication' by 'strong enhancement'.

You summarise nicely some previous relevant studies in the Introduction. It would be helpful to identify the gap in our knowledge or understanding? i.e. to state at the end of each Intro paragraph what we don't know, rather than only what we do know? And clarify how your model experiments build upon or compare with previous studies (e.g. you might do the same experiment but with a different model, or different forcing, or run for longer, or whatever). E.g. in L52 you say that you investigate triggering mechanisms for the cold to warm regime shift, but as you just described various mechanisms, it would be good to state your goals and how it differs from the previous work?

We added more information on how most of the suggested criteria failed to accurately categorize our experiments and a very brief description of our model set-up (the specifics that sets it apart from previous studies). We hope that the reader will now have a clearer idea of the goal of our study.

L58-60 I would call these sections, not chapters.

We agree to the suggested change and changed the wording from 'chapter' to 'section'.

L90 I would say “used by” and “it led”.

Yes, thank you. We changed it accordingly.

L98-103 I agree with the reviewer’s thoughts about justifying the experiments you will show – we don’t actually need to know what your rationale was at the time, only how it forms the story you are telling now. If it was a surprise that the regime shift wasn’t triggered, that’s interesting isn’t it? How does your forcing compare with the other studies you mentioned in the Intro?

We added to the text, with more information on the meaning of a failure to trigger a regime shift with simple measures, and on alterations that were similar to (or inspired by) other studies.

Many thanks for sending this interesting paper to Ocean Science; I enjoyed reading it.

Thank you!