Response to reviewer #1

One major issue I have with this study is that the authors chose to focus on one of the most difficult regions in the World Ocean to test the method. The Antarctic continental shelf suffers from one of the poorest data sampling, making any climatology at best questionable. It is also a region with very complex bathymetry and complex dynamics making its representation in current ocean models utterly difficult. One may wonder why the authors should test their statistical method on such a complex region. This raises the question of what is the main goal of the study: validate the statistical method or discuss the representation of hydrographic properties in a climate model?

Thank you for your comments and suggestions. We intended this manuscript to be a methodology paper. However, the methodology is not simply clustering (which could be tested elsewhere), but also developing and evaluating a mechanism for comparing hydrographic characteristics regionally averaged quantities in regions that are sparsely sampled. Thus implementation in a poorly sampled region is, in our view, necessary. Our goal is to demonstrate that clustering identifies hydrographic regimes that are common to different source fields (model or data), while allowing for biases in other metrics (e.g., water mass core properties) and shifts in region boundaries. Although we did not apply this method to scattered observations, this clustering/aggregation approach will be of particular importance if we want to avoid interpolation artefacts and use scattered, sparse observations (as noted in the discussion Line 358-368). We emphasized these points in our revised manuscript (Line 81-84): "***In sparsely sampled regions, grid-point based comparisons (e.g., Little and Urban, 2016) are thus of limited utility, and may underestimate uncertainty in the reference (observational) product. We suggest that it is often more meaningful to assess GCMs using a regionally averaged approach.***"

If the main focus is on the statistical method, I suggest the authors present a similar study in a more favourable region. (We justify our choice of a "difficult" region in the previous comment.) At least they should use better products than WOA as the reference, such as SOSE or MIMOC for the data product and present a comparison between more products to give confidence that results have any degree of generality.

The WOA (version 2018) is the most recent version of this widely-used ocean state climatology dataset. It is exclusively based on the observational data from different platforms, while the SOSE is a model-based reanalysis dataset, and the MIMOC uses only the Argo float and shipboard CTD data (i.e., a subset of data used in WOA). The preliminary comparison indicates these other datasets are not markedly better, although a more complete analysis is a valuable suggestion, which we hope to pursue in future work.

The authors should also explain on which basis they have decided to compare each of the five WOA and CESM groups one by one. Comparisons of geographical distributions in Fig 3 and of properties in the metric space in Fig. 4 shows very little resemblance, so it is not clear at all that the two fields can be usefully compared at all.

We disagree with the reviewer that the geographical distributions show very little agreement, especially for the "optimum" case of K=5. Also, recall that our metrics are relational (e.g., we want to discriminate between low/high $S(T_{min})$ vs $S(T_{max})$) and we make no a priori assumption that CESM2 does look like WOA.

Although the CESM2 exhibits large biases, the primary hydrographic regimes, i.e. groups 1, 4 and 5 of the CESM2 in the Antarctic Continental Shelf Seas (ACSS) are still consistent with WOA (Figure 3e and 3f). The CESM2 clearly shows similar distributions for groups 1, 4 and 5 (Figure 4c) as the WOA (Figure 4d), although the CESM2 has a much coarser resolution. These consistent hydrographic regimes can also be seen from their T-S properties in Figure 8c and 8d. Because these primary hydrographic regimes at boundaries of the study domain are comparable between CESM2 and WOA, we further looked at their differences in water properties and locations. We added more details in section 3.4 about the possible missing parameterizations leading to misrepresentations of the CESM2 in the ACSS based on the T-S properties of the clustered groups.

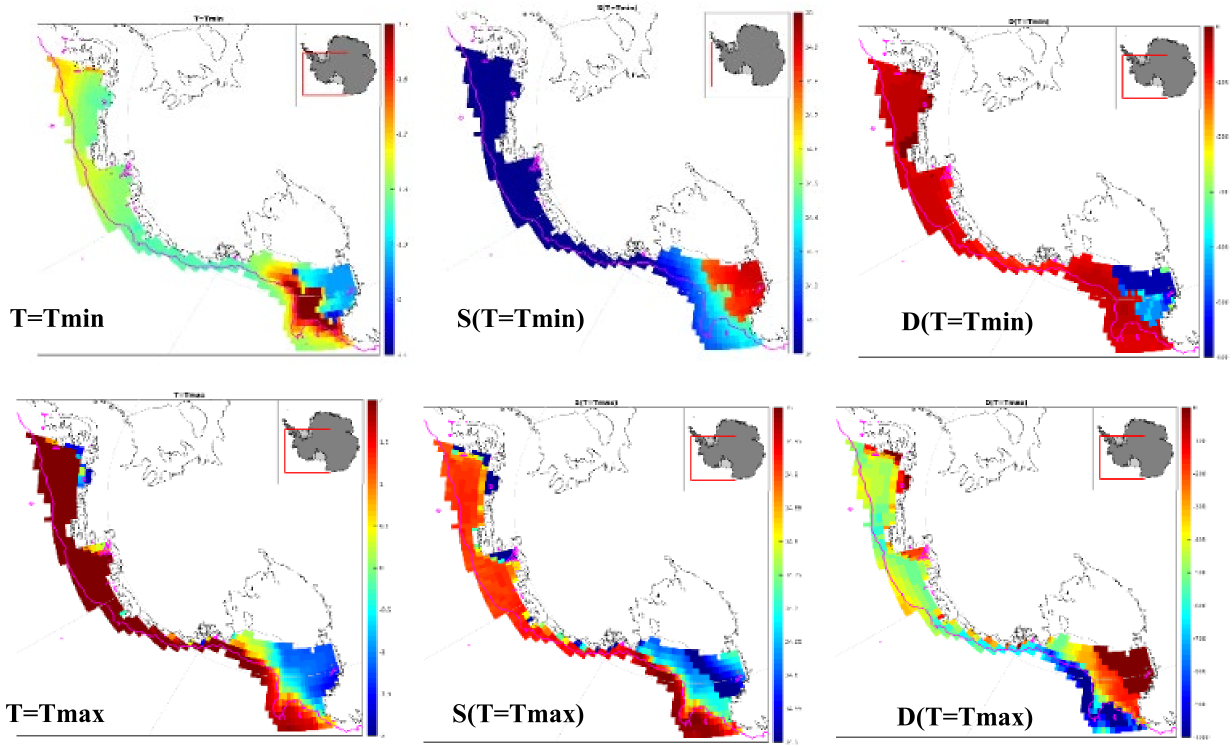
If the main focus is in discussing the degree of realism of CESM, important information should be added about this run. Technical details are missing, such as the vertical resolution of the ocean component, the type of atmosphere model, the representation of sea-ice in the model, and the parameterization of mixing processes.

The main focus of this manuscript is to introduce a clustering methodology to assess representations of hydrographic regimes in fields from diverse models and/or datasets. Our choice of metrics defines the properties of fields from different sources that we focus on, and we don't assume that perfect agreement is likely. But we also provided some additional explanations for why the groups' geographic distribution differs in the CESM2 (Line 293-296) "***Sea ice concentrations are biased low in CESM due to positive zonal wind stress biases in the Southern Ocean (Singh et al., 2020). This wind stress bias may, in turn, lead to an overestimate of the upwelling of warm and salty CDW onto the ACSS. The limited extent of the coastal fresh-water-enriched regime (group 1) in CESM2 may result from the absence of basal melt from ice shelves.***", and Line 304-308): "***It is possible that these differences result from the overflow parameterization in CESM2 (Briegleb, Danabasoglu, & Large, 2010). In this parameterization, locations of the on-shore source water at its formation regions and off-shore entrainment, which mixes with the source water to produce the final water mass, are defined, and overflow water is routed to fixed locations. While this parameterization allows transport of HSSW to the Southern Ocean, it is entirely artificial and does not represent on-shelf mixing processes.***". The references for the detailed CESM2 setups and CMIP6 forcing are also included (Line 119-126): "***We compare the Community Earth System Model version 2 (CESM2; Danabasoglu et al., 2019) to WOA for the same period and domain. The time-mean model salinity and temperature fields over the 1995-2004 period are calculated from the monthly output of the Coupled Model Intercomparison Project Phase 6 (CMIP6) historical simulation (experiment tag r1i1p1f) (Eyring et al., 2016) at the native ocean model resolution (roughly 1 degree in longitude and 0.5 degree in latitude). CESM2 uses the CICE5 (Hunke et al., 2015) sea ice model; however, dynamic and thermodynamic interactions with land ice are not represented (Danabasoglu et al., 2019). The CMIP6 forcing data is described in Eyring et al. (2016) and can be download from input4MIPs CoG (https://esgf-node.llnl.gov/search/input4MIPs).***".

Also, the equivalent of Figs. 1, 2 and 5 using CESM data is missing. I have the impression that most conclusions drawn in the study regarding CESM could be obtained by simple visual inspection of such added figures.

Firstly, we would like to clarify that the main goals of Fig. 1 are to provide the regional setting, demonstrate that neighboring regions can have similar (e.g., Bellingshausen and Amundsen seas) or very distinct (Amundsen and Ross seas) vertical profiles of T, S, and that we'd prefer to have an objective technique for defining "regions" rather than simply setting arbitrary boundaries. As the aim of Figure 1 to motivate the use of an objective method to segregate the ABRS into regimes, rather than to compare between model and data, we think it most relevant to show the data from the observations-based WOA (the best estimate of reality), rather than CESM2.

We include below the CESM2 version of Fig. 2. with the same arrangement of subplots and color scales. We do see some similarities in the cluster segregations, which confirms that these metrics are a relevant indicator of the various regimes.



Nevertheless, we argue that the clustering analysis allows us to automate the selection of each group, and thus allows us to examine water properties and biases independently of the bias in their location. Fig. 5 is a result delivered from the clustering analysis. The T-S diagrams, which show the water properties at every grid point in the WOA, more clearly display regime later properties of the layer averaged vertical profiles. Some analyses, such as recognizing the water mass with characteristic ranges of its T-S properties in 3.3, cannot be conducted without the detailed T-S diagrams of regimes. Once again, the purpose of this manuscript is to demonstrate the use of the clustering analysis (for a limited set of data), which we intend to use over a wider range of models.

Response to reviewer #2

Major comments:

My primary concern with the manuscript is the choice to apply the clustering technique to a vertical profile, rather than to individual water masses. I found the discussion of this confusing at various places throughout the text. Typically, we use tracers to define water masses with distinct properties, formed by a specific physical process in a specific place. These water masses are then used to trace circulation pathways as tracer properties become modified and ultimately destroyed during their transit to new formation regions (Groeskamp et al. 2019). Due to this lateral transport, throughout most of the ocean a vertical profile at a given latitude/longitude position samples a number of different water masses formed far from the profile itself. Consider an example: AAIW is formed on the norther side of the ACC and spreads into both the Atlantic and Pacific basins at intermediate depths. A clustering algorithm, similar to the one described here and performed on vertical profiles in these two regions, would produce two different hydrographic "groups" because the Atlantic profile would detect NADW whereas the Pacific profile would not. This would be true even if AAIW properties were the same. Thus the clustering method, defined this way at least, would not be helpful in identifying biases or differences in AAIW formation.

The authors need to more carefully state how this method can be used to address biases in water mass formation and circulation. Clustering on a vertical profile convolves these two processes, whereas clustering distinct groups of water properties would, I believe, provide a clearer assessment of the former. In particular, the authors refer to "core water masses" in Figure 8, but this is not an accurate description. In panels (b) and (c) of Figure 8, there are at least two water masses contributing to this clustering group, forms of CDW and WW. The formation of these two water masses happen through different processes and in quite different locations, and their changes need to be considered independently when trying to understand why five or six groups are selected by the clustering algorithm or why there are biases in the model data.

Ultimately, the choice to cluster on the profiles still provides information about model data Biases. I am not suggesting the authors need to revise their analysis. However, the motivation for this choice and the discussion of the manuscript's results could be improved. I provide a few more specific comments and suggestions below.

Thank you for your comments and suggestions. In the revised manuscript, we include more detailed descriptions about WOA climatology and CESM2 setup related to the ACSS physical processes. We also discuss further the physical processes misrepresented in the CESM2 based on the hydrographic regimes of the WOA from the clustering analysis.

Regarding your major concerns, we explained why we choose vertical profiles, rather than water masses for clustering analysis in this study. We summarize these reasons here:

1. Many important ocean processes in the ACSS depend on stratification, e.g., decoupling of deep onshore flows of CDW from wind forcing of the upper ocean. While the specifics of a model's representation of water masses are important, the stratification environment in which each water mass sits is also relevant to whether the model is correctly capturing the ocean dynamics of the ACSS.

2. The water mass properties do not need to be prescribed to find hydrographic regimes for clustering analysis. Indeed, the model's water mass properties can be very different (Figure 8c) or missing (Figure 8a) from the observations. We are unable to use cluster analyses on water masses in a specific model without first knowing how their properties are characterized in that model, because modeled water mass properties are often biased and could be very different from the observed values. It may be possible to define the "relative" water mass properties specifically for each model, such as in Sallée et al. (2013), but the water properties are much more complex on the Antarctic continental shelves than in the deep Southern Ocean because of the mixing induced by shallower water depth, complex bathymetry, and coastal boundaries, and the large spatial gradients of salt, freshwater and heat sources (e.g., intense cooling and sea ice formation in localized coastal polynyas, and outflows of cold and fresh water from ice shelves). Any misrepresentation of these complex processes in a model results in errors in water masses in the ACSS; Therefore, we don't feel that the methodology proposed by Sallée (2013) can be applied in this study.

the more detailed descriptions can be found in the revised manuscript (Line 91-105): "***The results of clustering analyses are dependent on the metrics chosen for the analysis. For example, metrics could be chosen as the layer thicknesses of water masses defined by T, S and neutral density. Schmidtko et al. (2014) partitioned water masses in the Southern Ocean into Winter Water (WW), CDW, and Antarctic Shelf Bottom Water (ASBW) using only temperature. However, their metrics of subsurface water temperature maxima and minima are ineffective on the continental shelf, where temperature profiles are often complex and show strong lateral variability in water properties (Figure1d). Sallée et al. (2013) proposed a method to use potential vorticity evaluated from density profiles and the local salinity minimum at 30^{o}S to distinguish vertical water masses in the Southern Ocean.***

On the ACSS, however, hydrographic structure is complicated not only by variability of primary water masses but also by transport, mixing, and strong and highly localized interactions between the atmosphere, ocean, sea ice and ice shelves. Each of these processes is sensitive to vertical and horizontal density gradients and gradients in bathymetry. Metrics that capture the importance of stratification concurrently with dominant water mass characteristics provide the best test of whether a model is representing the principal dynamical processes governing hydrographic variability in the ACSS. Here, we develop new metrics targeted at ACSS hydrography and assess the utility of a clustering-based approach for model-data comparison."

Following is the detailed response for each of your comment:

- Line 46: "These errors may influence the future rate of regional warming," Be clear you mean warming in the model here.

Revised as "***These modern-state biases suggest the potential for large uncertainties in the projected ocean state, including the vertical and horizontal distribution of ocean heat, with significant consequences for the accuracy of projections of the effect of the ACSS on other climate components (e.g., Sallée et al., 2013; Agosta, Fettweis and Datta, 2015).***"

- Line 56: "Strong gradients are evident . . . "Perhaps give a few examples?

We added citations to two papers that provide examples of high spatial gradients of hydrographic properties; Orsi and Wiederwohl (2009) and Thompson et al. (2018).

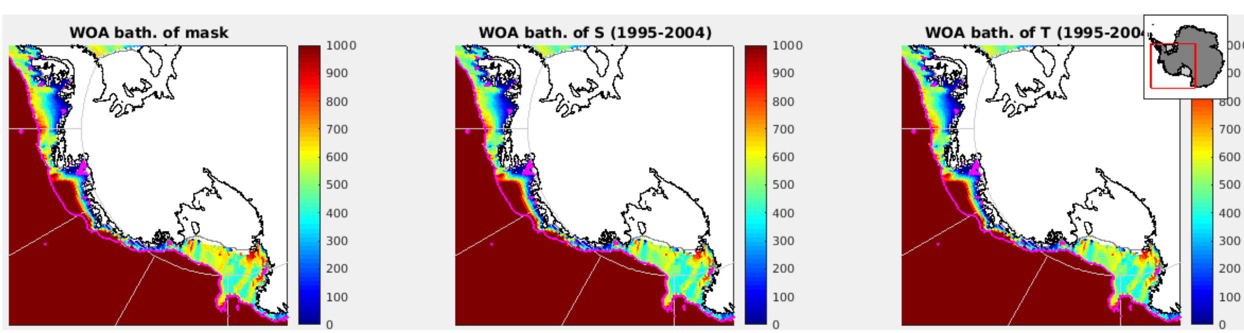
- The paper should include a more detailed description of the WOA and CESM2 hydrography used in the study as well as the surface forcing for the latter. For instance, it would be helpful for the reader to know how meltwater fluxes are parameterized or applied in CESM2. I assume there is no representation of ice shelf cavities. Similarly, a brief description of the types of data that is included in the WOA would be helpful: Are Argo floats included in the 2000-2500 m depth ranges? Are seal data from the MEOP data base included (e.g. Pellichero et al. 2017)?

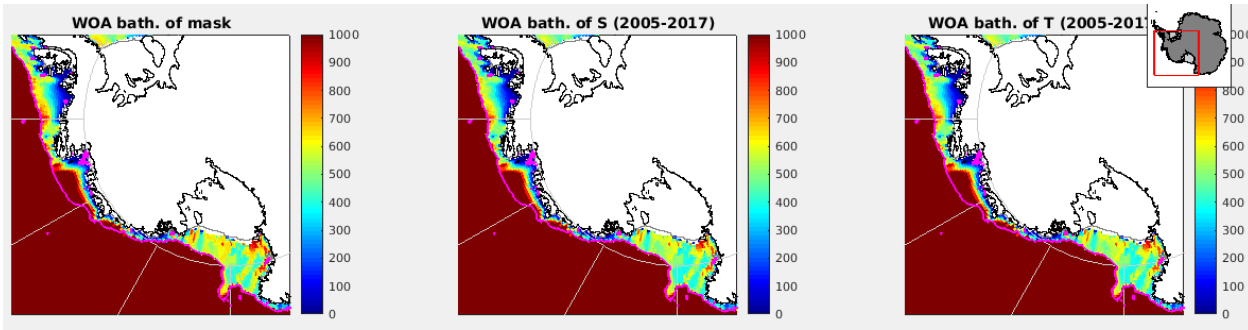
Revised in paragraph 2.1 by adding some information about the CESM setup and WOA data (Line 114-115): "***The data sources, quality controls, and processing procedures of the WOA are detailed in Locarnini et al. (2019) for temperature and Zweng et al. (2019) for salinity.***" and (Line 123-126): "***CESM2 uses the CICE5 (Hunke et al., 2015) sea ice model; however, dynamic and thermodynamic interactions with land ice are not represented (Danabasoglu et al., 2019). The CMIP6 forcing data is described in Eyring et al. (2016) and can be download from input4MIPs CoG (https://esgf-node.llnl.gov/search/input4MIPs).***". The documents, including more detailed information, are cited for the CESM and WOA. The reference describes the CMIP6 historical forcing data and its links for downloading, which are added.

In fact, using data that only goes up to 2004 is really not ideal considering how much effort there has been to improve observational coverage in West Antarctica over the last decade and a half.

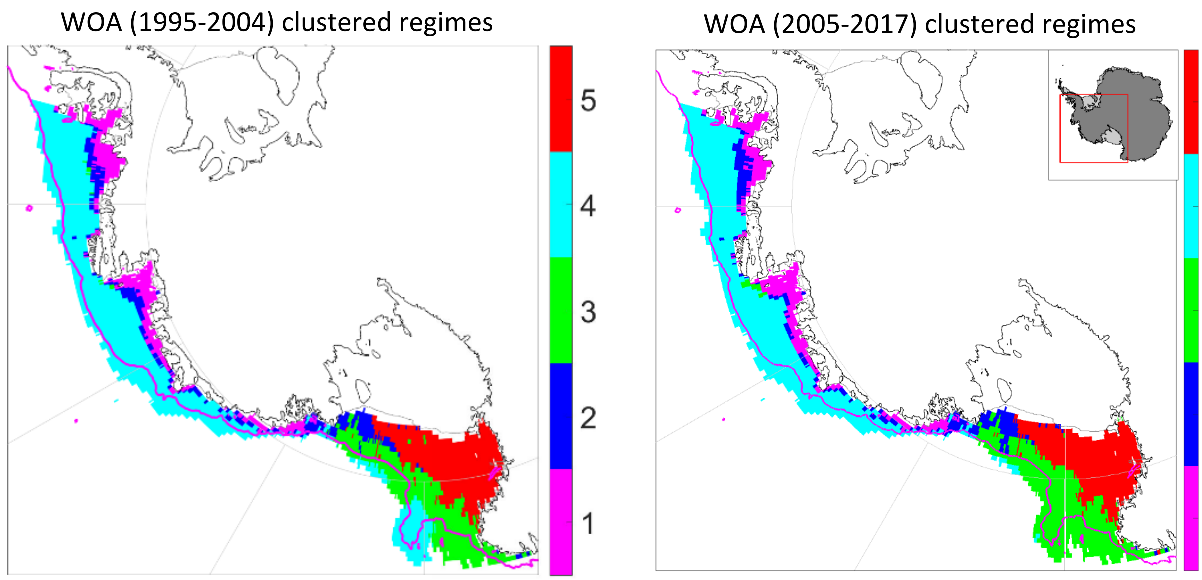
We agree with the reviewer's point that observational coverage has improved in the last decade and a half. To clarify, we repeated the analysis over the latest decade of WOA2018, and found no significant qualitative change to our conclusions (see below). We also add these comparison results in the manuscript (Line 394-398): "***Finally, we note that the clustering results for the ACSS based on the WOA decadal data (1995-2004) are consistent with the results based on the most modern WOA decadal data (2005-2017). However, clustering, applied to a variety of metrics, provides the potential to identify more subtle temporal changes in hydrographic fields such as changes in regime extent in the absence of significant changes in water mass characteristics in the ACSS.*"**. In particular, it appears that the loss of information due to the spatial interpolation of WOD, and the use of a flawed bathymetry (which we highlight in Fig. 10 of the manuscript) remains an issue for WOA 2018.

Regarding bathymetry/bottom data, we compare below the WOA bathymetry by using the WOA provided mask file, and the deepest T/S data from two different decadal mean products (1995-2004, and 2005-2017). They are exactly the same. We conclude that the WOA bathymetry issues stem from its mask file, rather than the availability of data.





Secondly, we recalculated the clusters using the most modern decadal WOA product (2005-2017). The segregation of the ABRS is qualitatively similar to that presented in the paper (for 1995-2004). The additional data in the modern period do not significantly impact the conclusions of this paper.



- Line 116: "We wish to identify regions that exhibit a similar vertical structure". Here the authors should provide additional justification for this approach considering formation sites are spatially distinct. In fact, it would be equally interesting to perform the clustering analysis on individual density surfaces. Combining different clustering analyses may provide complimentary information.

We agree with the reviewer that other choices of metrics are possible and, for certain purposes, could be preferable. We now try to better justify the current approach of focusing on S(Tmin) and S(Tmax) (see lines 99-105): "***On the ACSS, however, hydrographic structure is complicated not only by variability of primary water masses but also by transport, mixing, and strong and highly localized interactions between the atmosphere, ocean, sea ice and ice shelves. Each of these processes is sensitive to vertical and horizontal density gradients and gradients in bathymetry. Metrics that capture the importance of stratification concurrently with dominant water mass characteristics provide the best test of whether a model is representing the principal dynamical processes governing hydrographic variability in the ACSS. Here, we develop new metrics targeted at ACSS hydrography and***

assess the utility of a clustering-based approach for model-data comparison." based on the critical role of stratification on the dynamics of hydrographic variability in the ACSS.

We have also added a comment about potential different metric choices in the Discussion, to motivate other analyses (Line 372-374): "***It will also be interesting to track water masses and their pathways with metrics based on their characteristic properties. However, we note that comparisons of the locations of groups could become complex if the approach is applied to multiple models with substantial biases between their representations of specific water masses.***".

- Line 126: "rearranging the data nearest them." I did not understand how (or why) this rearrangement was carried out.

The clustering algorithm is iterative: when iteratively adjusting the centroids, we recalculate the data distance to the centroids, and thus which group a particular datapoint is included in. We clarify the language in the revised manuscript: "by adjusting the centroids, recalculating the distances, and rearranging data points among the groups. "

- Section 3.3: It would be useful to know how much seasonality exists at the target depths that have been selected for the clustering analysis (either variability over a year in the model output or data availability for the WOA).

We agree that seasonality of hydrography in the ACSS is an important issue, and we are working on analyses of the complete WOD data set to see whether true seasonality can be extracted from data that have a strong bias towards summer acquisitions. Our preliminary assessment is that seasonality is likely to be small below about 100-150 m depth; however, it could be quite large (e.g., a T range of 2-3 deg C) near the surface. However, except for a few well-sampled regions and at locations of moorings, seasonality is almost unrecorded unless high-resolution ocean models are used.

Here, we focus on the mean state in this paper, as cast profiles from the World Ocean Database shows that observations are overwhelmingly in southern summer months. Although we do not expect significant variations below the permanent pycnocline (200-300 m), we are skeptical that sufficient data is available to perform a detailed investigation.

- Line 255-256: "These three groups (1, 4 and 5) represent the three "source" ABRS hydrographic regimes." Based on my major comments above, I do not like the use of the term here "source" because these regions are not necessarily isolating water mass formation processes. The authors should be clearer about what this grouping represents.

We agree, and have changed the word "source" to "primary" to avoid confusion. We name the group 1, 4 and 5 as primary hydrographic regimes.

- Section 3.4: This manuscript would be improved if the authors went beyond simply stating the differences between the CESM2 and WOA clustering and provided some explanations for why the geographic distribution of the groups differ.

Revised. We added the discussion about the wind stress bias (Line293-296) "***Sea ice concentrations are biased low in CESM due to positive zonal wind stress biases in the Southern Ocean (Singh et al., 2020).***

This wind stress bias may, in turn, lead to an overestimate of the upwelling of warm and salty CDW onto the ACSS. The limited extent of the coastal fresh-water-enriched regime (group 1) in CESM2 may result from the absence of basal melt from ice shelves. ” and overflow parameterization (Line 304-308): “***It is possible that these differences result from the overflow parameterization in CESM2 (Briegleb, Danabasoglu, & Large, 2010). In this parameterization, locations of the on-shore source water at its formation regions and off-shore entrainment, which mixes with the source water to produce the final water mass, are defined, and overflow water is routed to fixed locations. While this parameterization allows transport of HSSW to the Southern Ocean, it is entirely artificial and does not represent on-shelf mixing processes.*** ” in the CESM2, which likely results in the mismatches in the location of clustering groups and water properties.

- Line 292: "The region identified as HSSW (group 5), in the southwestern Ross Sea, remains." Again, I would not identify a cluster group as a water mass. You might say, "The region identified as Group 5 in the SW Ross Sea, which is associated with HSSW formation, remains."

Thank you, we agree. Revised.

- Line 358: "Our comparison suggests that mean-state biases of CESM2 on the ACSS result from both local and remote processes." The authors do a good job of explaining the changes in T/S properties that give rise the distinct clustering groups, but the discussion of physical processes either missing or misrepresented in the models receives less attention. This statement in the conclusion is quite broad and not well justified. As mentioned above, further discussion of how the model represents glacial melting, sea-ice formation, surface fluxes, interior mixing, etc., and how this might impact tracer properties, would strengthen the manuscript.
Revised. We added a summary of the discussion about the possible melt effects in conclusion (Line 386-389): “***… including overestimated zonal winds in the Southern Ocean, unrepresented thermodynamic interactions with ice shelves, and the inadequate representation of overflows in the Ross Sea. A more specific investigation of coastal processes, Southern Ocean dynamics, and atmospheric forcing will help further identify the cause of these biases.***”.