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.

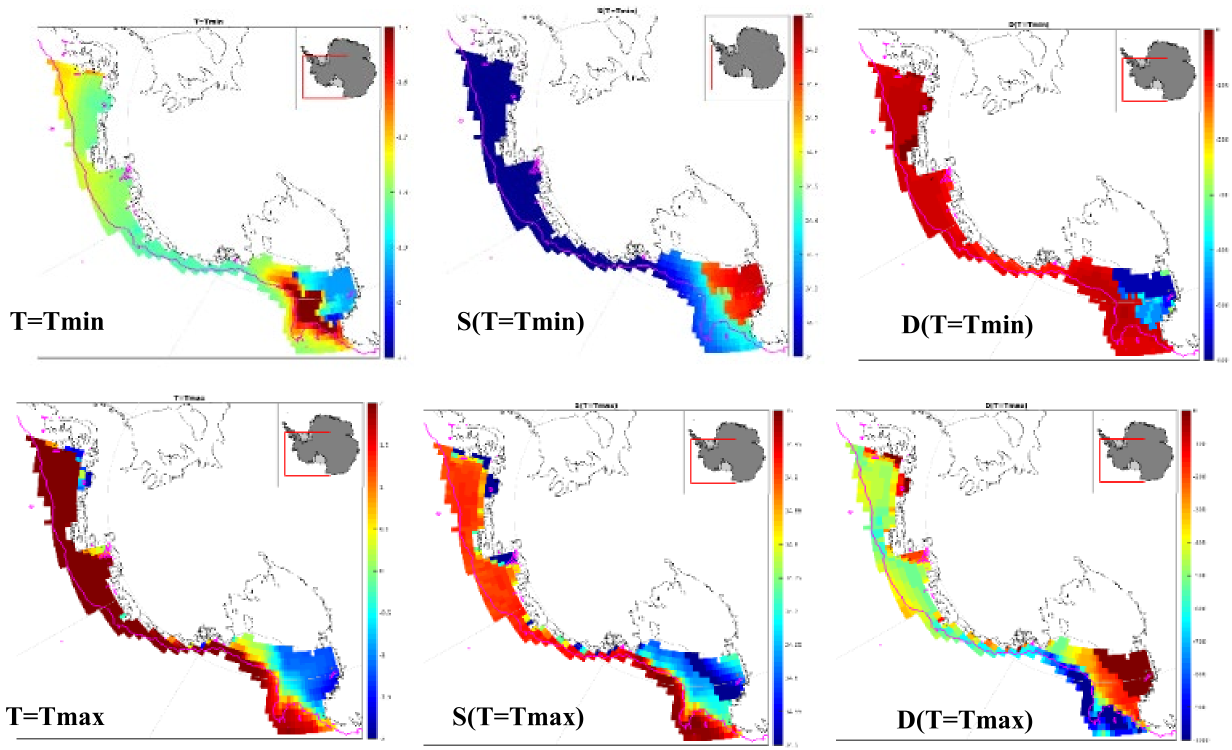
If the main focus is in discussing the degree of realism of CESM, important information should be added about this run. Technical details are missing, such as the vertical resolution of the ocean component, the type of atmosphere model, the representation of sea-ice in the model, and the parameterization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