A clustering-based approach to ocean model-data comparison around Antarctica

Qiang Sun[1], Christopher M. Little[1], Alice M. Barthel[2] and Laurie Padman[3]

1 Atmospheric and Environmental Research, Inc., Lexington, MA 02421, USA
2 Los Alamos National Laboratory, Los Alamos, NM 87545
3 Earth and Space Research, 3350 SW Cascade Ave., Corvallis, OR 97333, USA
Correspondence to: Qiang Sun, (qsun@aer.com)

10   **Abstract**

11   The Antarctic Continental Shelf Seas (ACSS) are a critical, rapidly-changing element of the Earth system.

12   Analyses of global-scale general circulation model (GCM) simulations, including those available through the

13   Coupled Model Intercomparison Project, Phase 6 (CMIP6), can help reveal the origins of observed changes and

14   predict the future evolution of the ACSS. However, an evaluation of ACSS hydrography in GCMs is vital: previous

15   CMIP ensembles exhibit substantial mean-state biases (reflecting, for example, misplaced water masses) with a wide

16   inter-model spread. Because the ACSS is also a sparely sampled region, grid-point based model assessments are of

17   limited value. Our goal is to demonstrate the utility of clustering tools for identifying hydrographic regimes that are

18   common to different source fields (model or data), while allowing for biases in other metrics (e.g., water mass core

19   properties) and shifts in region boundaries. We apply K-means clustering to hydrographic metrics based on the

20   stratification from one GCM (Community Earth System Model version 2; CESM2) and one observation-based

21   product (World Ocean Atlas 2018; WOA), focusing on the Amundsen, Bellingshausen, and Ross Seas. When

22   applied to WOA temperature and salinity profiles, clustering identifies "primary" and "mixed" regimes that have

23   physically interpretable bases. For example, meltwater-freshened coastal currents in the Amundsen Sea, and a region

24   of high salinity shelf water formation in the southwestern Ross Sea emerge naturally from the algorithm. Both

25   regions also exhibit clearly differentiated inner- and outer-shelf regimes. The same analysis applied to CESM2

26   demonstrates that, although mean-state model biases in water mass T-S characteristics can be substantial, using a

27   clustering approach highlights that the relative differences between regimes, and the locations where each regime

28   dominates, are well represented in the model. CESM2 is generally fresher and warmer than WOA and has a limited

29   fresh-water-enriched coastal regimes. Given the sparsity of observations on the ACSS, this technique is a promising

30   tool for the evaluation of a larger model ensemble (e.g., CMIP6) on a circum-Antarctic basis.

31

32

1. Introduction

The Antarctic Continental Shelf Seas (ACSS, defined here as the ocean regions adjacent to Antarctica with water depth shallower than 2,500 m) are critical components of the climate system, playing an essential role in ice sheet mass balance, sea ice formation, and ocean circulation (Rignot et al., 2008; Hobbs et al., 2016; Bindoff, Rosenberg and Warner, 2000). ACSS ocean state, and the climate system components that are coupled to it, are changing rapidly. In the Amundsen-Bellingshausen Seas sectors, the atmosphere (Bromwich, et al., 2013) and subsurface ocean (Schmidtko et al., 2014) are warming, the sea ice-free period is rapidly increasing (Stammerjohn et al., 2012), ice shelves are thinning (Rignot et al., 2013; Paolo, Fricker and Padman, 2015; Adusumilli et al., 2020), and the grounded portion of the ice sheet is losing mass at an accelerating rate (Shepherd et al., 2018; Sutterley et al., 2014; Gardner et al., 2018). The Ross Sea has also experienced long-term changes in fresh water content (Jacobs and Giulivi, 2010; Castagno et al., 2019) and an increase in sea ice production and extent (Parkinson, 2019; M. Holland et al., 2017).

Assessing the causes of observed changes in climate and the coastal cryosphere, and their future evolution, requires coupled, global, atmosphere-ocean general circulation models (GCMs). However, recent GCMs exhibit large biases relative to modern observations, and a wide inter-model spread (Agosta, Fettweis and Datta, 2015; Sallée et al., 2013; Rickard and Behrens, 2016; Hosking et al., 2016; Little and Urban, 2016; Barthel et al., 2019). 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). For example, DeConto and Pollard (2016) projected extreme rates of 21st-century ice sheet mass loss from the Pacific sector for a high-emission scenario. However, their projections were forced using a single GCM (CCSM4) that required a +3°C correction to subsurface water temperatures in the Amundsen Sea to match observed hydrography and modern ice shelf melt rates. This significant bias correction indicates an underlying mean-state error (e.g., a misplaced water mass) that casts substantial doubt on the projected future ocean state in that specific model.

The first step toward identifying the physical processes underlying GCM representation errors is assessing the magnitude and spatial distribution of biases. However, such a strategy must account for strong horizontal and vertical gradients in ACSS hydrographic properties (see, e.g., Orsi and Wiederwohl, 2009; Thompson et al., 2018), and the sparseness and variable quality of available observations (e.g., Schmidtko et al., 2014). Strong gradients are

61 evident in the Amundsen, Bellingshausen, and Ross Seas (ABRS) sector of the ACSS. There, the time-mean ocean

62 state of the objectively analyzed temperature and salinity field, as represented in the 0.25-degree World Ocean Atlas

63 version 2018 (WOA hereafter), suggests that the ABRS can be roughly separated into two geographical regions, the

64 Amundsen-Bellingshausen Seas and the Ross Sea (Figure 1a). In the Ross Sea, dense water formation occurs

65 locally, through brine rejection from winter sea ice formation in coastal polynyas, resulting in regionally averaged

66 water well below 0°C at water depths of 100 m to 700 m (Figure 1c). At the same depth range in the Amundsen-

67 Bellingshausen Seas, water temperatures can reach +1.2°C due to the presence of Circumpolar Deep Water (CDW).

68 In addition to these stark contrasts in regional mean temperature (and salinity), there is also significant

69 spatial variability within each region of the ABRS, and across the continental shelf break. For example, Figure 1

70 indicates a high standard deviation in ocean temperature on the continental shelf with water depth shallower than

71 700 m (0.5°C in the Amundsen-Bellingshausen Seas and 1.4°C in the Ross Sea). Much of this variability is

72 attributable to the lateral temperature gradient from the subsurface layer of CDW over the continental slope to the

73 modified (cooled) water masses inshore. In the alongshore direction, vertical profiles of water properties in the

74 Amundsen-Bellingshausen Seas are similar, with cold and fresh water overlying relatively warm and salty water. In

75 the Ross Sea, water properties are different on its southwestern and eastern sides, mainly distinguished by their

76 salinity (Figure 1d).

77 The sparseness of measurements on the ACSS also aggravates errors associated with gridded observational

78 products. Coastal regions, in particular, are subject to substantial errors. Sun et al. (2019) showed that salinity biases

79 between WOA objective analysis and the World Ocean Database increase toward coastlines. The gridded objective

80 analysis field neglects the dynamical processes governing water mass modifications and circulations induced by

81 complex continental shelf bathymetry (Dunn and Ridgway, 2002; Schmidtko, Johnson and Lyman, 2013). In

82 sparsely sampled regions, grid-point based comparisons (e.g., Little and Urban, 2016) are thus of limited utility, and

83 may underestimate uncertainty in the reference (observational) product. We suggest that it is often more meaningful

84 to assess GCMs using a regionally averaged approach.

85 Previous model-data comparisons on the ACSS have employed strategies such as averaging over a-priori

86 defined regions (e.g., Barthel et al., 2019). Such methods are ill-equipped to assess model biases resulting from

87 misplaced water masses. An alternative method is objective clustering, which can be used to identify regions of

88 similar hydrographic metrics. For example, Hjelmervik and Hjelmervik (2013) demonstrated the application of a

89   clustering-based approach using Argo profiles to segregate the North Atlantic into groups with similar vertical T and

90   S profiles separated by fronts.

91   The results of clustering analyses are dependent on the metrics chosen for the analysis. For example,

92   metrics could be chosen as the layer thicknesses of water masses defined by T, S and neutral density. Schmidtko et

93   al. (2014) partitioned water masses in the Southern Ocean into Winter Water (WW), CDW, and Antarctic Shelf

94   Bottom Water (ASBW) using only temperature. However, their metrics of subsurface water temperature maxima

95   and minima are ineffective on the continental shelf, where temperature profiles are often complex and show strong

96   lateral variability in water properties (Figure 1d). Sallée et al. (2013) proposed a method to use potential vorticity

97   evaluated from density profiles and the local salinity minimum at 30ºS to distinguish vertical water masses in the

98   Southern Ocean.

99   On the ACSS, however, the hydrographic structure is complicated not only by the variability of primary

100   water masses but also by transport, mixing, and strong and highly localized interactions between the atmosphere,

101   ocean, sea ice and ice shelves. Each of these processes is sensitive to vertical and horizontal density gradients and

102   gradients in bathymetry. Metrics that capture the importance of stratification concurrently with dominant water mass

103   characteristics provide the best test of whether a model is representing the principal dynamical processes governing

104   hydrographic variability in the ACSS. Here, we develop new metrics targeted at ACSS hydrography and assess the

105   utility of a clustering-based approach for model-data comparison.

106   **2.   Methods**

107   In this paper, we identify hydrographic regimes and their T-S properties using metrics derived from three-

108   dimensional grids of measured and modeled temperature and salinity (section 2.1) using a K-means clustering

109   method (section 2.2). We then apply a clustering algorithm based on data density to exclude outliers (2.3) from the

110   resulting "groups".

111

112   **2.1 Data and processing**

113   We use decadal-mean, objectively analyzed T and S fields from WOA for 1995-2004, with 0.25-degree

114   resolution in both latitude and longitude. The data sources, quality controls, and processing procedures of the WOA

115   are detailed in Locarnini et al. (2019) for temperature and Zweng et al. (2019) for salinity. This study focuses on the

5

116 domain from the west of Cape Adare (163°E) on the western side of the Ross Sea to the southern end of Alexander

117 Island (76°W), at depths between 0 and 2,500 m. The landward limit of the study domain is the Antarctic coast and

118 the ice shelf edges as identified in Figure 1a.

119 We compare the Community Earth System Model version 2 (CESM2; Danabasoglu et al., 2019) to WOA

120 for the same period and domain. The time-mean model salinity and temperature fields over the 1995-2004 period are

121 calculated from the monthly output of the Coupled Model Intercomparison Project Phase 6 (CMIP6) historical

122 simulation (experiment tag r1i1p1f) (Eyring et al., 2016) at the native ocean model resolution (roughly 1 degree in

123 longitude and 0.5 degree in latitude). CESM2 uses the CICE5 (Hunke et al., 2015) sea ice model; however, dynamic

124 and thermodynamic interactions with land ice are not represented (Danabasoglu et al., 2019). The CMIP6 forcing

125 data is described in Eyring et al. (2016) and can be download from input4MIPs CoG (https://esgf-

126 node.llnl.gov/search/input4MIPs).

127 We used the Gibbs SeaWater (GSW) Oceanographic Toolbox of TEOS–10 (McDougall and Barker, 2011)

128 to calculate seawater properties. The absolute salinity (S_A) has unit of g/kg, and conservative temperature (Θ) is in

129 °C. All seawater temperatures are referenced to the sea surface.

2.2 Prototype-based clustering technique (K-means)

131 The K-means clustering analysis used in this study is an unsupervised learning technique that classifies data

132 into meaningful groups based on their similarity. In this study, the similarity is defined by two metrics of the water

133 column: 1) salinity at the temperature minimum; and 2) salinity at the temperature maximum. The rationale for

134 these choices is discussed in section 3.1.

135 The K-means algorithm is initialized by randomly selecting data in N dimensions (here, N=2, for the two

136 specified metrics) for a specified number of groups (K). For each group (k_i), the Sum of Squared Distance (SSD) of

137 each data point (ξ) to the group's centroid (c_i) is calculated:

138

$$SSD = \sum_{i=1}^{K} \sum_{\xi \in k_i} dist(\boldsymbol{c_i}, \boldsymbol{\xi})^2 \quad with \quad \boldsymbol{c_i} = \frac{1}{m_i} \sum_{\xi \in k_i} \boldsymbol{\xi} \qquad \text{Eqn. 1}$$

139

140 where *dist* is the standard distance between data and centroid in N-dimensional Euclidean data space and m_i is the

141 total number of data points in group k_i. The algorithm iterates to minimize SSD by adjusting the centroids,

6

142 recalculating the distances, and redistributing data points among the groups. The K-means algorithm will have

143 multiple solutions because it is initialized with randomly selected data. We apply the K-means 1,000 times and

144 choose the solution with the lowest *SSD* for analysis.

145   The K-means algorithm requires specification of the number of groups (K). We use Silhouette scores $s_i(n)$

146 (Eqn. 2) to assess the appropriate values of K.

$$s_i(n) = \frac{b(\xi) - a(\xi)}{max\{a(\xi), b(\xi)\}} \qquad\qquad \text{Eqn. 2}$$

147

148 In Eqn. 2, n represents the number of data points in group k_i, $a(\xi)$ is the mean *dist* from a data point ξ to all other

149 data points within the group k_i, and $b(\xi)$ is mean *dist* from ξ to all other data points <u>outside</u> the group k_i. Silhouette

150 scores are evaluated for each data point ξ in the group k_i and range between -1 and 1. If ξ lays perfectly at the

151 centroid of group k_i, then $s_i(n)=1$.

152   A rigid interpretation of the Silhouette algorithm would choose the value of K that corresponds to the

153 highest mean value of $s_i(n)$. However, the optimal K value can vary with different clustering evaluation methods

154 (e.g., Elbow method: Thorndike, 1953) and different domains. The selection of K is thus based not only on the

155 results of Silhouette assessment but also on the ability to interpret the groups as representative of different

156 underlying physical processes (see section 3).

157 **2.3 Density-based clustering technique**

158   In subsequent sections, we use a *T-S* diagram to compare the properties of groups given by the K-means

159 algorithm. We applied a data density-based clustering technique (DBSCAN) (Ester et al., 1996) to define the "core"

160 of a group and to exclude outliers on the *T-S* diagram. Note that DBSCAN is only used to highlight the core of a

161 given group and facilitate comparisons of water properties between WOA and CESM2.

162   The *T-S* core of each hydrographic regime identified by the K-means clustering is determined by the

163 DBSCAN algorithm using two parameters: a radius (ε), and a minimum number of neighboring points (MinPts).

164 The DBSCAN algorithm builds up pools of data by initially choosing a random data point. If the initially chosen

165 data point has less than MinPts within ε, then it is defined as an outlier. If this data point has more than MinPts

166 within ε, then a pool of data is initialized consisting of the initial point and the points within ε (neighbors). The pool

167 grows by continually clustering neighboring points until these points have fewer than MinPts within ε. The

168 algorithm continues until all data points are either clustered into pools of data or labeled as outliers. In the current

169 study, we choose MinPts = 10 and $\varepsilon = \sqrt{S^2 + T^2}$. The value of ε is then selected (Table 1) so that the largest pool of

170 data contains at least 97% of non-outlier points (Table 2). This pool of data constitutes the core of each group.

3 Results

3.1 Defining water column metrics

173 Our goal in this analysis is to utilize key features of local water columns to identify regions with similar

174 hydrographic properties. Such metrics must be able to capture stratification, and the changes in T and S in both

175 along- and cross-shelf directions. For the ACSS, the metrics must include salinity because it is the dominant factor

176 influencing water column stability and reflects critical processes such as brine input during sea ice formation, and

177 freshwater inputs from melting sea ice and ice shelves. By itself, however, salinity poorly represents the vertical

178 composition of water masses since it increases monotonically with water depth over most of the ACSS (Figure 1);

179 salinity alone is insufficient to identify regimes with sub-surface heat reservoirs that are characteristic of regions

180 with high ice shelf basal melt rates (Rignot et al., 2013; Dinniman et al., 2016; Holland et al., 2020; Adusumilli et al.,

181 2020). The metrics we use in this study – salinity at the vertical temperature minimum and salinity at the vertical

182 temperature maximum – are similar to those used by Timmermans et al. (2014) to segregate surface water from

183 Alaska coastal water in the Central Canada Basin of the Arctic Ocean.

184 Along-shelf variations of water properties are evident in salinity at the vertical temperature minimum

185 (Figure 2b). In the Amundsen-Bellingshausen Seas, the depth of minimum temperature (Figure 2c) is commonly

186 above 200 m, where salinity is often less than 34.2 g/kg. In contrast, in the southwestern Ross Sea, the minimum

187 temperature is usually located below 350 m and coincides with much higher salinity (>34.8 g/kg). The northwestern

188 Ross Sea contains a regime with a local temperature minimum at shallower depths approaching the shelf break, but

189 its salinity (between 34.2 to 34.6 g/kg) is higher than near-surface water in the Amundsen-Bellingshausen Seas.

190 The salinity at the vertical temperature maximum shows pronounced variations in the cross-shelf direction

191 (Figure 2d-f). The maximum water temperature (Figure 2d) is commonly found at depths above 200 m close to the

192 coast and ice shelves (Figure 2f), and deeper toward the shelf break and over the continental slope where the water

193 depth increases. The salinity at the vertical maximxal temperature (Figure 2e) shows similar variations in the cross-

194 shelf direction, with lower salinity (<34.7 g/kg) near the coast and ice shelves and higher salinity (>34.8 g/kg) on the

195 continental shelves and near the shelf break.

3.2 Evaluating the optimum number of groups

197 We used the mean value of Silhouette score $s_i(n)$ in Eqn. 2 to evaluate an appropriate number of groups (K)

198 for WOA and CESM2, testing $2 \leq K \leq 13$ (Figure 3). For WOA, the highest value of s_i occurs when $K=3$; for CESM2,

199 $K=6$ has the highest Silhouette score (Figure 3a-b). The spatial distribution of groups 3, 5 and 6 in the ABRS are

200 shown in Figure 3c-h.

201 When WOA data are clustered into three groups (Figure 3c), the K-means algorithm segregates the water

202 close to the Antarctic coast from the water on the shelf and continental slope. The coastal domains are further

203 distinguished into Amundsen-Bellingshausen coastal waters and Ross coastal waters. By increasing the number of

204 groups to five (Figure 3e), a narrow domain between coastal and shelf waters emerges. In the Ross Sea, waters on

205 the shelf and across the shelf break are segregated into two groups. For $K=6$ (Figure 3g), the southeastern coastal

206 domain of the Ross Sea (orange) is further separated from the narrow domain between coastal and shelf waters in

207 the Amundsen-Bellingshausen Seas, while the locations of the other groups are generally unchanged.

208 Examining the groups with respect to the two metrics used in the K-means clustering (Figure 4) shows that,

209 when $K=3$, the groups are separated by the perpendicular lines from the incenter of the triangular T-S distributions

210 (Figure 4a). As the total number of groups increases, data points are progressively divided into smaller subsets, with

211 an asymmetry that is influenced by their original distribution in our two-metric parameter space, as well as gaps and

212 discontinuities (Figure 4c and e).

213 In CESM2, the clusters in the ABRS differ from those for WOA, particularly for $K=3$ and $K=6$. For $K=3$

214 (Figure 3d), the entire Amundsen-Bellingshausen Seas region is segregated from the Ross Sea, while the

215 southwestern Ross Sea is still recognized as an independent group. For $K=6$, the Amundsen Sea is segregated from

216 the Bellingshausen Sea. With $K=5$ (Figure 3f), CESM2 clustering is qualitatively similar to WOA, with a coastal

217 group emerges in the Amundsen-Bellingshausen Seas; however, its areal extent is much smaller than in WOA. In

218 the Ross Sea, the water on the continental shelf is separated from the water on the continental slope, similar to WOA.

219 CESM2 shows a similar range to WOA in metric space (Figure 4), although with much larger gaps. In particular,

220 CESM2 has substantially fewer data points with intermediate and low salinity (Figure 4b). Increasing K for

9

221   clustering analysis of CESM2 output subdivides high salinity regimes at T_{max} based on the distribution of salinity at

222   T_{min} (Figure 4d and f).

223   Based on the Silhouette scores, the optimum clustering for CESM2 is 6 groups; however, the WOA data

224   have a maximum Silhouette score for $K=3$. Segregating the WOA into 5 or 6 groups is reasonable, as the clustering

225   algorithm continually distinguishes finer differences in the coastal regimes (Figure3e and g). But the segregation of

226   CESM2 into 6 groups (Figure3h) is physically unfair since the water properties below the surface layers are nearly

227   indistinguishable between the Amundsen and Bellingshausen Seas (Figure 1d). Figure 4 also indicates that the

228   segregation of Amundsen-Bellingshausen Seas regions in CESM2 is a result of discontinuities between groups 1 and

229   5 (Figure 4f). We thus choose to use 5 groups for the rest of the study. Our findings from analyzing the temperature

230   and salinity properties in the following sections further support this decision.

231   **3.3 Physical interpretation of WOA groups**

232   Vertical profiles of temperature and salinity are shown for each WOA group in Figure 5. The mean vertical

233   structure of each group is clearly different; furthermore, the standard deviations at each depth within groups are

234   much smaller than those of regional mean profiles (Table 3). With these vertical structures as context, we examine

235   T-S properties at all depths from each WOA group in Figure 6. The DBSCAN algorithm is used to identify the "core"

236   of non-outlier data in each group, shown with dark shading in Figure 6.

237   Group 1, which occupies the inshore regions of the Amundsen-Bellingshausen Seas (Figure 3e), is

238   characterized by weak vertical gradients in both T and S over the ~400 m water column (Figure 5a). The water in

239   this group has relatively low salinity (33.8 to 34.5 g/kg), temperature close to the freezing point (generally lower

240   than -1°C) and low density (26.9 and 27.5 kg/m^3) (Figure 6a), which suggests that the water in this regime is

241   strongly influenced by coastal fresh water input (Moffat et al., 2008; Jacobs and Giulivi, 2010; Jourdain et al., 2017).

242   Group 2, which is spatially located between the coastal waters (groups 1 and 5) and outer continental shelf

243   waters (groups 3 and 4), represents a narrow domain of mixing (Figure 3e). This regime is characterized by

244   relatively high standard deviations in salinity and temperature at depths between 100 m and 700 m, indicating that

245   the location and shape of the thermocline and halocline above the typical depth of the shelf break vary within this

246   group (Figure 5b). Below 700 m, the range of salinity and temperature are relatively small, due to reduced the

247   limited amount of data at these depths over the relatively narrow continental slope. In the upper ocean, group 2 has a

248   salinity from 33.8 to 34.7 g/kg, temperature -2 to -0.5°C and density 27.1 to 27.8 kg/m^3 (Figure 5b and Figure 6b),

249   lying between the properties of surface waters in groups 1 and 5. In the subsurface, group 2 has a temperature above

250   -0.5°C and salinity above 34.5 g/kg, which represents modified CDW on the shelf (Carmack, 1970; Orsi and

251   Wiederwohl, 2009; Emery, 2011).

252   Group 3, which is found on the outer continental shelf and the continental slope of the Ross Sea (Figure 3e),

253   shows high standard deviations in temperature above 700 m (Figure 5c), similar to group 2. However, the water in

254   this regime is generally denser than group 2. The surface water in group 3 is fresher than that of group 5 (Figure5c,

255   Figure 6c and f), which may result from sea ice melt and/or lateral mixing with fresher shelf water originating in the

256   Amundsen-Bellingshausen Seas (Assmann, Hellmer and Jacobs, 2005; Porter et al., 2019). The subsurface water

257   (between 100 and 600 m) of group 3 (Figure 5c and Figure 6c) does not have a clear vertical water mass transition,

258   and denser water exhibits a wide temperature range (-1.5 to +1.5°C) with relatively high salinity (34.6 to 35 g/kg),

259   suggestive of mixing between High Salinity Shelf Water (HSSW) and CDW.

260   Group 4, on the continental shelf of the Amundsen-Bellingshausen Seas and along most of the continental

261   slope of the ABRS (Figure 3e), exhibits properties consistent with off-shelf Southern Ocean water as noted by

262   Schmidtko et al. (2014). It has a well-defined vertical temperature structure with limited spatial variability (Figure

263   5d). In this region, Winter Water (WW) with salinity 33.8-34.5 g/kg, temperature -2 to -0.5°C and density 27 to 27.5

264   g/m^3, overlays CDW (salinity 34.6 to 36.8 g/kg, temperature 0 to +2°C and density 27.8 to 27.9 g/m^3), with a mean

265   profile showing a clear transition between them (Figure 6d).

266   Group 5, in the southwestern Ross Sea with some extensions to the southeast (Figure 3e), has higher

267   salinity than other groups (Figure 6). The almost uniform vertical temperature profile (Figure 5e) is identified as

268   HSSW. It is characterized by salinity 34.3 to 35.1 g/kg, temperature close to the freezing point, and density of 27.5

269   to 28.1 kg/m^3 (Figure 6e), resulting from brine rejection in the polynyas along the coast and Ross Ice Shelf front

270   (Foster and Carmack, 1976). The surface portion of the waters in group 5 with salinity lower than 34.62 g/kg is often

271   defined as Low Salinity Shelf Water (LSSW) in the Ross Sea shelf, but we generally refer to group 5 as HSSW

272   because its volume is much higher than the LSSW (Orsi and Wiederwohl, 2009).

273   Overall, groups 1 and 5 (Figures 6a and 6e) show relatively homogeneous salinity and temperature, while

274   group 4 has a pronounced thermocline and halocline at shallow depth. These three groups (1, 4 and 5) represent the

275   three "primary" ABRS hydrographic regimes. In contrast to these primary regimes, groups 2 and 3 have more

276 complex vertical structures, more spatial variability in thermocline at depths above about 600 m (roughly the shelf

277 break) and can be considered as "mixed" regimes.

3.4 Assessing groups in CESM2

279 To identify hydrographic regimes in CESM2, we conduct the same analyses as described for WOA in the

280 previous section, focusing on results for K=5 (Figure 3f). The T-S properties of each group in CESM2 are shown in

281 Figure 7. CESM2 results are similar to WOA's in that three primary regimes are present (group 1, coastal fresh-

282 water-enriched; group 4, off-shelf; and group 5, HSSW), but they show differences in their spatial extent (Figure 3e

283 vs. f), volume (Table 4), and T-S properties (Figure 8).

284 As in WOA, HSSW (group 5) of CESM2 is localized in the southwestern Ross Sea, but its eastward

285 extension into the southeastern Ross Sea is missing in CESM2 (Figure 3e and f), resulting in a reduced HSSW

286 volume (Table 4). The coastal fresh water-enriched regime (group 1) is mostly absent in CESM2 and is replaced by

287 the off-shelf regime in the Amundsen Sea.

288 Mismatches between CESM2 and WOA are also evident in the T-S properties of these primary regimes. In

289 general, HSSW in CESM2 has a fresh and warm bias relative to WOA (Figure 8d). Combined with its reduced

290 volume relative to WOA, this bias in CESM2 HSSW properties suggests that weak modeled katabatic winds in the

291 southwestern Ross Sea may limit sea ice production and export. Group 4 (the off-shelf regime) exhibits a fresh bias

292 in WW in the upper water column, but the densest off-shelf water in group 4, i.e., CDW, is saltier and warmer

293 (Figure 8c). Sea ice concentrations are biased low in CESM due to positive zonal wind stress biases in the Southern

294 Ocean (Singh et al., 2020). This wind stress bias may, in turn, lead to an overestimate of the upwelling of warm and

295 salty CDW onto the ACSS. The limited extent of the coastal fresh-water-enriched regime (group 1) in CESM2 may

296 result from the absence of basal melt from ice shelves.

297 The mixed regimes shift geographic location in CESM2. The narrow mixing zone (group 2) between

298 coastal fresh-water-enriched and off-shelf regimes in the Amundsen-Bellingshausen Seas is not evident in CESM2

299 (Figure 3e and f); the CESM2 is likely too coarse to resolve these mixing fronts. In the Ross Sea, groups are

300 separated into on-shelf (group 2) and off-shelf (group 3) approximately along the 1,000 m isobath (Figure 3f).

301 CESM2 fails to show the path of export of Ross on-shelf water (group 2, Figure 3f) along the northwestern

302 continental slope (Orsi, Johnson and Bullister, 1999), as it is seen in the WOA (group 3, Figure 3e). The core of on-

303 shelf water (group 2) also has less overlap with HSSW (group 5) in the T-S diagram in CESM2 (Figure 7f)

304 compared to WOA (Figure 6f). It is possible that these differences result from the overflow parameterization in

305 CESM2 (Briegleb, Danabasoglu, & Large, 2010). In this parameterization, locations of the on-shore source water at

306 its formation regions and off-shore entrainment, which mixes with the source water to produce the final water mass,

307 are defined, and overflow water is routed to fixed locations. While this parameterization allows transport of HSSW

308 to the Southern Ocean, it is entirely artificial and does not represent on-shelf mixing processes.

309 **3.5 Assessing clustering over the ACSS**

310 As the K-means algorithm is based on purely statistical criteria (centroid and minimized *SSD* in Eqn. 1)

311 applied to specific metrics, it is valuable to assess whether clustering results are sensitive to different study domains.

312 As a test case, we apply the same algorithm to WOA over the entire circumpolar ACSS where total water depth is

313 less than 2,500 m. The metrics used as input for the K-means analysis, as well as the total number of groups (*K*=5),

314 are unchanged. The use of the uniformly-gridded WOA product, rather than observational data, avoids the

315 possibility that the comparison is biased by regional variations in data density.

316 The location of five clustered water groups over the entire ACSS is shown in Figure 9a. Within the ABRS

317 domain, the geographic locations of all groups are almost unchanged, indicating the clustering results in the ABRS

318 are insensitive to substantial enlargement of the domain. The region identified as group 5 in the southwestern Ross

319 Sea, which is associated with HSSW formation, remains. Outside the ABRS, the clustering approach identifies water

320 of similar properties to group 5 in the Weddell Sea near the Filchner-Ronne Ice Shelf, the George V Coast near the

321 Mertz Glacier tongue, and Bransfield Strait and south of Trinity Peninsula (regions marked on Figure 9b). The

322 southern Weddell Sea experiences similar conditions to the southwestern Ross Sea, with HSSW formation in winter

323 due to brine rejection from sea ice formation enhanced by katabatic winds and tides driving a narrow but persistent

324 along-ice-front polynya (Nicholls et al., 2009). Along the George V Coast, HSSW is also generated by similar

325 processes acting near the Mertz Glacier ice tongue (Bindoff, Rosenberg and Warner, 2000; Post et al., 2011).

326 The waters in the subsurface of Bransfield Strait and south of the Trinity Peninsula are also grouped with

327 the HSSW regions, although their surface water is warmer and fresher than that of other HSSW regions around

328 Antarctica. Cook et al. (2016) showed that the regional water properties around the tip of the Antarctica Peninsula,

329 based on the World Ocean Database, are very similar to HSSW. Gordon et al. (2000) also noted that the water

330 properties in the center of Bransfield Strait are similar to HSSW in the Weddell Sea; they inferred that these waters

331 are formed in western Weddell Sea coastal polynyas and flow into Bransfield Strait.

13

4 Discussion

We have shown that the ABRS can be clustered into different regions based on salinity at the vertical water temperature minimum and maximum. This technique can help identify regions, in model and observational datasets, in which water properties are controlled by similar physical processes. It contrasts with traditional grid point-based comparisons, which do not adequately account for misplaced water masses.

In this study, WOA has been employed to assess CESM2 results. However, the hydrographic regimes identified in WOA may be misleading if they result from interpolation/extrapolation artefacts associated with non-uniform sampling of data in time and space, or if the water column structures are not adequately represented in WOA. One source of uncertainty in WOA arises from differences between true and gridded bathymetry, complicating interpolation and extrapolation of sparely sampled data into deeper portions of the water column. In Figure 10, we compare the depths of the deepest available data in WOA and CESM2 with water depths in the International Bathymetric Chart of the Southern Ocean (IBCSO, Arndt, et al., 2013). WOA has a clear misrepresentation of the Amundsen-Bellingshausen Seas continental shelf bathymetry (Figure 10b). First, the 1000 m isobath is shifted substantially landward in the Amundsen Sea. Second, deep across-shelf troughs (e.g., in Figure 10a) are not represented in the inner shelf of WOA, which possibly affects the value of salinity at the temperature maximum because the CDW is missing in these regions of the Amundsen-Bellingshausen Seas.

It is, therefore, unclear whether groups 1 and 2 are separated from the shelf and continental slope waters of group 4 in WOA (Figure 3e) due to their upper-ocean fresh water enrichment relative to other groups, or if the groups are influenced by under-sampling of hydrography in deep troughs of the Amundsen-Bellingshausen Seas. We note that the bathymetry of CESM2 has similar issues as WOA in the Amundsen-Bellingshausen Seas (Figure 10c). Neither WOA nor CESM2 represents the water in deep troughs below about 300 m in these regions, so the differences in the groups between WOA and CESM2, i.e., the missing group 1 in the Amundsen coast and narrow group 2 in the Bellingshausen Sea, are unlikely to be due to the bathymetric misrepresentation (Figure 3e and f). We suggest, instead, that the mismatch of water properties is likely to be induced by the misrepresentation of fresh water input, or unresolved coastal currents, in CESM2 (Tseng, Bryan and Whitney, 2016; Sun et al., 2017).

We have highlighted a key advantage to assessing models with clustering-based approaches compared to traditional grid point-based methods; the ability to identify geographic displacements of hydrographic regimes and to distinguish these displacements from biases in water mass T-S properties. In addition, this approach minimizes

14

360 potential biases introduced during gridding or re-gridding of data and models to a common grid for comparison

361 studies. For example, it is possible to circumvent interpolation-related issues associated with using scattered and/or

362 sparse data. Such datasets might include individual observations, or model output on a native grid. For example, the

363 deepest observational temperature measurements in the World Ocean Database 2018 (WOD), even at a 1-degree

364 resolution, show that observations are available in coastal Amundsen-Bellingshausen Seas troughs that are not

365 present in IBCSO (compare Figure 10d with Figure 10a); see, also, Padman et al. (2010). More broadly, the WOD-

366 based salinity and temperature climatology of Sun et al. (2019) reveals that its use can avoid biases created by

367 spatial interpolation of shelf water with off-shelf water.

368 The success of this technique at identifying locations and properties of HSSW regimes at other locations

369 on the Antarctic continental shelf suggests that it might be used to evaluate other global and/or regional models on a

370 circum-Antarctic basis. Other metrics might be employed depending on specific research goals. For example, the

371 pycnocline depth, or the mean or maximum temperature below a fixed depth, may be better metrics of subsurface

372 water masses. It will also be interesting to track water masses and their pathways with metrics based on their

373 characteristic properties. However, we note that comparisons of the locations of groups could become complex if the

374 approach is applied to multiple models with substantial biases between their representations of specific water masses.

375 **5 Conclusions**

376 We have demonstrated the utility and sensitivity of a clustering-based approach for assessing hydrographic

377 regimes and their water properties on the Antarctic continental shelf, using the World Ocean Atlas objective analysis

378 product (WOA) and numerical model output from the Community Earth System Model version2 (CESM2). We

379 segregated the waters in the ABRS into 5 physically interpretable groups using the salinity at the minimum and

380 maximum temperature of each water column in the domain. The method identifies High Salinity Shelf Water

381 (HSSW), coastal fresh-water-enriched, and off-shelf hydrographic regimes in observations and the model. Water on

382 the continental shelf and upper continental slope in the ABRS generally show a warm bias in CESM2 compared to

383 WOA. The near-surface ocean in CESM2 is generally fresher than in WOA but lacks a well-defined fresh-water-

384 enriched coastal current. In the subsurface, CESM2 is saltier in regions of Circumpolar Deep Water, but fresher than

385 WOA in HSSW formation regions. Our comparison suggests that mean-state biases of CESM2 on the ACSS result

386 from both local and remote processes, including overestimated zonal winds in the Southern Ocean, unrepresented

387   thermodynamic interactions with ice shelves, and the inadequate representation of overflows in the Ross Sea. A

388   more specific investigation of coastal processes, Southern Ocean dynamics, and atmospheric forcing will help

389   further identify the cause of these biases.

390   The clustered hydrographic regimes in the ABRS are largely unchanged when our method is applied to the

391   entire circum-Antarctic Continental Shelf Seas. HSSW-characterized regimes emerge in WOA in the southern

392   Weddell Sea, near Mertz Glacier tongue, and in Bransfield Strait. Future work will focus on applying this approach

393   to a wider range of models (e.g., CMIP6 output and circum-Antarctic simulations) and establishing techniques to

394   work with scattered observational data. Finally, we note that the clustering results for the ACSS based on the WOA

395   decadal data (1995-2004) are consistent with the results based on the most modern WOA decadal data (2005-2017).

396   However, clustering, applied to a variety of metrics, provides the potential to identify more subtle temporal changes

397   in hydrographic fields such as changes in regime extent in the absence of significant changes in water mass

398   characteristics in the ACSS.

405

References

Adusumilli, S., Fricker, H. A., Medley, B., Padman, L., & Siegfried, M. R. (2020). Interannual variations in meltwater input to the Southern Ocean from Antarctic ice shelves. *Nature Geoscience, 13*(9), 616-620.

Agosta, C., Fettweis, X., & Datta, R. (2015). Evaluation of the CMIP5 models in the aim of regional modelling of the Antarctic surface mass balance. *The Cryosphere, 9*(6), 2311-2321.

Arndt, J. E., Schenke, H. W., Jakobsson, M., Nitsche, F. O., Buys, G., Goleby, B., . . . Wigley, R. (2013). The International Bathymetric Chart of the Southern Ocean (IBCSO) Version 1.0—A new bathymetric compilation covering circum-Antarctic waters. *Geophysical Research Letters, 40*(12), 3111-3117.

Arthur, D., & Vassilvitskii, S. (2007). k-means++: The advantages of careful seeding. *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms* (pp. 1027-1035). Society for Industrial and Applied Mathematics.

Assmann, K. M., Hellmer, H. H., & Jacobs, S. S. (2005). Amundsen Sea ice production and transport. *Journal of Geophysical Research: Oceans, 110*(C12).

Barthel, A., Agosta, C., Little, C. M., Hatterman, T., Jourdain, N. C., Goelzer, H., . . . Bracegirdle, T. J. (2019). CMIP5 model selection for ISMIP6 ice sheet model forcing: Greenland and Antarctica. *The Cryosphere Discuss*. Retrieved from https://doi.org/10.5194/tc-2019-191

Bindoff, N. L., Rosenberg, M. A., & Warner, M. J. (2000). On the circulation and water masses over the Antarctic continental slope and rise between 80 and 150 E. *Deep Sea Research Part II: Topical Studies in Oceanography, 47*(12-13), 2299-2326.

Briegleb, B. P., Danabasoglu, G., & Large, W. G. (2010). An Overflow parameterization for the ocean component of the Community Climate System Model. *University Corporation for Atmospheric Research, No. NCAR/TN-481+STR*. doi:10.5065/D69K4863

Bromwich, D. H., Nicolas, J. P., Monaghan, A. J., Lazzara, M. A., Keller, L. M., Weidner, G. A., & Wilson, A. B. (2013). Central West Antarctica among the most rapidly warming regions on Earth. *Nature Geoscience, 6*(2), 139.

Carmack, E. C. (1970). Water characteristics of the Southern Ocean south of the Polar Front. In M. Angel , *A voyage of Discovery: George Deacon 70th Anniversary* (pp. 15-42).

Cook, A. J., Holland, P. R., Meredith, M. P., Murray, T., Luckman, A., & Vaughan, D. G. (2016). Ocean forcing of glacier retreat in the western Antarctic Peninsula. *Science, 353*(6296), 283-286.

Danabasoglu, G., Lamarque, J. F., Bachmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., . . . Strand, W. G. (2019). The Community Earth System Model version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, submitted.

DeConto, R. M., & Pollard, D. (2016). Contribution of Antarctica to past and future sea-level rise. *Nature, 531*(7596), 591.

Dunn, J. R., & Ridgway, K. R. (2002). Mapping ocean properties in regions of complex topography. *Deep Sea Research Part I: Oceanographic Research Papers, 49*(3), 591-604.

Emery, W. J. (2011). Water types and water masses. *Encyclopedia of ocean sciences, 6*, 3179-3187.

Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *Kdd, 96*(34), 226-231.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development, 9*, 1937-1958. doi:10.5194/gmd-9-1937-2016

Foster, T. D., & Carmack, E. C. (1976). Frontal zone mixing and Antarctic Bottom Water formation in the southern Weddell Sea. *Deep Sea Research and Oceanographic Abstracts, 23*(4), 301-317.

Gardner, A. S., Moholdt, G., Scambos, T., Fahnstock, M., Ligtenberg, S., van den Broeke, M., & Nilsson, J. (2018). Increased West Antarctic and unchanged East Antarctic ice discharge over the last 7 years. *Cryosphere, 12*(2), 521-547.

Gordon, A. L., Mensch, M., Zhaoqian, D., Smethie Jr., W. M., & De Bettencourt, J. (2000). Deep and bottom water of the Bransfield Strait eastern and central basins. *Journal of Geophysical Research: Oceans, 105*(C5), 11337-11346.

Hjelmervik, K. T., & Hjelmervik, K. (2013, June). Improved estimation of oceanographic climatology using empirical orthogonal functions and clustering. *In 2013 MTS/IEEE OCEANS-Bergen*, 1-5.

Hobbs, W. R., Massom, R., Stammerjohn, S., Reid, P., Williams, G., & Meier, W. (2016). A review of recent changes in Southern Ocean sea ice, their drivers and forcings. *Global and Planetary Change, 143*, 228-250.

460 Holland, M. M., Landrum, L., Raphael, M., & Stammerjohn, S. (2017). Springtime winds drive Ross Sea ice
461   variability and change in the following autumn. *Nature communications, 8*(1), 1-8.
462 Hosking, J. S., Orr, A., Bracegirdle, T. J., & Turner, J. (2016). Future circulation changes off West Antarctica:
463   Sensitivity of the Amundsen Sea Low to projected anthropogenic forcing. *Geophysical Research Letters,*
464   *43*(1), 367-376.
465 Hunke, E. C., Lipscomb, W. H., Turner, A. K., Jeffery, N., & Elliott, S. (2015). CICE: The Los Alamos Sea Ice
466   Model. Documentation and Software User's Manual. Version 5.1. *T-3 Fluid Dynamics Group, Los Alamos*
467   *National Laboratory, Tech. Rep. LA-CC-06-012*.
468 Jacobs, S. S., & Giulivi, C. F. (2010). Large multidecadal salinity trends near the Pacific–Antarctic continental
469   margin. *Journal of Climate, 23*(17), 4508-4524.
470 Jourdain, N. C., Mathiot, P., Merino, N., Durand, G., Le Sommer, J., Spence, P., . . . Madec, G. (2017). Ocean
471   circulation and sea-ice thinning induced by melting ice shelves in the Amundsen Sea. *Journal of*
472   *Geophysical Research: Oceans, 122*(3), 2550-2573.
473 Little, C. M., & Urban, N. M. (2016). CMIP5 temperature biases and 21st century warming around the Antarctic
474   coast. *Annals of Glaciology, 57*(73), 69-78.
475 Locarnini, R. A., Mishonov, A. V., Baranova, O. K., Boyer, T. P., Zweng, M. M., Garcia, H. E., . . . Smolyar, I. V.
476   (2019). World Ocean Atlas 2018, Volume 1: Temperature. In A. Mishonov, *NOAA Atlas NESDIS 81.*
477 McDougall, T. J., & Barker, P. M. (2011). Getting started with TEOS-10 and the Gibbs Seawater (GSW)
478   oceanographic toolbox. *SCOR/IAPSO WG, 127*, 1-28.
479 Moffat, C., Beardsley, R. C., Owens, B., & Van Lipzig, N. (2008). A first description of the Antarctic Peninsula
480   Coastal Current. *Deep Sea Research Part II: Topical Studies in Oceanography, 55*(3-4), 277-293.
481 Nicholls, K. W., Østerhus, S., Makinson, K., Gammelsrød, T., & Fahrbach, E. (2009). Ice-ocean processes over the
482   continental shelf of the southern Weddell Sea Antarctica: A review. *Reviews of Geophysics, 47*(3).
483 Orsi, A. H., & Wiederwohl, C. L. (2009). A recount of Ross Sea waters. *Deep Sea Research Part II, 56*(13-14), 778-
484   795.
485 Orsi, A. H., & Wiederwohl, C. L. (2009). A recount of Ross Sea waters. *Deep Sea Research Part II: Topical Studies*
486   *in Oceanography, 56*(13-14), 778-795.
487 Orsi, A. H., Johnson, G. C., & Bullister, J. L. (1999). Circulation, mixing, and production of Antarctic Bottom
488   Water. *Progress in Oceanography, 43*(1), 55-109.
489 Padman, L., Costa, D. P., Bolmer, S. T., Goebel, M. E., Huckstadt, L. A., Jenkins, A., . . . Shoosmith, D. R. (2010).
490   Seals map bathymetry of the Antarctic continental shelf. *Geophysical Research Letters, 37*(21).
491 Paolo, F. S., Fricker, H. A., & Padman, L. (2015). Volume loss from Antarctic ice shelves is accelerating. *Science,*
492   *348*(6232), 327-331.
493 Parkinsona, C. L. (2019). A 40-y record reveals gradual Antarctic sea ice increases followed by decreases at rates far
494   exceeding the rates seen in the Arctic. *Proceedings of the National Academy of Sciences, 116*(29), 14414-
495   14423.
496 Porter, F. D., Springer, S. R., Padman, L., Fricker, H. A., Tinto, K. J., Riser, S. C., . . . the ROSETTA-Ice Team.
497   (2019). volution of the Seasonal Surface Mixed Layerof the Ross Sea, Antarctica, ObservedWith
498   Autonomous Profiling Floats. *Journal of GeophysicalResearch: Oceans, 124*(7), 4934–4953.
499 Porter, S. E., Parkinson, C. L., & Mosley-Thompson, E. (2016). Bellingshausen Sea ice extent recorded in an
500   Antarctic Peninsula ice core. *Journal of Geophysical Research: Atmospheres, 121*(23), 13-886.
501 Post, A. L., Beaman, R. J., O'Brien, P. E., Eléaume, M., & Riddle, M. J. (2011). Community structure and benthic
502   habitats across the George V Shelf, East Antarctica: trends through space and time. *Deep Sea Research*
503   *Part II: Topical Studies in Oceanography, 58*(1-2), 105-118.
504 Rickard, G., & Behrens, E. (2016). CMIP5 Earth system models with biogeochemistry: A Ross Sea assessment.
505   *Antarctic Science, 28*(5), 327-346.
506 Rignot, E., Bamber, J. L., Van Den Broeke, M. R., Davis, C., Li, Y., Van De Berg, W. J., & Van Meijgaard, E.
507   (2008). Recent Antarctic ice mass loss from radar interferometry and regional climate modelling. *Nature*
508   *geoscience, 1*(2), 106-110.
509 Rignot, E., Jacobs, S., Mouginot, E., & Scheuchl, B. (2013). Ice-shelf melting around Antarctica. *Science,*
510   *341*(6143), 266-270.
511 Sallée, J. B., Shuckburgh, E., Bruneau, N., Meijers, A. J., Bracegirdle, T. J., Wang, Z., & Roy, T. (2013).
512   Assessment of Southern Ocean water mass circulation and characteristics in CMIP5 models: Historical bias
513   and forcing response. *Journal of Geophysical Research: Oceans, 118*(4), 1830-1844.
514 Schmidtko, S., Heywood, K. J., Thompson, A. F., & Aoki, S. (2014). Multidecadal warming of Antarctic waters.
515   *Science, 346*(6214), 1227-1231.

516   Schmidtko, S., Johnson, G. C., & Lyman, J. M. (2013). MIMOC: A global monthly isopycnal upper-ocean
517   climatology. *Journal of Geophysical Research: Oceans, 118*, 1658–1672.
518   Shepherd, A., Ivins, E. R., Geruo, A., Barletta, V. R., Bentley, M. J., Bettadpur, S., . . . Horwath, M. (2012). A
519   reconciled estimate of ice-sheet mass balance. *Science, 338*(6111), 1183-1189.
520   Shepherd, A., Ivins, E., Rignot, E., Smith, B., van den Broeke, M., Velicogna, I., . . . Wouters, B. (2018). Mass
521   balance of the Antarctic Ice Sheet from 1992 to 2017. *Nature*(558), 219-222.
522   Singh, H. K., Landrum, L., Holland, M. M., Bailey, D. A., & DuVivier, A. K. (2020). An Overview of Antarctic Sea
523   Ice in the Community Earth System Model version 2, Part I: Analysis of the Seasonal Cycle in the Context
524   of Sea Ice Thermodynamics and Coupled Atmosphere-Ocean-Ice Processes. *Journal of Advances in*
525   *Modeling Earth Systems*, e2020MS002143.
526   Stammerjohn, S., Massom, R., Rind, D., & Martinson, D. (2012). Regions of rapid sea ice change: An inter-
527   hemispheric seasonal comparison. *Geophysical Research Letters, 39*(6).
528   Sun, Q., Whitney, M. M., Bryan, F. O., & Tseng, Y.-h. (2017). A box model for representing estuarine physical
529   processes in Earth system models. *Ocean Modelling, 112*, 139-153.
530   Sun, Q., Whitney, M. M., Bryan, F. O., & Tseng, Y.-h. (2019). Assessing the skill of the improved treatment of
531   riverine freshwater in the Community Earth System Model relative to a new salinity climatology. *Journal*
532   *of Advances in Modeling Earth Systems*. doi:10.1029/2018MS001349
533   Sutterley, T. C., Velicogna, I., Rignot, E., Mouginot, J., Flament, T., Van Den Broeke, J. M., . . . Reijmer, C. H.
534   (2014). Mass loss of the Amundsen Sea Embayment of West Antarctica from four independent techniques.
535   *Geophysical Research Letters, 41*(23), 8421-8428.
536   Thompson, A. F., Stewart, A. L., Spence, P., & Heywood, K. J. (2018). The Antarctic Slope Current in a changing
537   climate. *Reviews of Geophysics, 56*(4), 741-770.
538   Thorndike, R. L. (1953). Who Belongs in the Family? *Psychometrika, 18*(4), 267–276.
539   Timmermans, M. L., Proshutinsky, A., Golubeva, E., Jackson, J. M., Krishfield, R., McCall, M., . . . Nishino, S.
540   (2014). Mechanisms of Pacific summer water variability in the Arctic's Central Canada Basin. *Journal of*
541   *Geophysical Research: Oceans, 119*(11), 7523-7548.
542   Tseng, Y.-H., Bryan, F. O., & Whitney, M. M. (2016). Impacts of the representation of riverine freshwater input in
543   the community earth system model. *Ocean Modelling, 105*, 71-86.
544   Zweng, M. M., Reagan, J. R., Seidov, D., Boyer, T. P., Locarnini, R. A., Garcia, H. E., . . . Smolyar, I. V. (2019).
545   World Ocean Atlas 2018, Volume 2: Salinity. In A. Mishonov, *NOAA Atlas NESDIS 82.*
546
547

Tables and Figures

Table 1 The radius ε used in the DBSCAN for WOA and CESM2

	Group 1	Group 2	Group 3	Group 4	Group 5
WOA	0.05	0.05	0.04	0.03	0.03
CESM2	0.045	0.04	0.06	0.035	0.04

Table 2 The coverage (%) of the majority group of DBSCAN in the total non-outlier data

	Group 1	Group 2	Group 3	Group 4	Group 5
WOA	99.6	97.9	99.9	100	100
CESM2	100	97.3	99.5	99.9	99.7

Table 3 The salinity and temperature standard deviation of WOA (at depth of 500 m if not specified)

		Salinity (g/kg)		Temperature (°C)	
		Amundsen & Bellingshausen	Ross	Amundsen & Bellingshausen	Ross
Geography		0.16 (200 m)	0.11	0.84 (200 m)	1.37
		0.10		1.42	
K-means groups	1	0.10 (200 m)	N/A	0.22 (200 m)	N/A
	2	0.07		1.34	
	3	N/A	0.08	N/A	0.97
	4	0.08		0.44	
	5	N/A	0.10	N/A	0.17

Table 4 The percentage of clustered water in the total ocean volume in the ABRS.

	Group 1	Group 2	Group 3	Group 4	Group 5
WOA	1.0	3.6	21.0	62.1	12.3
CESM2	0.3	7.2	33.2	50.4	8.9

563

564

(a)

(b)

(c)

(d)

565
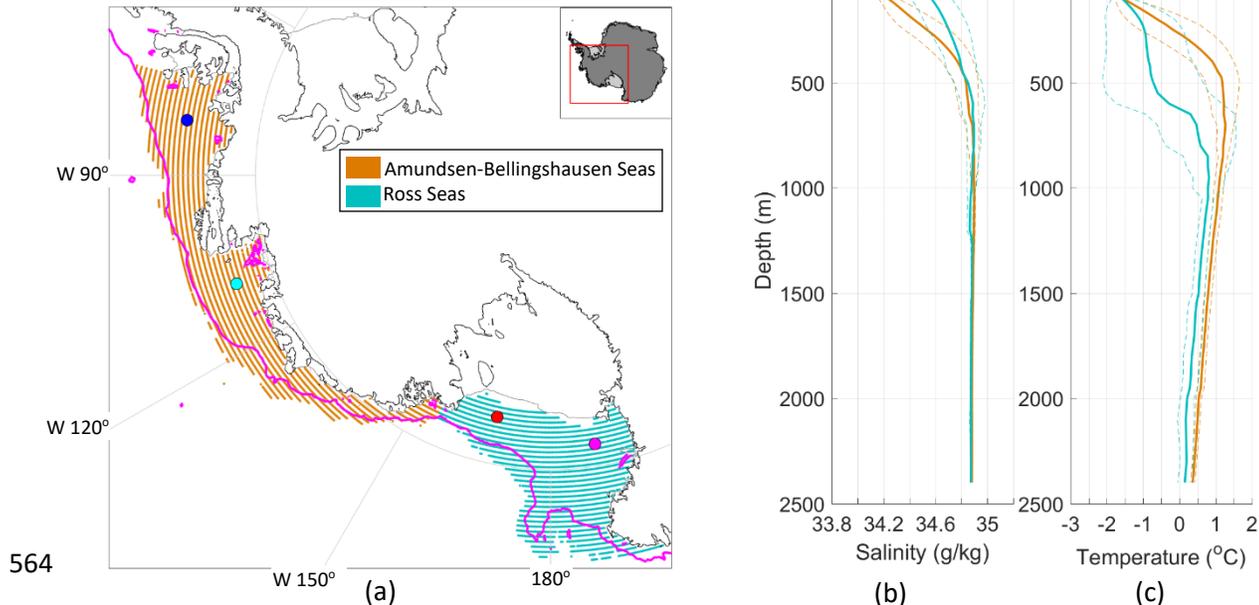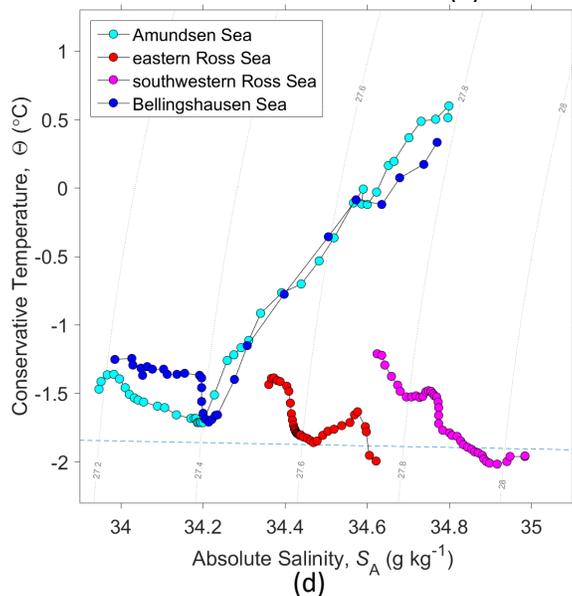
Figure 1 (a) the study domain of Amundsen-Bellingshausen Seas and Ross Sea with bathymetry above 2,500 m. The magenta line indicates the 1,000 m IBCSO depth contour. (b) and (c) show geographically averaged decadal (1995-2004) WOA salinity and temperature profiles in the Amundsen-Bellingshausen Seas (orange; corresponding to the orange stippled region in (a)) and the Ross Sea (cyan; corresponding to the cyan stippled region in (a)). Dashed lines indicate ±1 standard deviation of values at each depth in each region. d) T-S properties of selected water columns (corresponding to colored circles in panel (a)).

572
573
574
575
576
577
578

21

579

580

581
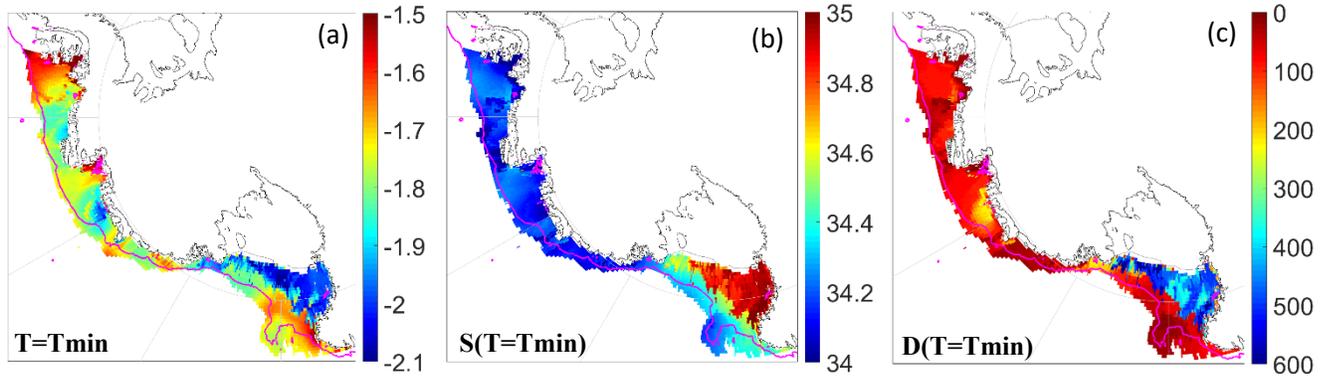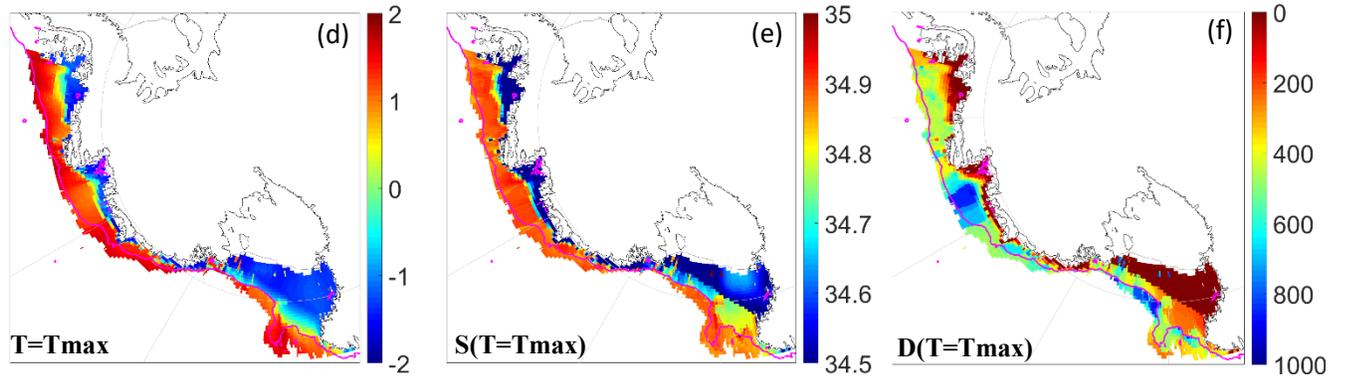
Figure 2 Clustering metrics in WOA. Minimum temperature at each grid point (a), and the salinity (b) and water depth at the minimum temperature. d-f) as a-c, but for quantities at the temperature maximum.

584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
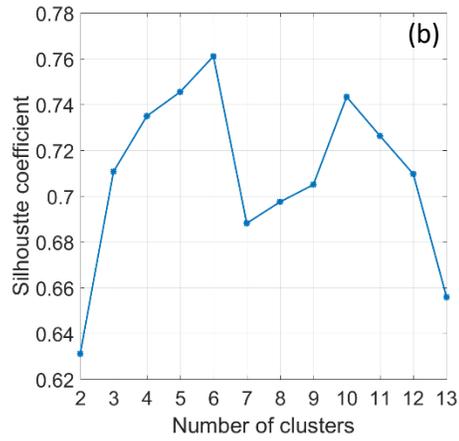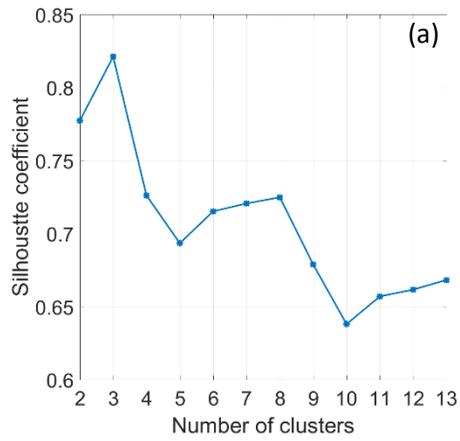
Figure 3 K-means clustering evaluation for WOA and CESM. Silhouette analysis is shown in (a) and (b) for WOA and CESM, respectively. The geographic regions corresponding to 3, 5 and 6 groups for WOA, (shown in c, e and g). and for CESM (shown in d, f and h).

609

610

611

612

23

613

614

615
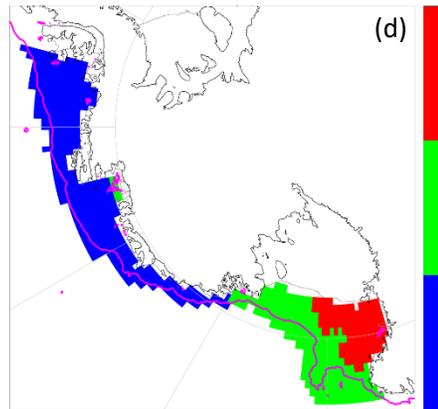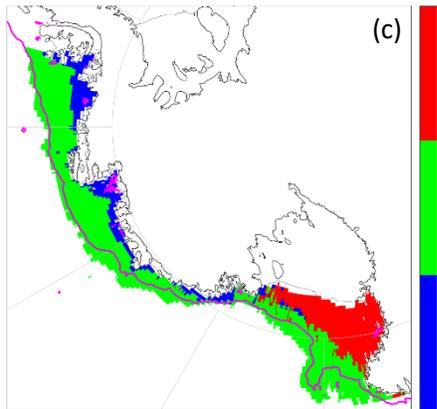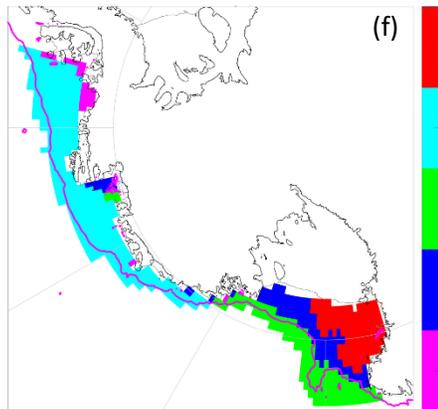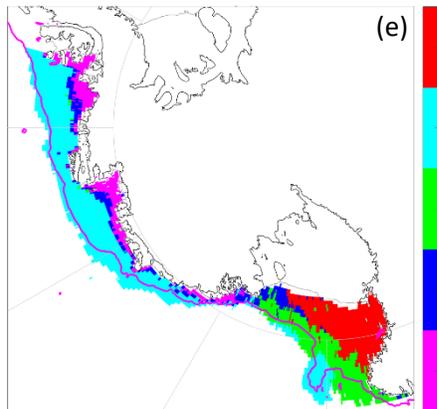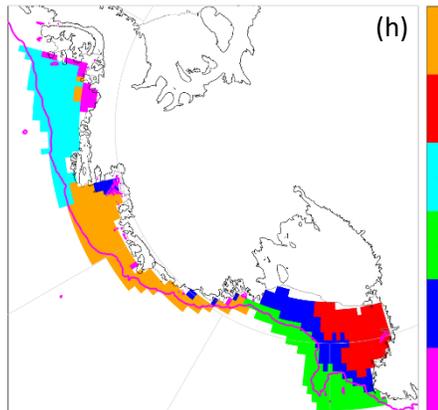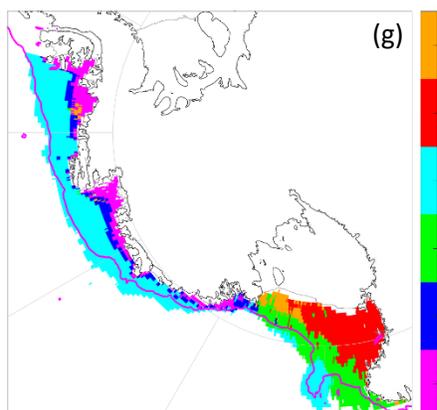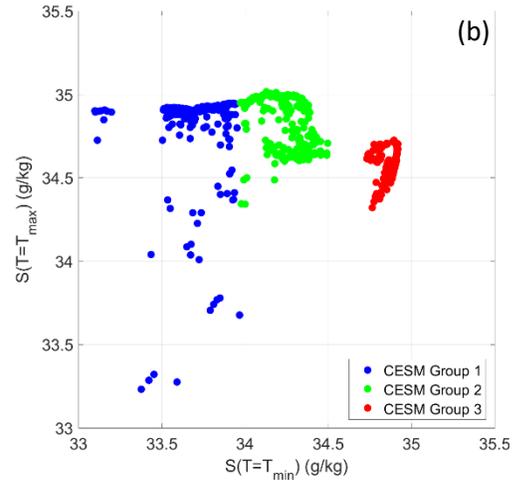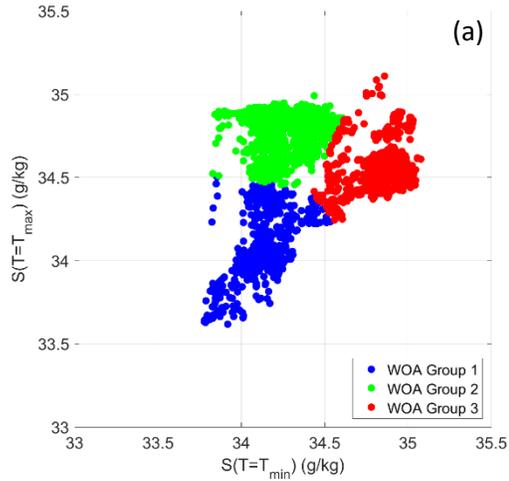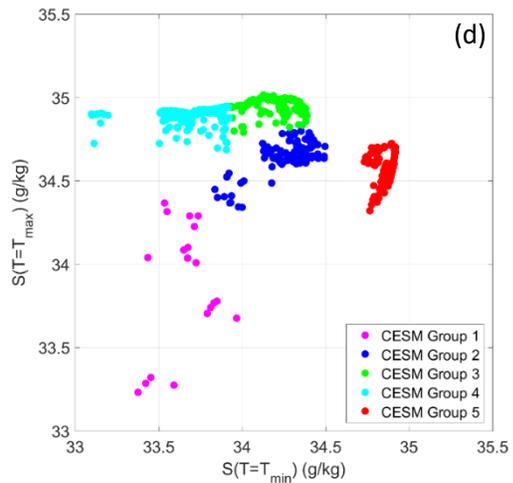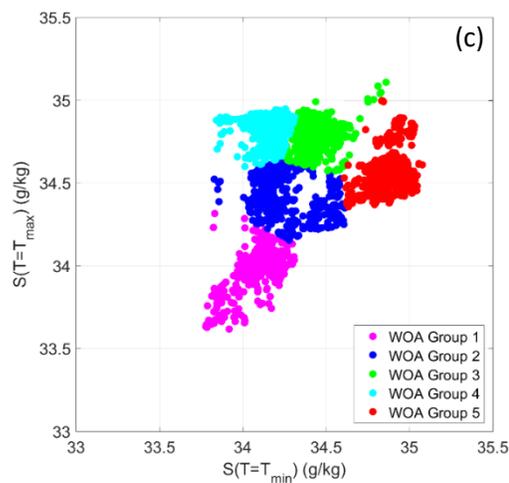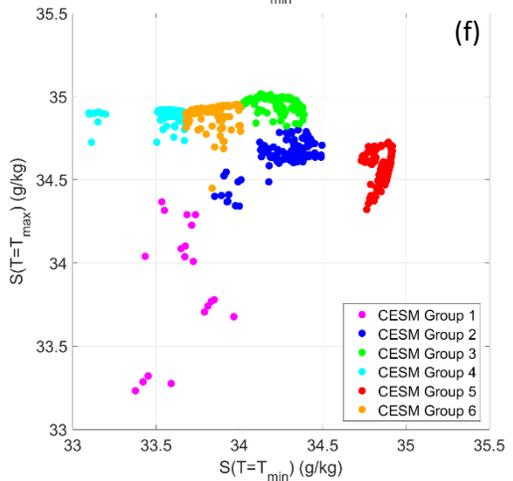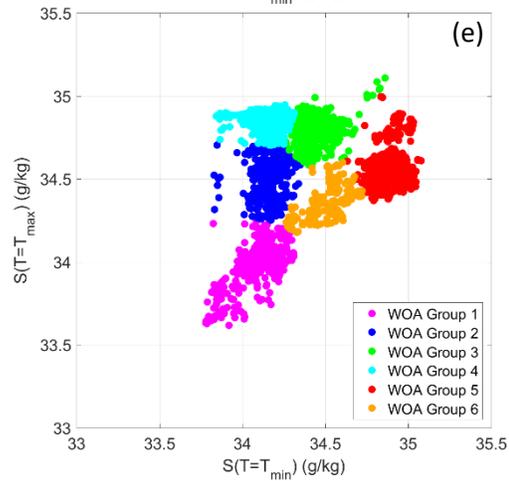
616
617 **Figure 4. ABRS groups in metric space. Each point corresponds to a grid point, with a color corresponding to its group**
618 **number, for K=3, 5 and 6, for WOA, (shown in a, c and e). and for CESM2 (shown in b, d and f).**

619
620

621

622

623
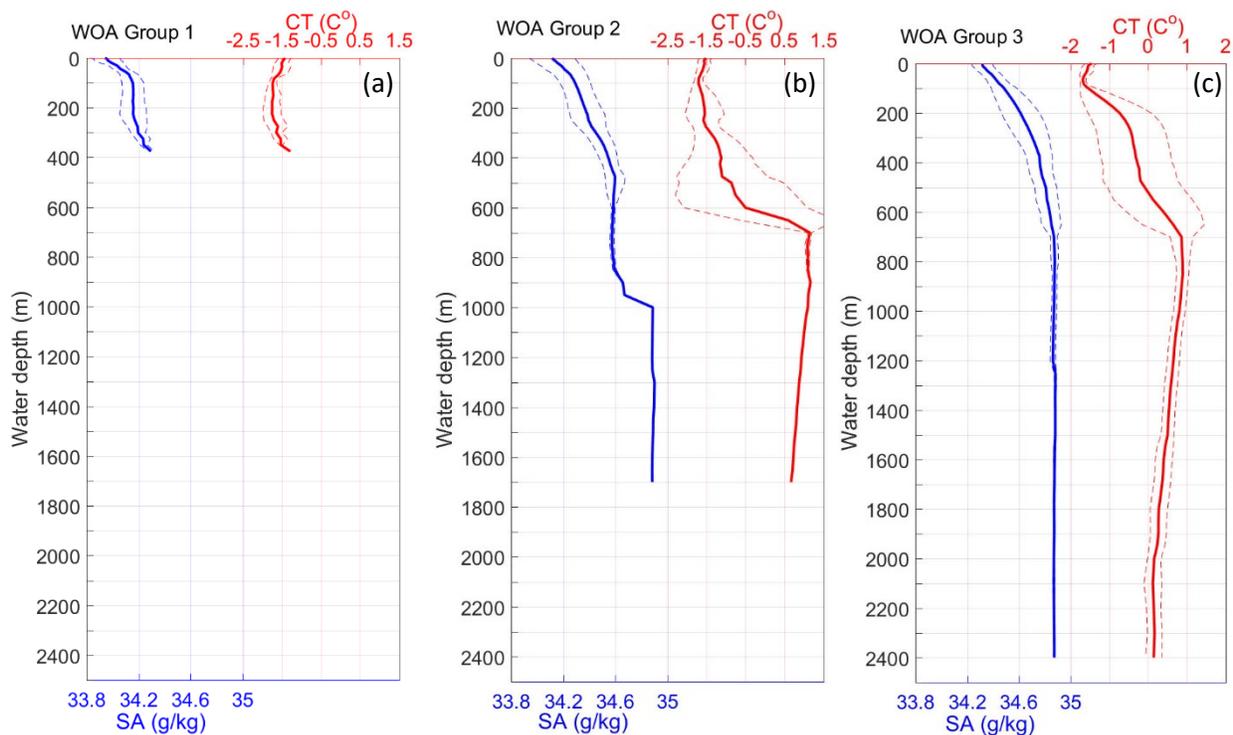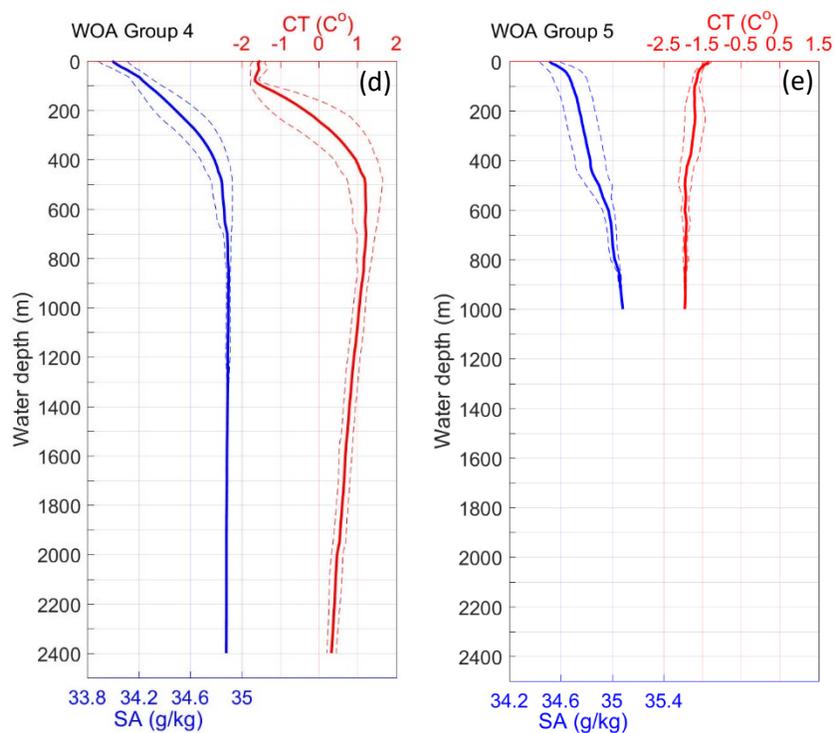
Figure 5 Mean (solid lines) WOA salinity (in blue) and temperature (in red) profiles for five groups (from a to e) shown in Figure 3 (e). ±1 standard deviation at each depth is shown with dashed lines.

626

627

628

629
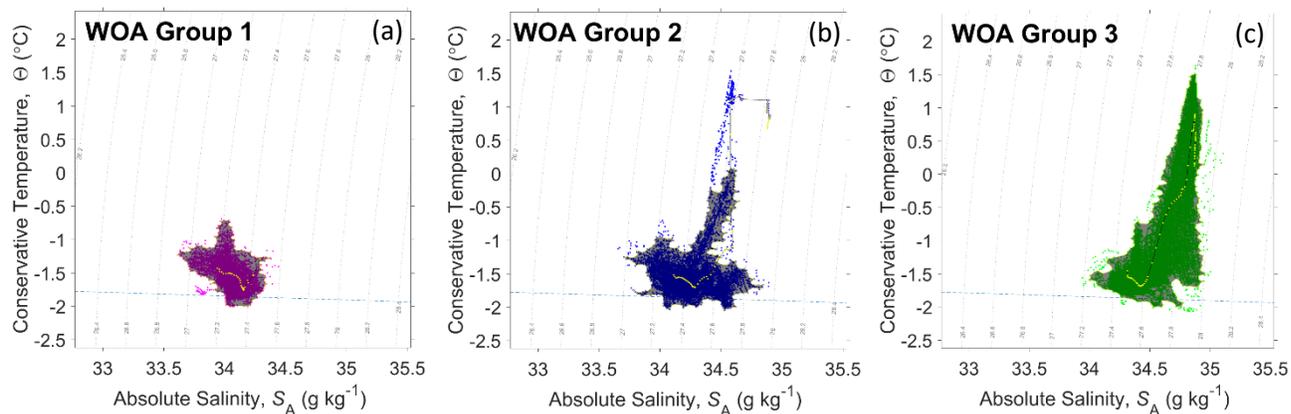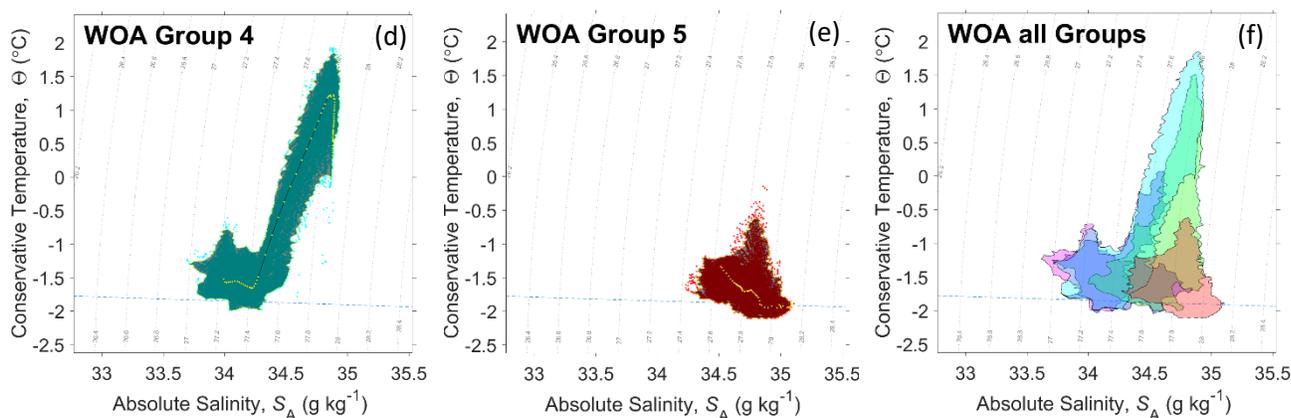
Figure 6 *T-S* properties for the five WOA groups (from a to e) shown in Figure 3 (e). The yellow dotted lines show the profile of mean temperature and salinity in each group, and the dark shaded areas are the cores of water property from the density-based clustering results. The cores of all five groups are overlaid on the same plot in (f).

630
631
632

633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654

655
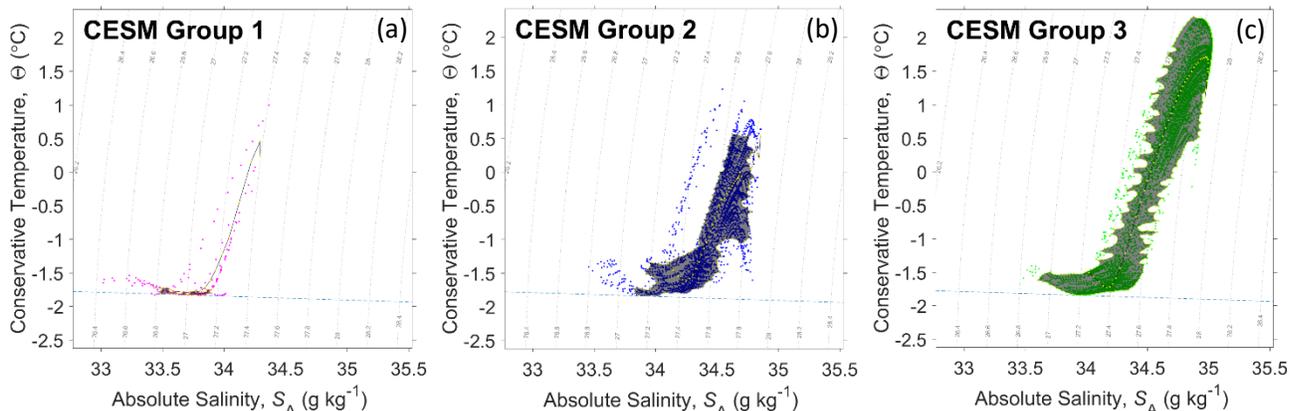
656
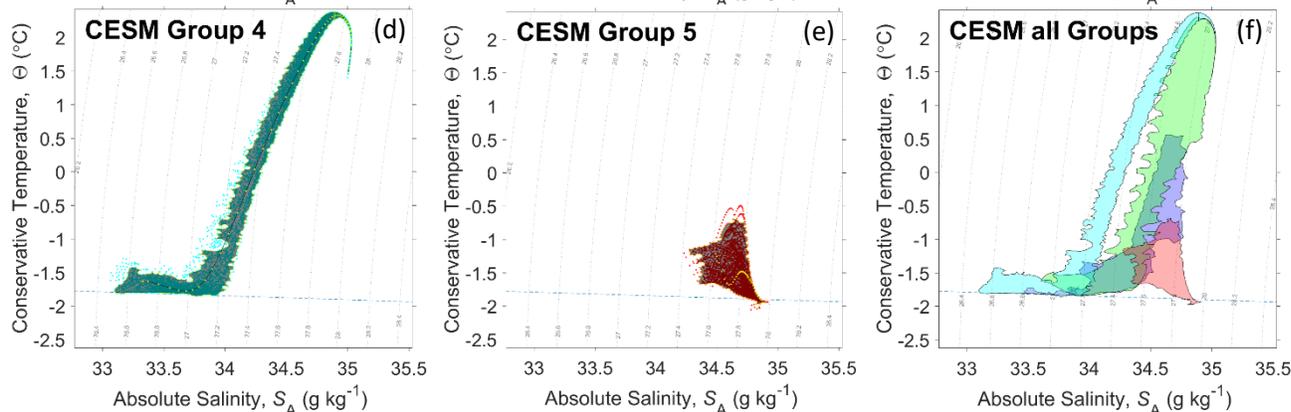
657
658 **Figure 7 As Figure 6, but for the five groups identified in CESM2 (from a to e) shown in Figure 3 (f).**

659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684

685

686

687
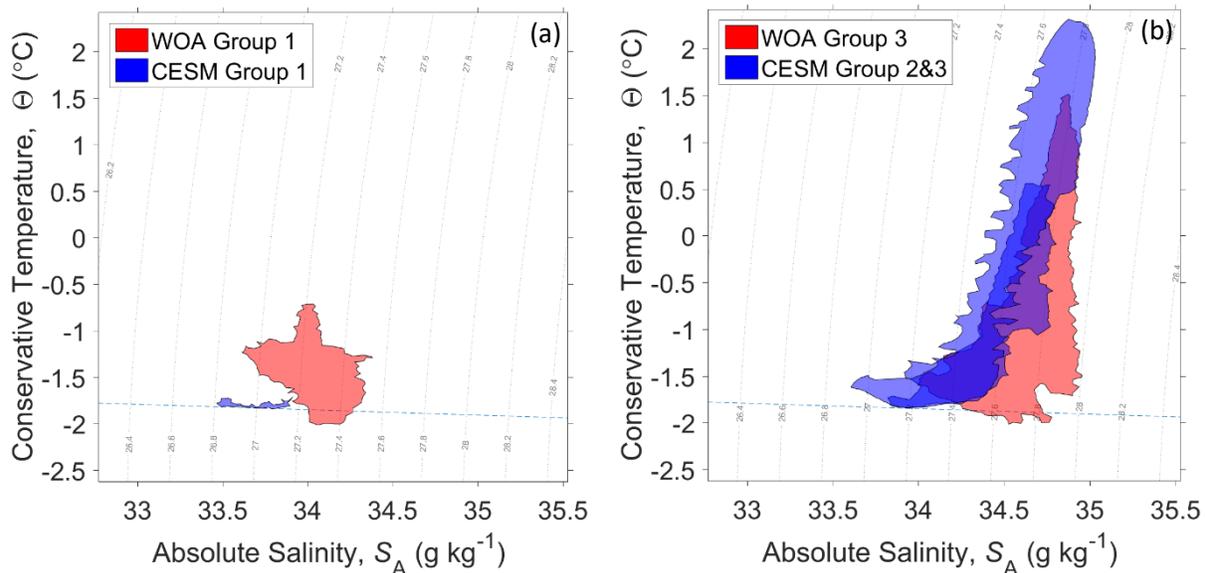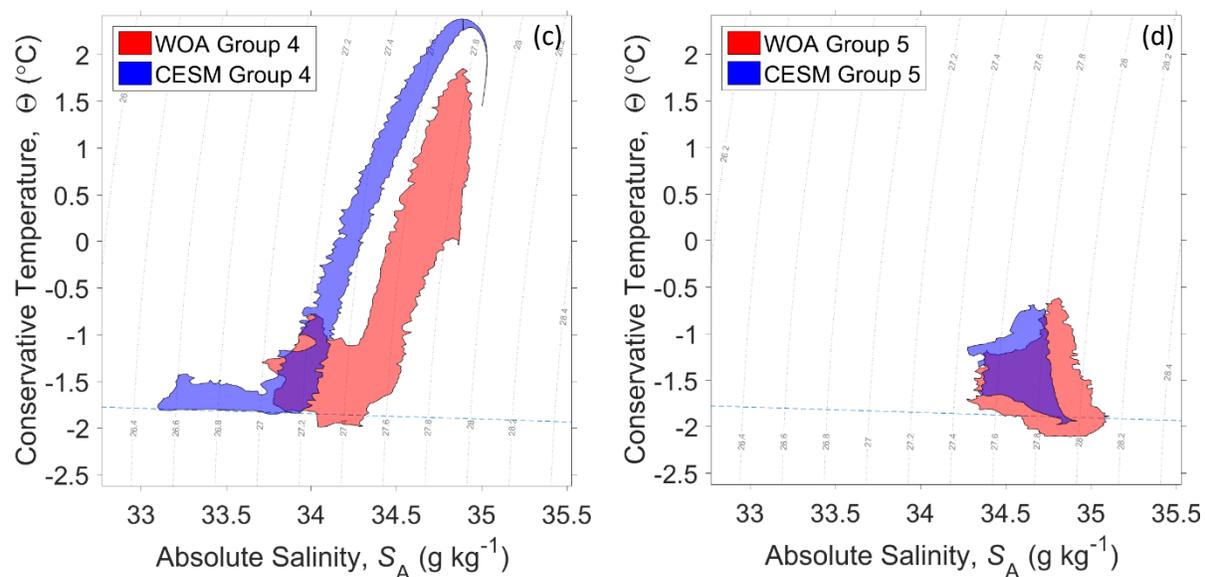
Figure 8 The core of water properties in WOA (red) and CESM2 (blue). Note that groups 2 and 3 have been combined for CESM2.

688
689

690
691
692
693
694
695
696
697
698
699
700

701

702

703
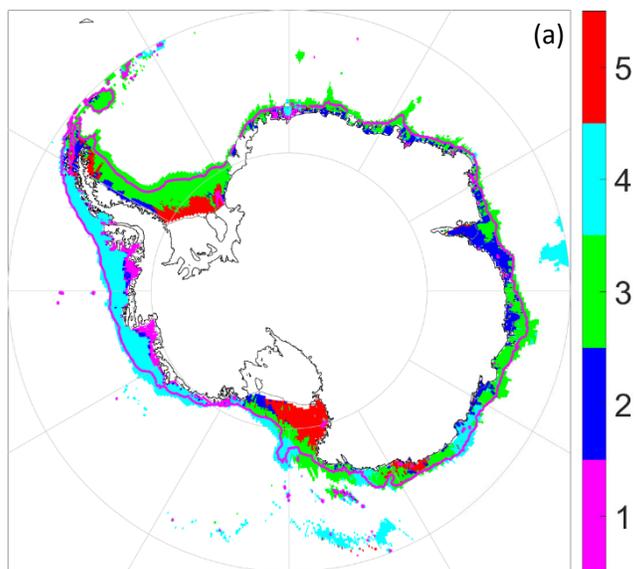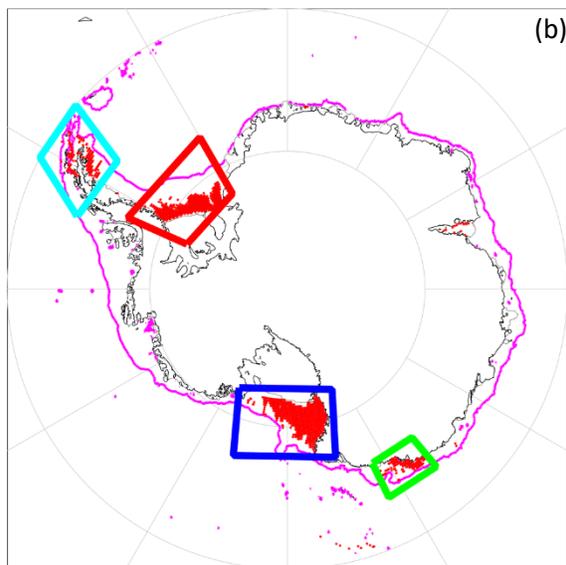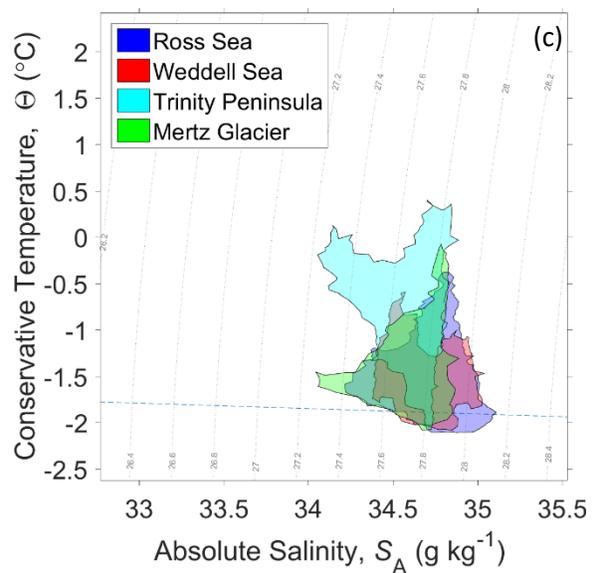
Figure 9 (a) WOA-based groups on the entire ACSS (same color code as Figure 3e). (b) Four places are identified as HSSW regime with color codes blue: southwestern Ross Sea; red: Weddell Sea near the Filchner-Ronne Ice Shelf, the George V Coast; cyan: Bransfield Strait and south of Trinity Peninsula; and green: Mertz Glacier tongue. (c)*T-S* **properties of group 5 (HSSW) regions, with their geographic location and color code matched in (b).**

708
709
710
711
712
713
714
715
716
717

718

719

720
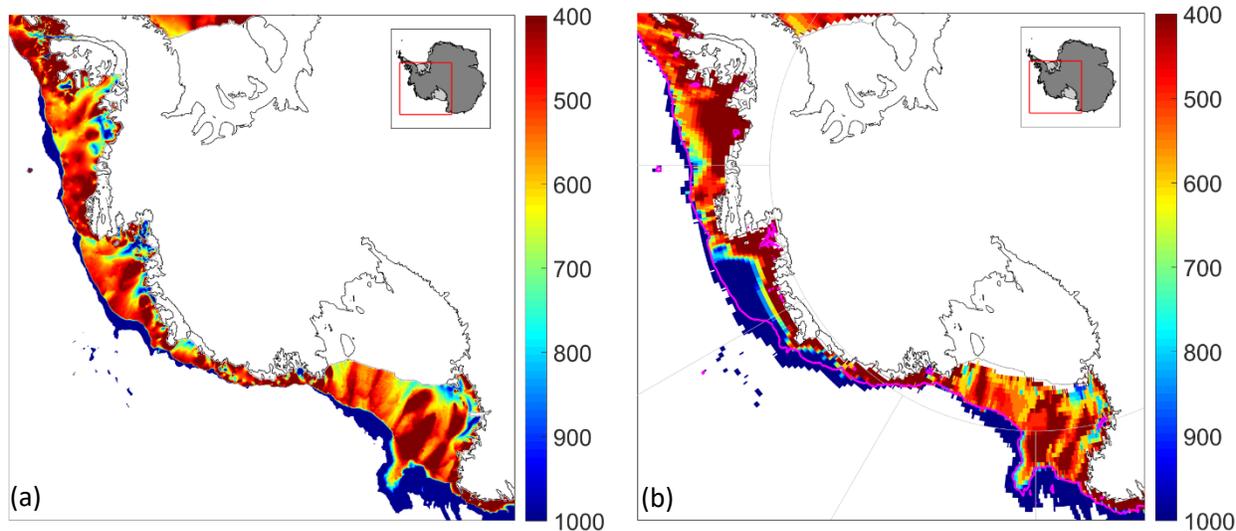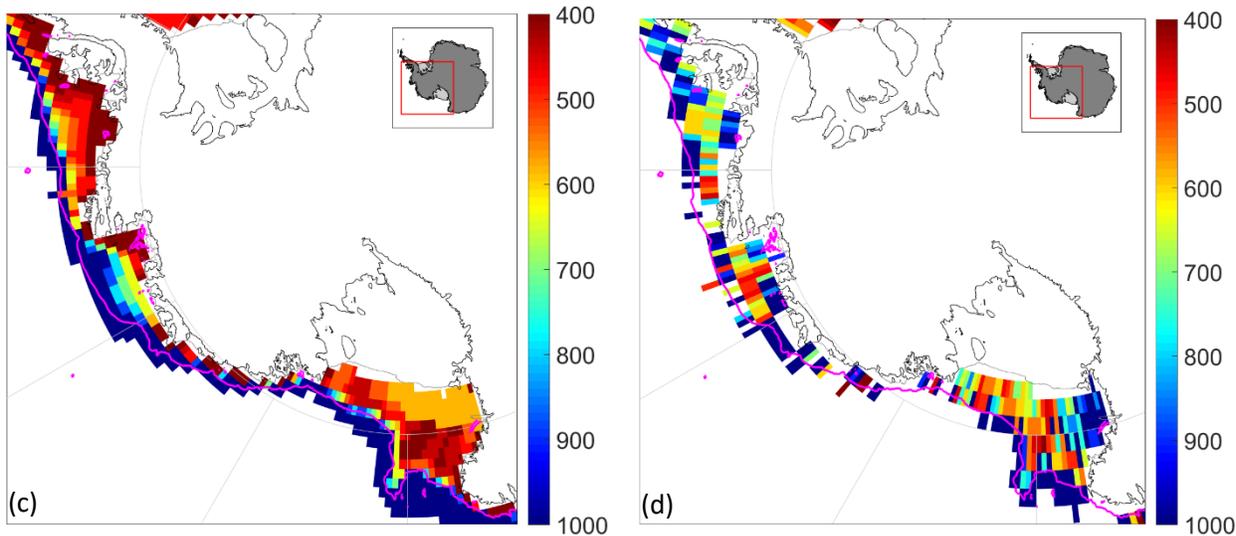
Figure 10 Bathymetry between $400\ m$ and $1,000\ m$ of: (a) IBCSO (500 m horizontal resolution), (b) WOA (0.25-degree horizontal resolution), (c) CESM2 (1 x 0.5-degree lon/lat resolution), and (d) the WOD bin-averaged into 1-degree horizontal resolution with all types of instrument with temperature measurements). The magenta line indicates the 1,000m IBCSO depth contour.