

Unsupervised classification identifies warm, fresh and dense regimes of the Antarctic margins

T. Sohail^{a,b} and J.D. Zika^{a,b,c}

^a *School of Mathematics and Statistics, University of New South Wales, Sydney, Australia*

^b *Australian Centre for Excellence in Antarctic Science (ACEAS), University of New South Wales,
NSW, Australia*

^c *UNSW Data Science Hub, University of New South Wales, Sydney, Australia*

Corresponding author: Taimoor Sohail, t.sohail@unsw.edu.au

9 ABSTRACT: The ocean surrounding Antarctica, also known as the Antarctic margins, is char-
10 acterised by complex and heterogeneous processes interactions which have major impacts on the
11 global climate. A common way to understand changes in the Antarctic margins is to categorise
12 regions into similar ‘regimes’, thereby guiding process-based studies and observational analyses.
13 However, this categorisation is traditionally largely subjective and based on temperature, density
14 and bathymetric criteria that are bespoke to the dataset being analysed. In this work, we introduce a
15 method to classify Antarctic shelf regimes using unsupervised learning. We apply *Gaussian Mix-*
16 *ture Modelling* to the across shelf temperature and salinity properties along the Antarctic margins
17 from a high-resolution ocean model, ACCESS-OM2-01. Three clusters are found to be optimum
18 according to the Bayesian Information Criterion. The three clusters correspond to the *fresh*, *dense*
19 and *warm* regimes identified canonically via subjective approaches. Our objective analysis allows
20 us to track changes to these regimes in a future projection of the ACCESS-OM2-01 model. We
21 identify the future collapse of dense water formation, and the merging of dense and fresh shelf
22 regions into a single fresh regime that covers the entirety of the Antarctic shelf except for the West
23 Antarctic. Our assessment of these objective clusters indicates that the Antarctic margins will shift
24 into a two-regime system in the future, consisting only of a strengthening warm shelf in the West
25 Antarctic and a fresh shelf regime everywhere else.

26 SIGNIFICANCE STATEMENT: The Antarctic margins are characterised by complex interac-
27 tions of surface and ocean processes, producing distinct regions or ‘regimes’. Understanding where
28 these regimes are and their future state is critical to understanding climate change. Based on a sub-
29 jective assessment of ocean conditions, past research has identified fresh, dense and warm regimes
30 in the Antarctic margins. In this work, we use an unsupervised classification tool, *Gaussian Mixture*
31 *Modelling*, to *objectively* identify regimes around the Antarctic margins. Our objective method
32 detects three regimes in an ocean model which match the subjectively identified fresh, dense and
33 warm regimes, and indicates a future shrinking of the dense regime. Our method is adaptable to
34 multiple datasets, enabling us to identify trends and processes in the Antarctic margins.

35 1. Introduction

36 The Antarctic margins play a crucial role in the climate system, and changes to the ocean
37 surrounding Antarctica can have major ramifications on the global impacts of climate change. In
38 the Antarctic shelf region, ice melting, sea ice formation, dense shelf water formation, circumpolar
39 deep water upwelling, surface winds and large-scale currents all interact to create a spatially
40 heterogeneous and ever-changing set of conditions (Stewart and Thompson 2015; Huneke et al.
41 2022; Neme et al. 2022; Thompson et al. 2018; Huguenin et al. 2022; Morrison et al. 2020; Darelius
42 et al. 2023; Oppenheimer et al. 2019). It is important to categorise different regions in the Antarctic
43 margins into similar ‘regimes’, given the complexity of the processes and climatic conditions there.
44 Doing so not only frames our understanding of the processes that govern flow, but also guides
45 process-based studies and observational mapping methods of the Antarctic margins.

46 Thompson et al. (2018) (hereafter T18) classifies thermodynamic conditions in the Antarctic
47 margins into ‘fresh’, ‘dense’ and ‘warm’ shelf regimes based on a review of ocean observations
48 (see figure 1 for examples of each regime in the model used in this study). The fresh regime is
49 characterised by a surface layer of freshwater that shoals onto the shelf and denies access to the
50 warm, salty Circumpolar Deep Water (CDW) which can enhance ice melt. In the dense regime, high
51 rates of sea ice formation and/or atmospheric cooling lead to salty, cold convective plumes, known
52 as Dense Shelf Water (DSW), downwelling and flowing off the shelf (Morrison et al. 2020; Li et al.
53 2023). In the warm regime, the warm, salty CDW flows onto the shelf unencumbered, enhancing
54 basal melting and threatening the stability of the ice sheet (Herraiz-Borreguero and Garabato 2022).

55 Knowledge of these thermodynamic regimes enables further idealised Antarctic studies (e.g., Daae
56 et al. (2017); Stewart and Thompson (2015)) and allows for additional categorisation of important
57 shelf processes, such as the Antarctic Slope Current (as in Huneke et al. (2022)).

58 The T18 criteria for defining Antarctic shelf regimes is based on whether the 28 kgm^{-3} neutral
59 surface extends onto the shelf, and whether temperatures greater than 0.5°C can be found on the
60 shelf. When Moorman et al. (2020) subsequently categorised a model output in the same way as
61 T18, they modified the criteria to reflect the prevailing conditions in the simulation. Furthermore,
62 they proposed the addition of a fourth, *Cool* regime. This redefinition is illustrative of the fact that
63 present regime classifications are inherently subjective, and thus need to be modified to fit different
64 datasets, models, and future scenarios. To enhance the utility of such shelf regime definitions across
65 Antarctic studies, there is a need to *objectively* define Antarctic shelf regimes in an unsupervised
66 way. Enabling such an unsupervised classification is the primary focus of this study.

72 Clustering algorithms have been increasingly used to classify the earth system into discrete
73 regions depending on the purpose of the clustering. For instance, Sohail et al. (2023) used *Binary*
74 *Space Partitioning* to cluster the global ocean into equal-volume regions. In the Antarctic context,
75 a number of unsupervised clustering methods have been used to categorise oceanic regimes.
76 Pauthenet et al. (2021) used *Principal Component Analysis* to identify the principal modes of
77 variability in the Antarctic shelf, and thus to develop a new gridded hydrographic dataset for
78 the region. Jones et al. (2019, 2023); Rosso et al. (2020); Boehme and Rosso (2021) used a
79 Profile Classification Method (PCM) based on *Gaussian Mixture Modelling* (GMM) to identify
80 key regions in the Southern Ocean and Antarctic basins from temperature and/or salinity profiles,
81 and Sun et al. (2020) used *K-means* clustering to categorise shelf regions in $T - S$ space.

82 Such past clustering efforts have been hampered by the fact that most off-the-shelf classification
83 algorithms cannot handle missing data as an input (e.g. due to bathymetric incrops and features).
84 This necessitates clustering the data in phase space (as in Sun et al. (2020)) or removing temper-
85 ature/salinity profiles above or below a certain depth (as in Jones et al. (2019); Pauthenet et al.
86 (2021), amongst others). In addition, no prior clustering work has worked towards updating the
87 T18 criteria using unsupervised methods.

88 In this work, we use unsupervised learning to objectively classify the Antarctic shelf region into
89 thermodynamic regimes in a high-resolution ocean model. In section 2, we introduce the ocean

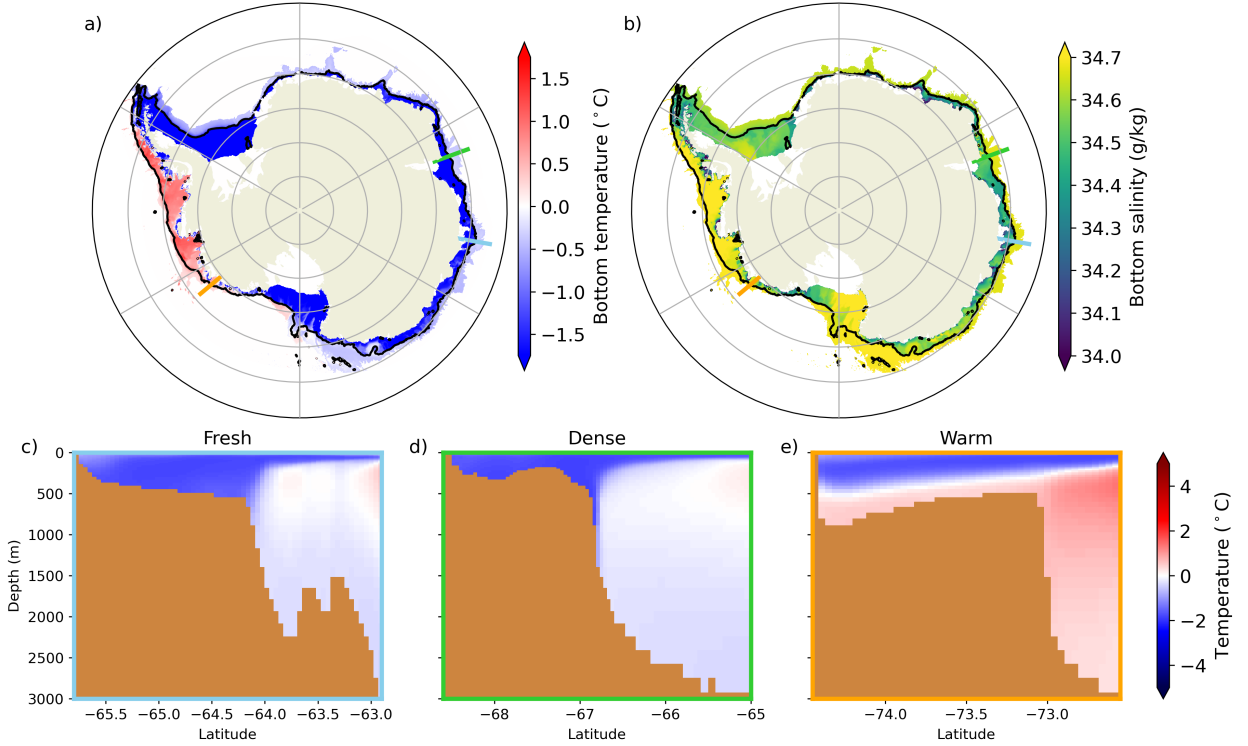


FIG. 1. a) Bottom temperature and b) bottom salinity in the Antarctic margins for regions with bottom depths less than 3000m from the ACCESS-OM2-01 RYF simulation used in this study. The black contour line shows the 1000m isobath, and the blue, green and orange lines show the longitudes at which a slice is taken to illustrate the fresh, dense and warm regimes, respectively. Temperature field in the c) fresh, d) dense and e) warm longitudinal slices. Bathymetric features are coloured in brown.

model and simulations assessed in this work. In section 3, we explain the input data treatment, clustering algorithm and classification methods employed in this study. In section 4 we compare our objective clusters to prior classification efforts by T18 for a ‘neutral’ climate state, and in section 5 we assess the changes to these clusters in the future. Finally, in section 6 we summarise our results and lay out a road map for applying this method to a variety of data sets and temporal periods to enable such a classification.

2. Model data: ACCESS-OM2-01

The primary tool used in this analysis is the ACCESS-OM2-01 model [see Kiss et al. (2019) for details]. ACCESS-OM2-01 is an ocean-sea ice model with 0.1° horizontal resolution and 75 vertical

99 z^* levels, composed of the MOM5.1 ocean model (Griffies 2012) coupled with the CICE5.1.2 sea
100 ice model (Hunke et al. 2015), and forced with the JRA55-do v1.3 atmospheric forcing product
101 (Tsujino et al. 2018). The ACCESS-OM2-01 model has been shown to accurately represent dense
102 shelf water formation and Antarctic shelf circulation processes (Morrison et al. 2020; Huneke et al.
103 2022). Further information on the bathymetric product and the characterisation of ice shelf and
104 melting processes in the model is provided in Kiss et al. (2019). We analyse temperature and
105 salinity in the ACCESS-OM2-01 model in two simulations, which are detailed below.

106 We analyse a *Repeat-Year Forcing* (RYF) simulation, in which the model is forced by 55 km and a
107 3-hourly temporal resolution JRA55-do v1.3 forcing cycling over the 1 May 1990 to 30 April 1991
108 time period. This year is characterised by having close to neutral conditions, that is, with dominant
109 modes of climate variability having little influence on the climate (Stewart et al. 2020). The RYF
110 simulation is spun up for 180 years, and we use monthly-averaged temperature and salinity from
111 the final 10 years of the simulation for our analysis.

112 We also analyse a *Future* perturbation simulation from Li et al. (2023). In this simulation,
113 the model is first spun up for 200 years with the RYF forcing, as above. The simulation is
114 then continued with perturbed anomalous wind, temperature and meltwater forcing fields. The
115 perturbed surface forcing fields are sourced from the historical JRA55-do v1.3 forcing field (from
116 1991-2019) and from CMIP6 future projections (from 2020-2050). In this analysis, we analyse the
117 monthly-averaged temperature and salinity from 2040-2050 in the future perturbation experiment.
118 Further details on the model set up and experimental design for this future run are provided in Li
119 et al. (2023). Note that in both the RYF and Future simulations model drift is minimal (Li et al.
120 2023).

121 3. Methods

122 In this section, we lay out the workflow used to obtain and interpret objective clusters of the
123 Antarctic margins. Specifically, we detail the a) choice and treatment of input data, b) classification
124 of input data using a clustering algorithm, and c) interpretation of objective clusters.

125 *a. Choice and treatment of input data*

126 Following T18, we cluster the full-depth temperature, T and salinity, S profiles in the Antarctic
 127 margins. We source data from the final ten years of the ACCESS-OM2-01 RYF simulation, and
 128 cluster the profiles of T and S as a function of latitude, y , longitude, x and depth, z . All profiles
 129 are time-averaged over the final ten years of the simulation.

130 A common issue with the application of tools such as Gaussian Mixture Modelling to ocean
 131 data near continental shelves is that it is not straightforward to compare deep and shallow profiles
 132 at constant depth or density when one profile is shallower than another. Deeper data for the
 133 shallow profile is effectively ‘missing’. Here, so that on-shelf and off-shelf profiles can be analysed
 134 equivalently and to avoid infilling, we interpolate the time-mean conservative temperature $\bar{T}(x, y, z)$
 135 and absolute salinity $\bar{S}(x, y, z)$ profiles onto a normalised, terrain-following depth, \hat{z} from 0 to 1.
 136 Figure 2 shows the result of this transformation onto terrain-following co-ordinates for the ‘fresh’
 137 latitudinal slice in figure 1. Both on-shelf profiles (in this case, where bathymetry, $h < 1000$ m) and
 138 off-shelf profiles (where bathymetry is $1000 < h < 3000$ m) are stretched in this new co-ordinate
 139 system.

144 A major objective of this study is to identify thermodynamic regimes in the spirit of T18. One
 145 of the fundamental controls on the regimes identified by T18 is the difference between T and S on
 146 and off the Antarctic shelf. For instance, in the fresh shelf regime, fresh water shoals the shelf, but
 147 is saltier off the shelf, while in the dense shelf regime, dense water overflows generally homogenise
 148 the salinity both on and off the shelf. To reflect the on and off shelf difference that motivates the
 149 T18 classifications, we average $\bar{T}(x, y, \hat{z})$ and $\bar{S}(x, y, \hat{z})$ profiles meridionally on- and off-shelf. This
 150 yields a series of on-shelf and off-shelf T and S profiles, defined as $\langle \bar{T}_{on} \rangle(x, \hat{z})$, $\langle \bar{S}_{on} \rangle(x, \hat{z})$,
 151 $\langle \bar{T}_{off} \rangle(x, \hat{z})$ and $\langle \bar{S}_{off} \rangle(x, \hat{z})$, respectively, as shown in figure 3. To ensure that changes
 152 in T are weighted equally to changes in S during clustering, we divide the normalised profiles
 153 $\langle \bar{T}_{on} \rangle(x, \hat{z})$, $\langle \bar{S}_{on} \rangle(x, \hat{z})$, $\langle \bar{T}_{off} \rangle(x, \hat{z})$ and $\langle \bar{S}_{off} \rangle(x, \hat{z})$ with their standard deviation
 154 over all longitudes and depths. These normalised profiles form the input data that is fed into the
 155 clustering algorithm to determine objective clusters of the Antarctic margins.

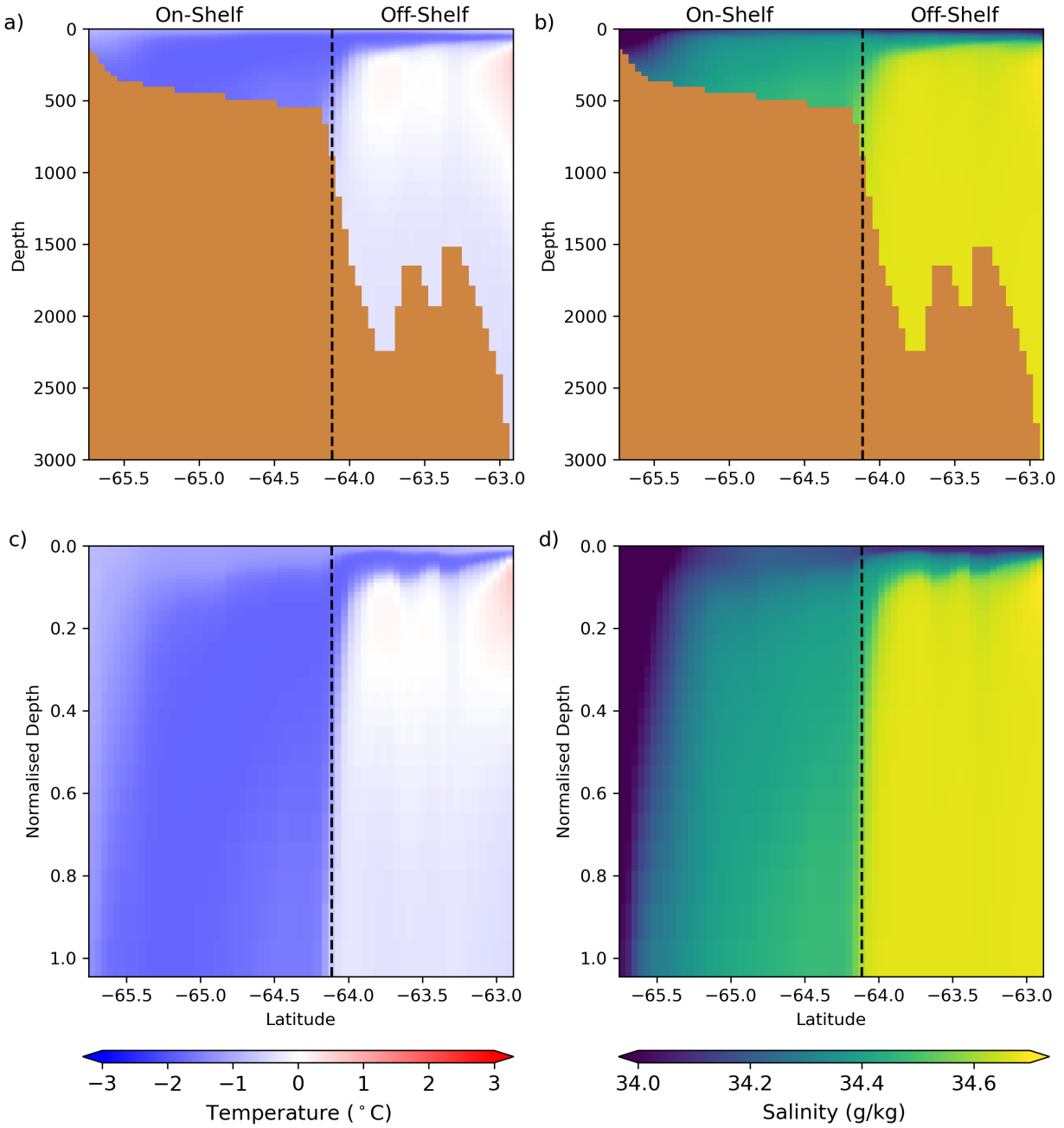


FIG. 2. a) Temperature and b) salinity in z co-ordinates for an exemplary ‘fresh’ shelf regime zonal slice (as shown in figure 1c). c) Temperature and d) salinity in normalised-depth \hat{z} co-ordinates for the same zonal slice. Dashed vertical line shows the latitude at the 1000 m isobath, which represents the boundary between the on-shelf and off-shelf region in this study.

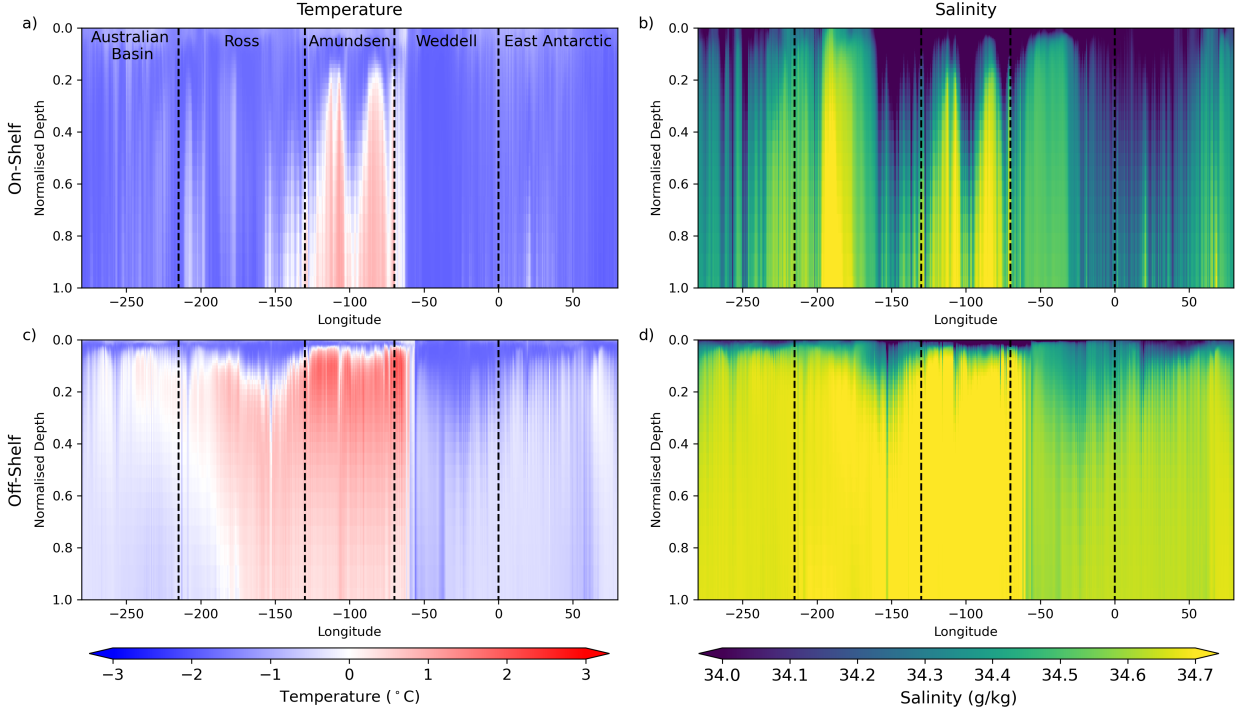


FIG. 3. Meridionally-averaged on-shelf a) temperature and b) salinity and meridionally-averaged off-shelf c) temperature and d) salinity in the ACCESS-OM2-01 RYF simulation. All profiles are shown in normalised depth co-ordinates.

b. Classification using Gaussian Mixture Modelling

The clustering algorithm chosen for this study is Gaussian Mixture Modelling (GMM). GMM is a probabilistic approach which fits the input data to a linear combination of k Gaussian functions, where k is the number of classes in the model. Below we summarise the governing equations of GMM analysis, as laid out previously from an oceanographic perspective in Jones et al. (2019). The Gaussian density function, \mathcal{N} is generally defined as:

$$\mathcal{N}(\mathbf{x}|\mu_k, \Sigma_k) = \frac{\exp\left(-\frac{1}{2\Sigma}(x - \mu_k)^T(x - \mu_k)\right)}{\sqrt{(2\pi)^D|\Sigma_k|}} \quad (1)$$

where \mathbf{x} is a given profile, μ_k is the mean of the profiles and Σ_k is the covariance matrix of the profiles. The probability that a profile \mathbf{x} belongs to a class k is given by:

$$p(k|\mathbf{x}) = \frac{\lambda_k \mathcal{N}(\mathbf{x}|\mu_k, \Sigma_k)}{\sum_{k=1}^K \lambda_k \mathcal{N}(\mathbf{x}|\mu_k, \Sigma_k)} \quad (2)$$

where λ_k is the probability distribution of a given class, k . GMM uses an *Expectation-Maximisation* algorithm to maximise the log of the probability density function $p(\mathbf{X})$ of all input profiles, \mathbf{X} . To achieve this, the GMM iteratively modifies the mean and covariance μ_k and Σ_k beginning from a random initial value until the log of equation (2) converges to a maximum for \mathbf{X} (Maze et al. 2017). In *Python*, the *sklearn.mixture.GaussianMixture* package is used to classify T and S profiles through the *gmm.fit_predict()* function. The GMM defines a set of fits that describe the input data well, and uses these fits to classify the data into a specified number of classes, K .

1) CHOOSING AN APPROPRIATE NUMBER OF CLUSTERS

While there are some subjective choices to be made as to how the GMM analysis is carried out, one of the most important is how many classes represent the input data well (without over-fitting onto the input data). In order to assess the appropriate number of clusters for our analysis, we rely on two quality tests - the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). AIC and BIC are defined as:

$$AIC(K) = 2K - 2\mathcal{L}(K), \& \quad (3)$$

$$BIC(K) = K\ln(n) - 2\mathcal{L}(K), \quad (4)$$

where $\mathcal{L}(K)$ is the log-likelihood of the model, K is the number of clusters and n is the sample size. In *Python*, the *gmm.aic()* and *gmm.bic()* functions in the *sklearn.mixture.GaussianMixture* package natively output AIC and BIC for a given model and number of clusters. Generally, the lower the AIC and BIC, the better the model fit is to the input data. BIC tends to penalise model complexity more than AIC due to the addition of the $K\ln(n)$ term that depends on the sample size.

186 *c. Interpretation of classes*

187 The GMM classifies the input data into a set of discrete classes with an associated probability.
188 The final step in the analysis is to assess these classes to ensure they are physically meaningful,
189 and to track changes to these classes over time. Learnings from this step feed back into the choice
190 and treatment of input data and the classification conducted using GMM.

191 **4. Classification of RYF simulation**

192 The first objective of this study is to use GMM to determine the thermodynamic regimes in
193 a ‘neutral’ ocean, as simulated in the RYF case for ACCESS-OM2-01. We feed the time- and
194 spatially-averaged input T and S data (described in section a) into the GMM algorithm to determine
195 the appropriate number of clusters. Figure 4 shows the AIC and BIC (calculated from equation [3])
196 for the input data. AIC suggests the optimal number of clusters is twelve, while BIC determines the
197 optimal number of clusters as three. While three clusters is a clear winner for BIC, AIC flattens out
198 between nine and fifteen clusters, indicating that number in this range could be suitable. However,
199 as mentioned in section 3, BIC penalises model complexity more harshly than AIC. For our
200 purposes, BIC is a more preferable indicator as the simpler model enables a simpler framework for
201 understanding and communicating changes in the Antarctic margins. Therefore, for the remainder
202 of this analysis, we will focus on three classes. A further exploration of what these three clusters
203 represent is provided below, alongside a comparison of what the clusters would look like if two or
204 four classes were selected instead.

207 In T18 and Moorman et al. (2020), the shelf regimes are spatially well-delineated, with the warm
208 regime isolated to West Antarctica, and the dense regime in the Weddell and Ross seas. Our GMM
209 results align well with the T18 and Moorman et al. (2020) regimes, indicating that the GMM is
210 able to objectively isolate the relevant regimes well (see figure 5). Indeed, Moorman et al. (2020)
211 performed the regime analysis on the same RYF model simulation, and the results are remarkably
212 similar (compare figure 5 to the control regimes in figure 4a in Moorman et al. (2020)). Note,
213 however, that the GMM tends to classify broader dense shelf regions in the East Antarctic when
214 compared with the Moorman et al. (2020) analysis.

218 The labelling of each cluster into warm, dense and fresh regimes is done through a qualitative
219 analysis of the T and S properties in each class. Figure 6 shows the T and S profiles in each shelf

Determination of optimal number of clusters

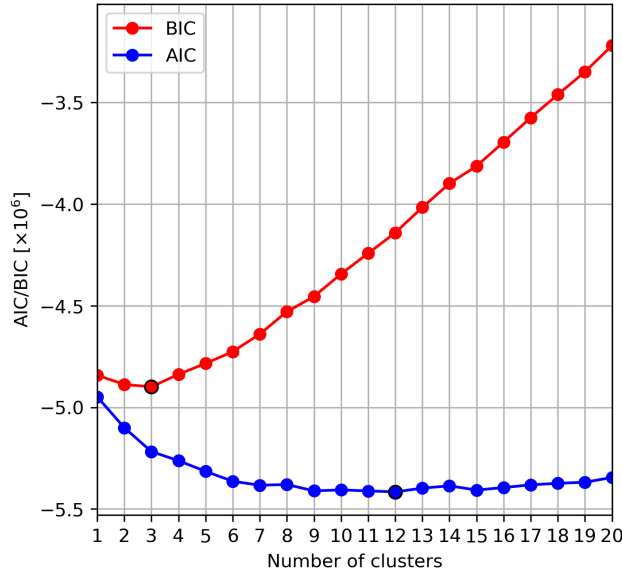


FIG. 4. Indicators of the optimal number of clusters in the time-mean RYF simulation based on the AIC and BIC metrics. Markers with black edges are the lowest AIC and BIC for the range of clusters explored here.

regime (grey) and the mean T and S profile for each class (blue, orange and green). The fresh profiles show the characteristically cold and fresh water on the shelf (solid blue lines) compared to the warmer and saltier off-shelf region. The warm shelf profiles (orange lines) highlight the warm CDW on the shelf (see temperature profile where $\hat{z} > 0.1$). The dense shelf profiles (green lines) have the smallest difference between the on-shelf salinity and the off-shelf salinity, indicating a higher rate of dense shelf water overflows. Based on this analysis, we classify the three classes as fresh, warm and dense.

One way to assess cluster properties over time is to track the bottom on- and off-shelf T and S in each shelf regime (see figure 7). The thermohaline regimes explored here are characterised by fresh, dense or warm water flowing onto the shelf region, as shown in figure 1. Therefore, the bottom properties in the Antarctic margins highlight key differences between different Antarctic shelf regimes and are a good metric to track over time. The bottom temperature, salinity and T and S difference are well differentiated in the classes identified in the RYF simulation (figure 7). Bottom temperature tends to be warmest (both on- and off-shelf) in the warm shelf regime (orange cluster in figure 7a) with fresh and dense bottom temperatures matching each other relatively

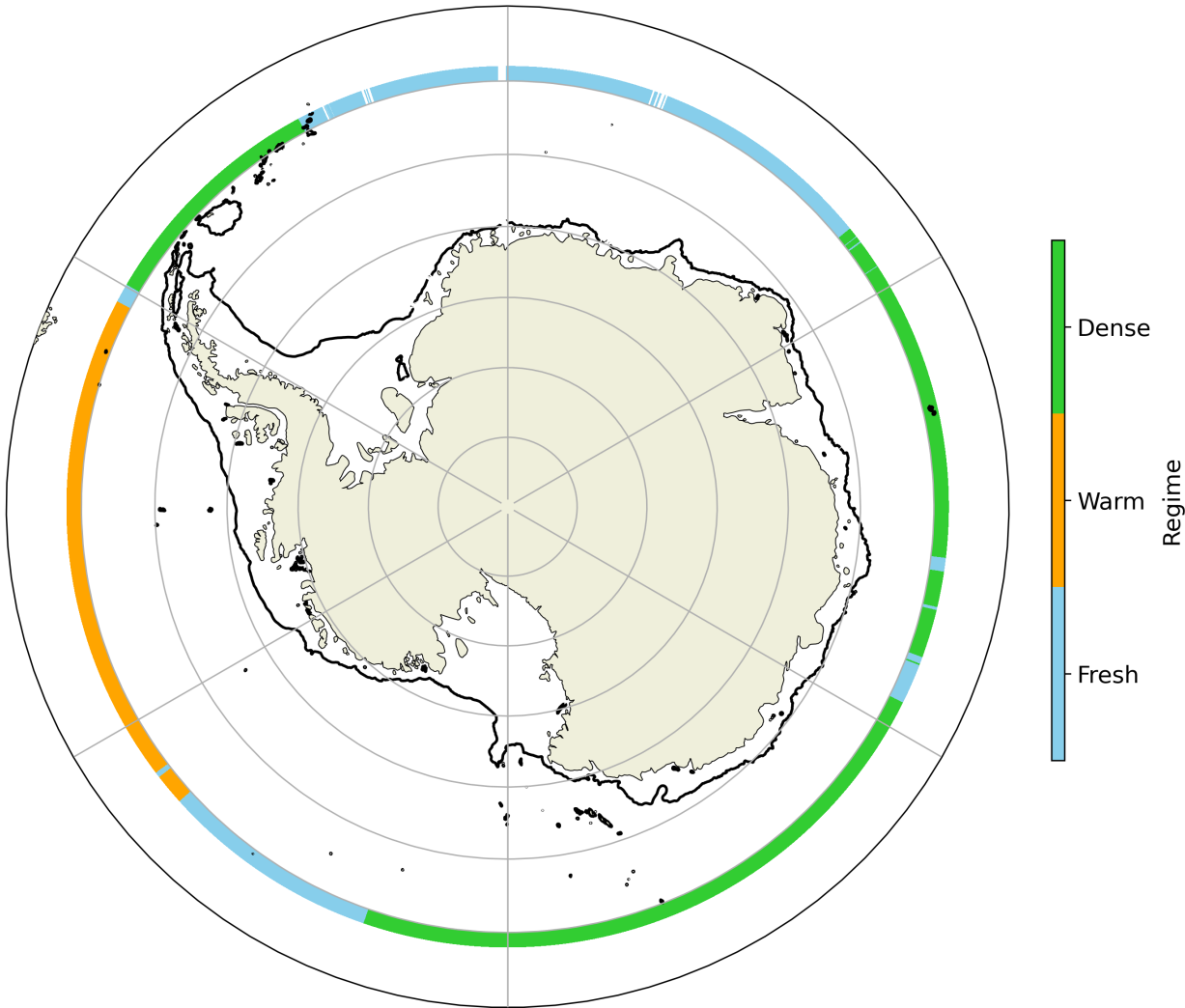


FIG. 5. Objectively classified shelf regimes in the ACCESS-OM2-01 RYF simulation as a function of longitude. Longitudes where no class is identified (shown in white) have no ‘on-shelf’ region (i.e., there are no isobaths shallower than 1000 m). The black contour line shows the 1000 m isobath.

closeby. Indeed, the difference between fresh and dense shelf waters appears when looking at on- and off-shelf salinity, in figure 7b. Generally speaking, the dense shelf regime has higher on- and off-shelf bottom salinity compared with the fresh regime. Dense shelf water formation regions are

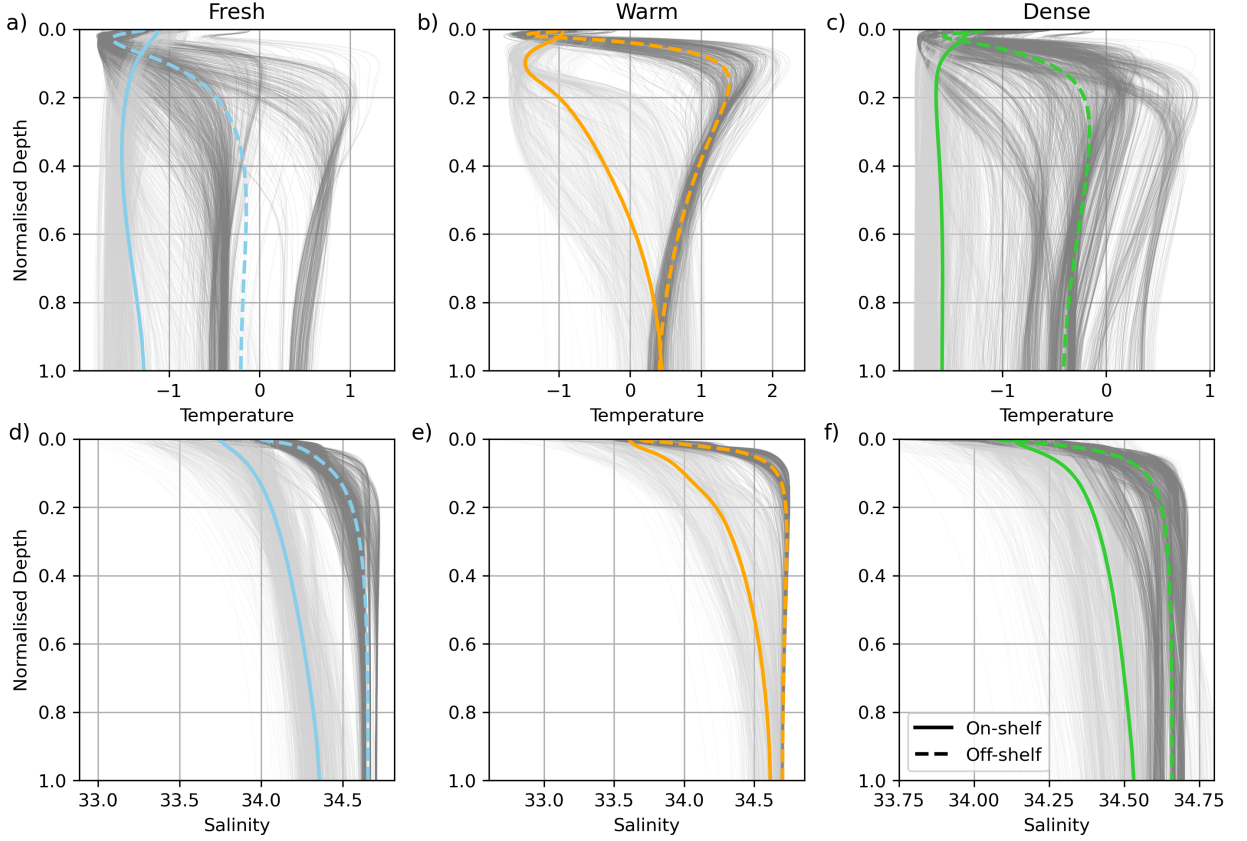


FIG. 6. Meridionally-averaged on- and off-shelf temperature and salinity profiles in normalised depth coordinates grouped by their shelf regime. On-shelf profiles are shown in light grey, and off-shelf profiles are plotted in dark grey. The on-shelf mean profiles for a given regime are shown in thick solid lines and off-shelf mean profiles are shown in thick dashed lines. Blue, orange and green lines correspond to the means for the fresh, warm and dense regimes, respectively.

also highlighted which have higher on-shelf bottom salinity, but have off-shelf bottom salinities that match the fresh shelf regime.

The difference between on- and off-shelf bottom T and S (figure 7c) clearly shows the difference between the dense, fresh and warm regimes identified by GMM. Bottom properties in the warm shelf regime tend to align along a single $\Theta_{off} - \Theta_{on}/S_{off} - S_{on}$ slope, and in general on-shelf bottom salinity is lower than off-shelf bottom salinity. Bottom properties in the fresh shelf regime are more broadly distributed, and reflect the shoaling of cold, fresh water on the shelf; on-shelf bottom temperature and salinity are lower than their off-shelf counterparts. In the dense shelf regime, dense water formation is evident as some profiles have an on-shelf bottom salinity that is

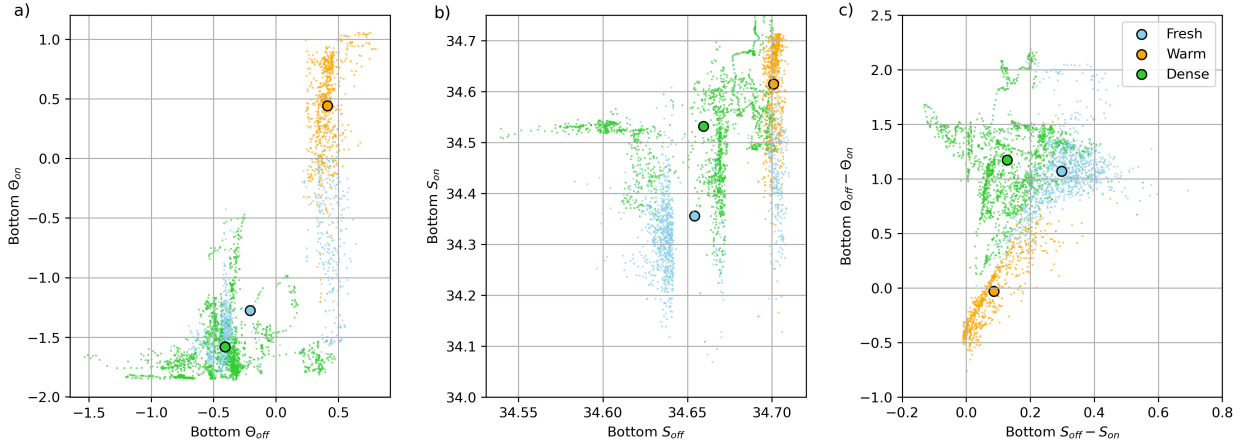


FIG. 7. a) Bottom on-shelf vs off-shelf temperature, b) salinity, and c) the difference between on- and off-shelf bottom temperature and on- and off shelf bottom salinity, grouped by shelf regime in the ACCESS-OM2-01 RYF simulation. Large coloured markers show the mean bottom properties for each regime.

greater than the off-shelf salinity, and on-shelf bottom temperatures are colder than off-shelf bottom temperatures. The difference between on- and off-shelf bottom salinity is also smaller in the dense shelf regions compared with the fresh shelf regime. This is because dense shelf water mixes with relatively salty circumpolar deep water as it overflows, increasing off-shelf bottom salinity but not to the extent of the fresh shelf regions.

We now compare the bottom properties in our ‘optimal’ (according to BIC) case where three classes are selected to the bottom properties corresponding to the case where two or four clusters are selected, shown in figure 8. When directed to classify two distinct classes from the input data, the GMM algorithm opts to separate the warm, West Antarctic region from the rest of the Antarctic margins, lumping together fresh and dense shelf water formation regions (figure 8a). The GMM defines the fourth class (shown in purple in figure 8c) by combining some profiles from the fresh and warm shelf regimes along the $\Theta_{off} - \Theta_{on} / S_{off} - S_{on}$ slope discussed earlier. Notably, there is very little overlap between points belonging to different classes in the two or three class case (figure 8a and b), but with the addition of a fourth class there is now substantial overlap between classes one, three or four. Class four occupies much of the warm shelf region as well as the fresh region in the Ross Sea (not shown), but there does not appear to be an insightful or physically relevant meaning to this region based on a qualitative assessment. These results add further credence to our choice of three as the appropriate number of regimes in the Antarctic margins.

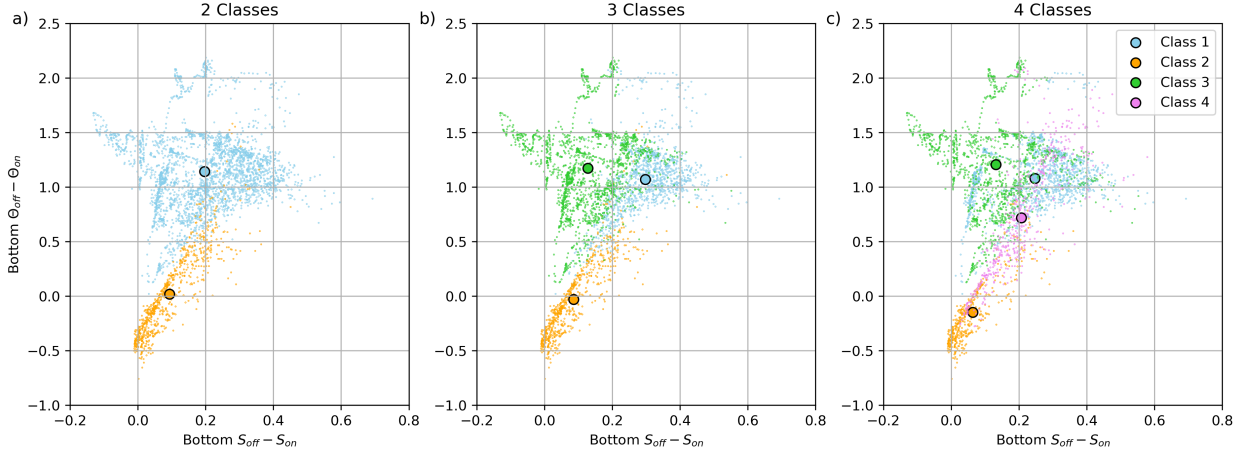


FIG. 8. As in figure 7c above, but for a) two classes or b) four classes obtained from the GMM output.

5. Future trends in shelf regimes

A key advantage of using an objective classification technique is the ability to track changes to thermohaline classes in space and time. Below we assess the changes to the warm, fresh and dense regimes (as classified using the RYF simulation) in the final 10 years of a future perturbation simulation of the ACCESS-OM2-01 model, from Li et al. (2023). In order to conduct this analysis, we maintain the same geographical definitions for the three classes (as determined in the RYF simulation and shown in figure 5), and simply quantify the changes to these geographical regions in the future simulation. Bottom temperatures increase for all regimes in the future, with dense regions experiencing the largest off-shelf warming relative to on-shelf temperatures, and warm regions experiencing approximately equal on- and off-shelf warming (see figure 9a). The broad-scale warming across the Antarctic margins reflects projected global ocean warming due to climate change, and is particularly strong in ventilating (i.e., dense shelf) regions. The on-shelf salinity decreases for all classes, and is particularly strong for the dense shelf regions (figure 9b). This is likely due to the strong influx of surface freshwater due to ice melting.

The difference between on- and off-shelf bottom T and S (figure 9c) shows the merging of the dense and fresh shelf regimes into a single cluster of bottom T and S in future simulations (see movement of light green and light blue centroids to dark green and dark blue centroids). This result reveals the disappearance of the dense shelf water class in the future, and aligns with the key findings of Li et al. (2023) which saw a slowdown in the formation of bottom water (the successor of dense shelf water). Overall, the bottom T and S appear to align more closely to a

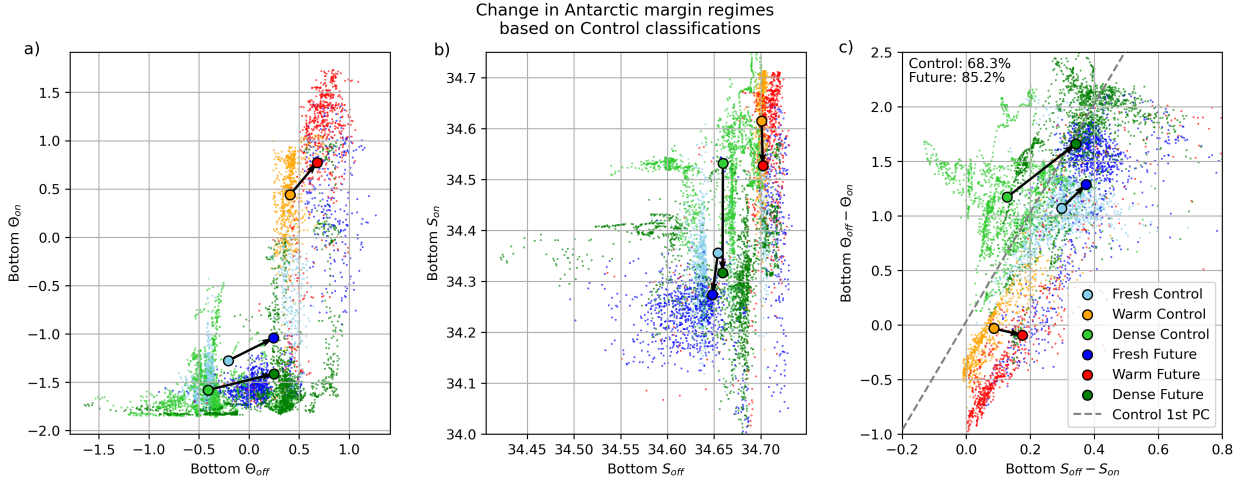


FIG. 9. As in figure 7 above, with the addition of the time-mean properties in the final ten years of the ACCESS-OM2-01 future projection, grouped into their respective regimes (as obtained from the RYF classification step). The future fresh, warm and dense regimes are plotted in dark blue, red and dark green, respectively. Arrows show the change in the mean properties from the RYF simulation to the future projection. Dashed grey line in c) shows the first principal component of the RYF simulation in phase space, and the proportion of variance explained by the first principal component is provided for both the RYF and future simulations.

single $\Theta_{off} - \Theta_{on} / S_{off} - S_{on}$ slope. In order to quantify this increased alignment, we plot the first principal component of the bottom clusters in the RYF simulation (dashed line in figure 9c), which appears to correspond to the $\Theta_{off} - \Theta_{on} / S_{off} - S_{on}$ slope along which the bottom properties orient. In the RYF simulation, the first principal component explains 68% of the total variance in the scatter plot. However, in the future, this number rises to 85%, illustrating that the plot has become more aligned to a single line with slope $\Theta_{off} - \Theta_{on} / S_{off} - S_{on}$.

Changes to bottom T and S indicate a collapse in dense shelf water formation in the future perturbation experiments. The T and S properties averaged along isobaths illustrate this collapse further, as shown in figure 10 and 11. In the RYF simulation, the temperature contours show the shoaling of the warm (figure 10a) and cold, fresh (figure 10b) waters in the warm and fresh regimes, respectively. The overflowing cold, salty dense waters are also evident in the temperature contours in the dense regime, which align with the continental shelf below the 1000 m isobath (figure 10c). In the future simulation, the temperature in the warm shelf is significantly enhanced due to ocean warming (figure 10d), and the shoaling cold, fresh water deepens substantially due to the influx of

313 meltwater (figure 10e). The overflowing cold, salty dense shelf water is no longer present, with the
314 dense shelf regions resembling the fresh shelf regime (compare figure 10e and f). These results
315 further illustrate the collapse of dense shelf water formation in the future, and the merging of the
316 fresh and dense shelf regimes into one.

322 A similar picture emerges for the salinity averaged along isobaths (albeit with smaller salinity
323 variations between the RYF and future simulations; figure 11). The shoaling of the salty Circumpo-
324 lar Deep Water is evident in the RYF simulation in figure 11a. Interestingly, the freshening signal
325 due to the meltwater perturbation is more strongly visible in the salinity field (figure 11d) than
326 the temperature field (figure 10d). This is because the meltwater is added at the local temperature
327 upon entry at the surface of the ocean, which means it is more likely to be seen in the salinity field
328 if it does not cause significant circulation changes. The thickening of the fresh layer in the fresh
329 regime is also shown in figure 11e, and the transformation of the dense shelf regime into something
330 resembling the fresh shelf regime is also visible in figure 11f. Overall, the strong ocean freshening
331 across the Antarctic margins is shown in the anomaly plots in figure 11g - i.

335 6. Discussion & Conclusions

336 This work presents an objective way to characterise oceanic regimes in the Antarctic shelf region
337 using unsupervised clustering methods. We introduce a workflow that applies *Gaussian Mixture*
338 *Modelling* (GMM) to modelled profiles of shelf temperature and salinity and groups the complex
339 Antarctic shelf region into *dense*, *fresh* and *warm* shelf regimes, in line with Thompson et al. (2018).
340 The objective nature of our workflow enables us identify regimes around the Antarctic margins,
341 compare these regimes with prior (subjective) methods, and track changes to these regimes in
342 future projections. Using a ‘neutral’ ocean model (i.e., with minimal influence of modes of climate
343 variability on the ocean state), the Bayesian Information Criterion metric indicates that the optimal
344 number of oceanic regimes in the Antarctic margins is three, in alignment with Thompson et al.
345 (2018). An assessment of the temperature and salinity profiles corresponding to these regimes
346 shows that the GMM suitably partitions the ocean into fresh, warm and dense regimes, but without
347 a subjective set of criteria that changes based on the data being analysed. Finally, upon obtaining an
348 objectively partitioned set of clusters in the Antarctic margins, we track the changes to these clusters
349 in a future ocean projection (from Li et al. (2023)). Over time, the dense and fresh shelf regimes

Temperature in thermodynamic regimes in ACCESS-OM2-01

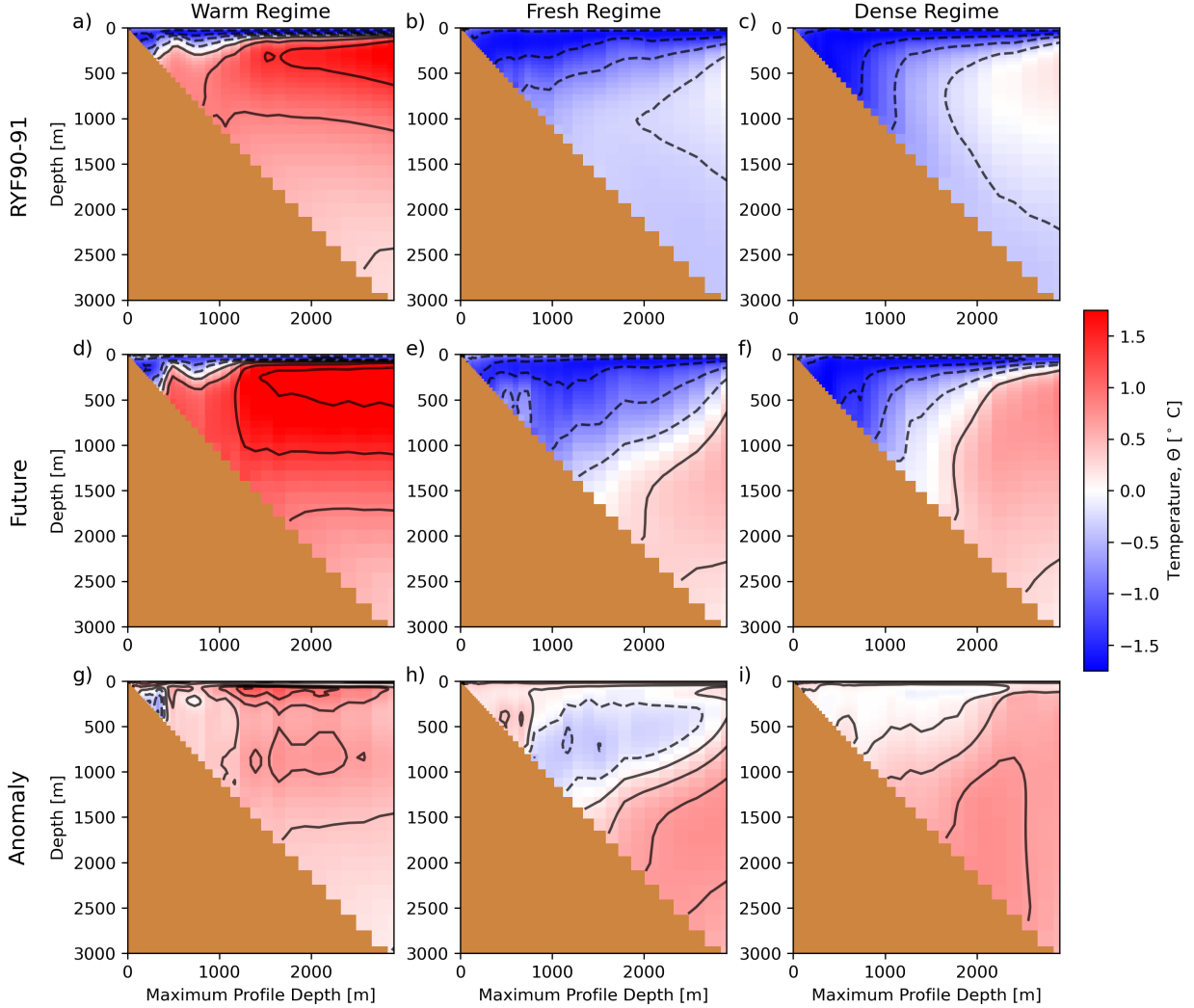


FIG. 10. Temperature averaged along isobaths around the Antarctic margins, grouped into the regimes identified by GMM, in a - c) the ACCESS-OM2-01 RYF simulation, d - f) the ACCESS-OM2-01 future projection. g - i) Shows the difference between the RYF and future simulations. In a - f), the black contour lines show 8 evenly spaced isotherms between 2 and -2°C, and in g - i) the contour lines show 16 evenly spaced isotherms between 2 and -2°C. Negative values are dashed and positive values are solid.

merge to form a single fresh regime which is present everywhere except the Ross, Amundsen and Bellingshausen seas due to the strong ice melt-driven meltwater flux into the ocean. In the Ross,

Salinity in thermodynamic regimes in ACCESS-OM2-01

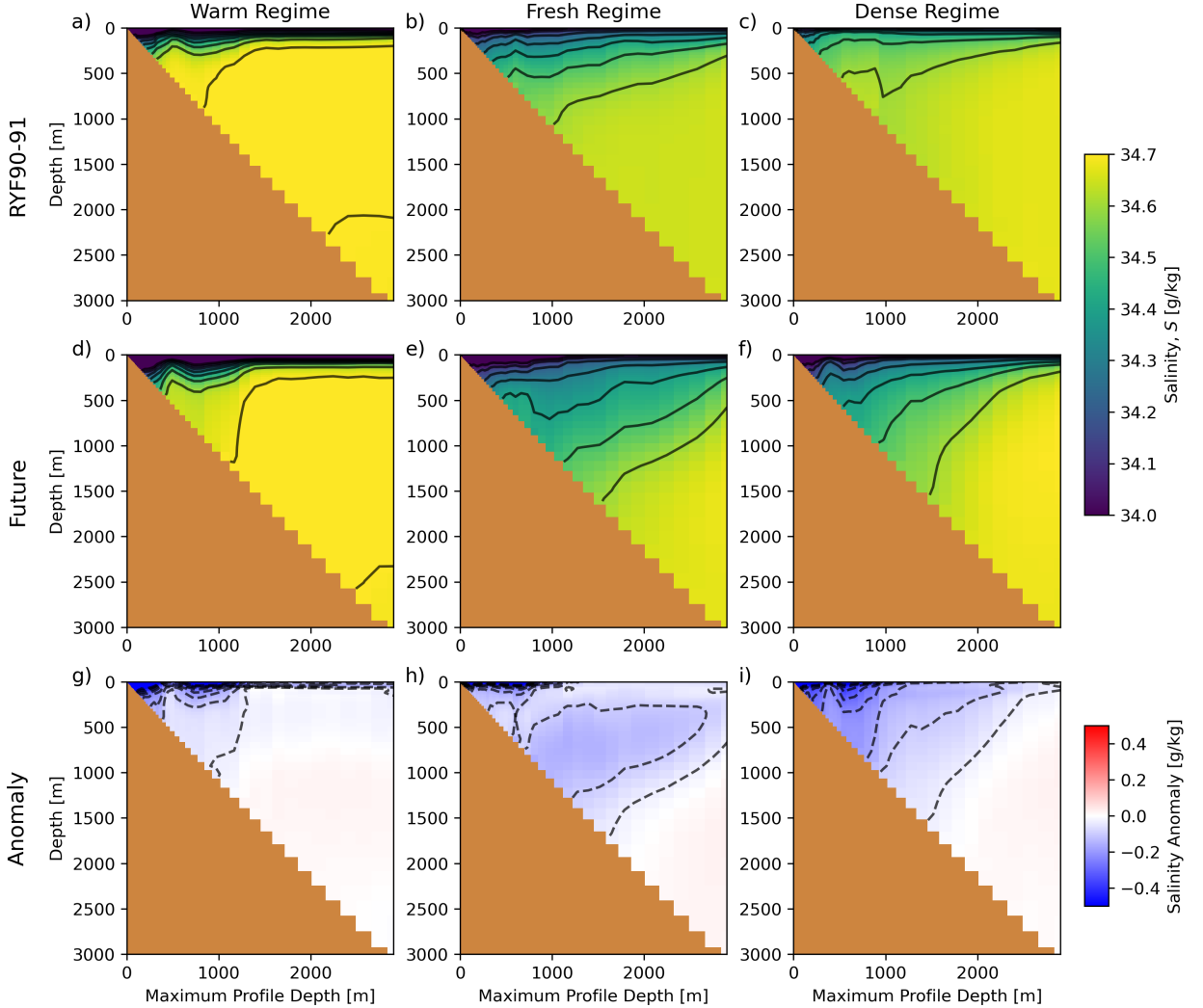


FIG. 11. As in figure 10 but for salinity. In a - f), the black contour lines show 8 evenly spaced isohalines between 34 and 34.7 g/kg, and in g - i) the contour lines show 16 evenly spaced isohalines between -0.5 and 0.5 g/kg. Negative values are dashed and positive values are solid.

Amundsen and Bellingshausen seas, the warm regime experiences stronger warming, with warmer Circumpolar Deep Water shoaling onto the shelf over time.

The transition of the Antarctic shelf region to a simpler two-regime system is evident upon repeating our method of classification (using GMM) for individual years in the future projection. Figure 12 shows the BIC over selected years in the future projection simulation, as calculated from

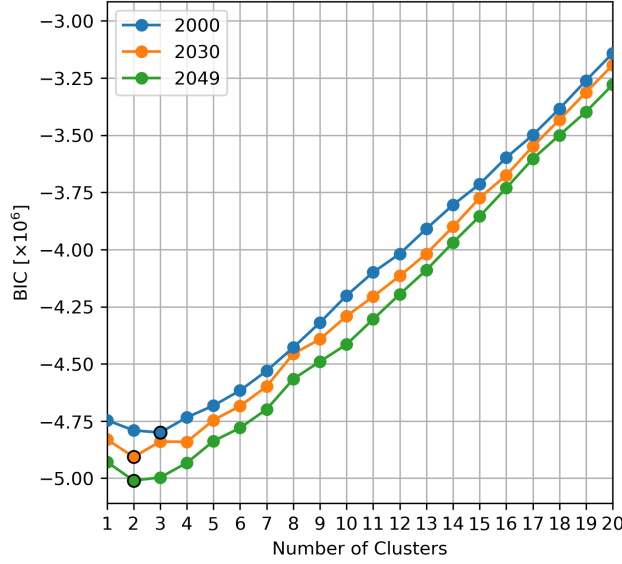


FIG. 12. Bayesian Information Criterion (BIC) in select years in the ACCESS-OM2-01 future projection for 1 to 20 clusters. The lowest BIC is highlighted in each year by a marker with a black edge. Note that the BIC is calculated by performing the classification step using the annual-mean temperature and salinity in the given year of the future projection.

equation (3). The BIC points to the optimal number of clusters being two by 2030 in the meltwater, thermal and wind perturbation simulations.

While we strive to remove as much subjectivity as possible from the clustering process in our workflow, some free parameters remain that can change the outputs of the clustering algorithm. First, the choice and treatment of input data is subjective, and here we choose on- and off-shelf temperature and salinity to obtain clusters that align with Thompson et al. (2018). Additional input data, such as the latitude or longitude of each profile, or process information, such as wind or current speeds, could be used. Second, different combinations of input data can be used for the classification step. In this work, we conduct the classification using the RYF simulation (or in the case of figure 12 above, using the future projection) as input data. Instead, we could calculate fits using GMM applied to the RYF simulation and apply these fits to the future projection data, amongst other combinations. Finally, the optimal number of classes is somewhat subjective. We are guided by metrics AIC and BIC and existing analysis to arrive at three as an appropriate number

374 of clusters. Alternative criteria, relevant to the specific problem approached by the user, could also
375 guide the number of clusters.

376 Our results provide a road-map for similar clustering analyses to be applied to the Antarctic
377 margins in the future. This unsupervised learning analysis provides the basis for future idealised
378 studies of the Antarctic margins (e.g., Stewart and Thompson (2015)). In addition, this knowledge
379 of similar regions around the Antarctic shelf will help to infill scarce observations to produce
380 a gridded observational product. Overall, our clustering analysis provides an important frame-
381 work for understanding dominant processes, model biases and future changes in a complex and
382 heterogeneous Antarctic.

383 *Acknowledgments.* The authors gratefully acknowledge Matthew England and Adele Morrison
384 for fruitful discussions about the model data and clustering workflow employed in the study.
385 We also acknowledge the Consortium of Ocean and Sea Ice Modelling in Australia (COSIMA)
386 and the authors of Li et al. (2023) for furnishing and hosting the ACCESS-OM2-01 RYF and
387 future projection simulations. The authors were funded by the Australian Center for Excellence in
388 Antarctic Science (ACEAS; grant number SR200100008).

389 *Data availability statement.* The raw model output is extremely large and we are unable to
390 share this using traditional public repositories. However, the input data used for the classifica-
391 tion in the GMM is archived for both the RYF simulation and the future projection on *Zenodo*
392 (doi:10.5281/zenodo.8206990).

393 **References**

- 394 Boehme, L., and I. Rosso, 2021: Classifying Oceanographic Structures in the Amundsen Sea,
395 Antarctica. *Geophysical Research Letters*, **48** (5), <https://doi.org/10.1029/2020gl089412>.
- 396 Daae, K., T. Hattermann, E. Darelius, and I. Fer, 2017: On the effect of topography and wind
397 on warm water inflow—An idealized study of the southern Weddell Sea continental shelf
398 system. *Journal of Geophysical Research: Oceans*, **122** (3), 2622–2641, [https://doi.org/10.](https://doi.org/10.1002/2016jc012541)
399 [1002/2016jc012541](https://doi.org/10.1002/2016jc012541).

400 Darelus, E., and Coauthors, 2023: Observational evidence for on-shelf heat transport driven by
 401 dense water export in the Weddell Sea. *Nature Communications*, **14** (1), 1022, [https://doi.org/](https://doi.org/10.1038/s41467-023-36580-3)
 402 10.1038/s41467-023-36580-3.

403 Griffies, S. M., 2012: Elements of the Modular Ocean Model (MOM). *GFDL Ocean Group Tech.*
 404 *Rep.*, **7**, 620.

405 Herraiz-Borreguero, L., and A. C. N. Garabato, 2022: Poleward shift of Circumpolar Deep Water
 406 threatens the East Antarctic Ice Sheet. *Nature Climate Change*, **12** (8), 728–734, [https://doi.org/](https://doi.org/10.1038/s41558-022-01424-3)
 407 10.1038/s41558-022-01424-3.

408 Huguenin, M. F., R. M. Holmes, and M. H. England, 2022: Drivers and distribution of global ocean
 409 heat uptake over the last half century. *Nature Communications*, **13** (1), 4921, [https://doi.org/](https://doi.org/10.1038/s41467-022-32540-5)
 410 10.1038/s41467-022-32540-5.

411 Huneke, W. G. C., A. K. Morrison, and A. M. Hogg, 2022: Spatial and Subannual Variabil-
 412 ity of the Antarctic Slope Current in an Eddying Ocean–Sea Ice Model. *Journal of Physical*
 413 *Oceanography*, **52** (3), 347–361, <https://doi.org/10.1175/jpo-d-21-0143.1>.

414 Hunke, E., W. Lipscomb, P. Jones, A. Turner, N. Jeffery, and S. Elliott, 2015: CICE, the Los
 415 Alamos sea ice model. Tech. rep., Los Alamos National Lab.(LANL), Los Alamos, NM (United
 416 States).

417 Jones, D. C., H. J. Holt, A. J. S. Meijers, and E. Shuckburgh, 2019: Unsupervised Clustering of
 418 Southern Ocean Argo Float Temperature Profiles. *Journal of Geophysical Research: Oceans*,
 419 **124** (1), 390–402, <https://doi.org/10.1029/2018jc014629>.

420 Jones, D. C., and Coauthors, 2023: Unsupervised classification identifies coherent thermohaline
 421 structures in the Weddell Gyre region. *Ocean Science*, **19** (3), 857–885, [https://doi.org/10.5194/](https://doi.org/10.5194/os-19-857-2023)
 422 os-19-857-2023.

423 Kiss, A. E., and Coauthors, 2019: ACCESS-OM2 v1.0: a global ocean–sea ice model
 424 at three resolutions. *Geoscientific Model Development*, **13** (2), 401–442, [https://doi.org/](https://doi.org/10.5194/gmd-13-401-2020)
 425 10.5194/gmd-13-401-2020.

Li, Q., M. H. England, A. M. Hogg, S. R. Rintoul, and A. K. Morrison, 2023: Abyssal ocean overturning slowdown and warming driven by Antarctic meltwater. *Nature*, **615** (7954), 841–847, <https://doi.org/10.1038/s41586-023-05762-w>.

Maze, G., H. Mercier, R. Fablet, P. Tandeo, M. L. Radcenco, P. Lenca, C. Feucher, and C. L. Goff, 2017: Coherent heat patterns revealed by unsupervised classification of Argo temperature profiles in the North Atlantic Ocean. *Progress in Oceanography*, **151**, 275–292, <https://doi.org/10.1016/j.pocean.2016.12.008>.

Moorman, R., A. K. Morrison, and A. M. Hogg, 2020: Thermal responses to Antarctic ice shelf melt in an eddy rich global ocean–sea-ice model Thermal responses to Antarctic ice shelf melt in an eddy rich global ocean–sea-ice model. *Journal of Climate*, **33** (15), 6599–6620, <https://doi.org/10.1175/jcli-d-19-0846.1>.

Morrison, A. K., A. M. Hogg, M. H. England, and P. Spence, 2020: Warm Circumpolar Deep Water transport toward Antarctica driven by local dense water export in canyons. *Science Advances*, **6** (18), eaav2516, <https://doi.org/10.1126/sciadv.aav2516>.

Neme, J., M. H. England, and A. M. Hogg, 2022: Projected Changes of Surface Winds Over the Antarctic Continental Margin. *Geophysical Research Letters*, **49** (16), <https://doi.org/10.1029/2022gl098820>.

Oppenheimer, M., and Coauthors, 2019: Sea Level Rise and Implications for Low-Lying Islands, Coasts and Communities. Tech. rep., Cambridge University Press, Cambridge, UK and New York, NY, USA, 321–445 pp. URL <https://doi.org/10.1017/9781009157964.006>.

Pauthenet, E., J.-B. Sallée, S. Schmidtko, and D. Nerini, 2021: Seasonal Variation of the Antarctic Slope Front Occurrence and Position Estimated from an Interpolated Hydrographic Climatology. *Journal of Physical Oceanography*, **51** (5), 1539–1557, <https://doi.org/10.1175/jpo-d-20-0186.1>.

Rosso, I., M. R. Mazloff, L. D. Talley, S. G. Purkey, N. M. Freeman, and G. Maze, 2020: Water Mass and Biogeochemical Variability in the Kerguelen Sector of the Southern Ocean: A Machine Learning Approach for a Mixing Hot Spot. *Journal of Geophysical Research: Oceans*, **125** (3), <https://doi.org/10.1029/2019jc015877>.

- 454 Sohail, T., R. M. Holmes, and J. D. Zika, 2023: Watermass Co-Ordinates Isolate the Historical
455 Ocean Warming Signal. *Journal of Climate*, 1–40, <https://doi.org/10.1175/jcli-d-22-0363.1>.
- 456 Stewart, A. L., and A. F. Thompson, 2015: Eddy-mediated transport of warm Circumpolar
457 Deep Water across the Antarctic Shelf Break. *Geophysical Research Letters*, **42** (2), 432–440,
458 <https://doi.org/10.1002/2014gl062281>.
- 459 Stewart, K., and Coauthors, 2020: JRA55-do-based repeat year forcing datasets for driving
460 ocean–sea-ice models. *Ocean Modelling*, **147**, 101–157, [https://doi.org/10.1016/j.ocemod.2019.](https://doi.org/10.1016/j.ocemod.2019.101557)
461 101557.
- 462 Sun, Q., C. M. Little, A. M. Barthel, and L. Padman, 2020: A clustering-based approach to ocean
463 model–data comparison around Antarctica. *Ocean Science*, **17** (1), 131–145, [https://doi.org/](https://doi.org/10.5194/os-17-131-2021)
464 10.5194/os-17-131-2021.
- 465 Thompson, A. F., A. L. Stewart, P. Spence, and K. J. Heywood, 2018: The Antarctic Slope
466 Current in a Changing Climate. *Reviews of Geophysics*, **56** (4), 741–770, [https://doi.org/10.](https://doi.org/10.1029/2018rg000624)
467 1029/2018rg000624.
- 468 Tsujino, H., and Coauthors, 2018: JRA-55 based surface dataset for driving ocean–sea-ice models
469 (JRA55-do). *Ocean Modelling*, **130**, 79–139, <https://doi.org/10.1016/j.ocemod.2018.07.002>.