



HAL
open science

Representation of Southern Ocean Properties across Coupled Model Intercomparison Project Generations: CMIP3 to CMIP6

Rebecca L Beadling, J. L Russell, R. J Stouffer, M. Mazloff, L. Talley, P. J
Goodman, Jean-Baptiste Sallée, H. T Hewitt, P. Hyder, Amarjiit Pandde

► **To cite this version:**

Rebecca L Beadling, J. L Russell, R. J Stouffer, M. Mazloff, L. Talley, et al.. Representation of Southern Ocean Properties across Coupled Model Intercomparison Project Generations: CMIP3 to CMIP6. *Journal of Climate*, 2020, 33 (15), pp.6555-6581. 10.1175/JCLI-D-19-0970.1 . hal-03045644

HAL Id: hal-03045644

<https://hal.science/hal-03045644v1>

Submitted on 8 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



1 **Representation of Southern Ocean properties across Coupled Model Intercomparison Project**
2 **generations: CMIP3 to CMIP6**

3
4 **R. L. Beadling^a, J. L. Russell^a, R. J. Stouffer^a, M. Mazloff^b, L. D. Talley^b, P. J. Goodman^a, J. B.**
5 **Sallée^c, H. T. Hewitt^d, P. Hyder^d, and Amarjiit Pandde^a**

6
7 ^a Department of Geosciences, The University of Arizona, Tucson, Arizona, USA

8 ^b Scripps Institution of Oceanography, University of California, San Diego, California, USA

9 ^c Sorbonne Université, UPMC Univ., Paris 06, UMR 7159, LOCEAN-IPSL, F-75005, Paris, France

10 ^d Met Office Hadley Centre, FitzRoy Road, Exeter, UK

11

12 Corresponding author address: Rebecca L. Beadling, Department of Geosciences, The University of
13 Arizona, 1040 E. 4th Street, Tucson, AZ 85721.

14 E-mail: Beadling@email.arizona.edu

15 **Abstract**

16 The air-sea exchange of heat and carbon in the Southern Ocean (SO) plays an important role in
17 mediating the climate state. The dominant role the SO plays in storing anthropogenic heat and carbon is
18 a direct consequence of the unique and complex ocean circulation that exists there. Previous
19 generations of climate models have struggled to accurately represent key SO properties and processes
20 that influence the large-scale ocean circulation. This has resulted in low confidence ascribed to 21st
21 century projections of the state of the SO from previous generations of models. This analysis provides a
22 detailed assessment of the ability of models contributed to the sixth phase of the Coupled Model
23 Intercomparison Project (CMIP6) to represent important observationally-based SO properties.
24 Additionally, a comprehensive overview of CMIP6 performance relative to CMIP3 and CMIP5 is
25 presented. CMIP6 models show improved performance in the surface wind stress forcing, simulating
26 stronger and less equatorward-biased wind fields, translating into an improved representation of the
27 Ekman upwelling over the Drake Passage latitudes. An increased number of models simulate an
28 Antarctic Circumpolar Current (ACC) transport within observational uncertainty relative to previous
29 generations, however, several models exhibit extremely weak transports. Generally, the upper SO
30 remains biased warm and fresh relative to observations, and Antarctic sea ice extent remains poorly
31 represented. While generational improvement is found in many metrics, persistent systematic biases are
32 highlighted that should be a priority during model development. These biases need to be considered
33 when interpreting projected trends or biogeochemical properties in this region.

34 **Introduction.**

35 The sequestration and ventilation of heat and carbon that occurs in the Southern Ocean (SO)
36 plays a key role in global climate change. Observational and modeling studies over the last several
37 decades continue to highlight the dominant role that the SO plays in the oceanic uptake of heat and
38 carbon in present day climate (Frölicher et al. 2015; Roemmich et al. 2015; Talley et al. 2016; Meredith
39 et al. 2019). The disproportionate role that this region plays in the planetary heat and carbon budget is
40 linked to the unique and complex physical circulation that exists in the SO. Coupled models that
41 contributed to previous generations of the Coupled Model Intercomparison Project (CMIP) have shown
42 large disagreements in their ability to represent the large-scale circulation and associated properties and
43 processes in this region (e.g., Russell et al. 2006a; Sen Gupta et al. 2009; Kuhlbrodt et al. 2012;
44 Meijers et al. 2012; Bracegirdle et al. 2013; Heuzé et al. 2013; Sallée et al. 2013a,b; Meijers 2014;
45 Frölicher et al. 2015; Shu et al. 2015; Ivanova et al. 2016; Hyder et al. 2018; Russell et al. 2018;
46 Beadling et al. 2019). This has led to low confidence ascribed to CMIP model-based projections of
47 future trends in the SO (Meredith et al. 2019). Given the vital role that this region plays in moderating
48 climate globally, such a large disagreement in model performance is alarming and has implications for
49 interpreting projected trends not just in the SO, but globally.

50 The coordinated multi-model experiments forced by observed changes in the Earth system from
51 ~1850 through the early 21st century, “twentieth-century control” (20C3M) in CMIP3 and “historical”
52 simulations in CMIP5-6, allows climate and Earth System Models (ESMs) to be scrutinized for how
53 well they simulate key aspects of the climate system relative to observable quantities over the
54 instrumental record. Through this process, a large number of systematic model biases pertaining to the
55 simulation of the SO have been identified and discussed at length in the IPCC-AR5 report (Flato et al.
56 2013) and summarized by Meijers (2014). Given the persistence of particular biases from one
57 generation to the next, one of the three major scientific questions in the design of CMIP6 was to

58 “investigate the origins and consequences of systematic model biases” with important and long-
59 standing biases in the SO simulation highlighted as an area that needs to be addressed (Stouffer et al.
60 2017). A review by Meijers (2014) painted the CMIP5 SO simulation as a ‘*better CMIP3*’ but lacking
61 dramatic advancements that may have been expected given the larger ensemble and much more
62 sophisticated models that participated relative to CMIP3.

63 One well known example of a persistent bias in the Southern Hemisphere climate is the
64 equatorward-biased westerly jet position (Russell et al. 2006a; Sen Gupta et al. 2009; Bracegirdle et. al
65 2013, Beadling et al. 2019). Beadling et al. (2019) showed how different combinations of biases in the
66 strength and position of the jet across the CMIP5 ensemble result in very different patterns of
67 integrated wind stress curl (WSC) over the SO. This is important because the pattern and strength of
68 the WSC forcing exerts a strong control on the resulting properties in the SO through its influence on
69 the SO meridional overturning circulation (MOC), water mass structure, and the strength and position
70 of the Southern Hemisphere subtropical and subpolar gyres that provide the meridional boundaries of
71 the Antarctic Circumpolar Current (ACC). Additionally, energy imparted to the ocean by the wind field
72 feeds the formation of mesoscale eddies through baroclinic instability (Olbers et al. 2004; Rintoul
73 2018). Thus, biases in the location and intensity of the momentum forcing from the overlying winds
74 may lead to inaccuracies in ocean mixing. The bias in westerly jet position has also been identified as
75 an emergent constraint, where models with weak and more equatorward biased Southern Hemisphere
76 westerly jets tend to exhibit the largest increase and poleward shift under increased warming (Kidston
77 and Gerber 2010; Bracegirdle et al. 2013). Mean state representation and warming driven changes in
78 midlatitude westerly winds have important ramifications for the ventilation of heat and carbon in the
79 Southern Hemisphere (Russell et al. 2006b; Waugh et al. 2019).

80 Coupled models have also consistently simulated upper-ocean temperatures in the SO that are
81 too warm relative to modern observations (Russell et al. 2006a; Sen Gupta et al. 2009; Sallée et al.

82 2013a,b; Beadling et al. 2019; Hyder et al. 2018). This influences geostrophic ocean circulation,
83 surface heat fluxes and water mass transformations, surface carbon fluxes, and the ability to accurately
84 represent Antarctic sea ice extent (SIE). The properties of the upper ocean are influenced by a complex
85 interplay between oceanic, atmospheric, and ice processes. This makes the upper ocean properties
86 highly sensitive to biases in individual model components (i.e., the atmosphere, ocean, or sea ice
87 models used in the coupled configuration). For example, Hyder et al. (2018) provided strong evidence
88 that sea surface temperature (SST) biases in the region 40°S – 60°S across the CMIP5 ensemble are
89 primarily the result of net flux biases in the stand-alone atmospheric model linked to poor
90 representation of clouds, cloud properties, and shortwave radiation errors.

91 The historical representation of the properties and large-scale circulation in the SO may play a
92 role in determining a model’s projected response to increased radiative forcing. Thus, a reduction in
93 uncertainty of future trends in the SO and globally may be achieved through improvement and detailed
94 understanding of mean state biases. Furthermore, as arguments grow *against* the idea of considering
95 the results of all model projections equally viable (“*model democracy*”), whereby uncertainties of the
96 trajectories of the climate system are assessed from a simple multi-model-mean approach (Knutti 2010;
97 Knutti et al. 2017; Eyring et al. 2019), it is vital to assess and interpret projected trends among models
98 with knowledge of biases in their historical simulations.

99 The evaluation presented here provides a robust and comprehensive assessment of key
100 observable metrics of SO properties and circulation in the historical simulations across a large
101 ensemble of CMIP6 models. Observable metrics assessed include surface momentum forcing, ACC
102 transport, density, salinity, and temperature patterns and gradients, and representation of Antarctic
103 seasonal SIE. The analysis has been done in a way that allows consistent assessment across model
104 generations from CMIP3 (Russell et al. 2006a; Sen Gupta et al. 2009) to CMIP5 (Beadling et al. 2019)
105 highlighting areas of model improvement and areas where systematic biases persist. For cohesiveness,

106 for each results section we first provide a discussion of the performance of that particular metric across
107 model generations and then present the CMIP6 results. We bring all of our results together in an overall
108 evaluation of cross generational performance, and suggest next steps in the conclusions and summary
109 section. As the results from the 21st century projections under various shared socio-economic pathways
110 (SSPs) for the CMIP6 scenarios have recently come online, it is urgently important to provide
111 comprehensive documentation of model biases in this climatically important region.

112 **Methods.**

113 *a) CMIP6 model output*

114 Thirty-four CMIP6 models are included in this assessment based on the availability of output
115 provided for each model’s “historical simulation” in the Earth System Grid Federation (ESGF) CMIP6
116 data archive at the time of publication. In CMIP6, the historical simulation spans 1850 to 2014 and is
117 forced by observed anthropogenic and natural sources of atmospheric composition changes and time-
118 evolving land cover (Eyring et al. 2016). The first ensemble member for each model’s historical
119 simulation is analyzed. In some cases, the “r1i1p1f1” member was not provided and another
120 appropriate member was analyzed. Table 1 lists the models, ensemble members, and additional details
121 regarding their ocean component. The analyses for all metrics presented here are performed on the
122 model’s native grid unless otherwise noted and follow the same exact computational methods as that
123 summarized in the methods section of Beadling et al. (2019). All metrics are computed as a time-
124 average of all monthly data spanning January 1986 to December 2005.

125 *b) Observational metrics*

126 Several observational metrics have been updated from Beadling et al. (2019). The World Ocean
127 Atlas 2018 (WOA18) product (Locarini et al. 2018; Zweng et al. 2018) is used to assess biases in the
128 density, salinity, and potential temperature differences across the ACC relative to those estimated from
129 observations. For comparison to the 1986 – 2005 time period simulated in the models, the 1985-94 and

130 1995-2004 decadal climatologies are averaged from the WOA18 product. For assessment of the surface
131 momentum exchange from atmospheric wind stress, the European Centre for Medium-Range Weather
132 Forecasts (ECMWF) ERA5 atmospheric reanalysis product is used (<https://cds.climate.copernicus.eu/>).
133 The ERA5 product, with increased spatial and temporal resolution and other improved features, is con-
134 sidered to be an improvement to ERA-Interim which ended production in August 2019 ([https://conflu-
135 ence.ecmwf.int/display/CKB/ERA5+data+documentation](https://confluence.ecmwf.int/display/CKB/ERA5+data+documentation)). ERA-Interim was used in the assessment of
136 SO surface momentum forcing by Beadling et al. (2019) given its proven reliability in representing
137 wind fields over the SO relative to other contemporary reanalysis products (Swart and Fyfe 2012;
138 Bracegirdle and Marshall 2012; Bracegirdle et al. 2013). For the model analysis, monthly atmospheric
139 tauu and tauv or ocean tauuo and tauvo output are used. If tauu and tauv are used, values over land are
140 masked out prior to computations.

141 For the assessment of Antarctic SIE, the monthly Sea Ice Index version 3 data product
142 (<https://nsidc.org/data/G02135/versions/3>) provided by the National Snow and Ice Data Center
143 (NSIDC; Fetterer et al. 2017) is used. This data is derived from satellite passive microwave data and
144 spans the time period 1978 to present day. In our assessment, we present the time-averaged monthly
145 data from 1986 to 2005. To be consistent with the NSIDC data product, we only consider grid cells
146 with a sea ice concentration greater than 15 % since satellite passive microwave instruments cannot ac-
147 curately measure concentrations below this value. To be consistent for comparison with Beadling et al.
148 (2019), after masking out values below 15%, the sea ice concentration data (*siconc*) is regridded to a
149 standard 1° horizontal resolution grid for models which have non-standard curvilinear horizontal grids.
150 This was done in Beadling et al. (2019) due to a lack of information about grid cell area to allow com-
151 putations on models with non-regular grids.

152 For the ACC transports, the net volume transport through the Drake Passage (DP; closest grid
153 cells to ~69°W) is computed in each model from their reported zonal velocity (*uo*) or mass transport

154 (umo) output. Velocities were only used if the mass transport output was not provided. Mass transport
155 is preferred over velocity for a number of reasons discussed in detail in Beadling et al. (2019) and ulti-
156 mately provides for a better representation of the true time-averaged flow. If mass transport was used,
157 the total transport was divided by a constant density of 1035 kg m^{-3} . For the observational benchmark,
158 we use the $173.3 \pm 10.7 \text{ Sv}$ estimate from the cDrake array experiment which was carried out from
159 2007 to 2011 (Chereskin et al. 2012; Chidichimo et al. 2014; Donohue et al. 2016). This value is the
160 sum of the $127 \pm 5.9 \text{ Sv}$ baroclinic transport (Chidichimo et al. 2014) and $45.6 \pm 8.9 \text{ Sv}$ barotropic
161 component computed from the cDrake array observations (Donohue et al. 2016). Please see the de-
162 tailed discussion in section “2a. Transport of the ACC through the Drake Passage” of Beadling et al.
163 (2019) regarding observational DP estimates over the last several decades. The larger transport com-
164 puted from the cDrake array relative to the canonical value of 134 Sv derived from hydrographic esti-
165 mates (Whitworth et al. 1982; Whitworth 1983; Whitworth and Peterson 1985; Cunningham et al.
166 2003) is purely attributable to the higher spatial and temporal resolution observations that allowed
167 strong barotropic currents near the bottom of the ocean to be resolved.

168 c) *B-SOSE Iteration 133*

169 Monthly ocean fields from the Iteration 133 solution of the Biogeochemical Southern Ocean
170 State Estimate (B-SOSE) at $1/6^\circ$ horizontal resolution spanning January 2013 to December 2018 are
171 also analyzed in this study (http://sose.ucsd.edu/BSOSE6_iter133_solution.html). This is an update
172 from the $1/3^\circ$ horizontal resolution Iteration 105 B-SOSE solution spanning January 2008 to December
173 2012 used in Beadling et al. (2019). B-SOSE, produced as part of the Southern Ocean Carbon Climate
174 Observational Modeling (SOCCOM) project, assimilates observations from shipboard data, profiling
175 floats, underway measurements, and satellites into a numerical model to produce a state estimate for
176 the SO. In B-SOSE, the MIT general circulation model (MITgcm) is fully coupled to the Nitrogen-ver-
177 sion of the Biogeochemistry with Light, Iron, Nutrients, and Gases (N-BLING) model (evolved from

178 Galbraith et al. 2010). Given the nature of B-SOSE, where the MITgcm is brought into consistency
179 with available observational data via an adjoint data assimilation approach, we expect B-SOSE to per-
180 form well in regions with a high density of observational measurements such as in the upper ocean and
181 along transects with repeat ship-based observations. B-SOSE is constrained by satellite measurements
182 of sea surface height (SSH) and mean dynamic topography, thus we expect B-SOSE to capture surface
183 currents and geostrophic flows consistent with other independent observations. It is important to note
184 that none of the hydrographic and velocity observations from the cDrake array (Chereskin et al. 2012;
185 Chidichimo et al. 2014; Donohue et al. 2016) are assimilated in B-SOSE, and thus the B-SOSE
186 transport provides an independent estimate. The momentum forcing at the ocean surface in B-SOSE is
187 derived from hourly ERA5 atmospheric winds, which are then adjusted throughout the assimilation to
188 achieve consistency with the ocean state. Please refer to Verdy and Mazloff (2017) for additional de-
189 tails on B-SOSE and a complete list of observational constraints used.

190 **3. Results**

191 *a) ACC transport*

192 The ACC transport is influenced by a large number of properties and processes in the SO,
193 including momentum input at the ocean surface from the overlying winds, the meridional gradient in
194 density across the current, interactions with bottom topography, mesoscale eddies, the position of the
195 subtropical and subpolar gyres (Meijers et al. 2012), and internal mixing processes, etc. Owing to this
196 complexity, achieving an accurate ACC strength has proven to be a difficult task in coupled models
197 (Russell et al. 2006a; Sen Gupta et al. 2009; Meijers et al. 2012; Beadling et al. 2019). The CMIP3
198 generation of models exhibited a very wide range of transports (Fig. 1, Table 2), with an inter-model
199 spread (1σ) of 71 to 77 Sv. Only three of the CMIP3 models studied collectively by Russell et al.
200 (2006a) and Sen Gupta et al. (2009) have mean values that fall within the observational uncertainty of
201 the Donohue et al. (2016) cDrake array ACC estimate (Fig. 1). From CMIP3 to CMIP5 there is a large

202 improvement in the range of the transports, with the inter-model spread (1σ) decreasing by ~ 36 Sv.

203 This improvement in ACC strength was also noted in the analysis by Meijers et al. (2012).

204 The spread in the ACC transport has increased in CMIP6 relative to CMIP5, ranging from 38
205 Sv simulated by INM-CM4-8 to 197 Sv simulated by GISS-E2-1-H. While there are no longer models
206 that have an ACC transport that is much too strong, several models exhibit an extremely weak transport
207 ($>7\sigma$ outside of the observational uncertainty; E3SM-1-0, MIROC-ES2L, CNRM-CM6-1-HR,
208 HadGEM3-GC31-MM, INM-CM4-8), reducing the multi-model-mean (MMM) by ~ 10 Sv from
209 CMIP5. Four models exhibit a transport weaker than any found in CMIP5. Interestingly, two of the
210 models with an exceptionally weak ACC transport have 0.25° horizontal resolutions (CNRM-CM6-1-
211 HR, HadGEM3-GC31-MM), with their 1° resolution versions (CNRM-CM6-1, HadGEM3-GC31-LL)
212 simulating transports on the order of 50 – 80 Sv stronger. Similar, but not as extreme, behavior is found
213 when comparing the 0.25° GFDL-CM4 (132 Sv) simulation with that of the 0.50° GFDL-ESM4
214 simulation (175 Sv).

215 Despite the range in ACC transport across the model ensemble increasing due to several models
216 with extremely weak magnitudes, CMIP6 is an overall improvement from CMIP5, with a larger
217 fraction of models falling within observational uncertainty (within 2σ of the Donohue et al. (2016)
218 estimate). Of the 31 CMIP5 models studied by Beadling et al. (2019), 10 models (32%) fell within
219 observational uncertainty. Of the 34 models studied here, 17 (50%) have transports within uncertainty.
220 It is important to note the caveat, as also discussed regarding the CMIP5 models in Beadling et al.
221 (2019), that most of the CMIP6 models analyzed here were developed *before* the Donohue et al. (2016)
222 ACC estimate was known. Thus, during modeling development, transports at lower values were likely
223 deemed reasonable. While we consider the Donohue et al. (2016) estimate as our benchmark for
224 diagnosing the model simulations, the margin of uncertainty may be slightly larger in reality than used
225 here given uncertainty associated with the fact that these observations were only collected from 2007-

226 2011. An independent estimate from assimilation using B-SOSE suggests a slightly lower ACC value
227 of 164 Sv (Table 2, Fig. 1).

228 Given this caveat, we consider the five models that fall just outside of the uncertainty bounds in
229 Figure 1 (and marked with an * in Table 2) to be simulating reasonable transports given that they only
230 differ from the models above them in Table 2 by a few Sverdrups. These models may fall within the
231 lower bound of the observational range if more than one ensemble member were included in the
232 analysis. The GFDL-CM4 simulation is an example of another caveat that complicates diagnosing the
233 accuracy of the ACC strength in coupled models, in that it exhibits significant centennial-scale
234 variability throughout the entire historical period, on the order of 30 Sv (Table S1, Fig. S1). If another
235 ensemble member were used, GFDL-CM4 can have an ACC value within the Donohue et al. (2016)
236 uncertainty range (other ensemble members not shown here). The BCC-CSM2-MR model is another
237 example which contains significant historical ACC variability and, when multiple ensemble members
238 are used, it can have an ACC transport that overlaps with the Donohue et al. (2016) range.

239 Several models exhibit significant decadal to multi-decadal variability in the ACC strength
240 throughout the entire historical period, on the order of 10 – 30 Sv (Table S1, Fig. S1). Such variability
241 may be associated with unrealistic quasiperiodic “superpolynya events” in the SO which alter the
242 density structure of the SO through intense open ocean convection. This problem is documented in the
243 GFDL-CM4 model (Held et al. 2019), where superpolynyas in the Ross Sea drive large centennial-
244 scale variability in the Southern Hemisphere climate. The lack of observational evidence of large
245 polynya events in the SO, with the exception of the 1974-1976 (Gordon 1978; Carsey 1980) and 2016-
246 2017 polynyas in the Weddell Sea (Campbell et al. 2019), and no similar events in the Ross Sea,
247 suggests that these frequent simulated “superpolynya events” are unrealistic and need to be improved
248 upon in future model development. In addition to the extreme rarity of these events in the real world,
249 the modeled convection, ventilation, and climate change that occurs during these simulated polynyas

250 bears no resemblance to observed polynya events (Gordon 1978; de Lavergne et al. 2014; Dufour et al.
251 2016; Campbell et al. 2019). In the model, when these events occur, the formation of dense water in the
252 subpolar SO translates into an increased ACC transport. If the simulation contains these events, the
253 ACC assessment is sensitive to the 20-yr period in the historical simulation chosen for analysis (Fig.
254 S1).

255 There is evidence of improvement at individual modeling centers with two examples being the
256 IPSL and NASA-GISS models. In CMIP3 and CMIP5 the IPSL models had some of the weakest ACC
257 transports amongst all models, with transports on the order of 34 Sv in CMIP3 and 94 - 108 Sv in
258 CMIP5. In CMIP6, the IPSL-CM6A-LR transport has increased dramatically to 147 ± 6.7 Sv (2σ),
259 falling just within 2σ of the Donohue et al. (2016) estimate. The GISS-R models were biased much too
260 strong in previous generations (266 Sv CMIP3; 246 Sv CMIP5). The GISS-E2-G and GISS-E2-G-CC
261 (E2-G but with interactive carbon cycle) models, which are an updated and improved version of the
262 GISS-E2-R model used in CMIP5 now simulate ACC transports of 148 and 146 Sv, respectively. The
263 improvement of the ACC transport in the IPSL and GISS-E2-R (now G) families of models appears to
264 be directly due to improvement in all of the metrics summarized in Table 2 of the present study and
265 Table 2 of Beadling et al. (2019).

266 In the real ocean, the ACC flow through the DP is composed of strong and narrow eastward
267 flowing jets that extend from the surface to the bottom of the ocean. Through the passage, strong
268 bottom eastward velocities that average 1.3 cm s^{-1} provide an additional 45.6 ± 8.9 Sv barotropic flow
269 to the 127 ± 5.9 Sv baroclinic transport (Chidichimo et al. 2014; Donohue et al. 2016). The CMIP6
270 models and B-SOSE have very different zonal velocity structures through the DP (Fig. 2, Fig. S2). The
271 eddy-permitting $1/6^\circ$ B-SOSE and the $1/4^\circ$ GFDL-CM4 simulations yield very similar structures, with
272 finely spaced jets that extend from the surface to the ocean bottom at most latitudes and increase in
273 strength northwards within the passage. These jets correspond to the dynamical ACC fronts

274 characteristic of the observed flow within the DP (Lenn et al. 2007; Firing et al. 2011), with the
275 majority of the eastward flow concentrated in the Subantarctic Front (SAF) along the northern
276 boundary of the DP and the Polar Front (PF) just to its south.

277 Interestingly, the GFDL-CM4, GFDL-ESM4, CNRM-CM6-1-HR, HadGEM3-GC31-MM, and
278 INM-CM5-0 models, which all have resolutions $1/2^\circ$ or finer, exhibit counterflowing westward
279 velocities northward of 56°S at depth. This may be the result of recirculation in this region that is
280 resolved as horizontal resolution increases. None of the coarser resolution models have significant
281 westward flows at depth (Fig. 2, Fig. S2). There is some observational evidence that suggests some
282 mean westward flow through the DP exists (evidenced where the transport stops accumulating or dips
283 in Figure 10 of Firing et al. (2011) and Figure 3 of Donohue et al. (2016)). However, these observed
284 westward transports are relatively weak and not close to the magnitude of those found in these models.
285 The measurements used to estimate the transport through the DP from the cDrake array in Donohue et
286 al. (2016) and from shipboard acoustic Doppler current profiler (ADCP) velocity data by Firing et al.
287 (2011) are not collected along a single meridional cross-section (as the CMIP models are analyzed at
288 69°W) but generally span the longitudes of $\sim 66^\circ\text{W}$ to 60°W . To assess the accuracy of these large
289 westward velocities found in the higher resolution models, the models need to be sampled in a fashion
290 that is more aligned with the manner in which the measurements were collected.

291 In most models, the majority of the ACC flow is concentrated in a single broad jet on the
292 northern side of the passage, rather than being distributed among several finely resolved jets as
293 observed in the real ocean. Many of the CMIP6 models exhibit shallow flow, with strong zonal jets that
294 do not extend to the seafloor. The diversity in vertical extent of the flow may have important
295 implications for the varied ACC strength found across models given the influence that interactions of
296 the mean flow with bottom topography have on the ACC structure and strength through topographic
297 steering, and dynamics such as momentum dissipation via bottom friction, etc. The degree to which

298 topography and surface forcing impacts the mean flow likely differs markedly across models. For
299 example, models with shallow jets are less likely to be influenced by interactions with bottom
300 topography, i.e., less bottom drag will be exerted on the large-scale flow than models which have
301 strong jets that extend to the sea floor. It appears that many of the errors in the ACC structure are
302 related to the model's horizontal resolution since only the highest resolution models begin to capture
303 the observed jets and vertical extent.

304 The ACC strength is influenced by the competition between wind-driven surface divergence
305 which acts to steepen isopycnals across the DP channel and baroclinic eddies which act to reduce the
306 isopycnal slopes. The parameterized mixing schemes and coefficients employed in model simulations
307 exerts a strong control on the resulting ACC strength and its vertical structure (Gent et al. 2001; Ragen
308 et al. 2020). The $\sim 1^\circ$ resolution CMIP6 models all employ various subgrid-scale mixing schemes with
309 different coefficients for parameterized eddy fluxes (Table 1), likely explaining a large portion of the
310 diversity found across the CMIP6 models, as was the case in previous model generations (Kuhlbrodt et
311 al. 2012). With the large diversity in sophisticated mixing schemes and magnitudes of eddy-induced
312 advection coefficients employed across these models (Table 1), we do not attempt to quantify the role
313 of subgrid-scale mixing on the model diversity in these simulations. We note that in the 0.25°
314 resolution models (GFDL-CM4, HadGEM3-GC31-MM, and CNRM-CM6-1-HR), there are no
315 parameterizations of mesoscale eddy transports and mesoscale eddies are explicitly, yet incompletely
316 resolved at high latitudes (Table 1). We refer readers to Adcroft et al. (2019) for a detailed discussion
317 on the effect of representing versus parameterizing mesoscale eddies on the ACC structure and other
318 features of the SO circulation in the GFDL-CM4 (explicitly resolves mesoscale eddies) and GFDL-
319 ESM4 (parameterized mesoscale eddy transport) models.

320 *b) Surface momentum forcing of the Southern Ocean and near-surface properties*

321 1) SURFACE MOMENTUM FORCING FROM WIND STRESS

322 The surface momentum forcing from the overlying wind stress provides a frictional force at the
323 ocean surface in the SO, “pushing” the ACC, while also “pulling” dense water from the deep ocean to
324 the surface through strong wind-driven surface divergence determined by the WSC magnitude.
325 Equatorward of the westerly wind stress maximum ($\tau_{max}^{(x)}$), buoyant surface waters are pushed
326 northward in the Ekman layer and subsequently downwelled into the interior ocean as a result of strong
327 positive WSC. These waters are subducted northward into the Southern Hemisphere subtropical gyres
328 as Subantarctic Mode Water (SAMW) and Antarctic Intermediate Water (AAIW) (Hanawa and Talley
329 2001). Poleward of $\tau_{max}^{(x)}$, the surface divergence results in steeply sloped isopycnals which drive
330 dense, deep water from the interior ocean polewards and toward the sea surface around Antarctica via
331 Ekman upwelling. These dynamics set up a strong meridional density gradient across the latitudes of
332 the ACC, providing the conditions to drive a strong eastward geostrophic flow.

333 Previous generations of models have struggled with achieving accurate wind stress forcing at
334 the ocean surface, with many models typically having relatively weak and equatorward biased $\tau_{max}^{(x)}$
335 values (Russell et al. 2006a; Sen Gupta et al. 2009; Meijers et al. 2012; Bracegirdle et al. 2013; Flato et
336 al. 2013; Russell et al. 2018; Beadling et al. 2019) (Table 2, Fig. 3c,e). In CMIP3, several models had
337 $\tau_{max}^{(x)}$ values on the order of 0.10 – 0.11 N m⁻² (Russell et al. 2006a), relative to the ERA5 value of
338 0.1788 N m⁻² used here as our observational benchmark. The CMIP5 generation showed improvement
339 with a minimum $\tau_{max}^{(x)}$ value on the order of 0.14 N m⁻² and most models simulating values within
340 observational uncertainty (Beadling et al. 2019). This improvement is noted in the shift from a CMIP3
341 MMM $\tau_{max}^{(x)}$ value of 0.1591 N m⁻² to 0.1829 N m⁻² in CMIP5 (Table 2), fewer models outside the
342 lower bound of the ERA5 range in Fig. 3c, and a better agreement in the pattern of the zonally-
343 averaged wind stress over the SO (compare Fig. 3a to Fig. 2 of Beadling et al. 2019). The CMIP6
344 ensemble tends to exhibit stronger $\tau_{max}^{(x)}$ than CMIP5, yielding a slightly increased MMM value (Table

345 2). Considering the mean and associated standard deviation (2σ) over the 20-yr period, only two
346 models have $\tau_{max}^{(x)}$ values that do not overlap with the ERA5 range. The BCC-ESM1 and MIROC-
347 ES2L models exhibit too strong and too weak $\tau_{max}^{(x)}$ values, respectively. The BCC-CSM-MR and
348 NESM3 models have strong wind stress relative to ERA5, with the lower bound of their standard
349 deviations (2σ) barely overlapping with that of ERA5.

350 The latitudinal location of $\tau_{max}^{(x)}$ ($\sim 52^\circ\text{S}$) is another metric that has been consistently analyzed
351 across model generations and has improved. Yet, a systematic equatorward bias has persisted. The
352 CMIP3 ensembles studied by Russell et al. (2006a) and Sen Gupta et al. (2009) had models with $\tau_{max}^{(x)}$
353 locations ranging from $\sim 42^\circ\text{S}$ to 54°S (Fig. 3e), with 63% (10 out of 16) to 78% (14 out of 18) of the
354 models having their mean $\tau_{max}^{(x)}$ located north of 50°S . The CMIP5 models showed a clear
355 improvement, with only 36% (11 out of 31) of models studied exhibiting locations north of 50°S and a
356 narrowing inter-model spread (Fig. 3e). The CMIP6 models show even more improvement, with only
357 5 out of 34 (15%) models having their $\tau_{max}^{(x)}$ north of 50°S . MIROC-ES2L, with its $\tau_{max}^{(x)}$ at 46.11°S
358 barely overlaps with the ERA5 mean and standard deviation (2σ) and is a clear outlier amongst the
359 other CMIP6 models with respect to its zonal wind structure across the SO (Fig 3a,e). The MMMs
360 across generations summarized in Table 2, clearly indicates an improvement in both the mean location
361 of $\tau_{max}^{(x)}$ and in the inter-model spread, going from $47.57 \pm 2.80^\circ\text{S}$ to $51.29 \pm 1.68^\circ\text{S}$ from CMIP3 to
362 CMIP6.

363 The poleward shift in the location of $\tau_{max}^{(x)}$ from CMIP5 to CMIP6 has resulted in an increase
364 in the total wind stress forcing (both from $\tau^{(x)}$ and WSC) over the open DP latitudes (Table 2). The
365 structure of the zonally-integrated WSC over the SO has improved from CMIP5 to CMIP6 (compare
366 Fig. 3b to Fig. 2 of Beadling et al. 2019). The CMIP6 models are generally getting an accurate
367 magnitude of Ekman suction / pumping over approximately the right locations, with the exception of

368 MIROC-ES2L. The region south of 55°S yields the largest disagreement in the magnitude of the
369 zonally-integrated WSC among models, likely linked to diverging representation of the polar easterlies
370 along the Antarctic margin. Noting this disagreement, we considered two additional metrics in this
371 analysis to characterize the WSC field, the magnitude and location of the minimum WSC in the SO,
372 corresponding to the magnitude and location of the maximum Ekman suction. Nine CMIP6 models
373 yield a minimum zonally-integrated WSC that is too weak, and five models have their latitudinal
374 location too far equatorward relative to ERA5 (Table 2). Given that these WSC metrics are tied to the
375 magnitude and location of where dense, carbon-rich water from the deep ocean is being pumped to the
376 sea surface, the divergence in performance here may be linked to divergence in the simulated SO
377 carbon budget in the CMIP6 ensemble. This is a topic of future study.

378 Given the fact that the magnitude and location of the wind stress plays an important role in
379 setting the isopycnal slopes across the ACC as discussed at the start of this section, one may expect
380 $\tau_{max}^{(x)}$ to be significantly correlated with the strength of the ACC. However, a statistically significant
381 relationship between $\tau_{max}^{(x)}$ and ACC strength is *not* found when considering the entire model ensemble
382 in CMIP3 or CMIP5 models (Kuhlbrodt et al. 2012; Meijers et al. 2012; Beadling et al. 2019). A
383 statistically insignificant relationship is also found between the latitudinal location of $\tau_{max}^{(x)}$ and ACC
384 strength in CMIP5 (Beadling et al. 2019). This lack of strong correlation between these wind metrics
385 and the ACC strength suggest that, while the momentum forcing by the Southern Hemisphere westerly
386 winds is undoubtedly a major driver of the ACC, other factors are exerting a strong influence on ACC
387 strength. The buoyancy forcing across the current also influences the meridional tilt of the isopycnals
388 in this region. Thus, the temperature and salinity properties from the surface to the interior ocean plays
389 a major role in setting the ACC strength. In CMIP5, several models simulated exceptionally strong
390 (weak) ACC flow with weak (strong) $\tau_{max}^{(x)}$. In some of these cases, errors in the buoyancy structure of
391 the SO compensated for the wind stress forcing, allowing steep isopycnal slopes to be maintained in

392 the absence of strong wind forcing or vice versa. Examples from CMIP5 included the GISS-E2-R-CC
393 and HadGEM3 models with ACC transports in excess of 240 Sv stemming from excessive density
394 gradients driven by large errors in the SO temperature and salinity structure (see detailed discussion in
395 Beadling et al. 2019).

396 The CMIP6 ensemble exhibits a stronger relationship (yet still not statistically significant,
397 $p=0.15$) relative to previous model generations between the ACC and $\tau_{max}^{(x)}$ (Fig. 4a). No statistically
398 significant relationships are found between ACC strength and the other wind stress metrics in Table 2
399 (not shown). The MIROC6 model has one of the weakest $\tau_{max}^{(x)}$ magnitudes with a position that is
400 slightly equatorward shifted relative to ERA5, yet one of the strongest ACCs out of the ensemble. The
401 E3SM-1-0, HadGEM-GC31-MM, INM-CM4-8, and CNRM-CM6-1-HR models all have reasonable
402 representation of both the position and strength of $\tau_{max}^{(x)}$, yet yield extremely weak ACC transports.

403 In addition to the buoyancy structure, explicitly resolved (in the case of HadGEM3-GC31-MM,
404 GFDL-CM4, and CNRM-CM6-1-HR) or parametrized mixing associated with mesoscale eddies also
405 plays a role in explaining the lack of a linear correlation between wind stress forcing and the ACC
406 strength in the CMIP6 models. Stronger wind forcing acts to enhance the meridional isopycnal tilt
407 across the ACC. However, in a regime of eddy saturation (Munday et al. 2013), this isopycnal tilt can
408 be counterbalanced by the production of eddies which then act to minimize the meridional density
409 gradient. Given that many of the CMIP3 models employed fixed eddy-induced advection coefficients,
410 Kuhlbrodt et al. (2014) showed that the sensitivity of the ACC transport to the magnitude of the eddy-
411 induced advection coefficient was larger than to the zonal wind stress maximum in CMIP3 models. As
412 discussed in the previous section, the implementation of parameterized sub-grid scale mixing
413 associated with eddy fluxes differs widely across the CMIP6 models (Table 1), likely contributing to
414 the lack of a strong relationship between the wind stress metrics and the ACC.

415 2) NEAR-SURFACE BIASES

416 The near-surface (0 – 100 m) thermal structure of the SO is characterized by a large meridional
417 gradient in temperature ranging from less than -1°C near the Antarctic coast to $\sim 20^{\circ}\text{C}$ in the
418 subtropical gyres at 30°S (Fig. 5a, Fig. 6a). Achieving accuracy in representing the near-surface
419 temperature structure in the SO, particularly south of 40°S , has proven to be a significant challenge for
420 the climate modeling community, with significant warm biases persisting across model generations.
421 Generally, above 200 m, the MMM CMIP3 and CMIP5 temperatures were biased slightly cold in the
422 global ocean with the exception of the SO where the upper ocean has been characterized by consistent
423 excessive warm biases (Flato et al. 2013). When considering the zonal-mean upper ocean temperature
424 distribution, CMIP3 and CMIP5 models generally represented the structure well (Russell et al. 2006a;
425 Sen Gupta et al. 2009; Beadling et al. 2019) with biases generally within $\sim 1^{\circ}\text{C}$ of that observed at a
426 given latitude. However as noted by Sen Gupta et al. (2009), zonal-averaging obscures regional biases
427 in upper ocean temperature that in some cases exceed 5°C . In CMIP3, major biases in the upper ocean
428 were found south of $\sim 45^{\circ}\text{S}$ in the region of the ACC and along the eastern boundaries of the basins, and
429 were attributed to poor representation of eastern boundary currents (Sen Gupta et al. 2009). Across the
430 CMIP5 ensemble, excessive surface temperatures translated into consistent warm biases found in
431 ventilated layers of the SO including surface subtropical, mode, and intermediate waters (Sallée et al.
432 2013b).

433 Excessive SO surface temperatures in CMIP5 originated to some extent from excessive
434 downward shortwave radiation related to poor representation of clouds and cloud properties, with a
435 strong correlation found between shortwave cloud forcing and the modeled spread in SO surface air
436 temperatures (Ceppi et al. 2012; Schneider and Reusch 2016). Work by Hyder et al. (2019) on the
437 CMIP5 ensemble showed that these biases in cloud-related shortwave radiation were mostly due to
438 errors in the stand-alone atmospheric model components used. Additionally, inaccuracies in the
439 representation of the large-scale ocean circulation in the SO including the location of the ACC, strength

440 and location of subtropical gyre boundary currents, wind-driven upwelling and associated mixing of
441 interior ocean waters with the sea surface, as well as eddy-induced transports and their
442 parameterizations, all play a role in explaining regional temperature biases. Propagation of deep ocean
443 warming to the sea surface related to climate model drift present in some model simulations also
444 contributes to the near-surface temperature biases. An example of this from the CMIP3 and CMIP5
445 ensemble was the GFDL-CM3 simulation with excessively warm biases in the abyssal ocean that
446 impacted the SO surface climate in the historical simulations (Griffies et al. 2011).

447 Noting that large local errors can be obscured by only assessing models according to their
448 zonal-mean properties, we present the upper 100 m temperature and salinity biases (Fig. 6) for the
449 entire SO. For comparison to previous studies (Russell et al. 2006a; Sen Gupta et al. 2009; and
450 Beadling et al. 2019), we also provide the zonally-averaged structure (Fig. 5). With the exception of
451 four models (MIROC6, MIROC-ES2L, GISS-E2-1-H, and EC-Earth3-Veg), the CMIP6 zonal-mean
452 temperature structure shows improved agreement across all latitudes relative to the spread found across
453 models in CMIP3 and CMIP5 (Fig. 4c in Beadling et al. 2019 and Fig. 4a in Russell et al. 2006a).
454 Model agreement tightens up north of 50°S. The MIROC6, MIROC-ES2L, GISS-E2-1-H, INM-CM4-8
455 and EC-Earth3-Veg models stand out in Fig. 6a and Fig. S3 with excessive warm biases that dominate
456 the entire circumpolar SO south of 40°S. Many other models exhibit regions with biases in excess of
457 3°C that are mostly concentrated within or just north of the ACC region, with the warmest biases
458 appearing mostly in the South Atlantic and Indian basins. All CMIP6 models tend to exhibit some
459 degree of temperature bias along the southern margin of the subtropical gyres or along the boundary
460 current regions such as the Brazil – Malvinas confluence zone and Agulhas retroflexion. Regional
461 temperature biases along the ACC margin and in the boundary current regions are likely related to
462 discrepancies in model representation of the pathways of these large-scale currents as was the case in
463 CMIP3 (Sen Gupta et al. 2009). The B-SOSE and GFDL-CM4 simulations are very similar, with the

464 exception of more intense cold biases in the subtropical regions in GFDL-CM4. These patterns and
465 magnitudes of biases in near-surface temperatures have implications for the accurate representation of
466 surface heat fluxes, SO ventilation of heat and carbon, water mass characteristics, and on the ability to
467 accurately represent the Antarctic SIE.

468 The near-surface SO salinity structure is characterized by relatively fresh water in the latitude
469 band of the ACC where buoyancy is gained at the ocean surface via precipitation and sea ice meltwater
470 is entrained into the Ekman drift (Fig. 5b, Fig. 6b). The upper ocean is slightly more saline south of the
471 ACC particularly in the Weddell and Ross Seas and along the Antarctic coast. North of the ACC, high
472 salinity subtropical waters dominate the upper ocean. In CMIP3 and CMIP5, models showed
473 very wide discrepancies in the representation of upper ocean salinity, even in the zonal-mean (Russell
474 et al. 2006a; Sen Gupta et al. 2009; Beadling et al. 2019). On average, both the CMIP3 and CMIP5
475 ensembles yielded fresh biases throughout the entire SO (Sen Gupta et al. 2009; Beadling et al. 2019),
476 with the largest fresh biases in the upper ocean north of 50°S. CMIP5 models generally agreed with
477 one another (but with a significant fresh bias) north of ~50°S where precipitation minus evaporation
478 dominates the freshwater budget. Much larger inter-model spread was found in the zonal-mean upper
479 ocean salinity in the seasonal sea ice zone (Beadling et al. 2019). In CMIP5, the fresh biases in
480 combination with the warm biases discussed above resulted in water masses that were too light in the
481 ventilated layers of the SO (Sallée et al. 2013b).

482 In CMIP6, the near-surface zonal-mean salinity structure (Fig 5b) shows improvement in the
483 inter-model spread across all latitudes relative to previous generations (Fig. 4d in Beadling et al. 2019
484 and Fig. 4b in Russell et al. 2006a). However, the systematic bias of models generally being too fresh
485 in the upper ocean persists into CMIP6. Many models exhibit intense fresh biases exceeding 0.50,
486 concentrated mostly north of 45°S in the subtropical zones (Fig. 6b, Fig S4), however several models
487 have fresh biases of these magnitudes along the Antarctic coast. Several models exhibit large saline

488 biases along Antarctic margin or within the ACC region. The circumpolar nature of many of these
489 saline biases suggest they may be linked to the upwelling of saline North Atlantic Deep Water
490 (NADW) within the ACC. These near-surface salinity biases combined with those of near-surface
491 temperature (Fig 6) suggest the water mass characteristics in the ventilated layers of the SO may differ
492 widely among models, similar to CMIP5 (Sallée et al. 2013b). Given that the formation and properties
493 of SO water masses are intimately tied to global ocean heat and carbon uptake, the substantial biases in
494 the near-surface properties in CMIP6 are potentially worrying from a global climate modeling
495 perspective.

496 *c) Interior ocean properties and gradients across the ACC region*

497 The strength and structure of the ACC is largely determined by the meridional density gradient
498 according to thermal wind balance. Thus, the ACC simulation is highly sensitive to biases in the
499 interior ocean properties from the surface to the abyssal ocean. In coupled models the ACC generally
500 spans the latitudes $\sim 65^{\circ}\text{S}$ to 45°S and the meridional difference in properties across these latitudes has
501 been used as a metric to assess model performance (Russell et al. 2006a; Farneti et al. 2015; Beadling
502 et al. 2019). We note that this is a crude simplification given that the meridional boundaries of the ACC
503 are highly dependent on longitude and differ across models (Sen Gupta et al. 2009; Meijers et al.
504 2012). While not a precise computation according to thermal wind, where meridional gradients in
505 density are vertically integrated to solve for baroclinic velocities, the zonally-averaged and depth-
506 averaged density contrast from 65°S to 45°S used as a model metric serves as a proxy for the intensity
507 of the isopycnal slopes across the current, which generally reflect the strength of the ACC.

508 Russell et al. (2006a) considered the zonally-averaged and depth-averaged (0 to 1500 m; above
509 the sill depth of the DP) difference in meridional potential density ($\Delta\rho$) from 65°S to 45°S for the 18
510 CMIP3 models studied. Relative to the WOA 2001 value of 0.58 kg m^{-3} (this value remains the same
511 when calculated from the WOA18 product) used as the observational benchmark, a range from 0.18 kg

512 m^{-3} to 0.97 kg m^{-3} was found across models. Considering a window of error of 25% on either side of
513 the WOA 2001 value, 39% of the models simulated a $\Delta\rho$ within the observational range, 50% had a
514 weak $\Delta\rho$, and 11% had excessively strong $\Delta\rho$ across the ACC. The UKMO-HadCM3 model, which
515 yielded the most excessive $\Delta\rho$ at 0.97 kg m^{-3} did so as the result of a very large meridional difference
516 in salinity (ΔS). The GISS-E2 model was an example of a model that yielded an excessive $\Delta\rho$ due to a
517 large meridional difference in temperature (ΔT) and a weak and opposite sign ΔS . Beadling et al.
518 (2019) considered the zonally-averaged and full-depth averaged $\Delta\rho$ (referenced to the surface) across
519 65°S to 45°S in 31 CMIP5 models, yielding a range of 0.13 kg m^{-3} (BNU-ESM) to 0.47 kg m^{-3}
520 (HadCM3) relative to the 0.25 kg m^{-3} value from the WOA 13 product. Considering this metric in
521 CMIP5, 55% of the models fell within the WOA13 25% error margin, 23% were too weak and 23%
522 were too strong. Similar to CMIP3, the IPSL models produced some of the weakest $\Delta\rho$, while the
523 HadCM3 and GISS models produced some of the strongest, with these biases driving errors in ACC
524 strength, despite accuracy in the wind stress forcing at the ocean surface (Beadling et al. 2019).

525 Following Beadling et al. (2019), the full-depth averaged, zonally-averaged $\Delta\rho$ (referenced to
526 the surface), ΔT , and ΔS across 65°S to 45°S are computed for the CMIP6 models. The ensemble
527 exhibits a slightly smaller spread in simulated $\Delta\rho$ relative to CMIP5, ranging from 0.13 kg m^{-3}
528 (MIROC-ES2L) to 0.39 kg m^{-3} (GISS-E2-1-H, ACCESS-CM2). Considering a 25% error margin
529 relative to the WOA18 $\Delta\rho$, 71% of the models simulate a $\Delta\rho$ within error, 23% are too strong, and 6%
530 are too weak (Table 2). Comparing the percentage of models that simulate a $\Delta\rho$ within the 25%
531 WOA18 error margin, there is a clear improvement from CMIP5 to CMIP6, with models converging
532 towards the WOA18 value.

533 Similar to previous model generations, the strength of the ACC generally scales with the
534 magnitude of $\Delta\rho$ across 65°S to 45°S (Fig. 4b). Upon breaking down the density difference into the
535 contributing differences in ΔT and ΔS , a wider spread in model performance is found (Table 2; Fig.

536 4c,d). Of the models that exhibit an excessive $\Delta\rho$, five are the result of having a much too large ΔS
537 (GISS-E2-1-H, INM-CM5-0, SAM0-UNICON, CESM2, CESM2-WACCM), two are the result of too
538 large ΔT (ACCESS-CM2, ACCESS-ESM1-5), and one exhibits large biases in both ΔS and ΔT
539 (CanESM5).

540 Three of the five models that fall into the category of yielding an excessive $\Delta\rho$ due to large
541 biases in ΔS share the Parallel Ocean Program version 2 (POP2) ocean model component (Table 1)
542 coupled with a different atmospheric model (CESM2:CAM6; CESM2-WACCM:WACCM6; SAM0-
543 UNICON:CAM5.3 with UNICON). The models with the POP2 ocean all appear to have excessively
544 saline water seemingly originating in the deep Atlantic that penetrates the upper ocean south of the
545 ACC region, and a thick fresh bias at the surface layer, penetrating through the upper few thousand
546 meters in the subtropical regions in all basins (Fig. 7b). The INM-CM5-0 model with excessive $\Delta\rho$ due
547 to large biases in ΔS shows a very similar pattern to the POP2 biases described above, but with more
548 pronounced saline biases in the deep ocean and in the upper ocean south of the ACC (Fig. 7b). The
549 GISS-E2-1-H model appears to only have a too strong ΔS due to thick fresh biases throughout the
550 water column north of $\sim 45^\circ\text{S}$ (Fig. 7b). The two ACCESS models that fall into the category of
551 excessive $\Delta\rho$ due to large biases in ΔT , which share the MOM5 ocean component, yield large cold
552 biases in the upper ocean south of the ACC and thick layers of warm biases in the upper few thousand
553 meters of the ocean in the subtropical regions (Fig 7a, Fig. S5). Similar to the POP2 models, the
554 ACCESS models appear to have biases that originate in the region where NADW enters the SO.

555 Several models simulate a ΔS in the opposite direction from WOA18, with fresh biases in the
556 water column on the southern edge of the ACC often accompanied by saline biases in the water column
557 on the northern edge, concentrated in the upper 1500 m in the subtropical gyre regions or NADW
558 regions (Fig. 7b). In many cases, biases in ΔT and ΔS compensate for one another yielding an accurate
559 $\Delta\rho$. Many of the model biases in the temperature and salinity structure of the SO appear to be

560 concentrated in the recently ventilated layers or in the deep Atlantic suggesting most stem from
561 inaccuracies in the surface climate (reflected in the biases in Fig. 6, Fig. S3, and Fig. S4) or
562 inaccuracies in the simulation of NADW. An interesting relationship found in this analysis is a strong
563 correlation between $\Delta\rho$ and the minimum Antarctic SIE (Fig. 4e), suggesting the representation of the
564 Antarctic SIE is important for achieving an accurate SO density structure likely through its influence
565 on upper ocean salinity and water mass transformation processes. However, the mechanisms behind
566 this deserve a detailed follow up study.

567 *d) Antarctic Sea Ice Extent*

568 Antarctic sea ice exerts a strong influence on the SO through its impacts on air-sea heat
569 exchange, local surface albedo, and the upper ocean freshwater budget. Recent work using numerical
570 models and observations have highlighted that freshwater fluxes associated with buoyancy loss during
571 sea ice formation and freshwater gain in the upper ocean from sea ice melt are dominant components in
572 the transformation of water masses in the SO (Abernathey et al. 2016; Pellichero et al. 2018). Wind-
573 driven sea ice export and subsequent delivery of freshwater have played a role in determining the
574 observed mean salinity distribution in the upper SO (Haumann et al. 2016; Cerovecki et al. 2019). The
575 representation of Antarctic SIE in models is complicated by the fact that sea ice is highly sensitive to
576 both atmospheric and ocean forcing. Errors in the representation of Antarctic SIE can *result in or be the*
577 *result of* errors in biases in SO temperature and salinity structure, patterns of surface wind stress
578 forcing, water mass properties and location of deep ocean upwelling, and geostrophic ocean circulation
579 such as the strength and pathway of the ACC. Furthermore, the representation of Antarctic sea ice in
580 historical simulations of coupled models has been shown to be linked to projected changes in the
581 Southern Hemisphere westerly jet, where models with larger Antarctic sea ice area in their historical
582 simulations exhibit more sea ice retreat and less strengthening of the jet under increased radiative
583 forcing throughout the 21st century (Flato 2004; Bracegirdle et al. 2015; Bracegirdle et al. 2018).

584 Biases in Antarctic sea ice representation can translate into errors in air-sea gas exchange and simulated
585 heat and carbon storage through its impact on water mass transformation and through its complex
586 coupling between surface air temperature, net precipitation, and the strength of the Southern
587 Hemisphere westerly jet.

588 Accurate representation of mean-state Antarctic sea ice extent, area, and distribution has been
589 an existing challenge in the climate modeling community, with Antarctic SIE generally poorly
590 represented (Parkinson et al. 2006; Flato et al. 2013; Shu et al. 2015). Across generations, improvement
591 in Antarctic SIE has been marginal relative to Arctic (Mahlstein et al. 2013). CMIP3 and CMIP5
592 models have shown errors in representing the overall Antarctic SIE and regional distributions, the
593 seasonal cycle, and observed trends (Parkinson et al. 2006; Connolley and Bracegirdle 2007; Sen
594 Gupta et al. 2009; Turner et al. 2012; Flato et al. 2013; Mahlstein et al. 2013; Shu et al. 2015; Ivanova
595 et al. 2016; Roach et al. 2018). When the ensemble average of SIE is used as a metric, it appears that
596 models perform well and that significant improvements have been made from CMIP3 to CMIP5 (Sen
597 Gupta et al. 2009; Flato et al. 2013; Shu et al. 2015). However, this averaging is severely misleading;
598 large biases are seen among individual models. In many cases, even when the magnitude of SIE is
599 numerically well represented, the spatial patterns are often completely unrealistic (Connolley and
600 Bracegirdle 2007). Using an ensemble average of SIE as a metric is also misleading due to the differing
601 degrees of strong model internal variability in this region across models (Mahlstein et al. 2013; Deser
602 et al. 2010).

603 It is difficult to make a direct “cross generational” performance conclusion of the representation
604 of Antarctic SIE due to differing choices in sea ice metrics used in previous analysis on the CMIP3 and
605 CMIP5 ensemble (sea ice area, thickness, volume, differing representative time periods: annual SIE,
606 maximum or minimum SIE, September or February values, or summer / winter seasonal averages).

607 The analysis by Mahlstein et al. (2013) concludes that “*The representations of Antarctic sea ice in*
608 *CMIP5 models have not improved compared to CMIP3 and show an unrealistic spread in the mean*
609 *state that may influence future sea ice behavior*”. However, the IPCC-AR5 report notes that “*The*
610 *CMIP5 multi-model ensemble exhibits improvements over CMIP3 in simulation of sea ice extent in*
611 *both hemispheres*” (Flato et al. 2013).

612 In our analysis of the Antarctic SIE in CMIP6 models, we perform the same computations
613 across the same metrics as in the CMIP5 analysis by Beadling et al. (2019). Similar to previous model
614 generations, the CMIP6 models have an accurate seasonal cycle with a minimum occurring in February
615 and a maximum in September (Table 3; Fig. 8a). The CNRM-CM6-1-HR and CNRM-CM6-1 models
616 are the only models that simulate a maximum SIE in October rather than September. The annual SIE
617 simulated across models ranges from 1.70 million km² (MIROC-ES2L) to 13.29 million km²
618 (NorCPM1). Considering the observed and modeled annual mean and standard deviation (2σ) over the
619 20-yr period, one model (NorCPM1) simulates a too great SIE while 25 models simulate a too small
620 annual SIE. Relative to CMIP5, the CMIP6 ensemble has fewer models with annual Antarctic SIE
621 values that are excessive relative to observed, evidenced by fewer models falling outside of the gray
622 shading on the righthand side of Fig. 8b. However, fewer CMIP6 models fall within the observational
623 range than CMIP5, and the overwhelming majority of the models are still simulating much too limited
624 annual SIE.

625 There are some clear outliers among the models, including the MIROC6, MIROC-ES2L, EC-
626 Earth3-Veg, MPI-ESM-1-2-HAM, MPI-ESM1-2-LR, HadGEM3-GC31-MM, and INM-CM4-8 models
627 which simulate extremely low maximum SIE values (Table 3; Fig. 8a). The MIROC-ES2L and
628 MIROC6 models have lower annual SIE values than any CMIP5 model (Fig. 8b) skewing the CMIP6
629 MMM toward lower values. Models with exceptionally low annual SIE values tend to have very large

630 upper and interior ocean temperature biases ($>3^{\circ}\text{C}$, Fig. 6a, Fig. S3, Fig. S5, Fig. 7a), likely explaining
631 the lack of seasonal sea ice.

632 The magnitude of the standard deviation (2σ) of the annual means over the 20-yr period differ
633 widely across models and in most cases are often much larger than observed (Fig. 8b), suggesting the
634 internal variability differs markedly across models and is different from that observed. For example,
635 compare the large variability in GISS-E2-1H, UK-ESM1-0-LL, HadGEM3-GC31-LL, IPSL-CM6A-
636 LR, and BCC-ESM1 to that of NorESM2-LM, MIROC-ES2L, and MIROC6. This brings up the
637 question of the representativeness of the time period chosen when making assessments of a model's
638 performance across Antarctic sea ice metrics since there seems to be a significant contribution of the
639 model's internal variability to the performance, as was the case in previous generations (Mahlstein et
640 al. 2013; Deser et al. 2010). Additionally, the presence of any large open-ocean polynyas (see
641 discussion in Section 3a) in a model's simulation would significantly impact the overall sea ice
642 representation. The CMIP5 and CMIP6 inter-model spread for all SIE metrics in Table 3 is much larger
643 than the observed interannual variability over this time period. Overall, it appears that the
644 representation of Antarctic SIE still remains a systematic bias from CMIP5 to CMIP6.

645 **4. Summary and conclusions**

646 The SO is a dynamically complex region where the global ocean water masses converge and the
647 world's strongest current, the ACC, exchanges properties between basins. Strong wind-driven surface
648 divergence drives deep waters to the ocean surface on the southern side of the ACC, resulting in
649 strongly sloped isopycnals that act as a vehicle for the exchange of deep ocean properties with that of
650 the upper ocean and overlying atmosphere. A portion of the circumpolar deep water (CDW) that is
651 upwelled along these steeply sloped isopycnals is transformed into denser Antarctic Bottom Water
652 (AABW) and exported northward, filling the abyssal ocean. Buoyancy gained at the ocean surface
653 through the incorporation of precipitation, glacial runoff, and sea ice melt, transforms the remaining

654 upwelled CDW into lighter intermediate and mode waters that are exported northward in the upper
655 ocean subtropical gyres. The ACC and intense water mass transformations that characterize the SO are
656 intimately coupled together, and these complicated dynamics are directly tied to the exchange of heat
657 and carbon with the atmosphere and the export of nutrients to the rest of the global ocean.

658 Lack of historical observations coupled with complex and not fully understood inter-connected
659 processes tied to the ocean eddy field, surface wind stress forcing, air-sea heat and freshwater fluxes
660 impacted by both the ocean circulation and overlying atmospheric properties, properties of upwelled
661 water, and the destruction and transformation of seasonal sea ice make the SO very difficult to
662 accurately represent in climate models. Since these dynamics are highly coupled, a well-represented
663 SO requires accuracy to be achieved in the atmosphere, ocean, and sea ice model subcomponents and
664 their exchanges since errors in any of these regimes can propagate into the other. An important example
665 discussed here is that in previous model generations, errors in representation of Southern Hemisphere
666 clouds and cloud properties in the atmospheric model led to significant errors in the SO upper ocean
667 thermal structure. Given that the surface temperature gradient is tightly coupled to the position and
668 strength of the surface wind stress forcing, this can propagate into surface forcing biases. Additionally,
669 as was shown here, near-surface property biases directly impact the ability to accurately represent
670 Antarctic SIE, where models with an exceptionally warm SO simulated extremely low SIE. The upper
671 ocean properties can also be impacted internally by poorly represented water mass properties, such as
672 too warm NADW or abyssal warming due to climate drift.

673 Despite these challenges, the analyses presented here in conjunction with those performed on
674 ensembles of CMIP3 and CMIP5 models, show some consistent improvements across generations. We
675 have focused on observable metrics that have been analyzed in previous model ensembles as well as
676 included additional metrics pertaining to characterizing the wind stress forcing. The results suggest the
677 following regarding model performance:

678 1) *The simulation of the strength of the ACC has improved from CMIP3 to CMIP6, with*
679 *modeled ACC strength converging toward the magnitude of observed net flow through the*
680 *Drake Passage as estimated by the cDrake array.* While there are no CMIP6 models that
681 exhibit unrealistically strong transports like previous generations, there are still several models
682 producing exceptionally weak transports ($>7\sigma$ outside of the observational uncertainty). An area
683 identified as a need for improvement is in the unrealistic multi-decadal variability observed in
684 the ACC transport for a number of the CMIP6 models. This behavior may be linked to the
685 presence of unrealistic “superpolynya events” in the pre-industrial control simulations
686 associated with quasiperiodic episodes of intense open-ocean convection which impact the
687 interior ocean density structure and thus the ACC. The underlying mechanisms causing these
688 events need to be further studied to improve the simulations. The coarse resolution models ($\sim 1^\circ$
689 or coarser) all employ various parameterization schemes and coefficient magnitudes for
690 subgrid-scale ocean mixing and this likely plays a role in explaining such diversity in the
691 vertical and horizontal structure of the ACC found across models. This resolution and
692 parameterization choice dependence needs to be explored in detail across models of varying
693 resolution.

694 2) *All metrics pertaining to the surface wind stress forcing have improved.* Notable
695 improvements are found in the strength and position of the zonally-averaged westerly wind
696 stress maximum ($\tau_{max}^{(x)}$) relative to the ERA5 reanalysis product. The persistent systematic bias
697 of equatorward winds in previous model generations has improved. These improvements have
698 yielded a much more realistic pattern of wind stress curl over the SO and a narrowing of spread
699 across the model ensemble. Noting that the magnitude and location of where wind-driven
700 surface divergence pulls deep carbon-rich water to the surface is potentially important for the
701 SO carbon budget, we have further characterized the wind stress forcing in the CMIP6 models

702 by computing the magnitude and location of minimum wind stress curl over the SO (magnitude
703 of maximum Ekman upwelling). These two metrics show the largest inter-model spread relative
704 to the other wind stress metrics considered; a follow up study will identify if this plays a role in
705 explaining model spread in biogeochemical performance here. Relative to CMIP3 and CMIP5,
706 a stronger correlation, although still not statistically significant, is found for the relationship
707 between ACC strength and $\tau_{max}^{(x)}$. Previous generations had many models that exhibited extreme
708 biases in their temperature and salinity gradients across the ACC that allowed for compensation
709 for errors in the surface wind stress forcing. These errors are not as prevalent in the CMIP6
710 ensemble.

711 **3) *The upper ocean remains biased too fresh and too warm relative to observed. There has***
712 ***not been a clear improvement in ensemble performance relative to CMIP3 and CMIP5.*** There
713 are still several models with exceptionally warm upper SOs with errors that translate into very
714 poor representations of Antarctic SIE. All CMIP6 models are generally too fresh in the upper
715 ocean. These errors likely translate into biases in the ventilated layers of the SO; a detailed
716 water mass analysis similar to Sallée et al. (2013b) should be performed to identify these biases
717 and compare them to that of previous generations and to understand how they impact heat and
718 carbon fluxes and storage.

719 **4) *The representation of the difference in density across the latitudes of the ACC has***
720 ***improved in CMIP6.*** Considering the simulated zonally-averaged, full-depth averaged
721 difference in potential density referenced to the surface ($\Delta\rho$) across the ACC, CMIP6 models
722 have improved in performance relative to that observed. The majority of models (71%) simulate
723 a $\Delta\rho$ within a 25% error margin of the WOA18 value. Two models yield a $\Delta\rho$ that is too weak
724 and eight models yield a $\Delta\rho$ much greater than the WOA18 value. It still remains fairly

725 common for models to achieve an accurate $\Delta\rho$ as a result of compensating errors in the
726 temperature and salinity structure.

727 **5) Antarctic sea ice extent (SIE) representation remains a systematic bias from CMIP5 to**
728 **CMIP6.** A few models appear to be performing worse than their CMIP5 predecessors, while
729 others have improved. While SIE is well observed relative to other metrics related to the sea ice
730 simulation such as sea ice volume, it might not be the best metric for overall model
731 performance. Thus, a detailed analysis of the regional distribution and other sea ice
732 characteristics should be performed to truly assess model performance. The correlation found
733 between the density gradient across the ACC and the minimum Antarctic SIE, highlights the
734 need for investigating how the sea ice simulation impacts the density structure and climatically
735 relevant properties in the SO such as carbon and heat storage. Additionally, there may be
736 important links in CMIP6 between Antarctic sea ice representation in historical simulations and
737 21st century projected change in surface temperature, precipitation, and westerly jet position as
738 was the case in previous generations (Flato 2004; Bracegirdle et al. 2015; Bracegirdle et al.
739 2018). If these emergent constraint relationships still exist, a general lack of improvement from
740 CMIP5 to CMIP6 and wide inter-model spread in simulated SIE suggests this source of
741 projection uncertainty may be a large contributor to the model spread in the trajectory of
742 Southern Hemisphere climate under 21st century forcing.

743 The analysis presented here provides critical information on improved and existing biases in
744 observable properties in the SO for the climate models that will provide projections of the climate
745 system for the Intergovernmental Panel on Climate Change's Sixth Assessment Report (IPCC-AR6).
746 Given the dominant role that the SO plays relative to other ocean basins in the oceanic storage of heat
747 and carbon, it is important for climate models to represent this region well in order to provide
748 meaningful simulations of transient climate change. While models have generally improved across

749 many metrics, the remaining biases associated with the temperature and salinity structure of the SO and
750 sea ice representation may have serious implications for climate projections. The impacts of these
751 biases on simulated ocean heat and carbon storage requires a detailed assessment.

752 In the conclusions outlined above, we have highlighted several paths forward for additional
753 analysis of the simulations here including extending such analysis to the higher resolution simulations,
754 determining the dependence of the representation of surface momentum forcing on the heat and carbon
755 budget, carrying out a detailed water mass analysis, and performing a detailed evaluation of the role
756 that sea ice representation plays in determining the SO density structure and in water mass
757 transformations. Additionally, while we examine a 20-yr average period to be consistent with the time
758 period evaluated in the CMIP3 and CMIP5 studies, studies should be performed to assess whether
759 these models capture observed historical trends such as observed changes in the mid-latitude westerlies
760 over the SO. We are currently working to extend this assessment to the results of the 21st century SSP
761 experiments to investigate how these properties change under continued warming and if robust model
762 agreement is found. In future analyses, the performance across the metrics presented here can
763 potentially be used to develop model weighting schemes to provide a constraint on the uncertainty of
764 global climate projections.

765

766 *Acknowledgments*

767 We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling
768 responsible for CMIP. We thank the modeling groups listed in Table 1 for producing and making
769 available their output, the Earth System Grid Federation (ESGF) for archiving the output and providing
770 access, and the multiple funding agencies who support CMIP and ESGF. For CMIP, the U.S.
771 Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides
772 coordinating support and led development of software infrastructure in partnership with the Global

773 Organization for Earth System Science Portals. We acknowledge the use of the PyFerret program from
774 NOAA's Pacific Marine Environmental Laboratory for analysis and graphics
775 (<http://ferret.pmel.noaa.gov/Ferret/>). We thank John Krasting at the Geophysical Fluid Dynamics
776 Laboratory for assistance in obtaining the native grid GFDL-ESM4 output, Teresa Chereskin at Scripps
777 Institution of Oceanography and Kathleen Donohue at the University of Rhode Island for discussion
778 regarding observations of the ACC transport, and Brandon Bui at the University of Arizona for
779 assistance in downloading data. We thank the editor and three anonymous reviewers for their
780 comments and suggestions during the review of this manuscript. This work was funded by NSF's
781 Southern Ocean Carbon and Climate Observations and Modeling (SOCCOM) Project under NSF
782 Award PLR-1425989, with additional support from NOAA and NASA. Logistical support for
783 SOCCOM in the Antarctic was provided by the U.S. NSF through the U.S. Antarctic Program. RB was
784 supported by the U.S. EPA Assistance Agreement FP-91780701-0. This publication has not been
785 reviewed by the EPA and the views expressed herein are solely those of the authors. We also thank the
786 Thomas R. Brown foundation at the University of Arizona for supporting JR and PJ. HH and PH were
787 supported by the Met Office Hadley Centre Climate Programme funded by BEIS and Defra
788 (GA01101).

789

790 *Data Availability Statement*

791 All of the CMIP6 model output analyzed in this manuscript is openly available from the Earth System
792 Grid Federation (ESGF) CMIP6 archive: <https://esgf-node.llnl.gov/projects/cmip6/>. In the case of the
793 GFDL-ESM4 model, native grid output for the uo and umo variables analyzed in this manuscript are
794 available upon request from the Geophysical Fluid Dynamics Laboratory (GFDL). The B-SOSE
795 Iteration 133 output used in this analysis are available at Scripps Institution of Oceanography:
796 http://sose.ucsd.edu/BSOSE6_iter133_solution.html. The WOA18 output is openly available from

797 NOAA's National Centers for Environmental Information: <https://www.nodc.noaa.gov/OC5/woa18/>.
798 The European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5 atmospheric reanalysis
799 product is openly available from: (<https://cds.climate.copernicus.eu/>). The Antarctic sea ice extent data
800 is openly available from the National Snow and Ice Data Center (NSIDC):
801 <https://nsidc.org/data/G02135/versions/3>.

802

803 **References.**

- 804 Abernathey, R., I. Cerovecki, P. R. Holland, E. Newsom, M. Mazloff, and L. D. Talley, 2016:
805 Water-mass transformation by sea ice in the upper branch of the Southern Ocean
806 overturning. *Nature Geosci.*, **9**, 596–601, doi:10.1038/ngeo2749.
- 807 Adcroft, A., and J.-M. Campin, 2004: Rescaled height coordinates for accurate representation of
808 free-surface flows in ocean circulation models. *Ocean Model.*, **7**, 269–284, doi:10.1016/J.OCE-
809 MOD.2003.09.003.
- 810 Adcroft, A., and Coauthors, 2019: The GFDL global ocean and sea ice model OM4.0: Model
811 description and simulation features. *J. Adv. in Model. Earth Sys.*, **11**, 3167– 3211,
812 <https://doi.org/10.1029/2019MS001726>.
- 813 Beadling, R.L., J.L. Russell, R.J. Stouffer, P.J. Goodman, and M. Mazloff, 2019: Assessing the
814 quality of Southern Ocean circulation in CMIP5 AOGCM and Earth system model
815 simulations. *J. Climate*, **32**, 5915–5940, <https://doi.org/10.1175/JCLI-D-19-0263.1>
- 816 Bracegirdle, T. J., and G. J. Marshall, 2012: The reliability of Antarctic tropospheric pressure
817 and temperature in the latest global reanalyses. *J. Climate*, **25**, 7138–7146,
818 doi:10.1175/JCLI-D-11-00685.1.
- 819 Bracegirdle T., E. Shuckburgh, J-B. Sallée, Z. Wang, A. J. S. Meijers, N. Bruneau, and L.
820 Wilcox, 2013: Assessment of surface winds over the Atlantic, Indian and Pacific Ocean

821 sectors of the Southern Hemisphere in CMIP5 models: historical bias, forcing response,
822 and state dependency. *J. Geophys. Res.*, **118**, 547–562, doi:10.1002/jgrd.50153.

823 Bracegirdle, T. J., D. B. Stephenson, J. Turner, and T. Phillips, 2015: The importance of sea ice area
824 biases in 21st century multimodel projections of Antarctic temperature and precipitation,
825 *Geophys. Res. Lett.*, **42**, 10,832– 10,839, doi:10.1002/2015GL067055.

826 Bracegirdle, T.J., P. Hyder, and C.R. Holmes, 2018: CMIP5 diversity in Southern Westerly jet
827 projections related to historical sea ice area: strong link to strengthening and weak link to shift.
828 *J. Climate*, **31**, 195–211, <https://doi.org/10.1175/JCLI-D-17-0320.1>.

829 Campbell, E.C., E. A. Wilson, G. W. K. Moore, S. C. Riser, C. E. Brayton, M. R. Mazloff, and L. D.
830 Talley, 2019: Antarctic offshore polynyas linked to Southern Hemisphere climate anomalies.
831 *Nature*, **570**, 319–325, doi:10.1038/s41586-019-1294-0.

832 Carsey, F. D., 1980: Microwave observation of the Weddell polynya. *Mon. Weath. Rev.* **108**,
833 2032–2044.

834 Ceppi, P., Y-T., Hwang, D. M. W. Frierson, and D. L. Hartman, 2012: Southern Hemisphere jet
835 latitude biases in CMIP5 models linked to short wave cloud forcing, *Geophys. Res. Lett.*, **39**,
836 L19708.

837 Cerovečki, I., A.J. Meijers, M.R. Mazloff, S.T. Gille, V.M. Tamsitt, and P.R. Holland, 2019: The
838 effects of enhanced sea ice export from the Ross Sea on recent cooling and freshening of
839 the southeast Pacific. *J. Climate*, **32**, 2013–2035, [https://doi.org/10.1175/JCLI-D-18-](https://doi.org/10.1175/JCLI-D-18-0205.1)
840 0205.1.

841 Chereskin, T. K., K. A. Donohue, and D. R. Watts, 2012: cDrake: Dynamics and transport of the
842 Antarctic Circumpolar Current in Drake Passage. *Oceanography*, **25**, 1FF34–135,
843 doi:<https://doi.org/10.5670/oceanog.2012.86>.

844 Chidichimo, M. P., K. A. Donohue, D. R. Watts, and K. L. Tracey, 2014: Baroclinic transport

845 time series of the Antarctic Circumpolar Current measured in Drake Passage. *J. Phys.*
846 *Oceanogr.*, **44**, 1829–1853, doi:10.1175/JPO-D-13-071.1.

847 Connolley, W.M., and T. J. Bracegirdle, 2007: An Antarctic assessment of IPCC AR4 coupled
848 models. *Geophys. Res. Letts.*, **34**, L22505, doi:10.1029/2007GL031648.

849 Cunningham, S., S. Alderson, B. King, and M. Brandon, 2003: Transport and variability of the
850 Antarctic Circumpolar Current in Drake Passage. *J. Geophys. Res.*, **108**, 8084,
851 doi:10.1029/2001JC001147.

852 Danabasoglu, G., and Coauthors, 2012: The CCSM4 ocean component. *J. Climate*, **25**, 1361–1389.
853 doi: 10.1175/JCLI-D-11-00091.1.

854 Danabasoglu, G., and Coauthors, 2020: The Community Earth System Model Version 2 (CESM2).
855 *J. Adv. Model. Earth Sy.*, **12**, e2019MS001916. doi: <https://doi.org/10.1029/2019MS001916>.

856 de Lavergne, C., J. Palter, E. Galbraith, R. Bernardello, and I. Marinov, 2014: Cessation of deep
857 convection in the open Southern Ocean under anthropogenic climate change. *Nature Clim.*
858 *Change*, **4**, 278–282, doi:10.1038/nclimate2132.

859 Deser, C., A. Phillips, V. Bourdette, and H. Teng, 2010: Uncertainty in climate change
860 projections: The role of internal variability, *Clim. Dyn.*, **38**, 527–546.

861 Donohue, K. A., K. L. Tracey, D. R. Watts, M. P. Chidichimo, and T. K. Chereskin, 2016: Mean
862 Antarctic Circumpolar Current transport measured in Drake Passage. *Geophys. Res.*
863 *Letts.*, **43**, 11,760–11,767, doi:10.1002/2016GL070319.

864 Dufour, C.O., A.K. Morrison, S.M. Griffies, I. Frenger, H. Zanowski, and M. Winton, 2017:
865 Preconditioning of the Weddell Sea polynya by the ocean mesoscale and dense water over
866 flows. *J. Climate*, **30**, 7719–7737, <https://doi.org/10.1175/JCLI-D-16-0586.1>.

867 Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor,
868 2016: Overview of the Coupled Model Intercomparison Project phase 6 (CMIP6)

869 experimental design and organization, *Geosci. Model Dev.*, **9**, 1937–1958,
870 <https://doi.org/10.5194/gmd-9-1937-2016>.

871 Eyring, V., and Coauthors, 2019: Taking climate model evaluation to the next level. *Nat. Clim.*
872 *Change.*, **9**, 102-110, <https://doi.org/10.1038/s41558-018-0355-y>.

873 Farneti, R., and Coauthors, 2015: An assessment of Antarctic Circumpolar Current and Southern
874 Ocean meridional overturning circulation during 1958-2007 in a suite of interannual
875 CORE-II simulations. *Ocean Model.*, **93**, 84-120, doi:10.1016/j.ocemod.2015.07.009.

876 Fetterer, F., K. Knowles, W. N. Meier, M. Savoie, and A. K. Windnagel, 2017: Updated daily.
877 Sea Ice Index, Version 3. [January1986 – December 2005, monthly data]. Boulder, Colorado
878 USA. NSIDC: National Snow and Ice Data Center. doi: <https://doi.org/10.7265/N5K072F8>.
879 [November 17th, 2019].

880 Firing, Y. L., T. K. Chereskin, and M. R. Mazloff, 2011: Vertical structure and transport of the
881 Antarctic Circumpolar Current in the Drake Passage from direct velocity measurements.
882 *J. Geophys. Res.*, **116**, C08015, doi: <https://doi.org/10.1029/2011JC006999>.

883 Flato, G.M., and Participating CMIP Modelling Groups Climate Dynamics, 2004: Sea-ice and its
884 response to CO₂ forcing as simulated by global climate models. *Clim. Dyn.*, **23**, 229-241,
885 <https://doi.org/10.1007/s00382-004-0436-7>.

886 Flato, G., and Coauthors, 2013: Evaluation of climate models. In: Climate change 2013: The
887 physical science Basis. Contribution of Working Group I to the Fifth Assessment Report
888 of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner,
889 M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)].
890 Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

891 Frölicher, T.L., J.L. Sarmiento, D.J. Paynter, J.P. Dunne, J.P. Krasting, and M.

892 Winton, 2015: Dominance of the Southern Ocean in Anthropogenic Carbon and Heat Uptake in
893 CMIP5 Models. *J. Climate*, **28**, 862–886, <https://doi.org/10.1175/JCLI-D-14-00117.1>.

894 Galbraith, E. D., A. Gnanadesikan, J. P. Dunne, and M. R., Hiscock, 2010: Regional impacts of
895 iron-light colimitation in a global biogeochemical model. *Biogeosciences*, **7**, 1043–1064,
896 <https://doi.org/10.5194/bg-7-1043-2010>, 2010.

897 Gent, P. R., W. G. Large, and F. O. Bryan, 2001: What sets the mean transport through the
898 Drake Passage? *J. Geophys. Res.*, **106**, 2693–2712, doi: <https://doi.org/10.1029/2000JC900036>.

899 Golaz, J.-C., and Coauthors, 2019: The DOE E3SM Coupled Model Version 1: overview and
900 evaluation at standard resolution. *J. Adv. Model. Earth Sy.*, **11**, 2089–2129.
901 doi: <https://doi.org/10.1029/2018MS001603>.

902 Gordon, A.L., 1978: Deep Antarctic Convection West of Maud Rise. *J. Phys. Oceanogr.*, **8**, 600–612,
903 [https://doi.org/10.1175/1520-0485\(1978\)008<0600:DACWOM>2.0.CO;2](https://doi.org/10.1175/1520-0485(1978)008<0600:DACWOM>2.0.CO;2).

904 Griesel, A., M. R. Mazloff, S. T. Gille, 2012: Mean dynamic topography in the Southern Ocean:
905 Evaluating Antarctic circumpolar transport. *J. Geophys. Res.*, **117**, C01020,
906 doi:10.1029/2011JC007573.

907 Griffies, S.M., and Coauthors, 2011: The GFDL CM3 Coupled Climate Model: Characteristics
908 of the Ocean and Sea Ice Simulations. *J. Climate*, **24**, 3520–3544,
909 <https://doi.org/10.1175/2011JCLI3964.1>.

910 Hanawa, K., and L. D. Talley, 2001: Mode waters. *Ocean Circulation and Climate*, G. Siedler,
911 J. Church, and J. Gould, Eds., Academic Press, 373–386.

912 Haumann, F., N. Gruber, M. Münnich, I. Frenger, and S. Kern, 2016: Sea-ice transport driving
913 Southern Ocean salinity and its recent trends. *Nature*, **537**, 89–92, doi:10.1038/nature19101.

914 Held, I. M., and Coauthors, 2019: Structure and performance of GFDL's CM4.0 climate model.
915 *Jour. of Adv. in Model. Earth Sys.*, **11**, <https://doi.org/10.1029/2019MS001829>.

916 Heuzé, C., K. J. Heywood, D. P. Stevens, and J. K. Ridley, 2013: Southern Ocean bottom water
917 characteristics in CMIP5 models, *Geophys. Res. Lett.*, **40**, 1409– 1414,
918 doi:10.1002/grl.50287.

919 Hyder, P., and Coauthors, 2018: Critical Southern Ocean climate model biases traced to
920 atmospheric model cloud errors. *Nat. Commun.* **9**, doi: [https://doi.org/10.1038/s41467-](https://doi.org/10.1038/s41467-018-05634-2)
921 [018-05634-2](https://doi.org/10.1038/s41467-018-05634-2).

922 Ivanova, D.P., P.J. Gleckler, K.E. Taylor, P.J. Durack, and K.D. Marvel, 2016: Moving beyond
923 the Total Sea Ice Extent in Gauging Model Biases. *J. Climate*, **29**, 8965–8987, doi:
924 <https://doi.org/10.1175/JCLI-D-16-0026.1>.

925 Jungclaus, J. H., and Coauthors, 2013: Characteristics of the ocean simulations in MPIOM, the ocean
926 component of the MPI- Earth system model, *J. Adv. Model. Earth Syst.*, **5**, 422– 446,
927 doi:10.1002/jame.20023.

928 Kidston, J., and E. P. Gerber, 2010: Intermodel variability of the poleward shift of the austral jet
929 stream in CMIP3 integrations linked to biases in 20th century climatology. *Geophys. Res.*
930 *Letts.*, **37**, L09708, doi:10.1029/2010GL042873.

931 Kiss, A. E., and Coauthors, 2020: ACCESS-OM2 v1.0: a global ocean-sea ice model at three
932 resolutions. *Geosci. Model. Dev.*, **13**, 401 – 441. doi: <https://doi.org/10.5194/gmd-13-401-2020>.

933 Knutti, R., 2010: The end of model democracy? *Clim. Change*, **102**, 395–404.

934 Knutti, and Coauthors, 2017: A climate model projection weighting scheme accounting for
935 performance and interdependence. *Geophys. Res. Lett.* **44**, 1909–1918.

936 Kuhlbrodt, T., R. S. Smith, Z. Wang, and J. M. Gregory, 2012: The influence of eddy
937 parameterizations on the transport of the Antarctic circumpolar current in coupled climate mod-
938 els. *Ocean Model.* **52 – 53**, 1 – 8, doi: <https://doi.org/10.1016/j.ocemod.2012.04.006>.

939 Kuhlbrodt, T., and Coauthors, 2018: The low-resolution version of HadGEM3 GC3.1: development

940 and evaluation for global climate. *J. Adv. Model. Earth Sy.*, **10**, 2865-2888, doi:
941 <https://doi.org/10.1029/2018MS001370>.

942 Lenn, Y. D., T. K. Chereskin, J. Sprintall, and E. Firing, 2007: Mean jets, mesoscale variability
943 and eddy momentum fluxes in the surface- layer of the Antarctic Circumpolar Current in
944 Drake Passage, *J. Mar. Res.*, **65**, 27– 58, <https://doi.org/10.1357/002224007780388694>.

945 Locarnini, R. A., and Coauthors, 2018: World Ocean Atlas 2018, Volume 1: Temperature.
946 A. Mishonov Technical Ed.; NOAA Atlas NESDIS 81, 52 pp.

947 Mahlstein, I., P. R. Gent, and S. Solomon, 2013: Historical Antarctic mean sea ice area, sea ice
948 trends, and winds in CMIP5 simulations. *J. Geophys. Res.*, **118**, 5105-5110,
949 doi:10.1002/jgrd.50443

950 Mazloff, M. R., P. Heimbach, and C. Wunsch, 2010: An Eddy-Permitting Southern Ocean State
951 Estimate. *J. Phys. Oceanogr.*, **40**, 880 – 899, doi: 10.1175/2009JPO4236.1.

952 Meijers A., E. Shuckburgh, N. Bruneau, J. B. Sallée, T. Bracegirdle, and Z. Wang, 2012:
953 Representation of the Antarctic circumpolar current in the CMIP5 climate models and
954 future changes under warming scenarios. *J. Geophys. Res.*, **117**, 547–562,
955 doi:10.1029/2012JC008412.

956 Meijers, A. J. S., 2014: The Southern Ocean in the Coupled Model Intercomparison Project
957 phase 5. *Philos. Trans. A. Math Phys. Eng. Sci.*, **372**,
958 doi: <https://doi.org/10.1098/rsta.2013.0296>.

959 Meredith, M., and Coauthors, 2019: Polar Regions. In: *IPCC Special Report on the Ocean and*
960 *Cryosphere in a Changing Climate* [H.-O. Pörtner, D.C. Roberts, V. Masson-Delmotte,
961 P. Zhai, M. Tignor, E. Poloczanska, K. Mintenbeck, A. Alegría, M. Nicolai, A. Okem, J.
962 Petzold, B. Rama, N.M. Weyer (eds.)]. In press.

963 Munday, D. R., H. L. Johnson, and D. P. Marshall, 2013: Eddy saturation of equilibrated circumpolar

964 currents. *J. Phys. Oceanogr.*, **43**, 507 – 532.

965 Olbers, D., D. Borowski, C. Völker, and J.-O. Wölf, 2004: The dynamical balance, transport and
966 circulation of the Antarctic Circumpolar Current. *Antarct. Sci.* **16**, 439–470.
967 doi: 10.1017/S0954102004002251.

968 Parkinson, C. L., K. Y. Vinnikov, and D. J. Cavalieri, 2006: Evaluation of the simulation of the
969 annual cycle of Arctic and Antarctic sea ice coverages by 11 major global climate
970 models. *J. Geophys. Res.*, **111**, C07012, doi:10.1029/2005JC003408.

971 Pellichero, V., J.-B., Sallée, C. C. Chapman, and S. M. Downes, 2018: The Southern Ocean
972 meridional overturning in the sea-ice sector is driven by freshwater fluxes. *Nat Commun.*,
973 **9**, 1789, doi:10.1038/s41467-018-04101-2.

974 Ragen, S., M. Pradal, and A. Gnanadesikan, 2020: The impact of parameterized lateral mixing on the
975 Antarctic Circumpolar Current in a coupled climate model. *J. Phys. Oceanogr.*, **50**, 965–982,
976 doi: <https://doi.org/10.1175/JPO-D-19-0249.1>.

977 Rintoul, S.R., 2018: The global influence of localized dynamics in the Southern Ocean. *Nature*, **558**,
978 209–218, doi: <https://doi.org/10.1038/s41586-018-0182-3>.

979 Roach, L. A., S. M. Dean, and J. A. Renwick, 2018: Consistent biases in Antarctic sea ice
980 concentration simulated by climate models. *The Cryosphere*, **12**, 365-383,
981 <https://doi.org/10.5194/tc-12-365-2018>.

982 Roemmich, D., J. Church, J. Gilson, D. Monselesan, P. Sutton, and S. Wijffels, 2015: Unabated
983 planetary warming and its structure since 2006. *Nat. Clim. Change.*, **5**, 240 – 245,
984 doi:10.1038/NCLIMATE2513.

985 Russell, J.L., R.J. Stouffer, and K.W. Dixon, 2006a: Intercomparison of the Southern Ocean
986 Circulations in IPCC Coupled Model Control Simulations. *J. Climate*, **19**, 4560–4575, doi:
987 <https://doi.org/10.1175/JCLI3869.1>.

988 Russell, J.L., K.W. Dixon, A. Gnanadesikan, R.J. Stouffer, and J.R. Toggweiler, 2006b: The
989 Southern Hemisphere Westerlies in a Warming World: Propping Open the Door to the Deep
990 Ocean. *J. Climate*, **19**, 6382–6390, doi: <https://doi.org/10.1175/JCLI3984.1>.

991 Sallée J-B., E. Shuckburgh, N. Bruneau, A. J. S. Meijers, Z. Wang, and T. Bracegirdle, 2013a
992 Assessment of the Southern Ocean mixed-layer depth in CMIP5 models: historical bias
993 and forcing response. *J. Geophys. Res.*, **118**, 1845–1862, doi:10.1002/jgrc.20157.

994 Sallée, J.- B., E. Shuckburgh, N. Bruneau, A. J. S. Meijers, T. J. Bracegirdle, Z. Wang, and T.
995 Roy, 2013b: Assessment of Southern Ocean water mass circulation and characteristics in
996 CMIP5 models: Historical bias and forcing response. *J. Geophys. Res. Oceans*, **118**,
997 1830–1844, doi:10.1002/jgrc.20135.

998 Schneider, D.P. and D.B. Reusch, 2016: Antarctic and Southern Ocean surface temperatures in
999 CMIP5 models in the context of the surface energy budget. *J. Climate*, **29**, 1689–1716,
1000 <https://doi.org/10.1175/JCLI-D-15-0429.1>.

1001 Sen Gupta, A., A. Santoso, A.S. Taschetto, C.C. Ummenhofer, J. Trevena, and M.H. England,
1002 2009: Projected Changes to the Southern Hemisphere Ocean and Sea Ice in the IPCC
1003 AR4 Climate Models. *J. Climate*, **22**, 3047–3078,
1004 doi:<https://doi.org/10.1175/2008JCLI2827.1>.

1005 Shu, Q., Z. Song, F. Qiao, 2015: Assessment of sea ice simulations in the CMIP5 models. *The*
1006 *Cryosphere*, **9**, 399-409, doi:<https://doi.org/10.5194/tc-9-399-2015>.

1007 Stouffer, R.J., V. Eyring, G.A. Meehl, S. Bony, C. Senior, B. Stevens, and K.E. Taylor, 2017: CMIP5
1008 scientific gaps and recommendations for CMIP6. *Bull. Amer. Meteor. Soc.*, **98**,
1009 95–105, <https://doi.org/10.1175/BAMS-D-15-00013.1>.

1010 Swart, N. C., and J. C. Fyfe, 2012: Observed and simulated changes in the Southern Hemisphere
1011 surface westerly wind- stress. *Geophys. Res. Lett.*, **39**, L16711, doi:10.1029/2012GL052810.

1012 Talley, L. D., and Coauthors, 2016: Changes in ocean heat, carbon content, and ventilation: a
1013 review of the first decade of GO-SHIP global repeat hydrography. *Ann. Rev. of Mar. Sci.*,
1014 **8**, 185-215, doi: <https://doi.org/10.1146/annurev-marine-052915-100829>.

1015 Taylor, K.E., R.J. Stouffer, and G.A. Meehl, 2012: An Overview of CMIP5 and the Experiment
1016 Design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, doi:[https://doi.org/10.1175/BAMS-D-11-](https://doi.org/10.1175/BAMS-D-11-00094.1)
1017 [00094.1](https://doi.org/10.1175/BAMS-D-11-00094.1).

1018 Turner, J., T. Bracegirdle, T. Phillips, G. J. Marshall, and J. S. Hosking, 2012: An initial
1019 assessment of Antarctic sea ice extent in the CMIP5 models, *J. Climate*, **26**, 1473–1484.

1020 Verdy, A. and M. R. Mazloff, 2017: A data assimilating model for estimating Southern Ocean
1021 Biogeochemistry. *J. Geophys. Res. Oceans*, **122**, 6968–6988, doi:10.1002/2016JC012650.

1022 Waugh, D.W., A. McC. Hogg, P. Spence, M.H. England, and T.W. Haine, 2019: Response of
1023 Southern Ocean ventilation to changes in midlatitude westerly winds. *J. Climate*, **32**,
1024 5345–5361, <https://doi.org/10.1175/JCLI-D-19-0039.1>

1025 Whitworth, T., 1983: Monitoring the transport of the Antarctic Circumpolar Current at Drake
1026 Passage. *J. Phys. Oceanogr.*, **13**, 2045–2057.

1027 Whitworth, T., W. Nowlin, and S. Worley, 1982: The net transport of the Antarctic Circumpolar
1028 Current through Drake Passage. *J. Phys. Oceanogr.*, **12**, 960–971.

1029 Whitworth, T., and R. Peterson, 1985: Volume transport of the Antarctic Circumpolar Current
1030 from bottom pressure measurements. *J. Phys. Oceanogr.*, **15**, 810–816.

1031 Yukimoto, S., and Coauthors, 2019: The Meteorological Research Institute Earth System Model
1032 Version 2.0, MRI-ESM2.0: description and basic evaluation of the physical component.
1033 *J. Meteorol. Soc. JPN*, **97**, 931-965. Doi: <https://doi.org/10.2151/jmsj.2019-051>.

1034 Zweng, M. M., and Coauthors, 2018: World Ocean Atlas 2018, Volume 2: Salinity.
1035 A. Mishonov Technical Ed.; NOAA Atlas NESDIS 82, 50 pp.

1036 **TABLE 1.** Details of the ocean model components in the CMIP6 models used in this study. Details of the ocean component, resolu-
1037 tion, and vertical levels were obtained from the header information in the netcdf files used for the analyses. Information regarding the
1038 eddy-induced advection coefficient was obtained from ES-DOCs (<https://search.es-doc.org/>), through personal communication with
1039 the modeling centers, or from model documentation literature, as noted by the subscript next to each entry. V corresponds to a coeffi-
1040 cient that is either 2D/3D/time-varying and F corresponds to a fixed coefficient. If the required information could not be obtained or
1041 confirmed at the time of publication, it is left blank (--). The ocean vertical coordinates are defined as follows: z, traditional depth co-
1042 ordinates; σ_2 , isopycnal vertical coordinates; z^* , rescaled geopotential vertical coordinate for better representation of free-surface vari-
1043 ations (Adcroft and Campin 2004); sigma, terrain-following coordinates; hybrid $\sigma_2 - z^*$, isopycnal coordinates in the interior ocean
1044 and a z^* coordinate in the mixed layer (Adcroft et al. 2019); hybrid z - σ_2 , isopycnal coordinates in the interior ocean and z coordinates
1045 in the mixed layer; hybrid z- σ_2 -sigma, z coordinates in the mixed layer, isopycnal coordinates in the open stratified ocean, and sigma
1046 coordinates in shallow coastal regions; hybrid z - sigma, sigma coordinates between the sea surface and a fixed geopotential depth
1047 (~50 m) in the upper ocean and z coordinates below this depth. The historical experiments span 1850 to 2014 and are forced by ob-
1048 served changes in atmospheric composition due to anthropogenic and natural sources over the entire historical period. These forcings
1049 are updated from the CMIP5 historical forcings (Taylor et al. 2012) and extended to 2014. Please refer to Eyring et al. (2018) and ref-
1050 erences therein for details of the exact forcing datasets used for the historical simulations in CMIP6. The indices describing the ensem-
1051 ble member correspond to the model realization number (r), initialization method (i), physics index (p), and forcing index (f) used in
1052 the experiment. Please see the CMIP6 guidance for data users for more documentation on this (<https://pcmdi.llnl.gov/CMIP6/Guide/>).

	CMIP6 Model	Ensemble member	Ocean component	Nominal ocean resolution (lon X lat)	Ocean vertical coordinate and levels	Eddy-induced advection coefficient (m ² s ⁻¹)	Modeling center
1	ACCESS-CM2	r1ilp1f1	MOM5	1.0 X 1.0	z* (50)	V; 100 – 1,200 ^{pc}	CSIRO-ARCCSS-BoM
2	ACCESS-ESM1-5	r1ilp1f1	MOM5	1.0 X 1.0	z* (50)	V; 50 – 600 ^a	CSIRO-ARCCSS-BoM
3	BCC-CSM2-MR	r1ilp1f1	MOM4-L40	1.0 X 1.0	z (40)	--	BCC-CMA
4	BCC-ESM1	r1ilp1f1	MOM4-L40	1.0 X 1.0	z (40)	--	BCC-CMA
5	CanESM5	r1ilp1f1	NEMO3.4.1	1.0 X 1.0	z (45)	V; 100 – 2,000 ^{pc}	CCCma
6	CESM2	r1ilp1f1	POP2	1.0 X 1.0	z (60)	V; 300 – 3,000 ^b	NCAR
7	CESM2-WACCM	r1ilp1f1	POP2	1.0 X 1.0	z (60)	V; 300 – 3,000 ^b	NCAR
8	CNRM-CM6-1	r1ilp1f2	NEMO3.6	1.0 X 1.0	z* (75)	V ^{pc}	CNRM-CERFACS
9	CNRM-CM6-1-HR	r1ilp1f2	NEMO3.6	0.25 X 0.25	z* (75)	None ^{pc}	CNRM-CERFACS
10	CNRM-ESM2-1	r1ilp1f2	NEMO3.6	1.0 X 1.0	z* (75)	V ^{pc}	CNRM-CERFACS
11	E3SM-1-0	r1ilp1f1	MPAS-Ocean	1.0 X 1.0	z* (60)	F; 1,800 ^c	E3SM-Project
12	EC-Earth3-Veg	r1ilp1f1	NEMO3.6	1.0 X 1.0	z* (75)	--	EC-Earth-Consortium
13	GFDL-CM4	r1ilp1f1	MOM6	0.25 X 0.25	hybrid $\sigma_2 - z^*$ (75)	None ^d	NOAA-GFDL
14	GFDL-ESM4	r1ilp1f1	MOM6	0.50 X 0.50	hybrid $\sigma_2 - z^*$ (75)	V; ~0 – 2,000 ^{pc}	NOAA-GFDL
15	GISS-E2-1-G	r1ilp1f1	GISS Ocean	1.25 X 1.0	z (40)	--	NASA-GISS
16	GISS-E2-1-G-CC	r1ilp1f1	GISS Ocean	1.25 X 1.0	z (40)	--	NASA-GISS
17	GISS-E2-1-H	r1ilp1f1	HYCOM	1.0 X 1.0	hybrid z- σ_2 -sigma (32)	--	NASA-GISS
18	HadGEM3-GC31-LL	r1ilp1f3	NEMO-HadGEM3-GO6.0	1.0 X 1.0	z* (75)	V; $\leq 1,000^e$	MOHC
19	HadGEM3-GC31-MM	r1ilp1f3	NEMO-HadGEM3-GO6.0	0.25 X 0.25	z* (75)	None ^{pc}	MOHC
20	INM-CM4-8	r1ilp1f1	INM-OM5	1.0 X 1.0	sigma (40)	None ^{pc}	INM
21	INM-CM5-0	r1ilp1f1	INM-OM5	0.5 X 0.25	sigma (40)	None ^{pc}	INM
22	IPSL-CM6A-LR	r1ilp1f1	NEMO3.6	1.0 X 1.0	z*(75)	V ^{ES}	IPSL
23	MCM-UA-1-0	r1ilp1f1	MOM1 plus	1.88 X 2.25	z (18)	None ^{ES}	University of Arizona
24	MIROC6	r1ilp1f1	COCO4.9	1.0 X 1.0	hybrid z - sigma (62)	F; 300 ^{ES}	JAMSTEC
25	MIROC-ES2L	r1ilp1f2	COCO4.9	1.0 X 1.0	hybrid z - sigma (62)	F; 300 ^{pc}	JAMSTEC
26	MPI-ESM1-2-HAM	r1ilp1f1	MPIOM1.6.3	1.5 X 1.5	z (40)	F; ~94 ^f	HAMMOZ-Consortium
27	MPI-ESM1-2-LR	r1ilp1f1	MPIOM1.6.3	1.5 X 1.5	z (40)	F; ~94 ^f	MPI-M
28	MRI-ESM2-0	r1ilp1f1	MRICOM4.4	1.0 X 0.5	z* (60)	V; 300 – 1500 ^g	MRI
29	NESM3	r1ilp1f1	NEMO3.4	1.0 X 1.0	z (46)	--	NUIST
30	NorCPM1	r1ilp1f1	MICOM1.1	1.0 X 1.0	hybrid z – σ_2 (53)	V; 100 – 1,500 ^{pc}	NCC
31	NorESM2-LM	r1ilp1f1	MICOM	1.0 X 1.0	hybrid z – σ_2 (53)	V; 100 – 1,500 ^{pc}	NCC
32	NorESM2-MM	r1ilp1f1	MICOM	1.0 X 1.0	hybrid z – σ_2 (53)	V; 100 – 1,500 ^{pc}	NCC
33	SAM0-UNICON	r1ilp1f1	POP2	1.0 X 1.0	z (60)	V; 300 – 3,000 ^{ES}	Seoul National University
34	UKESM1-0-LL	r1ilp1f2	NEMO-HadGEM3-GO6.0	1.0 X 1.0	z* (75)	V; $\leq 1,000^e$	NERC

1054 ^a Kiss et al. 2020.
1055 ^b Danabasoglu et al. 2012, Danabasoglu et al. 2020.
1056 ^c Golaz et al. 2019.
1057 ^d Adcroft et al. 2019.
1058 ^e Kuhlbrodt et al. 2018; and personal communication for maximum value allowed.
1059 ^f Jungclaus et al. 2013 and personal communication.
1060 ^g Yukimoto et al. 2019.
1061 ^{pc} Personal communication.
1062 ^{ES} Obtained from ES-DOCS: <https://search.es-doc.org/>

1063 **TABLE 2.** Metrics related to the strength of the ACC. The potential density (referenced to the surface) ($\Delta\rho$), salinity (ΔS), and
1064 potential temperature (ΔT) differences are the zonally and full-depth-averaged difference between 65°S and 45°S. The properties are
1065 first zonally and depth-averaged prior to computing the difference between 65°S and 45°S. The observational estimates are calculated
1066 from the WOA18 product (averaged 1985–94 and 1995–2004 decadal climatologies). The observational estimate for the surface wind
1067 stress parameters is the ERA5 monthly product averaged over the January 1986–December 2005 period with its associated interannual
1068 variability (standard deviation of the annual means over the 20-yr period) (1σ). The observed ACC estimate is that reported by
1069 Donohue et al. (2016) based on measurements from the cDrake array (Chereskin et al. 2012) over the 2007–11 period with its reported
1070 uncertainty (1σ). Please see the methods section and additional details in Beadling et al. (2019) on how this estimate is obtained.
1071 When computing metrics over the Drake Passage (DP) region, 55°S to 64°S is used as the latitudinal bounds. All model values are the
1072 time-averaged output from the last 20 years of the historical simulation (January 1986–December 2005). B-SOSE values are computed
1073 from the time-averaged output from January 2013 to December 2018 for the Iteration 133 solution. Given the likely large uncertainty
1074 from the pre-Argo era and the output provided at only decadal resolution, we do not attempt to provide an uncertainty for the property
1075 differences in the three columns of the table marked with an asterisk (*). For B-SOSE and the CMIP6 models, considering the mean \pm
1076 2σ (standard deviation of annual means for the time-period considered) of both the model and observational estimate, values which lie
1077 2σ outside the observational estimate are in bold and values that lie 2σ below are bold and italicized. Considering a 25% window of
1078 error for the WOA18 product, values are in bold if the mean is too strong and bold and italicized if the difference is too weak using
1079 this criterion. For the salinity difference, values that have an asterisk (*) show a gradient in the opposite direction than WOA18. The

1080 multi-model-mean (MMM) and inter-model-spread (1σ) for the CMIP6 models and that reported from the analysis presented in
1081 Beadling et al. (2019; B2019), Sen Gupta et al. (2009; S2009), and Russell et al. (2006a; R2006) is summarized at the bottom of the
1082 table. Please see Section 3a for discussion of models with their ACC values labeled with an asterisk (*).

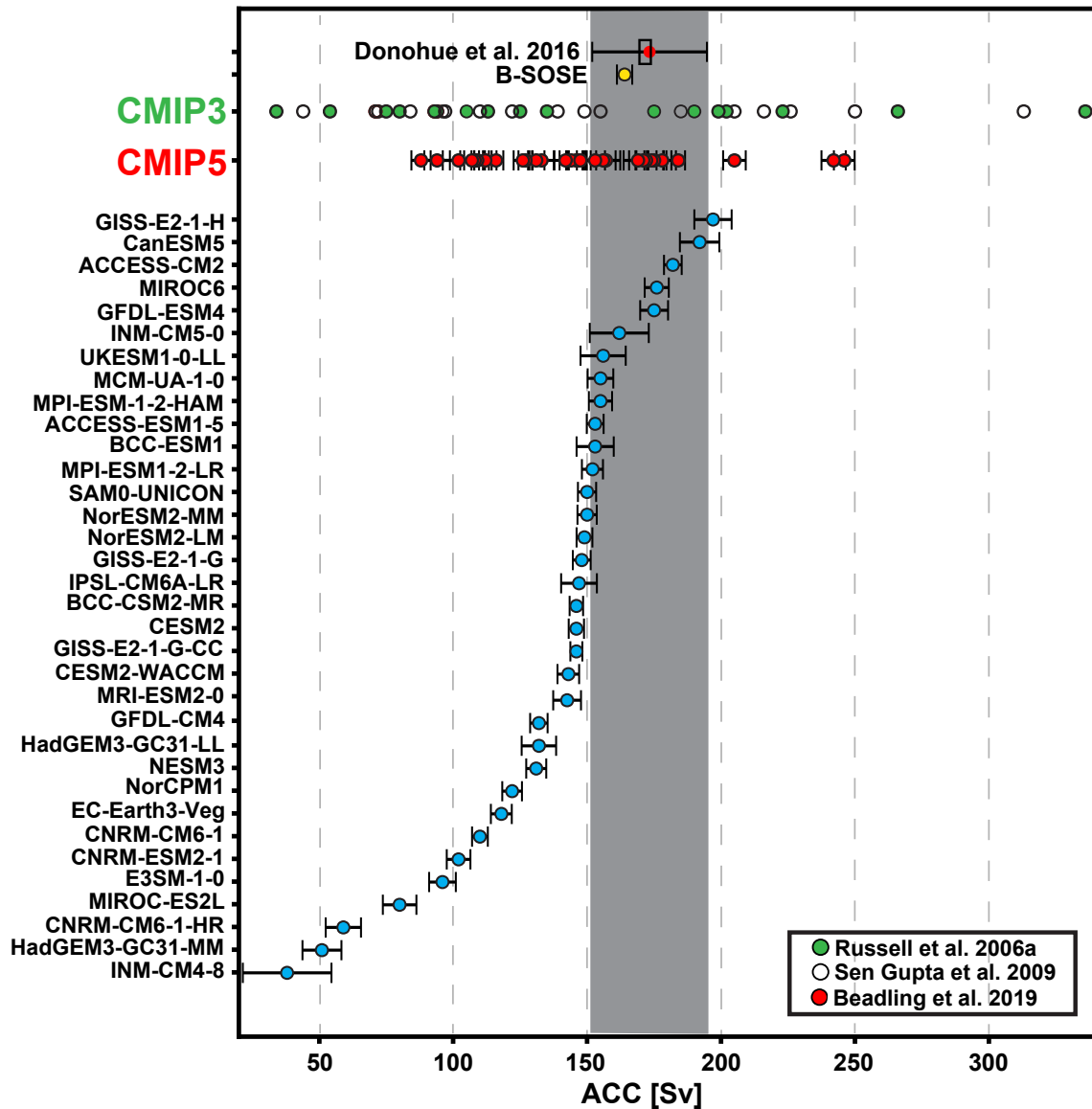
Model	ACC (Sv)	$\tau_{max}^{(x)}$ (N m ⁻²)	Lat. of $\tau_{max}^{(x)}$ (°S)	Total westerly $\tau^{(x)}$ (10 ¹² N)	Total $\tau^{(x)}$ in DP (10 ¹² N)	Total WSC in DP (10 ⁶ N m ⁻¹)	Min. zonally- integrated WSC (N m ⁻²)	Lat. of min WSC (°S)	$\Delta\rho$ (kg m ⁻³) *	ΔS *	ΔT (°C) *
Observational estimate	173.3 ± 10.7	0.1788 ± 0.01	52.37 ± 1.02	10 ± 0.5	2.4 ± 0.3	-2.57 ± 0.19	-5.32 ± 0.31	64.14 ± 0.47	0.25	0.06	-2.5
B-SOSE	164	0.1517	53.49	8.2	2.1	-2.20	-4.20	62.84	0.28	0.06	-2.3
GISS-E2-1-H	197	0.1760	52.45	9.7	2.4	-2.38	-5.13	64.63	0.39	0.20	-2.2
CanESM5	192	0.2049	50.68	11	2.4	-3.30	-5.03	62.20	0.35	0.10	-3.1
ACCESS-CM2	182	0.1944	50.96	11	2.2	-2.91	-4.56	61.90	0.39	0.01	-4.0
MIROC6	176	0.1537	50.10	9.0	1.8	-2.26	-4.37	62.91	0.23	-0.08*	-3.0
GFDL-ESM4	175	0.1901	51.09	11	2.4	-2.69	-4.33	62.35	0.23	-0.001*	-2.8
INM-CM5-0	162	0.2046	51.60	11	2.6	-3.14	-5.01	63.36	0.38	0.24	-2.1
UKESM1-0-LL	156	0.1980	51.78	11	2.6	-2.91	-4.69	62.46	0.29	0.05	-2.8
MCM-UA-1-0	155	0.1766	50.35	9.3	1.7	-2.71	-3.34	59.10	0.23	-0.06*	-3.3
MPI-ESM1-2-HAM	155	0.1824	51.05	10	2.2	-2.66	-3.89	62.06	0.26	-0.11*	-3.7
ACCESS-ESM1-5	153	0.2004	51.37	11	2.3	-3.22	-5.00	62.47	0.34	-0.001*	-3.7
BCC-ESM1	153	0.2263	51.86	12	2.7	-3.63	-4.65	61.22	0.27	-0.08*	-3.5
MPI-ESM1-2-LR	152	0.1997	49.16	12	2.1	-2.88	-4.02	61.56	0.24	-0.11*	-3.5
SAM0-UNICON	150	0.1910	53.35	10	2.7	-2.75	-4.54	62.92	0.32	0.15	-2.3
NorESM2-MM	150	0.1919	53.81	10	2.9	-2.58	-5.10	63.81	0.28	-0.03*	-3.7
NorESM2-LM	149	0.1914	53.70	9.9	2.8	-2.68	-4.23	62.77	0.31	-0.005*	-3.6
GISS-E2-1-G	148	0.1923	52.20	11	2.6	-2.70	-5.34	64.63	0.22	-0.04*	-2.9
IPSL-CM6A-LR	147	0.1785	50.47	10	1.9	-2.78	-4.09	61.80	0.26	0.02	-3.0
BCC-CSM2-MR	<i>146*</i>	0.2260	52.52	12	3.1	-3.14	-5.31	62.98	0.27	-0.05*	-3.3
CESM2	<i>146*</i>	0.2041	53.20	11	2.9	-3.15	-5.52	63.59	0.32	0.18	-1.9
GISS-E2-1-G-CC	<i>146*</i>	0.1884	52.24	10	2.6	-2.71	-5.35	64.51	0.21	-0.06*	-2.9
CESM2-WACCM	<i>143*</i>	0.2050	53.54	11	2.9	-3.26	-5.58	63.31	0.32	0.18	-1.9
MRI-ESM2-0	<i>143*</i>	0.1948	49.52	11	1.9	-3.13	-4.96	62.16	0.19	-0.002*	-2.3
GFDL-CM4	132	0.1787	50.63	11	2.1	-2.53	-4.32	61.78	0.21	0.05	-2.1
HadGEM3-GC31-LL	132	0.1903	51.38	11	2.4	-2.82	-4.61	62.15	0.29	0.05	-2.7
NESM3	131	0.2234	51.46	12	3.0	-2.92	-4.49	62.80	0.17	-0.13*	-2.6

NorCPM1	122	0.1855	53.86	9.4	2.6	-2.71	-4.08	62.73	0.27	0.08	-2.2
EC-Earth3-Veg	118	0.1790	51.21	10	2.3	-2.76	-5.20	63.80	0.29	0.02	-2.7
CNRM-CM6-1	110	0.1635	49.65	9.3	1.6	-2.75	-4.00	61.68	0.24	0.06	-2.0
CNRM-ESM2-1	102	0.1648	50.31	9.4	1.8	-2.71	-4.00	61.97	0.19	0.02	-1.8
E3SM-1-0	96	0.2000	51.79	11	2.6	-2.73	-4.49	62.06	0.20	-0.04*	-2.3
MIROC-ES2L	80	0.1280	46.11	7.0	0.5	-1.91	-3.05	57.70	0.13	-0.08*	-1.8
CNRM-CM6-1-HR	59	0.1645	47.46	9.8	1.3	-2.67	-4.52	61.93	0.19	0.03	-1.7
HadGEM3-GC31-MM	51	0.1791	50.49	10	2.1	-2.91	-5.45	62.76	0.23	0.05	-1.9
INM-CM4-8	38	0.2090	52.58	11	2.9	-3.03	-5.16	64.28	0.30	0.20	-1.4
R2006 CMIP3 MMM	144 ± 77	0.1591 ± 0.03	47.57 ± 2.80	NA	NA	NA	NA	NA	NA	NA	NA
S2009 CMIP3 MMM	142 ± 71	NA	49.44 ± 2.84	NA	NA	NA	NA	NA	NA	NA	NA
B2019 CMIP5 MMM	148 ± 38	0.1829 ± 0.02	50.36 ± 2.28	10 ± 1	1.9 ± 0.6	-2.67 ± 0.36	NA	NA	0.27 ± 0.08	0.05 ± 0.10	-2.6 ± 0.8
CMIP6 MMM	137 ± 37	0.1893 ± 0.02	51.29 ± 1.68	10 ± 1	2.3 ± 0.5	-2.82 ± 0.32	-4.63 ± 0.61	62.48 ± 1.4	0.27 ± 0.06	0.02 ± 0.10	-2.7 ± 0.7

1083

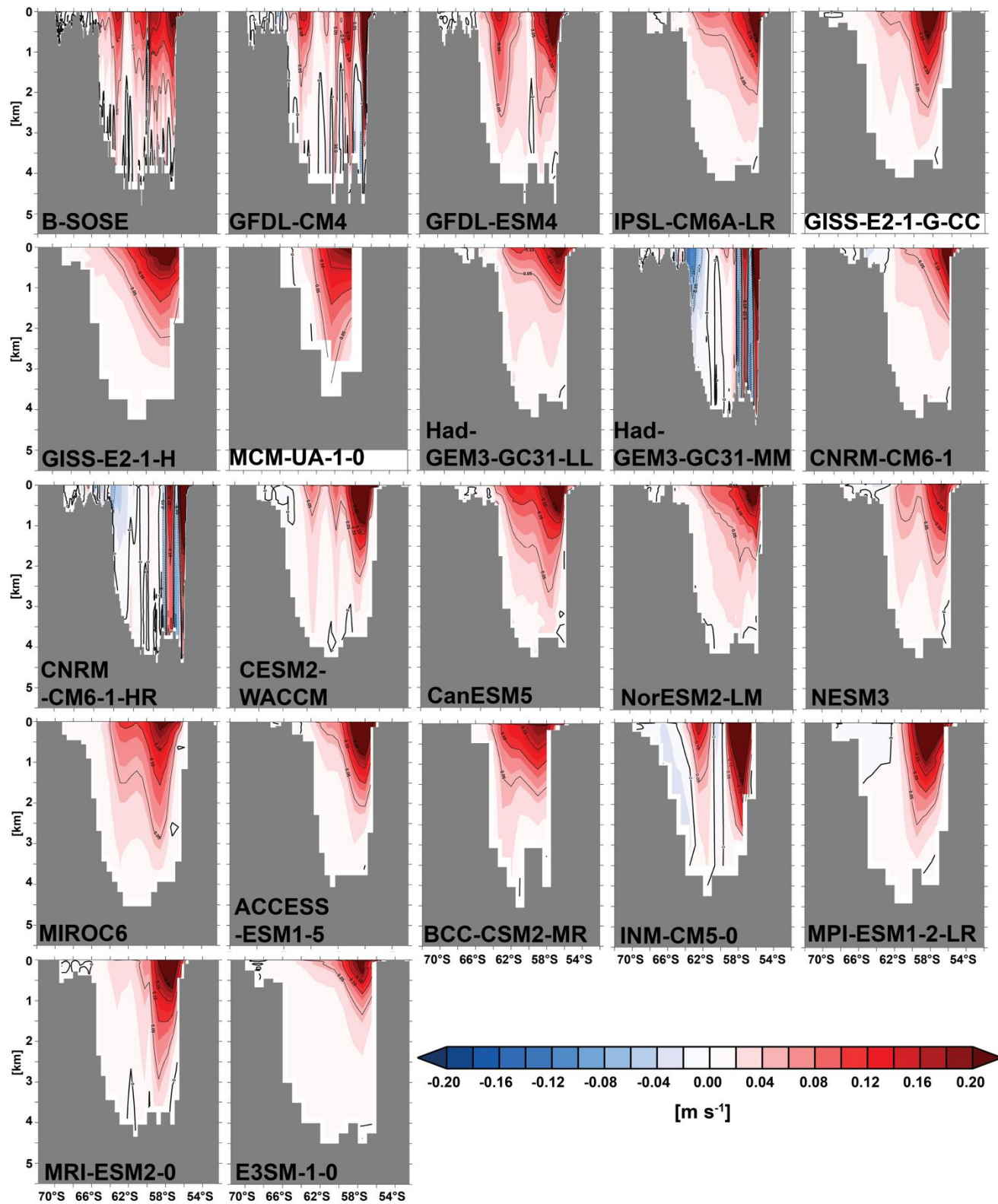
1084 **TABLE 3.** Metrics related to the representation of Antarctic Sea Ice Extent (SIE) (million km²)
1085 for the models considered in this study. The observed SIE metrics and their associated standard
1086 deviation (1σ) of annual means for the January 1986 to December 2005 time period are from the
1087 monthly sea ice extent data set (<https://nsidc.org/data/g02135>) from the National Snow and Ice
1088 Data Center Sea Ice Index. All CMIP6 values are time-averaged from January 1986 to December
1089 2005. B-SOSE values are computed from the time-averaged output from January 2013 to
1090 December 2018 for the Iteration 133 B-SOSE solution. In all calculations, only grid cells which
1091 had a sea ice concentration greater than 15% were included. Considering the standard deviations
1092 of the annual means over the time period of analysis for both the observational data and model
1093 output, values that lie 2σ outside the mean of the observational metric are in bold and values
1094 which lie 2σ below are bold and italicized. The B2019 CMIP5 multi-model mean (MMM) and
1095 inter-model spread (1σ) is from the analysis presented in Table S1 of Beadling et al. (2019). For
1096 the maximum and minimum SIE, the month when this occurs is also listed. *The MCM-UA-1-0
1097 model only provides ice thickness and concentration is either 1 or 0, preventing accurate
1098 assessment of monthly SIE.

Model	Annual SIE million km ²	Maximum SIE million km ²	Minimum SIE million km ²	Max – Min Million km ²
Observational estimate	11.58 ± 0.26	18.42 ± 0.34, SEP	3.09 ± 0.35, FEB	15.32 ± 0.52
B-SOSE	10.68	18.79, SEP	1.24, FEB	17.55
GISS-E2-1-H	11.21	18.91, SEP	2.86, FEB	16.06
CanESM5	12.10	18.23, SEP	4.12, FEB	14.11
ACCESS-CM2	7.97	14.30, SEP	0.60, FEB	13.70
MIROC6	1.91	4.09, SEP	0.04, FEB	4.05
GFDL-ESM4	9.00	16.66, SEP	0.82, FEB	15.84
INM-CM5-0	6.22	11.90, SEP	0.88, FEB	11.02
UKESM1-0-LL	9.99	15.47, SEP	2.68, FEB	12.79
MCM-UA-1-0	NA*	NA*	NA*	NA*
MPI-ESM-1-2-HAM	3.78	8.04, SEP	0.23, FEB	7.81
ACCESS-ESM1-5	8.83	14.73, SEP	2.48, FEB	12.25
BCC-ESM1	8.14	13.34, SEP	0.92, FEB	12.42
MPI-ESM1-2-LR	4.89	10.73, SEP	0.39, FEB	10.34
SAM0-UNICON	12.68	18.42, SEP	4.04, FEB	14.38
NorESM2-MM	6.85	12.12, SEP	1.80, FEB	10.32
NorESM2-LM	6.76	11.93, SEP	1.83, FEB	10.10
GISS-E2-1-G	8.19	15.02, SEP	0.64, FEB	14.38
IPSL-CM6A-LR	11.13	19.48, SEP	1.79, FEB	17.69
BCC-CSM2-MR	7.11	12.89, SEP	0.30, FEB	12.59
CESM2	9.79	15.35, SEP	1.81, FEB	13.54
GISS-E2-1-G-CC	7.88	14.68, SEP	0.64, FEB	14.04
CESM2-WACCM	10.38	15.91, SEP	2.27, FEB	13.64
MRI-ESM2-0	13.26	21.33, SEP	3.15, FEB	18.18
GFDL-CM4	10.33	19.00, SEP	0.68, FEB	18.32
HadGEM3-GC31-LL	8.82	14.33, SEP	1.80, FEB	12.53
NESM3	8.65	15.39, SEP	0.46, FEB	14.93
NorCPM1	13.29	19.70, SEP	4.38, FEB	15.32
EC-Earth3-Veg	4.97	10.63, SEP	0.29, FEB	10.34
CNRM-CM6-1	9.71	18.22, OCT	0.98, FEB	17.24
CNRM-ESM2-1	7.70	15.25, SEP	0.40, FEB	14.85
E3SM-1-0	8.92	15.97, SEP	1.04, FEB	14.93
MIROC-ES2L	1.70	4.11, SEP	0.04, FEB	4.07
CNRM-CM6-1-HR	8.76	16.87, OCT	0.48, FEB	16.39
HadGEM3-GC31-MM	6.40	10.36, SEP	1.62, FEB	8.74
INM-CM4-8	4.47	9.07, SEP	0.27, FEB	8.80
B2019 CMIP5 MMM	9.44 ± 3.35	16.28 ± 4.19	2.17 ± 1.97	14.11 ± 3.14
CMIP6 MMM	8.24 ± 2.88	14.32 ± 4.10	1.42 ± 1.22	12.90 ± 3.50



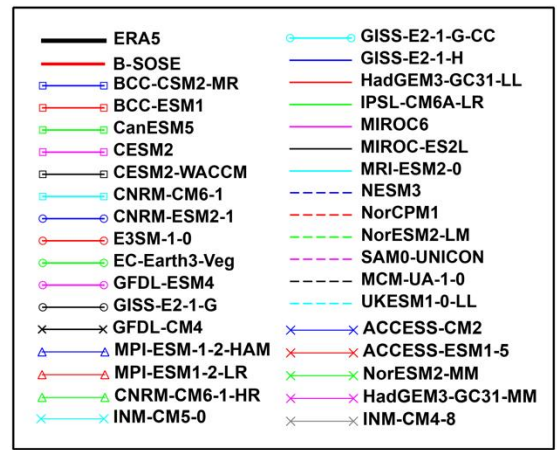
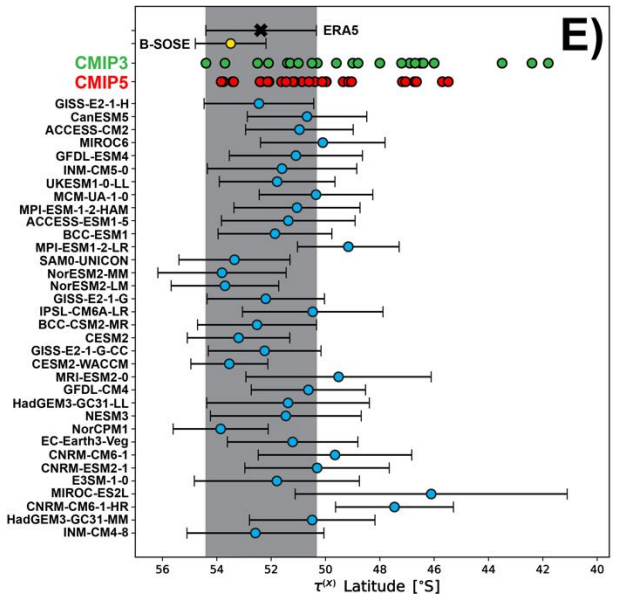
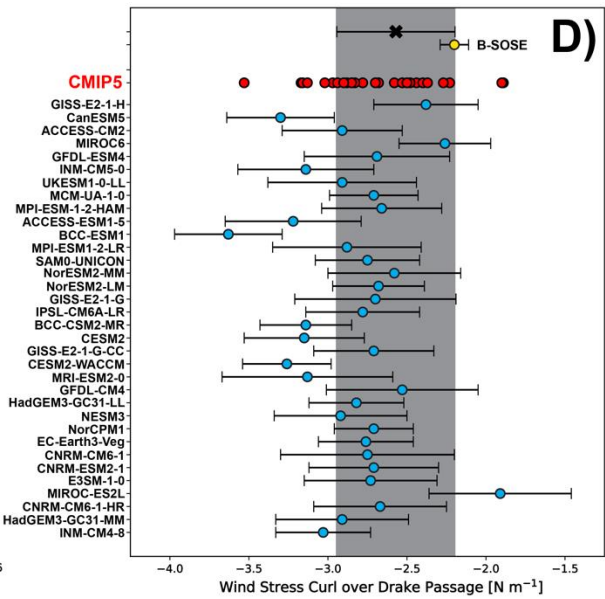
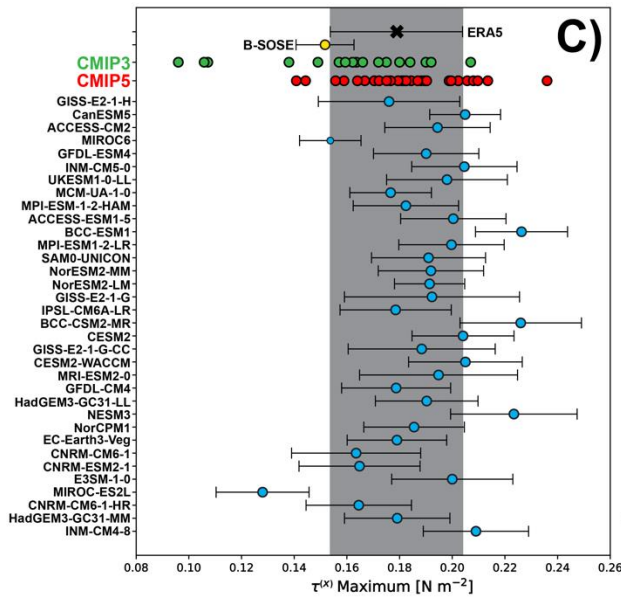
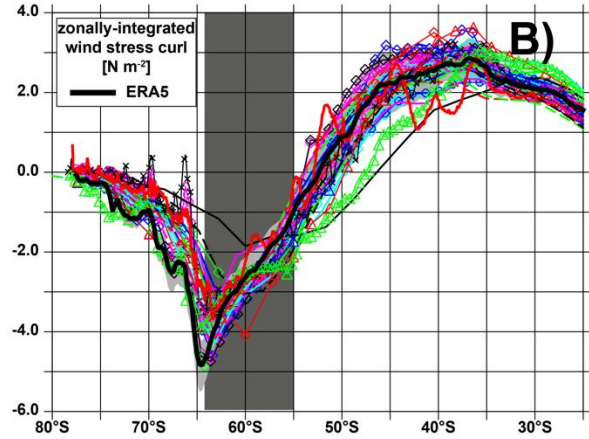
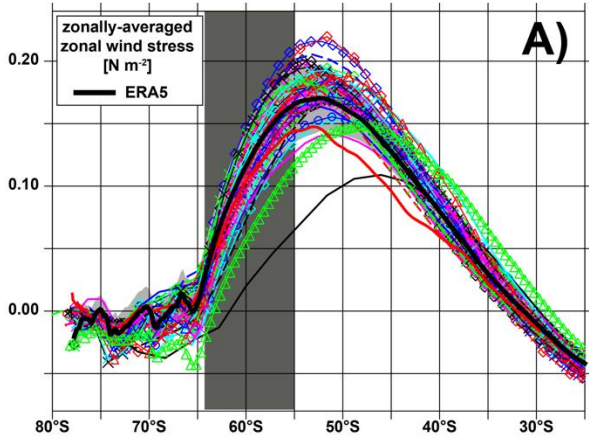
1101
 1102 **FIG. 1.** Volume transport of the ACC through the Drake Passage (Sv) as observed from the
 1103 cDrake experiment (Donohue et al. 2016) and as simulated across multiple generations of cli-
 1104 mate models. Gray shading corresponds to the observational uncertainty (2σ) of the ACC
 1105 transport reported in Donohue et al. (2016). The B-SOSE transport is computed from the Itera-
 1106 tion 133 solution at $1/6^{\text{th}}$ degree resolution over the January 2013 to December 2018 time period.
 1107 The CMIP6 model transports are calculated as the January 1986 to December 2005 time-aver-
 1108 aged net transport through the Drake Passage. The CMIP5 values are taken from Beadling et al.

1109 (2019). The CMIP6 transports are computed identically to that of Beadling et al. (2019), over the
1110 same time period in the historical simulations. Error bars correspond to the 2σ standard deviation
1111 of annual values about the mean. The CMIP3 values are taken from Table 1 of Russell et al.
1112 (2006a) and Table 3 of Sen Gupta et al. (2009). Russell et al. (2006a) computes flow through the
1113 Drake Passage from the last 20 years of each model's piControl integration, while Sen Gupta et
1114 al. (2009) computes flow from the last 20 years of each model's twentieth century control run.

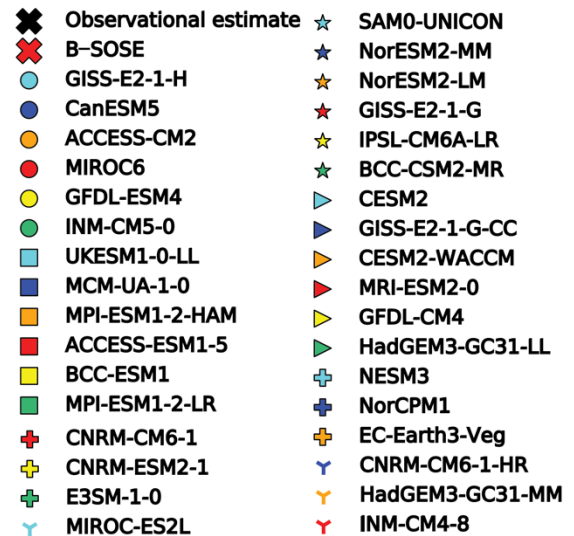
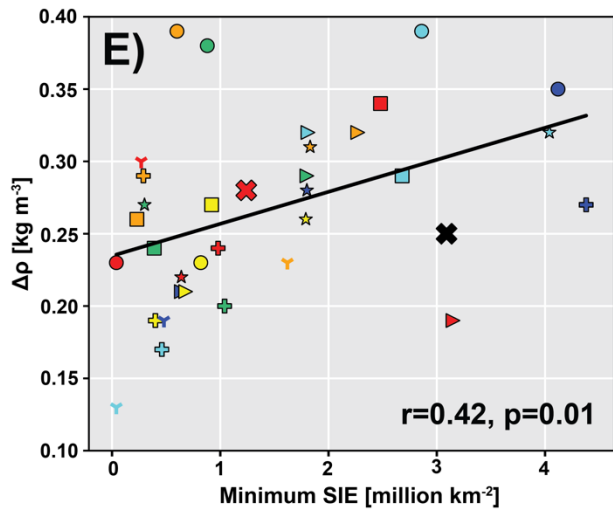
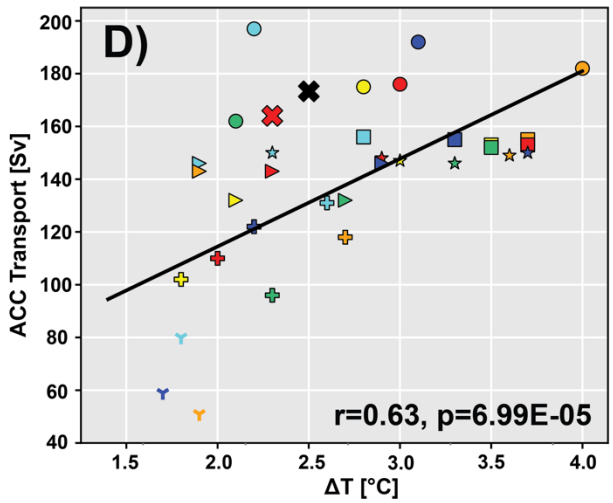
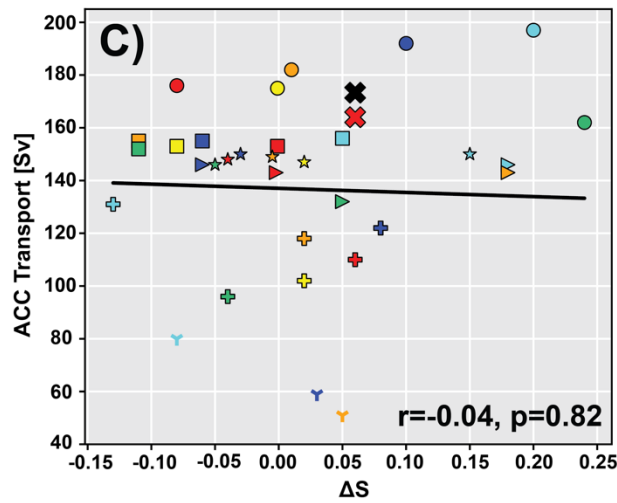
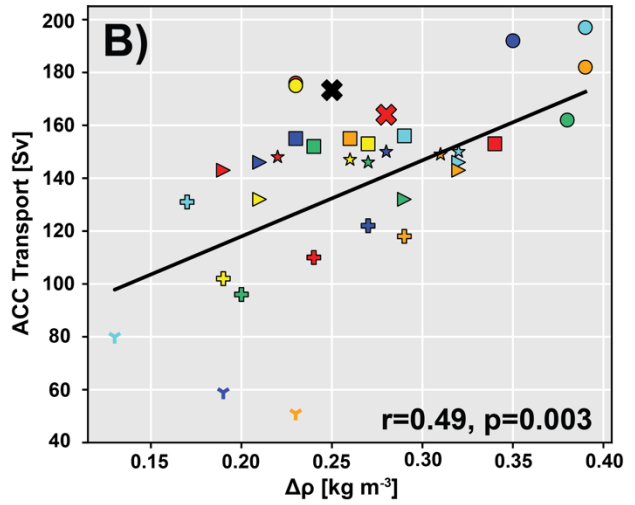
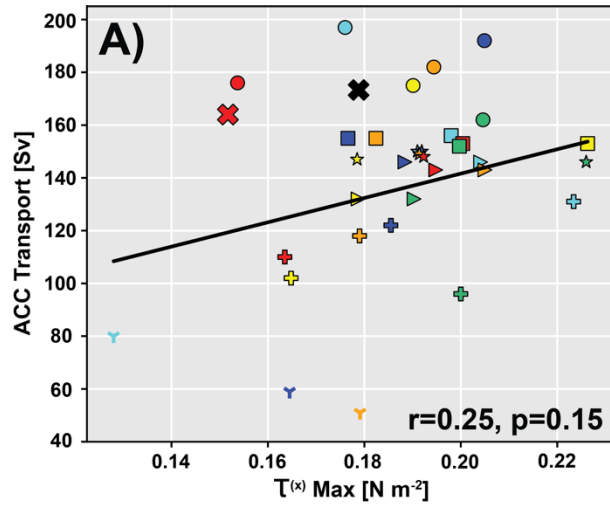


1115
1116

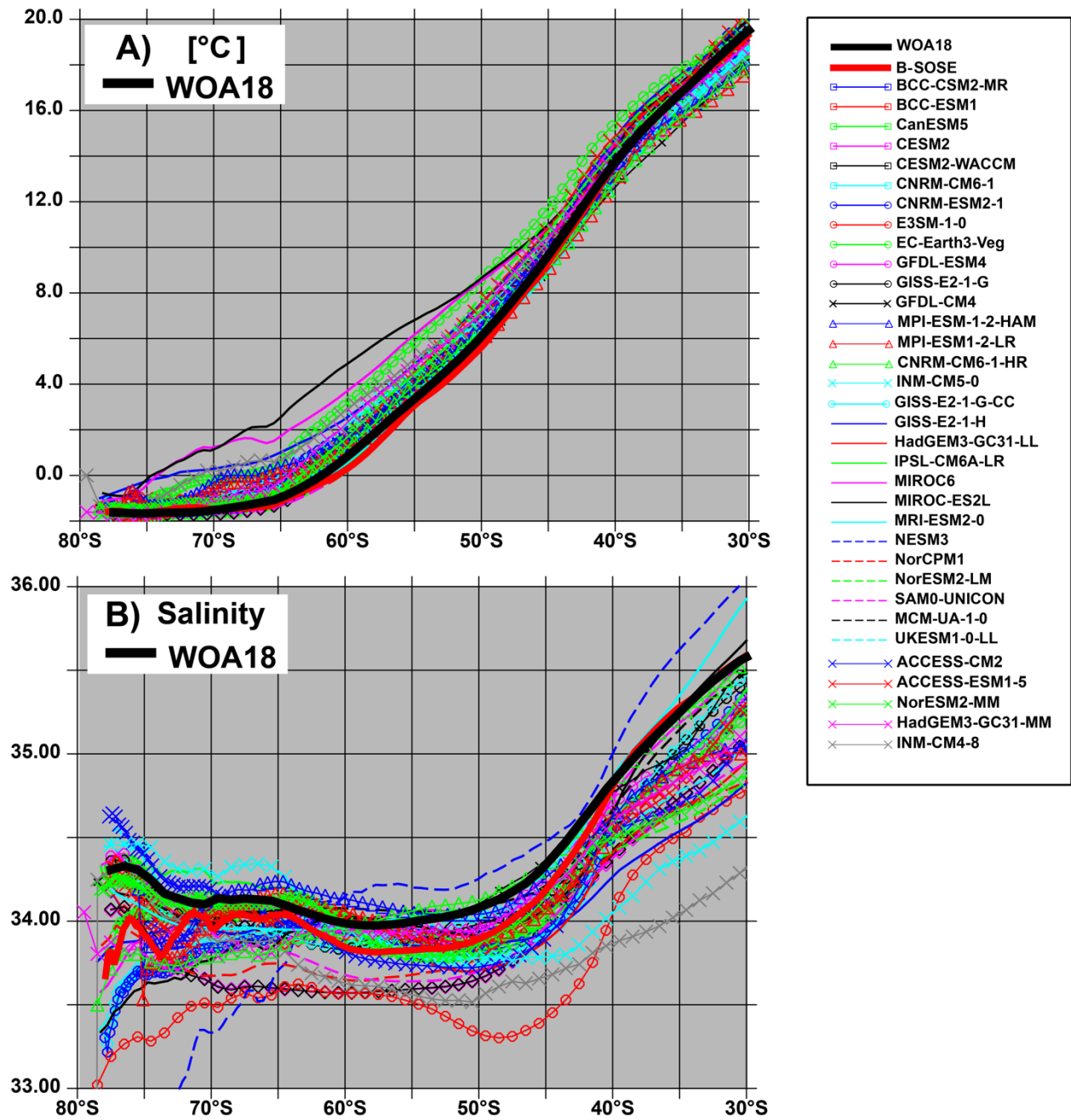
1117 **FIG. 2.** Zonal velocity (m s^{-1}) through the Drake Passage ($\sim 69^\circ\text{W}$). Red values indicate eastward
1118 velocities, and blue values indicate westward velocities. Velocities are contoured from -0.20 to
1119 0.20 m s^{-1} at intervals of 0.05 m s^{-1} . B-SOSE velocity is computed from the Iteration 133 solution
1120 at $1/6^{\text{th}}$ degree resolution over the January 2013 to December 2018 time period. All CMIP6
1121 model values are averaged from January 1986 to December 2005. Additional models are shown
1122 in the Supplemental Material.



1124 **FIG 3.** a) Zonally averaged zonal wind stress (N m^{-2}) and b) zonally integrated annual mean
1125 wind stress curl (N m^{-2}) from the ERA5 global atmospheric reanalysis product time-averaged
1126 from January 1986 to December 2005, from January 2013 to December 2018 for the Iteration
1127 133 B-SOSE solution, and from January 1986 to December 2005 for each CMIP6 model. The
1128 Drake Passage latitudinal band is the dark gray shading in panels A and B. The light gray shad-
1129 ing about the ERA5 mean corresponds to the interannual variability (2σ) of the zonally averaged
1130 zonal wind stress (a) or zonally integrated wind stress curl (b) at each latitude. Panels C through
1131 E show the performance of models from CMIP3 through CMIP6 in their representation of (c) the
1132 maximum zonally averaged zonal wind stress, (d) the integrated wind stress curl over the Drake
1133 Passage latitudes, and (e) position of the peak wind stress, relative to ERA5. The CMIP3 values
1134 are those reported by Russell et al. (2006a) and Sen Gupta et al. (2009), and the CMIP5 values
1135 are reported from Beadling et al. (2019). The CMIP6 transports are computed identically to that
1136 of Beadling et al. (2019), over the same time period in the historical simulations. Gray shading in
1137 panels C through D corresponds to the interannual variability (2σ) about the ERA5 mean over
1138 the 20-year time period.

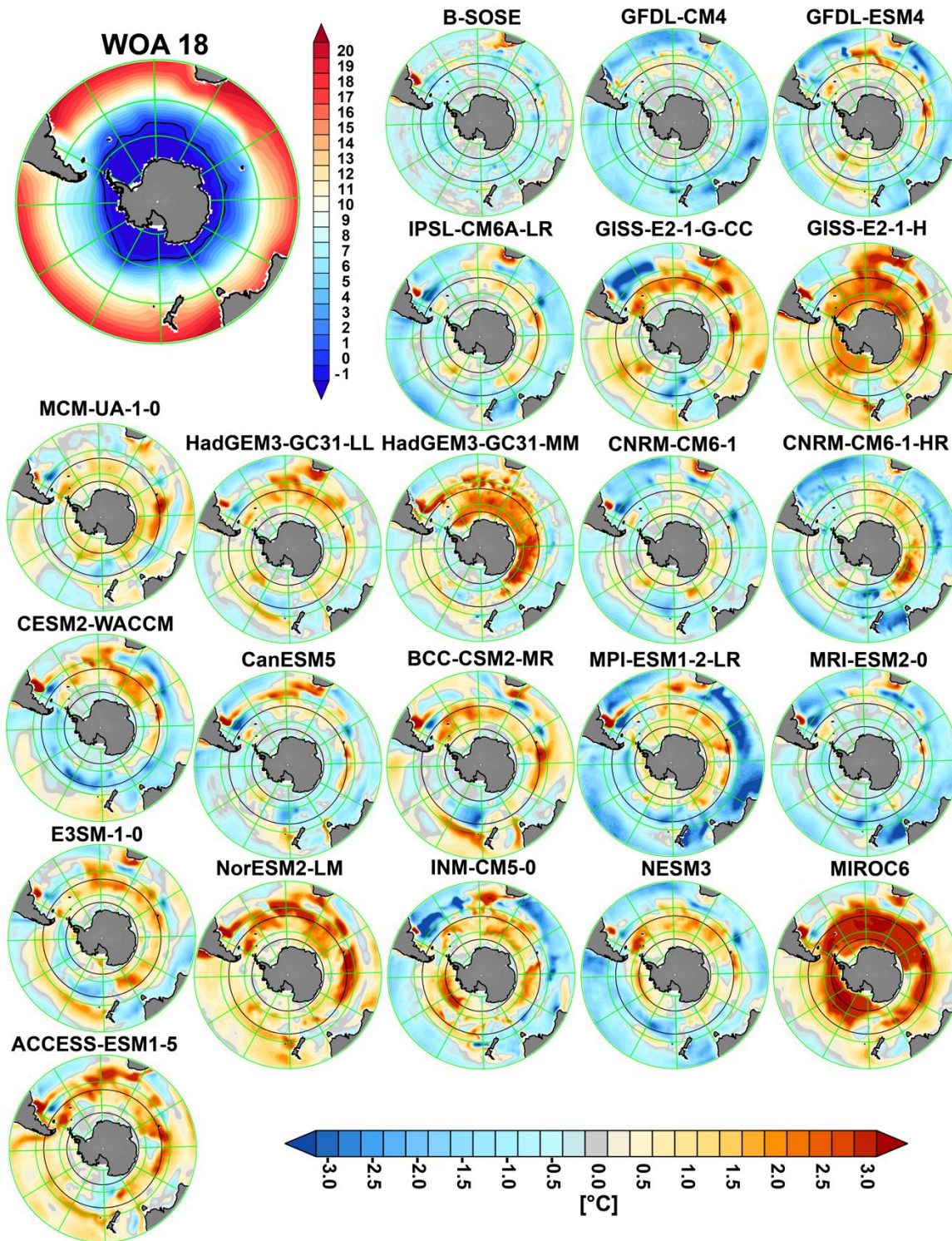


1140 **FIG. 4.** ACC transport vs. (a) zonally averaged maximum westerly wind stress, (b-d) full-depth-
1141 averaged, zonally averaged meridional potential density, potential temperature, and salinity dif-
1142 ference between 65°S and 45°S. (e) Full-depth-averaged meridional potential density difference
1143 between 65°S and 45°S vs. minimum Antarctic Sea Ice Extent (SIE). Mean observed and mod-
1144 eled values correspond to the values reported in Tables 2 and 3. The linear regression consider-
1145 ing only the CMIP6 models and the corresponding correlation coefficient and p value (n-2 de-
1146 grees of freedom) are displayed on each panel.



1147

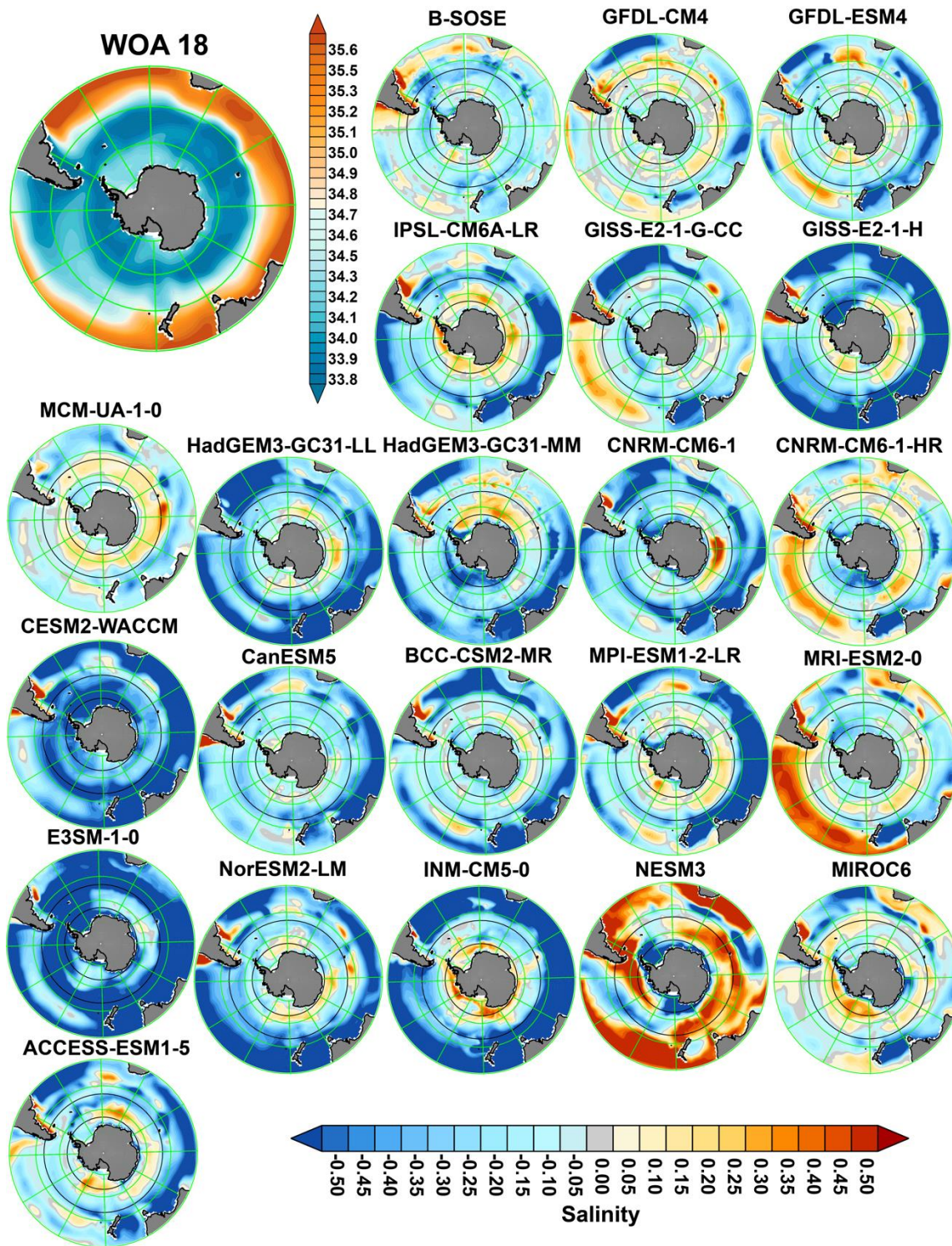
1148 **FIG. 5.** The zonally averaged and time-averaged near-surface (a) potential temperature (°C; 0–
 1149 100 m average), and (b) salinity (0–100-m average). The CMIP6 model output are time-averaged
 1150 from January 1986 to December 2005. The January 2013 – December 2018 time period is aver-
 1151 aged for the B-SOSE Iteration 133 solution. The WOA18 values are computed from the average
 1152 of the 1985–94 and 1995–2004 decadal climatologies.



1155 **FIG. 6a.** Upper ocean potential temperature ($^{\circ}\text{C}$) (0 -100 m average) in the Southern Ocean from

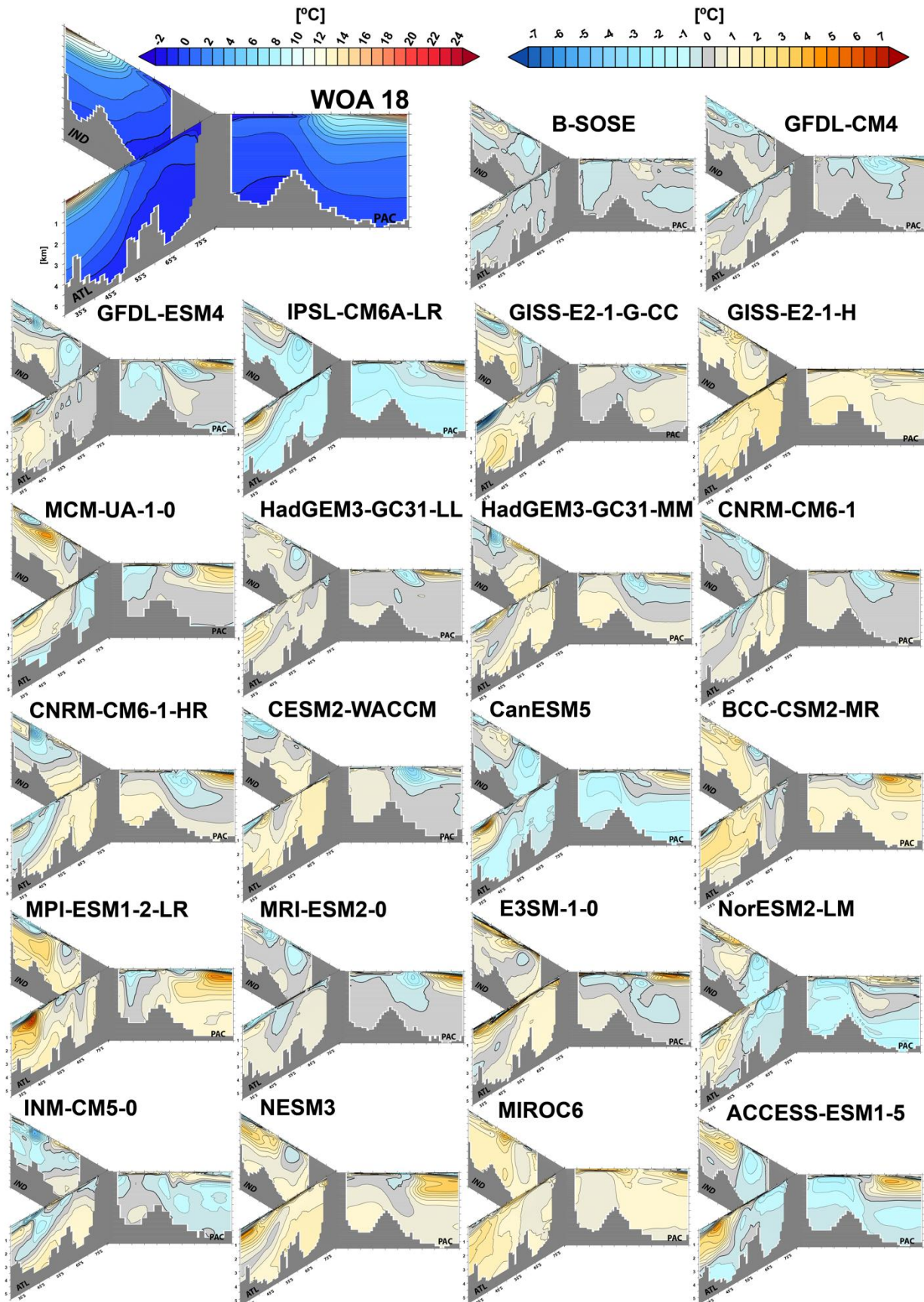
1156 the WOA18 climatological mean (large panel at top left; computed from the average of the

1157 1985–94 and 1995–2004 decadal climatologies). The difference between the simulated and ob-
1158 served temperature (°C) for the B-SOSE Iteration 133 solution (time-averaged from January
1159 2013 to December 2018), and for each CMIP6 model (time-averaged from January 1986 to De-
1160 cember 2005). Positive (from yellow to red) values indicate the model is warmer than observed.
1161 Negative (blue) values indicate the model is colder than observed. Additional models are shown
1162 in the Supplemental Material.

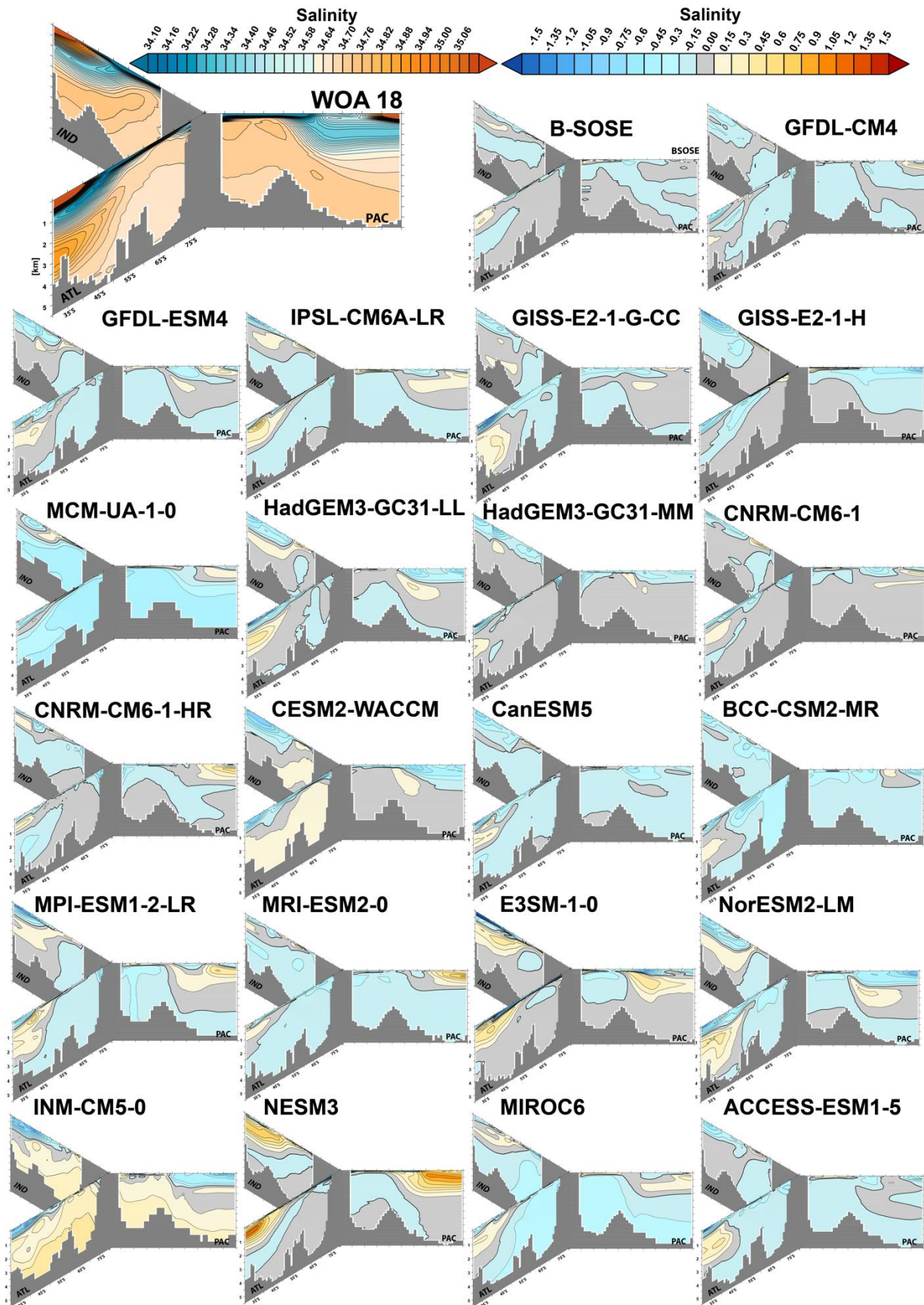


1163

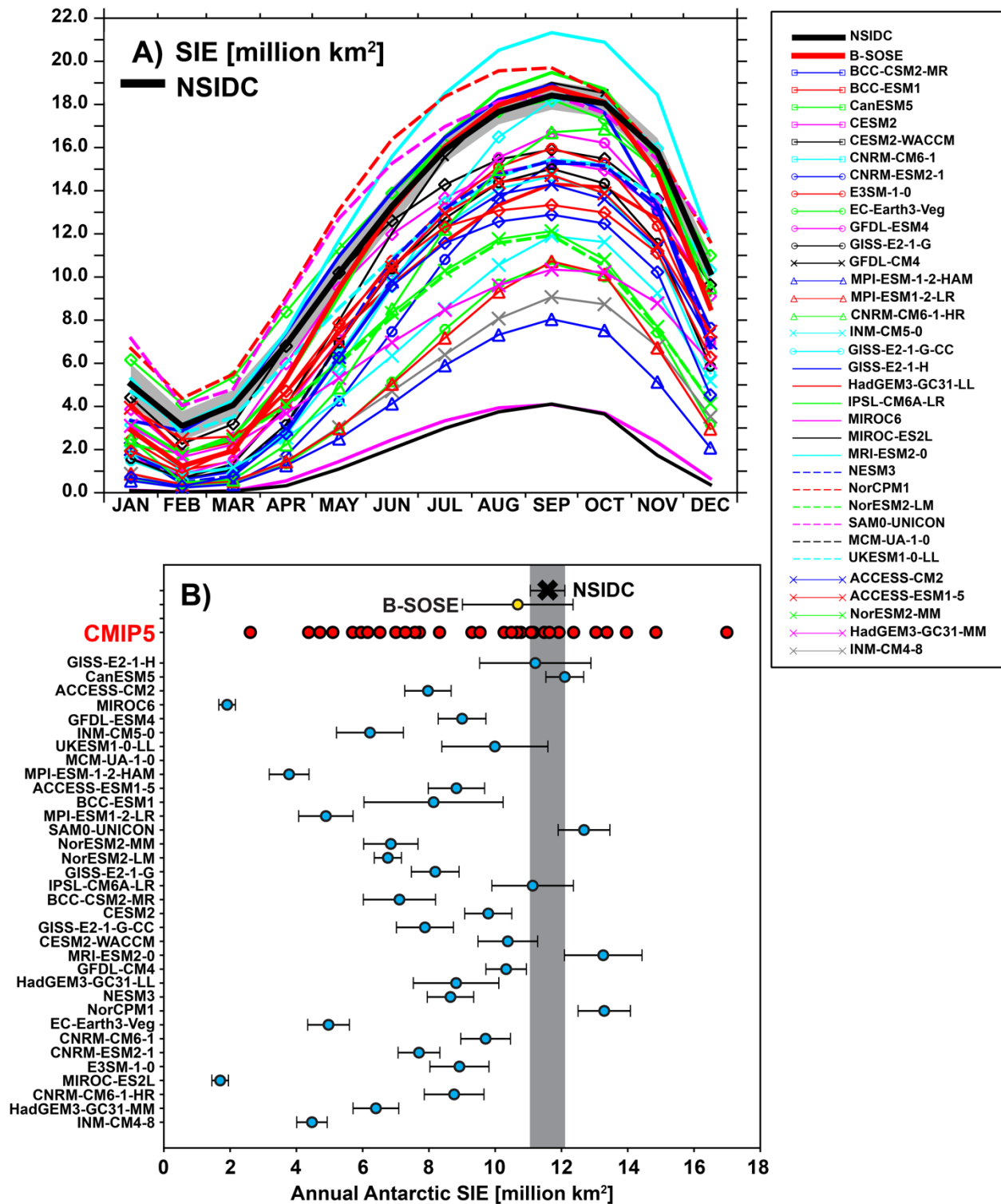
1164 **FIG. 6b.** As in 6a, but for salinity in the upper ocean (0 – 100 m average). Positive (from yellow
 1165 to red) values indicate the model is more saline than observed. Negative (blue) values indicate
 1166 the model is fresher than observed. Additional models are shown in the Supplemental Material.



1168 **FIG. 7a.** Potential temperature ($^{\circ}\text{C}$) in the Southern Ocean from the WOA18 climatological
1169 mean (large panel at top left; computed from the average of the 1985–94 and 1995–2004 decadal
1170 climatologies) in the Atlantic (30°W), Indian (90°E), and Pacific Ocean (150°W) and the differ-
1171 ence between the simulated and observed temperature for the B-SOSE Iteration 133 solution
1172 (time-averaged from January 2013 to December 2018), and for each CMIP6 model (time-aver-
1173 aged from January 1986 to December 2005). Positive (from yellow to red) values indicate the
1174 model is warmer than observed. Negative (blue) values indicate the model is colder than ob-
1175 served. Additional models are shown in the Supplemental Material.



1177 **FIG. 7b.** As in 7a, but for salinity in the Southern Ocean. Positive (from yellow to red) values
1178 indicate the model is more saline than observed. Negative (blue) values indicate the model is
1179 fresher than observed. Additional models are shown in the Supplemental Material.



1180

1181 **FIG. 8.** (a) Seasonal climatology of Antarctic Sea Ice Extent (SIE; million km²) observed by satellite and reported by the National Snow and Ice Data Center Sea Ice (NSIDC) and as simulated
 1182
 1183 in each CMIP6 model. The light gray shading about the NSIDC mean is the standard deviation

1184 (2σ) of the SIE for that month over the 20-year period. Data from observations and model output
1185 are time-averaged from January 1986 to December 2005. Only grid cells which have a sea ice
1186 concentration greater than 15% are included in the computation of SIE. (b) Annual Antarctic SIE
1187 simulated across model generations; CMIP5 through CMIP6. The CMIP5 values are taken from
1188 Beadling et al. (2019). The CMIP6 values are computed identically to that of Beadling et al.
1189 (2019), over the same time period in the historical simulations.