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

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. SALLÉE,[c] H. T. HEWITT,[d] P. HYDER,[d] AND AMARJIIT PANDDE[a]

[a] *Department of Geosciences, The University of Arizona, Tucson, Arizona*
[b] *Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California*
[c] *Sorbonne Université, UPMC University, Paris 06, UMR 7159, LOCEAN-IPSL, Paris, France*
[d] *Met Office Hadley Centre, Exeter, United Kingdom*

(Manuscript received 20 December 2019, in final form 28 April 2020)

## ABSTRACT

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

## 1. Introduction

The sequestration and ventilation of heat and carbon that occurs in the Southern Ocean (SO) plays a key role in global climate change. Observational and modeling studies over the last several decades continue to highlight the dominant role that the SO plays in the oceanic uptake of heat and carbon in present-day climate (Frölicher et al. 2015; Roemmich et al. 2015; Talley et al. 2016; Meredith et al. 2019). The disproportionate role that this region plays in the

planetary heat and carbon budget is linked to the unique and complex physical circulation that exists in the SO. Coupled models that contributed to previous generations of the Coupled Model Intercomparison Project (CMIP) have shown large disagreements in their ability to represent the large-scale circulation and associated properties and processes in this region (e.g., Russell et al. 2006a; Sen Gupta et al. 2009; Kuhlbrodt et al. 2012; Meijers et al. 2012; Bracegirdle et al. 2013; Heuzé et al. 2013; Sallée et al. 2013a,b; Meijers 2014; Frölicher et al. 2015; Shu et al. 2015; Ivanova et al. 2016; Hyder et al. 2018; Russell et al. 2018; Beadling et al. 2019). This has led to low confidence ascribed to CMIP model-based projections of future trends in the SO (Meredith et al. 2019). Given the vital role that this region plays in moderating climate globally, such a large disagreement in model performance is alarming and has implications for

---

interpreting projected trends not just in the SO, but globally.

The coordinated multimodel experiments forced by observed changes in the Earth system from ~1850 through the early twenty-first century, the "twentieth-century control" (20C3M) in CMIP3 and "historical" simulations in CMIP5 and CMIP6, allow climate and Earth system models (ESMs) to be scrutinized for how well they simulate key aspects of the climate system relative to observable quantities over the instrumental record. Through this process, a large number of systematic model biases pertaining to the simulation of the SO have been identified and discussed at length in the IPCC AR5 report (Flato et al. 2013) and summarized by Meijers (2014). Given the persistence of particular biases from one generation to the next, one of the three major scientific questions in the design of CMIP6 was to "*investigate the origins and consequences of systematic model biases*" with important and long-standing biases in the SO simulation highlighted as an area that needs to be addressed (Stouffer et al. 2017). A review by Meijers (2014) painted the CMIP5 SO simulation as a "*better CMIP3*" but lacking dramatic advancements that may have been expected given the larger ensemble and much more sophisticated models that participated relative to CMIP3.

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

heat and carbon in the Southern Hemisphere (Russell et al. 2006b; Waugh et al. 2019).

Coupled models have also consistently simulated upper-ocean temperatures in the SO that are too warm relative to modern observations (Russell et al. 2006a; Sen Gupta et al. 2009; Sallée et al. 2013a,b; Beadling et al. 2019; Hyder et al. 2018). This influences geostrophic ocean circulation, surface heat fluxes and water mass transformations, surface carbon fluxes, and the ability to accurately represent Antarctic sea ice extent (SIE). The properties of the upper ocean are influenced by a complex interplay between oceanic, atmospheric, and ice processes. This makes the upper ocean properties highly sensitive to biases in individual model components (i.e., the atmosphere, ocean, or sea ice models used in the coupled configuration). For example, Hyder et al. (2018) provided strong evidence that sea surface temperature (SST) biases in the region 40°–60°S across the CMIP5 ensemble are primarily the result of net flux biases in the stand-alone atmospheric model linked to poor representation of clouds, cloud properties, and shortwave radiation errors.

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

The evaluation presented here provides a robust and comprehensive assessment of key observable metrics of SO properties and circulation in the historical simulations across a large ensemble of CMIP6 models. Observable metrics assessed include surface momentum forcing, ACC transport, density, salinity, and temperature patterns and gradients, and representation of Antarctic seasonal SIE. The analysis has been done in a way that allows consistent assessment across model generations from CMIP3 (Russell et al. 2006a; Sen Gupta et al. 2009) to CMIP5 (Beadling et al. 2019), highlighting areas of model improvement and areas where systematic biases persist. For cohesiveness, for each results section we first provide a discussion of the performance of that particular metric across model generations and then present the CMIP6 results. We bring all of our results together in an overall evaluation of cross-generational performance, and suggest next steps in the conclusions and summary section. As the results from the

twenty-first-century projections under various shared socioeconomic pathways (SSPs) for the CMIP6 scenarios have recently come online, it is urgently important to provide comprehensive documentation of model biases in this climatically important region.

## 2. Methods

### a. CMIP6 model output

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

### b. Observational metrics

Several observational metrics have been updated from Beadling et al. (2019). The *World Ocean Atlas 2018* (*WOA18*) product (Locarnini et al. 2018; Zweng et al. 2018) is used to assess biases in the density, salinity, and potential temperature differences across the ACC relative to those estimated from observations. For comparison to the 1986–2005 time period simulated in the models, the 1985–94 and 1995–2004 decadal climatologies are averaged from the *WOA18* product. For assessment of the surface momentum exchange from atmospheric wind stress, the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5 atmospheric reanalysis product is used (https://cds.climate.copernicus.eu/). The ERA5 product, with increased spatial and temporal resolution and other improved features, is considered to be an improvement to ERA-Interim, which ended production in August 2019 (https://confluence.ecmwf.int/display/CKB/ERA5%3A+data+documentation). ERA-Interim was used in the assessment of SO surface momentum forcing by Beadling et al. (2019) given its proven reliability in representing wind fields over the SO relative to other contemporary reanalysis products (Swart and Fyfe

2012; Bracegirdle and Marshall 2012; Bracegirdle et al. 2013). For the model analysis, monthly atmospheric or oceanic zonal ($\tau^x$) and meridional ($\tau^y$) wind stress fields are used. If the atmospheric wind stress fields are used, values over land are masked out prior to computations.

For the assessment of Antarctic SIE, the monthly Sea Ice Index version 3 data product (https://nsidc.org/data/G02135/versions/3) provided by the National Snow and Ice Data Center (NSIDC; Fetterer et al. 2017) is used. The data are derived from satellite passive microwave data and span the time period 1978 to the present day. In our assessment, we present the time-averaged monthly data from 1986 to 2005. To be consistent with the NSIDC data product, we only consider grid cells with a sea ice concentration greater than 15% since satellite passive microwave instruments cannot accurately measure concentrations below this value. To be consistent for comparison with Beadling et al. (2019), after masking out values below 15%, the sea ice concentration data (*siconc*) are regridded to a standard 1° horizontal resolution grid for models that have nonstandard curvilinear horizontal grids. This was done in Beadling et al. (2019) due to a lack of information about grid cell area to allow computations on models with nonregular grids.

For the ACC transports, the net volume transport through the Drake Passage (DP; closest grid cells to ~69°W) is computed in each model from their reported zonal velocity (uo) or mass transport (umo) output. Velocities were only used if the mass transport output was not provided. Mass transport is preferred over velocity for a number of reasons discussed in detail in Beadling et al. (2019) and ultimately provides for a better representation of the true time-averaged flow. If mass transport was used, the total transport was divided by a constant density of 1035 kg m$^{-3}$. For the observational benchmark, we use the 173.3 ± 10.7 Sv (1 Sv ≡ 10$^6$ m$^3$ s$^{-1}$) estimate from the cDrake array experiment, which was carried out from 2007 to 2011 (Chereskin et al. 2012; Chidichimo et al. 2014; Donohue et al. 2016). This value is the sum of the 127 ± 5.9 Sv baroclinic transport (Chidichimo et al. 2014) and 45.6 ± 8.9 Sv barotropic component computed from the cDrake array observations (Donohue et al. 2016). Please see the detailed discussion in section 2a of Beadling et al. (2019) (titled "Transport of the ACC through the Drake Passage") regarding observational DP estimates over the last several decades. The larger transport computed from the cDrake array relative to the canonical value of 134 Sv derived from hydrographic estimates (Whitworth et al. 1982; Whitworth 1983; Whitworth and Peterson 1985; Cunningham et al. 2003) is purely attributable to the higher spatial and temporal resolution observations that allowed strong barotropic currents near the bottom of the ocean to be resolved.

TABLE 1. Details of the ocean model components in the CMIP6 models used in this study. Details of the ocean component, resolution, and vertical levels were obtained from the header information in the netcdf files used for the analyses. Information regarding the eddy-induced advection coefficient was obtained from ES-DOCs (https://search.es-doc.org/), through personal communication with the modeling centers, or from model documentation literature, as noted by the superscript next to each entry. The term $V$ corresponds to a coefficient that is either 2D/3D/time-varying, and $F$ corresponds to a fixed coefficient. If the required information could not be obtained or confirmed at the time of publication, it is left blank (—). The ocean vertical coordinates are defined as follows: $z$, traditional depth coordinates; $\sigma_2$, isopycnal vertical coordinates; $z^*$, rescaled geopotential vertical coordinate for better representation of free-surface variations (Adcroft and Campin 2004); sigma, terrain-following coordinates; hybrid $\sigma_2$–$z^*$, isopycnal coordinates in the interior ocean and a $z^*$ coordinate in the mixed layer (Adcroft et al. 2019); hybrid $z$–$\sigma_2$, isopycal coordinates in the interior ocean and $z$ coordinates in the mixed layer; hybrid $z$–$\sigma_2$–sigma, $z$ coordinates in the mixed layer, isopycnal coordinates in the open stratified ocean, and sigma coordinates in shallow coastal regions; hybrid $z$–sigma, sigma coordinates between the sea surface and a fixed geopotential depth ($\sim$50 m) in the upper ocean and $z$ coordinates below this depth. The historical experiments span 1850–2014 and are forced by observed changes in atmospheric composition due to anthropogenic and natural sources over the entire historical period. These forcings are updated from the CMIP5 historical forcings (Taylor et al. 2012) and extended to 2014. Please refer to Eyring et al. (2016) and references therein for details of the exact forcing datasets used for the historical simulations in CMIP6. The indices describing the ensemble member correspond to the model realization number (r), initialization method (i), physics index (p), and forcing index (f) used in 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 × lat) | Ocean vertical coordinate and levels | Eddy-induced advection coefficient (m² s⁻¹) | Modeling center |
|---|---|---|---|---|---|---|---|
| 1 | ACCESS-CM2 | r1i1p1f1 | MOM5 | 1.0 × 1.0 | $z^*$ (50) | $V$; 100–1200[a] | CSIRO-ARCCSS-BoM |
| 2 | ACCESS-ESM1.5 | r1i1p1f1 | MOM5 | 1.0 × 1.0 | $z^*$ (50) | $V$; 50–600[b] | CSIRO-ARCCSS-BoM |
| 3 | BCC-CSM2-MR | r1i1p1f1 | MOM4-L40 | 1.0 × 1.0 | $z$ (40) | — | BCC-CMA |
| 4 | BCC-ESM1 | r1i1p1f1 | MOM4-L40 | 1.0 × 1.0 | $z$ (40) | — | BCC-CMA |
| 5 | CanESM5 | r1i1p1f1 | NEMO3.4.1 | 1.0 × 1.0 | $z$ (45) | $V$; 100–2000[c] | CCCma |
| 6 | CESM2 | r1i1p1f1 | POP2 | 1.0 × 1.0 | $z$ (60) | $V$; 300–3000[d] | NCAR |
| 7 | CESM2-WACCM | r1i1p1f1 | POP2 | 1.0 × 1.0 | $z$ (60) | $V$; 300–3000[d] | NCAR |
| 8 | CNRM-CM6.1 | r1i1p1f2 | NEMO3.6 | 1.0 × 1.0 | $z^*$ (75) | $V$[e] | CNRM-CERFACS |
| 9 | CNRM-CM6.1-HR | r1i1p1f2 | NEMO3.6 | 0.25 × 0.25 | $z^*$ (75) | None[e] | CNRM-CERFACS |
| 10 | CNRM-ESM2.1 | r1i1p1f2 | NEMO3.6 | 1.0 × 1.0 | $z^*$ (75) | $V$[e] | CNRM-CERFACS |
| 11 | E3SM-1.0 | r1i1p1f1 | MPAS-Ocean | 1.0 × 1.0 | $z^*$ (60) | $F$; 1800[f] | E3SM-Project |
| 12 | EC-Earth3-Veg | r1i1p1f1 | NEMO3.6 | 1.0 × 1.0 | $z^*$ (75) | — | EC-Earth-Consortium |
| 13 | GFDL-CM4 | r1i1p1f1 | MOM6 | 0.25 × 0.25 | Hybrid $\sigma_2$–$z^*$ (75) | None[g] | NOAA-GFDL |
| 14 | GFDL-ESM4 | r1i1p1f1 | MOM6 | 0.50 × 0.50 | Hybrid $\sigma_2$–$z^*$ (75) | $V$; $\sim$0–2000[h] | NOAA-GFDL |
| 15 | GISS-E2.1-G | r1i1p1f1 | GISS Ocean | 1.25 × 1.0 | $z$ (40) | — | NASA-GISS |
| 16 | GISS-E2.1-G-CC | r1i1p1f1 | GISS Ocean | 1.25 × 1.0 | $z$ (40) | — | NASA-GISS |
| 17 | GISS-E2.1-H | r1i1p1f1 | HYCOM | 1.0 × 1.0 | Hybrid $z$–$\sigma_2$–sigma (32) | — | NASA-GISS |
| 18 | HadGEM3-GC31-LL | r1i1p1f3 | NEMO-HadGEM3-GO6.0 | 1.0 × 1.0 | $z^*$ (75) | $V$; $\leq$ 1000[i] | MOHC |
| 19 | HadGEM3-GC31-MM | r1i1p1f3 | NEMO-HadGEM3-GO6.0 | 0.25 × 0.25 | $z^*$ (75) | None[j] | MOHC |
| 20 | INM-CM4.8 | r1i1p1f1 | INM-OM5 | 1.0 × 1.0 | Sigma (40) | None[k] | INM |
| 21 | INM-CM5.0 | r1i1p1f1 | INM-OM5 | 0.5 × 0.25 | Sigma (40) | None[l] | INM |
| 22 | IPSL-CM6A-LR | r1i1p1f1 | NEMO3.6 | 1.0 × 1.0 | $z^*$ (75) | $V$[l] | IPSL |
| 23 | MCM-UA-1.0 | r1i1p1f1 | MOM1 plus | 1.88 × 2.25 | $z$ (18) | None[l] | University of Arizona |
| 24 | MIROC6 | r1i1p1f1 | COCO4.9 | 1.0 × 1.0 | Hybrid $z$–sigma (62) | $F$; 300[l] | JAMSTEC |
| 25 | MIROC-ES2L | r1i1p1f2 | COCO4.9 | 1.0 × 1.0 | Hybrid $z$–sigma (62) | $F$; 300[m] | JAMSTEC |
| 26 | MPI-ESM-1.2-HAM | r1i1p1f1 | MPIOM1.6.3 | 1.5 × 1.5 | $z$ (40) | $F$; $\sim$94[n] | HAMMOZ-Consortium |
| 27 | MPI-ESM1.2-LR | r1i1p1f1 | MPIOM1.6.3 | 1.5 × 1.5 | $z$ (40) | $F$; $\sim$94[n] | MPI-M |
| 28 | MRI-ESM2.0 | r1i1p1f1 | MRI.COM4.4 | 1.0 × 0.5 | $z^*$ (60) | $V$; 300–1500[i] | MRI |

TABLE 1. (*Continued*)

| CMIP6 model | Ensemble member | Ocean component | Nominal ocean resolution (lon × lat) | Ocean vertical coordinate and levels | Eddy-induced advection coefficient ($m^2\,s^{-1}$) | Modeling center |
|---|---|---|---|---|---|---|
| 29 NESM3 | r1i1p1f1 | NEMO3.4 | 1.0 × 1.0 | $z$ (46) | — | NUIST |
| 30 NorCPM1 | r1i1p1f1 | MICOM1.1 | 1.0 × 1.0 | Hybrid $z$–$\sigma_2$ (53) | $V$; 100–1500[p] | NCC |
| 31 NorESM2-LM | r1i1p1f1 | MICOM | 1.0 × 1.0 | Hybrid $z$–$\sigma_2$ (53) | $V$; 100–1500[p] | NCC |
| 32 NorESM2-MM | r1i1p1f1 | MICOM | 1.0 × 1.0 | Hybrid $z$–$\sigma_2$ (53) | $V$; 100–1500[p] | NCC |
| 33 SAM0-UNICON | r1i1p1f1 | POP2 | 1.0 × 1.0 | $z$ (60) | $V$; 300–3000[l] | Seoul National University |
| 34 UKESM1.0-LL | r1i1p1f2 | NEMO-HadGEM3-GO6.0 | 1.0 × 1.0 | $z^*$ (75) | $V$; $\leq$ 1000[i] | NERC |

[a] Personal communication (D. Bi, 2020).
[b] Kiss et al. (2020).
[c] Personal communication (N. Swart, 2020).
[d] Danabasoglu et al. (2012, 2020).
[e] Personal communication (A. Voldoire, 2020).
[f] Golaz et al. (2019).
[g] Adcroft et al. (2019).
[h] Personal communication (J. Krasting, 2020).
[i] Kuhlbrodt et al. (2018), and personal communication (P. Hyder, 2020) for maximum value allowed.
[j] Personal communication (P. Hyder, 2020).
[k] Personal communication (A. Volodin, 2020).
[l] Obtained from ES-DOCS: https://search.es-doc.org/.
[m] Personal communication (H. Tatebe, 2020).
[n] Jungclaus et al. (2013) and personal communication (D. Neubauer and H. Haak, 2020).
[o] Yukimoto et al. (2019).
[p] Personal communication (M. Bentsen, 2020).

### c. B-SOSE Iteration 133

Monthly ocean fields from the Iteration 133 solution of the Biogeochemical Southern Ocean State Estimate (B-SOSE) at 1/6° horizontal resolution spanning January 2013–December 2018 are also analyzed in this study (http://sose.ucsd.edu/BSOSE6_iter133_solution.html). This is an update from the 1/3° horizontal resolution Iteration 105 B-SOSE solution spanning January 2008–December 2012 used in Beadling et al. (2019). B-SOSE, produced as part of the Southern Ocean Carbon Climate Observational Modeling (SOCCOM) project, assimilates observations from shipboard data, profiling floats, underway measurements, and satellites into a numerical model to produce a state estimate for the SO. In B-SOSE, the MIT general circulation model (MITgcm) is fully coupled to the Nitrogen-version of the Biogeochemistry with Light, Iron, Nutrients, and Gases (N-BLING) model [evolved from Galbraith et al. (2010)]. Given the nature of B-SOSE, where the MITgcm is brought into consistency with available observational data via an adjoint data assimilation approach, we expect B-SOSE to perform well in regions with a high density of observational measurements such as in the upper ocean and along transects with repeat ship-based observations. B-SOSE is constrained by satellite measurements of sea surface height (SSH) and mean dynamic topography, thus we expect B-SOSE to capture surface currents and geostrophic flows consistent with other independent observations. It is important to note that none of the hydrographic and velocity observations from the cDrake array (Chereskin et al. 2012; Chidichimo et al. 2014; Donohue et al. 2016) are assimilated in B-SOSE, and thus the B-SOSE transport provides an independent estimate. The momentum forcing at the ocean surface in B-SOSE is derived from hourly ERA5 atmospheric winds, which are then adjusted throughout the assimilation to achieve consistency with the ocean state. Please refer to Verdy and Mazloff (2017) for additional details on B-SOSE and a complete list of observational constraints used.

## 3. Results

### a. ACC transport

The ACC transport is influenced by a large number of properties and processes in the SO, including momentum input at the ocean surface from the overlying winds, the meridional gradient in density across the current, interactions with bottom topography, mesoscale eddies, the position of the subtropical and subpolar gyres (Meijers et al. 2012), internal mixing processes, and so on. Owing to this complexity, achieving an accurate ACC strength has proven to be a difficult task in coupled models (Russell
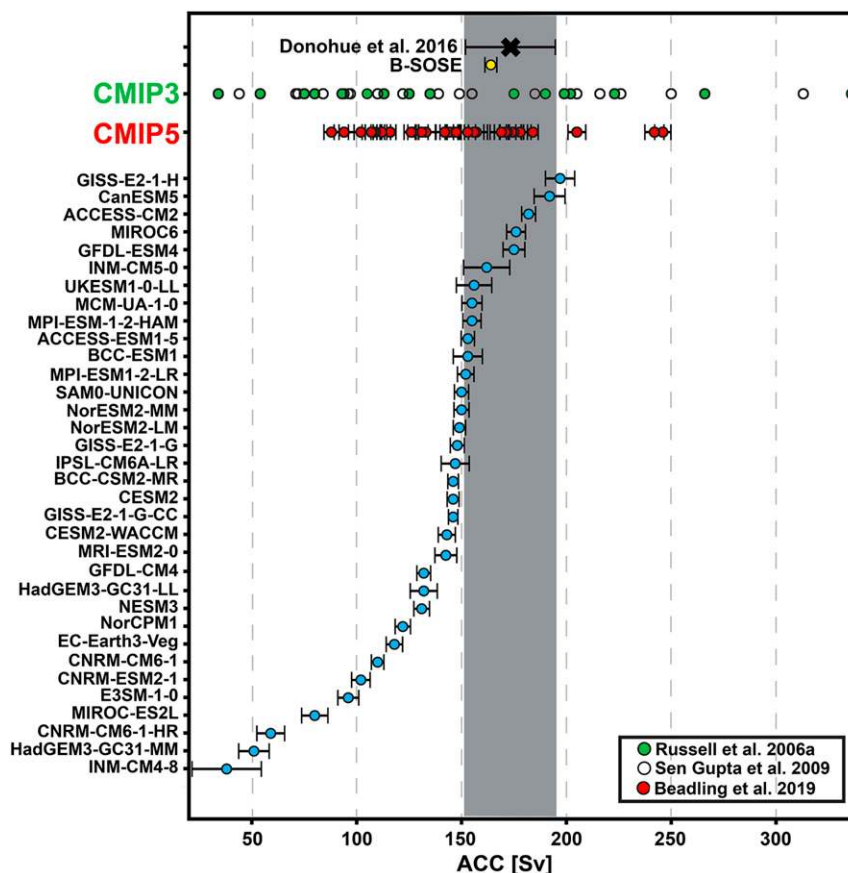
FIG. 1. Volume transport of the ACC through the Drake Passage (Sv) as observed from the cDrake experiment (Donohue et al. 2016) and as simulated across multiple generations of climate models. Gray shading corresponds to the observational uncertainty (2σ) of the ACC transport reported in Donohue et al. (2016). The B-SOSE transport is computed from the Iteration 133 solution at 1/6° resolution over the January 2013–December 2018 time period. The CMIP6 model transports are calculated as the January 1986–December 2005 time-averaged net transport through the Drake Passage. The CMIP5 values are taken from Beadling et al. (2019). The CMIP6 transports are computed identically to that of Beadling et al. (2019), over the same time period in the historical simulations. Error bars correspond to the 2σ standard deviation of annual values about the mean. The CMIP3 values are taken from Table 1 of Russell et al. (2006a) and Table 3 of Sen Gupta et al. (2009). Russell et al. (2006a) compute flow through the Drake Passage from the last 20 years of each model's piControl integration, while Sen Gupta et al. (2009) compute flow from the last 20 years of each model's twentieth-century control run.

et al. 2006a; Sen Gupta et al. 2009; Meijers et al. 2012; Beadling et al. 2019). The CMIP3 generation of models exhibited a very wide range of transports (Fig. 1; Table 2), with an intermodel spread (1σ) of 71–77 Sv. Only three of the CMIP3 models studied collectively by Russell et al. (2006a) and Sen Gupta et al. (2009) have mean values that fall within the observational uncertainty of the Donohue et al. (2016) cDrake array ACC estimate (Fig. 1). From CMIP3 to CMIP5 there is a large improvement in the range of the transports, with the intermodel spread (1σ) decreasing by ~36 Sv. This improvement in ACC strength was also noted in the analysis by Meijers et al. (2012).

The spread in the ACC transport has increased in CMIP6 relative to CMIP5, ranging from 38 Sv simulated

by INM-CM4.8 to 197 Sv simulated by GISS-E2.1-H. While there are no longer models that have an ACC transport that is much too strong, several models exhibit an extremely weak transport (>7σ outside of the observational uncertainty; E3SM-1.0, MIROC-ES2L, CNRM-CM6.1-HR, HadGEM3-GC31-MM, INM-CM4.8), reducing the multimodel mean (MMM) by ~10 Sv from CMIP5. Four models exhibit a transport weaker than any found in CMIP5. Interestingly, two of the models with an exceptionally weak ACC transport have 0.25° horizontal resolutions (CNRM-CM6.1-HR, HadGEM3-GC31-MM), with their 1° resolution versions (CNRM-CM6.1, HadGEM3-GC31-LL) simulating transports on the order of 50–80 Sv stronger. Similar, but not as extreme, behavior is found when comparing the 0.25°

GFDL-CM4 (132 Sv) simulation with that of the 0.50° GFDL-ESM4 simulation (175 Sv).

Despite the range in ACC transport across the model ensemble increasing due to several models with extremely weak magnitudes, CMIP6 is an overall improvement from CMIP5, with a larger fraction of models falling within observational uncertainty [within $2\sigma$ of the Donohue et al. (2016) estimate]. Of the 31 CMIP5 models studied by Beadling et al. (2019), 10 models (32%) fell within observational uncertainty. Of the 34 models studied here, 17 (50%) have transports within uncertainty. It is important to note the caveat, as also discussed regarding the CMIP5 models in Beadling et al. (2019), that most of the CMIP6 models analyzed here were developed *before* the Donohue et al. (2016) ACC estimate was known. Thus, during modeling development, transports at lower values were likely deemed reasonable. While we consider the Donohue et al. (2016) estimate as our benchmark for diagnosing the model simulations, the margin of uncertainty may be slightly larger in reality than used here given uncertainty associated with the fact that these observations were only collected from 2007–11. An independent estimate from assimilation using B-SOSE suggests a slightly lower ACC value of 164 Sv (Table 2; Fig. 1).

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

Several models exhibit significant decadal to multi-decadal variability in the ACC strength throughout the entire historical period, on the order of 10–30 Sv (Table S1, Fig. S1). Such variability may be associated with unrealistic quasiperiodic "superpolynya events" in the SO which alter the density structure of the SO through intense open ocean convection. This problem is documented in the GFDL-CM4 model (Held et al. 2019), where superpolynyas in the

Ross Sea drive large centennial-scale variability in the Southern Hemisphere climate. The lack of observational evidence of large polynya events in the SO, with the exception of the 1974–76 (Gordon 1978; Carsey 1980) and 2016–17 polynyas in the Weddell Sea (Campbell et al. 2019), and no similar events in the Ross Sea, suggests that these frequent simulated "superpolynya events" are unrealistic and need to be improved upon in future model development. In addition to the extreme rarity of these events in the real world, the modeled convection, ventilation, and climate change that occur during these simulated polynyas bear no resemblance to observed polynya events (Gordon 1978; de Lavergne et al. 2014; Dufour et al. 2017; Campbell et al. 2019). In the model, when these events occur, the formation of dense water in the subpolar SO translates into an increased ACC transport. If the simulation contains these events, the ACC assessment is sensitive to the 20-yr period in the historical simulation chosen for analysis (Fig. S1).

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

In the real ocean, the ACC flow through the DP is composed of strong and narrow eastward flowing jets that extend from the surface to the bottom of the ocean. Through the passage, strong bottom eastward velocities that average $1.3$ cm s$^{-1}$ provide an additional $45.6 \pm 8.9$ Sv barotropic flow to the $127 \pm 5.9$ Sv baroclinic transport (Chidichimo et al. 2014; Donohue et al. 2016). The CMIP6 models and B-SOSE have very different zonal velocity structures through the DP (Fig. 2; see also Fig. S2). The eddy-permitting 1/6° B-SOSE and the 1/4° GFDL-CM4 simulations yield very similar structures, with finely spaced jets that extend from the surface to the ocean bottom at most latitudes and increase in strength northward within the passage. These jets correspond to the dynamical ACC fronts characteristic of the observed flow within the DP (Lenn et al. 2007; Firing et al. 2011), with the majority of

TABLE 2. Metrics related to the strength of the ACC. The potential density (referenced to the surface) ($\Delta\rho$), salinity ($\Delta S$), and potential temperature ($\Delta T$) differences are the zonally and full-depth-averaged difference between 65° and 45°S. The properties are first zonally and depth averaged prior to computing the difference between 65° and 45°S. The observational estimates are calculated from the *WOA18* product (averaged 1985–94 and 1995–2004 decadal climatologies). The observational estimate for the surface wind stress parameters is the ERA5 monthly product averaged over the January 1986–December 2005 period with its associated interannual variability (standard deviation of the annual means over the 20-yr period) ($1\sigma$). The observed ACC estimate is that reported by Donohue et al. (2016) based on measurements from the cDrake array (Chereskin et al. 2012) over the 2007–11 period with its reported uncertainty ($1\sigma$). Please see the methods section and additional details in Beadling et al. (2019) on how this estimate is obtained. When computing metrics over the Drake Passage (DP) region, 55°–64°S is used as the latitudinal bounds. All model values are the time-averaged output from the last 20 years of the historical simulation (January 1986–December 2005). B-SOSE values are computed from the time-averaged output from January 2013 to December 2018 for the Iteration 133 solution. Given the likely large uncertainty from the pre-Argo era and the output provided at only decadal resolution, we do not attempt to provide an uncertainty for the property differences in the three columns of the table marked with an asterisk (*). For B-SOSE and the CMIP6 models, considering the mean $\pm 2\sigma$ (standard deviation of annual means for the time period considered) of both the model and observational estimate, values that lie $2\sigma$ outside the observational estimate are in bold and values that lie $2\sigma$ below are bold and italicized. Considering a 25% window of 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 this criterion. For the salinity difference, values that have an asterisk (*) show a gradient in the opposite direction than *WOA18*. The multimodel mean (MMM) and intermodel spread ($1\sigma$) for the CMIP6 models and that reported from the analysis presented in Beadling et al. (2019) (here B2019), Sen Gupta et al. (2009) (here S2009), and Russell et al. (2006a) (here R2006) is summarized at the bottom of the 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$^{-3}$)* | $\Delta S$* | $\Delta T$ (°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-ESM-1.2-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 | 10 | 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 | *146** | 0.2260 | 52.52 | **12** | 3.1 | −3.14 | −5.31 | 62.98 | 0.27 | −0.05* | **−3.3** |
| CESM2 | *146** | 0.2041 | 53.20 | 11 | 2.9 | −3.15 | −5.52 | 63.59 | **0.32** | **0.18** | −1.9 |
| GISS-E2.1-G-CC | *146** | 0.1884 | 52.24 | 10 | 2.6 | −2.71 | −5.35 | 64.51 | 0.21 | −0.06* | −2.9 |
| CESM2-WACCM | *143** | 0.2050 | 53.54 | 11 | 2.9 | **−3.26** | −5.58 | 63.31 | **0.32** | **0.18** | −1.9 |
| MRI-ESM2.0 | *143** | 0.1948 | 49.52 | 11 | 1.9 | −3.13 | −4.96 | 62.16 | 0.19 | −0.002* | −2.3 |

TABLE 2. (*Continued*)

| 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$^{-3}$)* | $\Delta S$* | $\Delta T$ (°C)* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 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 | — | — | — | — | — | — | — | — |
| S2009 CMIP3 MMM | 142 ± 71 | NA | 49.44 ± 2.84 | — | — | — | — | — | — | — | — |
| B2019 CMIP5 MMM | 148 ± 38 | 0.1829 ± 0.02 | 50.36 ± 2.28 | 10 ± 1 | 1.9 ± 0.6 | −2.67 ± 0.36 | −4.63 ± 0.61 | 62.48 ± 1.4 | 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 | — | — | 0.27 ± 0.06 | 0.02 ± 0.10 | −2.7 ± 0.7 |

the eastward flow concentrated in the Subantarctic Front (SAF) along the northern boundary of the DP and the Polar Front (PF) just to its south.

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

In most models, the majority of the ACC flow is concentrated in a single broad jet on the northern side of the passage, rather than being distributed among several finely resolved jets as observed in the real ocean. Many of the CMIP6 models exhibit shallow flow, with strong zonal jets that do not extend to the seafloor. The diversity in vertical extent of the flow may have important implications for the varied ACC strength found across models given the influence that interactions of the mean flow with bottom topography have on the ACC structure and strength through topographic steering, and dynamics such as momentum dissipation via bottom friction, etc. The degree to which topography and surface forcing impacts the mean flow likely differs markedly across models. For example, models with shallow jets are less likely to be influenced by interactions with bottom topography; that is, less bottom drag will be exerted on the large-scale flow than in models that have strong jets that extend to the sea floor. It appears that many of the errors in the ACC structure are related to the model's horizontal resolution since only the highest-resolution models begin to capture the observed jets and vertical extent.

The ACC strength is influenced by the competition between wind-driven surface divergence which acts to steepen isopycnals across the DP channel and baroclinic
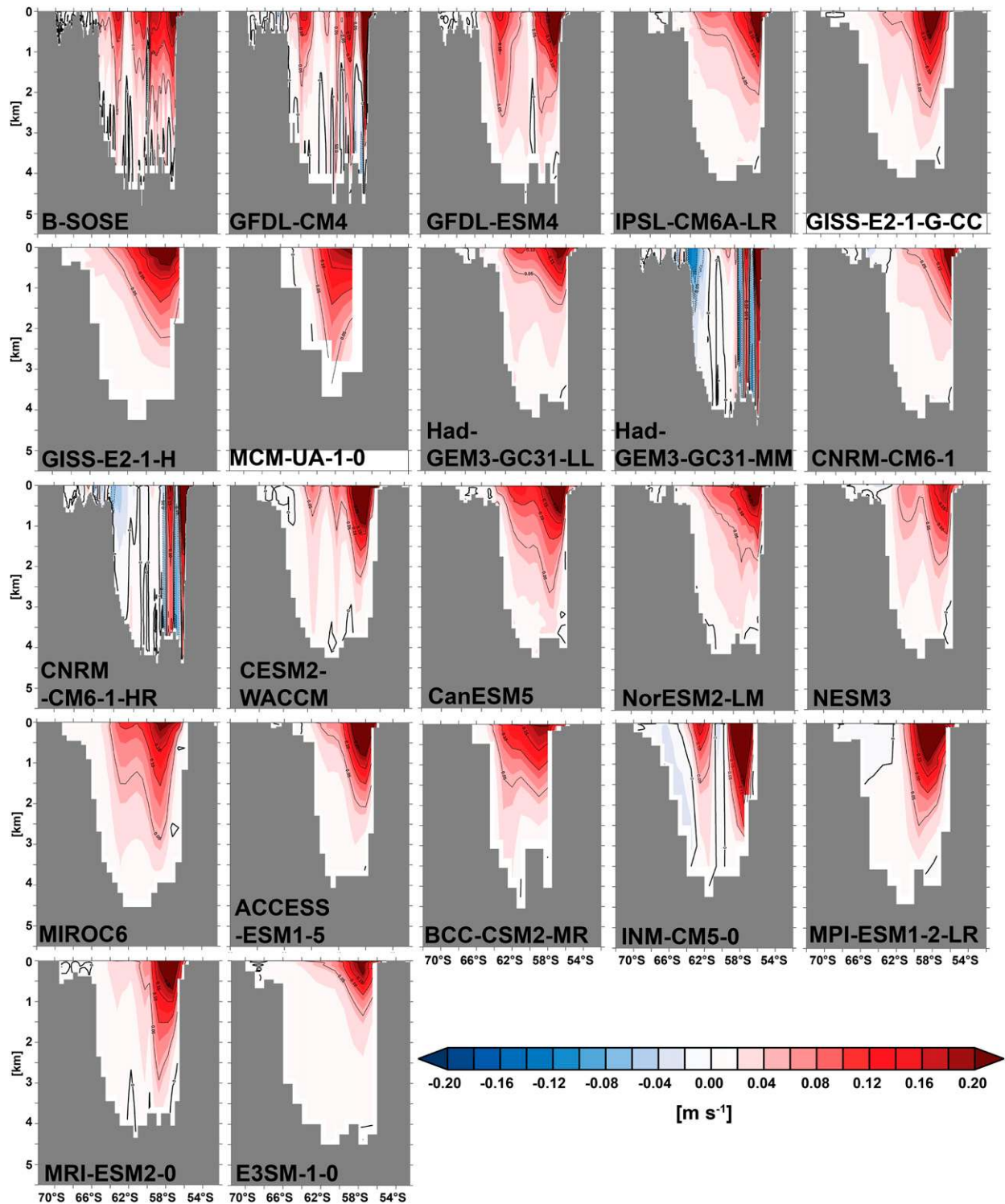
FIG. 2. Zonal velocity (m s$^{-1}$) through the Drake Passage (~69°W). Red values indicate eastward velocities, and blue values indicate westward velocities. Velocities are contoured from −0.20 to 0.20 m s$^{-1}$ at intervals of 0.05 m s$^{-1}$. B-SOSE velocity is computed from the Iteration 133 solution at 1/6° resolution over the January 2013–December 2018 time period. All CMIP6 model values are averaged from January 1986 to December 2005. Additional models are shown in the online supplemental material.

eddies which act to reduce the isopycnal slopes. The parameterized mixing schemes and coefficients employed in model simulations exerts a strong control on the resulting ACC strength and its vertical structure (Gent et al. 2001; Ragen et al. 2020). The ~1° resolution CMIP6 models all employ various subgrid-scale mixing schemes with different coefficients for parameterized eddy fluxes (Table 1), likely explaining a large portion of the diversity found across the CMIP6 models, as was the case in previous model generations (Kuhlbrodt et al. 2012). With the large diversity in sophisticated mixing schemes and magnitudes of eddy-induced advection coefficients employed across these models (Table 1), we do not attempt to quantify the role of subgrid-scale mixing on the model diversity in these simulations. We note that in the 0.25° resolution models (GFDL-CM4, HadGEM3-GC31-MM, and CNRM-CM6.1-HR), there are no parameterizations of mesoscale eddy transports and mesoscale eddies are explicitly, yet incompletely resolved at high latitudes (Table 1). We refer readers to Adcroft et al. (2019) for a detailed discussion on the effect of representing versus parameterizing mesoscale eddies on the ACC structure and other features of the SO circulation in the GFDL-CM4 (explicitly resolves mesoscale eddies) and GFDL-ESM4 (parameterized mesoscale eddy transport) models.

## b. Surface momentum forcing of the Southern Ocean and near-surface properties

### 1) SURFACE MOMENTUM FORCING FROM WIND STRESS

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

Previous generations of models have struggled with achieving accurate wind stress forcing at the ocean surface, with many models typically having relatively weak and equatorward biased $\tau_{max}^{(x)}$ values (Russell et al. 2006a; Sen Gupta et al. 2009; Meijers et al. 2012; Bracegirdle et al. 2013; Flato et al. 2013; Russell et al. 2018; Beadling et al. 2019) (Table 2, Figs. 3c,e). In CMIP3, several models had $\tau_{max}^{(x)}$ values on the order of 0.10–0.11 N m$^{-2}$ (Russell et al. 2006a), relative to the ERA5 value of 0.1788 N m$^{-2}$ used here as our observational benchmark. The CMIP5 generation showed improvement with a minimum $\tau_{max}^{(x)}$ value on the order of 0.14 N m$^{-2}$ and most models simulating values within observational uncertainty (Beadling et al. 2019). This improvement is noted in the shift from a CMIP3 MMM $\tau_{max}^{(x)}$ value from 0.1591 to 0.1829 N m$^{-2}$ in CMIP5 (Table 2), fewer models outside the lower bound of the ERA5 range in Fig. 3c, and a better agreement in the pattern of the zonally averaged wind stress over the SO [cf. Fig. 3a herein to Fig. 2 of Beadling et al. (2019)]. The CMIP6 ensemble tends to exhibit stronger $\tau_{max}^{(x)}$ than CMIP5, yielding a slightly increased MMM value (Table 2). Considering the mean and associated standard deviation ($2\sigma$) over the 20-yr period, only two models have $\tau_{max}^{(x)}$ values that do not overlap with the ERA5 range. The BCC-ESM1 and MIROC-ES2L models exhibit too strong and too weak $\tau_{max}^{(x)}$ values, respectively. The BCC-CSM-MR and NESM3 models have strong wind stress relative to ERA5, with the lower bound of their standard deviations ($2\sigma$) barely overlapping with that of ERA5.

The latitudinal location of $\tau_{max}^{(x)}$ ($\sim$52°S) is another metric that has been consistently analyzed across model generations and has improved. Yet, a systematic equatorward bias has persisted. The CMIP3 ensembles studied by Russell et al. (2006a) and Sen Gupta et al. (2009) had models with $\tau_{max}^{(x)}$ locations ranging from $\sim$42° to 54°S (Fig. 3e), with 63% (10 out of 16) to 78% (14 out of 18) of the models having their mean $\tau_{max}^{(x)}$ located north of 50°S. The CMIP5 models showed a clear improvement, with only 36% (11 out of 31) of models studied exhibiting locations north of 50°S and a narrowing intermodel spread (Fig. 3e). The CMIP6 models show even more improvement, with only 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 barely overlaps with the ERA5 mean and standard deviation ($2\sigma$) and is a clear outlier among the other CMIP6 models with respect to its zonal wind structure across the SO (Figs. 3a,e). The MMMs across generations, summarized in Table 2, clearly indicate an improvement in both the mean location of $\tau_{max}^{(x)}$ and in the intermodel spread, going from 47.57° $\pm$ 2.80° to 51.29° $\pm$ 1.68°S from CMIP3 to CMIP6.

The poleward shift in the location of $\tau_{max}^{(x)}$ from CMIP5 to CMIP6 has resulted in an increase in the total wind stress forcing (both from $\tau^{(x)}$ and WSC) over the open DP latitudes (Table 2). The structure of the zonally integrated WSC over the SO has improved from CMIP5 to CMIP6 [cf. Fig. 3b herein to Fig. 2 of Beadling et al.
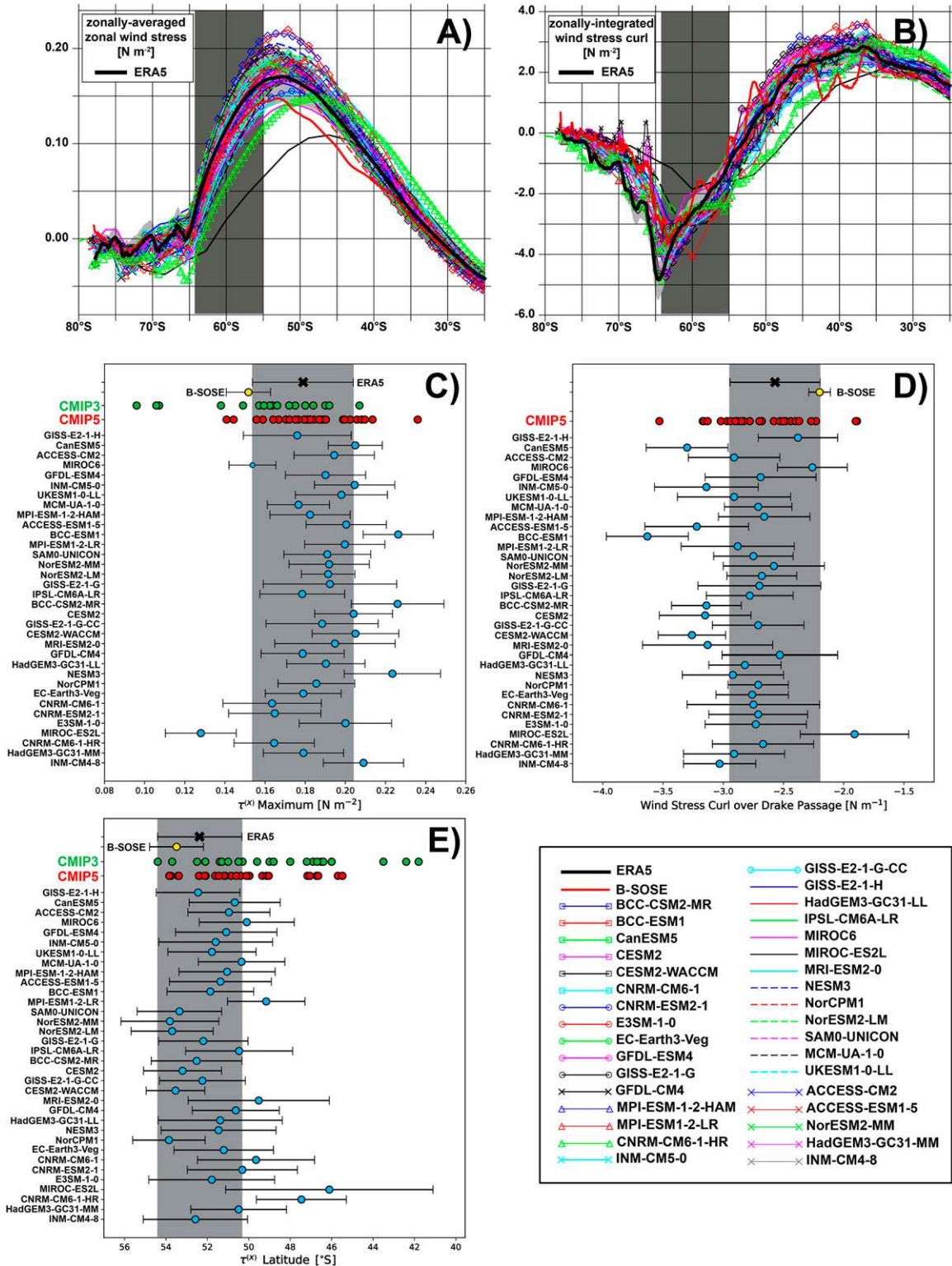
FIG. 3. (a) Zonally averaged zonal wind stress ($N\,m^{-2}$) and (b) zonally integrated annual mean wind stress curl ($N\,m^{-2}$) from the ERA5 global atmospheric reanalysis product time-averaged from January 1986 to December 2005, from January 2013 to December 2018 for the Iteration 133 B-SOSE solution, and from January 1986 to December 2005 for each CMIP6 model. The Drake Passage latitudinal band is the dark gray shading in (a) and (b). The light gray shading about the ERA5 mean corresponds to the interannual variability ($2\sigma$) of the zonally averaged zonal wind stress or zonally integrated wind stress curl at each latitude. Also shown is the performance of models from

(2019)]. The CMIP6 models are generally getting an accurate magnitude of Ekman suction/pumping over approximately the right locations, with the exception of MIROC-ES2L. The region south of 55°S yields the largest disagreement in the magnitude of the zonally integrated WSC among models, likely linked to diverging representation of the polar easterlies along the Antarctic margin. Noting this disagreement, we considered two additional metrics in this analysis to characterize the WSC field, the magnitude and location of the minimum WSC in the SO, corresponding to the magnitude and location of the maximum Ekman suction. Nine CMIP6 models yield a minimum zonally integrated WSC that is too weak, and five models have their latitudinal location too far equatorward relative to ERA5 (Table 2). Given that these WSC metrics are tied to the magnitude and location of where dense, carbon-rich water from the deep ocean is being pumped to the sea surface, the divergence in performance here may be linked to divergence in the simulated SO carbon budget in the CMIP6 ensemble. This is a topic of future study.

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

forcing, allowing steep isopycnal slopes to be maintained in the absence of strong wind forcing or vice versa. Examples from CMIP5 included the GISS-E2-R-CC and HadGEM3 models with ACC transports in excess of 240 Sv stemming from excessive density gradients driven by large errors in the SO temperature and salinity structure [see detailed discussion in Beadling et al. (2019)].

The CMIP6 ensemble exhibits a stronger relationship (yet still not statistically significant; $p = 0.15$) relative to previous model generations between the ACC and $\tau_{\max}^{(x)}$ (Fig. 4a). No statistically significant relationships are found between ACC strength and the other wind stress metrics in Table 2 (not shown). The MIROC6 model has one of the weakest $\tau_{\max}^{(x)}$ magnitudes with a position that is slightly equatorward shifted relative to ERA5, yet one of the strongest ACCs out of the ensemble. The E3SM-1.0, HadGEM-GC31-MM, INM-CM4.8, and CNRM-CM6.1-HR models all have reasonable representation of both the position and strength of $\tau_{\max}^{(x)}$, yet yield extremely weak ACC transports.

In addition to the buoyancy structure, explicitly resolved (in the case of HadGEM3-GC31-MM, GFDL-CM4, and CNRM-CM6.1-HR) or parameterized mixing associated with mesoscale eddies also plays a role in explaining the lack of a linear correlation between wind stress forcing and the ACC strength in the CMIP6 models. Stronger wind forcing acts to enhance the meridional isopycnal tilt across the ACC. However, in a regime of eddy saturation (Munday et al. 2013), this isopycnal tilt can be counterbalanced by the production of eddies, which then act to minimize the meridional density gradient. Given that many of the CMIP3 models employed fixed eddy-induced advection coefficients, Kuhlbrodt et al. (2012) showed that the sensitivity of the ACC transport to the magnitude of the eddy-induced advection coefficient was larger than to the zonal wind stress maximum in CMIP3 models. As discussed in the previous section, the implementation of parameterized subgrid-scale mixing associated with eddy fluxes differs widely across the CMIP6 models (Table 1), likely contributing to the lack of a strong relationship between the wind stress metrics and the ACC.

2) NEAR-SURFACE BIASES

The near-surface (0–100 m) thermal structure of the SO is characterized by a large meridional gradient in

---

← 

CMIP3 through CMIP6 in their representation of (c) the maximum zonally averaged zonal wind stress, (d) the integrated wind stress curl over the Drake Passage latitudes, and (e) position of the peak wind stress, relative to ERA5. The CMIP3 values are those reported by Russell et al. (2006a) and Sen Gupta et al. (2009), and the CMIP5 values are reported from Beadling et al. (2019). The CMIP6 transports are computed identically to that of Beadling et al. (2019), over the same time period in the historical simulations. Gray shading in (c) and (d) corresponds to the interannual variability ($2\sigma$) about the ERA5 mean over the 20-yr time period.

temperature ranging from less than −1°C near the Antarctic coast to ~20°C in the subtropical gyres at 30°S (Figs. 5a and 6a ). Achieving accuracy in representing the near-surface temperature structure in the SO, particularly south of 40°S, has proven to be a significant challenge for the climate modeling community, with significant warm biases persisting across model generations. Generally, above 200 m, the MMM CMIP3 and CMIP5 temperatures were biased slightly cold in the global ocean with the exception of the SO, where the upper ocean has been characterized by consistent excessive warm biases (Flato et al. 2013). When considering the zonal-mean upper ocean temperature distribution, CMIP3 and CMIP5 models generally represented the structure well (Russell et al. 2006a; Sen Gupta et al. 2009; Beadling et al. 2019) with biases generally within ~1°C of that observed at a given latitude. However as noted by Sen Gupta et al. (2009), zonal-averaging obscures regional biases in upper ocean temperature that in some cases exceed 5°C. In CMIP3, major biases in the upper ocean were found south of ~45°S in the region of the ACC and along the eastern boundaries of the basins, and were attributed to poor representation of eastern boundary currents (Sen Gupta et al. 2009). Across the CMIP5 ensemble, excessive surface temperatures translated into consistent warm biases found in ventilated layers of the SO including surface subtropical, mode, and intermediate waters (Sallée et al. 2013b).

Excessive SO surface temperatures in CMIP5 originated to some extent from excessive downward shortwave radiation related to poor representation of clouds and cloud properties, with a strong correlation found between shortwave cloud forcing and the modeled spread in SO surface air temperatures (Ceppi et al. 2012; Schneider and Reusch 2016). Work by Hyder et al. (2018) on the CMIP5 ensemble showed that these biases in cloud-related shortwave radiation were mostly due to errors in the stand-alone atmospheric model components used. Additionally, inaccuracies in the representation of the large-scale ocean circulation in the SO including the location of the ACC, strength and location of subtropical gyre boundary currents, and wind-driven upwelling and associated mixing of interior ocean waters with the sea surface, as well as eddy-induced transports and their parameterizations, all play a role in explaining regional temperature biases. Propagation of deep ocean warming to the sea surface related to climate model drift present in some model simulations also contributes to the near-surface temperature biases. An example of this from the CMIP3 and CMIP5 ensemble was the GFDL CM3 simulation with excessively warm biases in the abyssal ocean that impacted the SO surface climate in the historical simulations (Griffies et al. 2011).

Noting that large local errors can be obscured by only assessing models according to their zonal-mean properties, we present the upper 100-m temperature and salinity biases (Fig. 6) for the entire SO. For comparison to previous studies (Russell et al. 2006a; Sen Gupta et al. 2009; Beadling et al. 2019), we also provide the zonally averaged structure (Fig. 5). With the exception of four models (MIROC6, MIROC-ES2L, GISS-E2.1-H, and EC-Earth3-Veg), the CMIP6 zonal-mean temperature structure shows improved agreement across all latitudes relative to the spread found across models in CMIP3 and CMIP5 [Fig. 4c in Beadling et al. (2019) and Fig. 4a in Russell et al. (2006a)]. Model agreement tightens up north of 50°S. The MIROC6, MIROC-ES2L, GISS-E2.1-H, INM-CM4.8, and EC-Earth3-Veg models stand out in Fig. 6a and Fig. S3 with excessive warm biases that dominate the entire circumpolar SO south of 40°S. Many other models exhibit regions with biases in excess of 3°C that are mostly concentrated within or just north of the ACC region, with the warmest biases appearing mostly in the South Atlantic and Indian basins. All CMIP6 models tend to exhibit some degree of temperature bias along the southern margin of the subtropical gyres or along the boundary current regions such as the Brazil–Malvinas confluence zone and Agulhas retroflection. Regional temperature biases along the ACC margin and in the boundary current regions are likely related to discrepancies in model representation of the pathways of these large-scale currents as was the case in CMIP3 (Sen Gupta et al. 2009). The B-SOSE and GFDL-CM4 simulations are very similar, with the exception of more intense cold biases in the subtropical regions in GFDL CM4. These patterns and magnitudes of biases in near-surface temperatures have implications for the accurate representation of surface heat fluxes, SO ventilation of heat and carbon, water mass characteristics, and the ability to accurately represent the Antarctic SIE.

The near-surface SO salinity structure is characterized by relatively fresh water in the latitude band of the ACC where buoyancy is gained at the ocean surface via precipitation and sea ice meltwater is entrained into the Ekman drift (Figs. 5b and 6b). The upper ocean is slightly more saline south of the ACC, particularly in the Weddell and Ross Seas and along the Antarctic coast. North of the ACC, high-salinity subtropical waters dominate the upper ocean. In CMIP3 and CMIP5, models showed very wide discrepancies in the representation of upper ocean salinity, even in the zonal mean (Russell et al. 2006a; Sen Gupta et al. 2009; Beadling et al. 2019). On average, both the CMIP3 and CMIP5 ensembles yielded fresh biases throughout the entire SO (Sen Gupta et al. 2009; Beadling et al. 2019), with the largest fresh biases in the upper ocean north of
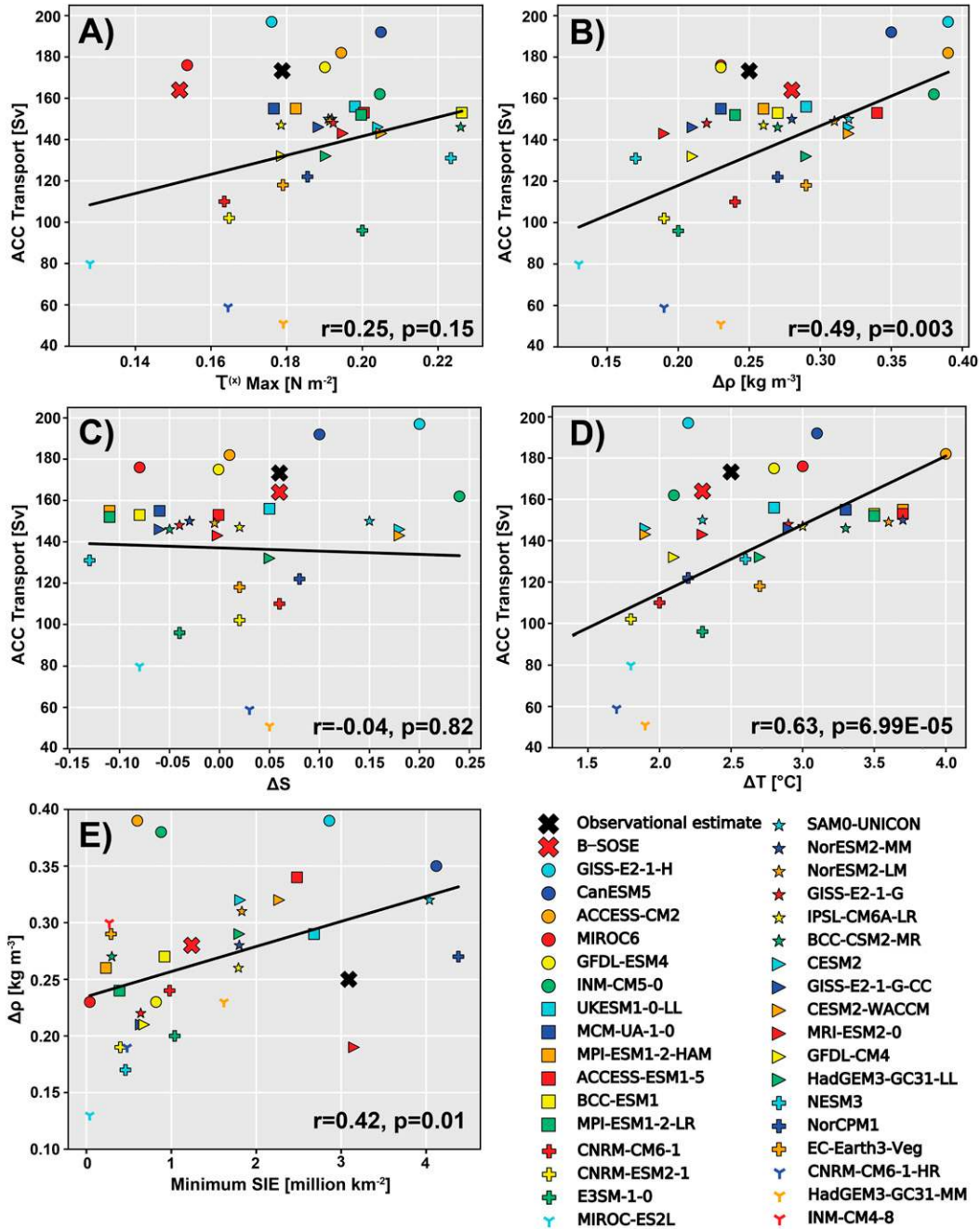
FIG. 4. ACC transport vs (a) zonally averaged maximum westerly wind stress and ACC transport vs full-depth-averaged, zonally averaged (b) meridional potential density, (c) potential temperature, and (d) salinity difference between 65° and 45°S. (e) Full-depth-averaged meridional potential density difference between 65° and 45°S vs minimum Antarctic sea ice extent (SIE). Mean observed and modeled values correspond to the values reported in Tables 2 and 3. The linear regression considering only the CMIP6 models and the corresponding correlation coefficient and *p* value (*n* − 2 degrees of freedom) are displayed on each panel.

50°S. CMIP5 models generally agreed with one another (but with a significant fresh bias) north of ~50°S where precipitation minus evaporation dominates the freshwater budget. Much larger intermodel spread

was found in the zonal-mean upper ocean salinity in the seasonal sea ice zone (Beadling et al. 2019). In CMIP5, the fresh biases in combination with the warm biases discussed above resulted in water masses that
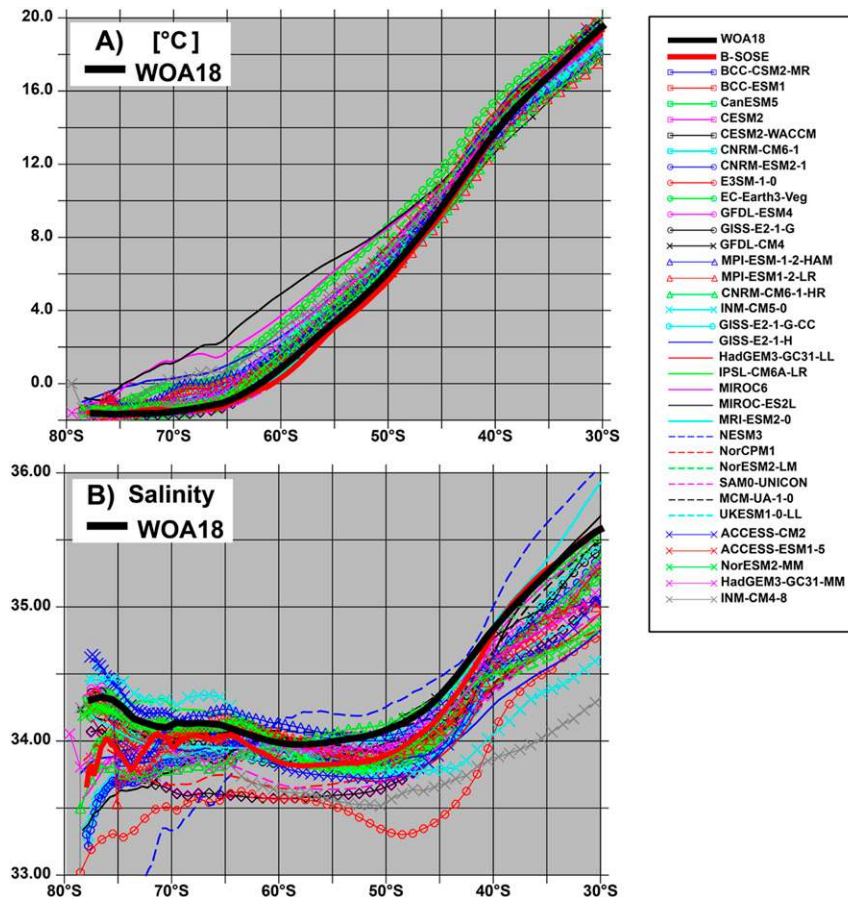
FIG. 5. The zonally averaged and time-averaged near-surface (a) potential temperature (°C; 0–100 m average) and (b) salinity (0–100-m average). The CMIP6 model output are time-averaged from January 1986 to December 2005. The January 2013–December 2018 time period is averaged for the B-SOSE Iteration 133 solution. The *WOA18* values are computed from the average of the 1985–94 and 1995–2004 decadal climatologies.

were too light in the ventilated layers of the SO (Sallée et al. 2013b).

In CMIP6, the near-surface zonal-mean salinity structure (Fig. 5b) shows improvement in the intermodel spread across all latitudes relative to previous generations [Fig. 4d in Beadling et al. (2019) and Fig. 4b in Russell et al. (2006a)]. However, the systematic bias of models generally being too fresh in the upper ocean persists into CMIP6. Many models exhibit intense fresh biases exceeding 0.50, concentrated mostly north of 45°S in the subtropical zones (Fig. 6b; see also Fig S4), however several models have fresh biases of these magnitudes along the Antarctic coast. Several models exhibit large saline biases along the Antarctic margin or within the ACC region. The circumpolar nature of many of these saline biases suggest they may be linked to the up-welling of saline North Atlantic Deep Water (NADW) within the ACC. These near-surface salinity biases combined with those of near-surface temperature (Fig. 6)

suggest that the water mass characteristics in the ventilated layers of the SO may differ widely among models, similar to CMIP5 (Sallée et al. 2013b). Given that the formation and properties of SO water masses are intimately tied to global ocean heat and carbon uptake, the substantial biases in the near-surface properties in CMIP6 are potentially worrying from a global climate modeling perspective.

### c. Interior ocean properties and gradients across the ACC region

The strength and structure of the ACC is largely determined by the meridional density gradient according to thermal wind balance. Thus, the ACC simulation is highly sensitive to biases in the interior ocean properties from the surface to the abyssal ocean. In coupled models the ACC generally spans the latitudes ~65°–45°S and the meridional difference in properties across these latitudes has been used as a metric to assess model
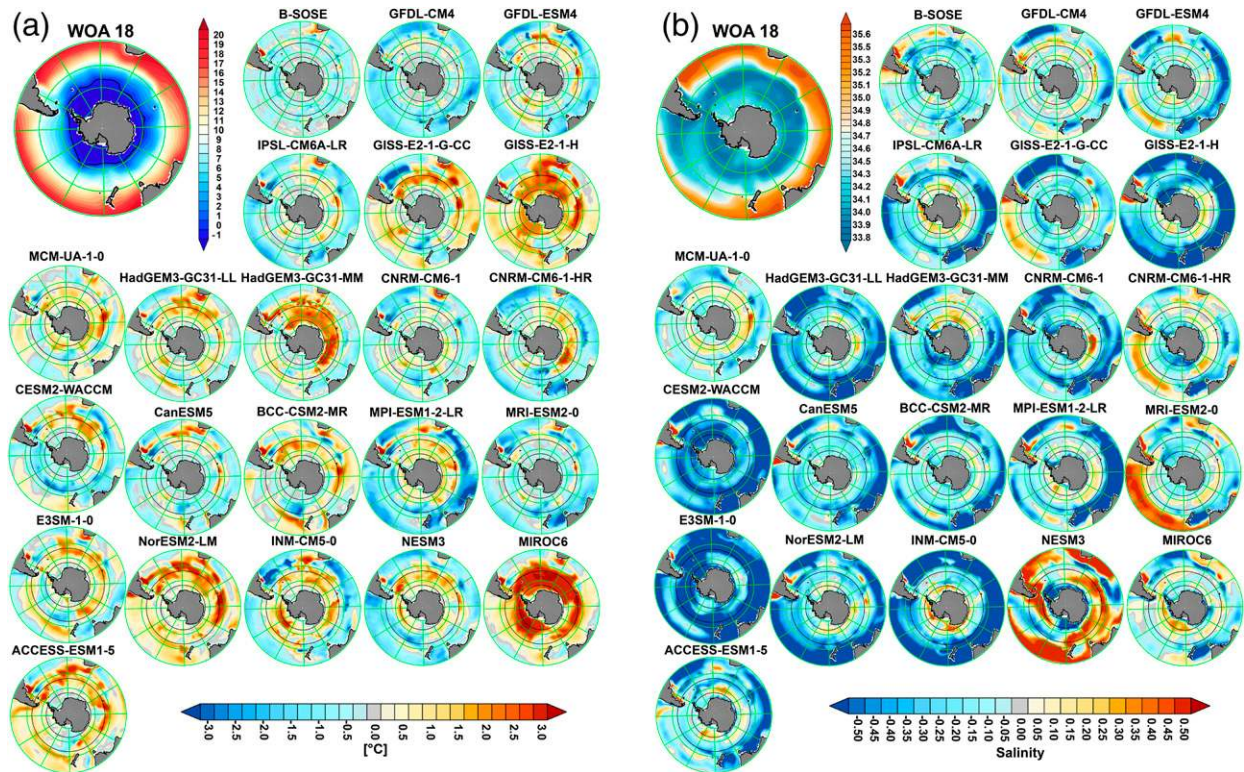
FIG. 6. (a) Upper-ocean potential temperature (°C) (0–100-m average) in the Southern Ocean from the *WOA18* climatological mean (large panel at top left; computed from the average of the 1985–94 and 1995–2004 decadal climatologies). The difference between the simulated and observed temperature (°C) for the B-SOSE Iteration 133 solution (time-averaged from January 2013 to December 2018) and for each CMIP6 model (time averaged from January 1986 to December 2005). Positive (from yellow to red) values indicate the model is warmer than observed. Negative (blue) values indicate the model is colder than observed. Additional models are shown in the supplemental material. (b) As in (a), but for salinity in the upper ocean (0–100-m average). Positive (from yellow to red) values indicate the model is more saline than observed. Negative (blue) values indicate the model is fresher than observed. Additional models are shown in the supplemental material.

performance (Russell et al. 2006a; Farneti et al. 2015; Beadling et al. 2019). We note that this is a crude simplification given that the meridional boundaries of the ACC are highly dependent on longitude and differ across models (Sen Gupta et al. 2009; Meijers et al. 2012). While not a precise computation according to thermal wind, where meridional gradients in density are vertically integrated to solve for baroclinic velocities, the zonally averaged and depth-averaged density contrast from 65° to 45°S used as a model metric serves as a proxy for the intensity of the isopycnal slopes across the current, which generally reflect the strength of the ACC.

Russell et al. (2006a) considered the zonally averaged and depth-averaged (0–1500 m; above the sill depth of the DP) difference in meridional potential density ($\Delta\rho$) from 65° to 45°S for the 18 CMIP3 models studied. Relative to the *WOA 2001* value of $0.58 \, \mathrm{kg \, m^{-3}}$ (this value remains the same when calculated from the *WOA18* product) used as the observational benchmark, a range from 0.18 to $0.97 \, \mathrm{kg \, m^{-3}}$ was found across models. Considering a

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

driving errors in ACC strength, despite accuracy in the wind stress forcing at the ocean surface (Beadling et al. 2019).

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

Similar to previous model generations, the strength of the ACC generally scales with the magnitude of $\Delta\rho$ across 65°–45°S (Fig. 4b). Upon breaking down the density difference into the contributing differences in $\Delta T$ and $\Delta S$, a wider spread in model performance is found (Table 2; Figs. 4c,d). Of the models that exhibit an excessive $\Delta\rho$, five are the result of having a much too large $\Delta S$ (GISS-E2.1-H, INM-CM5.0, SAM0-UNICON, CESM2, CESM2-WACCM), two are the result of too large $\Delta T$ (ACCESS-CM2, ACCESS-ESM1.5), and one exhibits large biases in both $\Delta S$ and $\Delta T$ (CanESM5).

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

Several models simulate a $\Delta S$ in the opposite direction from WOA18, with fresh biases in the water column on the southern edge of the ACC often accompanied by saline biases in the water column on the northern edge, concentrated in the upper 1500 m in the subtropical gyre regions or NADW regions (Fig. 7b). In many cases, biases in $\Delta T$ and $\Delta S$ compensate for one another yielding an accurate $\Delta\rho$. Many of the model biases in the temperature and salinity structure of the SO appear to be concentrated in the recently ventilated layers or in the deep Atlantic, suggesting most stem from inaccuracies in the surface climate (reflected in the biases in Fig. 6, Fig. S3, and Fig. S4) or inaccuracies in the simulation of NADW. An interesting relationship found in this analysis is a strong correlation between $\Delta\rho$ and the minimum Antarctic SIE (Fig. 4e), suggesting that the representation of the Antarctic SIE is important for achieving an accurate SO density structure likely through its influence on upper ocean salinity and water mass transformation processes. However, the mechanisms behind this deserve a detailed follow-up study.

### d. Antarctic sea ice extent

Antarctic sea ice exerts a strong influence on the SO through its impacts on air–sea heat exchange, local surface albedo, and the upper ocean freshwater budget. Recent work using numerical models and observations have highlighted that freshwater fluxes associated with buoyancy loss during sea ice formation and freshwater gain in the upper ocean from sea ice melt are dominant components in the transformation of water masses in the SO (Abernathey et al. 2016; Pellichero et al. 2018). Wind-driven sea ice export and subsequent delivery of freshwater have played a role in determining the observed mean salinity distribution in the upper SO (Haumann et al. 2016; Cerovečki et al. 2019). The representation of Antarctic SIE in models is complicated by the fact that sea ice is highly sensitive to both atmospheric and ocean forcing. Errors in the representation of Antarctic SIE can *result in* or *be the result of* errors in biases in SO temperature and salinity structure, patterns of surface wind stress forcing, water mass properties and location of deep ocean upwelling, and geostrophic ocean circulation such as the strength and pathway of the ACC. Furthermore, the representation of Antarctic sea ice in historical simulations of coupled models has been shown to be linked to projected changes in the Southern Hemisphere westerly jet, where models with larger Antarctic sea ice area in their historical simulations exhibit more sea ice retreat and less strengthening of the jet under increased radiative forcing throughout the twenty-first century (Flato 2004; Bracegirdle et al. 2015; Bracegirdle et al. 2018). Biases in Antarctic sea ice representation can
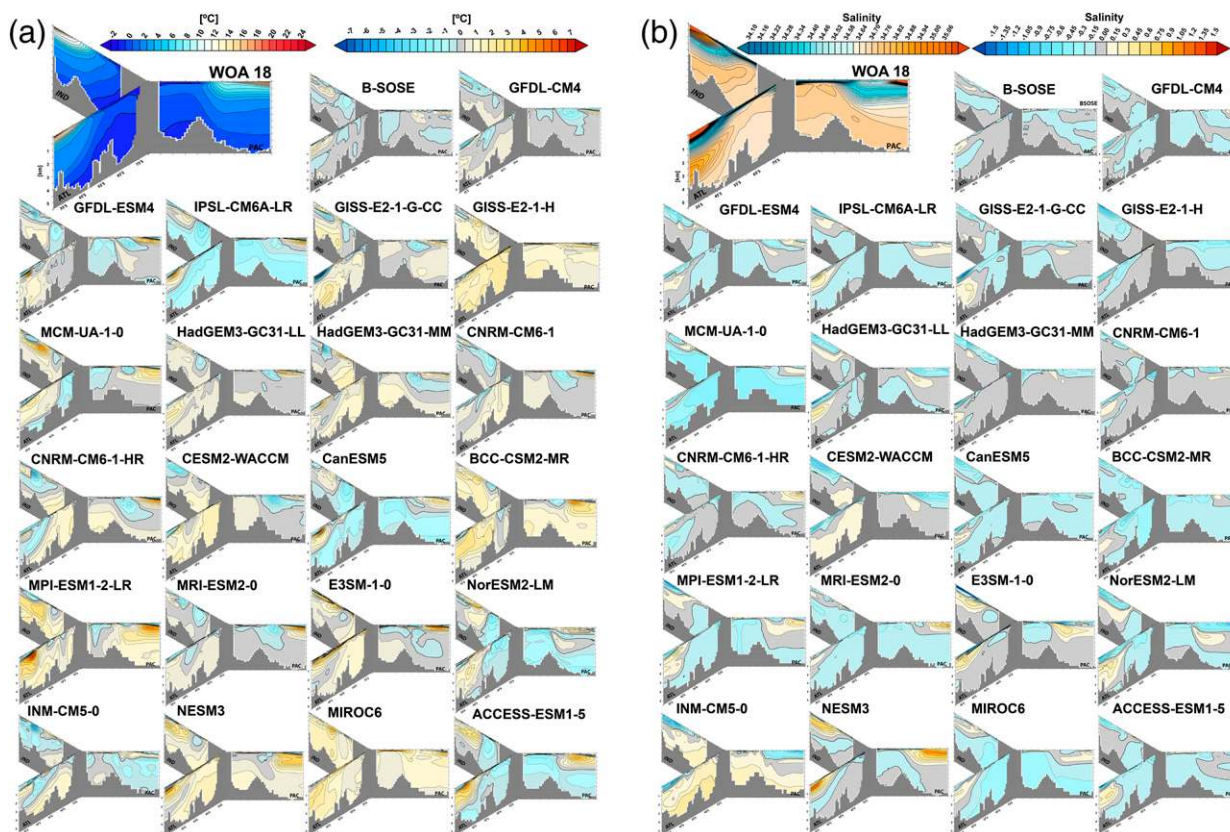
FIG. 7. (a) Potential temperature (°C) in the Southern Ocean from the *WOA18* climatological mean (large panel at top left; computed from the average of the 1985–94 and 1995–2004 decadal climatologies) in the Atlantic (30°W), Indian (90°E), and Pacific Ocean (150°W) and the difference between the simulated and observed temperature for the B-SOSE Iteration 133 solution (time averaged from January 2013 to December 2018), and for each CMIP6 model (time-averaged from January 1986 to December 2005). Positive (from yellow to red) values indicate the model is warmer than observed. Negative (blue) values indicate the model is colder than observed. Additional models are shown in the supplemental material. (b) As in (a), but for salinity in the Southern Ocean. Positive (from yellow to red) values indicate the model is more saline than observed. Negative (blue) values indicate the model is fresher than observed. Additional models are shown in the supplemental material.

translate into errors in air–sea gas exchange and simulated heat and carbon storage through its impact on water mass transformation and through its complex coupling between surface air temperature, net precipitation, and the strength of the Southern Hemisphere westerly jet.

Accurate representation of mean-state Antarctic sea ice extent, area, and distribution has been an existing challenge in the climate modeling community, with Antarctic SIE generally poorly represented (Parkinson et al. 2006; Flato et al. 2013; Shu et al. 2015). Across generations, improvement in Antarctic SIE has been marginal relative to Arctic (Mahlstein et al. 2013). CMIP3 and CMIP5 models have shown errors in representing the overall Antarctic SIE and regional distributions, the seasonal cycle, and observed trends (Parkinson et al. 2006; Connolley and Bracegirdle 2007; Sen Gupta et al. 2009; Turner et al. 2013; Flato et al. 2013; Mahlstein et al. 2013; Shu et al. 2015; Ivanova et al. 2016; Roach et al. 2018). When the ensemble average of SIE

is used as a metric, it appears that models perform well and that significant improvements have been made from CMIP3 to CMIP5 (Sen Gupta et al. 2009; Flato et al. 2013; Shu et al. 2015). However, this averaging is severely misleading; large biases are seen among individual models. In many cases, even when the magnitude of SIE is numerically well represented, the spatial patterns are often completely unrealistic (Connolley and Bracegirdle 2007). Using an ensemble average of SIE as a metric is also misleading due to the differing degrees of strong model internal variability in this region across models (Mahlstein et al. 2013; Deser et al. 2010).

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

The analysis by Mahlstein et al. (2013) concludes that "the representations of Antarctic sea ice in CMIP5 models have not improved compared to CMIP3 and show an unrealistic spread in the mean state that may influence future sea ice behavior" (p. 5105). However, the IPCC AR5 report notes that "the CMIP5 multi-model ensemble exhibits improvements over CMIP3 in simulation of sea ice extent in both hemispheres" (Flato et al. 2013, p. 788).

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

There are some clear outliers among the models, including MIROC6, MIROC-ES2L, EC-Earth3-Veg, MPI-ESM-1.2-HAM, MPI-ESM1.2-LR, HadGEM3-GC31-MM, and INM-CM4.8, that simulate extremely low maximum SIE values (Table 3; Fig. 8a). The MIROC-ES2L and MIROC6 models have lower annual SIE values than any CMIP5 model (Fig. 8b) skewing the CMIP6 MMM toward lower values. Models with exceptionally low annual SIE values tend to have very large upper and interior ocean temperature biases ($>3°C$; Figs. 6a and 7a; see also Figs. S3 and S5), likely explaining the lack of seasonal sea ice.

The magnitude of the standard deviation ($2\sigma$) of the annual means over the 20-yr period differ widely across models and in most cases are often much larger than observed (Fig. 8b), suggesting that the internal variability differs markedly across models and is different from that observed. For example, compare the large variability in GISS-E2.1H, U.K.-ESM1.0-LL, HadGEM3-GC31-LL, IPSL-CM6A-LR, and BCC-ESM1 to that of NorESM2-LM, MIROC-ES2L, and MIROC6. This brings up the question of the representativeness of the time period chosen when making assessments of a model's performance across Antarctic sea ice metrics since there seems to be a significant contribution of

the model's internal variability to the performance, as was the case in previous generations (Mahlstein et al. 2013; Deser et al. 2010). Additionally, the presence of any large open-ocean polynyas (see discussion in section 3a) in a model's simulation would significantly impact the overall sea ice representation. The CMIP5 and CMIP6 intermodel spread for all SIE metrics in Table 3 is much larger than the observed interannual variability over this time period. Overall, it appears that the representation of Antarctic SIE still remains a systematic bias from CMIP5 to CMIP6.

## 4. Summary and conclusions

The SO is a dynamically complex region where the global ocean water masses converge and the world's strongest current, the ACC, exchanges properties between basins. Strong wind-driven surface divergence drives deep waters to the ocean surface on the southern side of the ACC, resulting in strongly sloped isopycnals that act as a vehicle for the exchange of deep ocean properties with that of the upper ocean and overlying atmosphere. A portion of the Circumpolar Deep Water (CDW) that is upwelled along these steeply sloped isopycnals is transformed into denser Antarctic Bottom Water (AABW) and exported northward, filling the abyssal ocean. Buoyancy gained at the ocean surface through the incorporation of precipitation, glacial runoff, and sea ice melt transforms the remaining upwelled CDW into lighter intermediate and mode waters that are exported northward in the upper ocean subtropical gyres. The ACC and intense water mass transformations that characterize the SO are intimately coupled together, and these complicated dynamics are directly tied to the exchange of heat and carbon with the atmosphere and the export of nutrients to the rest of the global ocean.

Lack of historical observations coupled with complex and not fully understood interconnected processes tied to the ocean eddy field, surface wind stress forcing, air–sea heat and freshwater fluxes impacted by both the ocean circulation and overlying atmospheric properties, properties of upwelled water, and the destruction and transformation of seasonal sea ice make the SO very difficult to accurately represent in climate models. Since these dynamics are highly coupled, a well-represented SO requires accuracy to be achieved in the atmosphere, ocean, and sea ice model subcomponents and their exchanges since errors in any of these regimes can propagate into the other. An important example discussed here is that in previous model generations, errors in representation of Southern Hemisphere clouds and cloud properties in the atmospheric model led to significant errors in the SO upper ocean thermal structure. Given that the surface temperature gradient is tightly coupled to the position and strength

TABLE 3. Metrics related to the representation of Antarctic sea ice extent (SIE) (million km$^2$) for the models considered in this study. The observed SIE metrics and their associated standard deviation (1$\sigma$) of annual means for the January 1986–December 2005 time period are from the monthly sea ice extent dataset (https://nsidc.org/data/g02135) from the National Snow and Ice Data Center Sea Ice Index. All CMIP6 values are time averaged from January 1986 to December 2005. B-SOSE values are computed from the time-averaged output from January 2013 to December 2018 for the Iteration 133 B-SOSE solution. In all calculations, only grid cells that had a sea ice concentration greater than 15% were included. Considering the standard deviations of the annual means over the time period of analysis for both the observational data and model output, values that lie 2$\sigma$ outside the mean of the observational metric are in bold and values that lie 2$\sigma$ below are bold and italicized. The B2019 CMIP5 multimodel mean (MMM) and intermodel spread (1$\sigma$) is from the analysis presented in Table S1 of Beadling et al. (2019). For the maximum and minimum SIE, the month when this occurs is also listed. Note that the MCM-UA-1.0 model, which has missing values marked with asterisks (*), only provides ice thickness and the concentration is either 1 or 0, preventing accurate 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 | —* | —* | —* | —* |
| 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 |

of the surface wind stress forcing, this can propagate into surface forcing biases. Additionally, as was shown here, near-surface property biases directly impact the ability to accurately represent Antarctic SIE, where models with an exceptionally warm SO simulated extremely low SIE. The upper ocean properties can also be impacted internally by poorly represented water

mass properties, such as too warm NADW or abyssal warming due to climate drift.

Despite these challenges, the analyses presented here, in conjunction with those performed on ensembles of CMIP3 and CMIP5 models, show some consistent improvements across generations. We have focused on observable metrics that have been analyzed in previous
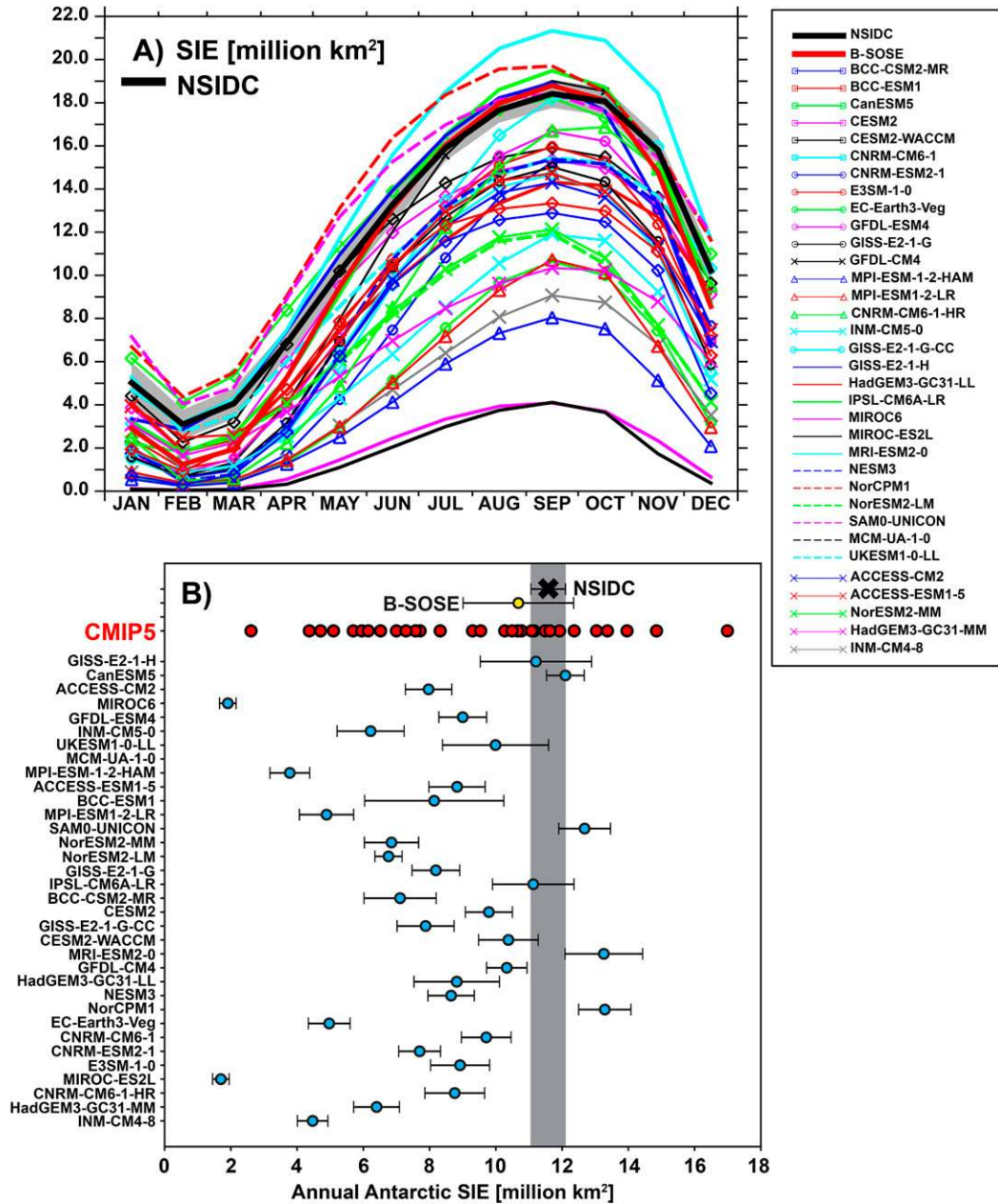
FIG. 8. (a) Seasonal climatology of Antarctic sea ice extent (SIE; million km$^2$) observed by satellite and reported by the National Snow and Ice Data Center Sea Ice (NSIDC) and as simulated in each CMIP6 model. The light gray shading about the NSIDC mean is the standard deviation ($2\sigma$) of the SIE for that month over the 20-yr period. Data from observations and model output are time-averaged from January 1986 to December 2005. Only grid cells that have a sea ice concentration greater than 15% are included in the computation of SIE. (b) Annual Antarctic SIE simulated across model generations: CMIP5 through CMIP6. The CMIP5 values are taken from Beadling et al. (2019). The CMIP6 values are computed identically to that of Beadling et al. (2019), over the same time period in the historical simulations.

model ensembles as well as included additional metrics pertaining to characterizing the wind stress forcing. The results suggest the following regarding model performance:

1) *The simulation of the strength of the ACC has improved from CMIP3 to CMIP6*, with modeled ACC strength converging toward the magnitude of observed net flow through the Drake Passage as

estimated by the cDrake array. While there are no CMIP6 models that exhibit unrealistically strong transports like previous generations, there are still several models producing exceptionally weak transports ($>7\sigma$ outside of the observational uncertainty). An area identified as a need for improvement is in the unrealistic multidecadal variability observed in the ACC transport for a number of the CMIP6 models. This behavior may be linked to the presence of unrealistic "superpolynya events" in the preindustrial control simulations associated with quasiperiodic episodes of intense open-ocean convection, which impact the interior ocean density structure and thus the ACC. The underlying mechanisms causing these events need to be further studied to improve the simulations. The coarse-resolution models ($\sim 1°$ or coarser) all employ various parameterization schemes and coefficient magnitudes for subgrid-scale ocean mixing and this likely plays a role in explaining such diversity in the vertical and horizontal structure of the ACC found across models. This resolution and parameterization choice dependence needs to be explored in detail across models of varying resolution.

2) *All metrics pertaining to the surface wind stress forcing have improved.* Notable improvements are found in the strength and position of the zonally averaged westerly wind stress maximum $\tau_{\max}^{(x)}$ relative to the ERA5 reanalysis product. The persistent systematic bias of equatorward winds in previous model generations has improved. These improvements have yielded a much more realistic pattern of wind stress curl over the SO and a narrowing of spread across the model ensemble. Noting that the magnitude and location of where wind-driven surface divergence pulls deep carbon-rich water to the surface is potentially important for the SO carbon budget, we have further characterized the wind stress forcing in the CMIP6 models by computing the magnitude and location of minimum wind stress curl over the SO (magnitude of maximum Ekman upwelling). These two metrics show the largest intermodel spread relative to the other wind stress metrics considered; a follow up study will identify if this plays a role in explaining model spread in biogeochemical performance here. Relative to CMIP3 and CMIP5, a stronger correlation, although still not statistically significant, is found for the relationship between ACC strength and $\tau_{\max}^{(x)}$. Previous generations had many models that exhibited extreme biases in their temperature and salinity gradients across the ACC that allowed for compensation for errors in the surface wind stress forcing. These errors are not as prevalent in the CMIP6 ensemble.

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

4) *The representation of the difference in density across the latitudes of the ACC has improved in CMIP6.* Considering the simulated zonally averaged, full-depth averaged difference in potential density referenced to the surface ($\Delta\rho$) across the ACC, CMIP6 models have improved in performance relative to that observed. The majority of models (71%) simulate a $\Delta\rho$ within a 25% error margin of the *WOA18* value. Two models yield a $\Delta\rho$ that is too weak and eight models yield a $\Delta\rho$ much greater than the *WOA18* value. It still remains fairly common for models to achieve an accurate $\Delta\rho$ as a result of compensating errors in the temperature and salinity structure.

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

The analysis presented here provides critical information on improved and existing biases in observable properties in the SO for the climate models that will provide projections of the climate system for the Intergovernmental Panel on Climate Change's Sixth Assessment Report (IPCC AR6). Given the dominant role that the SO plays relative to other ocean basins in the oceanic storage of heat and carbon, it is important for climate models to represent this region well in order to provide meaningful simulations of transient climate change. While models have generally improved across many metrics, the remaining biases associated with the temperature and salinity structure of the SO and sea ice representation may have serious implications for climate projections. The impacts of these biases on simulated ocean heat and carbon storage requires a detailed assessment.

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

## Data availability statement.

All of the CMIP6 model output analyzed in this manuscript is openly available from the Earth System Grid Federation (ESGF) CMIP6 archive: https://esgf-node.llnl.gov/projects/cmip6/. In the case of the GFDL-ESM4 model, native grid output for the uo and umo variables analyzed in this manuscript is available upon request from the Geophysical Fluid Dynamics Laboratory (GFDL). The B-SOSE Iteration 133 output used in this analysis is available at Scripps Institution of Oceanography: http://sose.ucsd.edu/BSOSE6_iter133_solution.html. The *WOA18* output is openly available from NOAA's National Centers for Environmental Information: https://www.nodc.noaa.gov/OC5/woa18/. The European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5 atmospheric reanalysis product is openly available from https://cds.climate.copernicus.eu/. The Antarctic sea ice extent data are openly available from the National Snow and Ice Data Center (NSIDC): https://nsidc.org/data/G02135/versions/3.

REFERENCES

Abernathey, R., I. Cerovecki, P. R. Holland, E. Newsom, M. Mazloff, and L. D. Talley, 2016: Water-mass transformation by sea ice in the upper branch of the Southern Ocean, overturning. *Nat. Geosci.*, **9**, 596–601, https://doi.org/10.1038/ngeo2749.

Adcroft, A., and J.-M. Campin, 2004: Rescaled height coordinates for accurate representation of free-surface flows in ocean circulation models. *Ocean Modell.*, **7**, 269–284, https://doi.org/10.1016/j.ocemod.2003.09.003.

——, and Coauthors, 2019: The GFDL global ocean and sea ice model OM4.0: Model, description and simulation features. *J. Adv. Model. Earth Syst.*, **11**, 3167–3211, https://doi.org/10.1029/2019MS001726.

Beadling, R. L., J. L. Russell, R. J. Stouffer, P. J. Goodman, and M. Mazloff, 2019: Assessing the quality of Southern Ocean circulation in CMIP5 AOGCM and Earth system model simulations. *J. Climate*, **32**, 5915–5940, https://doi.org/10.1175/JCLI-D-19-0263.1.

Bracegirdle, T. J., and G. J. Marshall, 2012: The reliability of Antarctic tropospheric pressure and temperature in the latest global reanalyses. *J. Climate*, **25**, 7138–7146, https://doi.org/10.1175/JCLI-D-11-00685.1.

——, E. Shuckburgh, J.-B. Sallée, Z. Wang, A. J. S. Meijers, N. Bruneau, and L. Wilcox, 2013: Assessment of surface winds over the Atlantic, Indian, and Pacific Ocean sectors of the Southern Ocean in CMIP5 models: Historical bias, forcing response, and state dependence. *J. Geophys. Res. Atmos.*, **118**, 547–562, https://doi.org/10.1002/jgrd.50153.

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

——, P. Hyder, and C. R. Holmes, 2018: CMIP5 diversity in southern westerly jet projections related to historical sea ice area: Strong link to strengthening and weak link to shift. *J. Climate*, **31**, 195–211, https://doi.org/10.1175/JCLI-D-17-0320.1.

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

Carsey, F. D., 1980: Microwave observation of the Weddell polynya. *Mon. Wea. Rev.*, **108**, 2032–2044, https://doi.org/10.1175/1520-0493(1980)108<2032:MOOTWP>2.0.CO;2.

Ceppi, P., Y.-T. Hwang, D. M. W. Frierson, and D. L. Hartman, 2012: Southern Hemisphere jet latitude biases in CMIP5 models linked to shortwave cloud forcing. *Geophys. Res. Lett.*, **39**, L19708, https://doi.org/10.1029/2012GL053115.

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

Chereskin, T. K., K. A. Donohue, and D. R. Watts, 2012: cDrake: Dynamics and transport of the Antarctic Circumpolar Current in Drake Passage. *Oceanography*, **25**, 134–135, https://doi.org/10.5670/oceanog.2012.86.

Chidichimo, M. P., K. A. Donohue, D. R. Watts, and K. L. Tracey, 2014: Baroclinic transport time series of the Antarctic Circumpolar Current measured in Drake Passage. *J. Phys. Oceanogr.*, **44**, 1829–1853, https://doi.org/10.1175/JPO-D-13-071.1.

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

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

Danabasoglu, G., S. C. Bates, B. P. Briegleb, S. R. Jayne, M. Jochum, W. G. Large, S. Peacock, and S. G. Yeager, 2012: The CCSM4 ocean component. *J. Climate*, **25**, 1361–1389, https://doi.org/10.1175/JCLI-D-11-00091.1.

——, and Coauthors, 2020: The Community Earth System Model version 2 (CESM2). *J. Adv. Model. Earth Syst.*, **12**, e2019MS001916, https://doi.org/10.1029/2019MS001916.

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

Deser, C., A. Phillips, V. Bourdette, and H. Teng, 2010: Uncertainty in climate change projections: The role of internal variability. *Climate Dyn.*, **38**, 527–546, https://doi.org/10.1007/s00382-010-0977-x.

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

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

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

——, and Coauthors, 2019: Taking climate model evaluation to the next level. *Nat. Climate Change*, **9**, 102–110, https://doi.org/10.1038/s41558-018-0355-y.

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

Fetterer, F., K. Knowles, W. N. Meier, M. Savoie, and A. K. Windnagel, 2017: Sea ice index, version 3. NSIDC: National Snow and Ice Data Center, accessed 17 November 2019, https://doi.org/10.7265/N5K072F8.

Firing, Y. L., T. K. Chereskin, and M. R. Mazloff, 2011: Vertical structure and transport of the Antarctic circumpolar current in Drake Passage from direct velocity observations. *J. Geophys. Res.*, **116**, C08015, https://doi.org/10.1029/2011JC006999.

Flato, G. M., 2004: Sea-ice and its response to $CO_2$ forcing as simulated by global climate models. *Climate Dyn.*, **23**, 229–241, https://doi.org/10.1007/s00382-004-0436-7.

——, and Coauthors, 2013: Evaluation of climate models. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 741–866.

Frölicher, T. L., J. L. Sarmiento, D. J. Paynter, J. P. Dunne, J. P. Krasting, and M. Winton, 2015: Dominance of the Southern Ocean in anthropogenic carbon and heat uptake in CMIP5 models. *J. Climate*, **28**, 862–886, https://doi.org/10.1175/JCLI-D-14-00117.1.

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

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

Golaz, J.-C., and Coauthors, 2019: The DOE E3SM coupled model version 1: Overview and evaluation at standard resolution.

*J. Adv. Model. Earth Syst.*, **11**, 2089–2129, https://doi.org/10.1029/2018MS001603.

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

Griffies, S. M., and Coauthors, 2011: The GFDL CM3 coupled climate model: Characteristics of the ocean and sea ice simulations. *J. Climate*, **24**, 3520–3544, https://doi.org/10.1175/2011JCLI3964.1.

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

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

Held, I. M., and Coauthors, 2019: Structure and performance of GFDL's CM4.0 climate model. *J. Adv. Model. Earth Syst.*, **11**, 3691–3727, https://doi.org/10.1029/2019MS001829.

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

Hyder, P., and Coauthors, 2018: Critical Southern Ocean climate model biases traced to atmospheric model cloud errors. *Nat. Commun.*, **9**, 3625, https://doi.org/10.1038/s41467-018-05634-2.

Ivanova, D. P., P. J. Gleckler, K. E. Taylor, P. J. Durack, and K. D. Marvel, 2016: Moving beyond the total sea ice extent in gauging model biases. *J. Climate*, **29**, 8965–8987, https://doi.org/10.1175/JCLI-D-16-0026.1.

Jungclaus, J. H., and Coauthors, 2013: Characteristics of the ocean simulations in the Max Planck Institute Ocean Model (MPIOM) the ocean component of the MPI-Earth system model. *J. Adv. Model. Earth Syst.*, **5**, 422–446, https://doi.org/10.1002/jame.20023.

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

Kiss, A. E., and Coauthors, 2020: ACCESS-OM2 v1.0: A global ocean–sea ice model at three resolutions. *Geosci. Model Dev.*, **13**, 401–442, https://doi.org/10.5194/gmd-13-401-2020.

Knutti, R., 2010: The end of model democracy? *Climatic Change*, **102**, 395–404, https://doi.org/10.1007/s10584-010-9800-2.

——, J. Sedláček, B. M. Sanderson, R. Lorenz, E. M. Fischer, and V. Eyring, 2017: A climate model projection weighting scheme accounting for performance and interdependence. *Geophys. Res. Lett.*, **44**, 1909–1918, https://doi.org/10.1002/2016GL072012.

Kuhlbrodt, T., R. S. Smith, Z. Wang, and J. M. Gregory, 2012: The influence of eddy parameterizations on the transport of the Antarctic Circumpolar Current in coupled climate models. *Ocean Modell.* **52–53**, 1–8, https://doi.org/10.1016/j.ocemod.2012.04.006.

——, and Coauthors, 2018: The low-resolution version of HadGEM3 GC3.1: Development and evaluation for global climate. *J. Adv. Model. Earth Syst.*, **10**, 2865–2888, https://doi.org/10.1029/2018MS001370.

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

Locarnini, R. A., and Coauthors, 2018: *Temperature*. Vol. 1, *World Ocean Atlas 2018*, NOAA Atlas NESDIS 81, 52 pp.

Mahlstein, I., P. R. Gent, and S. Solomon, 2013: Historical Antarctic mean sea ice area, sea ice trends, and winds in

CMIP5 simulations. *J. Geophys. Res. Atmos.*, **118**, 5105–5110, https://doi.org/10.1002/jgrd.5044.

Meijers, A. J. S., 2014: The Southern Ocean in the Coupled Model Intercomparison Project phase 5. *Philos. Trans. Roy. Soc.*, **372A**, 20130296, https://doi.org/10.1098/RSTA.2013.0296.

——, E. Shuckburgh, N. Bruneau, J. B. Sallée, T. Bracegirdle, and Z. Wang, 2012: Representation of the Antarctic Circumpolar Current in the CMIP5 climate models and future changes under warming scenarios. *J. Geophys. Res.*, **117**, 547–562, https://doi.org/10.1029/2012JC008412.

Meredith, M., and Coauthors, 2019: Polar regions. *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*, H.-O. Pörtner et al., Eds., https://www.ipcc.ch/srocc/, in press.

Munday, D. R., H. L. Johnson, and D. P. Marshall, 2013: Eddy saturation of equilibrated circumpolar currents. *J. Phys. Oceanogr.*, **43**, 507–532, https://doi.org/10.1175/JPO-D-12-095.1.

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

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

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

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

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

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

Roemmich, D., J. Church, J. Gilson, D. Monselesan, P. Sutton, and S. Wijffels, 2015: Unabated planetary warming and its ocean structure since 2006. *Nat. Climate Change*, **5**, 240–245, https://doi.org/10.1038/nclimate2513.

Russell, J. L., R. J. Stouffer, and K. W. Dixon, 2006a: Intercomparison of the Southern Ocean circulations in IPCC coupled model control simulations. *J. Climate*, **19**, 4560–4575, https://doi.org/10.1175/JCLI3869.1.

——, K. W. Dixon, A. Gnanadesikan, R. J. Stouffer, and J. R. Toggweiler, 2006b: The Southern Hemisphere westerlies in a warming world: Propping open the door to the deep ocean. *J. Climate*, **19**, 6382–6390, https://doi.org/10.1175/JCLI3984.1.

——, and Coauthors, 2018: Metrics for the evaluation of the Southern Ocean in coupled climate models and Earth system models. *J. Geophys. Res. Oceans*, **123**, 3120–3143, https://doi.org/10.1002/2017JC013461.

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

——, ——, ——, ——, T. J. Bracegirdle, Z. Wang, and T. Roy, 2013b: Assessment of Southern Ocean water mass circulation

and characteristics in CMIP5 models: Historical bias and forcing response. *J. Geophys. Res. Oceans*, **118**, 1830–1844, https://doi.org/10.1002/jgrc.20135.

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

Sen Gupta, A., A. Santoso, A. S. Taschetto, C. C. Ummenhofer, J. Trevena, and M. H. England, 2009: Projected changes to the Southern Hemisphere ocean and sea ice in the IPCC AR4 climate models. *J. Climate*, **22**, 3047–3078, https://doi.org/10.1175/2008JCLI2827.1.

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

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

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

Talley, L. D., and Coauthors, 2016: Changes in ocean heat, carbon content, and ventilation: A review of the first decade of GO-SHIP global repeat hydrography. *Ann. Rev. Mar. Sci.*, **8**, 185–215, https://doi.org/10.1146/ANNUREV-MARINE-052915-100829.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, https://doi.org/10.1175/BAMS-D-11-00094.1.

Turner, J., T. Bracegirdle, T. Phillips, G. J. Marshall, and J. S. Hosking, 2013: An initial assessment of Antarctic sea ice extent in the CMIP5 models. *J. Climate*, **26**, 1473–1484, https://doi.org/10.1175/JCLI-D-12-00068.1.

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

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

Whitworth, T., 1983: Monitoring the transport of the Antarctic Circumpolar Current at Drake Passage. *J. Phys. Oceanogr.*, **13**, 2045–2057, https://doi.org/10.1175/1520-0485(1983)013<2045:MTTOTA>2.0.CO;2.

——, and R. Peterson, 1985: Volume transport of the Antarctic Circumpolar Current from bottom pressure measurements. *J. Phys. Oceanogr.*, **15**, 810–816, https://doi.org/10.1175/1520-0485(1985)015<0810:VTOTAC>2.0.CO;2.

——, W. Nowlin, and S. Worley, 1982: The net transport of the Antarctic Circumpolar Current through Drake Passage. *J. Phys. Oceanogr.*, **12**, 960–971, https://doi.org/10.1175/1520-0485(1982)012<0960:TNTOTA>2.0.CO;2.

Yukimoto, S., and Coauthors, 2019: The Meteorological Research Institute Earth System Model version 2.0, MRI-ESM2.0: Description and basic evaluation of the physical component. *J. Meteor. Soc. Japan*, **97**, 931–965, https://doi.org/10.2151/jmsj.2019-051.

Zweng, M. M., and Coauthors, 2018: *Salinity*. Vol. 2, *World Ocean Atlas 2018*, NOAA Atlas NESDIS 82, 50 pp.