Assessing the Quality of Southern Ocean Circulation in CMIP5 AOGCM and Earth System Model Simulations

R. L. BEADLING, J. L. RUSSELL, R. J. STOUFFER, AND P. J. GOODMAN

Department of Geosciences, The University of Arizona, Tucson, Arizona

M. MAZLOFF

Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California

(Manuscript received 5 April 2019, in final form 6 June 2019)

ABSTRACT

The Southern Ocean (SO) is vital to Earth's climate system due to its dominant role in exchanging carbon and heat between the ocean and atmosphere and transforming water masses. Evaluating the ability of fully coupled climate models to accurately simulate SO circulation and properties is crucial for building confidence in model projections and advancing model fidelity. By analyzing multiple biases collectively across large model ensembles, physical mechanisms governing the diverse mean-state SO circulation found across models can be identified. This analysis 1) assesses the ability of a large ensemble of models contributed to phase 5 of the Coupled Model Intercomparison Project (CMIP5) to simulate observationally based metrics associated with an accurate representation of the Antarctic Circumpolar Current (ACC), and 2) presents a framework by which the quality of the simulation can be categorized and mechanisms governing the resulting circulation can be deduced. Different combinations of biases in critical metrics including the magnitude and position of the zonally averaged westerly wind stress maximum, wind-driven surface divergence, surface buoyancy fluxes, and properties and transport of North Atlantic Deep Water entering the SO produce distinct mean-state ACC transports. Relative to CMIP3, the quality of the CMIP5 SO simulations has improved. Eight of the thirty-one models simulate an ACC within observational uncertainty (2σ) for approximately the right reasons; that is, the models achieve accuracy in the surface wind stress forcing and the representation of the difference in the meridional density across the current. Improved observations allow for a better assessment of the SO circulation and its properties.

1. Introduction

The Southern Ocean (SO) plays a dominant role in the uptake of excess planetary heat, accounting for 67%–98% of the observed increase in heat content of the global ocean (Roemmich et al. 2015). Furthermore, considering heat and carbon uptake in historical simulations from coupled climate models contributed to phase 5 of the Coupled Model Intercomparison Project (CMIP5), the SO accounts for ~43% of the anthropogenic CO₂ uptake and ~75% of the excess heat

uptake by the ocean over the historical period (1861–2005) (Frölicher et al. 2015). Given the key role it plays in the planetary carbon and heat budgets, an accurate simulation of SO circulation by fully coupled climate and Earth system models (ESMs) is vital for building confidence in model projections of future climate under increased radiative forcing and advancing model fidelity.

The SO is a dynamically complex region where deep waters that have been isolated from the atmosphere for hundreds of years are upwelled to the ocean surface as a result of strong wind-driven surface divergence. The surface forcing from the strong westerly wind stress, properties of the upwelled water, and surface buoyancy forcing result in the strong eastward Antarctic Circumpolar Current (ACC), which wraps around the Antarctic continent, transporting large amounts of heat and freshwater both zonally and meridionally (Tamsitt et al. 2016)

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JCLI-D-19-0263.s1>.

Corresponding author: Rebecca L. Beadling, Beadling@email.arizona.edu

DOI: 10.1175/JCLI-D-19-0263.1

© 2019 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

and transforming global ocean water masses. The complex horizontal and vertical circulation that exists in the SO, coupled with the lack of historical observations of SO properties and dynamics, has led to large disagreements among coupled climate and ESMs with respect to the simulation of physical and biogeochemical processes in this region (Russell et al. 2006a; Sen Gupta et al. 2009; Kuhlbrodt et al. 2012; Meijers et al. 2012; Bracegirdle et al. 2013; Sallée et al. 2013a,b; Frölicher et al. 2015; Russell et al. 2018).

For accurate future projections of global climate change over the twenty-first century under increased radiative forcing, it is crucial that a strong understanding of the mean-state circulation and the mechanisms governing it across climate models is obtained. Russell et al. (2006b) showed, using two coupled climate models that were nearly identical except for distinct westerly wind regimes over the SO, that the biases in the mean-state representation of SO divergence directly impact the amount of future oceanic heat and carbon storage under increased radiative forcing. Thus, models with distinctly different ocean circulation structures due to differing patterns and magnitudes of surface wind forcing or other physical properties, such as surface heat and freshwater fluxes or water mass structures, are expected to respond differently in response to increasing atmospheric greenhouse gas concentrations. In other words, errors in the representation of the mean-state ocean circulation across models will contribute to large uncertainties and intermodel spread in the response to increased radiative forcing.

Several previous studies have assessed various aspects of SO mean-state circulation and associated properties across various ensembles of CMIP5 models (Meijers et al. 2012; Bracegirdle et al. 2013; Downes and Hogg 2013; Sallée et al. 2013a,b; Frölicher et al. 2015; Russell et al. 2018), the majority of which are summarized in a review paper by Meijers (2014). The assessment of individual biases is crucial; however, only by assessing biases in key metrics in combination with one another can a framework be built to aid in identifying and understanding the major contributors to a model's ability to achieve an accurate SO simulation. For example, it is not enough just to know whether a model performs well at simulating a certain physical process or property, but it is crucial to be able to decipher whether its accurate performance is *right for the right reasons* or if its accuracy is achieved through compensating errors, that is, *right for the wrong reasons*. If a model has an inaccurate simulation, it is important to understand the physical mechanisms governing this divergence from realism.

A study by Russell et al. (2006a) analyzed physically based metrics in combination with one another to determine

what drives and determines the ACC in the preindustrial control integrations of 18 CMIP3 models. The metrics included the representation of the surface westerly wind stress forcing, surface freshwater and heat fluxes, differences in the meridional density, temperature, and salinity above the sill depth of the Drake Passage (DP), and the properties of North Atlantic Deep Water (NADW) export from the Atlantic into the SO. Russell et al. (2006a) found that the most important indicator for assessing the likely SO simulation performance was the strength of the westerly wind forcing over the open DP latitudes and the most significant internal ocean contributor to the differing ACC structures across models was the properties and transport of saline NADW into the SO from the Atlantic. By assessing the performance of multiple parameters in combination with one another across an ensemble of models and categorizing the models according to which criteria they failed to meet, Russell et al. (2006a) was able to present plausible mechanisms that helped to explain the intermodel differences in the ability to simulate the ACC.

This study is motivated by the analysis and framework presented by Russell et al. (2006a) in their assessment of CMIP3 preindustrial control simulations. Using more recent generation CMIP5 models and updated observational estimates, this study assesses the SO simulation in the historical experiments of 31 different CMIP5 models, with a particular focus on the representation of the ACC transport. Importantly, this study utilizes the updated estimate of the ACC transport through the DP obtained from measurements from the cDrake array (Chereskin et al. 2012; Chidichimo et al. 2014; Donohue et al. 2016), which was not available at the time of the CMIP3 assessment or for early assessments of CMIP5 models. The high temporal and spatial resolution measurements from the cDrake experiment allowed strong near-bottom currents in the DP to be resolved for the first time, shifting the canonical benchmark for the ACC transport from 134 Sv ($1 \text{ Sv} \equiv 10^6 \text{ m}^3 \text{ s}^{-1}$) (Whitworth and Peterson 1985; Cunningham et al. 2003) to $173.3 \pm 10.7 \text{ Sv}$ (Donohue et al. 2016).

We evaluate key metrics associated with the ability to accurately represent the ACC including the simulation of surface wind stress, differences in the meridional properties across the ACC, surface buoyancy fluxes, temperature and salinity biases, and transport and properties of NADW entering the SO. Given the sparse nature of SO observed quantities such as those derived from the pre-Argo era, we also analyze the results from the Biogeochemical Southern Ocean State Estimate (BSOSE), which incorporates modern in situ measurements to provide a state estimate against which we can additionally evaluate climate models.

Using physical reasoning, we categorize the CMIP5 models based on their deficiencies in the simulation of key observationally based metrics that influence the ACC. The analysis across such a large ensemble provides us with a means to understand the diversity in the ability to simulate the large-scale SO circulation. Furthermore, using such a large model ensemble and a large number of observational metrics provides an important resource to the community and a comprehensive baseline of SO metrics against which the next generation of CMIP6 simulations can be assessed for model improvements.

2. Methods

a. Model data

Thirty-one CMIP5 models are chosen for use in this study based on the availability of data provided from each model's "historical" simulation provided in the Earth System Grid Federation (ESGF) CMIP5 data archive. The historical simulation covers much of the industrial period (~ 1860 –2005) and is forced by observed anthropogenic and natural sources of atmospheric composition changes and time-evolving land cover (Taylor et al. 2012). Only the first ensemble member for each model's historical simulation is analyzed. All model output presented is a 20-yr average of monthly data from January 1986 to December 2005 of the historical simulation. Table 1 lists the models and contains additional details regarding their ocean component. All calculations were performed on each model's native grid, or the grid on which the data were reported at ESGF for isopycnal coordinate models. For all models considered, the horizontal ocean grid is uniform south of 30°S and thus no regridding is needed prior to integration for transports or calculation of zonal-mean fields.

Some of the models used in this analysis share the same atmosphere and/or ocean components and/or similar parameterizations and thus may contain similar biases, and therefore should not be viewed as completely independent from one another (Knutti et al. 2010). Furthermore, each model has differing times used for their respective spinup simulations, resulting in varying degrees of equilibrium reached by the time of the branching of the historical simulation from its parent preindustrial control integration (piControl). Given that the historical simulation is on the order of ~ 156 years (1850–2005), the integrations presented have "built in" deep ocean biases from their respective piControl and spinup integrations since the historical period is much too short for the deep ocean properties to differ much from their initial states at the start of the historical

experiment. Furthermore, each modeling center performs a different model spinup procedure before the start of the piControl integration, using varying initial conditions for their baseline climate state. The above-mentioned caveats should be kept in mind when assessing the differences among models in the simulated properties at the end of the historical simulations relative to that observed.

b. Observational estimates

In situ temperature and salinity output from the *World Ocean Atlas 2013* (WOA13) product (Locarnini et al. 2013; Zweng et al. 2013) is used to compute the temperature, salinity, and potential density differences across the SO against which the simulations are compared. To be consistent with the time frame of the historical simulations, the climatologies are computed from the average of the 1985–94 and 1995–2004 decadal climatologies provided by the WOA13 product. The WOA13 in situ temperature is converted to potential temperature prior to calculations.

When assessing the simulation of ocean surface heat flux and surface freshwater flux in the CMIP5 models, we consider monthly output from two different reanalysis products over the 1986 to 2005 period: the National Centers for Environmental Prediction's (NCEP) Climate Forecast Reanalysis (CFSR) product (Saha et al. 2010) and the European Centre for Medium-Range Weather Forecasts' (ECMWF) ERA-Interim global atmospheric reanalysis (Dee et al. 2011). For the total ocean surface heat flux, the following terms were combined: surface sensible heat flux, surface latent heat flux, surface net shortwave radiation, and surface net longwave radiation. For the ocean surface freshwater flux, only precipitation minus evaporation (PmE) is considered.

In addition to the CFSR and ERA-Interim reanalysis products, we also analyzed NASA's Modern-Era Retrospective Analysis for Research and Applications (MERRA) atmospheric reanalysis (Rienecker et al. 2011). A third product was analyzed due to the large downward trend in the wind stress magnitude found in the CFSR reanalysis product from 1986 to 2005. The MERRA product also exhibited a downward trend over the same time period but of a lesser magnitude than that in CFSR. The downward trends found in the MERRA and CFSR data are inconsistent with documented strengthening of the westerly wind stress over the SO found in other reanalysis products and observational datasets (Swart and Fyfe 2012).

Bracegirdle et al. (2013) use ERA-Interim as their benchmark to assess the skill of CMIP5 models in their simulation of Southern Hemisphere winds, citing indications that "ERA-Interim is the most reliable of

TABLE 1. CMIP5 models used in this study. The table includes the ocean component and its version, ocean horizontal resolution (longitude $\times$ latitude), vertical coordinate with the number of layers/levels, and modeling center. The vertical coordinates are defined as follows: z , traditional depth coordinates; σ_2 , isopycnal vertical coordinates; z^* , rescaled geopotential vertical coordinate for better representation of free-surface variations (Adcroft and Campin 2004). The historical experiments are simulations from 1850 to 2005 forced by observed changes in atmospheric composition due to anthropogenic and natural sources. These changes include anthropogenic and volcanic activity, solar forcing, emissions or concentrations of short-lived species and natural and anthropogenic aerosols or their precursors, and land use change (Taylor et al. 2012). (Expansions of acronyms are available online at <http://www.ametsoc.org/PubsAcronymList>.)

	CMIP5 model	Ocean component	Ocean resolution (lon $\times$ lat)	Ocean vertical coordinate and levels	Modeling center
1	ACCESS1.0	MOM4.1	$1.0^\circ \times 1.0^\circ$	z^* (50)	CSIRO-BOM
2	ACCESS1.3	MOM4.1	$1.0^\circ \times 1.0^\circ$	z^* (50)	CSIRO-BOM
3	BCC_CSM1.1	MOM4-L40v1	$1.0^\circ \times 1.0^\circ$	z (40)	BCC-CMA
4	BCC_CSM1.1(m)	MOM4-L40v2	$1.0^\circ \times 1.0^\circ$	z (40)	BCC-CMA
5	BNU-ESM	MOM4.1	$1.0^\circ \times 1.0^\circ$	z^* (50)	BNU
6	CanESM2	NCOM	$1.41^\circ \times 0.94^\circ$	z (40)	CCCma
7	CCSM4	POP2	$1.11^\circ \times 0.54^\circ$	z (60)	NCAR
8	CESM1(CAM5)	POP2	$1.11^\circ \times 0.54^\circ$	z (60)	NSF-DOE-NCAR
9	CNRM-CM5	NEMO3.2	$1.0^\circ \times 0.65^\circ$	z (42)	CNRM-CERFACS
10	CNRM-CM5.2	NEMO3.2	$1.0^\circ \times 0.65^\circ$	z (42)	CNRM-CERFACS
11	CSIRO Mk3.6.0	MOM4.2	$1.88^\circ \times 0.93^\circ$	z^* (31)	CSIRO-QCCCE
12	FGOALS-g2	LICOM2	$1.0^\circ \times 1.0^\circ$	z (30)	LASG/IAP
13	FGOALS-s2	LICOM2	$1.0^\circ \times 1.0^\circ$	z (30)	LASG/IAP
14	GFDL-ESM2G	GOLD	$1.0^\circ \times 1.0^\circ$	σ_2 (59 + 4)	NOAA/GFDL
15	GFDL-ESM2M	MOM4.1	$1.0^\circ \times 1.0^\circ$	z^* (50)	NOAA/GFDL
16	GFDL CM2.1	MOM4.0	$1.0^\circ \times 1.0^\circ$	z (50)	NOAA/GFDL
17	GFDL CM3	MOM4.1	$1.0^\circ \times 1.0^\circ$	z^* (50)	NOAA/GFDL
18	GISS-E2-H-CC	HYCOM	$1.0^\circ \times 1.0^\circ$	$z - \sigma_2$ (26)	NASA/GISS
19	GISS-E2-R-CC	Russell Ocean	$1.25^\circ \times 1.0^\circ$	z^* (32)	NASA/GISS
20	HadCM3	HadOM3	$1.25^\circ \times 1.25^\circ$	z (20)	MOHC
21	HadGEM2-CC	HadGOM1	$1.0^\circ \times 1.0^\circ$	z (40)	MOHC
22	HadGEM2-ES	HadGOM1	$1.0^\circ \times 1.0^\circ$	z (40)	MOHC
23	IPSL-CM5A-LR	NEMO3.2	$1.98^\circ \times 1.30^\circ$	z (31)	IPSL
24	IPSL-CM5A-MR	NEMO3.2	$1.98^\circ \times 1.30^\circ$	z (31)	IPSL
25	IPSL-CM5B-LR	NEMO3.2	$1.98^\circ \times 1.30^\circ$	z (31)	IPSL
26	MIROC-ESM	COCO3.4	$1.4^\circ \times 1.4^\circ$	$z - \sigma$ (44)	JAMSTEC
27	MIROC-ESM-CHEM	COCO3.4	$1.4^\circ \times 1.4^\circ$	$z - \sigma$ (44)	JAMSTEC
28	MRI-CGCM3	MRI.COM3	$1.0^\circ \times 0.5^\circ$	z (50)	MRI
29	MRI-ESM1	MRI.COM3	$1.0^\circ \times 0.5^\circ$	z (50)	MRI
30	NorESM1-M	MICOM	$1.13^\circ \times 0.54^\circ$	σ_2 (53)	NCC
31	NorESM1-ME	MICOM	$1.13^\circ \times 0.54^\circ$	σ_2 (53)	NCC

contemporary reanalyses” over the SO. More specifically, the detailed assessment of various reanalysis products and observations by Swart and Fyfe (2012) indicates that ERA-Interim is a reliable product over the SO for wind stress. Therefore, when assessing the skill of the CMIP5 models in simulating parameters related to the representation of the zonal wind stress forcing at the ocean surface, we only compare the simulated values against ERA-Interim. All reanalysis output was downloaded from NASA’s Collaborative REAnalysis Technical Environment (CREATE) data system (Potter et al. 2018).

Ocean fields from the iteration 105 solution of the BSOSE are also analyzed in this study (http://sose.ucsd.edu/BSOSE_iter105_solution.html). BSOSE assimilates observations from profiling floats, shipboard data, underway measurements, and satellites into a numerical model to

produce a state estimate for the SO from January 2008 to December 2012 (Verdy and Mazloff 2017). No in situ data from the cDrake array (Chereskin et al. 2012) are assimilated into BSOSE. The SOSE is based on the MIT general circulation model (MITgcm) with the Biogeochemistry with Light, Iron, Nutrients, and Gases (BLING) model implemented (Galbraith et al. 2010). For additional details on BSOSE see Verdy and Mazloff (2017). In this analysis we present the time-averaged (monthly data from January 2008 to December 2012) fields from BSOSE at $1/3^\circ$ resolution.

c. Calculation of volume transports

To compute volume transports associated with the ACC and meridional transports across 32°S in the Atlantic in BSOSE, the time-averaged (January 2008–December 2012) meridional and zonal velocities were used. For the CMIP5

models, meridional and zonal mass transport data (vmo, umo) time averaged from January 1986 to December 2005 of the historical experiment were used (velocities were used if mass transport data were not provided). Mass transport is preferred here for transport calculations for the following reasons: 1) The vmo diagnostic is defined as the velocity times mass sampled at each time step, providing a better representation of the true time-averaged flow; 2) this diagnostic automatically accounts for any partial cells at the bottom of the ocean due to topography and at the top due to each model's free surface layer; 3) vmo is computed on each model's native vertical coordinate, reducing analytical errors from interpolating from nonstandard depth coordinates (e.g., isopycnal coordinates) to the standard depth coordinates that the output is reported on in the CMIP5 archive; and 4) analytical errors associated with the tracer grid points being located at different locations than the velocity grid points are minimized.

Meridional volume transports across 32°S in the Atlantic are calculated for BSOSE and the CMIP5 models by using the reported temperature and salinity fields to first bin the grid points closest to 32°S into density classes (σ_θ , σ_2 , and σ_4) according to the layer definitions provided in Table 14 of Talley (2008). Each layer is further subdivided into four equal sublayers to account for any canceling within the layer. The transports through these binned layers are then computed from the respective meridional velocity or mass transport. If mass transport is used, each layer is divided by a uniform density of 1035 kg m^{-3} . Salt transport (kg s^{-1}) is computed by multiplying each grid cell's salinity by the volume transport through the grid cell and summing through the layer. Temperature transport ($1 \text{ PWT} = 10^{15} \text{ J s}^{-1}$) in each layer is calculated using meridional mass transport (or velocity if mass transport is not provided), temperature (°C), in situ density (kg m^{-3}), and specific heat ($\text{J kg}^{-1} \text{ °C}^{-1}$). This quantity is referred to as temperature transport as defined by Hall and Bryden (1982) and used by Talley (2003) and Talley (2008) rather than heat transport ($1 \text{ PW} = 10^{15} \text{ J s}^{-1}$) because mass is *not* conserved in the individual layers.

d. Transport of the ACC through the Drake Passage

We consider the volume transport through the DP to be representative of the strength of the ACC as in Russell et al. (2006a) and Russell et al. (2018). For each CMIP5 model and BSOSE, the volume transport is computed as the net transport time averaged through the DP at the closest grid cell to 69°W. The flow through the DP provides an important observational metric due to the repeat measurements made in this region over the last several decades.

Early observational estimates of the net transport through the DP calculated using data from the International Southern Ocean Study (ISOS) (Whitworth et al. 1982; Whitworth 1983; Whitworth and Peterson 1985) yielded three separate estimates from a yearlong deployment (1979–80) with an average value of $134 \pm 11.2 \text{ Sv}$ (standard deviation of the yearlong transport). The later occupation of the World Ocean Circulation Experiment (WOCE) SR1b line between 1993 and 2000 yielded a net transport through the DP of $136 \pm 7.8 \text{ Sv}$ (Cunningham et al. 2003). Furthermore, combining repeat occupations of the DP region over the 1975–2000 time period, Cunningham et al. (2003) showed that over this 25-yr period, the baroclinic transport of the ACC above and relative to 3000 dbar demonstrated long-term stability. These early studies did not have sufficient spatial or temporal resolution required to resolve the strong near-bottom currents flowing through the DP.

Recently, an improved estimate of the total transport of the ACC through the DP has become available as a result of the cDrake experiment carried out from 2007 to 2011 (Chereskin et al. 2012; Chidichimo et al. 2014; Donohue et al. 2016). The flow through the DP was measured at high temporal and spatial resolution, resolving strong near-bottom currents, which were not resolved in earlier estimates, shifting the net transport estimate from 134 Sv (Whitworth and Peterson 1985; Cunningham et al. 2003) to $173.3 \pm 10.7 \text{ Sv}$ (Donohue et al. 2016). This estimate is the result of combining the $127.7 \pm 5.9 \text{ Sv}$ baroclinic transport (Chidichimo et al. 2014) and the $45.6 \pm 8.9 \text{ Sv}$ barotropic transport calculated from the cDrake array (Donohue et al. 2016). Importantly, the baroclinic component from ISOS (Whitworth and Peterson 1985), SR1b (Cunningham et al. 2003), and cDrake (Chidichimo et al. 2014) agree well.

The larger total transport from the cDrake experiment (Donohue et al. 2016) relative to earlier estimates is attributed to the depth-independent or barotropic flow that was able to be resolved due to the large improvement in the temporal and spatial resolution of sampling. Donohue et al. (2016) note that the lower bound of their DP estimate is consistent with the maximum transport value of Cunningham et al.'s (2003) error corrected DP estimate of 163 Sv , and agrees well with two modern, independent ACC estimates (Firing et al. 2011; Colin de Verdière and Ollitrault 2016).

Donohue et al. (2016) do not ascribe the increase to a climatologically forced signal from increased westerly winds based on supporting research using satellite altimetry records and model simulations by Hogg et al. (2015) and Bishop et al. (2016). This lack of increase in the ACC in response to the observed poleward shift and

intensification of SO westerly wind stress is understood through the theory of eddy saturation in which the additional energy imparted by the wind stress has increased the ocean eddy kinetic energy, opposing the steepening of isopycnals (and thus the acceleration of the ACC) expected from the increased wind forcing (Hallberg and Gnanadesikan 2006; Böning et al. 2008; Meredith et al. 2012; Munday et al. 2013; Hogg et al. 2015).

Prior to the cDrake experiment, the mean ACC transport estimate of 134 Sv served as an observational metric to strive for in model simulations including the CMIP5 models analyzed here. After the CMIP5 models were developed, the results from the cDrake experiment became available, providing a much higher benchmark for the ACC flow. Given this scenario, it is expected (and shown below) that most of the CMIP5 models simulate a weak ACC compared to the new cDrake value, which we use in this study to assess the model's performance. Given the lack of evidence of any long-term trend in the observed ACC flow through the DP as discussed above, we do not find issue with comparing the 1986–2005 time average for the models with the cDrake estimate from 2007–11.

3. Results

a. ACC transport

There is a wide range of ACC transports simulated by the CMIP5 models, ranging from the exceptionally high value of 246 Sv simulated by GISS-E2-R-CC to the exceptionally weak transports of 88 and 94 Sv simulated by the CNRM-CM5 and IPSL-CM5B-LR models, respectively (Fig. 1, Table 2). Only 10 of the CMIP5 models have an ACC transport that lies within the uncertainty (2σ) of the Donohue et al. (2016) estimate. BSOSE also has an ACC within the range of Donohue et al. (2016). As noted previously, BSOSE does not assimilate any in situ data from the cDrake array and thus represents an independent estimate. Several models cluster just below the range of observational uncertainty (2σ), falling outside by only a few Sverdrups (FGOALS-g2, CESM1(CAM5), ACCESS1.0, GFDL CM2.1, and CNRM-CM5.2). Overall, 3 models have a too strong transport and 13 have much too weak transport. Relative to the CMIP3 generation of models (Russell et al. 2006a; Sen Gupta et al. 2009), the 31 models considered here show improvement with respect to the strength of the ACC.

We note the caveat here that the ACC transport through the DP exhibits long-term variability (multi-decadal or longer) for several of the models when

considering the entire historical period (not shown here). This variability also appears in the piControl integrations and the reason for this behavior likely varies from model to model. For other models, the ACC transport remains stationary, exhibiting no long-term variability, throughout the entire period. Given this behavior, the use of a single ensemble member and only considering the 20-yr period at the end of the historical simulation adds additional uncertainty to the analysis. The five models that lie only outside of the 2σ range of the observed mean by <10 Sv might not be statistically different from the observed ACC transport if more than one ensemble member were used.

b. Surface forcing of the Southern Ocean

1) SURFACE MOMENTUM FORCING FROM WIND STRESS

In CMIP3 models, the strength of the ACC in piControl integrations generally scaled with the strength of the Southern Hemisphere zonally averaged westerly wind stress maximum (Russell et al. 2006a). Given this strong relationship found in the previous generation of climate models, we begin our examination of the representation of key metrics associated with the ACC in CMIP5 models by analyzing the structure of the wind stress forcing over the SO. As our benchmark for assessing the performance of the CMIP5 models, we compare the model-simulated wind stress metrics against the mean values from the ERA-Interim reanalysis product over the 1986–2005 time period. As discussed previously (see section 2), ERA-Interim is the most reliable reanalysis product over the SO (Bracegirdle and Marshall 2012; Swart and Fyfe 2012; Bracegirdle et al. 2013).

For each metric, we consider a model to be consistent with ERA-Interim if its mean value falls within 2σ (standard deviation of the annual means over the 1986–2005 period) of the ERA-Interim mean value for the metrics considered. We consider the following metrics in our analysis: the strength and position of the zonally averaged maximum westerly wind stress, integrated zonal wind stress over the open DP latitudes, integrated westerly wind stress south of 30°S, and integrated wind stress curl (WSC) over the open DP latitudes (Table 2).

The CMIP5 models perform well in their ability to accurately simulate the magnitude of the zonally averaged westerly wind stress maximum over the SO, with 26 out of the 31 models simulating a magnitude within 2σ of the ERA-Interim mean (Table 2, Fig. 2b). Relative to ERA-Interim, four models simulate a much too weak

FIG. 1. Volume transport of the ACC through the Drake Passage (Sv). The transport as observed from the cDrake experiment by Donohue et al. (2016) is plotted at the top with gray shading corresponding to its associated 2σ error. The CMIP5 transports are the January 1986–December 2005 time-averaged net transport through the Drake Passage and their associated 2σ standard deviation of annual values about the mean. The BSOSE transport is the December 2008–January 2012 time-averaged net transport through the Drake Passage and its associated 2σ standard deviation of annual values about the mean. For comparison to transports in the CMIP3 generation of models, the mean transports through the Drake Passage reported in Table 1 of Russell et al. (2006a) from the last 20 years of each model's piControl integration and the mean transports reported in Table 3 of Sen Gupta et al. (2009) from the last 20 years of each model's twentieth-century control run are plotted above the CMIP5 results.

zonally averaged westerly wind stress maximum and only one model simulates a magnitude that is much too strong (Table 2, Figs. 2a,c). The region with the largest disagreement among CMIP5 models in their simulated zonal wind stress is in the open latitudinal band of the DP (Fig. 2). The total wind stress over this region differs widely across models and is unsurprisingly significantly correlated ($r = 0.94$, $p = 3.0 \times 10^{-15}$) to the position

of the maximum wind stress. Models with their peak wind stress shifted farther equatorward intuitively have weaker surface forcing by the westerly winds over the DP latitudes (Table 2). The magnitude of the zonally averaged westerly wind stress maximum in BSOSE is significantly weaker relative to that of ERA-Interim, which is likely related to the different choice of bulk formula used to compute wind stress. This disagreement

TABLE 2. Parameters related to the strength of the ACC. The potential density ($\Delta\rho$), salinity (ΔS), and potential temperature (ΔT) differences are the zonally and full-depth-averaged difference between 65° and 45°S. The observational estimates of the differences against which the models are compared are calculated from the *WOA13* product (averaged 1985–94 and 1995–2004 decadal climatologies) (Locarnini et al. 2013; Zweng et al. 2013). The observational estimate for the surface wind stress parameters presented here is the ERA-Interim 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) (1σ). 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 (1σ). Please see the methods section for more details on how this estimate is obtained. Total τ_x and total wind stress curl are the integrals of the zonal wind stress and its associated wind stress curl over the open Drake Passage latitudes (55°–64°S). All model values are the time-averaged data from the last 20 years of the historical simulation (January 1986–December 2005). BSOSE values are computed from the time-averaged data from January 2008 to December 2012 for the BSOSE iteration 105 solution. Given the lack of annual resolution in the *WOA13* product and likely large uncertainty from the pre-Argo era, we do not attempt to provide an uncertainty for the property differences in the three columns of the table marked with an asterisk (*). For BSOSE and the CMIP5 models, mean values that are 2σ above the mean of the observational metric are in bold, and values that lie 2σ below are bold and italicized. Considering a window of error of 25% of the mean *WOA13* value for the density difference, values are in bold if the difference is too strong and bold and italicized if the difference is too weak using this criterion. Please see the discussion section (section 4) for a description and discussion of the criteria used to determine which category the CMIP5 models fall into.

Model	ACC (Sv)	Max τ_x (N m^{-2})	Total wind stress curl in DP (10^6 N m^{-1})	$\Delta\rho$ (kg m^{-3})*	ΔS (psu)*	ΔT (°C)*	Total τ_x in DP (10^{12} N)	Total westerly τ_x (10^{12} N)	Lat of max τ_x (°S)
Observational estimate	173.3 ± 10.7	0.1898 ± 0.01	-2.72 ± 0.21	0.27	0.06	2.5	2.5 ± 0.3	11 ± 0.5	52.27 ± 1.05
BSOSE	181	0.1505	-2.01	0.26	0.06	2.6	2.1	8.5	52.22
Reasonable ACC for approximately the right reasons									
CCSM4	169	0.2135	-2.97	0.34	0.17	2.4	3.1	11	53.76
CanESM2	153	0.2080	-2.91	0.32	0.17	2.2	2.5	12	50.86
HadGEM2-CC	178	0.1759	-2.78	0.29	0.01	3.2	1.9	9.8	50.39
HadGEM2-ES	172	0.1844	-2.85	0.30	0.02	3.1	2.1	10	50.66
MIROC-ESM	174	0.1796	-2.49	0.28	0.11	2.1	1.1	11	46.73
MIROC-ESM-CHEM	175	0.1808	-2.58	0.29	0.12	2.2	1.2	11	47.17
ACCESS1.3	184	0.2023	-3.17	0.34	0.03	3.9	2.3	11	51.62
GFDL CM3	156	0.1817	-2.43	0.27	-0.04	3.2	2.2	10	50.89
Accurate simulation of key metrics but weak ACC									
ACCESS1.0	146	0.1883	-2.88	0.30	0.002	3.4	2.2	11	51.44
GFDL CM2.1	144	0.1825	-2.70	0.23	-0.01	2.6	2.2	10	51.14
FGOALS-s2	133	0.1988	-3.13	0.23	0.04	2.6	1.6	11	49.35
NorESM1-M	127	0.1755	-2.40	0.25	-0.01	2.9	2.4	9.2	53.38
NorESM1-ME	126	0.1903	-2.53	0.27	0.06	2.5	2.7	9.5	53.84
CSIROMk3.6.0	109	0.1996	-2.85	0.25	0.13	1.6	1.8	11	49.15
Accurate wind stress forcing with errors in density gradient									
HadCM3	242	0.1755	-2.40	0.47	0.34	2.4	1.9	9.8	49.99
GISS-E2-H-CC	205	0.1640	-2.43	0.40	0.18	2.8	2.1	9.7	51.19
CESM1(CAM5)	146	0.1896	-2.68	0.36	0.19	2.3	2.7	10	53.42
GFDL-ESM2M	131	0.1704	-2.44	0.19	-0.05	2.4	2.1	9.6	51.45
GFDL-ESM2G	112	0.1722	-2.48	0.15	-0.12	3.1	2.3	9.8	52.13
MRI-CGCM3	116	0.1794	-2.94	0.19	0.01	1.9	2.1	9.7	52.41
MRI-ESM1	112	0.1766	-2.83	0.19	0.01	1.9	2.0	9.6	52.08
IPSL-CM5A-MR	108	0.1869	-2.90	0.18	0.02	2.1	1.1	10	47.05
IPSL-CM5A-LR	102	0.1751	-2.37	0.14	-0.02	2.0	0.59	9.6	45.48
BNU-ESM	107	0.2097	-3.02	0.13	-0.01	1.4	1.5	11	49.04
Errors in wind stress forcing with accurate density gradient									
BCC_CSM1.1(m)	171	0.2360	-3.53	0.29	-0.01	3.6	3.1	13	52.37
BCC_CSM1.1	157	0.2057	-3.16	0.27	-0.01	3.3	1.8	11	50.12
FGOALS-g2	147	0.1443	-1.89	0.26	0.07	2.7	0.59	8.3	45.70
CNRM-CM5-2	142	0.1589	-2.27	0.25	0.02	2.7	1.9	9.1	51.20
IPSL-CM5B-LR	94	0.1408	-1.90	0.20	-0.06	2.4	0.86	8.5	46.65
CNRM-CM5	88	0.1557	-2.50	0.29	0.14	1.7	1.7	9.0	50.61
Errors in wind stress forcing and errors in density gradient									
GISS-E2-R-CC	246	0.1668	-2.23	0.38	-0.11	5.2	2.0	10	49.97

FIG. 2. The (left) zonally averaged zonal wind stress (N m^{-2}) and (right) zonally integrated annual mean wind stress curl (N m^{-2}) from the ERA-Interim global atmospheric reanalysis (Dee et al. 2011) product time averaged over the January 1986–December 2005 time period (thick black line), from January 2008 to December 2012 of iteration 105 of the BSOSE solution, and from the January 1986 to December 2005 of the historical simulation for each CMIP5 model. The latitude band of the Drake Passage is indicated by the dark gray shading in each panel. Models are separated by having (a) weak, (b) accurate, and (c) excessive simulated peak zonal wind stress magnitude relative to the ERA-Interim mean and its associated interannual variability (2σ) (standard deviation of annual means over the 20-yr period). The light gray shading about the ERA-Interim mean represents the interannual variability (2σ) of the zonally averaged zonal wind stress in the left panel or zonally integrated wind stress curl in the right panel at each latitude.

does not appear to be an issue of comparison of different time periods between the two products.

While only a few models exhibit biases in the strength of the zonally averaged maximum westerly wind stress, the latitudinal position of the maximum is not well represented for many models (Table 2, Fig. 2). Twelve of the models have their peak zonally averaged westerly

wind stress biased equatorward relative to ERA-Interim, with 11 out of the 12 having a position north of 50°S . None of the models exhibit a poleward bias.

The varied skill across models in representing both the strength of the zonally averaged westerly wind stress maximum and where it is located leads to very different patterns of zonally integrated WSC across the SO

(Fig. 2). The pattern of WSC sets the amount of upwelling and downwelling south and north of the westerly wind stress maximum and thus has major influences on the density gradients across the ACC. Relative to ERA-Interim, four models exhibit too weak and three models exhibit too strong integrated WSC over the open DP latitudes. Interestingly, many models that exhibit significant equatorward biases in the position of their zonally averaged westerly wind stress maximum are still able to maintain a strong meridional gradient in the zonal wind stress and thus accurate Ekman suction over the open DP latitudes. With the exception of CNRM-CM5, the models with a weak zonally averaged westerly wind stress maximum tend to have weak integrated WSC over the open DP latitudes (Table 2, Fig. 2a).

Considering the overall relationships of the wind stress metrics considered and the strength of the ACC, a relatively weak and statistically insignificant relationship is found between the ACC strength and the magnitude of the zonally averaged westerly wind stress maximum ($r = 0.10$, $p = 0.58$) (Fig. 3a). The correlations between the ACC strength and the total westerly wind stress over the open DP latitudes and the ACC and the total westerly wind stress over the entire SO yield slightly larger but still statistically insignificant positive correlations ($r = 0.24$, $p = 0.19$ and $r = 0.26$, $p = 0.16$, respectively). The relationship between ACC strength and the latitude of the zonally averaged westerly wind stress maximum is also statistically insignificant ($r = 0.10$, $p = 0.59$). Several models have excessively strong ACC transports [>195 Sv ($>2\sigma$)] in the absence of strong westerly wind stress. Even with these outliers removed, the correlations do not become exceptionally strong. These relatively weak correlations indicate there must be other processes at play in these models that compensate for the weak wind stress biases, allowing a strong current to be maintained in the presence of weak mechanical forcing at the ocean surface.

2) NET SURFACE HEAT AND FRESHWATER FLUXES

The net surface heat flux at the ocean surface consists mainly of radiative (net shortwave and longwave radiation) and turbulent (latent and sensible) heat fluxes in addition to those associated with the formation and melt of sea ice and to a lesser extent those associated with precipitation/evaporation and river runoff. Here we consider the *net* surface heat flux, not the individual components. This is reported in the CMIP5 archive as heat flux downward at the ocean surface (hfds) from the ocean model output that contains *all* the components that contribute to the net flux of heat at the ocean surface. If the model reported *all* of the individual terms instead of hfds, all the terms were summed to obtain the net ocean surface heat flux. The same is true for the net surface freshwater flux.

The majority of the CMIP5 models report a single variable, water flux at the ocean surface (wfo), which contains all the sources of freshwater at the surface, such as that from sea ice formation/melt, river runoff, icebergs (if included in the model), precipitation, and evaporation.

Freshwater and heat fluxes in the SO from observationally based products are *highly* uncertain and different flux products can show quite different magnitudes and patterns. Thus, the ERA and CFSR reanalysis products presented should be taken as a “best guess” but not necessarily an accurate estimate as direct observations still remain too sparse to confirm these values. For this reason, we do not assess the skill of the models in representing the surface heat or freshwater fluxes relative to either reanalysis product. Instead we simply highlight here the disagreement among models in their zonally averaged heat and freshwater fluxes across the SO.

In the reanalysis products, the zonally averaged (across all basins) net surface heat flux shows the general structure of ocean heat loss south of $\sim 63^\circ$ S, ocean heat gain from $\sim 63^\circ$ to 43° S, and ocean heat loss north of 43° S (Fig. 4a). However, there are clear inconsistencies between the two different products throughout the entire SO. The surface heat flux simulated by BSOSE is inconsistent with both of the reanalysis products, with a wavier structure in the region of heat gain, with an inflection in the pattern occurring between 60° and 55° S likely a result of the Malvinas Current. A large spread is found in the surface heat fluxes simulated by the CMIP5 models south of 55° S, potentially associated with the differing simulation of sea ice extent and surface radiation balance.

A large disagreement in the surface freshwater fluxes among CMIP5 models is also seen south of 55° S, supporting the inference that this spread in both the heat and freshwater fluxes is likely linked to diversity in the skill of the CMIP5 models in their representation of the seasonal sea ice extent and the associated drift of and subsequent melt of sea ice away from the coast (Figs. 4a,b). Large inconsistencies among CMIP5 models in their representation of the annual mean Southern Hemisphere sea ice extent and its seasonal amplitude have been documented in other studies (Shu et al. 2015; Ivanova et al. 2016). An assessment of parameters related to the representation of Antarctic sea ice extent for the 31 models considered in this study are summarized in Table S1 in the online supplemental material.

Outside of the seasonal sea ice zone, where the freshwater flux is governed by precipitation minus evaporation (PmE), the CMIP5 models are generally consistent with one another, with the two reanalysis products falling within the model spread (Fig. 4b). BSOSE shows a starkly different pattern of the zonally averaged freshwater flux relative to the reanalysis data and the CMIP5 models, even outside of the sea ice zone.

FIG. 3. ACC transport (Sv) and (a) zonally averaged maximum westerly wind stress, full-depth-averaged, zonally averaged meridional (b) potential density, (c) potential temperature, and (d) salinity difference between 65° and 45°S. Mean observed and modeled values correspond to the values reported in Table 2. The linear regression considering only the CMIP5 models and the corresponding correlation coefficient and p value (29 degrees of freedom) are displayed on each panel. The model groups correspond to the category that the models fall into in Table 2 [see the discussion section (section 4) for details of model groupings].

Despite disagreements among the surface heat fluxes in BSOSE and the CMIP5 models, a very consistent structure of the zonally averaged upper 100-m temperature is found among models (Fig. 4c). BSOSE simulated upper 100-m

temperature is nearly identical to the WOA13 product, as expected given the assimilation of temperature data into the state estimate. Much greater inconsistencies are seen in the structure of the zonally averaged upper 100-m salinity

FIG. 4. The zonally averaged and time-averaged (a) surface heat flux, (b) surface freshwater flux, (c) near-surface potential temperature ($^{\circ}\text{C}$; 0–100-m average), and (d) near-surface salinity (psu; 0–100-m average). The CFSR and ERA-Interim reanalysis products and the CMIP5 model data are time averaged from January 1986 to December 2005. The January 2008–December 2012 time period is averaged for the BSOSE Iteration 105 solution. The *WOA13* output is computed from the average of the 1985–94 and 1995–2004 decadal climatologies provided in the *WOA13* product (see the methods section).

among models (Fig. 4d). Again, BSOSE simulates the upper 100-m salinity in high agreement with the *WOA13* product. All CMIP5 models except the IPSL models generally underestimate the upper 100-m salinity north of 55°S , outside of the DP latitudinal band, particularly in the subtropical gyres. The large deviations between models in their upper 100-m salinity south of 55°S are likely related to the large disagreements in the surface freshwater fluxes found here (Fig. 4b) combined with the differing properties of the upwelled water, particularly NADW, within the ACC and south of the DP.

c. Difference in meridional density across the ACC and its structure

The meridional density gradient across the ACC is important in setting and maintaining the strong eastward

flow of the current through thermal wind balance. A very strong relationship ($r = 0.83$, $p = 8.5 \times 10^{-9}$) exists between the density difference ($\Delta\rho$) across the current [full-depth-averaged, zonally averaged difference in potential density (referenced to the surface) from 65° – 45°S] and the ACC strength (Fig. 3b). We note here that the northward extent of the ACC is specific to each model and varies with longitude, with notable northward excursions over part of the Atlantic; however, 65° – 45°S captures the majority of the eastward flow across the models studied here and has been used in previous studies when assessing properties across the ACC (Russell et al. 2006a; Meijers et al. 2012). When $\Delta\rho$ is broken down into the contributing differences in temperature (ΔT) and salinity (ΔS) (Figs. 3c,d), the relationship between the ACC and ΔT yields a robust statistically significant positive correlation ($r = 0.62$,

$p = 0.0002$), while the correlation between the ACC and ΔS exhibits a weaker and not statistically significant positive correlation ($r = 0.33$, $p = 0.07$). The properties contributing to each model's resulting $\Delta\rho$ differ from model to model (Table 2, Figs. 5a,b).

The CMIP5 models differ greatly in their ability to accurately simulate $\Delta\rho$ relative to that observed (Table 2). The models exhibit a large intermodel spread in $\Delta\rho$, ranging from 0.47 kg m^{-3} (HadCM3) to 0.13 kg m^{-3} (BNU-ESM). The uncertainty in $\Delta\rho$ computed from the WOA13 product from the average of the 1985–94 and 1995–2004 decadal climatologies is not well constrained. Considering a window of 25% error ($0.20\text{--}0.34\text{ kg m}^{-3}$) relative to the WOA13 $\Delta\rho$ (0.27 kg m^{-3}), 20 CMIP5 models have a reasonable $\Delta\rho$, 4 have a too strong $\Delta\rho$, and 7 have a too weak $\Delta\rho$ (Table 2).

Examining the structure of the salinity and temperature differences (columns 6 and 7 of Table 2) that contribute to the resulting $\Delta\rho$, we see a wide diversity across models (Table 2, Figs. 5a,b). Unsurprisingly, all of the CMIP5 models show some degree of bias with respect to the full-depth-averaged SO temperature and salinity structure relative to the WOA13 product (Figs. 5a,b). Several of the models that have a $\Delta\rho$ within 25% of the WOA13 value have biases in ΔT and ΔS that compensate for one another. For example, the CSIRO-Mk3.6.0 and CNRM-CM5 models exhibit differences in salinity that are more than twice as large as the ΔS calculated from the WOA13 product, attributed to both too saline and too fresh waters south and north of the ACC, respectively. These larger ΔS magnitudes are offset by weak ΔT . Other models exhibit the opposite pattern, with excessive ΔT coupled with relatively weak ΔS [GFDL CM3, ACCESS1.0, BCC_CSM1.1, BCC_CSM1.1(m)]. In some cases, models fall at the upper or lower bounds of the 25% $\Delta\rho$ error window due to an excessive or weak difference in one property yet a reasonable difference in the other (CCSM4, CanESM2, IPSL-CM5B-LR, ACCESS1.3).

Of the four models that exhibit excessive $\Delta\rho$ ($>0.34\text{ kg m}^{-3}$), three are dominated by too strong ΔS and one is dominated by an excessive ΔT . The HadCM3 and GISS-E2-R-CC models stand out in this group (Figs. 5a,b). The HadCM3 model has a ΔS of 0.34 PSU relative to the WOA13 value of 0.06 PSU, with this excessive difference clearly attributed to too saline waters south of the ACC (Fig. 5a). The HadCM3 ΔT across the ACC is well represented. The GISS-E2-R-CC model has a ΔT of 5.2°C relative to the WOA13 value of 2.5°C , with the major driver of this bias being a much too warm ocean on the ACC's northern edge (Fig. 5b).

Among the models with too weak $\Delta\rho$ ($<0.20\text{ kg m}^{-3}$) across the ACC, the contributions from biases in ΔT and ΔS differ across models (Table 2). The main contributor to the weak $\Delta\rho$ in the GFDL-ESM2M and

GFDL-ESM2G models appears to be fresh biases south of the ACC, resulting in a ΔS of opposite sign than found in WOA13. The other models that exhibit weak $\Delta\rho$ have both weak ΔT and ΔS across the current. The BNU-ESM model stands out in Fig. 5b due to the large warm biases in the ocean south of the ACC, which likely explains the exceptionally low ΔT relative to WOA13.

d. Atlantic transports and properties

A major internal contributor to the resulting density structure across the ACC is the transport and properties of the various water masses in the SO. Of particular importance is the simulation of the transport and properties of NADW entering the SO. Russell et al. (2006a) provided evidence that the most significant internal ocean contribution to the variation in the simulation of SO mean circulation in CMIP3 models is their ability to export salty NADW into the SO. The structure of the meridional flow observed in the South Atlantic across 32°S consists of 1) northward transport in the subducted thermocline (from the surface to $26.40\sigma_\theta$) in which the northward surface Ekman transport is embedded, 2) northward transport in the lower thermocline ($26.40\sigma_\theta\text{--}26.90\sigma_\theta$), 3) northward transport of Antarctic Intermediate Water (AAIW) ($26.90\sigma_\theta\text{--}27.40\sigma_\theta$), 4) southward transport of North Atlantic Deep Water (NADW)/Upper Circumpolar Deep Water (UCDW) ($27.40\sigma_\theta\text{--}45.86\sigma_4$), and 5) northward flow of Antarctic Bottom Water (AABW) ($>45.86\sigma_4$) (Talley 2008) (Fig. 6). For our purposes we will refer to the NADW/UCDW layer as NADW.

In the observed ocean, the southward-flowing NADW layer is bounded by the $27.40\sigma_\theta$ and $45.86\sigma_4$ isopycnals (Fig. 6), which are found on average at approximately 1300 and 3500 m, respectively. The NADW layer is clearly seen as a region of enhanced subsurface salinity below the northward-flowing fresh AAIW layer (Fig. 6a). The CMIP5 models differ widely in their ability to accurately simulate the salinity and density structure relative to that derived from the WOA13 product (Fig. 6a).

There is a clear disconnect between the $27.40\sigma_\theta$ and $45.86\sigma_4$ isopycnal bounds of the NADW in the observed ocean and the isopycnals that bound the NADW transport in the CMIP5 models and BSOSE as evidenced by the juxtaposition of the region of enhanced subsurface salinity and the $27.40\sigma_\theta$ and $45.86\sigma_4$ isopycnals (Fig. 6a). This “mismatch” between the density classes in which NADW is found in the observed ocean and those where the southward flow associated with its transport is found in the models is shown more clearly in the meridional volume transports across 32°S in Fig. 6b. The net southward transport associated with the NADW occurs in density classes below $45.86\sigma_4$ in BSOSE and in the majority of the CMIP5 models, indicating that most models are

FIG. 5. (a) Full-depth-averaged salinity (psu) in the Southern Ocean from the *WOA13* climatological mean (large panel at top left; computed from the average of the 1985–94 and 1995–2004 decadal climatologies), and the difference between the simulated and observed salinity (psu) for the BSOSE Iteration 105 solution (time averaged from January 2008 to December 2012), and for each CMIP5 model (time averaged from January 1986 to December 2005). The 65° and 45°S latitudes from which the property differences are calculated are contoured in black. Positive (from yellow to red) values indicate the model is more saline than observed. Negative (blue) values indicate the model is fresher than observed. The HadGEM2-CC, MIROC-ESM-CHEM, IPSL-CM5A-LR, MRI-CGCM3, and NorESM1-M models are not shown; their salinity structure closely resembles the other models from their respective modeling centers (HadGEM2-ES, MIROC-ESM, IPSL-CM5A-MR, MRI-ESM1, NorESM1-ME).

FIG. 5b. As in (a), but for full-depth-averaged potential temperature ($^{\circ}\text{C}$) in the Southern Ocean. Positive (yellow to red) values indicate the model is warmer than observed. Negative (blue) values indicate the model is colder than observed. The HadGEM2-CC, MIROC-ESM-CHEM, IPSL-CM5A-LR, MRI-CGCM3, and NorESM1-M models are not shown; their temperature structure closely resembles the other models from their respective modeling centers (HadGEM2-ES, MIROC-ESM, IPSL-CM5A-MR, MRI-ESM1, NorESM1-ME).

FIG. 6. (a) Salinity (psu) in the South Atlantic at 32°S from the *WOA13* climatological mean (shown at top left; computed from the average of the 1985–94 and 1995–2004 decadal climatologies), time averaged from January 2008 to December 2012 for the BSOSE iteration 105 solution, and time averaged from January 1986 to December 2005 for each CMIP5 model. The thick black contours overlaid on each plot are the isopycnal layers that define the major flow regimes at this latitude as defined in Talley (2008) and labeled in the *WOA13* panel; subducted thermocline (ST; from the surface to $26.40\sigma_{\theta}$), lower thermocline (LT; $26.40\sigma_{\theta}$ – $26.90\sigma_{\theta}$), Antarctic Intermediate Water (AAIW; $26.90\sigma_{\theta}$ – $27.40\sigma_{\theta}$), North Atlantic Deep Water (NADW; $27.40\sigma_{\theta}$ – $45.86\sigma_{\theta}$), and Antarctic Bottom Water (AABW; $>45.86\sigma_{\theta}$). Salinity contours (thin white lines) are drawn at intervals of 0.2. The HadGEM2-CC, MIROC-ESM-CHEM, IPSL-CM5A-LR, MRI-CGCM3, and NorESM1-M models are not shown; their salinity structure closely resembles the other models from their respective modeling centers (HadGEM2-ES, MIROC-ESM, IPSL-CM5A-MR, MRI-ESM1, NorESM1-ME).

FIG. 6b. Volume transport (Sv) across 32°S in the Atlantic in density layers based on the definitions in Talley (2008) time averaged from January 2008 to December for the BSOSE iteration 105 solution and from January 1986 to December 2005 for each CMIP5 model. The empty boxes are the transport for each layer reported by Talley (2008). The blue bars are the modeled transport in each density layer, and the smaller red bars are equal subdivisions within each density layer. For the Talley transports, the reported Ekman transport has been added into the reported surface layer transport. The gray shading indicates the location of the southward flowing deep water ($27.40\sigma_\theta$ – $45.86\sigma_\theta$) according to Talley (2008). The gray hatching indicates where net southward flow between the northward flowing AAIW and AABW is found in each model, corresponding to the “model specific” NADW in Table 3. For each model, the transports were calculated across the grid cell line closest to 32°S; values were not interpolated. Positive values are northward transport (Sv). The HadGEM2-CC, MIROC-ESM-CHEM, IPSL-CM5A-LR, MRI-CGCM3, and NorESM1-M models are not shown; their transport structure closely resembles the other models from their respective modeling centers (HadGEM2-ES, MIROC-ESM, IPSL-CM5A-MR, MRI-ESM1, NorESM1-ME).

TABLE 3. Parameters related to the transport of NADW at 32°S. Both the transports within the density layers defined by Talley (2008) (net transport bounded by the $27.40\sigma_\theta$ and $45.86\sigma_4$ isopycnals) and within each model's "specific" NADW layer are listed. The "model specific" NADW transports (values in parentheses) are defined in the layers where net southward flow between the northward flowing AAIW and AABW is found in each model (see Fig. 6b). Temperature transport values for the NADW layer are calculated as temperature transport (PWT) as defined originally by Hall and Bryden (1982) and used in the transport estimates by Talley (2003) and Talley (2008) (see methods section). The AMOC value listed in the table is the maximum of the meridional overturning streamfunction in latitude–depth space in the Atlantic. The observational estimate reference for this parameter spans the range of values determined from hydrographic measurements and inverse estimates reported by Lumpkin and Speer (2007) and Talley (2003) and direct measurements from the Rapid Climate Change (RAPID) mooring array (Smeed et al. 2018). NP indicates that neither the meridional overturning streamfunction data nor ocean basin mask was provided in the ESGF archive. NA indicates not applicable; BSOSE only extends to $\sim 30^\circ\text{S}$. All CMIP5 values are time averaged from January 1986 to December 2005 of the historical simulation. BSOSE values are time averaged from January 2008 to December 2012 for the BSOSE iteration 105 solution.

Model	ACC (Sv)	AMOC (Sv)	Talley NADW (Sv)	Talley NADW salinity	Talley NADW salt export (kg s^{-1})	Talley NADW temperature export (PWT)
Observational estimate	173.3 ± 10.7	13–23	-17.6 ± 8.13	34.77	-613 ± 283	−0.20
BSOSE	181	NA	−7.3 (−11.1)	34.76 (34.84)	−252 (−386)	−0.10 (−0.14)
GISS-E2-R-CC	246	27.2	−18.3 (−18.3)	34.95 (34.95)	−641 (−641)	−0.41 (−0.41)
HadCM3	242	NP	−7.4 (−14.0)	34.79 (34.81)	−260 (−491)	−0.15 (−0.25)
GISS-E2-H-CC	205	NP	−14.9 (−17.8)	34.59 (34.55)	−516 (−614)	−0.18 (−0.22)
ACCESS1.3	184	20.2	−14.9 (−21.4)	34.85 (34.84)	−522 (−746)	−0.27 (−0.32)
HadGEM2-CC	178	20.0	−20.8 (−21.9)	34.91 (34.91)	−727 (−767)	−0.35 (−0.36)
MIROC-ESM-CHEM	175	15.0	−11.2 (−15.2)	34.91 (34.93)	−394 (−534)	−0.23 (−0.28)
MIROC-ESM	174	15.5	−11.2 (−14.8)	34.91 (34.93)	−392 (−518)	−0.24 (−0.28)
HadGEM2-ES	172	19.1	−20.3 (−20.9)	34.90 (34.90)	−709 (−732)	−0.35 (−0.36)
BCC_CSM1.1(m)	171	23.3	−21.5 (−22.8)	34.82 (34.83)	−750 (−797)	−0.45 (−0.46)
CCSM4	169	23.0	−7.2 (−18.3)	34.78 (34.95)	−254 (−640)	−0.13 (−0.26)
BCC_CSM1.1	157	17.2	−15.7 (−18.4)	34.78 (34.71)	−550 (−642)	−0.38 (−0.41)
GFDL CM3	156	26.2	−19.6 (−19.7)	35.00 (35.01)	−689 (−694)	−0.49 (−0.49)
CanESM2	153	17.8	−11.0 (−15.6)	34.97 (34.90)	−386 (−547)	−0.23 (−0.30)
FGOALS-g2	147	28.8	−15.3 (−26.9)	35.14 (35.24)	−540 (−949)	−0.29 (−0.50)
CESM1(CAM5)	146	23.7	−10.9 (−19.0)	34.88 (34.97)	−382 (−668)	−0.19 (−0.29)
ACCESS1.0	146	19.9	−16.2 (−19.6)	34.83 (34.81)	−567 (−684)	−0.29 (−0.32)
GFDL CM2.1	144	23.9	−22.7 (−21.3)	34.97 (34.95)	−796 (−747)	−0.42 (−0.42)
CNRM-CM5.2	142	15.7	−15.6 (−15.6)	34.89 (34.89)	−544 (−544)	−0.23 (−0.23)
FGOALS-s2	133	19.9	−3.8 (−21.0)	34.85 (35.11)	−136 (−741)	−0.09 (−0.39)
GFDL-ESM2M	131	23.7	−19.3 (−20.0)	34.96 (34.97)	−677 (−702)	−0.39 (−0.40)
NorESM1-M	127	31.3	−18.6 (−23.3)	35.08 (35.08)	−657 (−823)	−0.40 (−0.47)
NorESM1-ME	126	29.9	−18.7 (−23.1)	35.05 (35.07)	−660 (−815)	−0.37 (−0.44)
MRI-CGCM3	116	15.1	−12.5 (−13.8)	34.82 (34.83)	−436 (−481)	−0.21 (−0.22)
GFDL-ESM2G	112	19.8	−13.0 (−19.8)	34.95 (35.01)	−457 (−696)	−0.22 (−0.31)
MRI-ESM1	112	15.2	−12.1 (−13.3)	34.82 (34.82)	−422 (−465)	−0.21 (−0.22)
CSIRO Mk3.6.0	109	21.9	−9.9 (−19.8)	34.93 (35.01)	−348 (−695)	−0.21 (−0.34)
IPSL-CM5A-MR	108	NP	−12.0 (−14.9)	34.83 (34.89)	−420 (−520)	−0.18 (−0.21)
BNU-ESM	107	NP	−25.2 (−25.2)	34.80 (34.80)	−880 (−880)	−0.49 (−0.49)
IPSL-CM5A-LR	102	NP	−3.4 (−11.8)	34.69 (34.78)	−120 (−413)	−0.06 (−0.13)
IPSL-CM5B-LR	94	NP	−1.43 (−1.84)	34.63 (34.66)	−50 (−64)	−0.02 (−0.02)
CNRM-CM5	88	13.5	−9.5 (−11.2)	34.77 (34.74)	−332 (−388)	−0.17 (−0.20)

producing too dense NADW likely due to this water mass being much too saline (Fig. 6a, Fig. S1a). The upper-bounding isopycnal of the southward-flowing NADW layer corresponds generally to that observed in most models ($27.40\sigma_\theta$). The IPSL-CM5B-LR model stands out relative to the other models with regard to its lack of a visible subsurface salinity maximum and extremely low net southward transports associated with its NADW (Fig. 6).

The volume, salt, and temperature transports for the net flow within the Talley (2008) NADW bounds ($27.40\sigma_\theta$ – $45.86\sigma_4$) and for the net transport within the model "specific" NADW bounds are summarized in Table 3. The IPSL-CM5B-LR model is excluded from the following summary due to its exceptionally weak NADW (−1.84 Sv) and associated properties. The CMIP5 models simulate a large range of NADW transports from −11.2 Sv (CNRM-CM5) to −26.9 Sv (FGOALS-g2).

BSOSE simulates a NADW transport on the low end of the Talley (2008) estimate and one of the lowest transports relative to the CMIP5 models related to errors beyond the scope this study. All but six of the CMIP5 models (GISS-E2-R-CC, CCSM4, GFDL CM3, CESM-CAM5, NorESM1-M, NorESM1-ME) have NADW transports across 32°S that are >80% of their maximum Atlantic meridional overturning circulation (AMOC) value. Thus, the majority of the NADW formed in the high-latitude North Atlantic is exported into the SO contributing to the properties here.

The salt transport associated with NADW at 32°S shows considerable spread among models, with transports from -388 kg s^{-1} (CNRM-CM5) to -949 kg s^{-1} (FGOALS-g2). While Talley (2008) does not report a salt transport estimate, we use the volume transport and the average salinity reported for the NADW layer in Table 10 of Talley (2008) as a crude estimate of the “observed” salt transport and its associated range. BSOSE, while appearing to have a salinity structure consistent with that observed at 32°S (Fig. 6a), has a relatively weak NADW volume transport, resulting in a salt transport of only -386 kg s^{-1} , a value lower than that simulated by any of the CMIP5 models (with the exception of IPSL-CM5B-LR).

The temperature transports associated with the southward NADW flow across 32°S also show a wide spread among CMIP5 models, with a simulated range from -0.13 PWT (IPSL-CM5A-LR) to -0.50 PWT (FGOALS-g2). Despite being able to accurately simulate the volume transport of NADW, the overwhelming majority of the CMIP5 models simulate a temperature transport greater than the Talley (2008) NADW temperature transport of -0.20 PWT due to the warm biases in their deep waters, which are often associated with model drift. In addition, some models (e.g., CNRM-CM5), despite their exceptionally weak NADW volume transport, are able to achieve accurate temperature transports due to these excessive warm biases.

4. Discussion

a. Metrics

By considering various observationally based metrics related to the ACC transport in combination with one another, the CMIP5 models can be categorized based on common deficiencies in their representation of their SO mean state. This approach can aid in building a framework to understand why a model’s ACC transport does or does not diverge from reality. Using physical reasoning, we group models by assessing how well their simulation passes the following tests of what we consider to be the most important observationally based metrics:

- 1) Accurate push of the ACC at the ocean surface: Is the peak wind stress magnitude within 2σ of the ERA-Interim mean?
- 2) Accurate pull via surface divergence: Is the integrated WSC over the open Drake Passage latitudes within 2σ of the ERA-Interim mean?
- 3) Accurate full-depth zonally averaged density difference across the ACC: Is the full-depth-averaged and zonally averaged density difference across the ACC ($\Delta\rho$, 65°–45°S) within 25% of the difference computed from the WOA13 data?

1) REASONABLE ACC FOR APPROXIMATELY THE RIGHT REASONS

The eight models that fall into this category (Table 2) have an ACC strength within the observational uncertainty (2σ) and achieve accuracy¹ in all three metrics: 1) accurate push of the ACC at the ocean surface, 2) accurate pull via surface divergence, and 3) accurate $\Delta\rho$ across the ACC. We have chosen to keep ACCESS1.3 within this group despite it having an integrated WSC over the DP that is just above the 2σ uncertainty of the ERA-Interim mean given that it achieves accuracy in all of the other metrics in Table 2 and that its ACC is above (but still within 2σ of) the Donohue et al. (2016) mean ACC. Thus, it makes physical sense that with a slightly larger WSC over the DP, the ACC transport should be slightly larger than the mean transport observed. Another caveat we note regarding the models in this group is that the GFDL CM3 model has a significant climate drift present in its simulation (Griffies et al. 2011). This leads to enhanced warming both north and south of the ACC, likely linked to a known extreme warm bias in its abyssal waters (particularly evidenced by the large NADW heat transport in Table 3), which warms the upper ocean as these waters are pulled to the surface. The drift results in a slight positive trend in the ACC over the 1986–2005 period (not shown here), increasing by $\sim 5 \text{ Sv}$ from 1986 to 2005, explaining why the interannual standard deviation over this time period is larger for this model (Fig. 1).

We emphasize here that while all of the models in this group meet the requirement to have their $\Delta\rho$ across the ACC within 25% of the difference computed from the WOA13 data, this accuracy is commonly achieved by compensating errors between ΔT and ΔS across the current (Table 2, columns 4–6). While the compensation

¹ Here we define accuracy as within 2σ of the mean observational value.

provides for a density difference that supports an accurate ACC magnitude, such details in the structure of the errors may become important when considering other climatologically important processes such as heat and carbon fluxes given that these errors are often the symptom of inaccuracies in other processes including the representation of surface fluxes, water mass formation, and/or water mass transport into the SO. We provide examples of compensating errors among this group in [section 4b](#) of the discussion.

2) ACCURATE SIMULATION OF METRICS BUT WEAK ACC

The models that fall into this category ([Table 2](#)) achieve accuracy in all three metrics described above but simulate an ACC strength below the range of observational uncertainty (2σ). These models have ACC transports ranging from an extremely low value of 109 Sv (CSIRO Mk3.6.0) to 146 Sv (ACCESS1.0). There is a clear separation in this group between models that only fall outside of the observational uncertainty by a few Sverdrups ([Fig. 1](#)) and those that have a much too weak ACC.

Given the accuracy in the wind stress forcing and $\Delta\rho$ across the ACC, the models in this category should have an ACC within the observational uncertainty; however, the fact that three of the models (NorESM1-M, NorESM1-ME, CSIRO Mk3.6.0) produce exceptionally weak transports suggests that other model errors may be at fault. Sources of such errors may include choices in the magnitude and scheme for parameterizing unresolved mixing. It is well known that the ACC simulation in climate models is very sensitive to the parameterization of eddy-induced transports ([Danabasoglu and McWilliams 1995](#)). For CMIP3 generation models, a large number of models used an implementation of [Gent and McWilliams \(1990\)](#) with a fixed eddy-induced transport coefficient κ (m^2s^{-1}), allowing the impact of the choice of coefficients used in the simulation to be diagnosed ([Kuhlbrodt et al. 2012](#)).

Such an analysis on the CMIP5 generation of models is nontrivial given the more sophisticated eddy parameterization schemes used, with only a small number of models employing fixed κ values [CanESM2 ($\kappa = 1000\text{m}^2\text{s}^{-1}$), FGOALS-g2 ($\kappa = 500\text{m}^2\text{s}^{-1}$), FGOALS-s2 ($\kappa = 500\text{m}^2\text{s}^{-1}$), MIROC-ESM ($\kappa = 700\text{m}^2\text{s}^{-1}$), and MIROC-ESM-CHEM ($\kappa = 700\text{m}^2\text{s}^{-1}$)]. The majority of CMIP5 models employ eddy parameterization schemes with κ values that vary in both space (horizontally and vertically) and time, preventing an analysis of the impacts of the eddy-induced transport without detailed knowledge of the exact implementation of the parameterization scheme in each simulation. While we do not attempt to identify the impact of the

eddy-induced transports on the resulting ACC, we do note that there does not seem to be any discernible patterns between groups with respect to commonalities in the parameterization schemes and κ values used. For example, even among the eight models in group 1 ([Table 2](#)), three have fixed κ values (CanESM2, MIROC-ESM, MIROC-ESM-CHEM; see values listed above), and the other five have varying κ from GFDL CM3's κ range of $100\text{--}600\text{m}^2\text{s}^{-1}$ to CCSM4's much larger κ range of $300\text{--}3000\text{m}^2\text{s}^{-1}$.

In addition to the impacts of the choice and implementation of the eddy-parameterization schemes discussed above, the models may achieve accuracy in all three metrics yet have a weaker ACC due to errors stemming from interactions of the ACC with the bottom topography. Such interactions impact local density gradients and exert a drag force on the large-scale flow.

3) ACCURATE WIND STRESS FORCING WITH ERRORS IN $\Delta\rho$

This category contains the greatest number of models relative to the other groups ([Table 2](#)). These models achieve accuracy in their wind stress forcing, yet either have a too strong (3 models) or too weak (7 models) $\Delta\rho$ across the ACC. The HadCM3 model has an excessive $\Delta\rho$ across the ACC, attributed to a much too large ΔS (0.34 psu compared to 0.06 psu from *WOA13*). This model has much too saline waters throughout the whole water column south of the DP ([Fig. 5a](#)). The HadCM3 model only has 20 grid cells from the surface to the bottom of the ocean ([Table 1](#)), and thus this model may have issues with vertical diffusion as a result of such coarse vertical resolution. The other two models in this category also exhibit excessive ΔS with close to observed ΔT across the ACC. The combination of accuracy in wind stress forcing with a too strong $\Delta\rho$ should lead to an excessive ACC strength relative to the [Donohue et al. \(2016\)](#) estimate. This is true for HadCM3 and GISS-E2-H-CC; however, this is not the case for CESM1(CAM5), which actually has an ACC just below the observational uncertainty (2σ) ([Table 2, Fig. 1](#)).

The models that have a too weak $\Delta\rho$ all exhibit weaker than observed ΔS combined with a too weak ΔT as well (except for GFDL-ESM2M and GFDL-ESM2G). The GFDL-ESM2M and GFDL-ESM2G models are too fresh within and south of the ACC, yielding a ΔS in the opposite direction than observed. However, their temperature structure across the ACC yields a slightly stronger than observed ΔT in the case of GFDL-ESM2G (too cold south of the DP) and close to observed in the case of GFDL-ESM2M (too warm everywhere, but ΔT remains close to observed). Most of the other models in this group exhibit both too fresh and too warm waters

within and south of the ACC. The BNU-ESM model in particular, has large warm biases within and south of the ACC coupled with too fresh waters in these regions as well, resulting in the weakest $\Delta\rho$ across the ACC out of all the models considered (Fig. 5). The combination of accurate wind stress forcing with a weak $\Delta\rho$ should yield a weaker ACC than that observed. This is the case for all of the models in this group, yielding ACC transports from 107 Sv (BNU-ESM) to 131 Sv (GFDL-ESM2M).

4) ERRORS IN WIND STRESS FORCING BUT ACCURATE $\Delta\rho$

The models in this category have an accurate $\Delta\rho$ across the ACC but exhibit errors in their simulation of the key wind stress metrics: accurate push of the ACC at the ocean surface or accurate pull via surface divergence. Two models in this category [BCC_CSM1.1 and BCC_CSM1.1(m)] have excessive wind stress forcing with regard to one or both of the wind stress metrics. The BCC_CSM1.1(m) model has a too strong zonally averaged westerly wind stress maximum and too strong integrated WSC over the DP latitudes. The BCC_CSM1.1 model only has a too strong integrated WSC over the DP. In neither of these models does the excessive wind stress forcing lead to an excessive ACC transport, with both model's ACC transport falling well within the observational uncertainty (2σ).

Three of the models in this category have both a too weak zonally averaged westerly wind stress maximum and too weak integrated WSC over the DP latitudes. The CNRM-CM5 model only has a too weak zonally averaged westerly wind stress maximum, while its integrated WSC over the DP is accurately represented. As expected, weaker forcing by the overlying wind stress leads to a weaker than observed ACC transport. The FGOALS-g2 model has an ACC with a mean that just falls outside the lower bound of the observational uncertainty (2σ).

5) ERRORS IN WIND STRESS FORCING AND ERRORS IN $\Delta\rho$

The only model that falls into this category is the GISS-E2-R-CC model. This model, despite its relatively weak wind stress forcing in both metrics (at the low end of the 2σ uncertainty for the peak wind stress magnitude, and below the 2σ uncertainty for the integrated WSC over the DP), simulates an excessively strong $\Delta\rho$ and exceptionally large ACC transport. This model also has an equatorward bias in its zonally averaged westerly wind stress maximum. The $\Delta\rho$ in this model is clearly driven by an excessive ΔT across the ACC (5.2° compared to 2.5°C

from WOA13). This model has exceptionally warm temperatures throughout the water column within and north of the ACC (Fig. 5b) and colder than observed temperatures south of the ACC. The ΔS across the ACC is relatively strong but in the opposite direction, with too fresh water south of the DP and slightly too salty water on the ACC's northern edge. Given the relatively weak wind stress forcing, the fact that this model gets such a strong thermally driven $\Delta\rho$ across the ACC indicates there must be errors associated with the buoyancy forcing of the current. Furthermore, documented errors associated with the parameterized mixing implemented in the GISS-E2-R-CC model (Schmidt et al. 2014) likely contribute to the exceptionally strong ACC.

b. Sources of errors in the meridional density difference across the ACC

In most cases, models achieve accuracy in $\Delta\rho$ across the ACC as a result of compensating errors between biases in ΔT and ΔS (Table 2, groups 1, 2, and 4). For example, the CanESM2 model, which falls into group 1, has a ΔS that is more than twice the magnitude of that from the WOA13 product. However, a slightly weaker ΔT across the ACC relative to WOA13 partially compensates for the salinity, producing a $\Delta\rho$ on the upper end, but within 25% error of the WOA13 value. The ACCESS1.3 and GFDL CM3 models provide examples of the opposite scenario, with strong ΔT across the ACC compensated by slightly weaker ΔS . In the case of GFDL CM3, the ΔS is actually opposite in sign, with the full-depth-averaged and zonally averaged salinity greater on the northern edge of the ACC.

If a model achieves accuracy in the wind stress forcing at the ocean surface yet has an inaccurate $\Delta\rho$ (group 3), or if a model achieves an accurate $\Delta\rho$ in the absence of accurate wind stress forcing (group 4), this is suggestive of inaccuracies in the horizontal and/or vertical buoyancy fluxes within the SO. The temperature and salinity structure of the SO in the region of the ACC is influenced by a number of factors that differ widely from model to model, including but not limited to representation of the annual mean and seasonal Antarctic sea ice extent (Shu et al. 2015; Ivanova et al. 2016; Table S1), local surface heat and freshwater fluxes (Figs. 4a,b), the injection of warm and saline water from the subtropical gyres adjacent to the ACC, and the properties and location in the water column of NADW as it is exported into the SO (Table 3, Fig. 6).

In the observed ocean, the NADW that is exported southward from the Atlantic into the region of the ACC is both warm and salty relative to the rest of the deep and abyssal ocean (Fig. S1). Thus, the amount of NADW entering the SO, its properties, and its vertical location in the water column exert a strong influence on the

density structure across the ACC. NADW that is dense enough to traverse the SO below the effective sill depth of the DP latitudinal band (where a net meridional geostrophic flow can be maintained) and upwell directly to the upper ocean south of the DP provides a significant source of saline waters on the ACC's southern edge (Talley 2013; Warren 1990). A strong ΔS is obtained in the upper ocean across the ACC with NADW/CDW flanking the current's southern boundary and fresh intermediate and mode waters dominating the boundary of the ACC to the north (Fig. S1a).

Models that simulate exceptionally weak or too fresh NADW transport into the SO cannot maintain a strong density gradient across the ACC even if this water were dense enough to cross the ACC below the sill depth and upwell south of the DP. Six of the CMIP5 models exhibit very weak NADW transports (<14 Sv) across 32°S (Table 3, Fig. 6b). Consistent with their weak NADW salt transport, the MRI-CGCM3, MRI-ESM1, and IPSL models are all associated with too fresh waters south of the ACC (Fig. 5a, Fig. S1a). Excluding the HadCM3 model, whose errors will be addressed later, the models with weak NADW transports are associated with exceptionally weak ACC transports, ranging from 88 to 116 Sv. While the majority of the CMIP5 models have too salty NADW, several of the models are too fresh. Similarly, it is common for a model's NADW to be much too warm (Fig. S1b), resulting in exceptionally large temperature transports when coupled with a strong southward flow (Table 3).

Exactly how these biases in the NADW's temperature and salinity impact the density structure, and thus the geostrophic flow, depends on the location of the NADW in the water column and where (and if) it upwells to the upper ocean. Several models exhibit strong NADW transports but with no or very little net southward flow in the densest class of NADW ($45.80\sigma_4$ – $45.86\sigma_4$; Fig. 6b); thus, their NADW is biased too light relative to observations. Unsurprisingly, these models all show large NADW temperature biases (Fig. S1b), an error that has an increasingly important impact on density with increasing depth. Too light NADW can impact the density gradient in two ways; 1) not providing enough (or any) NADW dense enough to traverse the sill-depth as mentioned above, and 2) sitting too high in the water column such that it mixes with and degrades the cold, fresh signature of the intermediate and mode waters on the ACC's northern edge.

The representation of Antarctic sea ice extent is another important process that influences the density structure across the ACC by impacting the surface heat and freshwater fluxes across the SO. The formation of sea ice near the Antarctic coast increases the salinity of

the water column south of the DP via brine rejection. The drift and subsequent melt of the sea ice in the surface Ekman layer provides a source of cold, fresh water that gets entrained in the intermediate and mode layers, influencing the salinity structure within and on the ACC's northern edge. In fact, the salinity and temperature averaged over the upper 1500 m of the ocean (where the majority of the ACC flow is concentrated) at 45°S is significantly correlated with the annual sea ice extent across models ($r = -0.41$, $p = 0.02$ and $r = -0.42$, $p = 0.02$; scatterplots not shown here). Generally, the larger the annual sea ice extent, the colder and fresher the intermediate and mode waters on northern sector of the ACC.

Only 4 out of the 31 models in this study accurately represent the annual sea ice extent, with the overwhelming majority having too little sea ice extent (Table S1). Too little sea ice extent is a common feature across CMIP5 models (Shu et al. 2015; Ivanova et al. 2016), likely attributable to too warm surface layers. Recent work suggests that the thermal structure of the ocean, which strongly influences the sea ice representation in this region, may be caused by atmospheric model net surface flux biases linked to model cloud errors (Hyder et al. 2018). The GISS-E2-R-CC and GFDL models have among the smallest annual sea ice extents, with maximum sea ice extents limited to the regions south of the DP (Shu et al. 2015). All of these models yield a too fresh water column on the ACC's southern edge (Fig. S1a).

The BNU-ESM model has one of the largest Antarctic sea ice extent biases out of all the CMIP5 models (Ji et al. 2014; Shu et al. 2015; Table S1). This model exhibits a thick fresh layer in its upper ocean and some of the warmest subsurface temperatures on the ACC's southern edge (Fig. S1a, Fig. 5b), resulting in an exceptionally weak density difference and ACC flow (107 Sv; Table 2). The HadCM3 model has a very large density difference across the ACC attributed to its excessive difference in salinity across the current (Table 2). The pattern of the bias in subsurface salinity and the fact that this model has weak NADW suggest that the enhanced salinity south of the ACC may actually be related to excessive brine rejection via sea ice formation near the Antarctic coast.

As highlighted by Russell et al. (2006a), the horizontal flux of warm and saline subtropical waters from the southward flowing western boundary currents over- or underinjecting into the upper ocean can significantly influence the buoyancy forcing across the ACC. This is driven by the combination of a weak or strong bias in the zonally averaged westerly wind stress maximum and a significant equatorward shift in its

latitudinal position that impacts the strength of the subtropical gyre western boundary currents. Furthermore, an equatorward biased position of the zonally averaged westerly wind stress maximum shifts the position of the zero WSC boundary northward, placing regions of Ekman pumping over warmer and saltier subtropical waters. Thus, more salt and heat are incorporated into the northward flowing intermediate and mode water layers on the ACC's northern edge. The pattern of the zonally integrated WSC over the SO suggests this may be an issue for several of the CMIP5 models (Fig. 2).

5. Summary and conclusions

The quality of the simulation of the mean ACC transport in climate models is an emergent feature of a large number of other physical processes and properties. An ACC transport that is “right for the right reasons” is hard to achieve due to the accuracy required in the simulated atmosphere, ocean, and sea ice components along with coupling between the domains. Owing to the complex nature of the SO circulation coupled with the lack of long-term, high-resolution observations over the historical period, this region has been an area that climate models have struggled to accurately represent (Meijers et al. 2012; Bracegirdle et al. 2013; Downes and Hogg 2013; Sallée et al. 2013a; Sallée et al. 2013b; Frölicher et al. 2015; Russell et al. 2018). Given the dominant role that the SO plays in the global carbon and heat budget and in the anthropogenic heat and carbon uptake relative to the other ocean basins (Frölicher et al. 2015; Roemmich et al. 2015), inaccuracies in the representation of SO processes and properties in the mean-state are likely to propagate into significant errors associated with a model's response to increased radiative forcing associated with global warming.

Relative to CMIP3, there have been significant model improvements in the simulation of the ACC and parameters associated with its flow. Considering the improved DP transport estimate of 173.3 ± 10.7 Sv (Donohue et al. 2016), which was not available at the time of the assessment of CMIP3 models, only 2 of the 18 CMIP3 models considered by Russell et al. (2006a) simulate an ACC transport that falls solidly within observational uncertainty (2σ). Similarly, the analysis of CMIP3 models by Sen Gupta et al. (2009) also yields only two models that would fall within today's observational uncertainty (2σ). In the analysis presented here, 10 out of the 31 models simulate a transport within observational uncertainty (2σ). Among the CMIP5 models, there are only two models that exhibit extreme biases ($>7\sigma$ outside of observational uncertainty) in

their ACC strength as opposed to the large fraction of CMIP3 models that exhibited extreme biases (Fig. 1).

Two other important metrics considered by Russell et al. (2006a), the latitude of the maximum zonal-mean westerly wind stress over the SO and its magnitude, show clear improvements in CMIP5. Of the models analyzed by Russell et al. (2006a), 14 of the 18 models have their maximum wind stress located equatorward of 50°S compared to only 11 out of the 31 CMIP5 models considered here. Furthermore, using the ERA-Interim analysis as our observational benchmark, a greater fraction of the CMIP5 models achieve a reasonable (within 2σ of the ERA-Interim mean) magnitude of the maximum zonal wind stress compared to that found in the analysis of CMIP3 models. Additional metrics related to the surface wind stress forcing considered in this study are within the uncertainty (2σ) of the ERA-Interim mean for the majority of the CMIP5 models. The CMIP5 models seem to be achieving accuracy in two key metrics that impact ACC strength: the “push” of the current at the ocean surface and the “pull” of dense isopycnals to the upper ocean via surface divergence over the open DP latitudes.

Eight of the CMIP5 models achieve a “reasonable ACC for approximately the right reasons,” meaning that they simulate an accurate ACC transport, accurate surface wind stress forcing, and a reasonable full-depth-averaged, zonally averaged density difference across the current. The models that fall into this category can be found at the top of Table 2. In the analysis by Russell et al. (2006a), only two models fell into the category of “about right for the present-day climate scenario,” one of which was the GFDL CM2.1 model, which also participated in CMIP5. The GFDL CM2.1 model in our CMIP5 analysis falls into the “accurate simulation of key metrics but weak ACC” category. The only reason it fell into the category it did in the CMIP3 analysis is that the benchmark for transport through the DP at the time of assessment was 134 Sv (Cunningham et al. 2003), compared to the updated value of 173 Sv determined from the high spatial and temporal resolution observations from the cDrake array (Donohue et al. 2016), which were not known at the time. As interpreted in detail in the methods section (section 2d), while the baroclinic transport estimates of the ACC from the early ISOS measurements in the late 1970s and 1980s (Whitworth and Peterson 1985), WOCE SR1b (1993–2000; Cunningham et al. 2003), and cDrake (2007–11; Chidichimo et al. 2014) agree remarkably well, the increased total transport is ascribed to the cDrake array's ability to resolve the depth-independent or barotropic component. These starkly different transport estimates provide two very different benchmarks for ACC transport whose uncertainties

barely overlap, highlighting the role that observational uncertainty plays when assessing the performance of climate models.

The full-depth-averaged, zonally averaged difference in density generally sets the ACC's strength, and many models are able to achieve reasonable density differences. However, the detailed look at the structure of the errors in the various parameters considered in our analysis reveals that even among models that achieve a reasonable density difference and thus a reasonable ACC, this is often from compensating errors in the temperature and salinity structure. Furthermore, many models have quite large errors in their density structure across the ACC even when their surface wind stress forcing is accurately represented or achieve an accurate density difference in the absence of accurate wind stress forcing. We have highlighted many examples among the CMIP5 models of various processes and biases that directly impact the density structure including representation of surface freshwater and heat fluxes, transport and properties of NADW exported into the SO from the Atlantic, inaccurate representations of Antarctic sea ice extent, and equatorward biased winds that impact the properties of the intermediate and mode waters that flank the ACC's northern edge.

By analyzing such a large suite of CMIP5 models and a large number of observationally based metrics known to impact the ACC, this study provides a useful resource to the community regarding the historical simulation of the SO. More importantly, by quantifying biases across a large number of parameters in combination with one another, we have provided a framework that aids in identifying and understanding the major contributors to a model's ability to accurately represent the SO circulation. Using physical reasoning, the CMIP5 models have been categorized according to their ability to accurately represent key metrics associated with the ACC. The various categories of the mean state SO presented here likely have large implications for how these models respond under increased radiative forcing. The buoyancy errors that have been highlighted here, while some compensate and allow for a reasonable ACC, likely have major implications for ocean mixing and thus the fluxes of heat and carbon within the SO. Quantifying and understanding these impacts is the subject of an ongoing study. The comprehensive assessment of SO metrics presented here provides a solid baseline against which the next generation of CMIP6 simulations can be evaluated for model improvements.

Acknowledgments. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling responsible for CMIP. We thank the

modeling groups listed in Table 1 for producing and making available their output. For CMIP, the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. The BSOSE data used in this analysis are available at Scripps Institution of Oceanography (<http://sose.ucsd.edu/>). The CFSR and ERA-Interim reanalysis data were downloaded from the NASA CREATE system (Potter et al. 2018; <https://esgf.nccs.nasa.gov/search/create-ip/>). The *World Ocean Atlas 13* (WOA13) data were downloaded from <https://www.nodc.noaa.gov/OC5/woa13/>. We acknowledge the use of the Ferret program from NOAA's Pacific Marine Environmental Laboratory for analysis and graphics (<http://ferret.pmel.noaa.gov/Ferret/>). We thank the editor and three anonymous reviewers for their comments and suggestions on earlier versions of this manuscript. This research was funded by the U.S. EPA Assistance Agreement FP-91780701-0. This publication has not been reviewed by the EPA and the views expressed herein are solely those of the authors. This work was funded by NSF's Southern Ocean Carbon and Climate Observations and Modeling (SOCCOM) Project under the NSF Award PLR-1425989, with additional support from NOAA and NASA. Logistical support for SOCCOM in the Antarctic was provided by the U.S. NSF through the U.S. Antarctic Program. We also thank the Thomas R. Brown foundation at the University of Arizona for supporting Joellen L. Russell and Paul J. Goodman.

REFERENCES

- 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>.
- Bishop, S. P., P. R. Gent, F. O. Bryan, A. F. Thompson, M. C. Long, and R. Abernathy, 2016: Southern Ocean overturning compensation in an eddy-resolving climate simulation. *J. Phys. Oceanogr.*, **46**, 1575–1592, <https://doi.org/10.1175/JPO-D-15-0177.1>.
- Böning, C. W., A. Disper, M. Visbeck, S. R. Rintoul, and F. U. Schwazkopf, 2008: The response of the Antarctic Circumpolar Current to recent climate change. *Nat. Geosci.*, **1**, 864–869, <https://doi.org/10.1038/ngeo362>.
- 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, T. Phillips, and L. Wilcox, 2013: Assessment of surface winds over the Atlantic, Indian and Pacific Ocean sectors of the Southern Hemisphere in CMIP5 models:

- Historical bias, forcing response, and state dependency. *J. Geophys. Res. Atmos.*, **118**, 547–562, <https://doi.org/10.1002/jgrd.50153>.
- 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>.
- Colin de Verdière, A., and M. Ollitrault, 2016: A direct determination of the World Ocean barotropic circulation. *J. Phys. Oceanogr.*, **46**, 255–273, <https://doi.org/10.1175/JPO-D-15-0046.1>.
- 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., and J. C. McWilliams, 1995: Sensitivity of the global ocean circulation to parameterizations of mesoscale tracer transports. *J. Climate*, **8**, 2967–2987, [https://doi.org/10.1175/1520-0442\(1995\)008<2967:SOTGOC>2.0.CO;2](https://doi.org/10.1175/1520-0442(1995)008<2967:SOTGOC>2.0.CO;2).
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, <https://doi.org/10.1002/qj.828>.
- 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>.
- Downes, S., and A. Hogg, 2013: Southern Ocean circulation and eddy compensation in CMIP5 models. *J. Climate*, **26**, 7198–7220, <https://doi.org/10.1175/JCLI-D-12-00504.1>.
- Firing, Y., T. K. Chereskin, and M. R. Mazloff, 2011: Vertical structure and transport of the Antarctic Circumpolar Current in Drake Passage from direct velocity measurements. *J. Geophys. Res.*, **116**, C08015, <https://doi.org/10.1029/2011JC006999>.
- 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., and J. C. McWilliams, 1990: Isopycnal mixing in ocean circulation models. *J. Phys. Oceanogr.*, **20**, 150–155, [https://doi.org/10.1175/1520-0485\(1990\)020<0150:IMOCM>2.0.CO;2](https://doi.org/10.1175/1520-0485(1990)020<0150:IMOCM>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>.
- Hall, M. M., and H. L. Bryden, 1982: Direct estimates and mechanisms of ocean heat transport. *Deep-Sea Res.*, **29**, 339–359, [https://doi.org/10.1016/0198-0149\(82\)90099-1](https://doi.org/10.1016/0198-0149(82)90099-1).
- Hallberg, R., and A. Gnanadesikan, 2006: The role of eddies in determining the structure and response of the wind-driven Southern Hemisphere overturning: results from the Modeling Eddies in the Southern Ocean (MESO) project. *J. Phys. Oceanogr.*, **36**, 2232–2252, <https://doi.org/10.1175/JPO2980.1>.
- Hogg, A. M., M. P. Meredith, D. P. Chambers, E. P. Abrahamson, C. W. Hughes, and A. K. Morrison, 2015: Recent trends in the Southern Ocean eddy field. *J. Geophys. Res. Oceans*, **120**, 257–267, <https://doi.org/10.1002/2014JC010470>.
- 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>.
- Ji, D., and Coauthors, 2014: Description and basic evaluation of Beijing Normal University Earth System Model (BNU-ESM) version 1. *Geosci. Model Dev.*, **7**, 2039–2064, <https://doi.org/10.5194/gmd-7-2039-2014>.
- Knutti, R., R. Furrer, C. Tebaldi, J. Cermak, and G. A. Meehl, 2010: Challenges in combining projections from multiple climate models. *J. Climate*, **23**, 2739–2758, <https://doi.org/10.1175/2009JCLI3361.1>.
- 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>.
- Locarnini, R. A., and Coauthors, 2013: *Temperature*. Vol. 1, *World Ocean Atlas 2013*, S. Levitus, Ed., NOAA Atlas NESDIS 73, 40 pp.
- Lumpkin, R., and K. Speer, 2007: Global ocean meridional overturning. *J. Phys. Oceanogr.*, **37**, 2550–2562, <https://doi.org/10.1175/JPO3130.1>.
- 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**, C12008, <https://doi.org/10.1029/2012JC008412>.
- Meredith, M. P., A. C. Naveira Garabato, A. M. Hogg, and R. Farneti, 2012: Sensitivity of the overturning circulation in the Southern Ocean to decadal changes in wind forcing. *J. Climate*, **25**, 99–110, <https://doi.org/10.1175/2011JCLI4204.1>.
- 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>.
- Potter, G. L., L. Carriere, J. Hertz, M. Bosilovich, D. Duffy, T. Lee, and D. N. Williams, 2018: Enabling reanalysis research using the Collaborative Reanalysis Technical Environment (CREATE). *Bull. Amer. Meteor. Soc.*, **99**, 677–687, <https://doi.org/10.1175/BAMS-D-17-0174.1>.
- Rienecker, M. M., and Coauthors, 2011: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Climate*, **24**, 3624–3648, <https://doi.org/10.1175/JCLI-D-11-00015.1>.
- Roemmich, D., J. Church, J. Gilson, D. Monselesan, P. Sutton, and S. Wijffels, 2015: Unabated planetary warming and its 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>.
- Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1058, <https://doi.org/10.1175/2010BAMS3001.1>.
- Sallée, J.-B., E. Shuckburgh, N. Bruneau, A. J. S. Meijers, T. J. Bracegirdle, and Z. Wang, 2013a: Assessment of the Southern Ocean mixed-layer depth in CMIP5 models: Historical bias and forcing response. *J. Geophys. Res.*, **118**, 1845–1862, <https://doi.org/10.1002/jgrc.20157>.
- , —, —, —, —, —, 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>.
- Schmidt, G. A., and Coauthors, 2014: Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive. *J. Adv. Model. Earth Syst.*, **6**, 141–184, <https://doi.org/10.1002/2013MS000265>.
- Sen Gupta, A., A. Santos, 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>.
- Smeed, D. A., and Coauthors, 2018: The North Atlantic is in a state of reduced overturning. *Geophys. Res. Lett.*, **45**, 1527–1533, <https://doi.org/10.1002/2017GL076350>.
- 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., 2003: Shallow, intermediate, and deep overturning components of the global heat budget. *J. Phys. Oceanogr.*, **33**, 530–560, [https://doi.org/10.1175/1520-0485\(2003\)033<0530:SIADOC>2.0.CO;2](https://doi.org/10.1175/1520-0485(2003)033<0530:SIADOC>2.0.CO;2).
- , 2008: Freshwater transport estimates and the global overturning circulation: Shallow, deep and throughflow components. *Prog. Oceanogr.*, **78**, 257–303, <https://doi.org/10.1016/j.pocean.2008.05.001>.
- , 2013: Closure of the global overturning circulation through the Indian, Pacific, and Southern Oceans: Schematics and transports. *Oceanography*, **26**, 80–97, <https://doi.org/10.5670/oceanog.2013.07>.
- Tamsitt, V., L. D. Talley, M. R. Mazloff, and I. Cerovečki, 2016: Zonal variations in the Southern Ocean heat budget. *J. Climate*, **29**, 6563–6579, <https://doi.org/10.1175/JCLI-D-15-0630.1>.
- 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>.
- 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>.
- Warren, B. A., 1990: Suppression of deep oxygen concentrations by Drake Passage. *Deep Sea Res.*, **37A**, 1899–1907, [https://doi.org/10.1016/0198-0149\(90\)90085-A](https://doi.org/10.1016/0198-0149(90)90085-A).
- 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](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](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](https://doi.org/10.1175/1520-0485(1982)012<0960:TNTOTA>2.0.CO;2).
- Zweng, M. M., and Coauthors, 2013: *Salinity*. Vol. 2, *World Ocean Atlas 2013*, S. Levitus, Ed., NOAA Atlas NESDIS 74, 39 pp.