Check for updates

ADVANCING EARTH AND SPACE SCIENCE

JGR Oceans

RESEARCH ARTICLE
10.1029/2022JC018880

Key Points:

- CMIP models ranked for their representation of the Ross Sea Region marine environment, and a "best" ensemble has been defined
- End of century projections show a warmer, fresher Ross Sea, and reduced sea ice concentration and mixed layer depths compared to present day
- Surface nutrients reduced by end of century, smaller changes to surface chlorophyll and Net Primary production, with increased Ross Gyre transport

Correspondence to:
G. J. Rickard,
graham.rickard@niwa.co.nz

Citation:
Rickard, G. J., Behrens, E., Bahamondes Dominguez, A. A., & Pinkerton, M. H. (2023). An assessment of the oceanic physical and biogeochemical components of CMIP5 and CMIP6 models for the Ross Sea Region. *Journal of Geophysical Research: Oceans*, *128*, e2022JC018880. https://doi.org/10.1029/2022JC018880

Received 19 MAY 2022
Accepted 28 FEB 2023

Author Contributions:
Conceptualization: Graham J. Rickard
Formal analysis: Graham J. Rickard
Funding acquisition: Matt H. Pinkerton
Investigation: Erik Behrens
Methodology: Graham J. Rickard, Erik Behrens, Angela A. Bahamondes Dominguez, Matt H. Pinkerton
Software: Graham J. Rickard
Writing – original draft: Graham J. Rickard
Writing – review & editing: Graham J. Rickard, Erik Behrens, Angela A. Bahamondes Dominguez, Matt H. Pinkerton

© 2023. The Authors.
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

An Assessment of the Oceanic Physical and Biogeochemical Components of CMIP5 and CMIP6 Models for the Ross Sea Region

Graham J. Rickard[1], Erik Behrens[1], Angela A. Bahamondes Dominguez[1], and Matt H. Pinkerton[1]

[1]National Institute of Water and Atmospheric Research, Wellington, New Zealand

Abstract The physical and biogeochemical performance of 16 CMIP5 and 16 CMIP6 Earth System models (ESM) are examined relative to present day (1976–2005) observational data sets for a Ross Sea Region (RSR) containing the Ross Gyre (RG) and the Ross Sea Continental Shelf. A relative ranking scheme using error metrics and published ESM properties (including climate sensitivity parameters and anomalous deep ocean convection statistics) enables identification of a "best" ensemble of models for the RSR. Over the RSR the CMIP6 models are generally found to have improved physical representations compared to the CMIP5 set (based on our metrics for sea ice concentration, sea surface temperature, salinity, and height, and mixed layer depth), but the CMIP5 and CMIP6 biogeochemical representations remain similar. Examination of mean properties for the period 2081 to 2100 for RCP8.5 and SSP585 for CMIP5 and CMIP6, respectively, shows significant surface temperature increases, with significant decreases in sea surface salinity, sea ice concentration, and mixed layer depth across the RSR. Biogeochemically, there are generally small increases in surface values for chlorophyll, integrated primary production, and zooplankton carbon concentrations. The projections also have robust reductions in surface nitrate, phosphate, and silicate across the RSR. For that part of the RG circulation to the east of 180°E—which we refer to as the "inner RG"—significant barotropic transport increases are found by the end of century.

Plain Language Summary Global climate simulation models have been tested to see how well they represent observations of the ocean in a region around Antarctica known as the Ross Sea. Here we are interested in both the physical (temperature, flows etc) and biogeochemical (chlorophyll, nutrients etc) properties of the models compared to the observations. This process allows us to identify the "best" models, or at least to rank them all from "best" to "worst." This then provides us with more confidence about their future predictions, the argument being that models that do well for our present ocean will do better into the future. By the end of the century we find that the predictions show the surface ocean is warmer and fresher than today, with much reduced sea ice, and surface nutrients are depleted but the surface chlorophyll not so. The large scale circulation is also increased. This shows that the Ross Sea is likely to experience significant physical and biological changes over the coming century relative to today's world.

1. Introduction

In a recent review Rogers et al. (2020) assess Southern Ocean (SO) ecosystem responses to changes in the physical and biogeochemical marine climate using present day observations, and for future states using RCP8.5 projections to 2100 from three CMIP5 Earth System models (ESM), building on earlier reviews such as Constable et al. (2014). Both these reviews highlight future ecosystem risks for the SO, and indeed Rogers et al. (2020) conclude: "The Antarctic therefore represents a challenging management scenario where rising pressure to exploit potentially increased availability of resources is combined with climate impacts that in turn may accelerate loss of the endemic biodiversity of the SO." Tables 1 and 2 in Rogers et al. (2020) summarize likelihoods, potential severity, and expert level of agreement of changing parameters on SO biological communities, where the "parameters" include both physical (e.g., ocean warming, sea ice decline etc), and biogeochemical (e.g., nutrient increase, primary production etc). In developing management strategies for the SO, the need for best advice in order to reduce uncertainties—in particular the least agreed upon red and yellow highlighted sections in Table 2 in Rogers et al. (2020)—remains critical. Here the aim is to narrow those uncertainties by assessment

of the physical and biogeochemical components of both CMIP5 and CMIP6 ESM for the Ross Sea Region (RSR) of the SO. For this study RSR spans 140°E to 99.25°W and 76°S to 58.25°S.

Our focus on the RSR reflects New Zealand's research interests in the Ross Sea Dependency, which includes the Balleny Islands, Scott Island (and other islands), and the Ross Ice Shelf. This region's importance was further highlighted by an international agreement to establish a Ross Sea MPA in December 2017 (CCAMLR (2016)) to improve protection of the unique Ross Sea region environment, with a research and monitoring plan developed collaboratively by members (Dunn et al. (2017)). The spatial domain of the RSR MPA extends from the continental shelf up to 60°S, making it the largest MPA in the world covering around 1.55 million km^2, and the research aims underpinning the agreement to evaluate its potential effectiveness require understanding of the physical and biogeochemical environment of this region and their impact on the lower and higher trophic levels of the ecosystem. In order to provide advice on future states of the RSR and the MPA for planning and policy, the analysis provided in this paper necessarily extends the earlier CMIP5 ESM study of Rickard and Behrens (2016) to cover not only the Ross Sea continental shelf, but the MPA region and also the wider gyral system of the Ross Sea. In Rickard and Behrens (2016) rankings for the 16 CMIP5 models used were produced, and from them best estimates of changes by end of century using RCP4.5 and RCP8.5 projections. Since then CMIP6 models have been made available, allowing for this updated assessment using a combined suite of CMIP5 and CMIP6 models.

The assessment process reported here mirrors Rickard et al. (2016) and Rickard and Behrens (2016) and more recently applied by Laurent et al. (2021), except here we include extra criteria associated with SO anomalous deep convection impacting Ross Sea region processes (e.g., see Heuzé (2021) and references therein), as well as the Equilibrium Climate Sensitivity (ECS) and Transient Climate Response (TCR) metrics measuring ESM climate sensitivity. These extra criteria do not obviously lend themselves to objective rankings, but we nevertheless factor them into our choices of model ensembles following, for example, Shiogama et al. (2021) who "exclude 5 [global climate models] GCMs with large warming tends in recent decades." In calculating future changes relative to our present day baseline of 1976–2005 we present results using four ensembles, our innermost ("best"), the inner ("intermediate"), our outer models ("worst"), and the full suite of models, so although the ranking process does identify an innermost set we provide projections using all the models. Part of the reason for retaining a "best" ensemble will be to identify what might be the most representative future states to use for downscaling, as present resources necessarily limit the processes that CMIP models can resolve, especially so for Antarctica and the role played by ice shelves, for example, Our process complements the advice of Brunner et al. (2020) that "Combining a range of relevant diagnostics is less prone to overconfidence, as the risk of upweighting a model because it "accidently" fits observations for one diagnostic while being far away from them in several others is greatly reduced."

The paper is structured as follows: In Section 2 the methods used to obtain our model rankings are detailed, in particular the spatial domains covering the RSR, and the framework for assessment based on comparison with climatologically relevant data sets. Section 3 presents the results of our assessment, in particular the models comprising each of our ensembles. Section 4 takes the results of the model ranking, and uses them to produce annual mean future changes by end of century based on RCP8.5/SSP585 projections, as well as seasonal cycles using our "best" ensemble for 20 year periods between 1850 and 2100 and comparing them to the observational data. Discussions and conclusions are then provided in Section 5.

2. Methods

Figure 1a shows our domain of interest, spanning southern limits of the Antarctic Circumpolar Current (ACC) to the north, the major gyres—plotted using the sea surface height (SSH) or ZOS as archived by CMIP models (see Griffies et al. (2016))—spanning the ocean between the ACC and the Ross Sea continental shelf and the land, and the Ross continental shelf itself. The major gyres evident in SSH we refer to as the "Balleny Gyre" and "inner RG" and are typically found in long term means to the west and east of Scott Island and associated seamounts along ≈180°E, respectively; see, for example, satellite and model studies of the gyres in Armitage et al. (2018) and Roach and Speer (2019), respectively. The white boxes in Figure 1a outline the RSR MPA zones that our domain extension now covers, in particular regions (i) the General Protection Zone, (ii) the Krill Research Zone, and (iii) the Special Research Zone. Figure 1b shows the sub-regions 1 to 4 we will be using for our assessment, bounded by the white contours, the land, and the domain boundary, plotted on top of January mean surface chlorophyll. In our earlier study (Rickard and Behrens (2016)) the analysis was mainly confined to the continental

21699291, 2023, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022JC018880 by Ministry Of Health, Wiley Online Library on [23/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

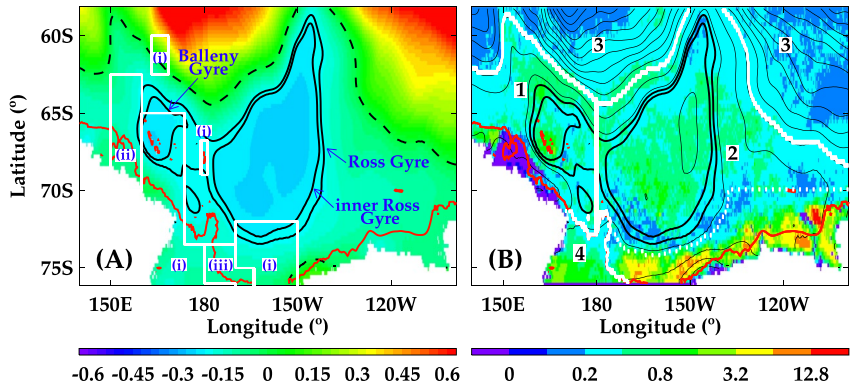

Figure 1. The Ross Sea Region (RSR) for this study spanning 140°E to 99.25°W and 76°S to 58.25°S. Color plots show (a) the mean sea surface height (SSH) (meters) with the area average adjusted to zero, and (b) the January mean surface chlorophyll (mg Chl-a m^{-3}) (plotted on a logarithmic scale and with negative values included for missing satellite data). The mean SSH in (a) is obtained from 6 years of altimetric observational data processed into monthly SSH means and provided by Ron Kwok (pers comm) from Kwok and Morison (2016). The January mean surface chlorophyll in (b) is from the MODIS-Aqua 4 km resolution product using the years 2002–2021, processed to produce an annual cycle climatology. The Ross Gyre (RG) here is defined by the last SSH continuous contour not intercepting land, approximated by the outer of the solid black curves at −0.19 m. The black dashed line locates the SSH at 0.0 m. The RG comprises two gyres separated by the Balleny Islands at 180°E, that we refer to as the "inner RG" (iRG) and the "Balleny Gyre" (BG), with their last approximate continuous contours shown by solid black contours at −0.21 m. Ross Sea MPA zones—outlined in white and bounded by land—are (i) the General Protection Zones (GPZ), (ii) the Krill Research Zone (KRZ), and (iii) the Special Research Zone (SRZ). The red contours locate 1000 m depths, showing the Ross Sea continental shelf region of interest that passes through the SRZ and part of the GPZ to the west of 180°E. Sub-regions 1 to 4 for our analysis are shown in (b), bounded by the white contours, the land, and the domain boundary. Dotted white line shows boundary between test areas for sub-region 2. The light black contours in (b) are the SSH at intervals of 0.08 m.

shelf; here we now account for oceanic influences on the shelf dynamics, and to also incorporate characterization of the recent RSR MPA. As a consequence sub-regions 1 to 3 are off-shelf and effectively cover the Balleny Gyre, the inner RG, and the ACC, respectively, while sub-region 4 covers the Ross Sea continental shelf bounded by the white contour along the 1,000 m isobath, the land, and the southern domain boundary. The southern boundary domain latitude of 76°S is a compromise over our CMIP5 and CMIP6 models to avoid coastline inconsistencies due to horizontal spatial resolution differences; and although this latitude would exclude the Ross Ice Shelf, it still enables consistent comparison between the models as none of the CMIP models analyzed here include ice shelves. Further, Figure 2 in Rickard and Behrens (2016) shows that the choice of 76°S still allows for analysis of the Ross Sea polynya in CMIP5 models, and Mohrmann et al. (2021) find that the "largest coastal polynya, the Ross Sea polynya, is represented in all models" for the 27 CMIP6 models they evaluated. The white contour forming the lower bound of sub-region 3 is defined by the black dashed line in Figure 1a along the SSH 0.0 m contour, where the Figure 1 area average SSH value has been set to zero; this contour approximately separates the gyral system from the stronger SSH gradients of the ACC in the north of the domain (shown by the color contours in Figure 1a and the light black contours in Figure 1b).

Table 1 shows the CMIP6 (models 1 to 16) and CMIP5 (models 17 to 32) ESM used in this study. Model datasets and user guidance can be found at https://esgf-node.llnl.gov/projects/cmip5/ and https://esgf-node.llnl.gov/projects/cmip6/. Each institution runs its own model a number of times to build up an ensemble defined by a "ripf" identifier representing realization, initialization, physics, and forcing. Here we choose one ensemble member per model based on "ripf" continuity between the historical and future forcing, and maximizing in particular the archiving of the BGC parameters in Table 1. For reference the r1i1p1f1 ensemble member is used for the CMIP6 models, except r1i1p2f1 for 2, r1i1p1f2 for 3, 6, and 7, r1i2p1f1 for 10, and r4i1p1f1 for 11. All CMIP5 models used the r1i1p1 ensemble. In each cell in column "Grid Cells" the left value ("1 + 2 + 3") shows the total number of grid model grid cells that cover (horizontally) sub-regions 1, 2, and 3, while the right value is the number of sub-region 4 grid cells (the sub-regions shown in Figure 1b). It is apparent that horizontal resolution has generally improved going from CMIP5 to CMIP6 (except for the MPI suite of models 8, 9, 30, 31 that retain their CMIP5 ocean resolution); indeed averaging these columns we see mean grid resolutions of (9531, 377) and (2463, 93) for CMIP6 and CMIP5, respectively, highlighting the investment in the CMIP6 ESM at these

21699291, 2023, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022JC018880 by Ministry Of Health, Wiley Online Library on [23/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 1
CMIP5 and CMIP6 Earth System Models Used in This Study With Their Respective Biogeochemical (BGC) Models and Which of the Primary Macronutrients Are Included in Each BGC Model

	Model Name	Grid cells 1 + 2 + 3, 4	BGC Model	NO_3	PO_4	SiO_4	Fe	Reference
1	CanESM5	4783,238	CMOC (1P,1Z)	Y	n	n	n	Swart et al. (2019)
2	CanESM5-CanOE	4783,238	CanOE (2P,2Z)	Y	n	n	Y	Swart et al. (2019)
3	CNRM-ESM2-1	4815,246	PISCESv2-gas (2P,2Z)	Y	Y	Y	Y	Aumont et al. (2015), Séférian et al. (2019)
4	ACCESS-ESM1-5	3699,216	WOMBAT (1P,1Z)	Y	Y	n	Y	Oke et al. (2013), Law et al. (2017), Ziehn et al. (2020)
5	IPSL-CM6A-LR	4815,246	NEMO-PISCES (2P,2Z)	Y	Y	Y	Y	Aumont et al. (2015), Boucher et al. (2020)
6	MIROC-ES2L	3446,138	OECO2 (2P,1Z)	Y	Y	n	Y	Hajima et al. (2020)
7	UKESM1-0-LL	4815,246	MEDUSA-2.1 (2P,2Z)	Y	n	Y	Y	Yool et al. (2021)
8	MPI-ESM1-2-HR	10859,350	HAMOCC (2P,1Z)	Y	Y	Y	Y	Ilyina et al. (2013), Paulsen et al. (2017)
9	MPI-ESM1-2-LR	1650,77	HAMMOC (2P,1Z)	Y	Y	Y	Y	Ilyina et al. (2013), Paulsen et al. (2017)
10	MRI-ESM2-0	3845,122	NPZD (1P,1Z)	Y	Y	n	n	Nakano et al. (2011)
11	CESM2	3244,101	MARBL (3P,1Z)	Y	Y	Y	Y	Danabasoglu et al. (2020)
12	CESM2-WACCM	3244,101	MARBL (3P,1Z)	Y	Y	Y	Y	Danabasoglu et al. (2020)
13	NorESM2-LM	4898,238	iHAMMOC (1P,1Z)	Y	Y	Y	Y	Tjiputra et al. (2020)
14	NorESM2-MM	4898,238	iHAMMOC (1P,1Z)	Y	Y	Y	Y	Tjiputra et al. (2020)
15	GFDL-CM4	70351,2543	BLINGv2 (2P,1Z)	n	Y	n	Y	Dunne et al. (2020)
16	GFDL-ESM4	18359,686	COBALTv2 (3P,3Z)	Y	Y	Y	Y	Stock et al. (2020)
17	CANESM2	1429,59	CMOC (1P,1Z)	Y	n	n	n	Zahariev et al. (2008), Christian et al. (2010)
18	CESM1BGC	3244,101	BEC (3P,1Z)	Y	Y	Y	Y	Moore et al. (2013)
19	CMCC-CESM	1180,54	PELAGOS (3P,3Z)	Y	Y	Y	Y	Vichi et al. (2007)
20	CNRM-CM5	4786,234	PISCES (2P,2Z)	Y	Y	Y	Y	Aumont and Bopp (2006), Séférian et al. (2013)
21	GFDL-ESM2G	2408,131	TOPAZ2 (3P,1Z)	Y	Y	Y	Y	Dunne et al. (2013)
22	GFDL-ESM2M	1968,64	TOPAZ2 (3P,1Z)	Y	Y	Y	Y	Dunne et al. (2013)
23	GISS-E2-H-CC	368,14	NOBM (4P,1Z)	Y	n	Y	Y	Gregg (2008)
24	GISS-E2-R-CC	368,14	NOBM (4P,1Z)	Y	n	Y	Y	Gregg (2008)
25	HadGEM2-CC	1883,55	Diat-HadOCC (2P,1Z)	Y	n	Y	Y	Palmer and Totterdell (2001)
26	HadGEM2-ES	1883,55	Diat-HadOCC (2P,1Z)	Y	n	Y	Y	Palmer and Totterdell (2001)
27	IPSL-CM5A-LR	1180,54	PISCES (2P,2Z)	Y	Y	Y	Y	Aumont and Bopp (2006), Séférian et al. (2013)
28	IPSL-CM5A-MR	1180,54	PISCES (2P,2Z)	Y	Y	Y	Y	Aumont and Bopp (2006), Séférian et al. (2013)
29	IPSL-CM5B-LR	1180,54	PISCES (2P,2Z)	Y	Y	Y	Y	Aumont and Bopp (2006), Séférian et al. (2013)
30	MPI-ESM-LR	1650,77	HAMOCC5.2 (1P,1Z)	Y	Y	Y	Y	Ilyina et al. (2013)

21699291, 2023, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022JC018880 by Ministry Of Health, Wiley Online Library on [23/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 1
Continued

	Model Name	Grid cells 1 + 2 + 3, 4	BGC Model	NO_3	PO_4	SiO_4	Fe	Reference
31	MPI-ESM-MR	10859,350	HAMOCC5.2 (1P,1Z)	Y	Y	Y	Y	Ilyina et al. (2013)
32	MRI-ESM1	3845,122	MRI.COM3+NPZD (1P,1Z)	Y	Y	n	n	Adachi et al. (2013) plus Oschlies (2001)

Note. "Grid Cells" counts the total number of horizontal grid cells over full domain (sub-region 0), shown in Figure 1. CMIP5 models use the r1i1p1 ensemble. CMIP6 models use r1i1p1f1 ensemble, except for r1i1p2f1 for 2, r1i1p1f2 for 3, 6, and 7, r1i2p1f1 for 10, and r4i1p1f1 for 11.

high latitudes. Biogeochemically the number of explicitly cycled nutrients is very similar, with 9 not cycled ("n") for CMIP5 compared to 12 for CMIP6. The range of complexity in CMIP5 and CMIP6 model representation of phytoplankton and zooplankton functional types remains comparable, with examples ranging from the "simplest" (1P, 1Z) (e.g., models 1, 4, 10, 13, 14, 17, 30, 31, and 32) up to the more "complex" (3P, 3Z) (e.g., models 16, 19) for both CMIP5 and CMIP6.

For comparison with previous studies—and because the CMIP5 models remain archived to a maximum historical year of 2005—we use output from both the CMIP5 and CMIP6 suites for the period 1976 to 2005. We have extracted 30 years of monthly means between 1976 and 2005 from the "historical" simulation of each model from the CMIP archives and constructed mean area average fields for each month for each sub-region (see Figure 1b), these 12 comprising a single representative (present-day) annual cycle. It is these mean fields we compare with observational annual cycles.

Historical data are from the World Ocean Atlas (WOA) 2009 database for salinity (Antonov et al. (2010)), temperature (Locarnini et al. (2010)), and the nutrients nitrate, phosphate, and silicate (Garcia et al. (2010)), and sea ice concentration (SIC) and sea surface temperature (SST) from HadISST (Rayner et al. (2003)). For MLD we use the monthly climatology of Boyer Montégut et al. (2004) based on a fixed density difference of 0.03 kgm^{-3}. For surface dissolved iron values in the top 10 m of the water column from the geotraces data set "Dissolved iron (2015 Version 2)" were used (Tagliabue et al. (2012)). Surface chlorophyll monthly means and standard deviations from the MODIS-Aqua 4 km resolution product using the years 2002–2021 have been processed to produce an annual cycle climatology (12 monthly means and standard deviations). For integrated Net Primary production (NPP) we use Vertical Generalized Production Model output (VGPM, Behrenfeld and Falkowski (1997)) and carbon-based productivity model CbPM (Westberry et al. (2008)), both obtained from the Ocean Productivity website https://www.science.oreganstate.edu/ocean.productivity. For zooplankton information, we use monthly climatologies available for 6 months (October to March) using the estimated total zooplankton abundance as detailed in Pinkerton et al. (2020). Note that our central results will be based on using WOA SST, and VGPM for NPP.

Metrics of area average root-mean-square-error (RMSE) and bias (BIAS, [model - observation]) using monthly mean model and observational data (see, e.g., Rickard and Behrens (2016)) are calculated for each sub-region 1 to 4. In order to quantify relative outlier models as a way of obtaining a model hierarchy we use the mean RMSE and BIAS values from each model from each sub-region for each variable, and a model outlier is one that has mean RMSE and BIAS values that lie outside the multi-model mean plus standard deviation. This process is repeated for each model for each variable for each sub-region to build up outlier distributions that can then be compared across all the models. For model 19 for surface phosphate for sub-regions 1 to 4, and for models 30 and 31 for mixed layer depth for sub-regions 1 to 2, their respective mean RMSE and BIAS errors are already substantially larger than any of the other model errors, and so their RMSE and BIAS contributions are not included in the multi-model means used to determine outliers, but they are of course recorded as outliers as well. For each variable we can also use the RMSE and BIAS values to obtain ranking positions (as in Rickard and Behrens (2016)), and these ranking positions are then summed and averaged to obtain an overall score per model per variable (where the minimum score is one, i.e., that model scores the lowest for RMSE and BIAS for each subregion, and the maximum score is equal to the number of models in the variable sample).

For transport estimates we have calculated RG transport (RGTr) using the CMIP variable "msftbarot" (the barotropic or quasi-barotropic streamfunction in Griffies et al. (2016)), and used values from the literature for ACC transport (ACCTr). Outlier models for these transports are deemed to be those outside twice the uncertainty for RGTr, and outside twice the observational uncertainty for ACCTr. Not all our sampled models provide msftbarot,

21699291, 2023, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022JC018880 by Ministry Of Health, Wiley Online Library on [23/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

so we additionally calculate the strength of the RG (RGStr) using the model SSH (the CMIP variable ZOS). RGStr is defined to be the height difference between the Gyre centre and the last bounding height contour of the Gyre.

For the ESM climate sensitivity metrics ECS and TCR we have obtained values from the literature, and deem those that are outliers to have ECS and TCR values beyond accepted bounds of 4.5°K and 2.5°K, respectively.

For quantifying anomalous deep SO model convection we have again taken values from literature for the CMIP5 and CMIP6 models and used approximate scalings to classify the models into non-convecting, intermediate convecting, and the remainder to be excessively convecting. While it is accepted that an excessively convecting model is not representative of the observations (see, e.g., De Lavergne et al. (2014) and Heuzé (2021)), this does not necessarily imply that a non-convecting model is generating Antarctic bottom water correctly or at all. Further, Gjermundsen et al. (2021) show that a model with acceptable climate sensitivity metrics ECS and TCR might be obtaining these via excessive convection in the first place. By classifying outliers for the climate sensitivity parameters and the anomalous convection properties, our ranking scheme aims to find those models that at least adequately satisfy both.

3. Results

To motivate the labeling in the figures to follow, we first present our result which is an ordering of our models into ensembles we label as "innermost" (the "best"), "outer" (the worst), and the remainder "inner" models. As noted, this is based on considering the distribution per model of outlier estimates shown in Figure 3. Those ensembles are listed in Table 2. The colors red, blue, and black will be used in Figures 3–7 to label the respective models based on the ensemble they are in.

Examples of RMSE and BIAS values for the models for different variables are shown in Figure 2. For example, Figure 2a is for surface chlorophyll in sub-region 1, and the outliers here are identified as models 8, 9, 30, and 31, as both their mean RMSE and BIAS values lie outside the multi-model mean plus standard deviation (the horizontal black and red dashed lines for RMSE and BIAS, respectively, in Figure 2). These models are then shown on our summary Figure 3 as red dots with a number 1 in row i for chlorophyll. Figures 2b to 2d reveal outliers 17 and 19 for surface nitrate (NO_3) in sub-region 2, models 11, 12, 17, 20, 23, and 29 for sea surface salinity (SSS) in sub-region 3, and models 6, 29, and 32 for sea ice concentration (SIC) in sub-region 4. Repeated application of this process results in the labeling of rows a to j for each model for each sub-region for each variable in rows a to j in summary Figure 3 (noting that models that do not explicitly cycle a variable, or a value could not be found at the time of writing, are labeled with "0" and "?," respectively). Note that row e is for "mlotst" which is the CMIP mixed layer depth based on a density difference of 0.03 kgm^{-3} (see Griffies et al. (2016)). Row j is for "Intpp" which is the CMIP variable for vertically integrated NPP (see Orr et al. (2017)).

Rows RGTr and ACCTr in Figure 3 rank model estimates of mean inner RG and ACC transport for the present day, with a red dot labeled R or A for those models considered relative outliers. Figure 4c plots ACC transports obtained from Table 2 in Beadling et al. (2019) and Beadling et al. (2020) for our models and ensembles, with outliers here having transports outside twice the observational uncertainty reported by Donohue et al. (2016) (the horizontal black dashed lines) about the Donohue et al. (2016) mean of 173.3 Sv (the horizontal black line). Model inner RG transport is calculated from the CMIP output msftbarot (the barotropic or quasi-barotropic streamfunction in Griffies et al. (2016)) and is taken to be the peak transport (westward) between the core of the inner RG and the coastline. Dotto et al. (2018) and references therein suggest a range of somewhere between 15 and 30 Sv as the best estimates (from models and inverse approaches); conservatively we approximate twice that uncertainty and define model outliers to have transports outside the range 7–39 Sv and these are shown as red dots labeled with R in Figure 3. Figure 4 frame (A) shows the mean barotropic stream function for model 10, and the vertical black line is the section for calculating the Gyre transport, and Figure 4d plots the values for our models and ensembles, with the mean of 23 Sv the approximate centre of the Dotto et al. (2018) references.

For row RGTr in Figure 3 many of the models—especially CMIP5—are labeled "?" showing that the CMIP parameter msftbarot was not available at the time of writing, and so we have supplemented those estimates with the strength of the inner RG RGStr using the CMIP SSH variable "ZOS," and Figure 4b is an example, again for model 10. The Gyre strength is defined to be the height difference between the Gyre centre and the last bounding contour of the Gyre and is shown as the red and white horizontal lines for the observational data and

21699291, 2023, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022JC018880 by Ministry Of Health, Wiley Online Library on [23/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

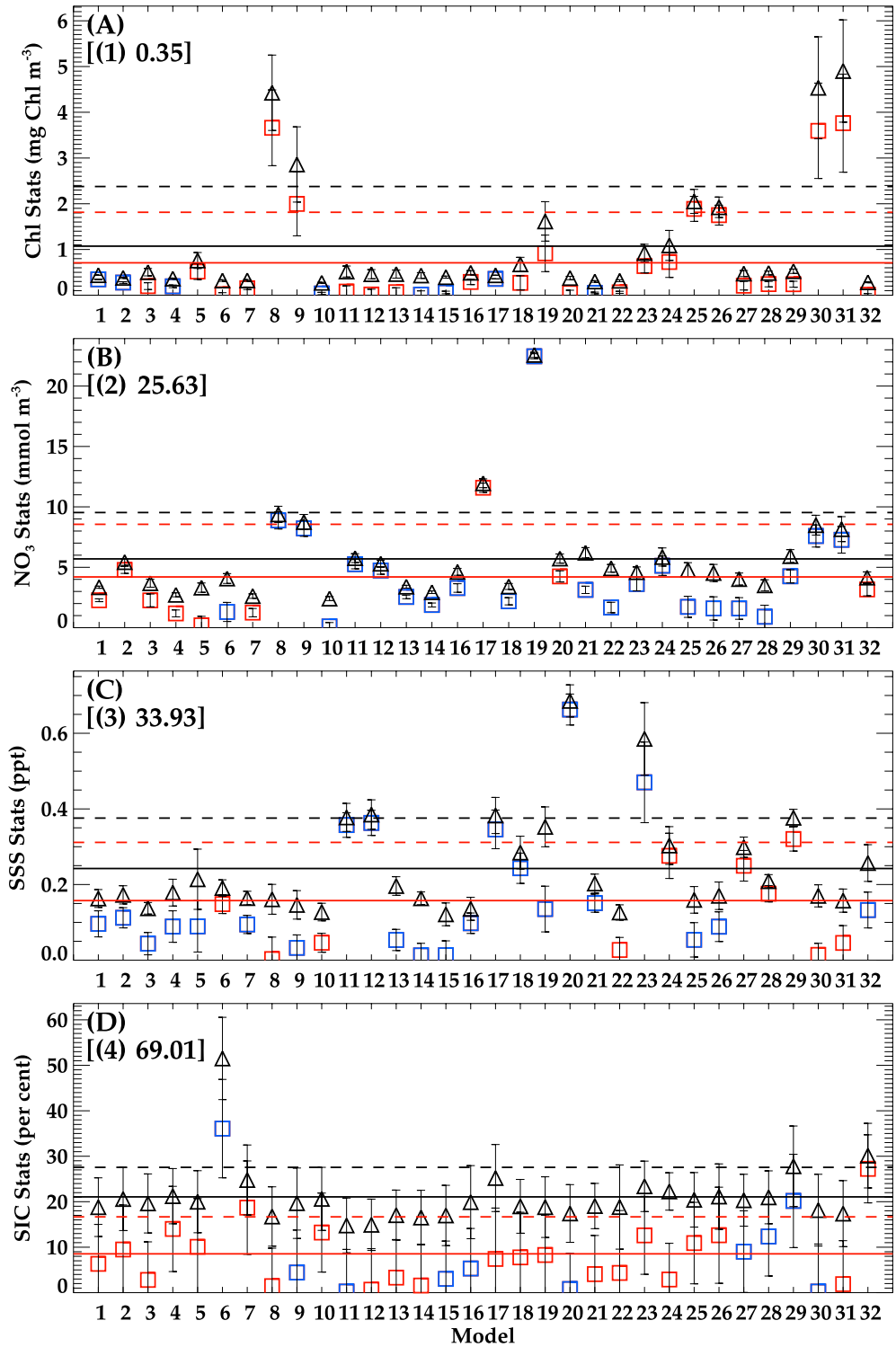

Figure 2. Mean RMSE (△) and BIAS (□) examples for each model for (a) chlorophyll, (b) nitrate, (c) SSS, and (d) SIC. In each frame the numbers in the square parentheses are the sub-region (in brackets) and the annual mean observed value for the respective variable. Horizontal black and red lines are the mean RMSE and BIAS for all the models, with the respective dashed lines at one standard deviation from the mean. Red and blue squares show positive and negative bias with respect to the mean observed value.

21699291, 2023, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022JC018880 by Ministry Of Health, Wiley Online Library on [23/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

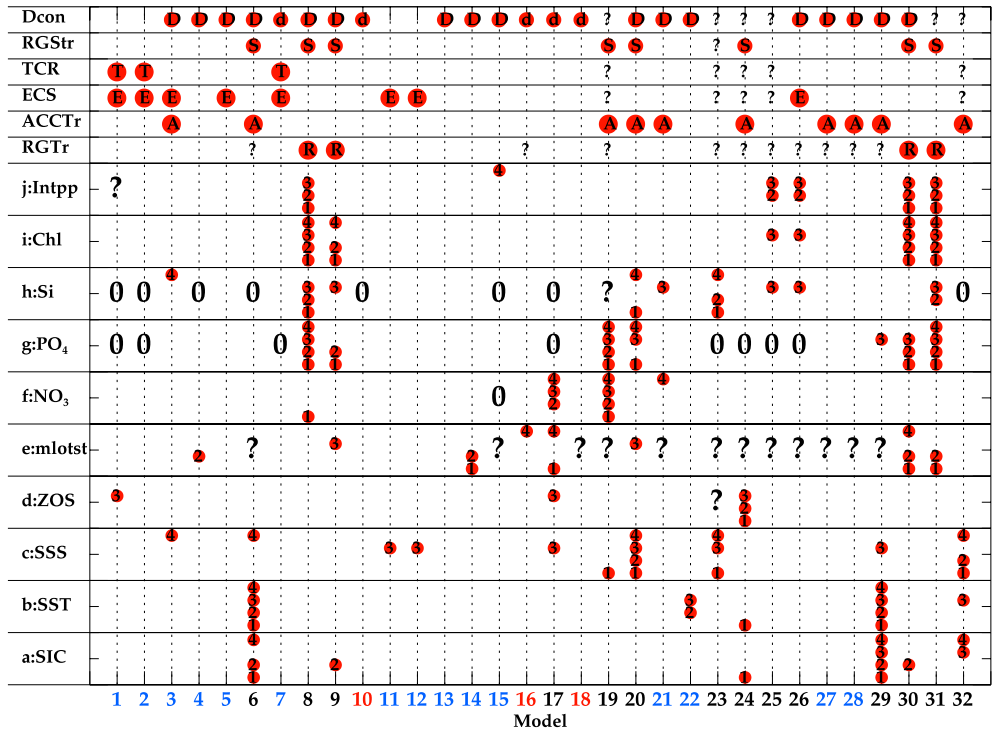

Figure 3. Plot of models against variables to identify outliers. Model number colors identify the ensembles in Table 1. Outlier models are labeled with red dot for anomalous Southern Ocean deep convection (Dcon), Ross Gyre (RG) strength (RGStr), Transient Climate Response (TCR), Equilibirum Climate Sensitivity (ECS), Antarctic Circumpolar Current transport (ACCTr), and RG transport (RGTr). The top row Dcon labels models having "excessive" or "intermediate" values of deep convection (see text) with a red dot containing D or d, respectively. The rows a to j are for sea ice concentration (SIC), sea surface temperature (SST), sea surface salinity (SSS), sea surface height (CMIP variable ZOS), mixed layer depth (CMIP variable mlotst), the surface nutrients nitrate (NO_3), phosphate (PO_4), silicate (Si), surface chlorophyll (Chl), and vertically integrated primary production (CMIP variable Intpp). For each model a "0" indicates a variable that is not explicitly represented by that model, and a "?" denotes that information for the variable at the time of writing could not be obtained from either the relevant CMIP archive or from the literature. For the rows a to j the red dots for each outlier are labeled with the sub-region 1 to 4 in which they are located.

the model, respectively. Observational analyses of the SSH are used to find a mean and standard deviation of the Gyre strength (using 6 years of monthly SSH means provided by Ron Kwok (pers comm) from Kwok and Morison (2016)), and as before conservative outlier models are those more than twice the standard deviation away from the mean, and here shown as red dots labeled with S in Figure 3. The mean model Gyre strengths are plotted in Figure 4e, with the mean observational strength found to be 0.06 m. As "ZOS" is available from more models compared to msftbarot, we see that we need only add models 6, 19, 20, and 24 as outliers characterizing the inner RG circulation pattern and transport. Comparing Figures 4d and 4e there is some correspondence between excess transport and strength for example, models 8, 9, 30, and 31, but conversely not so for model 20. Nevertheless, in this manuscript poor representation of either transport or strength determines a model outlier. Comparing Figures 4c and 4d we also see that there is no obvious relationship between the model ACC and RG transports with models 8, 9, 30, and 31 RG outliers but not for ACC, but conversely model 3 for ACC (on the low side) but not for the RG.

Table 2
Ensembles Ranked Into Innermost, Inner, and Outer, Their Reference Color for Plots Referring to the Model Numbers, and the Specific Models in Each Ensemble

Ensemble	Color	Models
Innermost	Red	10, 16, 18
Inner	Blue	1,2,3,4,5,7,11,12,13,14,15,21,22,27,28
Outer	Black	6,8,9,17,19,20,23,24,25,26,29,30,31,32

Rows ECS and TCR show models that exceed accepted values of 4.5°K and 2.5°K as red dots labeled E and T, respectively; models exceeding these thresholds are generally found to be inconsistent with observed global warming, for example, Nijsse et al. (2020). For the 32 models here we have first taken the ECS and TCR values from Nijsse et al. (2020). Gaps have then been filled from other sources, in particular Gjermundsen et al. (2021) and Tokarska et al. (2020) for ECS for MIROC-ES2L, ECS for MPI-ESM1-LR, NorESM2-MM, and IPSL-CM5B-MR from Gjermundsen et al. (2021), TCR

21699291, 2023, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022JC018880 by Ministry Of Health, Wiley Online Library on [23/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

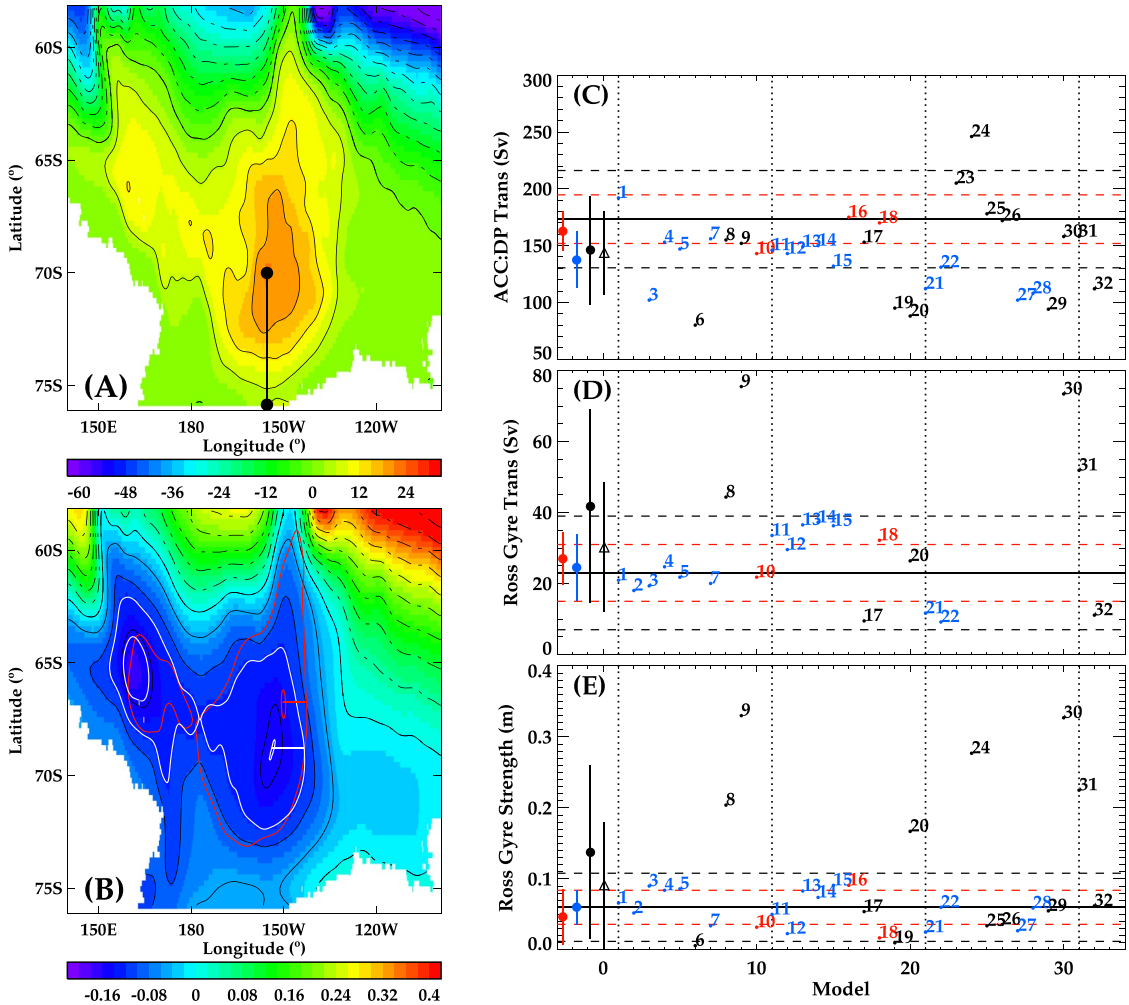

Figure 4. Color plots of mean historical barotropic streamfunction (Sv) and sea surface height (m) in (a) and (b), respectively, for model 10 (MRI-ESM2-0). In (a) contours every 6 Sv (dashed contours negative), and vertical solid line from Gyre centre to the coastline locates section across which the (westward) inner Ross Gyre (RG) transport for this model is calculated. In (b) black contours every 0.05 m (dashed contours positive), red and white contours show the central and bounding heights defining the inner RG for the observational data and for this model, respectively, and the strength of the inner RG is defined as the height difference between these contours (indicated by the horizontal red and white lines). For each model frames (c, d, and e) plot the mean Antarctic Circumpolar Current transport (Sv), the inner RG transport (Sv), and the inner RG strength (m), respectively, with the horizontal line the mean observational value, and the red and black dashed lines one and two standard error estimates, respectively. The vertical dotted lines indicate models 1, 11, 21, and 31. Ensemble averages and spreads are the vertical lines below zero in red (circle), blue (circle), and black (circle) for the innermost, inner, and outer models, respectively, and the black (triangle) for the full model ensemble.

for MPI-ESM1-LR and NorESM2-MM from Brunner et al. (2020), and ECS and TCR for CESM1BGC from Bacmeister et al. (2020). Comparing each of these sources for all models—including Forster et al. (2013) for CMIP5 models—the ECS and TCR are all consistent across the different sources such that any variations between them do not account for models exceeding the ECS and TCR thesholds. The result is—for models where we have a value of ECS or TCR—that 43% and 19% of our sampled CMIP6 models exceed the ECS and TCR thresholds, respectively, compared to 19% and 0% of CMIP5 models. Reasons for these changes between CMIP5 and CMIP6 have been attributed to cloud feedbacks (see, e.g., Meehl et al. (2020), Schlund et al. (2020), and Zelinka et al. (2020)), and the links to SO convection discussed by Gjermundsen et al. (2021).

The top row Dcon in Figure 3 labels models having "excessive" or "intermediate" (defined below) values of deep convection with a red dot containing D or d, respectively. This criterion is motivated by the study of Gjermundsen et al. (2021) (in particular their supplementary Tables 3 and 4) on the influence of SO deep convection on the resultant ECS of a model, namely that substantial deep ocean warming is generally associated with low (i.e., <4.5°K) ECS values. To explore this relationship for our model suite, Figure 5 replots Supplementary Figure 1a in Gjermundsen et al. (2021), as well as including estimates of maximal SO deep convection volumes (DMVmax)

21699291, 2023, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022JC018880 by Ministry Of Health, Wiley Online Library on [23/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

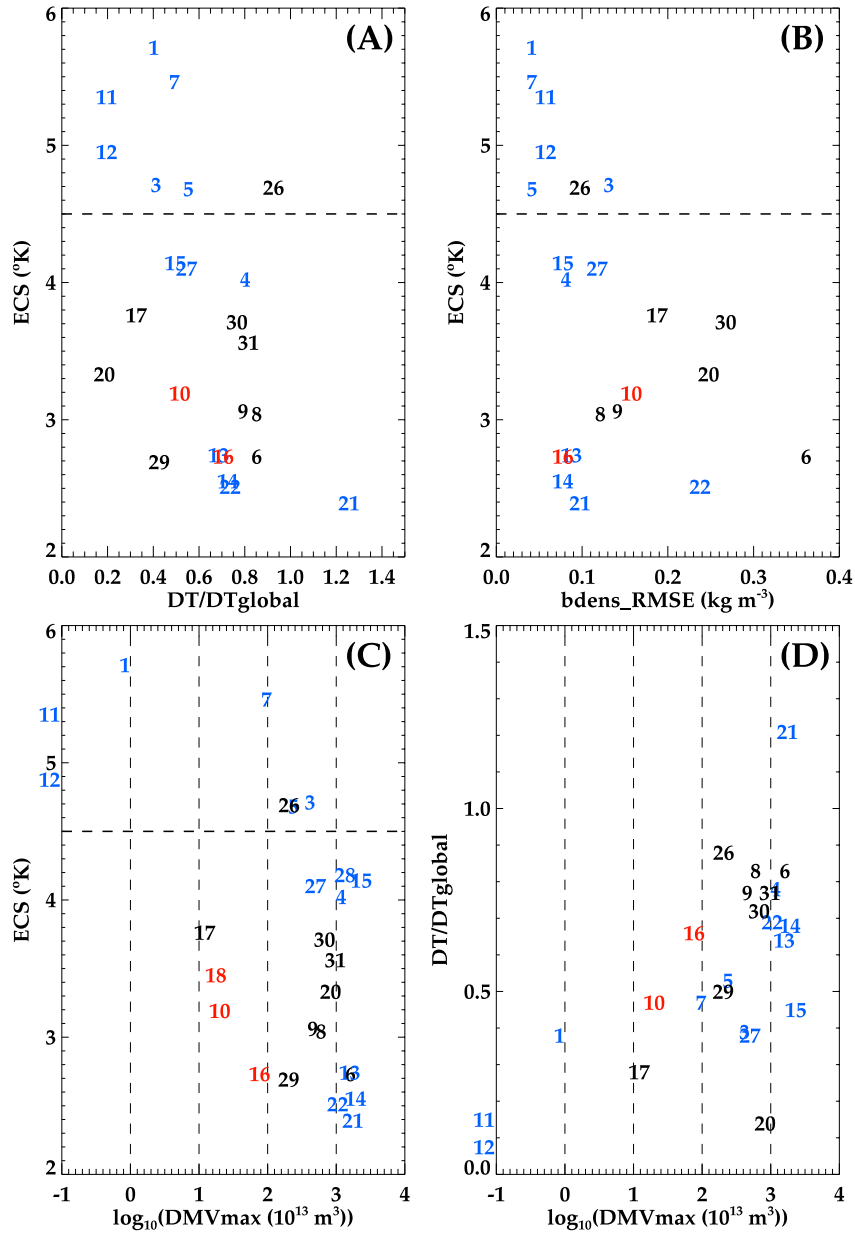

Figure 5. Model deep water properties influencing Equilibrium Climate Sensitivity (ECS) (°K), in particular (a) ECS against DT/DTglobal, (b) ECS against bdens_rmse (kg/m^3), (c) ECS against $\log_{10}(DMVmax)$ (10^{13} m^3), and (d) DT/DTglobal against $\log_{10}(DMVmax)$. Note that models 11 (CESM2) and 12 (CESM2-WACCM) have zero DMVmax and so are plotted outside frames (c, d). DT/DTglobal is the change in ocean temperature below 1,500 m and south of 50°S relative to global ocean temperature change (Gjermundsen et al. (2021)), bdens_rmse is the bottom density RMSE (Heuzé et al. (2013); Heuzé (2021)), and DMVmax is the maximal Southern Ocean deep convection volume (De Lavergne et al. (2014), Heuzé (2021).

for CMIP6 (see Table 2 in Heuzé (2021), where we have taken DMVmax to be the sum of the Weddell, Amery, and Ross maximums) and CMIP5 (see Table S1 in De Lavergne et al. (2014)) and bottom density root-mean-square errors (bdens_rmse) taken from Figures 1 and 2 in Heuzé (2021) and Heuzé et al. (2013), respectively. Note that Table S1 in De Lavergne et al. (2014) actually only provides the maximum convection area for MLD > 2,000 m over September (since maximum convection depths are typically found at austral winter end); here maximal volumes are estimated by scaling the areas with an average depth of ≈3,000 m for the SO. As a result, for example, the relatively small volumes we estimate for models 17 and 18 of 7.2 and 10.5 10^{13} m^3, respectively, are consistent with these models being considered non-convecting by De Lavergne et al. (2014), as well as the

21699291, 2023, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022JC018880 by Ministry Of Health, Wiley Online Library on [23/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

values in Table 2 of Behrens et al. (2016) based on a maximum MLD criterion of 500 m and using long-term means. Further consistency with these volume estimates is apparent when comparing the CMIP6 MPI DMVmax (10^{13} m^3) of 328.0 (LR) and 435.0 (HR) from Heuzé (2021) with CMIP5 MPI values of 407.7 (LR) and 585.6 (MR) from De Lavergne et al. (2014), since Mauritsen et al. (2019) note that their CMIP6 ocean and sea ice components "remained largely unchanged with respect to the MPI-ESM model version used during CMIP5."

Figure 5a for our model suite indeed shows the same tendency for deep SO trends in temperature of DT/DTglobal to be associated with relatively smaller values of ECS as found by Gjermundsen et al. (2021) (where the dashed horizontal line shows the 4.5°K threshold), where DT/DTglobal is the ratio of the ocean temperature change below 1500 m and south of 50°S relative to the global ocean temperature change. Gjermundsen et al. (2021) conclude that models (particularly CMIP6) might be achieving acceptable ECS values through unphysical (excessive) SO deep convection resulting in warming at depth. Figure 5b plots ECS against the model bottom density RMSE values and shows a similar trend to Figure 5a, ie, models with a poorer RMSE tend to have a lower ECS. Figures 5c and 5d highlight the ECS and deep SO temperature trend dependence on DMVmax, suggesting that indeed the larger DMVmax values are associated with reduced model ECS and relatively increased warming at depth.

Taking guidance from De Lavergne et al. (2014) that models 17 and 18 are "non-convecting," we denote intermediate convecting models as those in the range 1–100 10^{13} m^3 in Figure 5c; our excessive models then lie beyond this range. This ranking is consistent with Heuzé et al. (2013) who note that "most [CMIP5] models create deep water by open ocean deep convection, a process occurring rarely in reality," and that even for CMIP6 for the SO most models "form deep and bottom water via open-ocean deep convection too deeply, too often, and/or over too large an area" (Heuzé (2021)). This ranking has been used to populate the Dcon row in Figure 3. While it is beyond the scope of this paper to establish coupled process feedbacks, we note that of the five intermediate models ("d") in row Dcon, only model 7 has an ECS beyond the 4.5°K threshold. Note that Figures 5c and 5d plot models 11 (CESM2) and 12 (CESM2-WACCM) DMVmax values outside the frame; this is because these models use a parameterization to transport dense shelf water to the deep basins (Briegleb et al. (2010)) and therefore have no false deep convection.

In order to obtain a ranked hierarchy of models, we first consider rows a to j in Figure 3 and consider our "outer" models to be those that are outliers (red dots) for multiple variables over multiple sub-regions. Of the remainder, we consider our "innermost" models to be those with no red dots from row RGTr to RGStr, and at most a red dot with d ("intermediate convecting") in row Dcon. Those left over we consider as "inner" in that they perform well over rows a to j in terms of the surface metrics, but perform relatively poorly over the top 6 rows. As a result of this process we obtain the model hierarchy shown in Table 2. The ensemble colors have been used in Figure 5, for example, motivating some of the convection choices made.

Average rankings for the physical and biogeochemical variables for the shelf sub-region 4 and off-shelf sub-regions 1 to 3 are shown in Figure 6. The triangles show the average ranking over variables R_{AV}, with the individual ranking per variable for each model labeled with the lower case letters, with the models identified along the $y = 0$ axis in each frame. The vertical green bars per model in Figure 6 show the spread in rankings per variable. For the physical variables in Figures 6a and 6c the CMIP6 models are typically the best ranked, with 11 and 13 in the top 16, respectively. For the biogeochemical variables in Figures 6b and 6d the top rankings are shared between CMIP5 and CMIP6, but with the CMIP6 models occupying most of the best rankings off-shelf. Overall, the combined rankings show a definite improvement over the Ross Sea Region in CMIP6 compared to CMIP5, particularly off-shelf and more generally over the physical variables. The poorest CMIP6 model for physical variables is model 6 MIROC-ES2L—and is actually the lowest ranked model overall—and is more than likely due to the known warm biases over the SO (see Hajima et al., 2020); despite this MIROC-ES2L is actually one of the best ranked biogeochemical models. Using just these rankings then we would not necessarily have chosen models 10, 16, and 18 (red) to be our innermost models; indeed apart from model 10 off-shelf (frames (C) and (D)) these models rank relatively conservatively, and their innermost rating results from use of the information from the upper rows in Figure 3.

Figure 6e takes the mean values (the triangles) on the shelf from frames (A) and (B) and plots the results of the physical (the *x*-axis) against the biogeochemical (the *y*-axis) components, with the best ranked models being closest to the origin. The spread of models shows that our pre-selection does indeed result in the outermost models generally ranking less well than the innermost and inner models, except for models 25 and 26 that rank equally

21699291, 2023, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022JC018880 by Ministry Of Health, Wiley Online Library on [23/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

with our innermost choices. Off-shelf in frame (F) the separation is much clearer, with nearly all the outermost models ranking comparatively worse. The averages over frames (E) and (F) are shown in frame (G), and reinforce our selections of the innermost and inner models as providing the best ranked properties overall, and with 13 of the models in the innermost and inner cohort of 18 being CMIP6 suggests general improvements in model representation over the Ross Sea Region since CMIP5. As previously noted CMIP6 outlier model 6 suffers from its relatively poor physical scores here; models 8 and 9 are part of the MPI suite represented by models 30 and 31 in CMIP5 and each has a similar ranking (and as noted previously their respective ocean and sea ice components remain largely unchanged (Mauritsen et al. (2019))).

Sensitivity to the choice of observational data sets is examined by replacing our reference SST from WOA by SST from HadISST, replacing our reference Intpp from VGPM by CbPM, and by not considering the winter months July to September in NO_3 from WOA (referred to as $NO3^{[no789]}$). These choices are motivated by the low winter observational data points in WOA over sub-region 4 the Ross Sea continental shelf (see in particular the SST and NO_3 plots in Figures 9 and 10, respectively), and for comparisons of the differences in the VGPM and CbPM algorithms over the SO. For the critical identification of the outliers in Figure 3 the replacement data sets produce no changes to those already identified off shelf. On the shelf the outliers are identical for SST from WOA or HadISST; $NO3^{[no789]}$ is the same except model 8 is found to be an outlier rather than model 17; and for CbPM models 11, 12, 15, and 18 are now classified as outliers compared to only model 15 for VGPM. These overall small differences would make no substantive changes to our ensemble choices based on outliers.

There is, however, sensitivity to the data set changes in the model rank positions compared to our reference rankings plotted along $y = 0$ in Figure 6. Off shelf in frames (C) and (D) in Figure 6 the modified rank positions are shown when using HadISST and CbPM, respectively; it is apparent that the modified and reference rankings in each are very similar, and this similarity is quantified by the parameters that measure a model's change in rank position between the reference and modified rankings. We define a distance "Dist" to be a model's reference rank position minus the modified rank position. For example, in frame (C) model 18 ranks at positions 21 and 22, so that Dist $= -1$, that is, model 18 gets a slightly worse ranking when using HadISST. From the distribution of Dist values we compute a root mean squared distance (RMSD), the number of models that have Dist between -2 and 2 (nDist2), and the maximum Dist (DistMax). In frames (C) and (D) the reference and modified rankings are "close" in that the RMSD values are just over 1, the nDist2 values of 30 cover most models, and that the DistMax of 4 apply to relatively few models, so that off shelf the final rankings are not overly sensitive to the change in data sets. On shelf in frames (A) and (B) the use of HadISST and $NO3^{[no789]}$, respectively, produces comparably more sensitivity with (RMSD, nDist2) of (3,21) and (2.4,25), respectively, but the overall rankings can be seen to remain similar. However, for CbPM in frame (B) there is now a marked sensitivity with (RMSD, nDist2) of (5.6,10), and in fact with a DistMax of 13 model 4 is promoted from a ranking position of 17 to 4; however, note that of the 16 best models using VGPM, 14 of those remain in the best 16 using CbPM, so that the change in observational data set has not largely changed the balance of what we would rank as the "better" models (at least based on the metrics we have used).

Note that if rather than computing the average rank position based on the average rank score R_{AV} over the rank score per variable in Figure 6, we instead used the median of the rank scores per variable, then off shelf we find the reference and modified median rankings (not shown) have equally "close" Dist statistics to those shown in frames (C) and (D). The median statistics are also found to be similar to those shown for SST using HadISST in frame (A). In frame (B) the median Dist statistics using VGPM are (RMSD, nDist2) of (3.4,15), so less than the sensitivity when replacing VGPM with CbPM. Finally, comparing VGPM median with CbPM median statistics gives (RMSD, nDist2) of (5.7,15), so now comparable with the VGPM replaced with CbPM above. Importantly, however, for all these cases the majority (at least 14) of the reference top 16 models remain so. This means that either changing some of the reference data sets, or using statistics based on medians compared to averages, does not alter our conclusions about what we would consider the "better" ranked models compared to the "worst" ranked models.

4. Projection Results

Following Rickard and Behrens (2016) we present examples of end of century changes for the 20 year period 2081 to 2100 relative to the 30 years historical reference period 1976 to 2005 for scenarios RCP8.5 and SSP585 for CMIP5 and CMIP6, respectively, using our derived ensembles. As described by Bopp et al. (2013) using Meehl et al. (2007) a "robust" change has both the magnitude of the model mean change greater than the inter-model standard deviation (spread), and more than 80% of the models agree on the sign of the mean change; in the

21699291, 2023, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022JC018880 by Ministry Of Health, Wiley Online Library on [23/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

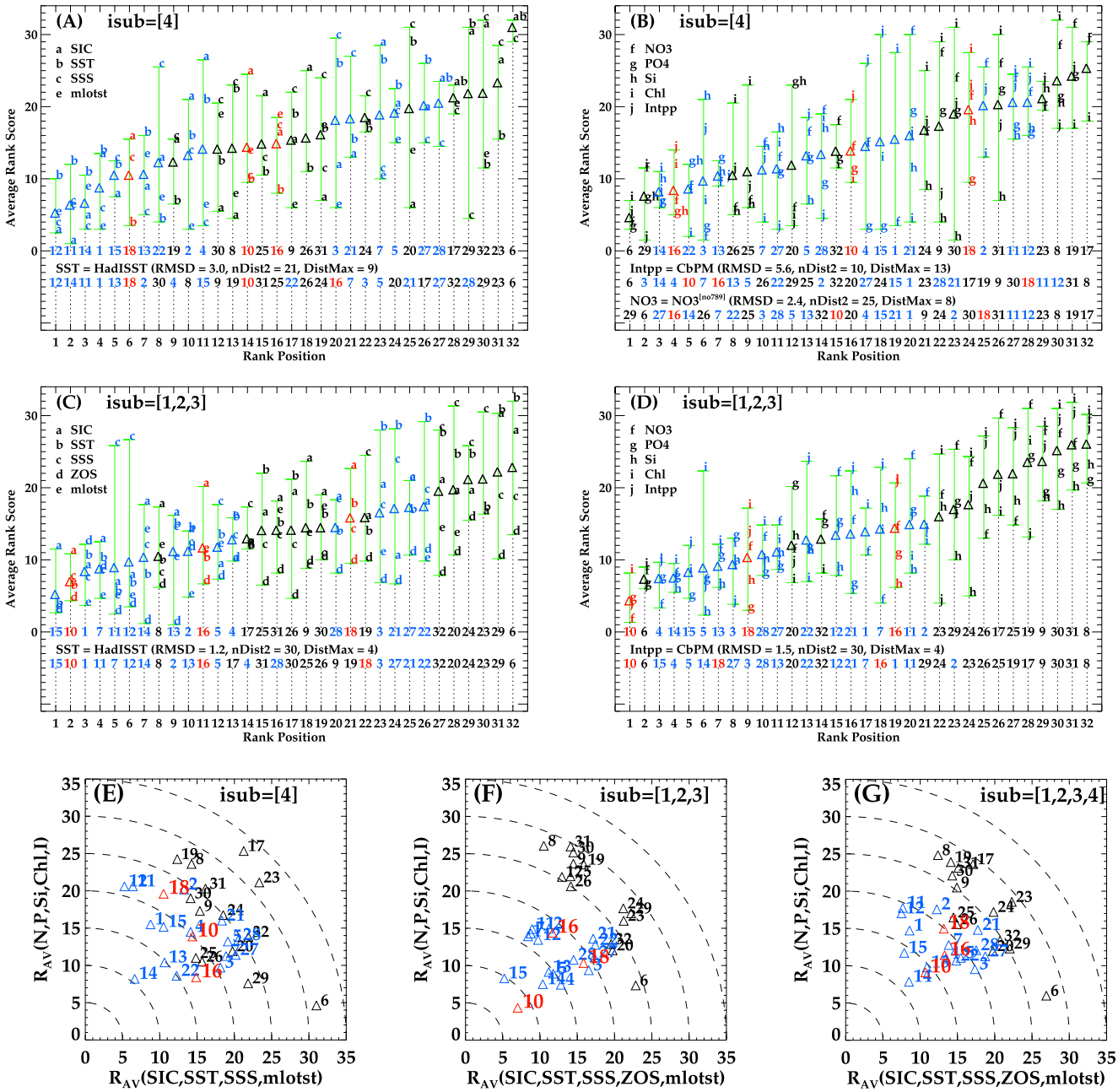

Figure 6. Rank score per model per variable for sub-region 4 for (a) [SIC (a), SST (b), SSS (c), mlotst (e)], (b) [NO3 (f), PO4 (g), Si (h), Chl (i), Intpp (j)], and for sub-regions 1 to 3 for (c) [SIC (a), SST (b), SSS (c), ZOS (d), mlotst (e)], and (d) [NO3 (f), PO4 (g), Si (h), Chl (i), Intpp (j)]. Rank position determined by the average rank score R_{AV} (Δ) over the respective variables. Red, blue, and black model numbers along $y = 0$ in frames (A) to (D) are for the innermost, the inner, and outer model ensembles, respectively. Vertical green lines span minimum to maximum rank scores per variable for each model, and the vertical black dotted lines link model number to its average score. Average rank scores R_{AV} per model for sub-regions (E) 4, (F) 1 to 3, and (G) 1 to 4. For each the y-axis R_{AV} is the average ranking over the biogeochemical variables NO3, PO4, Si, Chl, and Intpp; for frame (e) for sub-region 4 the x-axis R_{AV} average ranking for the physical variables does not include ZOS, whereas for frames (F) and (G) ZOS is included. In frames (A) to (D) the model rankings shown below $y = 0$ are for sensitivity checks to the change in the indicated variable (see text).

relevant figures robust changes are shown as black squares about the mean change. Figure 7 is for the physical variables, and shows that the ensembles consistently predict SST increases for each sub-region (frame (B)), with similarly consistent decreases in SIC (A), SSS (C), and mlotst (D). For the innermost ensemble (red) most of these changes are robust, except for sub-regions 1 and 4 for SIC. So the general physical picture is of a warmer, fresher, and more stable surface ocean, with significant decreases in the mean SIC.

21699291, 2023, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022JC018880 by Ministry Of Health, Wiley Online Library on [23/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

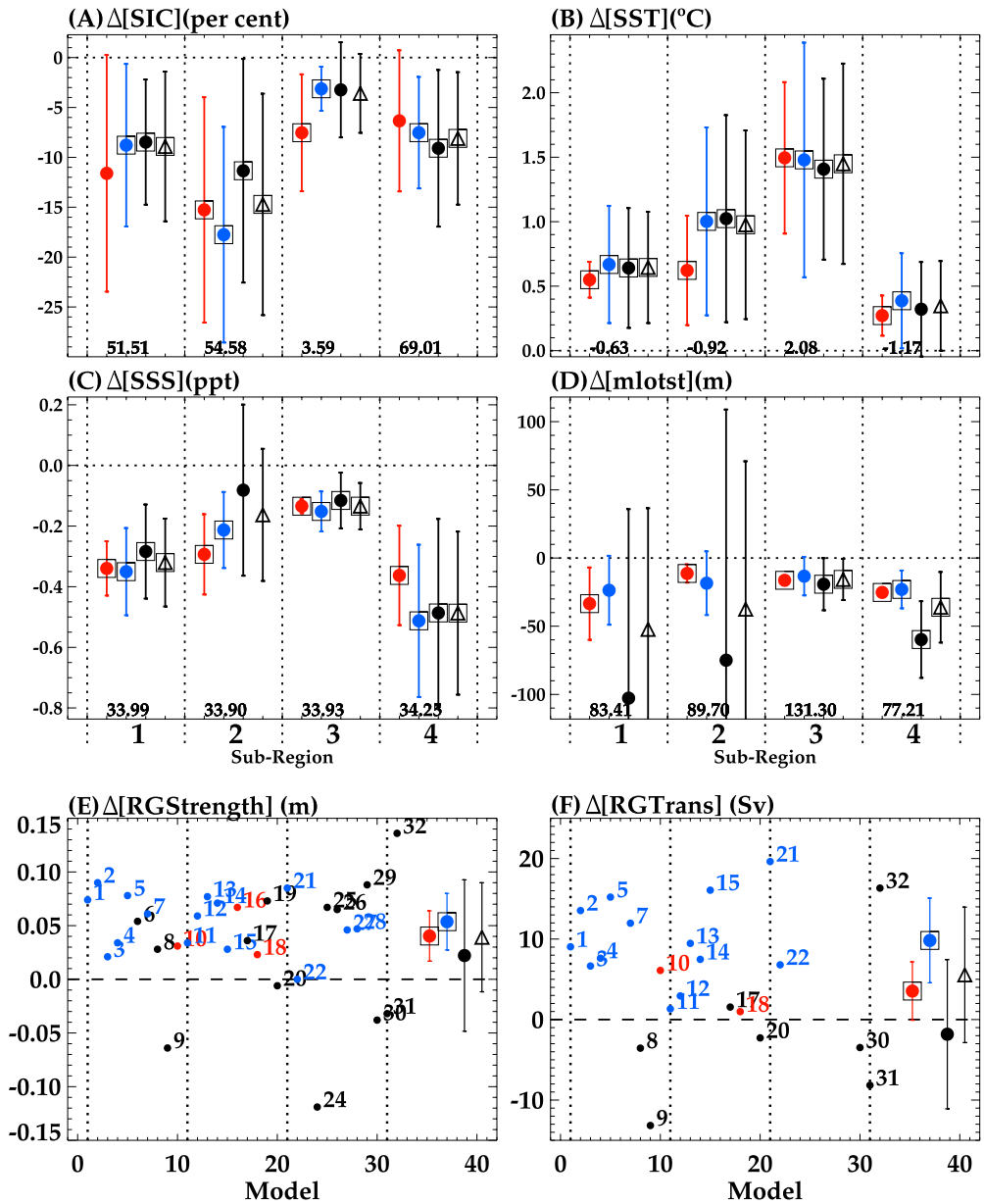

Figure 7. Ensemble changes for each sub-region labeled 1 to 4 for (a) SIC (per cent), (b) SST (°C), (c) SSS (ppt), and (d) mlotst (m). Vertical dotted lines separate sub-regions, and horizontal dotted lines show zero change. Ensemble averages and spreads are the vertical lines in red (circle), blue (circle), and black (circle) for the innermost, inner, and outer models, respectively, and the black (triangle) for the full model ensemble. Present day mean observational values per variable per sub-region included at bottom of each frame. Frames (E) and (F) show mean changes in Ross Gyre strength (m) and transport (Sv) per model, respectively, including ensemble averages and spreads, and with vertical dotted lines locating models 1, 11, 21, and 31. Changes are for RCP8.5/SSP585 for period 2081 to 2100 relative to historical period 1976 to 2005. Black squares on ensemble averages indicate "robust" changes.

Figures 7e and 7f show the model changes in RG strength (m) and transport (Sv), respectively, and we see robust increases in both according to our innermost (red) and inner (blue) ensemble averages. For the outer models we see—apart from model 8—that an increase/decrease in strength occurs for transport as well, generally reinforcing our link between the strength and transport in determining model outliers. The mean innermost changes by end of century are 0.04 m and 3.54 Sv compared to present day means of 0.06 m and 23 Sv, respectively.

For the biogeochemical changes in Figure 8 there are robust reductions in the innermost ensemble surface nutrients [NO_3, PO_4, Si] on the shelf (sub-region 4) and in sub-region 1 of around (−13.0, −12.0, −26.0)%

21699291, 2023, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022JC018880 by Ministry Of Health, Wiley Online Library on [23/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

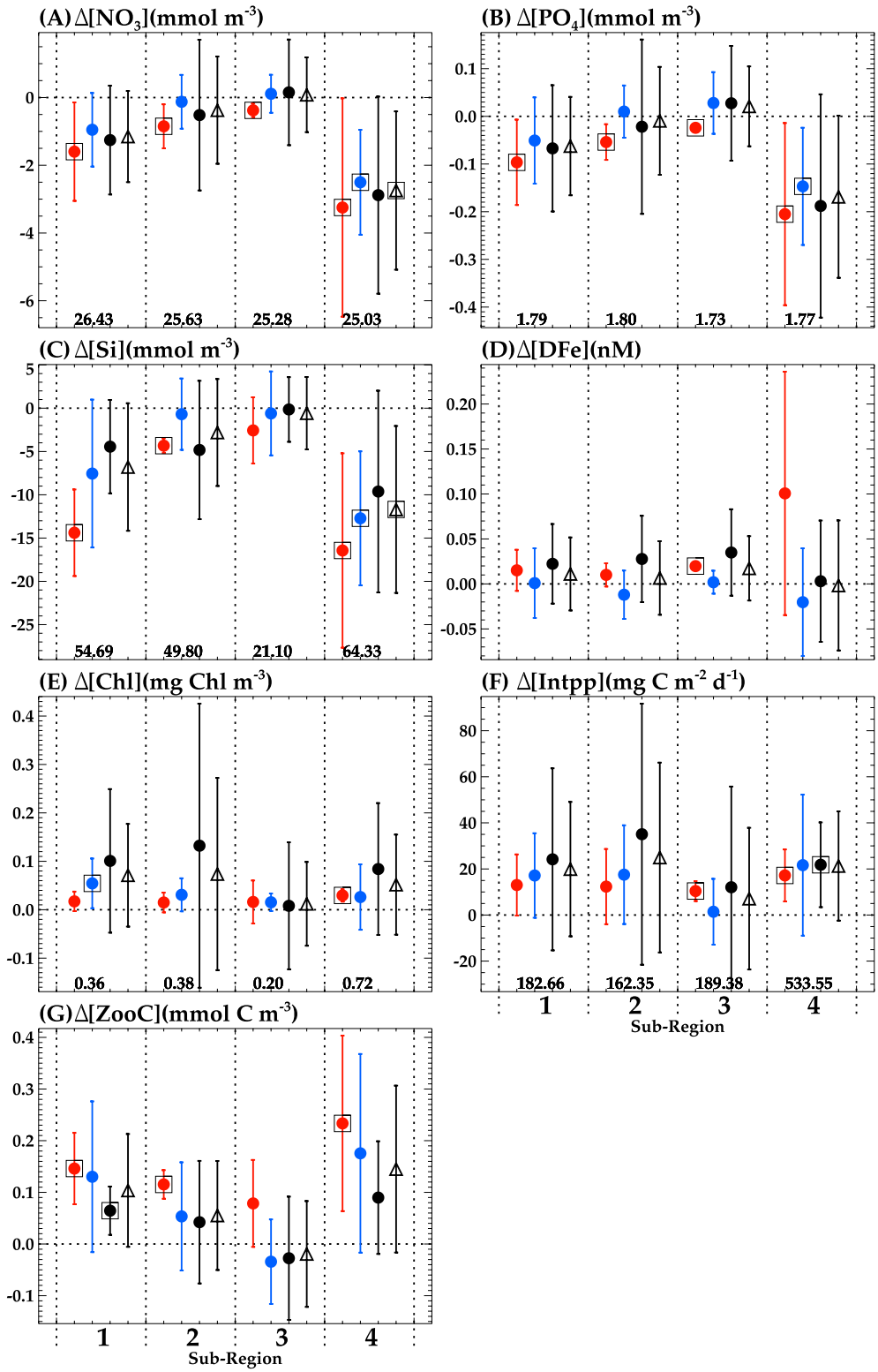

Figure 8. Ensemble changes for each sub-region labeled 1 to 4 for (a) NO_3 (mmol m^{-3}), (b) PO_4 (mmol m^{-3}), (c) Si (mmol m^{-3}), (d) DFe (nM), (e) Chl (mg Chl m^{-3}), (f) Intpp (mg C m^{-2} d^{-1}), and (g) ZooC (mmol C m^{-3}). Vertical dotted lines separate sub-regions, and horizontal dotted lines show zero change. Ensemble averages and spreads are the vertical lines in red (circle), blue (circle), and black (circle) for the innermost, inner, and outer models, respectively, and the black (triangle) for the full model ensemble. Present day mean observational values per variable per sub-region included at the bottom of each frame. Changes are for RCP8.5/SSP585 for period 2081 to 2100 relative to historical period 1976 to 2005. Black squares on ensemble averages indicate "robust" changes.

21699291, 2023, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022JC018880 by Ministry Of Health, Wiley Online Library on [23/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

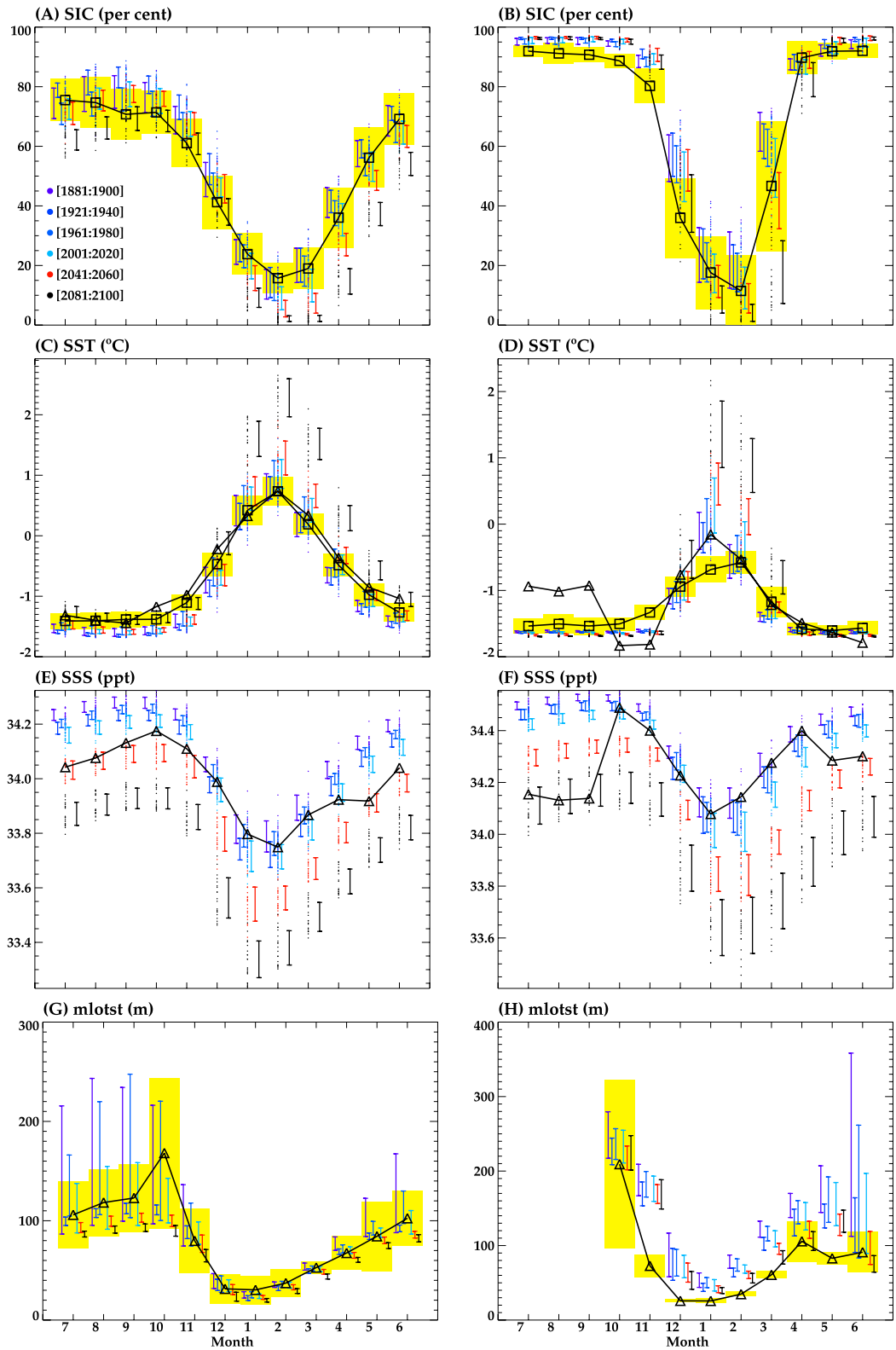

Figure 9.

21699291, 2023, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022JC018880 by Ministry Of Health, Wiley Online Library on [23/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

and (−6.0, −5.0, −26.0)%, respectively, relative to the annual mean observational values (in black at the bottom of each frame); the reductions in sub-region 2 are relatively smaller at (−3.0, −3.0, −9.0)%, and weaker (if at all) in sub-region 3. Robust increases are found for ZooC in frame (G) for sub-regions 1, 2, and 4 for the innermost ensemble (but we do not have observational estimates for ZooC to determine percentage changes). For Chl and Intpp the trend over the ensembles is for small increases in each sub-region, but these are only robust of magnitude +4.0% for Chl in sub-region 4, and (+5.5, +3.0)% in sub-regions 3 and 4, respectively, for Intpp. For DFe in frame (D) the only robust change across all the ensembles is for an increase in sub-region 3 for the innermost ensemble; otherwise there are no clear projected changes. Compared to the physical variables in Figure 7, the number of robust changes across all the ensembles is much reduced; for example, all the physical variables have at least 2 ensembles that agree on the sign of a robust change over all sub-regions, whereas this only happens for sub-region 4 for NO_3, PO_4, Si, and Intpp, and for ZooC in sub-region 1.

For each variable Figures 9–11 plot average monthly means and standard deviations for the innermost ensemble compared to present day observational estimates for sub-regions 1 (the left column in each Figure) and 4 (the right column). Where available observational standard deviations per month are shown as the yellow bands. The observational data stencil is used to calculate the observational and model values to account for any observational missing data so that, for example, mlotst in sub-region 4 in Figure 9h has no values for months 7–9 as there is no observational mlotst data available. The model ensemble monthly values are from a selection of 20 years windows over the period 1850 to 2100, in particular 1881 to 1900, 1921 to 1940, 1961 to 1980, 2001 to 2020, 2041 to 2060, and 2081 to 2100, and are color coded as shown in the legend in frame (A) in each figure. These figures provide comparisons of how well the innermost ensemble does in representing the observational data for present day conditions, and whether the projected values and their spreads for the future ranges 2041 to 2060 (red) and 2081 to 2100 (black) indeed exceed the model and data present day ranges, and whether there are robust seasonal as well as amplitude changes.

For SIC in Figure 9a the present day ensemble values match the data well, and the 2081 to 2100 projections show marked reductions for January to July. There are 2041–2060 reductions, but these still lie within the observational range. In frame (B) on the shelf the model ensemble overestimates June to November present day values, with 2081–2100 projections for SIC showing summer reductions but still within the observational range, except in March. For SST in Figure 9c the innermost ensemble seasonal cycle generally matches the WOA (Δ) and HadISST (□) data points well, but tends to be a little lower than the observations over the August to November period. For December to June there is a robust SST increase for 2081 to 2100, with clear 2041 to 2060 changes for February and March. On the shelf in frame (D) the SST ensemble and data agree well (except for the WOA winter values); and there are clear projection SST increases, especially in January and February. These SST increases are reflected in the robust annual mean changes shown in Figure 7b.

For SSS in Figure 9e there is a good comparison with data for December to April, but typically the ensemble is too salty otherwise. There is a clear freshening for both 2041 to 2060 and 2081 to 2100 for all months. In frame (F) the SSS ensemble and data agree well apart from the winter July to September period, and the freshening for all months is again seen. For mlotst in frame (G) the ensemble lies within the observational range over the whole year, and remains so even for the projection periods. Note the large present day model spreads for June to October which might be associated with false model convection that then actually appears to cease into the future. On the shelf in frame (H) the ensemble follows the seasonal data cycle well, but typically overestimates the depths. There are 2041–2060 and 2081 to 2100 mlotst reductions, but these are relatively modest. The large spread is apparent here in June, but again decays with time.

For NO_3 in Figure 10a the ensemble underestimates the observational values over most of the year, but particularly from December to April; the ensemble shows a clear reduction into the future over all months. On the shelf in frame (B) the ensemble seasonal cycle follows the data but is again relatively low in summer—and there is a suggestion of

Figure 9. Innermost ensemble monthly means and standard deviations (vertical colored bars) compared to observational means (Δ or □) and observational standard deviations (the yellow bands where available) for sub-regions 1 (the left column) and 4 (the right column). Missing observational data is accounted for in the model values by integrating both over the observational data stencil. Frames (A) and (B) for SIC, (C) and (D) for SST, (E) and (F) for SSS, and (G) and (H) for mlotst. In frames (A) to (D) the □ observational means are from HadISST, and for (C) to (F) the Δ are from World Ocean Atlas 2009. Progressing from left to right the model values for each month are for the periods 1881 to 1900, 1921 to 1940, and 1961 to 1980 to the left of each Δ, and 2001 to 2020, 2041 to 2060, and 2081 to 2100 to the right, and see legend in frame (A).

21699291, 2023, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022JC018880 by Ministry Of Health, Wiley Online Library on [23/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

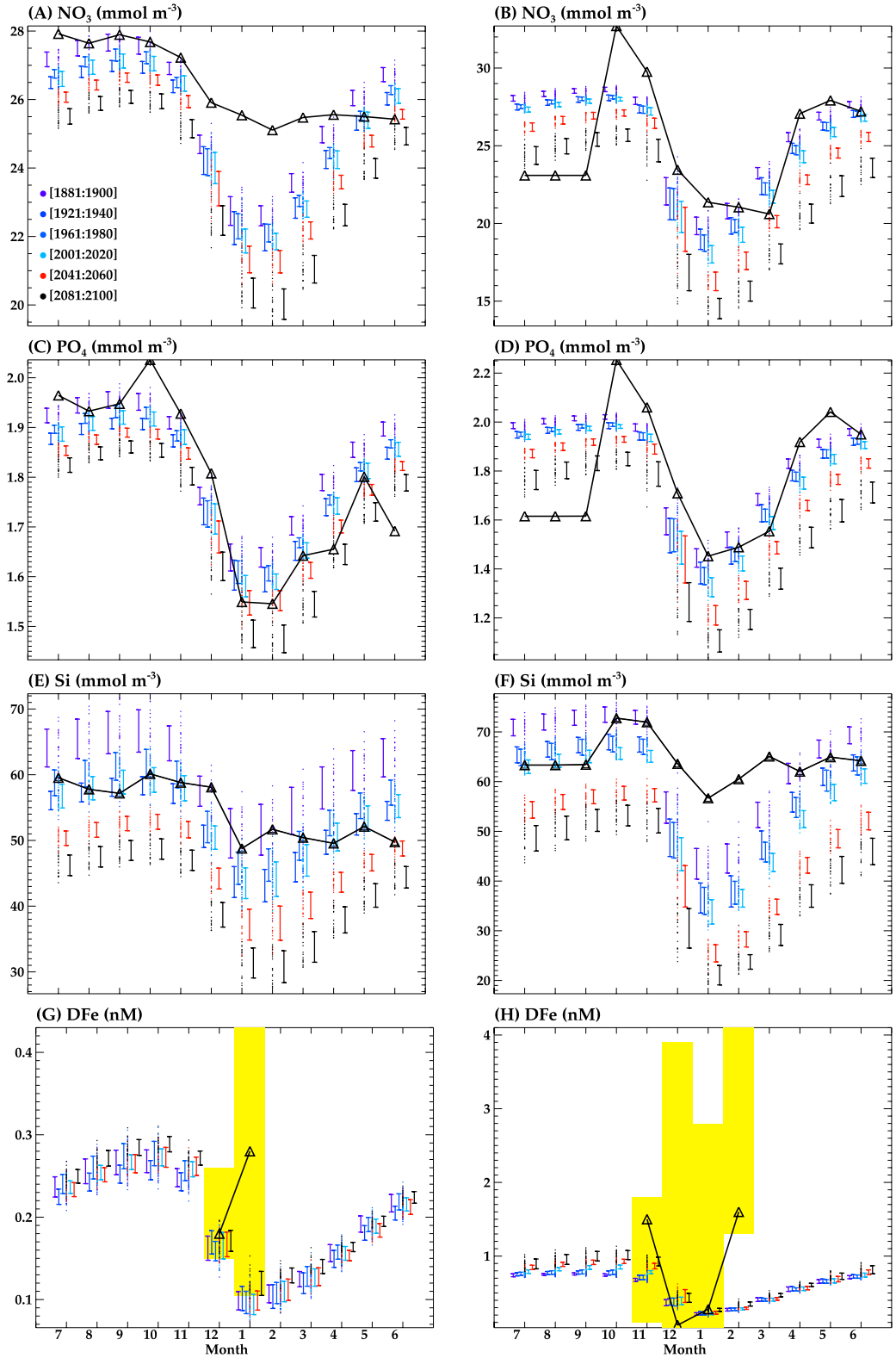

Figure 10.

21699291, 2023, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022JC018880 by Ministry Of Health, Wiley Online Library on [23/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

low data values over winter—and the reductions into the future mirror those in frame (A). For PO_4 in frames (C) and (D), the ensemble and data values actually agree well over the full season (except again in winter on the shelf), and the projection reductions follow those for NO_3. In frame (E) for Si in sub-region 1, the results mimic the equivalent PO_4 results, with again significant projection reductions into the future. On the shelf in frame (F) the results are similar to NO_3, with low values compared to the observations in summer, but again significant reductions into the future.

For surface iron in Figure 10g there is a clear ensemble seasonal cycle, and these span the lower range of the observations in December and January. Note here that the DFe data is not a climatology but rather a collection of all available individual observations, so that the yellow band ranges from minimum to maximum observed values, and the Δ is the observed median. There is a small but not significant increase over the seasonal cycle for the projections. On the shelf in frame (H) the ensemble follows the observations over the summer, but again at the lower end of the data values. As in frame (G) there are again small increases in the projections, but not significant.

For surface chlorophyll in Figure 11a the ensemble matches the observed seasonal cycle well but typically on the high side, especially from September to the peak in December. The projections suggest a slight but not significant increase in spring for 2041 to 2060 and 2081 to 2100. In frame (B) on the shelf, the results are similar to those in frame (A), except that the ensemble December peak is now much larger than suggested by the observations, and again the projections show no clear future trends. For Intpp in frames (C) and (D), the ensemble follows the observed seasonal cycle but tending to be on the high side for September to December off-shelf, and for December on-shelf. In frame (C) there is not really a projection trend, while on-shelf there is an increasing projected trend found but only for January to March relative to present day values.

Finally, for surface zooplankton concentration ZooC in Figures 11e and 11f we first note that the model and data values represent different ways of quantifying zooplankton, since the data values are obtained from abundance estimates (typically counts per segment length) to be contrasted with the model values of concentration of zooplankton carbon. To that end the plots show each normalized to their maximum value in order to provide some kind of seasonal cycle comparison. In frames (E) and (F) the ensemble shows a clear seasonal cycle that peaks in January, typically a month later than the equivalent peak in chlorophyll in frames (A) and (B). In comparison the abundance observations suggest peaks in November and December off- and on-shelf, respectively, followed by drops, with subsequent increases again in February and March. Consistent with the annual average changes shown in Figure 8g ZooC tends to increase compared to present day for the 2081 to 2100 averages, but these are only robust for October and November, and February and March, in Figures 11e and 11f, respectively.

5. Discussion and Conclusion

By including 16 CMIP5 and 16 CMIP6 ESM in our study it is hoped that future assessments for the SO following Rogers et al. (2020) and their Tables 1 and 2, for example, will be more confident on the likely physical and biogeochemical climate change impacts on Antarctic marine communities. This study reinforces and broadens the outcomes found by Rickard and Behrens (2016) in terms of RCP8.5/SSP585 projections by the increased model suite and the significantly extended spatial range to encompass the Ross Sea continental shelf as well as the wider Ross Sea encompassing the Ross Sea Region MPA. In terms of model ranking we now include consideration of ECS and TCR, as well as model false deep convection, in order to obtain our ("best") innermost ensemble, and the latter has been used to examine how well this ensemble monthly means represent the observations, and to look at projected seasonal cycles. Our ranking process using our specific metrics suggests that the physical representation (measured in terms of [SIC, SST, SSS, ZOS, mlotst]) of the CMIP6 models has improved on that from CMIP5—likely due in part to improved spatial resolution over the Ross Sea Region compared to CMIP5—but that there is not an equivalent improvement in the biogeochemical components. Physical gains have also been concluded by Beadling et al. (2020) with CMIP6 improvements compared to previous CMIP generations being found for the simulated ACC strength, all surface wind stress metrics, and density differences across the ACC latitudes. Conversely, however, a much higher proportion of our CMIP6 models exceed the ECS and TCR thresholds compared to CMIP5, and the issue of unphysical deep ocean convection persists from CMIP5 to CMIP6.

Figure 10. Innermost ensemble monthly means and standard deviations (vertical colored bars) compared to observational means (Δ) and observational standard deviations (the yellow bands where available) for sub-regions 1 (the left column) and 4 (the right column). Missing observational data is accounted for in the model values by integrating both over the observational data stencil. Frames (A) and (B) for NO_3, (C) and (D) for PO_4, (E) and (F) for Si, and (G) and (H) for DFe. Progressing from left to right the model values for each month are for the periods 1881 to 1900, 1921 to 1940, and 1961 to 1980 to the left of each Δ, and 2001 to 2020, 2041 to 2060, and 2081 to 2100 to the right, and see legend in frame (A).

21699291, 2023, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022JC018880 by Ministry Of Health, Wiley Online Library on [23/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

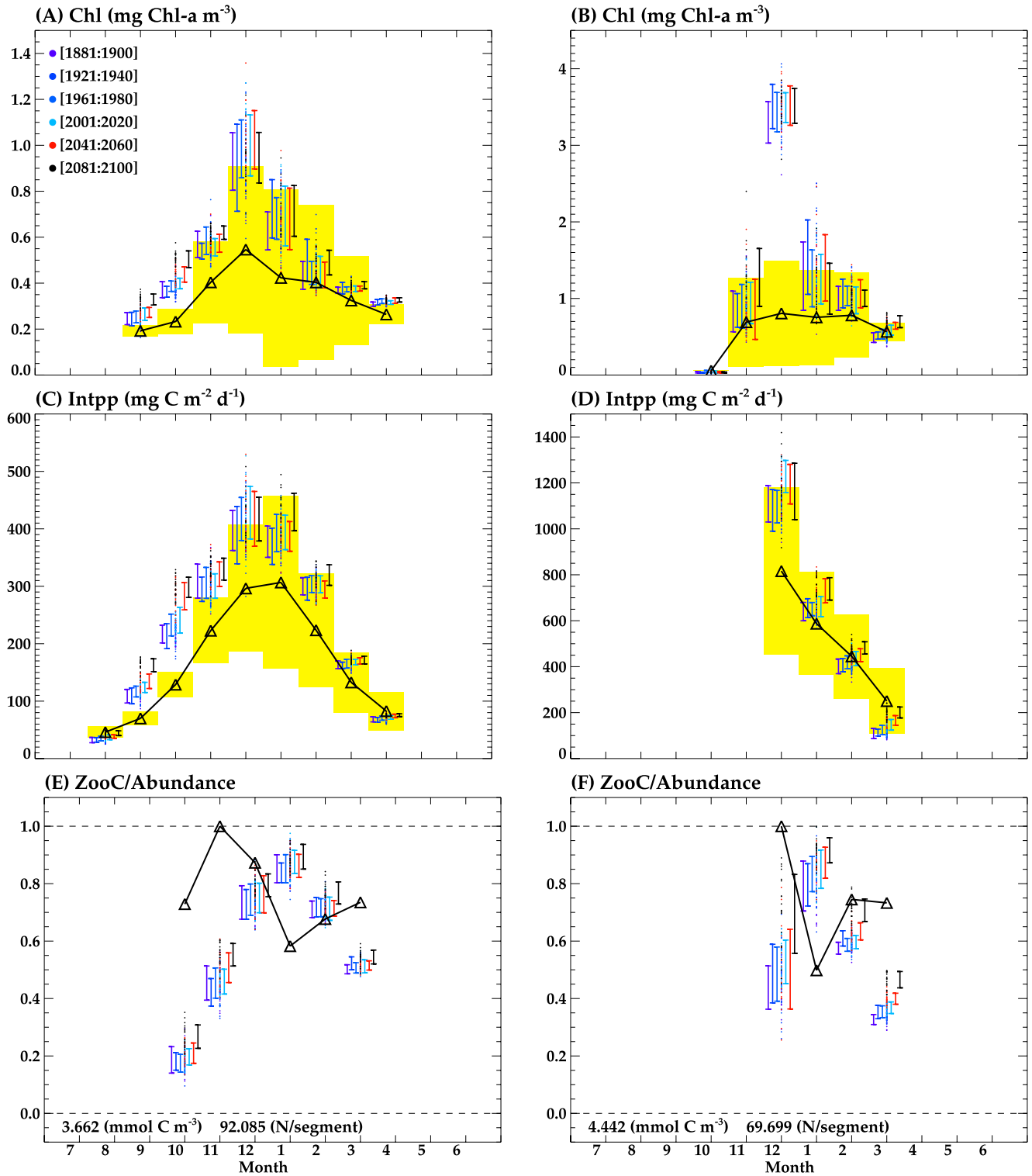

Figure 11. Innermost ensemble monthly means and standard deviations (vertical colored bars) compared to observational means (Δ) and observational standard deviations (the yellow bands where available) for sub-regions 1 (the left column) and 4 (the right column). Missing observational data is accounted for in the model values by integrating both over the observational data stencil. Frames (A) and (B) for surface chlorophyll (Chl), (C) and (D) for Intpp, and (E) and (F) for model surface zooplankton concentration (ZooC) and abundance data. Note that in frames (E) and (F) that normalized values have been plotted, with the normalizing values shown at the foot of each frame. Progressing from left to right the model values for each month are for the periods 1881 to 1900, 1921 to 1940, and 1961 to 1980 to the left of each Δ, and 2001 to 2020, 2041 to 2060, and 2081 to 2100 to the right, and see legend in frame (A).

21699291, 2023, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022JC018880 by Ministry Of Health, Wiley Online Library on [23/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Further, Mohrmann et al. (2021) find in their sample of CMIP6 models that those with an accurate ACC transport and wind stress curl tend to generate too many open water polynyas. Nevertheless, our Figure 3 shows that where there is information on ACC transport and deep convection, that 7 of our 16 CMIP6 models have both an adequate ACC and relative lack of deep convection compared to only 2 out of 10 for the CMIP5 models.

By end of century under RCP8.5/SSP585 our innermost ensemble predicts:

- Projected spin-up of the inner RG, with the mean strength and transport predicted to increase by 0.04 m and 3.54 Sv relative to present day values of 0.06 m and 23 Sv, respectively;
- Robust reductions in annual mean sea ice concentration, sea surface salinity, and mixed layer depth, over our four sub-regions, and sea surface temperature increases;
- For surface silicate there are robust declines of −26.0% in sub-regions 1 and 4, and around −9.0% in sub-region 2. In comparison both surface nitrate and phosphate have robust falls of around −12.0, −5.0, and −3.0% for sub-regions 4, 1, and 2, respectively. In contrast, surface zooplankton has robust increases in sub-regions 1, 2, and 4.
- Surface chlorophyll and NPP have small but not significant increases over all sub-regions, but generally robust but relatively small increases of the order of 3% over the continental shelf in sub-region 4 compared to present day values.

This pattern of surface changes—increased stratification, reduced nutrients—might be expected to result in reduced surface chlorophyll and NPP. However, at these high latitudes, the significant projected loss of seasonal sea ice cover allows for increased light during the growing season, and this factor would seem to be responsible for the increases in surface chlorophyll, zooplankton, and NPP shown in Figure 8. The relatively small percentage changes in chlorophyll and NPP (≈3%) presumably reflects a balance between the increased light and any limitation impacts from the reduced nutrients, although the surface dissolved iron shows small but not significant increases perhaps due to sources from melting sea ice (but of course dependent on the respective model parameterization). These ecosystem changes are further consistent with those inferred by Smith Jr et al. (2014) in a physical downscaling experiment using ECHAM5 (no biogeochemistry), a predecessor of the MPI CMIP5 and CMIP6 models analyzed here. We have documented the total changes in chlorophyll and zooplankton carbon using, for example, our innermost models, but changes in the seasonal succession of the functional groups would have implications for the Ross Sea ecosystem. To that end analyzing our innermost models for ecosystem changes would be a next useful step, even though they each have different levels of biological complexity (see Table 1 with model 10 (1P,1Z), 18 (3P,1Z) and 16 (3P,3Z)).

Our innermost ensemble comprises models MRI-ESM2-0 (10), GFDL-ESM4 (16), and CESM1BGC (18) (see Table 1), based on our multi-variable assessment (see in particular the summary in Figure 3). In terms of model convergence, a trend toward the higher resolution and more biogeochemically complex models being the "best" would provide confidence on the next developments for future CMIP work; unfortunately—at least for the Ross Sea Region and based on our metric choices—this has not yet been realised (although this is only the second time that biogeochemistry has been a full part of CMIP analysis), with these three models spanning both physical resolution and biogeochemical complexity (indeed, MRI-ESM2-0 is a NPZD model and does not explicitly cycle silicate or iron). Figure 3 shows that a lot of the models have been relatively relegated because of being outliers for ECS (E) and TCR (T), a trend actually worse at CMIP6, and further for false convection events in the Ross Sea region, the latter equally common across CMIP5 and CMIP6; to that end it would seem that efforts to reduce and/or eliminate these shortcomings in future models would go a long way to improving the overall Ross Sea region representation.

The spin up of the inner RG under SSP585 is likely due to projected deepening of the Amundsen Sea Low (ASL) in a warming world. Dotto et al. (2018) found that the size and intensity of the RG is linked to ASL forcing, and that future ASL deepening would likely further increase the gyre size and intensity. Further, Gao et al. (2021) show for a sample of CMIP6 models—including 8 models studied here, namely 1, 4, 5, 8, 10, 12, 13, and 16—that robust and significant deepening of the ASL central pressure is found for their ensemble over a range of future scenarios and warming (see in particular Figure 7 in Gao et al. (2021)). This is consistent with our results that show, of the 8 models analyzed here and by Gao et al. (2021), that 8 of them have increased strength and 6 increased barotropic transport of the gyre by end of century compared to present day (see Figure 7 frames (E) and (F)). Of the 32 models studied here the majority show both increased gyre strength and transport, and further

21699291, 2023, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022JC018880 by Ministry Of Health, Wiley Online Library on [23/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

analysis of the future changes to the ASL in these models could enhance understanding of this connection, and its consequences for the Ross Sea.

The remote and harsh environment of the SO, coupled with significant cloud and sea ice coverage, means that it remains relatively undersampled (both in-situ and by satellite), hindering biogeochemical model development for the region. Furthermore, the SO phytoplankton community poses specific challenges in terms of modeling, especially the representation of SO species and their carbon cycling impact (see, e.g., Nissen and Vogt (2021) and references therein), and this also complicates satellite retrievals for chlorophyll and models for estimating primary production. For example, Ferreira et al. (2022) note a range of some 18%–67% shortfall in chlorophyll derived from satellite measurements (see, e.g., Jena (2017)), but do note that new algorithms can do better in specific regions (e.g., Moutier et al. (2019) and Haëntjens et al. (2017)), and indeed Ferreira et al. (2022) introduce a new algorithm specific to the Western Antarctic Peninsula which provides for better local chlorophyll estimates. In terms of NPP, Carr et al. (2006) in a comparison of 24 NPP models—including VGPM—find that the "Equatorial Pacific and the Southern Oceans presented the worst results," and the "Southern Ocean data included both the lowest and highest values of primary production." Carr et al. (2006) find that VGPM generally overestimates primary production in the SO, and this is consistent with the recent study of Pinkerton et al. (2021) that finds both VGPM and CbPM look to be overestimating primary production in northern zone waters around Antarctica, but underestimating in zones closer to Antarctica (see Pinkerton et al. (2021) Supplementary information Figure 1 showing these zones).

To sample the Ross Sea Region using the SSH patterns for the gyres and the ACC, as well as the bathymetry for the Ross Sea continental shelf, our analysis used the sub-regions 1 to 4 shown in Figure 1b. However, Figure 1b shows that surface chlorophyll along the 1,000 m isobath in the south of sub-region 2 is substantially larger than the remainder of this sub-region. How sensitive might our results be to using sub-region 2 to cover both the inner RG and the waters closer to the continental shelf for surface chlorophyll? To test this sub-region 2 has been sub-divided into two new areas, one effectively covering the high chlorophyll region in Figure 1b, and a second covering the remaining waters spanned by the inner RG and within the upper boundary of sub-region 2; the dotted white line in sub-region 2 in Figure 1b shows the boundary between these new areas. We now repeat our surface chlorophyll analysis over these two new sub-regions, average the ranking scores, and compare the result with the original sub-region 2 outcomes. In each case the outlier models shown in Figure 3 are the same, that is, models 8, 9, 30, 31, so that our ensemble choices would not be changed. The "Dist" statistics between the reference and modified sub-region 2 rankings are (RMSD, nDist2) of (4.4,20), comparable to the sensitivity changes shown in Figure 6, and further 15 of the best ranked models for sub-region 2 remain so using the modified sub-regions. Overall this would make little difference to our ensemble choices. Does this mean that our sub-region choices are optimal? As Jolliff et al. (2009) highlight in their discussion with regard to their Figure 15, pattern and bias mis-matches can be imperfect metrics in their own right, with skill assessment ultimately dependent on researcher value judgments. In this regard we note that our skill is effectively summarized in Figure 3 which represents identification of outliers across multiple variables, and the various sensitivity tests discussed suggest we have not unfairly penalized outlier models.

The lower latitude of 76°S for our analysis (see Figure 1) has been chosen to accommodate the maximum number of the CMIP5 and CMIP6 models given their respective native resolutions and southern boundary latitudes. However, there is a relatively large region of continental shelf water between 76°S and the front of the Ross Ice Shelf and, importantly, Figures 13, 14, and 16 in Arrigo and Dijken (2004) show that significant chlorophyll production occurs over this region. Indeed, using our surface chlorophyll and NPP climatologies, we find—averaged over the months December, January, and February—that the continental shelf region between 76°S and depths shallower than 1,000 m accounts for 33% and 42% of the total surface chlorophyll and NPP calculated using the front of the Ross Ice Shelf as the southern boundary. Unlike our discussion for the choice of sub-region 2, a sensitivity analysis of model ranking for this unaccounted volume of water is complicated by how the different models locate and treat their southernmost boundary, as none of the present CMIP models directly model the Ross Ice Shelf. Nevertheless, an extension to our present study would be to rank the sub-set of models that have a southern boundary close to that of the Ross Ice Shelf and see how that reflects on our present rankings, and complementing the polynya study of Mohrmann et al. (2021) for CMIP6 models by focusing on processes close to the location of the Ross Ice Shelf.

While a large part of the effort here is toward the biogeochemistry, the physical system obviously cannot be ignored, especially any potential impacts of the Ross Sea region Balleny and inner gyres that separate the Ross Sea continental shelf from the ACC. For example, recent work by Behrens et al. (2021) shows a relationship between

21699291, 2023, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022JC018880 by Ministry Of Health, Wiley Online Library on [23/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

drifting larvae, the Ross Sea gyres, and sea ice, suggesting that future physical changes—such as increased gyre transport and reduced sea ice concentration—could have implications for ecosystem response. Further Dinniman et al. (2020) (and references therein) show how important iron sources for phytoplankton growth and circulation changes may be linked around Antarctica, with clear implications for understanding impacts from such projected changes.

Our focus has been almost solely on surface fields, but impact below the surface needs to be addressed as well. For the SO the growing number of Biogeochemical Argo floats deployed by the SO Carbon and Climate Observations and Modeling (SOCCOM) project enables significantly more profiles to be obtained, as well as extending coverage into the sparsely observed Southern Hemisphere winter. While using the SOCCOM data profiles for intercomparison with climatological models remains a challenge (due to the differing spatial and temporal characteristics of the models and the profiles), we nevertheless need to extend our work to look at the sub-surface impacts predicted by the models. We have also only indirectly assessed on-shelf dense water formation by consideration of deep water convection and bottom water properties off-shelf. Heuzé et al. (2013) shows that no CMIP5 model they analyzed correctly represents dense shelf overflows, and in Heuzé (2021) the CMIP6 models NorESM2-LM (13) and NorESM2-MM (14) show overflowing, models IPSL-CM6A-LR (5), UKESM1-0-LL (7), and GFDL-ESM4 (16) might (but requires more analysis to be sure), models CanESM5 (1), CanESM5-CanOE (2), and ACCESS-ESM1-5 (4) show an occasional plume, and of course models CESM2 (11) and CESM2-WACCM (12) have a parameterized "pipe" for dense water. This means that the models do indeed generate dense water on the shelf, but its properties will need improvement to consistently connect the shelf to the deep basins.

Presenting projections based on RCP8.5/SSP585 provides an example of use of our ensemble choices for our chosen Ross Sea region. However, it is well known that both projected model variability within a given scenario, and variability between scenarios, result in uncertainty in future projections. As an example, Wang et al. (2022) find that about 50% of the uncertainty in projected SO warming for CMIP6 models with SSP585 is associated with inter-model El Niño-Southern Oscillation variations, with factors such as model climate sensitivity, high latitude westerlies, and mesoscale eddy parameterization also contributing (see Wang et al. (2022) and references therein). In terms of variation between scenarios, Frölicher et al. (2016), for example, show that model and scenario uncertainty dominate end of century projections for potential ocean ecosystem stressors (in their case surface pH, SST, O_2, and NPP). Further, a recent study on projection downscaling for the California Current System by Pozo Buil et al. (2021) finds that (for their limited 3 model ensemble) scenario uncertainty is significant for SST but that it is model uncertainty for NPP (see their Figure S1).

Even if our list of metrics in Figure 3 was "complete"—which it clearly is not—this still would not guarantee that the "best" models would most accurately project into the future. As highlighted by Hall et al. (2019) (see their Figure 1) our "conventional evaluation" combines multiple metrics to quantify the ESM but only relates the metrics to the projections in an "ambiguous" way. Hall et al. (2019) contrast this with the emergent constraint approach that seeks to identify metrics relevant to the projections, but that typically only a single metric is used. In the context of the study presented here, we note that Hall et al. (2019) emphasize that emergent constraint methods are "complementary to traditional model evaluation, which the scientific community must keep doing." Could we have used an ecosystem emergent constraint method for our region? Presently, the answer would appear to be no, as according to Table 1 in Hall et al. (2019) only 1 out of 32 identified emergent constraints related to ocean biogeochemistry and that was associated with co-variation of tropical SST and NPP anomalies (Kwiatkowski et al. (2017)).

Specific Antarctic processes are not yet explicitly represented in the CMIP models, in particular ice shelves and the circulation beneath them and consequently feed backs on melting, and regional oceanic properties (especially dense water formation in polynyas and its off-shelf transport). Future CMIP developments and computational improvements may allow their inclusion, but for the time being a way to bridge the computational and resolution gap is via local grid refinement in global models (see, e.g., Behrens et al. (2020) and Behrens et al. (2021)) or downscaling. Downscaling of future states for the Ross Sea continental shelf has shown similar changes, driven by projected output from a single model (e.g., Smith Jr et al. (2014) and Dinniman et al. (2018)), and it is hoped that our analysis provides for improved confidence in the selection of physical and biogeochemical states to drive the regional models. Further, for those with access to sufficient computational resources, our work shows that the publicly available CMIP output made possible by the World Climate Research Programme's Working Group on Coupled Modeling is readily accessible, and represents a valuable source of forcing fields for future projections.

21699291, 2023, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022JC018880 by Ministry Of Health, Wiley Online Library on [23/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Data Availability Statement

CMIP5 and CMIP6 model fields used in this paper can be freely downloaded from https://doi.org/10.5281/zenodo.7145435, including the fields used to plot Figure 1.

Acknowledgments
We are grateful to the editors and reviewers for their constructive and insightful critique of our manuscript. This work was supported by NZ Ministry of Business, Innovation and Employment (MBIE) NIWA Strategic Investment Fund, MBIE C01X1710 Ross-RAMP, and the NZ Antarctic Science Platform (MBIE ANTA1801). We also acknowledge the use of New Zealand eScience Infrastructure (NeSI) high performance computing facilities; New Zealand's national facilities are provided by NeSI and funded jointly by NeSI's collaborator institutions and through the MBIE Research Infrastructure programme (https://www.nesi.org.nz). Thanks also to the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and further thank the climate modelling groups (whose models are in Table 1 of this paper) for producing and making their model output available. WOA2009 data accessed via https://www.nodc.noaa.gov/OC5/WOA09/pr_woa09.html, and MLD via https://cerweb.ifremer.fr/deboyer/mld/Surface_Mixed_Layer_Depth.php. We are grateful to the Ocean Biology Processing Group for the SeaWiFS data (NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing Group; (2014): Sea-viewing Wide Field-of-view Sensor (SeaWiFS) Ocean Color Data, NASA OB.DAAC (https://doi.org/10.5067/ORBVIEW-2/SEAWIFS_OC.2014.0). Also thanks to Cliff Law for feedback on the draft manuscript. Open access publishing facilitated by National Institute of Water and Atmospheric Research, as part of the Wiley - National Institute of Water and Atmospheric Research agreement via the Council of Australian University Librarians.

References

Adachi, Y., Yukimoto, S., Deushi, M., Obata, A., Nakano, H., Tanaka, T. Y., et al. (2013). Basic performance of a new Earth system model of the meteorological research institute (MRI-ESM1). *Papers in Meteorology and Geophysics*, *64*(0), 1–19. https://doi.org/10.2467/mripapers.64.1

Antonov, J., Seidov, D., Boyer, T., Locarnini, R., Mishonov, A., Garcia, H., et al. (2010). World ocean atlas 2009. In S. Levitus (Ed.), *Salinity. NOAA atlas NESDIS 69. Unpublished doctoral dissertation* (Vol. 2). US Government Printing Office.

Armitage, T. W., Kwok, R., Thompson, A. F., & Cunningham, G.(2018). Dynamic topography and sea level anomalies of the southern ocean: Variability and teleconnections. *Journal of Geophysical Research: Oceans*, *123*(1), 613–630. https://doi.org/10.1002/2017jc013534

Arrigo, K. R., & Dijken, G. L. (2004). Annual changes in sea-ice, chlorophyll a, and primary production in the Ross Sea, Antarctica. *Deep Sea Research Part II: Topical Studies in Oceanography*, *51*(1–3), 117–138. https://doi.org/10.1016/j.dsr2.2003.04.003

Aumont, O., & Bopp, L. (2006). Globalizing results from ocean in situ iron fertilization studies. *Global Biogeochemical Cycles*, *20*(2), GB2017. https://doi.org/10.1029/2005gb002591

Aumont, O., Éthé, C., Tagliabue, A., Bopp, L., & Gehlen, M. (2015). Pisces-v2: An ocean biogeochemical model for carbon and ecosystem studies. *Geoscientific Model Development*, *8*(8), 2465–2513. https://doi.org/10.5194/gmd-8-2465-2015

Bacmeister, J. T., Hannay, C., Medeiros, B., Gettelman, A., Neale, R., Fredriksen, H.-B., et al. (2020). CO_2 increase experiments using the CESM: Relationship to climate sensitivity and comparison of CESM1 to CESM2. *Journal of Advances in Modeling Earth Systems*, *12*(11), e2020MS002120. https://doi.org/10.1029/2020ms002120

Beadling, R., Russell, J., Stouffer, R., Goodman, P., & Mazloff, M.(2019). Assessing the quality of southern ocean circulation in CMIP5 AOGCM and Earth system model simulations. *Journal of Climate*, *32*(18), 5915–5940. https://doi.org/10.1175/jcli-d-19-0263.1

Beadling, R., Russell, J., Stouffer, R., Mazloff, M., Talley, L., Goodman, P., et al.(2020). Representation of southern ocean properties across coupled model intercomparison project generations: CMIP3 TO CMIP6. *Journal of Climate*, *33*(15), 6555–6581. https://doi.org/10.1175/jcli-d-19-0970.1

Behrenfeld, M. J., & Falkowski, P. G.(1997). Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnology & Oceanography*, *42*(1), 1–20. https://doi.org/10.4319/lo.1997.42.1.0001

Behrens, E., Pinkerton, M., Parker, S., Rickard, G., & Collins, C. (2021). The impact of sea-ice drift and ocean circulation on dispersal of toothfish eggs and juveniles in the Ross gyre and Amundsen Sea. *Journal of Geophysical Research: Oceans*, *126*(10), e2021JC017329. https://doi.org/10.1029/2021jc017329

Behrens, E., Rickard, G., Morgenstern, O., Martin, T., Osprey, A., & Joshi, M. (2016). Southern ocean deep convection in global climate models: A driver for variability of subpolar gyres and Drake Passage transport on decadal timescales. *Journal of Geophysical Research: Oceans*, *121*(6), 3905–3925. https://doi.org/10.1002/2015jc011286

Behrens, E., Williams, J., Morgenstern, O., Sutton, P., Rickard, G., & Williams, M. J. (2020). Local grid refinement in New Zealand's Earth system model: Tasman Sea ocean circulation improvements and super-gyre circulation implications. *Journal of Advances in Modeling Earth Systems*, *12*(7), e2019MS001996. https://doi.org/10.1029/2019ms001996

Bopp, L., Resplandy, L., Orr, J. C., Doney, S. C., Dunne, J. P., Gehlen, M., et al. (2013). Multiple stressors of ocean ecosystems in the 21st century: Projections with CMIP5 models. *Biogeosciences*, *10*(10), 6225–6245. https://doi.org/10.5194/bg-10-6225-2013

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., et al. (2020). Presentation and evaluation of the IPSL-CM6A-LR climate model. *Journal of Advances in Modeling Earth Systems*, *12*(7), e2019MS002010. https://doi.org/10.1029/2019MS002010

Boyer Montégut, C., Madec, G., Fischer, A. S., Lazar, A., & Iudicone, D. (2004). Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology. *Journal of Geophysical Research*, *109*(C12), C12003. https://doi.org/10.1029/2004jc002378

Briegleb, B., Danabasoglu, G., & Large, W. (2010). *An overflow parameterization for the ocean component of the community climate system model*. National Center for Atmospheric Research Tech. Note NCAR/TN-4811STR.

Brunner, L., Pendergrass, A. G., Lehner, F., Merrifield, A. L., Lorenz, R., & Knutti, R. (2020). Reduced global warming from CMIP6 projections when weighting models by performance and independence. *Earth System Dynamics*, *11*(4), 995–1012. https://doi.org/10.5194/esd-11-995-2020

Carr, M.-E., Friedrichs, M. A., Schmeltz, M., Aita, M. N., Antoine, D., Arrigo, K. R., et al. (2006). A comparison of global estimates of marine primary production from ocean color. *Deep Sea Research Part II: Topical Studies in Oceanography*, *53*(5–7), 741–770. https://doi.org/10.1016/j.dsr2.2006.01.028

CCAMLR. (2016). *Conservation measure 91-05*. Commission for the Conservation of Antarctic Marine Living Resources.

Christian, J., Arora, V., Boer, G., Curry, C., Zahariev, K., Denman, K., et al. (2010). The global carbon cycle in the Canadian Earth system model (CANESM1): Preindustrial control simulation. *Journal of Geophysical Research*, *115*(G3), G03014. https://doi.org/10.1029/2008jg000920

Constable, A. J., Melbourne-Thomas, J., Corney, S. P., Arrigo, K. R., Barbraud, C., Barnes, D. K., et al.(2014). Climate change and Southern Ocean ecosystems i: How changes in physical habitats directly affect marine biota. *Global Change Biology*, *20*(10), 3004–3025. https://doi.org/10.1111/gcb.12623

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D., DuVivier, A., Edwards, J., et al. (2020). The community Earth system model version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, *12*(2), e2019MS001916. https://doi.org/10.1029/2019ms001916

De Lavergne, C., Palter, J. B., Galbraith, E. D., Bernardello, R., & Marinov, I. (2014). Cessation of deep convection in the open southern ocean under anthropogenic climate change. *Nature Climate Change*, *4*(4), 278–282. https://doi.org/10.1038/nclimate2132

Dinniman, M. S., Klinck, J. M., Hofmann, E. E., & Smith, W. O.(2018). Effects of projected changes in wind, atmospheric temperature, and freshwater inflow on the Ross Sea. *Journal of Climate*, *31*(4), 1619–1635. https://doi.org/10.1175/jcli-d-17-0351.1

Dinniman, M. S., St-Laurent, P., Arrigo, K. R., Hofmann, E. E., & Dijken, G. L. (2020). Analysis of iron sources in Antarctic continental shelf waters. *Journal of Geophysical Research: Oceans*, *125*(5), e2019JC015736. https://doi.org/10.1029/2019jc015736

Donohue, K., Tracey, K., Watts, D., Chidichimo, M. P., & Chereskin, T. (2016). Mean Antarctic circumpolar current transport measured in Drake Passage. *Geophysical Research Letters*, *43*(22), 11–760. https://doi.org/10.1002/2016gl070319

Dotto, T. S., Naveira Garabato, A., Bacon, S., Tsamados, M., Holland, P. R., Hooley, J., et al. (2018). Variability of the Ross gyre, southern ocean: Drivers and responses revealed by satellite altimetry. *Geophysical Research Letters*, *45*(12), 6195–6204. https://doi.org/10.1029/2018gl078607

Dunn, A., Vacchi, M., & Watters, G. (2017). The Ross sea region marine protected area research and monitoring plan. In *CCAMLR document SCCAMLR-XXXVI/20*. Hobart.

21699291, 2023, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022JC018880 by Ministry Of Health, Wiley Online Library on [23/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Dunne, J. P., Bociu, I., Bronselaer, B., Guo, H., John, J., Krasting, J., et al.(2020). Simple global ocean biogeochemistry with light, iron, nutrients and gas version 2 (BLINGV2): Model description and simulation characteristics in GFDL's CM4. 0. *Journal of Advances in Modeling Earth Systems*, *12*(10), e2019MS002008. https://doi.org/10.1029/2019ms002008

Dunne, J. P., John, J. G., Shevliakova, E., Stouffer, R. J., Krasting, J. P., Malyshev, S. L., et al. (2013). GFDL's ESM2 global coupled climate–carbon Earth system models. Part II: Carbon system formulation and baseline simulation characteristics. *Journal of Climate*, *26*(7), 2247–2267. https://doi.org/10.1175/jcli-d-12-00150.1

Ferreira, A., Brito, A. C., Mendes, C. R., Brotas, V., Costa, R. R., Guerreiro, C. V., et al.(2022). Oc4-so: A new chlorophyll-a algorithm for the Western Antarctic peninsula using multi-sensor satellite data. *Remote Sensing*, *14*(5), 1052. https://doi.org/10.3390/rs14051052

Forster, P. M., Andrews, T., Good, P., Gregory, J. M., Jackson, L. S., & Zelinka, M. (2013). Evaluating adjusted forcing and model spread for historical and future scenarios in the cmip5 generation of climate models. *Journal of Geophysical Research: Atmospheres*, *118*(3), 1139–1150. https://doi.org/10.1002/jgrd.50174

Frölicher, T. L., Rodgers, K. B., Stock, C. A., & Cheung, W. W.(2016). Sources of uncertainties in 21st century projections of potential ocean ecosystem stressors. *Global Biogeochemical Cycles*, *30*(8), 1224–1243. https://doi.org/10.1002/2015gb005338

Gao, M., Kim, S.-J., Yang, J., Liu, J., Jiang, T., Su, B., et al. (2021). Historical fidelity and future change of Amundsen Sea low under 1.5 c–4 c global warming in CMIP6. *Atmospheric Research*, *255*, 105533. https://doi.org/10.1016/j.atmosres.2021.105533

Garcia, H., Locarnini, R., Boyer, T., Antonov, J., Zweng, M., Baranova, O., et al. (2010). World ocean atlas 2009, vol. 4. In *Nutrients (phosphate, nitrate, silicate), NOAA atlas NESDIS* (Vol. 71). U.S. Government Publishing Office.

Gjermundsen, A., Nummelin, A., Olivié, D., Bentsen, M., Seland, Ø., & Schulz, M. (2021). Shutdown of southern ocean convection controls long-term greenhouse gas-induced warming. *Nature Geoscience*, *14*(10), 724–731. https://doi.org/10.1038/s41561-021-00825-x

Gregg, W. W. (2008). Assimilation of seawifs ocean chlorophyll data into a three-dimensional global ocean model. *Journal of Marine Systems*, *69*(3–4), 205–225. https://doi.org/10.1016/j.jmarsys.2006.02.015

Griffies, S. M., Danabasoglu, G., Durack, P. J., Adcroft, A. J., Balaji, V., Böning, C. W., et al. (2016). OMIP contribution to CMIP6: Experimental and diagnostic protocol for the physical component of the ocean model intercomparison project. *Geoscientific Model Development*, *9*(9), 3231–3296. https://doi.org/10.5194/gmd-9-3231-2016

Haëntjens, N., Boss, E., & Talley, L. D. (2017). Revisiting ocean color algorithms for chlorophyll a and particulate organic carbon in the southern ocean using biogeochemical floats. *Journal of Geophysical Research: Oceans*, *122*(8), 6583–6593. https://doi.org/10.1002/2017jc012844

Hajima, T., Watanabe, M., Yamamoto, A., Tatebe, H., Noguchi, M. A., Abe, M., et al. (2020). Development of the MIROC-ES2l Earth system model and the evaluation of biogeochemical processes and feedbacks. *Geoscientific Model Development*, *13*(5), 2197–2244. https://doi.org/10.5194/gmd-13-2197-2020

Hall, A., Cox, P., Huntingford, C., & Klein, S. (2019). Progressing emergent constraints on future climate change. *Nature Climate Change*, *9*(4), 269–278. https://doi.org/10.1038/s41558-019-0436-6

Heuzé, C. (2021). Antarctic bottom water and north Atlantic deep water in CMIP6 models. *Ocean Science*, *17*(1), 59–90. https://doi.org/10.5194/os-17-59-2021

Heuzé, C., Heywood, K. J., Stevens, D. P., & Ridley, J. K. (2013). Southern ocean bottom water characteristics in CMIP5 models. *Geophysical Research Letters*, *40*(7), 1409–1414. https://doi.org/10.1002/grl.50287

Ilyina, T., Six, K. D., Segschneider, J., Maier-Reimer, E., Li, H., & Núñez-Riboni, I. (2013). Global ocean biogeochemistry model HAMOCC: Model architecture and performance as component of the MPI-Earth system model in different CMIP5 experimental realizations. *Journal of Advances in Modeling Earth Systems*, *5*(2), 287–315. https://doi.org/10.1029/2012ms000178

Jena, B.(2017). The effect of phytoplankton pigment composition and packaging on the retrieval of chlorophyll-a concentration from satellite observations in the southern ocean. *International Journal of Remote Sensing*, *38*(13), 3763–3784. https://doi.org/10.1080/01431161.2017.1308034

Jolliff, J. K., Kindle, J. C., Shulman, I., Penta, B., Friedrichs, M. A., Helber, R., & Arnone, R. A. (2009). Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment. *Journal of Marine Systems*, *76*(1–2), 64–82. https://doi.org/10.1016/j.jmarsys.2008.05.014

Kwiatkowski, L., Bopp, L., Aumont, O., Ciais, P., Cox, P. M., Laufkötter, C., et al. (2017). Emergent constraints on projections of declining primary production in the tropical oceans. *Nature Climate Change*, *7*(5), 355–358. https://doi.org/10.1038/nclimate3265

Kwok, R., & Morison, J. (2016). Sea surface height and dynamic topography of the ice-covered oceans from cryosat-2: 2011–2014. *Journal of Geophysical Research: Oceans*, *121*(1), 674–692. https://doi.org/10.1002/2015jc011357

Laurent, A., Fennel, K., & Kuhn, A. (2021). An observation-based evaluation and ranking of historical Earth system model simulations in the northwest North Atlantic Ocean. *Biogeosciences*, *18*(5), 1803–1822. https://doi.org/10.5194/bg-18-1803-2021

Law, R. M., Ziehn, T., Matear, R. J., Lenton, A., Chamberlain, M. A., Stevens, L. E., et al. (2017). The carbon cycle in the Australian community climate and Earth system simulator (access-ESM1)–part 1: Model description and pre-industrial simulation. *Geoscientific Model Development*, *10*(7), 2567–2590. https://doi.org/10.5194/gmd-10-2567-2017

Locarnini, R., Mishonov, A., Antonov, J., Boyer, T., Garcia, H., Baranova, O., et al. (2010). World ocean atlas 2009. In s. levitus (Ed.), *temperature, atlas NESDIS 68*, (Vol. 1). U.S. Government Publishing Office.

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., et al. (2019). Developments in the MPI-M Earth system model version 1.2 (MPI-ESM1. 2) and its response to increasing CO_2. *Journal of Advances in Modeling Earth Systems*, *11*(4), 998–1038. https://doi.org/10.1029/2018MS001400

Meehl, G., Stocker, T., Collins, W., Friedlingstein, P., Gaye, A., Gregory, J., et al. (2007). Chapter 10: Global climate projections. In *IPCC fourth assessment report*. Cambridge University Press.

Meehl, G. A., Senior, C. A., Eyring, V., Flato, G., Lamarque, J.-F., Stouffer, R. J., et al. (2020). Context for interpreting equilibrium climate sensitivity and transient climate response from the CMIP6 Earth system models. *Science Advances*, *6*(26), eaba1981. https://doi.org/10.1126/sciadv.aba1981

Mohrmann, M., Heuzé, C., & Swart, S. (2021). Southern ocean polynyas in cmip6 models. *The Cryosphere*, *15*(9), 4281–4313. https://doi.org/10.5194/tc-15-4281-2021

Moore, J. K., Lindsay, K., Doney, S. C., Long, M. C., & Misumi, K. (2013). Marine ecosystem dynamics and biogeochemical cycling in the community Earth system model [CESM1 (BGC)]: Comparison of the 1990s with the 2090s under the RCP4. 5 and RCP8. 5 scenarios. *Journal of Climate*, *26*(23), 9291–9312. https://doi.org/10.1175/jcli-d-12-00566.1

Moutier, W., Thomalla, S. J., Bernard, S., Wind, G., Ryan-Keogh, T. J., & Smith, M. E. (2019). Evaluation of chlorophyll-a and POC MODIS aqua products in the southern ocean. *Remote Sensing*, *11*(15), 1793. https://doi.org/10.3390/rs11151793

Nakano, H., Tsujino, H., Hirabara, M., Yasuda, T., Motoi, T., Ishii, M., & Yamanaka, G. (2011). Uptake mechanism of anthropogenic CO_2 in the kuroshio extension region in an ocean general circulation model. *Journal of Oceanography*, *67*(6), 765–783. https://doi.org/10.1007/s10872-011-0075-7

21699291, 2023, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022JC018880 by Ministry Of Health, Wiley Online Library on [23/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Nijsse, F. J., Cox, P. M., & Williamson, M. S. (2020). Emergent constraints on transient climate response (TCR) and equilibrium climate sensitivity (ECS) from historical warming in CMIP5 and CMIP6 models. *Earth System Dynamics*, *11*(3), 737–750. https://doi.org/10.5194/esd-11-737-2020

Nissen, C., & Vogt, M. (2021). Factors controlling the competition between phaeocystis and diatoms in the southern ocean and implications for carbon export fluxes. *Biogeosciences*, *18*(1), 251–283. https://doi.org/10.5194/bg-18-251-2021

Oke, P., Griffin, D., Schiller, A., Matear, R., Fiedler, R., Mansbridge, J., et al. (2013). Evaluation of a near-global eddy-resolving ocean model. *Geoscientific Model Development*, *6*(3), 591–615. https://doi.org/10.5194/gmd-6-591-2013

Orr, J. C., Najjar, R. G., Aumont, O., Bopp, L., Bullister, J. L., Danabasoglu, G., et al.(2017). Biogeochemical protocols and diagnostics for the CMIP6 Ocean model intercomparison project (OMIP). *Geoscientific Model Development*, *10*(6), 2169–2199. https://doi.org/10.5194/gmd-10-2169-2017

Oschlies, A. (2001). Model-derived estimates of new production: New results point towards lower values. *Deep Sea Research Part II: Topical Studies in Oceanography*, *48*(10), 2173–2197. https://doi.org/10.1016/s0967-0645(00)00184-3

Palmer, J., & Totterdell, I. (2001). Production and export in a global ocean ecosystem model. *Deep Sea Research Part I: Oceanographic Research Papers*, *48*(5), 1169–1198. https://doi.org/10.1016/s0967-0637(00)00080-7

Paulsen, H., Ilyina, T., Six, K. D., & Stemmler, I. (2017). Incorporating a prognostic representation of marine nitrogen fixers into the global ocean biogeochemical model HAMOCC. *Journal of Advances in Modeling Earth Systems*, *9*(1), 438–464. https://doi.org/10.1002/2016ms000737

Pinkerton, M. H., Boyd, P. W., Deppeler, S., Hayward, A., Höfer, J., & Moreau, S. (2021). Evidence for the impact of climate change on primary producers in the southern ocean. *Frontiers in Ecology and Evolution*, *9*, 592027. https://doi.org/10.3389/fevo.2021.592027

Pinkerton, M. H., Décima, M., Kitchener, J. A., Takahashi, K. T., Robinson, K. V., Stewart, R., & Hosie, G. W. (2020). Zooplankton in the southern ocean from the continuous plankton recorder: Distributions and long-term change. *Deep Sea Research Part I: Oceanographic Research Papers*, *162*, 103303. https://doi.org/10.1016/j.dsr.2020.103303

Pozo Buil, M., Jacox, M. G., Fiechter, J., Alexander, M. A., Bograd, S. J., Curchitser, E. N., et al. (2021). A dynamically downscaled ensemble of future projections for the California current system. *Frontiers in Marine Science*, *8*, 612874. https://doi.org/10.3389/fmars.2021.612874

Rayner, N., Parker, D. E., Horton, E., Folland, C. K., Alexander, L. V., Rowell, D., et al. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research*, *108*(D14), 4407. https://doi.org/10.1029/2002jd002670

Rickard, G., & Behrens, E. (2016). Cmip5 Earth system models with biogeochemistry: A Ross Sea assessment. *Antarctic Science*, *28*(5), 327–346. https://doi.org/10.1017/s0954102016000122

Rickard, G. J., Behrens, E., & Chiswell, S. M.(2016). Cmip5 Earth system models with biogeochemistry: An assessment for the southwest Pacific Ocean. *Journal of Geophysical Research: Oceans*, *121*(10), 7857–7879. https://doi.org/10.1002/2016jc011736

Roach, C. J., & Speer, K. (2019). Exchange of water between the Ross gyre and acc assessed by Lagrangian particle tracking. *Journal of Geophysical Research: Oceans*, *124*(7), 4631–4643. https://doi.org/10.1029/2018jc014845

Rogers, A., Frinault, B., Barnes, D., Bindoff, N., Downie, R., Ducklow, H., et al.(2020). Antarctic futures: An assessment of climate-driven changes in ecosystem structure, function, and service provisioning in the southern ocean. *Annual Review of Marine Science*, *12*(1), 87–120. https://doi.org/10.1146/annurev-marine-010419-011028

Schlund, M., Lauer, A., Gentine, P., Sherwood, S. C., & Eyring, V. (2020). Emergent constraints on equilibrium climate sensitivity in cmip5: Do they hold for CMIP6? *Earth System Dynamics*, *11*(4), 1233–1258. https://doi.org/10.5194/esd-11-1233-2020

Séférian, R., Bopp, L., Gehlen, M., Orr, J. C., Ethé, C., Cadule, P., et al. (2013). Skill assessment of three Earth system models with common marine biogeochemistry. *Climate Dynamics*, *40*(9), 2549–2573. https://doi.org/10.1007/s00382-012-1362-8

Séférian, R., Nabat, P., Michou, M., Saint-Martin, D., Voldoire, A., Colin, J., et al. (2019). Evaluation of CNRM Earth system model, CNRM-ESM2-1: Role of Earth system processes in present-day and future climate. *Journal of Advances in Modeling Earth Systems*, *11*(12), 4182–4227. https://doi.org/10.1029/2019ms001791

Shiogama, H., Ishizaki, N. N., Hanasaki, N., Takahashi, K., Emori, S., Ito, R., et al. (2021). *Selecting cmip6-based future climate scenarios for impact and adaptation studies*. SOLA.

Smith, W. O., Jr., Dinniman, M. S., Hofmann, E. E., & Klinck, J. M. (2014). The effects of changing winds and temperatures on the oceanography of the Ross sea in the 21st century. *Geophysical Research Letters*, *41*(5), 1624–1631. https://doi.org/10.1002/2014gl059311

Stock, C. A., Dunne, J. P., Fan, S., Ginoux, P., John, J., Krasting, J. P., et al. (2020). Ocean biogeochemistry in GFDL's Earth system Model 4.1 and its response to increasing atmospheric CO_2. *Journal of Advances in Modeling Earth Systems*, *12*(10), e2019MS002043. https://doi.org/10.1029/2019ms002043

Swart, N. C., Cole, J. N., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., et al. (2019). The Canadian Earth system model version 5 (CANESM5. 0.3). *Geoscientific Model Development*, *12*(11), 4823–4873. https://doi.org/10.5194/gmd-12-4823-2019

Tagliabue, A., Mtshali, T., Aumont, O., Bowie, A., Klunder, M., Roychoudhury, A., & Swart, S. (2012). A global compilation of dissolved iron measurements: Focus on distributions and processes in the Southern Ocean. *Biogeosciences*, *9*(6), 2333–2349. https://doi.org/10.5194/bg-9-2333-2012

Tjiputra, J. F., Schwinger, J., Bentsen, M., Morée, A. L., Gao, S., Bethke, I., et al. (2020). Ocean biogeochemistry in the Norwegian Earth system model version 2 (NORESM2). *Geoscientific Model Development*, *13*(5), 2393–2431. https://doi.org/10.5194/gmd-13-2393-2020

Tokarska, K. B., Stolpe, M. B., Sippel, S., Fischer, E. M., Smith, C. J., Lehner, F., & Knutti, R. (2020). Past warming trend constrains future warming in CMIP6 models. *Science Advances*, *6*(12), eaaz9549. https://doi.org/10.1126/sciadv.aaz9549

Vichi, M., Pinardi, N., & Masina, S. (2007). A generalized model of pelagic biogeochemistry for the global ocean ecosystem. Part I: Theory. *Journal of Marine Systems*, *64*(1–4), 89–109. https://doi.org/10.1016/j.jmarsys.2006.03.006

Wang, G., Cai, W., Santoso, A., Wu, L., Fyfe, J. C., Yeh, S.-W., et al. (2022). Future southern ocean warming linked to projected ENSO variability. *Nature Climate Change*, *12*(7), 649–654. https://doi.org/10.1038/s41558-022-01398-2

Westberry, T., Behrenfeld, M., Siegel, D., & Boss, E.(2008). Carbon-based primary productivity modeling with vertically resolved photoacclimation. *Global Biogeochemical Cycles*, *22*(2), GB2024. https://doi.org/10.1029/2007gb003078

Yool, A., Palmiéri, J., Jones, C. G., Mora, L., Kuhlbrodt, T., Popova, E. E., et al.(2021). Evaluating the physical and biogeochemical state of the global ocean component of UKESM1 in CMIP6 historical simulations. *Geoscientific Model Development*, *14*(6), 3437–3472. https://doi.org/10.5194/gmd-14-3437-2021

21699291, 2023, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022JC018880 by Ministry Of Health, Wiley Online Library on [23/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Zahariev, K., Christian, J. R., & Denman, K. L. (2008). Preindustrial, historical, and fertilization simulations using a global ocean carbon model with new parameterizations of iron limitation, calcification, and N_2 fixation. *Progress in Oceanography*, *77*(1), 56–82. https://doi.org/10.1016/j.pocean.2008.01.007

Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., et al. (2020). Causes of higher climate sensitivity in CMIP6 models. *Geophysical Research Letters*, *47*(1), e2019GL085782. https://doi.org/10.1029/2019gl085782

Ziehn, T., Chamberlain, M. A., Law, R. M., Lenton, A., Bodman, R. W., Dix, M., et al. (2020). The Australian Earth system model: Access-ESM1.5. *Journal of Southern Hemisphere Earth Systems Science*, *70*(1), 193–214. https://doi.org/10.1071/es19035

21699291, 2023, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022JC018880 by Ministry Of Health, Wiley Online Library on [23/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License