

The GFDL-CM4X climate model hierarchy, Part I: model description and thermal properties

Stephen M Griffies¹, Alistair Adcroft², Rebecca Lynn Beadling³, Mitchell Bushuk⁴, Chiung-Yin Chang², Henri Francois Drake⁵, Raphael Dussin⁶, Robert W. Hallberg⁷, William Hurlin⁸, Hemant Khatri⁹, John P Krasting¹⁰, Matthew Lobo², Graeme MacGilchrist¹¹, Brandon G Reichl¹, Aakash Sane², Olga V. Sergienko¹², Maike Sonnewald¹³, Jacob M. Steinberg¹⁰, Jan-Erik Tesdal², Matthew D Thomas¹⁴, Katherine Elise Turner¹⁵, Marshall L Ward¹⁶, Michael Winton⁴, Niki Zadeh⁸, Laure Zanna¹⁷, Rong Zhang⁴, Wenda Zhang², and Ming Zhao⁴

¹NOAA Geophysical Fluid Dynamics Laboratory

²Princeton University

³Temple University

⁴GFDL/NOAA

⁵University of California Irvine

⁶National Oceanic and Atmospheric Administration

⁷NOAA/Geophysical Fluid Dynamics Laboratory

⁸NOAA/GFDL

⁹University of Liverpool

¹⁰Geophysical Fluid Dynamics Laboratory

¹¹University of St Andrews

¹²AOS, Princeton University/GFDL

¹³University of California, Davis

¹⁴Centre for Environment, Fisheries and Aquaculture Science

¹⁵University of Arizona

¹⁶Australian National University

¹⁷New York University

November 28, 2024

Abstract

We present the GFDL-CM4X (Geophysical Fluid Dynamics Laboratory Climate Model version 4X) coupled climate model hierarchy. The primary application for CM4X is to investigate ocean and sea ice physics as part of a realistic coupled Earth climate model. CM4X utilizes an updated MOM6 (Modular Ocean Model version 6) ocean physics package relative to CM4.0, and there are two members of the hierarchy: one that uses a horizontal grid spacing of 0.25° (referred to as CM4X-p25) and the other that uses a 0.125° grid (CM4X-p125). CM4X also refines its atmospheric grid from the nominally 100° km (cubed sphere C96) of CM4.0 to 50° km (C192). Finally, CM4X simplifies the land model to allow for a more focused study of the role of ocean changes to global mean climate.

CM4X-p125 reaches a global ocean area mean heat flux imbalance of $-0.02^{\circ}\text{W}^{-2}\text{m}^{-2}$ within $\mathcal{O}(150)$

years in a pre-industrial simulation, and retains that thermally equilibrated state over the subsequent centuries. This 1850 thermal equilibrium is characterized by roughly $\$400 \text{ J}$ less ocean heat than present-day, which corresponds to estimates for anthropogenic ocean heat uptake between 1850 and present-day. CM4X-p25 approaches its thermal equilibrium only after more than 1000 years, at which time its ocean has roughly $\$1100 \text{ J}$ $\{\text{more}\}$ heat than its early 21st century ocean initial state. Furthermore, the root-mean-square sea surface temperature bias for historical simulations is roughly 20% smaller in CM4X-p125 relative to CM4X-p25 (and CM4.0). We offer the $\{\text{mesoscale dominance hypothesis}\}$ for why CM4X-p125 shows such favorable thermal equilibration properties.

The GFDL-CM4X climate model hierarchy, Part I: model description and thermal properties

Stephen M. Griffies^{1,2}, Alistair Adcroft², Rebecca L. Beadling³, Mitchell Bushuk¹, Chiung-Yin Chang², Henri F. Drake⁴, Raphael Dussin¹, Robert W. Hallberg^{1,2}, William J. Hurlin¹, Hemant Khatri⁵, John P. Krasting¹, Matthew Lobo², Graeme A. MacGilchrist⁶, Brandon G. Reichl¹, Aakash Sane², Olga Sergienko², Maike Sonnewald⁷, Jacob M. Steinberg¹, Jan-Erik Tesdal², Matthew Thomas⁸, Katherine E. Turner², Marshall L. Ward¹, Michael Winton¹, Niki Zadeh¹, Laure Zanna⁹, Rong Zhang^{1,2}, Wenda Zhang², Ming Zhao¹

¹NOAA Geophysical Fluid Dynamics Laboratory, Princeton, USA

²Princeton University Atmospheric and Oceanic Sciences Program, Princeton, USA

³Department of Earth and Environmental Science, Temple University, Philadelphia, USA

⁴Department of Earth System Science, University of California, Irvine, USA

⁵Department of Earth, Ocean and Ecological Sciences, University of Liverpool, UK

⁶School of Earth and Environmental Sciences, University of St. Andrews, UK

⁷Computer Science Department, University of California, Davis, USA

⁸Centre for Environment Fisheries and Aquaculture Science, Suffolk, UK

⁹Courant Institute of Mathematical Sciences, New York University, New York, USA

November 23, 2024

Key Points:

- CM4X-p25 and CM4X-p125 are designed to study ocean and sea ice physics, focusing on effects from eddies and boundary currents.
- CM4X-p125 reaches pre-industrial thermal equilibrium in 150 years whereas the coarser CM4X-p25 has yet to equilibrate after 1000 years.
- CM4X-p125's active eddies and negligible spurious mixing render an equilibrated pre-industrial ocean with 400ZJ less heat than present day.

28 **Abstract**

29 We present the GFDL-CM4X (Geophysical Fluid Dynamics Laboratory Climate
 30 Model version 4X) coupled climate model hierarchy. The primary application for CM4X
 31 is to investigate ocean and sea ice physics as part of a realistic coupled Earth climate
 32 model. CM4X utilizes an updated MOM6 (Modular Ocean Model version 6) ocean physics
 33 package relative to CM4.0, and there are two members of the hierarchy: one that uses
 34 a horizontal grid spacing of 0.25° (referred to as CM4X-p25) and the other that uses a
 35 0.125° grid (CM4X-p125). CM4X also refines its atmospheric grid from the nominally
 36 100 km (cubed sphere C96) of CM4.0 to 50 km (C192). Finally, CM4X simplifies the land
 37 model to allow for a more focused study of the role of ocean changes to global mean cli-
 38 mate.

39 CM4X-p125 reaches a global ocean area mean heat flux imbalance of -0.02 W m^{-2}
 40 within $\mathcal{O}(150)$ years in a pre-industrial simulation, and retains that thermally equilibrated
 41 state over the subsequent centuries. This 1850 thermal equilibrium is characterized by
 42 roughly 400 ZJ less ocean heat than present-day, which corresponds to estimates for an-
 43 thropogenic ocean heat uptake between 1850 and present-day. CM4X-p25 approaches
 44 its thermal equilibrium only after more than 1000 years, at which time its ocean has roughly
 45 1100 ZJ *more* heat than its early 21st century ocean initial state. Furthermore, the root-
 46 mean-square sea surface temperature bias for historical simulations is roughly 20% smaller
 47 in CM4X-p125 relative to CM4X-p25 (and CM4.0). We offer the *mesoscale dominance*
 48 *hypothesis* for why CM4X-p125 shows such favorable thermal equilibration properties.

49 **Plain Language Summary**

50 We detail a new climate model hierarchy, CM4X. CM4X has two model configu-
 51 rations, CM4X-p25 and CM4X-p125, that differ only in the ocean/sea ice horizontal grid
 52 spacing. CM4X-p125 outperforms CM4X-p25 for certain climate relevant processes, while
 53 maintaining skill levels seen in previous generations for other results. CM4X-p125 re-
 54 quires about 10 times less time than CM4X-p25 to reach pre-industrial control thermal
 55 equilibration. Also, CM4X-p125 equilibrates to an ocean state with roughly 400 ZJ less
 56 heat content than present-day, consistent with estimates of anthropogenic heat uptake
 57 since 1870, whereas CM4X-p25 equilibrates to a state with roughly 1100 ZJ *more* heat
 58 than present-day. Consequently, the CM4X-p125 ocean state has not drifted far from
 59 observational estimates, so that the ocean water mass structure is more accurate than
 60 models requiring $\mathcal{O}(1000)$ years to equilibrate. We propose the *mesoscale dominance hy-*
61 pothesis to interpret the relatively rapid thermal equilibration of CM4X-p125 to a cooler
 62 pre-industrial state. Such ocean models result from negligible spurious mixing (from nu-
 63 merical truncation errors) along with an active mesoscale transport and realistic par-
 64 ameterization of small scale (diapycnal) mixing. Because of their more rapid thermal equi-
 65 libration, mesoscale dominant ocean models reduce the computational energy footprint
 66 of models that are not mesoscale dominant.

67 **1 Introducing this paper and its thesis**

68 The ocean is the predominant reservoir in the Earth's climate system for enthalpy
 69 (heat), absorbing roughly 90% of the added heat due to anthropogenic increases in green-
 70 house gasses in the atmosphere (Fox-Kemper et al., 2021; von Schuckmann et al., 2023).
 71 Ocean circulation and its attendant physical processes redistribute both pre-industrial
 72 and anthropogenic heat laterally and vertically throughout the World Ocean, with as-
 73 sociated impacts on regional and global climate variations and change.

74 We developed a climate model hierarchy, CM4X, to support research into the ocean's
 75 role in climate. CM4X is comprised of two coupled climate models: CM4X-p25 and CM4X-

76 p125. These two models are identically configured in all aspects, except for their ocean
 77 (and sea ice) horizontal grid spacing and bottom topography. We document the devel-
 78 opment and performance of CM4X in this paper along with its companion in Part II (Griffies
 79 et al., 2024).

80 1.1 Content of this paper

81 We overview the CM4X project in Sections 2 and 3, touching on some of the high-
 82 lights as well as the unresolved problems. We focus in Sections 4 and 5 on the key in-
 83 novation emerging from the CM4X model development. Namely, the CM4X-p125 model
 84 exhibits a relatively rapid $\mathcal{O}(100)$ years thermal equilibration in its pre-industrial control
 85 (piControl) simulation. Furthermore, the global volume integrated ocean heat con-
 86 tent in CM4X-p125 is roughly 400 ZJ less than the present-day initial conditions, which
 87 corresponds to the estimates by Zanna et al. (2019) for the anthropogenic ocean heat
 88 uptake since 1870. In contrast, many other climate models (e.g., CM4.0 and CM4X-p25)
 89 approach a piControl equilibrium only after $\mathcal{O}(1000)$ years and with an ocean heat con-
 90 tent that is higher than present-day. There are relatively few field measurements of the
 91 1850 ocean. Even so, we are virtually certain that the 1850 ocean had less heat content
 92 than the early 21st century ocean. Although the equilibrated heat content is lower in
 93 CM4X-p125, its sea surface temperature is higher than CM4X-p25, which then results
 94 in a reduced bias in the CM4X-p125 historical SST relative to CM4X-p25.

95 We conjecture in Section 6 that the ocean component of CM4X-p125 approximates
 96 a novel class of ocean climate models that are *mesoscale dominant*, with mesoscale dom-
 97 inance described in the next subsection. We conclude the main portion of this paper in
 98 Section 7 by discussing the implications for climate modeling suggested by the CM4X
 99 results. Appendix A then details the CM4X model configuration, focusing on aspects
 100 that are distinct from the OM4.0 ocean/sea ice configuration of Adcroft et al. (2019) and
 101 the CM4.0 climate model of Held et al. (2019). In Part II of this paper (Griffies et al.,
 102 2024) we detail results from a suite of case studies targeting areas of planned and on-
 103 going research with the CM4X hierarchy.

104 1.2 Mesoscale dominance as the paper's central thesis

105 Mesoscale dominant ocean models are characterized by an explicit representation
 106 of vigorous ocean mesoscale features (or with a suitable mesoscale eddy parameteriza-
 107 tion), an accurate parameterization of small scale (diapycnal) mixing, and a low level
 108 of spurious numerical ocean mixing. We hypothesize that these three properties, in com-
 109 bination, form necessary (though not sufficient) conditions for a relatively rapid (cen-
 110 tennial rather than millennial) thermal equilibration of the global climate system under
 111 piControl forcing, and with the piControl equilibrium having less, not more, heat con-
 112 tent than the early 21st century ocean. Stated in a negative manner, we hypothesize that
 113 pi-Control thermal equilibration will require $\mathcal{O}(1000)$ years so long as the ocean model
 114 component offers a poor approximation to mesoscale dominance.

115 We are partly led to propose the mesoscale dominance hypothesis by acknowledg-
 116 ing the history of ocean model developments that have focused on the importance of mesoscale
 117 transport, along with maintaining physically relevant diapycnal mixing and negligible
 118 spurious mixing. Within the context of mesoscale eddy parameterization, Solomon (1971),
 119 Redi (1982), Gent and McWilliams (1990) and Gent et al. (1995) offer a starting point
 120 for this line of research, whereby their formulations stress the need to parameterize the
 121 mesoscale eddy transport while respecting water mass boundaries. The CM4X-p125 re-
 122 sults exemplify this approach through an explicit realization of the mesoscale rather than
 123 via its parameterization, and by employing an ocean model component that pays par-
 124 ticular attention to the numerical needs of reducing spurious mixing between water masses.
 125 Even so, we do not claim that CM4X-p125 perfectly realizes the three characteristics of

126 a mesoscale dominant ocean. Indeed, as documented in Part II of this study (Griffies
 127 et al., 2024), CM4X-p125 has many shortcomings. Rather, we contend that CM4X-p125
 128 approaches these characteristics sufficiently to render a rapid thermal equilibration into
 129 a cooler pre-industrial ocean state, thus offering a taste for what is possible from mesoscale
 130 dominant ocean climate models.

131 We acknowledge the difficulty of quantitatively testing whether a model can be char-
 132 acterized as mesoscale dominant. Indeed, it might be most practical to infer relative scales
 133 of mesoscale dominance based on comparisons across a model hierarchy, such as we do
 134 for CM4X. Developing direct ways of characterizing mesoscale dominance remains a topic
 135 of ongoing research, with some suggestions offered in Section 6 based on water mass anal-
 136 ysis. Even though difficult to test, we propose that the mesoscale dominance hypoth-
 137 esis provides a useful means to frame the question of climate model thermal equilibra-
 138 tion. It also offers targets for model development strategies. In particular, we empha-
 139 size the central role of reducing spurious numerical mixing in developing models capa-
 140 ble of realizing mesoscale dominance. Namely, a strongly energetic ocean mesoscale must
 141 be realized with negligible spurious mixing if the model is to meet the needs of, and the ben-
 142 efits from, mesoscale dominance.

143 Our focus throughout this model description paper concerns the ocean and sea ice.
 144 However, we fully acknowledge that atmospheric and land radiative properties have a
 145 substantial impact on the simulation of sea surface temperature, ocean volume mean tem-
 146 perature, and ocean heat uptake. Radiative properties, in turn, are significantly influ-
 147 enced by cloud parameterizations containing poorly constrained parameters. We doc-
 148 ment the tuning of CM4X cloud parameters in Section A1 of the appendix. Importantly,
 149 by setting all atmospheric model configuration details to be identical for CM4X-p25 and
 150 CM4X-p125, we enable a focus on the ocean’s role in affecting thermal equilibration of
 151 the climate model, as well as other aspects of the simulation.

ACRONYM	MEANING	CITATION OR SECTION
AMIP	Atmospheric Model Intercomparison Project	pcmdi.llnl.gov/mips/amip
AM4	GFDL Atmospheric Model 4.0	Zhao et al. (2018b, 2018a)
CM2-O	GFDL climate model hierarchy 2.0	Delworth et al. (2006), Griffies et al. (2015)
C96	AM4 with cubed-sphere (≈ 100 km)	Zhao et al. (2018b, 2018a)
C192	AM4 with cubed-sphere (≈ 50 km) in CM4X	Zhao (2020)
CMIP6	Coupled Model Intercomparison Project 6	Eyring et al. (2016)
CM4.0	GFDL Climate Model 4.0 (0.25° ocn & C96 atm)	Held et al. (2019)
CM4X	GFDL Climate Model hierarchy	this paper
CM4X-p25	CM4X w/ 0.25° ocn and C192 atm	this paper
CM4X-p125	CM4X with 0.125° ocn and C192 atm	this paper
ESM4.1	GFDL Earth System Model 4.1	Dunne et al. (2020)
GFDL	Geophysical Fluid Dynamics Laboratory	–
MOM6	Modular Ocean Model version 6	Adcroft et al. (2019), Griffies et al. (2020)
OM4.0	GFDL Ocean/Sea-ice Model 4.0 (0.25°)	Adcroft et al. (2019)
OHUE	ocean heat uptake efficiency	Section 5.7
SST	sea surface temperature	Section 4.2
TOA	top of atmosphere	Sections 5.6 and A1

Table 1. Acronyms used in this paper, their meaning, and relevant citation or section.

Acronyms in the upper part of the table refer to model names and configurations, whereas those in the lower portion refer to physical climate properties.

152 2 The CM4X model hierarchy

153 We here introduce the CM4X project and provide reasoning for particular development
 154 strategies and corresponding decisions.

155 2.1 The importance of humility in model development

156 In developing CM4X, we started from the CM4.0 climate model of Held et al. (2019),
 157 along with its OM4.0 ocean/sea ice component from Adcroft et al. (2019). Our initial
 158 development goal was to upgrade the ocean component of CM4.0 to use a finer horizontal
 159 grid for the ocean/sea ice components. In confronting that goal, we chose to revisit
 160 certain earlier choices that then led to further development. We also chose to refine the
 161 atmosphere grid relative to CM4.0 to enhance the representation of atmospheric storms,
 162 which are important not only for studies of high-impact weather and their associated pre-
 163 cipitation (Zhao, 2020, 2022b; Hsieh et al., 2020) but also for simulating ocean extremes
 164 (Yin et al., 2020). A version of CM4.0 with increased atmospheric resolution (C192AM4)
 165 was previously configured for GFDL’s participation in CMIP6 HighResMIP (Haarsma
 166 et al., 2016). Simulations for 1950 control, the historical period, and future projections
 167 up to 2050 were conducted following the CMIP6 HighResMIP specifications, with the
 168 output submitted to CMIP6. However, this coupled model has not been thoroughly ex-
 169 plored or documented.

170 Throughout the development process, we found it crucial to maintain a balance be-
 171 tween what is practical and what is aspirational. Along the way, we experienced the wis-
 172 dom of Hofstadter’s Law (Hofstadter, 1979), which, as paraphrased for the CM4X project,
 173 reads: *Climate model development always takes longer than you expect, even when you*
account for Hofstadter’s Law. Indeed, our original plans in early 2020 for a six-month
 174 project expanded into a multi-year development effort. Part of the extended time frame
 175 arose from expanded visions for the project (i.e., “mission creep”), the entrainment of
 176 more individuals with unique ideas, interests, and scientific needs for an ocean eddying
 177 coupled model, and more simply, from our experience with an application of Murphy’s
 178 Law: *Climate models are Murphy’s Law Machines: anything that can go wrong will go*
wrong.

181 2.2 Balancing the aspirational with the feasible

182 CM4X is designed for research concerning physical ocean and sea ice processes that
 183 manifest in regional and global weather and climate, extending in a seamless fashion from
 184 days (e.g., sea level and temperature extremes) to centuries (e.g., ocean overturning cir-
 185 culation variations and changes). A faithful numerical depiction of such phenomena re-
 186 quires coupled models with relatively fine grid spacing in both the horizontal and ver-
 187 tical. There is growing experience with the payoffs realized by refined ocean grid spac-
 188 ing, in particular for the representation of mesoscale processes and their role in long term
 189 climate such as ocean heat uptake (Griffies et al., 2015). Additionally, refined ocean grid
 190 spacing and subsequent improved topographic representation are essential for the study
 191 of ocean extremes near the coasts. For example, when coupled to an ocean model that
 192 accurately represents the geometry of ocean continental shelves, climate models are key
 193 to the study of coastal ocean heat waves (Bian et al., 2023). The study from Yin et al.
 194 (2020) pointed to the additional benefit of refined atmospheric grid spacing for captur-
 195 ing atmospheric cyclones largely responsible for extremes in coastal sea level. Indeed,
 196 Yin et al. (2020) and Zhao (2020, 2022b) demonstrate the enhanced realism of simulated
 197 tropical and mid-latitude cyclones when moving from a nominally 100 km (C96) version
 198 of the AM4 atmosphere (Zhao et al., 2018a, 2018b) used in CM4.0, to the nominally 50 km
 199 (C192) version of AM4 used by Zhao (2020). These studies, along with those by Hsieh
 200 et al. (2020), motivate us to use the C192 version of AM4 for CM4X. Improved simu-
 201 lations of atmospheric storms would also enable better exploration of the rectified im-

202 pact of small-scale storms on climate, although we do not claim that the model, at its
 203 current resolution, is near convergence.

204 For the ocean component of CM4X, we considered the practical question: *how far*
 205 *can we go in refining the grid spacing while producing a suitable science tool within an*
 206 *acceptable development timeframe?* Centennial-scale climate simulations are necessary
 207 to evaluate the fidelity of inter-annual to inter-decadal modes of coupled climate vari-
 208 ability, yet experience indicates that intellectual patience can greatly dissipate if these
 209 100-year simulations take longer than roughly two months of wall clock time. With this
 210 turnaround in mind, we determined that computer resources available for this project
 211 dictate that 0.125° ocean grid spacing is the finest horizontal grid that we can reason-
 212 ably afford.

213 Furthermore, we remained open to considering advancements across the entire suite
 214 of OM4.0 physical parameterizations documented in Adcroft et al. (2019). We deliber-
 215 ately and systematically advanced the ocean parameterizations (see details in Appendix
 216 A), with tests of changes requiring more than 3×10^3 simulation years in prototypes of
 217 CM4X-p25 and more than 1.7×10^3 simulation years in prototypes of CM4X-p125. This
 218 development process greatly benefited from the roughly 4.5×10^4 coupled model sim-
 219 ulation years used to develop CM4.0 (Adcroft et al., 2019).

220 2.3 Climate model hierarchy based on ocean horizontal grid scale

221 Simulation realism, including a comprehensive suite of physical, chemical, and bi-
 222 ological processes, often runs counter to the needs for unpacking the underlying mech-
 223 anisms rendering scientific insights and robust understanding. In support of understand-
 224 ing, one is motivated to develop models that focus on specific processes and to compro-
 225 mise on comprehensiveness. As emphasized by Held (2005), Jeevanjee et al. (2017), and
 226 Maher et al. (2019), model hierarchies provide a means to span the spectrum between
 227 comprehensive and ideal. Such hierarchies support the science forming the basis for com-
 228 prehensive earth system models, and enhance the science understanding emerging from
 229 model simulations. A model hierarchy can be realized in a variety of manners, for ex-
 230 ample, see Figure 6 in Maher et al. (2019) for atmospheric model hierarchies. The two
 231 CM4X models differ only by their ocean and sea ice grid spacing, along with the rep-
 232 resentation of bottom topography. All other model details are identical. Hence, CM4X
 233 is a hierarchy of scale, with a focus on questions related to the ocean and sea ice.

234 2.4 CM4X design decisions

235 The CM4X hierarchy is inspired by the GFDL CM2-O hierarchy of Griffies et al.
 236 (2015), which consists of three ocean/sea ice models of varying grid spacing that are cou-
 237 pled to the same land model and atmosphere model. The coarsest CM2-O model has a
 238 one degree ocean that uses a mesoscale eddy parameterization. The other two models
 239 have 0.25° and 0.1° ocean grids and use no mesoscale parameterizations (other than bi-
 240 harmonic friction). For the CM4X project, we targeted the role of explicitly represented
 241 mesoscale eddies and boundary currents in the eddy-admitting regime. Inclusion of a one-
 242 degree ocean component to the CM4X hierarchy would necessarily require development
 243 of an eddy parameterization for this model, with such research outside of our scope.

244 Further examples of ocean grid spacing hierarchies include the coupled climate model
 245 of Kirtman et al. (2012), which includes a 1° ocean (with mesoscale eddy parameteri-
 246 zations) and 0.1° ocean (without mesoscale eddy parameterizations). The hierarchy of
 247 Kiss et al. (2020) shares much with CM2-O, yet with a focus on forced ocean/sea ice sim-
 248 ulations using the JRA55-do atmospheric state from Tsujino et al. (2018). The Never-
 249 world2 hierarchy of Marques et al. (2022) consists of idealized stacked shallow water mod-

250 els using MOM6 and is designed to study adiabatic aspects of mesoscale eddies and their
 251 parameterization.

252 *Absence of mesoscale eddy parameterizations in CM4X*

253 We disabled mesoscale eddy parameterizations for CM4X, which also follows the
 254 choice of Adcroft et al. (2019) for OM4.0, as well as in the hierarchies of Griffies et al.
 255 (2015), Kirtman et al. (2012), Kiss et al. (2020), and Marques et al. (2022), for those con-
 256 figurations with 0.25° and finer grid spacing. Disabling mesoscale eddy parameteriza-
 257 tion helps to clarify the physics associated with distinct representations of ocean fluid
 258 mechanics. Even so, this design choice in no way presumes that the ocean mesoscale is
 259 accurately or completely represented by either the 0.25° or 0.125° grids in CM4X (e.g.,
 260 see Figure 1). Indeed, many authors to this paper are actively pursuing research into pa-
 261 rameterizations suited for the eddy active “gray zone” regime of CM4X. This regime re-
 262 mains challenging for the ocean modeling community, similar in vein to the atmospheric
 263 model quest for storm-resolving resolutions, yet it represents the foreseeable future for
 264 GFDL coupled models. The CM4X hierarchy informs this parameterization research.

265 *Component model simplifications to align with CM4X science goals*

266 The Earth system complexity is deliberately limited in the CM4X hierarchy rel-
 267 ative to more comprehensive climate or earth system models. In particular, CM4X does
 268 not include an interactive biogeochemistry component. Furthermore, in CM4X (as for
 269 CM4.0), we use prescribed glaciers and icecaps rather than interactive ice sheets (we choose
 270 distinct albedos relative to the CM4.0 values, as discussed in Appendix A3). Interactive
 271 land ice capabilities are at the cutting-edge of climate modeling science today, and they
 272 are outside of our scope. We made further simplifications to the CM4X land model rel-
 273 ative to CM4.0, with details provided in Appendix A2. In brief, CM4X makes use of static
 274 rather than dynamic land vegetation along with a land use forcing fixed at 1980 values,
 275 along with no CO₂ fertilization. These land model changes compromise CM4X’s abil-
 276 ity to simulate climate in the presence of known forcing agents used by CMIP6 models
 277 such as CM4.0.

278 *CM4X is not a comprehensive CMIP6-class climate model*

279 In some of our analysis, we include results from CM4.0 to benchmark the CM4X
 280 results. In such comparisons, it is important to note that the various configuration dif-
 281 ferences between CM4X and CM4.0 preclude an unambiguous attribution of causes for
 282 simulation differences. The only clean comparison in this paper is between CM4X-p25
 283 and CM4X-p125. Even so, in Appendix A3 we show that from a global mean heat bud-
 284 get perspective, the key difference between CM4.0 and CM4X-p25 concerns the distinct
 285 snow on glacier albedos.

286 Certain of the CM4X design choices, particularly those associated with the land
 287 model, preclude CM4X from being considered a comprehensive CMIP6-class climate model,
 288 whereas CM4.0 directly targeted CMIP6 questions. This limitation is important to ac-
 289 knowledge when comparing CM4.0 and CM4X during the historical simulation since CM4.0
 290 has the CMIP6 land forcings whereas CM4X does not. Simplifying the climate model
 291 by disabling elements of the land model reduced the risk of having the perfect be the en-
 292 emy of the good in our task of developing a fit-for-purpose coupled climate model hier-
 293 archy focused on ocean and sea ice physics.

294 **3 A survey of CM4X results**

295 We here offer a selection of results that highlight facets of the model simulations
 296 using pre-industrial, historical, and future scenario experiments from CMIP6 (Eyring et
 297 al., 2016). Further details for many of these diagnostics are presented in the various case
 298 studies of Part II (Griffies et al., 2024).

299 **3.1 Experiments, initialization, and analysis**

300 We analyze CM4X as realized in the following CMIP6 experiments (Eyring et al.,
 301 2016).

- 302 • **piControl:** Pre-industrial control with radiative forcing fixed at year 1850. This
 303 experiment illustrates how the models drift from their initial conditions, taken from
 304 the early 21st century, and approach thermal equilibrium under pre-industrial forc-
 305 ing.
- 306 • **Historical:** 01January of year 101 from the piControl is used to initialize a histor-
 307 ical simulation that is run from 1850 to 2014. In this historical simulation, we did
 308 not account for temporal evolution in vegetation, land use, or CO₂ fertilization.
- 309 • **SSP5-8.5:** 01January of year 2015 provides the initial condition for the CMIP6 SSP5-
 310 8.5 scenario experiment, which allows us to study how the CM4X models simu-
 311 late climate change through to 2100.

312 The piControl ocean was initialized from the 2013 World Ocean Atlas, which consists
 313 of Locarnini et al. (2018) for temperature and Zweng et al. (2018) for salinity. This anal-
 314 ysis is dominated by late 20th and early 21st century measurements, though with no-
 315 tably poor sampling of the deep ocean. The ocean model makes use of the pre-TEOS10
 316 equation of state from Wright (1997). We interpret the model temperature and salin-
 317 ity according to McDougall et al. (2021), whereby the model's prognostic temperature
 318 field is interpreted as Conservative Temperature. There is no hidden or undocumented
 319 spin-up phase for the ocean temperature and salinity. Rather, all piControl time series
 320 presented in this paper start with ocean initial conditions taken from World Ocean At-
 321 las.

322 For CM4X-p125, we took initial conditions for the land model from CM4X-p25 as
 323 remapped onto the CM4X-p125 land/sea mask grid. Sea ice, iceberg, and calving ini-
 324 tial conditions for CM4X-p125 and CM4X-p25 were taken from prototype configura-
 325 tions that ran for 100 and 70 years, respectively, under 1950 radiative forcing conditions. This
 326 choice avoids “cold starting” the sea ice component, which can introduce large initial salin-
 327 ity and temperature pulses into the high-latitude oceans due to sea ice formation. In both
 328 CM4X-p25 and CM4X-p125, initial conditions for sea ice concentration and sea ice vol-
 329 ume reasonably agree with present-day observations. The initial Pan-Arctic sea ice ex-
 330 tent is $16.6 \times 10^6 \text{ km}^2$ in CM4X-p125 and $14.4 \times 10^6 \text{ km}^2$ in CM4X-p25, and the ini-
 331 tial Pan-Antarctic sea ice extent is $11.4 \times 10^6 \text{ km}^2$ (CM4X-p125) and $7.9 \times 10^6 \text{ km}^2$
 332 (CM4X-p25).

333 For horizontal maps, we present results on the model's native horizontal grid. For
 334 the vertical, we follow the CMIP6 recommendations from Section A4 of Griffies et al.
 335 (2016) by generating results on a 35-level diagnostic z^* grid used in versions of the World
 336 Ocean Atlas (Locarnini et al., 2018). We do so by remapping, on every model time step,
 337 to the z^* grid from the model native grid.

338 **3.2 Climate modeling with an active ocean mesoscale**

339 Hallberg (2013) as well as Griffies and Treguier (2013) noted that the first baro-
 340 clinic Rossby radius, L_d , is a useful measure for gauging the degree by which the ocean

horizontal grid spacing is able to resolve the mesoscale ocean dynamics (see K. Stewart et al. (2017) for complementary statements about vertical grid requirements). The reason is that the most unstable baroclinic waves have a length scale proportional to the deformation radius (e.g., Vallis (2017)). The Rossby radius also determines the scale of coastal boundary waves. Following the methods from Hallberg (2013), we compute L_d as realized in CM4X-p125 (Figure 1). Also in Figure 1, we map the ratio, L_d/Δ , where Δ is the horizontal grid spacing given by $\Delta^2 = (\Delta x)^2 + (\Delta y)^2$. This figure reveals that the 0.125° grid is insufficient to fully resolve the first baroclinic Rossby radius in the middle to high latitudes as well as around the continental shelves, with the 0.25° grid (not shown) satisfying $L_d/\Delta > 2$ in a smaller region encompassing just the tropics.

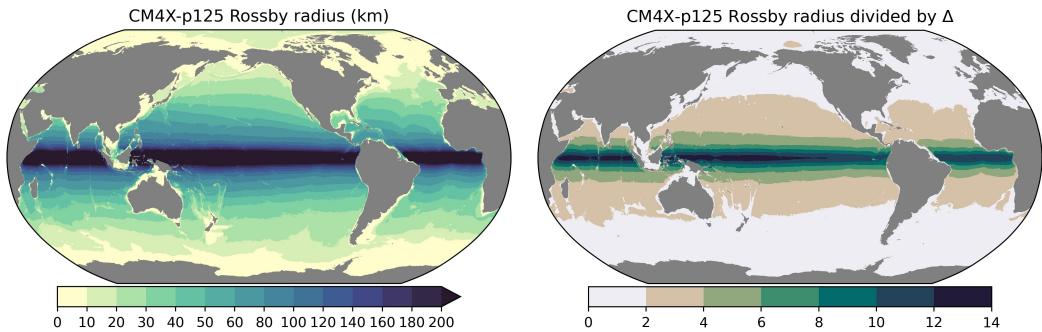


Figure 1. Left panel: first baroclinic Rossby radius, L_d , computed according to Hallberg (2013). The deformation radius is computed as a time average over years 1980-2009 in the historical CM4X-p125 simulation. Right panel: ratio of the first baroclinic Rossby radius to the horizontal grid spacing, Δ , with the grid spacing determined by $\Delta = \sqrt{(\Delta x)^2 + (\Delta y)^2}$. Following Hallberg (2013) and Griffies and Treguier (2013), regions where the ratio is less than two (i.e., the white regions) support a relatively poor representation of mesoscale processes (e.g., baroclinic instability), whose scale is determined by the deformation radius. Roughly 53% of the ocean surface area has $L_d/\Delta \geq 2$ for CM4X-p125, whereas for the same ocean state, CM4X-p25 has only 23% of the ocean area with $L_d/\Delta \geq 2$.

Figure 1 exemplifies the challenge posed by Hallberg (2013). Namely, the need to provide rational and accurate parameterizations of incompletely resolved mesoscale processes in an ocean where those processes have widely varying spatial (and temporal) scales, and as such, where those processes might be adequately resolved in some regions yet incompletely resolved in others. Consequently, there is a need for “scale-aware” physics parameterizations to accommodate the regional dependence of ocean dynamical scales. A further illustration of this challenge is revealed by Figure 2, which shows the magnitude of the gradient in the dynamic sea level as computed for 31 December 2014. According to the Hallberg (2013) criteria, the Rossby radius is unresolved by either model in the high latitudes (i.e., Figure 1). Even so, the dynamic sea level for both CM4X-p25 and CM4X-p125 contains many meandering mesoscale fronts and eddies in the high latitudes. Evidently, the deformation radius provides a scale for baroclinically unstable linear waves, and as such, it is important for the initiation of mesoscale eddy features. Yet, the deformation radius does not necessarily set the scale of statistically equilibrated non-linear mesoscale fronts and eddies. Indeed, mesoscale eddies are here found at scales larger than the deformation radius, which can be interpreted as due to the inverse cascade of kinetic energy occurring in geostrophic turbulence (e.g., Vallis (2017)). A mesoscale eddy parameterization ideally will not overly dampen such eddies, while accounting for those eddies that are missing.

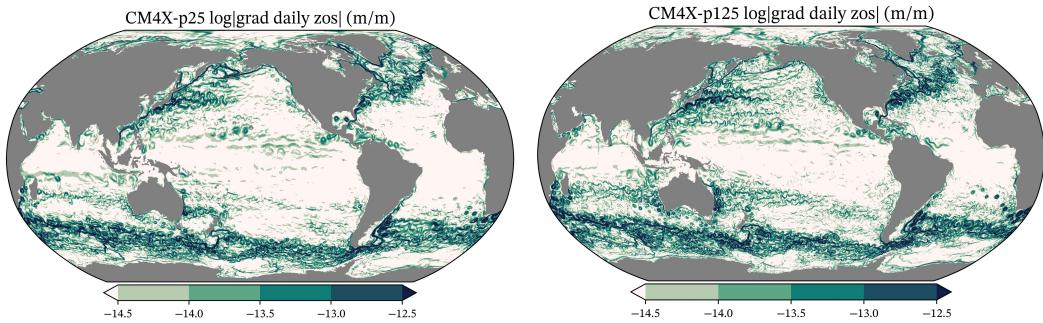


Figure 2. Magnitude of the gradient of the daily mean dynamic sea level for CM4X-p25 (left) and CM4X-p125 (right), as computed from 31 December 2014 in the historical simulations. Meandering mesoscale jets and eddies are represented in these simulations even though the models do not resolve the first baroclinic Rossby radius in most of the high latitudes, in particular within the Southern Ocean (see Figure 1). Note that the dynamic sea level is also known as `zos` within the CMIP community, with details provided in Appendix H of Griffies et al. (2016) and in Section N11 of Gregory et al. (2019).

370 *Eddy activity as revealed by sea level variability*

371 Through the connection between sea level and surface geostrophic currents, sea level
 372 variability provides an effective spatial measure of mesoscale eddy activity in the sur-
 373 face ocean in off-equatorial regions. We thus show in Figure 3 the standard deviation
 374 (root-mean-square) of the daily mean sea level relative to a climatological mean. This
 375 figure reveals the expected increase in variability when moving from CM4X-p25 to CM4X-
 376 p125, given the more active mesoscale eddies admitted by the finer grid spacing. This
 377 increased variability is revealed within the boundary current extensions as well as the
 378 Southern Ocean. For example, the Kuroshio Current in the west Pacific is notably stronger
 379 and extends further eastward in CM4X-p125, making it more consistent with the GLO-
 380 RYS12 analysis product. The Gulf Stream variability also increases in CM4X-p125.

381 Although CM4X-p125 shows more eddy variability than CM4X-p25, it is still far
 382 weaker than the GLORYS12 analysis. Furthermore, there is a notable bias in both CM4X
 383 simulations in which the Gulf Stream exhibits broad and diffuse variability close to the
 384 American coast, which contrasts to GLORYS12 in which variability extends further east-
 385 ward and along a more localized frontal region. We have more to say about deficiencies
 386 of the Gulf Stream in Sections 2 and 6 of Part II (Griffies et al., 2024).

387 *Enhanced role for boundary currents*

388 Along with the inclusion of open ocean transient mesoscale eddies, the CM4X mod-
 389 els capture strong boundary currents, thus allowing for more realistic interactions be-
 390 tween flows along the continental shelves and the open ocean (e.g., for the Antarctic Slope
 391 Current, see Beadling et al. (2022); Beadling (2023); Tesdal et al. (2023)). We illustrate
 392 these features for the depth averaged flow for CM4X-p125 in Figure 4, and then bring
 393 the zonal and meridional flows together in Figure 5 by showing the square root of the
 394 depth averaged kinetic energy.

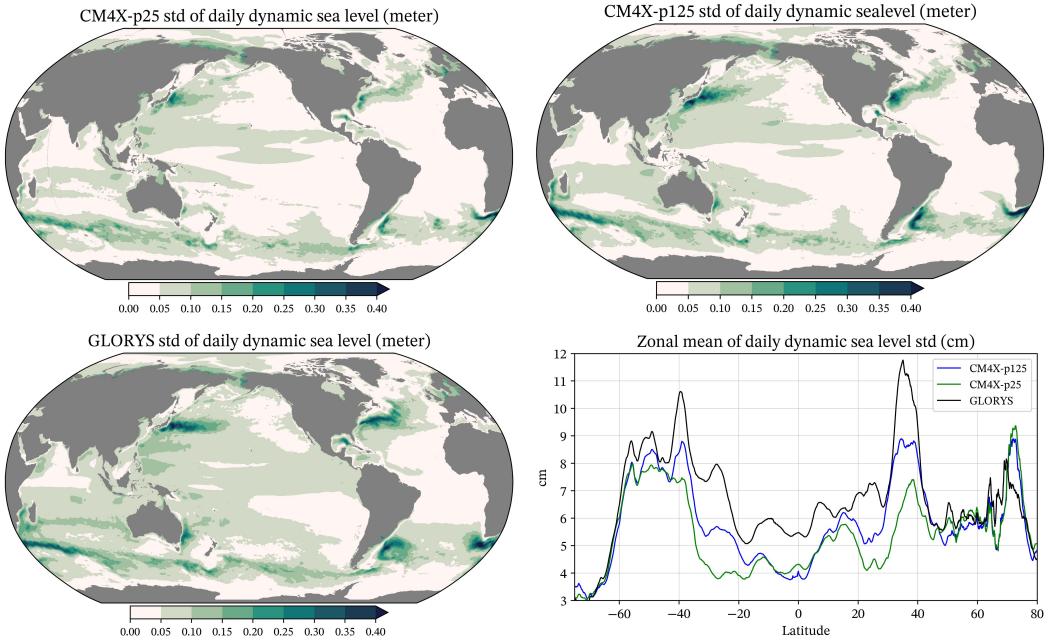


Figure 3. Standard deviation for the daily mean dynamic sea level. The top row shows CM4X-p25 and CM4X-p125, and the bottom row shows the 1/12° GLORYS12 analysis product (Lellouche et al., 2021) as well as the zonal mean for the three. Details for the standard deviation calculation are provided in Section 2 of Part II (Griffies et al., 2024).

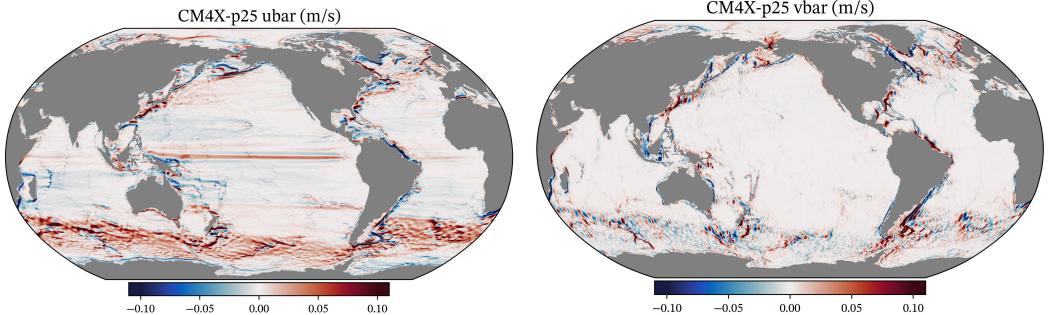


Figure 4. Depth averaged zonal velocity (left) and meridional velocity (right) for CM4X-p125, as computed from the time mean over model years 1980–2009 in the historical simulations. Note the boundary currents as well as the flows at the high latitudes. Additionally, along the equator, we find the zonal flows comprising the equatorial current system.

395

Barotropization particularly in the high latitudes

396

As part of the inverse cascade of geostrophic turbulence, we expect mesoscale eddying flows to carry a large portion of the kinetic energy within barotropic motions, and the process of transferring kinetic energy to the barotropic motions is known as “barotropization” (e.g., Charney (1971), Salmon (1980), Smith and Vallis (2001), Scott and Wang (2005), Kjellsson and Zanna (2017), Yankovsky et al. (2022)). Following Yankovsky et al. (2022), in Figure 5 we provide a rudimentary measure of barotropization given by the ratio of kinetic energy contained in the depth-averaged flow (which approximates the barotropic motion) to the depth-averaged kinetic energy. Note how kinetic energy in the lower lat-

397

398

399

400

401

402

403

404 latitudes is mostly baroclinic, including the equatorial currents. In contrast, high latitude
 405 currents contain a sizable fraction of their kinetic energy in barotropic motions, with this
 406 property related to the relatively weak vertical stratification found particularly in the
 407 high latitude Southern Ocean and North Atlantic. Strong barotropic motions are con-
 408 strained (at least in part) by the need to preserve planetary geostrophic potential vor-
 409 ticity; i.e., to follow contours of constant f/H , where f is the Coriolis parameter and H
 410 is the bottom depth. This property of such flows emphasizes the importance of topographic
 411 representation, as well as for capturing the strong flows that interact with the bottom.

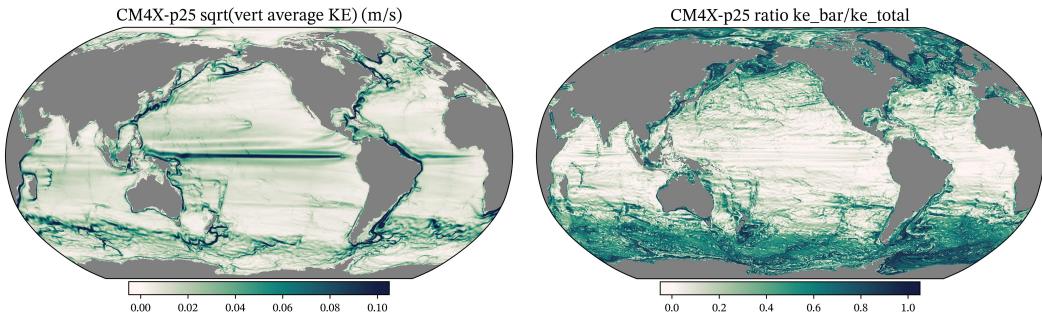


Figure 5. Left: square root of the depth averaged kinetic energy per mass, $1/2 \sum \Delta z \langle u \rangle \cdot \langle u \rangle / \sum \Delta z$, as computed using the 1980-2009 (in the historical simulations) time mean horizontal velocity for CM4X-p125, $\langle u \rangle$. Right: ratio of the kinetic energy contained in the depth averaged flow, $1/2 \overline{\langle u \rangle} \cdot \overline{\langle u \rangle}$, to the total kinetic energy. Here, we use the depth and time averaged flow as given by $\overline{\langle u \rangle} = \sum \Delta z \langle u \rangle / \sum \Delta z$. Regions where the ratio is large (e.g., Southern Ocean, North Atlantic, boundary currents) are regions where the depth averaged flow plays a leading role in determining the depth averaged kinetic energy.

412 3.3 Northward ocean heat transport

413 In Figure 6 we display the northward ocean heat transport (referenced to 0°C), as
 414 diagnosed from the CM4.0, CM4X-p25, and CM4X-p125 simulations averaged over sim-
 415 ulation years 1980-2009, as well as the reanalysis-based results from Trenberth and Caron
 416 (2001) and ocean *in situ* based analysis of Ganachaud and Wunsch (2000). The model
 417 results are closer to one another than to the reanalysis and *in situ* measures, and gen-
 418 erally show a somewhat reduced magnitude for poleward transport. Furthermore, the
 419 model results are quite close in the northern hemisphere, whereas CM4X-p125 exhibits
 420 somewhat more southward transport away from the tropics in the Indian-Pacific and global
 421 transports. Yet overall we conclude that the CM4.0 and CM4X models are closely aligned
 422 in their poleward heat transports. This result was also found in the CM2-O hierarchy
 423 by Griffies et al. (2015), where it was noted that the structure of the poleward ocean heat
 424 transport is dominated by the atmosphere, whereas the vertical heat transport, and hence
 425 the global balance of heat within the climate system, is dominated by ocean processes
 426 that affect a vertical exchange of heat, such as diapycnal mixing and mesoscale eddy trans-
 427 port.

428 3.4 Time scales for piControl thermal equilibration

429 The figures shown thus far in this section reveal that both CM4X-p25 and CM4X-
 430 p125 capture a nontrivial amount of mesoscale eddy activity (albeit weaker than the GLO-
 431 RYS12 analysis seen in Figure 3) as well as strong boundary currents, with such assess-
 432 ment relative to non-eddying one-degree models. A key question concerns the degree to

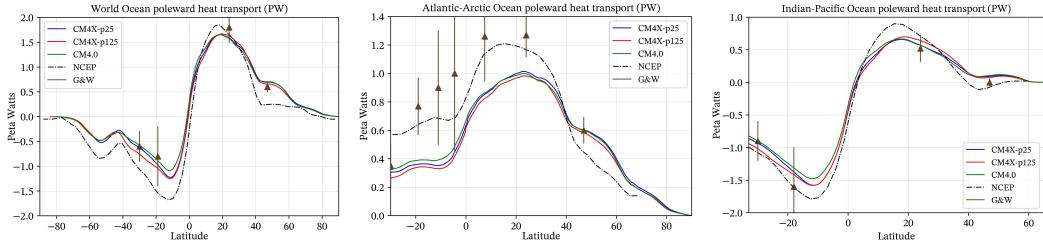


Figure 6. Northward ocean heat transport (referenced to 0°C) for the World Ocean, Atlantic-Arctic Oceans, and Indian-Pacific Oceans. We show results averaged over years 1980-2009 from CM4X-p25, and CM4X-p125, as well as implied ocean heat transports from Trenberth and Caron (2001) using the NCEP reanalysis, and the ocean *in situ* based analysis of Ganachaud and Wunsch (2000). Note the different ranges for the latitude and heat transports for the different panels. The CM4X results are reasonably close to one another, since the poleward heat transport is dominated by the same atmosphere model used in both of the CM4X models.

which these mesoscale features affect the long-term climate and response to anthropogenic warming. Asking this question in a more limited manner, we wish to know whether the strength of ocean eddy activity impacts the time scale for the climate system to reach thermal equilibrium in a piControl simulation. In anticipation of exploring this question in Sections 4, 5 and 6, we here present three time series in Figure 7 from the piControl simulations.

Figure 7 indicates that the area mean SST in the CM4.0 and CM4X models reach a quasi-steady thermal equilibrium after 150 years, though with differing longer term drifts. Note that the SST in CM4X-p25 is cooler than CM4X-p125 and CM4.0. The SST in CM4.0 exhibits a sizable centennial SST fluctuation after year 500, which is related to Southern Ocean super-polynyas in CM4.0 as discussed in Held et al. (2019) and Dunne et al. (2020). As seen in Section 5 of Part II (Griffies et al., 2024), we find no such fluctuations in either CM4X-p25 or CM4X-p125.

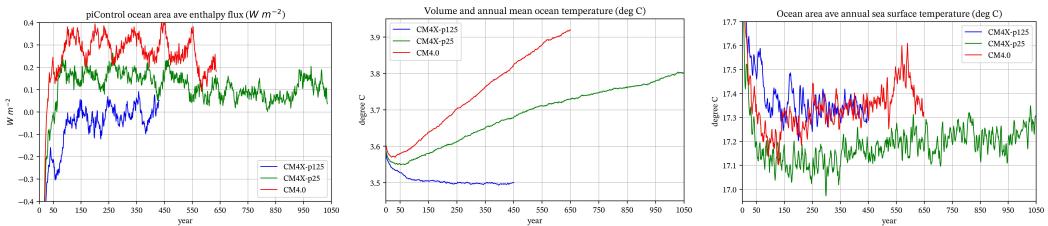


Figure 7. Ocean model diagnostics from the piControl simulations with CM4.0, CM4X-p25, and CM4X-p125. Left panel: net ocean area integrated boundary heat flux entering the ocean, including surface boundary (turbulent fluxes, radiative fluxes, enthalpy associated with mass transfer, and frazil sea ice formation), as well as the bottom geothermal heating. The ocean area provides the normalization for this heat flux, so to get the earth area normalized flux multiply by 0.71. The time series are smoothed with a 30-year rolling mean, hence the missing values near the start and end. Middle panel: time series for the volume mean and annual mean global ocean Conservative Temperature for CM4X-p25, CM4X-p125, and CM4.0. Right panel: time series for the ocean area mean 5-year rolling mean sea surface temperature.

SST is a key ocean field for coupled model simulations due to its interactions with the atmosphere in establishing feedback processes throughout the climate system. But to measure thermal equilibration of the climate system, we must consider the full ocean depth. The reason is the ocean is the major repository for excess enthalpy arising from anthropogenic impacts, thus motivating us to consider the net ocean boundary heat flux and volume mean ocean temperature. (Note that all heat fluxes in this section are reported with an ocean area normalization rather than global area mean.) Figure 7 indicates that the models possess two separate time scales for ocean boundary heat fluxes. During the first few years, there is a negative heat flux (heat leaving the ocean). This initial release of heat is expected since the atmosphere radiative forcing is pre-industrial yet the ocean is initialized with an estimate of the 2013 ocean conditions, with the 2013 ocean warmer than the pre-industrial ocean. Thereafter, the net heat fluxes in both CM4.0 and CM4X-p25 steadily increase, switching from negative to positive after 20 years (CM4.0) and 50 years (CM4X-p25). These models also exhibit centennial scale fluctuations of roughly $\pm 0.1 \text{ W m}^{-2}$ amplitude. The positive heat fluxes entering the ocean in CM4.0 and CM4X-p25 cause the ocean volume mean temperature to exhibit a quasi-linear rise. In contrast, the integrated boundary heat flux in CM4X-p125 increases more gradually over the first 100 years, and thereafter it reaches a near statistical steady state with fluctuations centered around zero. Consequently, the ocean volume mean temperature in CM4X-p125 is nearly steady after around 150 years.

The warming drift in the global volume mean ocean temperature for CM4X-p25 continues until the end of the piControl simulation at year 1050. The trend is generally declining, though with some centennial-scale fluctuations. In particular, during years 751-900, CM4X-p25 has an ocean area mean and time mean boundary heat flux of 0.07 W m^{-2} , whereas for years 901-1050 it is 0.10 W m^{-2} , though it appears to flatten just after year 1000. CM4.0 exhibits a relatively large warming drift, with roughly twice the net imbalance of boundary heat fluxes as found in CM4X-p25. CM4.0 was only run to year 650 due to the presence of Southern Ocean super-polynyas documented in Held et al. (2019) and Dunne et al. (2020). We thus do not know how long it would take CM4.0 to reach thermal equilibrium. In contrast to the long-term warming found in CM4X-p25 and CM4.0, the finer ocean grid spacing in CM4X-p125 leads to far less thermal drift, with roughly -0.02 W m^{-2} net imbalance as averaged over years 100-300 and averaged over the ocean surface area. As a result, CM4X-p125 reaches a nearly steady thermal equilibrium with a cooler ocean after roughly 150 years, and it remains that way for the remaining years of the simulation.

3.5 High latitude overturning and ventilation

Much of our planned research with CM4X concerns the study of physical processes that affect the ocean's meridional overturning circulation (Cessi, 2019; R. Zhang et al., 2019; Hirschi et al., 2020; A. L. Stewart et al., 2021), with those plans guiding the case studies presented in Part II (Griffies et al., 2024). To get a sense of the CM4X overturning circulation, we show the global ocean overturning streamfunction in Figure 8, following the analysis of A. L. Stewart et al. (2021). Both CM4X models display the expected pole-to-pole overturning, whereby the North Atlantic overturning (red clockwise cell) extends into the southern hemisphere and connects to Southern Ocean intermediate waters around 45°S to 50°S . The Atlantic cell is bounded below by the Antarctic Bottom Water cell (blue counter-clockwise cell) that floods the abyssal ocean. Above these cells we find the shallow Ekman-driven overturning cells in the lower latitudes. We have far more to say concerning the Atlantic and Southern Ocean overturning circulation in Sections 5 and 6 of Part II (Griffies et al., 2024).

The mixed layer depth provides a summary diagnostic of processes affecting ocean ventilation. In Figure 9, we show summer and winter climatology of mixed layer depth from the historical and SSP5-8.5 simulations, comparing CM4X-p125 to both Argo and

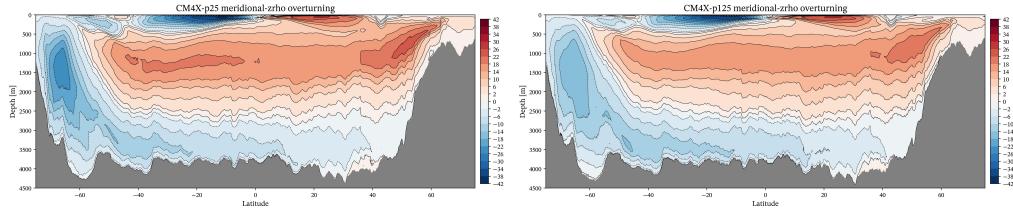


Figure 8. Meridional-density overturning circulation as a function of latitude and potential density referenced to 2000 dbar: $\psi(y, \sigma, t) = -\rho_0 \oint \left[\int_{z'=z_\sigma}^{z'=z_\sigma} v(x, y, z', t) dz' \right] dx$, where $\rho_0 = 1035 \text{ kg m}^{-3}$ is the Boussinesq reference density, σ is the potential density referenced to 2000 dbar, $z_\sigma(x, y, t)$ is the vertical position of the σ isosurface, $z = -H(x, y)$ is the ocean bottom, and v is the meridional velocity. Following A. L. Stewart et al. (2021), we map the circulation onto the time and zonal average depth of the potential density surfaces. Units are mass-Sverdrup (10^9 kg s^{-1}). The left panel shows CM4X-p25 and the right panel CM4X-p125. We make use of 1980-2009 time mean flows and density surfaces to compute this streamfunction. The meridional velocity is mapped to potential density surfaces online using each model time step. Hence, we are here showing the residual-mean overturning circulation.

498 CM4X-p25. Among the regions of notable biases, we see overly deep mixed layer depths
 499 in CM4X-p125 in the mode and intermediate water regions of the Southern Ocean, as
 500 well as in the Labrador Sea (Figure 9D). Biases in the Southern Ocean ventilation for
 501 CM4.0 are the subject of Krasting et al. (2024), with these biases also reflected in the
 502 CM4X simulations. We find a notable reduction in the Labrador Sea bias moving from
 503 CM4X-p25 to CM4X-p125, presumably due to the stronger mesoscale eddies contribut-
 504 ing to more restratification in CM4X-p125 (Figure 9E) (J. Marshall & Schott, 1999). Even
 505 so, the overly deep mixed layer biases in the Southern Ocean worsen in CM4X-p125, pos-
 506 sibly due to the eddy-induced deepening effects in the Indo-West Pacific region described
 507 by Q. Li and Lee (2017). The SSP5-8.5 scenario experiment finds a systematic reduc-
 508 tion in the mixed layer depths throughout most of the high latitude Southern Ocean and
 509 North Atlantic (Figure 9F).

510 3.6 Tropical variability

511 Figure 10 shows the power spectra for the Niño-3 SST index from the historical sim-
 512 ulation (years 1850-2014) with comparison to the NOAA reconstructed product over years
 513 1854-2014 from Huang et al. (2017). Although there is much spread in the spectrum,
 514 the CM4X models show consistently weaker spectral power than the reconstruction across
 515 most time scales. Weak tropical variability is also reflected in the skewness of the sea
 516 surface height (Figure 4 in Part II of Griffies et al. (2024)). We hypothesize that CM4X's
 517 weak tropical variability is related to changes in the ocean mixing parameterizations dis-
 518 cussed in Appendix A. This hypothesis is prompted by noting that the tropical variabil-
 519 ity is stronger (consistent with CM4.0 as documented in Section 5.5 of Held et al. (2019))
 520 when the C192 atmosphere (used here for CM4X) is coupled to the OM4.0 ocean for a
 521 150-year piControl experiment (not shown).

522 Interannual tropical variability can exhibit rather sizable centennial scale modu-
 523 lations (Wittenberg, 2009), so that quasi-stationary power spectra result only after run-
 524 ning multi-centennial scale coupled simulations and/or ensembles. Unfortunately, we could
 525 not afford to run long simulations to examine every separate change made to the ocean
 526 physical parameterizations. Consequently, we are unaware of the precise changes that
 527 led to the weaker variability. Unraveling this story remains an ongoing process. For ex-

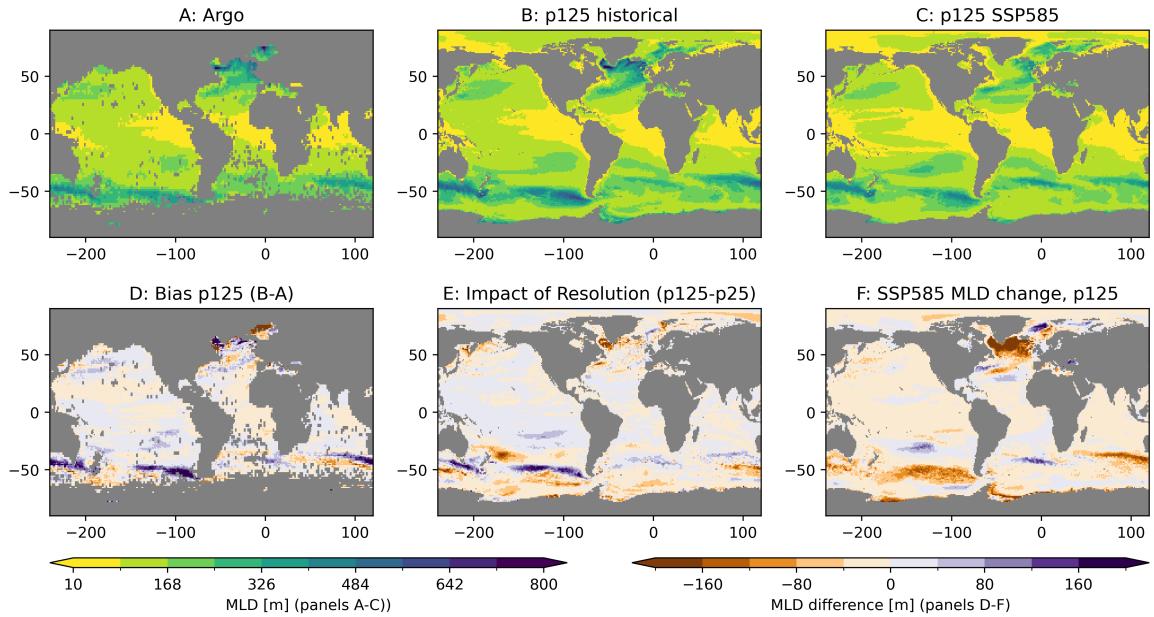


Figure 9. Monthly climatology of the maximum mixed layer depth from Argo (years 2004–2023) (Argo, 2023), as well as years 2000–2014 (historical simulation) and years 2085–2099 (SSP5-8.5 simulation) from CM4X-p25 and CM4X-p125. The mixed layer depth is diagnosed according to the potential energy method from Reichl et al. (2022) (using 2500 J m^{-2} energy criteria). Panel A: estimates from the Argo profiling floats; Panel B: results from CM4X-p125 historical experiment; Panel C: CM4X-p125 SSP5-8.5 simulation. Panel D shows the differences between CM4X-p125 and Argo (Panels B-A), whereas Panel E shows the difference between CM4X-p125 and CM4X-p25. Finally, Panel F shows the impacts from the SSP5-8.5 climate change, showing years 2084–2099 minus years 2000–2014 (Panels C-B) from CM4X-p125. Note that the differences documented in Panels D and E are robust to a longer time average over model years 1955–2014.

528 example, improved representation of upper ocean vertical mixing in this region was not included in CM4X (Reichl et al., 2024).

530 3.7 Coupled climate benchmarks

531 In Figure 11, we present benchmark performance metrics for years 1978–2014 of the
 532 historical simulations in the GFDL earth system model ESM4.1 (Dunne et al., 2020),
 533 along with the climate models CM4.0, CM4X-p25, and CM4X-p125. We compare mod-
 534 els with observations using the PCMDI Metrics Package version 3 (Lee et al., 2023) and
 535 present the results using the “portrait plot” format described in Gleckler et al. (2008).
 536 One caveat for these results is that we only have one ensemble member for each model,
 537 though a single realization is generally reliable for assessing these climatological biases.
 538 As noted by Lee et al. (2023), the GFDL models ESM4.1 and CM4.0 are among the top-
 539 performing CMIP6 models for these atmospheric benchmarks, thus providing a useful
 540 benchmark for the CM4X simulations.

541 Both CM4X-p25 and CM4X-p125 make use of a finer grid for the GFDL-AM4.0
 542 atmosphere relative to CM4.0 (C192 versus C96), which we propose accounts for an over-
 543 all reduction in model biases in both CM4X configurations compared to ESM4.1 and CM4.0.
 544 Furthermore, note how CM4X-p125 exhibits improved performance relative to CM4X-

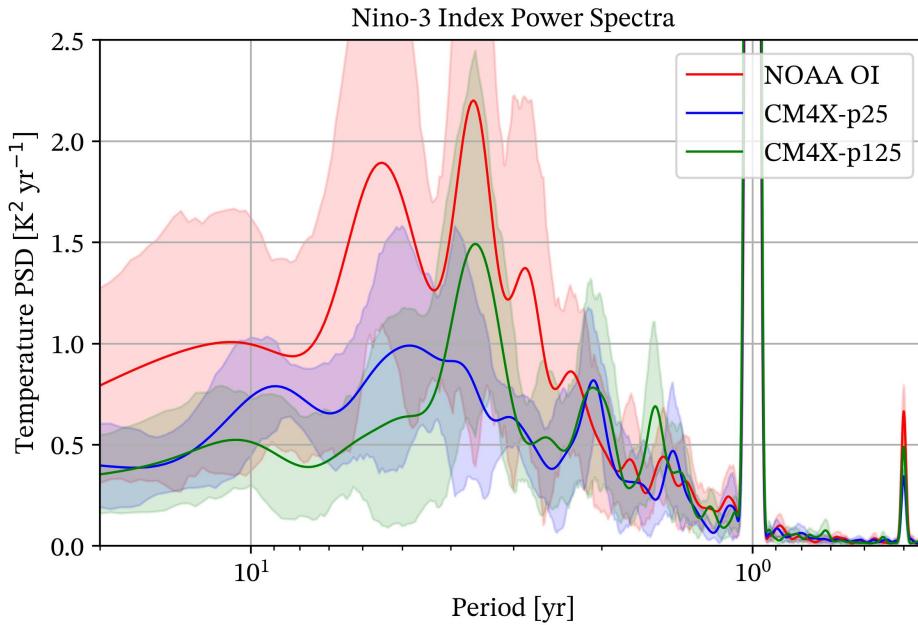


Figure 10. Power spectra for the Niño-3 index, which is computed from the monthly mean sea surface temperature area averaged over the region ($5^{\circ}\text{N} - 5^{\circ}\text{S}$, $150^{\circ}\text{W} - 90^{\circ}\text{W}$) in the Tropical Pacific. We show results from the CM4X simulations over years 1850-2014, as well the reconstructed NOAA temperatures over years 1854-2014 (Huang et al., 2017). The method of computation follows Welch (1967) as realized by the Matlab `pwelch` routine. We use 30-year windows for computing the periodogram and associated confidence intervals.

p25 in the mean mid-tropospheric atmospheric circulation (zg-500) and upper tropospheric temperatures (ta-200). This comparison exemplifies how coupling to the finer resolution ocean in CM4X-p125 can render benefits to the atmospheric simulation. Some seasonal differences are present, particularly in the boundary layer winds and temperatures in JJA. Treatment of the land model differs across ESM4.1, CM4.0, and CM4X (see Sections 2.4 and A2). CM4.0 uses an earlier version of the GFDL Land Model (LM4.0) with dynamic vegetation that responds to a changing climate, whereas ESM4.1 uses an updated version (LM4.1) that has advancements in hydrology and terrestrial vegetation geared towards Earth System Modeling (ESM) applications. CM4X, with its focus on exploring ocean questions, uses a version of LM4.0 with static vegetation (circa 1980). Land surface forcing differences also play a role in the coupled climate model performance (Ghimire et al., 2014; Zhao, 2022a), especially when evaluating model simulations of historical temperature trends. Despite these qualifications, the performance metrics in Figure 11, along with the broader assessment provided by Lee et al. (2023), suggest that CM4X historical simulations perform among the best-in-class for atmospheric metrics within the current generation of coupled climate models.

4 Ocean surface temperature properties

In this section we focus on the surface ocean temperature (SST) and related boundary fluxes.

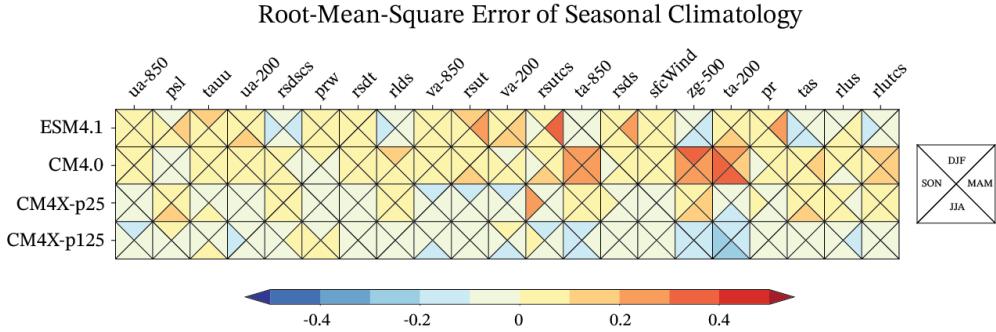


Figure 11. Performance metric portrait plot taken after Gleckler et al. (2008), here focused on atmospheric metrics from the ESM4.1, CM4.0, CM4X-p25, and CM4X-p125 historical simulations over years 1978-2014 of the historical simulations. The root-mean-square-error (RMSE) values are normalized across the models shown here, so we are showing relative error rather than absolute error. As per Table 1 in Lee et al. (2023), we compare the simulations to the precipitation from GPCP-2.3 (Adler et al., 2018); sea level pressure, temperatures, and winds from ERA5 (Hersbach et al., 2020); surface wind stress from ERA-Interim (Dee et al., 2011); and radiation from CERES-EBAF-4.1 (Loeb et al., 2018). Blue shading denotes improved model skill (relative to the models shown) while red shading denotes degraded model skill.

564

4.1 SST variability and turbulent heat fluxes

565
566
567
568
569
570
571
572
573
574
575
576

In Figure 12 we show one measure of transient variability in CM4X-p25 and CM4X-p125 is shown, here mapping the temporal standard deviation of the daily mean SST computed over years 1980-2009. Both models show the characteristic tropical Pacific pattern related to the El Niño-Southern Oscillation variability. In the middle latitudes, variability is largely concentrated near the western boundary currents, such as the Gulf Stream in the North Atlantic and Kuroshio in the North Pacific, with such variability largely associated with transient mesoscale eddies. Additional variations are seen in the Southern Ocean and are associated with mesoscale jets and eddies. In the polar regions, SST variability is muted due to the presence of sea ice and the freezing point lower bound on temperature. For each of the middle and high latitude patterns, CM4X-p125 shows a slight increase in the amplitude of the variability relative to CM4X-p25, along with an extension of the variability eastward away from the western boundaries.

577
578
579
580
581
582
583
584
585
586
587
588
589
590

Following Kirtman et al. (2012), we provide a measure of the impact of ocean variability on the atmosphere by computing the temporal correlations of the monthly SST anomalies with the monthly anomalies of the surface turbulent heat fluxes (sensible plus latent). These correlations are shown in Figure 13 for CM4X. As discussed by Kirtman et al. (2012) (and references therein), a negative correlation reveals regions where the ocean forces the atmosphere. Such regions are here seen to be closely tied to regions with large sea surface temperature variability. The western tropical Pacific is a notable region where the atmosphere forces the ocean, which corresponds to the region of strong tropical atmospheric convection. Kirtman et al. (2012) noted that for coarser models, such as the one-degree class of ocean models that do not admit mesoscale eddy fluctuations, most of the significant correlations are positive, indicating that the ocean is forced by the atmosphere. In contrast, the two CM4X simulations are dominated by correlations indicating that the atmosphere is forced by the ocean. Also observe that the polar regions show a large correlation and yet very little temperature variation. This be-

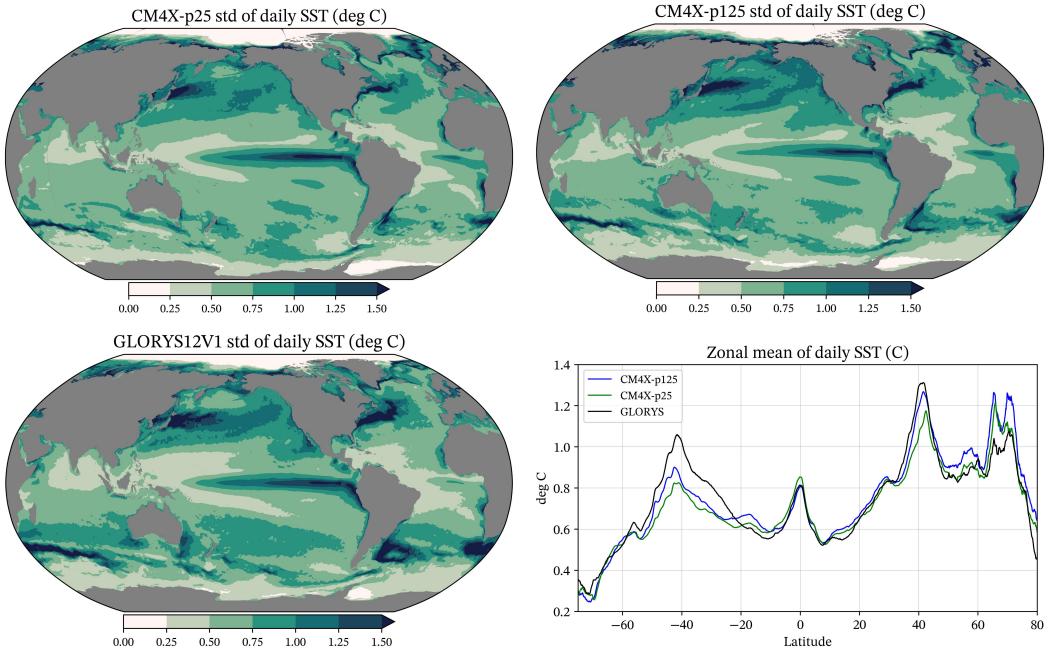


Figure 12. Standard deviation (degree C) of the sea surface temperature in CM4X-p25 and CM4X-p125 as computed over years 1980-2009, as well as the $1/12^\circ$ GLORYS12 analysis product (Lellouche et al., 2021) along with the zonal mean for the three. The standard deviation is computed by taking the mean squared difference of the daily mean SST from the climate mean for that day computed over years 1980-2009.

591 behavior reflects the dominance of sea ice formation/melt that produce latent heat fluxes
 592 yet little temperature changes.

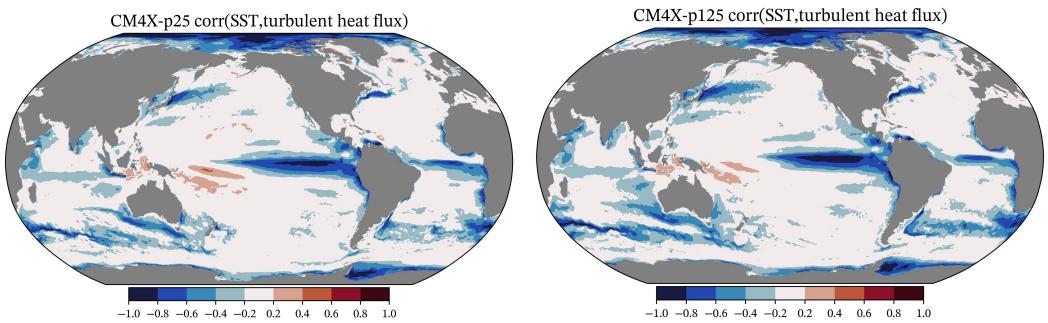


Figure 13. Temporal correlations between the monthly sea surface temperature and the turbulent surface heat fluxes (latent plus sensible). This correlation is computed over years 201-250 from the piControl simulations, with the SST and heat fluxes each computed relative to their climatological mean for each month. The sign convention is such that a positive surface heat flux warms the ocean and negative flux cools. We chose the color bar to emphasize correlations with amplitude larger than 0.2. These correlation maps can be compared to Figure 19 of Kirtman et al. (2012), though note that we chose the opposite sign convention for the heat fluxes.

593 **4.2 Historical sea surface temperature biases**

594 Figure 14 shows the sea surface temperature biases computed from the 30-year time
 595 mean of the simulations over years 1980-2009. There are many common biases across
 596 the models, such as the strong cold bias in the region east of Greenland and the warm
 597 bias in the Labrador Sea. We hypothesize that these North Atlantic biases arise, at least
 598 in part, from relatively poor representation of the mesoscale eddy processes and over-
 599 flow processes (R. Zhang et al., 2011). The eastern boundary regions off California as
 600 well as West Africa south of the equator both show warm biases that generally reflect
 601 biases in the stratocumulus clouds (e.g., see Wang et al. (2014); Richter (2015)). Notably,
 602 however, there is only a modest bias off the west coast of South America in the two CM4X
 603 simulations, which is significantly smaller than those found by Wang et al. (2014) and
 604 Richter (2015) in earlier climate models, and smaller than the CM4.0 bias in this region.
 605 This improvement is presumably due to the finer atmospheric grid in CM4X relative to
 606 CM4.0, though some studies suggest that finer ocean grid spacing can help reduce the
 607 stratocumulus cloud biases by enhancing coastal upwelling (Richter, 2015). We have more
 608 to say about these eastern boundary regions in Section 3 of Griffies et al. (2024).

609 Many of the SST biases in CM4X are broadly reflected in the CM4.0 biases, with
 610 the global root-mean-square bias of 0.92°C in CM4X-p25 close to the 0.90°C found in
 611 CM4.0. However, in nearly all regions in the World Ocean, SST biases in CM4X-p125
 612 are reduced relative to both CM4X-p25 and CM4.0, as reflected by the roughly 20% re-
 613 duction in the global root-mean-square bias (0.73°C for CM4X-p125) and 40% reduc-
 614 tion in the mean bias (-0.51° for CM4X-p25 and -0.29° for CM4X-p125). This bias re-
 615 duction is generally realized by a broadly warmer SST in CM4X-p125, thus reducing,
 616 for example, the broad-scale cool bias in the Southern Hemisphere middle latitudes. Fur-
 617 thermore, it is notable that both CM4X models display only modest biases in the South-
 618 ern Ocean relative to the larger warm bias in CM4.0 and even larger warm biases found
 619 in many other climate models (Sallée et al., 2013; Wang et al., 2014; Hyder et al., 2018;
 620 Beadling et al., 2020).

621 CM4X shows a general improvement with the overly cool eastern equatorial Pa-
 622 cific. This bias has been persistent in models, and is of importance due to its impact on
 623 the Hadley cell (G. Li et al., 2015). The Pacific subtropical gyre bias is also reduced, which
 624 supports a hypothesis from Burls et al. (2016) and Thomas and Federov (2017) that the
 625 equatorial and subtropical biases are connected. Finally, note the reduced bias in CM4X-
 626 p125 found in the western Atlantic near the U.S. coast, whereby the strong warm bias
 627 in CM4X-p25 (also seen in CM4.0) is largely absent in CM4X-p125. This reduced bias
 628 reflects differences in the Gulf Stream as represented in CM4X-p25 and CM4X-p125. We
 629 provide more analysis of the Gulf Stream in Section 6.2 of Part II (Griffies et al., 2024),
 630 where we identify a number of shortcomings in both CM4X models.

631 **4.3 Patterns of SST change under SSP5-8.5**

632 Figure 15 shows the change in SST under the SSP5-8.5 scenario. We compute this
 633 pattern as the difference between the final ten years of the scenario (years 2090-2099)
 634 and the 30-year mean over years 1980-2009 of the historical. There is broad agreement
 635 between CM4.0 and the two CM4X models, showing a general warming pattern in lower
 636 to middle latitude open ocean, with the exception of a muted warming in the southeast
 637 and south central Pacific. A muted southeast and south central Pacific warming has been
 638 seen in historical changes, with this pattern suggested to be related to changes in the
 639 Walker Circulation (Wills et al., 2022). However, a systematic comparison of the histor-
 640 ical SST trends would require multiple realizations for the historical period. The North
 641 Pacific shows general warming, with CM4X-p125 somewhat larger.

642 The Southern Ocean shows muted warming and some patches of cooling, with the
 643 exact positioning of these features likely dependent on decadal modes of climate vari-

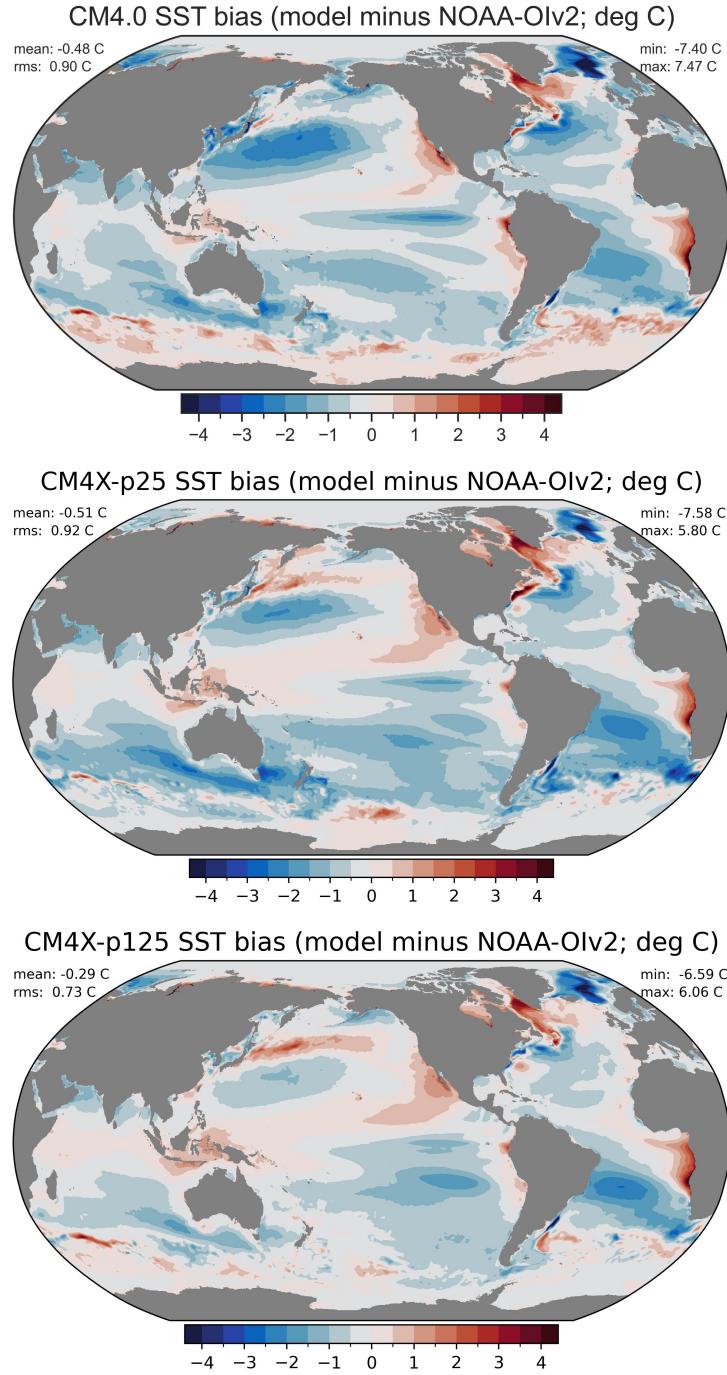


Figure 14. Sea surface temperature bias as defined by the difference between the model time mean over years 1980-2009 versus the NOAA OIv2 observational-based analysis (also averaged over 1980-2009) from Huang et al. (2020). The top panel shows CM4.0, whereas the lower two panels show CM4X-p25 and CM4X-p125. In each case, the NOAA analysis was interpolated to the model grid.

ability. This delayed Southern Ocean warming is related to the role of the deep ocean mixing, in which the Southern Ocean absorbs the bulk of anthropogenic heat and yet

644
645

most of that heat is transported to the north and exported to other ocean basins (Manabe et al., 1990, 1991; Gregory et al., 2016; Armour et al., 2016). The North Atlantic shows a distinctive tripole pattern, with warming next to the east coast of North America, cooling in the subpolar gyre, and warming to the east of Greenland. The cooling is related to a weakening of the Atlantic overturning circulation (Manabe et al., 1990, 1991; R. Zhang et al., 2019), whereas the coastal warming arises from associated changes in the Gulf Stream (Saba et al., 2016; Caesar et al., 2018). Note the larger warming off the American coast in CM4X-p125 relative to CM4X-p25, which is consistent with the CM2.5/CM2.6 results found by Saba et al. (2016), in which refined grid spacing produces both a more accurate simulation of coastal currents and a larger projection for coastal warming. Indeed, we find that CM4X-p125 shows increased SST warming relative to CM4X-p25 around the full North American continent (both Atlantic and Pacific, including the subpolar Pacific).

There is a rather large warming to the east of Greenland. This warming is over the anomalously cold region found in the historical experiment for CM4.0 and the two CM4X models (see Figure 14). This region also has anomalously large positive sea ice biases during the historical simulation (see Figure 8 in Part II (Griffies et al., 2024)). The sea ice melts back during the SSP5-8.5, which accounts for the large SST warming in Figure 15.

5 Ocean spin-up and historical/SSP5-8.5 response

In this section we build on the results from Section 3.4 by focusing on the timescales of equilibration in the piControl simulation. We also present a selection of enthalpy budgets from the historical and SSP5-8.5 simulations. As detailed in Section 3.1, the piControl starts from 21st century ocean observations of temperature and salinity (with zero ocean currents and sea ice spun-up from a previous simulation) and couples to an atmosphere with 1850 pre-industrial radiative forcing. Hence, the piControl experiment is an instantaneous cooling experiment. One may expect the piControl to expel the anthropogenic ocean heat and then equilibrate to a cooler volume mean ocean. Although much of the upper 1000 m of anthropogenic heat might be expected to release within $\mathcal{O}(100)$ years of pre-industrial forcing, it will presumably take far longer for deep ocean heat to be released. Even so, since the bulk of anthropogenic heat is in the upper 1000 m of the ocean, it is plausible that near thermal equilibration can arise on a timescale decoupled from the deep ocean. As shown in this section, this expectation is realized for the CM4X-p125 simulation, which reaches a cooler equilibrium state within 100-150 years, whereas the CM4X-p25 simulation drifts warm and has yet to thermally equilibrate (to the degree of CM4X-p125) even after 1000 years.

5.1 Volume mean Conservative Temperature

In Figure 16, we present the time series for annual mean and ocean volume mean Conservative Temperature from the piControl, historical, and SSP5-8.5 experiments for CM4X. The CM4X models display an initial cooling in their piControl simulations. Yet CM4X-p25 starts to warm after roughly 50 years, with a nearly linear warming trend thereafter. In contrast, CM4X-p125 cools for roughly 100 years and then stays relatively steady afterward for the duration of the experiment. Averaged over years 101-200, the net heat flux entering the ocean (positive flux enters the ocean) is 0.16 W m^{-2} for CM4X-p25. In contrast, for CM4X-p125 the imbalance is a mere -0.02 W m^{-2} , and this imbalance holds until the end of the piControl (we provide a more detailed global heat budget analysis in Section 5.6). Note that we use the ocean surface area to normalize these fluxes. Use of the earth's full surface area reduces the magnitude of the imbalance by roughly 30%.

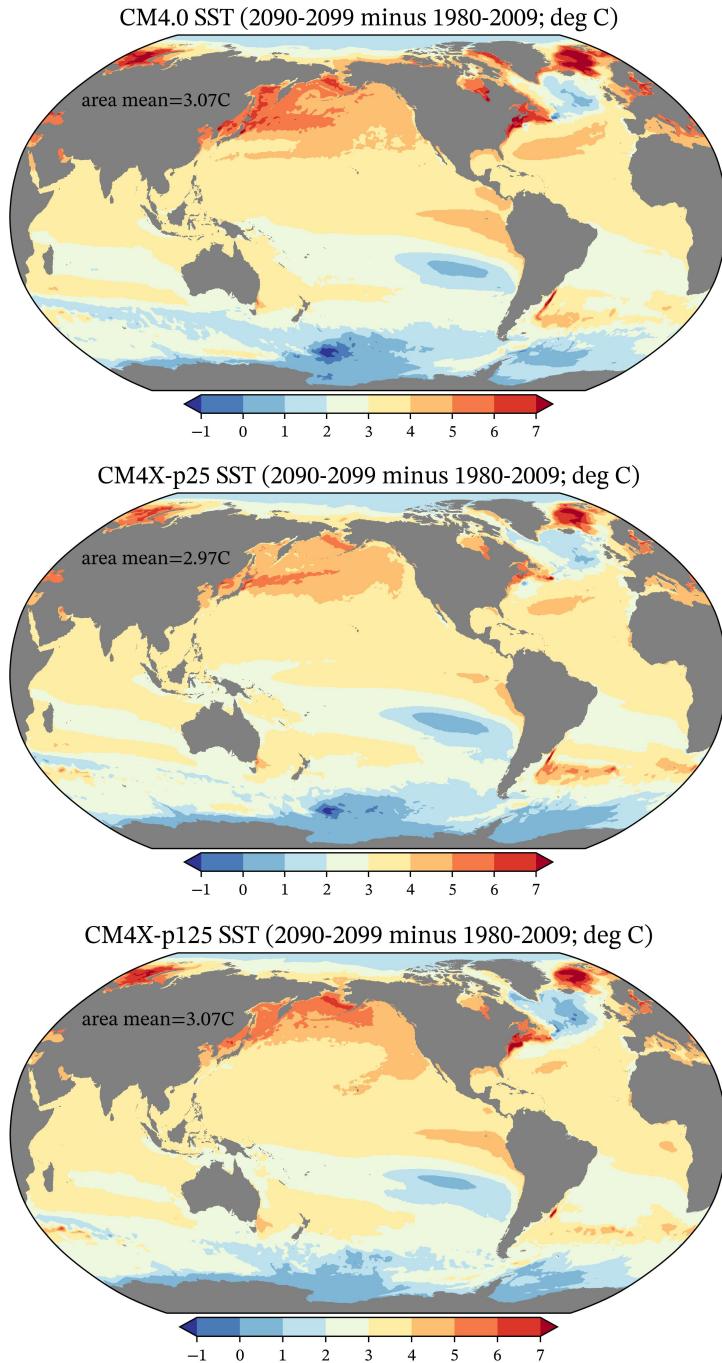


Figure 15. SST change under the SSP5-8.5 scenario, computed as the difference between the ten year mean from years 2090-2099, versus the 30-year mean of the historical 1980-2009. We insert the area SST change average over Asia. The top panel shows CM4.0 and lower panels show CM4X-p25 and CM4X-p125. Note the non-symmetric color scales, with only the dark blues reflecting cooling. Use of the final ten years of the SSP5-8.5 scenario (rather than, say, the final 30 years) enhances the signal from climate change, and it is commonly used for such purpose; e.g., Newsom et al. (2023).

695 The CM4X models include a prescribed geothermal heat flux at the ocean bottom
 696 with a global ocean area mean of 0.094 W m^{-2} (see Figure A2). After approximately
 697 100 years in the piControl simulation, the area integrated geothermal heat flux entering
 698 the ocean bottom in CM4X-p125 is nearly balanced by an equivalent heat flux leaving
 699 the ocean surface, thus leading to a very small residual. In contrast, CM4X-p25 re-
 700 alizes a net positive heating and thus a steady rise in global mean ocean Conservative
 701 Temperature. As a point of comparison, the CM2.5 (0.25° ocean) and CM2.6 (0.1° ocean)
 702 models from the CM2-O hierarchy (Griffies et al., 2015) have a far larger pre-industrial
 703 warming drift of roughly 1.5 W m^{-2} (CM2.5) and 0.5 W m^{-2} (CM2.6) during the sec-
 704 ond century in their piControl simulations.

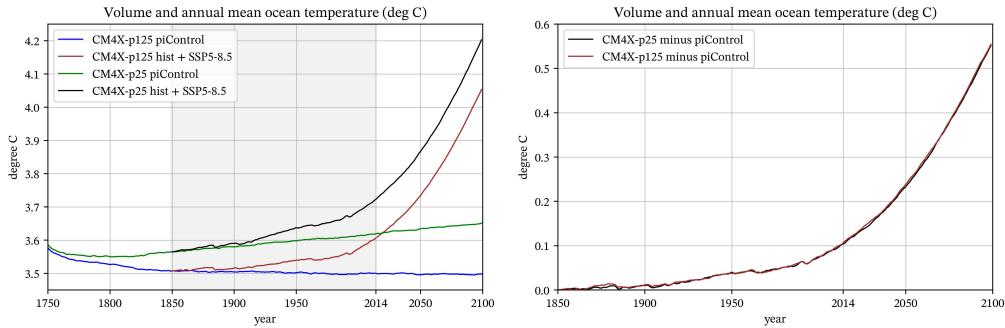


Figure 16. Left panel: time series for the volume mean and annual mean ocean Conservative Temperature for CM4X-p25 and CM4X-p125 when run for the piControl, historical (1850–2014 shaded), and SSP5-8.5 from 2014–2100. The piControl calendar years are nominal, here set to allow for alignment with the historical and SSP5-8.5 experiments. For the piControl experiments, both CM4X models are initialized from the 2013 World Ocean Atlas (Section 3.1), with slight disagreement between CM4X-p125 and CM4X-p25 at the initial time arising from the differences in ocean geometry. There are no "hidden" simulation years prior to those shown here. The historical simulations branch from the piControl from year 101 and run from 1850 to 2014 (gray shaded region). The SSP5-8.5 then starts from the historical at year 2015. Right panel: Historical and SSP5-8.5 time series for CM4X-p25 and CM4X-p125, relative to their respective piControl simulations shown in the left panel. Removing the piControl allows for time series for both models collapse on top of one another.

705 The historical simulations shown in Figure 16 branch at year 101 of the piControl
 706 and then run from year 1850 to 2014. A perfect historical simulation will return the ocean
 707 mean temperature to its 2013 initial condition when the simulation reaches year 2013.
 708 Since CM4X-p25 warms during its piControl, its 1850 initial condition is already at the
 709 2013 temperature, with the result being a simulated 2013 global mean temperature roughly
 710 0.15 K warmer than the 2013 measurements. This result has implications for state-dependent
 711 transient responses, such as those discussed in Stouffer et al. (2006) as well as for sea level
 712 rise (Hallberg et al., 2013). In contrast to CM4X-p25, Figure 16 shows that CM4X-p125
 713 cools during the first 100 years of its piControl, so that its simulated 2013 ocean tem-
 714 perature is only about 0.03 K warmer than 2013 measurements. Based on the further
 715 slight cooling from years 101–150 of the CM4X-p125 piControl, initializing the histor-
 716 ical at year 150 rather than year 100 could have reduced the already small CM4X-p125
 717 overshoot of 0.03 K to an even smaller value.

718 The SSP5-8.5 simulations shown in Figure 16 extend from year 2014 to year 2100.
 719 Removing their respective piControls (right panel of Figure 16), collapses the CM4X-

720 p25 and CM4X-p125 historical and SSP5-8.5 curves on top of one another, thus revealing
 721 the very similar climate sensitivities between the two CM4X models.

722 5.2 Area mean SST

723 In Figure 17, we show the time series for the area-mean SST. For the piControl sim-
 724 ulations, both CM4X models reach a statistically stationary state after roughly 100 years,
 725 with CM4X-p125 approximately 0.25°C warmer than CM4X-p25. The warmer SST in
 726 CM4X-p125 was indicated earlier by the SST bias patterns in Figure 14. For the SSP5-
 727 8.5 climate change simulation, the SST rises by roughly 3°C from 2014 to 2100, with both
 728 models exhibiting a nearly parallel increase. There is a notable cool bias in both CM4X
 729 models found during years 1950-2014. This cool bias is also seen in CM4.0 (see Figure
 730 12 in Held et al. (2019)), and it has been related to aerosol effects by Zhao et al. (2018b).

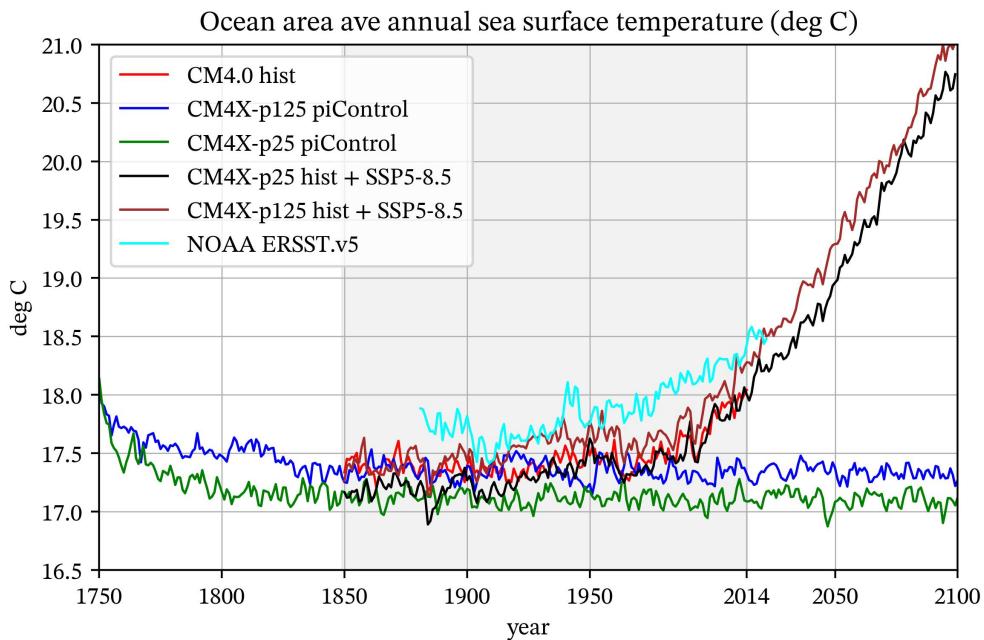


Figure 17. Time series for the ocean area mean annual sea surface temperature as in Figure 16. The historical simulations branch from the piControl and run from year 1850 to 2014 (gray shaded region). The cooling blips in the historical simulation result from volcanic aerosol forcing. The SSP5-8.5 scenario extends from year 2014 to 2100. We also show the NOAA ERSST v5 (Huang et al., 2017) (cyan) starting at year 1880.

731 5.3 Depth-time Conservative Temperature: World Ocean

732 We expose more details about the Conservative Temperature evolution in Figure
 733 18, which shows the depth-time diagram for the global horizontally averaged annual mean
 734 temperature in the piControl simulations. Both models cool in the upper ocean, consis-
 735 tent with the SST in Figure 17 and consistent with the pre-industrial forcing. However,
 736 both also show a warming trend in intermediate depths, and slight cooling again in the
 737 abyssal regions. The upper and intermediate depth trends are enhanced in CM4X-p25
 738 relative to CM4X-p125, thus leading to the cooler surface temperature and warmer in-
 739 termediate temperature in CM4X-p25. As for the global volume mean time series in Fig-

ure 16, it is remarkable how stationary CM4X-p125 is for all depths after roughly 100 years of the piControl.

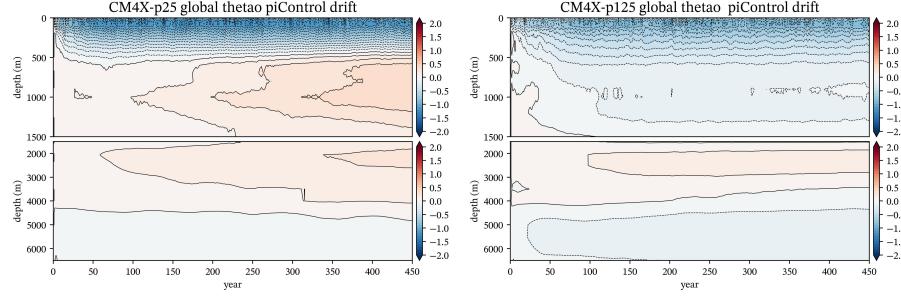


Figure 18. Depth-time drift for the global horizontally averaged annual mean Conservative Temperature. We show temperature evolution relative to the initial year as realized for the piControl simulations, computed as $\theta_{\text{drift}}(z, t) = \sum_{xy} \Delta V [\theta(x, y, z, t) - \theta(x, y, z)_{\text{init}}(x, y, z)] / \sum_{xy} \Delta V$, where $\theta(x, y, z)_{\text{init}}(x, y, z)$ is the annual mean of the first simulation year, and ΔV is the volume of a grid cell. Note the split between the upper ocean and deeper ocean, though with both regions having the same color range, with this range set to correspond to the basin time series shown in Figure 20.

Figure 19 plots centennial scale vertical profiles of the horizontally averaged Conservative Temperature, relative to the initial year. This figure emphasizes the warming found in CM4X-p25 below around 500 m depth and until roughly 4000 m. Furthermore, note how the vertical temperature anomaly in CM4X-p25 grows over the four centuries, partly due to the interior continuing to warm and partly due to the enhanced cooling of the upper ocean after the first century. An upper ocean cooling and interior warming results in a reduction in the surface-to-interior temperature difference. The volume mean ocean warming in CM4X-p25 is a result of a net heat going from the atmosphere into the ocean (see Figure 7 and the heat budget in Section 5.6). The thermal adjustment in the CM4X-p125 piControl is fundamentally distinct from CM4X-p25, in which CM4X-p125 locks into a nearly stationary vertical temperature profile during the second century and afterward, from the surface down to around 3000 m, with cooling down to around 1500 m, warming between 1500 m to 3500 m, and slight cooling below.

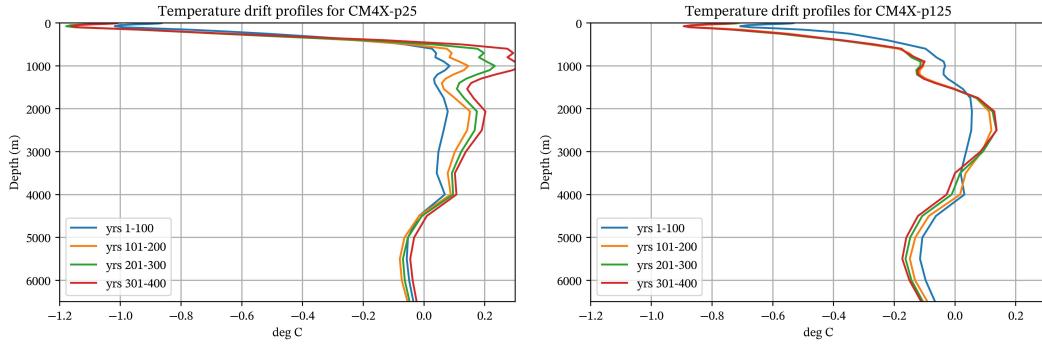


Figure 19. Vertical profiles of horizontally averaged Conservative Temperature relative to the initial year, computed as a time mean for years 1-100, 101-200, 201-300, and 301-400. The left panel shows CM4X-p25 and right panel shows CM4X-p125.

We interpret these results according to the mesoscale eddy effects described by Griffies et al. (2015), whereby mesoscale eddies regulate the drift in water mass properties through their systematic restratification effects (and corresponding shoaling of the pycnocline) and leading to a net upward transport of buoyancy (when integrated horizontally over the globe). Griffies et al. (2015) based this interpretation on simulations with CM2.5 and CM2.6, which are coupled climate models from an earlier generation of GFDL models, with CM2.5 and CM2.6 differing only via their ocean horizontal grid spacing (CM2.5 has a 0.25° grid whereas CM2.6 has a 0.10° grid). As with the CM4X models, the volume mean thermal drift in CM2.6 is significantly smaller than CM2.5.

In an ocean with relatively weak mesoscale eddies, such as CM4X-p25 and CM2.5, the upper ocean boundary forcing is only modestly offset by eddy transport. Surface cooling from the piControl atmosphere reduces the vertical temperature gradient, allowing for more mixing between the relatively warm upper ocean and the cooler interior ocean and accounting for the interior (below 500 m) warm drift. The atmosphere, in response to the relatively cool SSTs in CM4X-p25, enhances the heat going into the ocean surface, which in turn leads to even more heat entering the ocean interior even while the SSTs remain relatively cool.

The story in CM4X-p125 is distinct due to the stronger role of mesoscale eddy transport, whereby stronger eddies act to limit the surface cooling and interior warming after the first century. That is, the surface cooling during the first few decades of the CM4X-p125 piControl is met by an upward transport of ocean heat (presumably dominated by anthropogenically induced heat contained in the initial conditions), thus leading to a cooling of the CM4X-p125 ocean down to around 1500 m. In turn, the eddies partially counteract the destratification affected by the surface cooling. Once the bulk of the heat is released after $\mathcal{O}(100)$ years, the ocean temperature profile remains nearly stationary from the surface down to around 3000 m. This stationarity is matched by a tiny net heat crossing the ocean boundaries, with ocean heat gained at the bottom through geothermal heating (roughly $+0.07 \text{ W m}^{-2}$ averaged over the ocean), and ocean heat lost at the surface of nearly the same magnitude as the geothermal heating (we provide more details of the global heat budget in Section 5.6). In this manner, the CM4X-p125 ocean reaches near thermal equilibrium with the pre-industrial atmosphere after $\mathcal{O}(100)$ years.

5.4 Depth-time Conservative Temperature: basin averages

The depth-time patterns for the World Ocean temperature adjustments seen in Figure 18 arise from an array of water mass adjustments within and between ocean basins. Figure 20 exposes further structure by decomposing the adjustment into basins, and we also include results from CM4.0. This figure reveals that across all basins, CM4X-p125 has a muted adjustment amplitude relative to both CM4X-p25 and CM4.0, and CM4X-p125 shows a quicker equilibration to a quasi-stationary state. Furthermore, the global adjustment of temperature is dominated by the Pacific and, to some degree, the Indian basins (here shown combined), which comprise the bulk of the World Ocean volume. For the Atlantic and Southern Oceans, the surface cooling penetrates to roughly 1000 m, which contrasts with the more surface-focused cooling (upper 100 m) in the Indian and Pacific basins. The deeper penetration for the Atlantic and Southern Oceans reflects the presence of deep ventilation that allows for the release of heat from deeper waters.

CM4X-p25 and CM4.0 show a warming between 500 m-4000 m in the Atlantic/Arctic basins, whereas CM4X-p125 shows a warming starting around 1000 m or deeper. We suspect that the deep subsurface warming pattern, which also penetrates into the Southern Ocean, originates from a strengthened overturning circulation (due to the surface cooling) that brings more of the relatively warm upper ocean waters into the abyss. Finally, the Arctic basin (when shown separately) has a slight cooling in the upper 200m and a

805 larger deep warming signal below 500m in all three models. Ongoing research suggests
806 that this adjustment is related to ocean mixing biases.

807 As emphasized in Section 2.4, the comparison between CM4.0 and CM4X is not
808 clean, since there are many details across the models that differ. Even so, by exhibiting
809 the CM4.0 results in Figure 20, we are able to better gauge the improvements in CM4X-
810 p125 within the ocean interior. One further point of note concerns the hint of a centen-
811 nial time scale variations in the Southern Ocean, with such variations present in the up-
812 per 1000 m across the three models. These variations are also noticeable in the deeper
813 ocean, particularly in CM4.0. We return to such signals in Section 5 of Part II (Griffies
814 et al., 2024). In that discussion we note that this signal corresponds to Southern Ocean
815 polynya activity, with the CM4X signals far smaller and more realistic than the super-
816 polynyas found in CM4.0 and discussed by Held et al. (2019); Dunne et al. (2020); L. Zhang
817 et al. (2021).

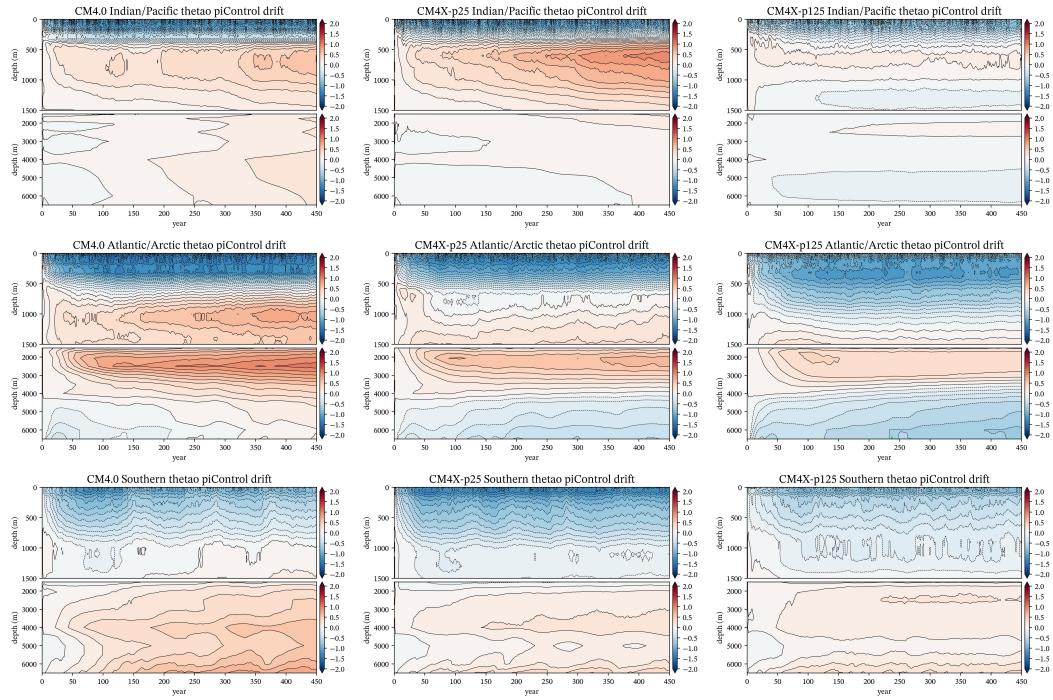


Figure 20. Depth-time adjustment for the horizontally averaged annual mean Conservative Temperature, as in Figure 18, but here decomposed into basins (Indian/Pacific, Atlantic/Arctic, and Southern) and showing CM4.0 (left column), CM4X-p25 (middle column), and CM4X-p125 (right column). Also note the expanded color range relative to Figure 18.

818 5.5 Depth-time salinity: World Ocean

819 The bulk of our study in this paper concerns temperature and heat. To check that
820 our focus is not missing something revealed by salinity, we display in Figure 21 the depth-
821 time adjustments for salinity, here including CM4.0 as well as the two CM4X models.
822 The near surface generally sees an increase in salinity during the first few years, as ex-
823 pected from the increases in sea ice due to the return to pre-industrial forcing in the pi-
824 Control. Below the surface salinity increase, and reaching down to a few hundred me-
825 ters, CM4X-p25 shows an initial freshening that gradually dissipates in favor of a grow-
826 ing salinification in the region between 500 m and 1000 m. In contrast, both CM4.0 and

CM4X-p125 show far less salinity increase in this depth range. In the deeper ocean in each model shows a gradual increase in salinity between 1500 m and 4000 m, and a muted freshening along with some salinification below.

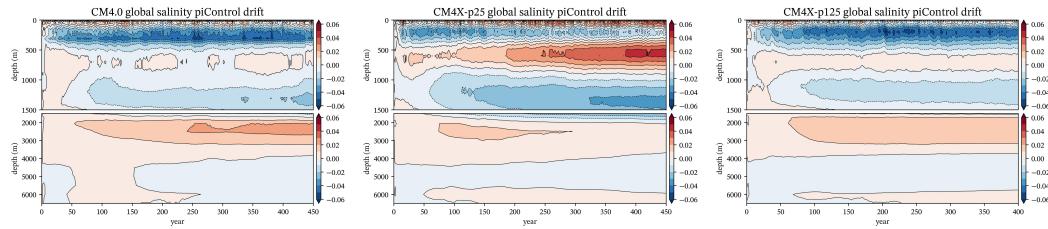


Figure 21. As in Figure 18, yet here for the salinity averaged over the World Ocean rather than temperature, and here including CM4.0 as well as CM4X-p25 and CM4X-p125. Units are g/kg or parts per thousand.

Fleshing out details for the salinity drift patterns requires a detailed analysis that is beyond our scope. The key conclusion we take is that CM4X-p125 shows noticeably less drift than either CM4.0 or CM4X-p25 after the initial 100-150 years. We thus find that the salinity drift patterns are consistent with those seen in temperature, further supporting the conclusion that there is a reduced drift in CM4X-p125 relative to CM4X-p25.

5.6 Global heat (enthalpy) budget

Global diagnostics of the piControl enthalpy budget are shown in Figure 22, with these budgets further illustrating how well CM4X-p125 has thermally equilibrated relative to CM4X-p25 and CM4.0. Thermal equilibrium for the ocean is realized by balancing the net heat flux crossing the ocean boundaries, both through a constant (in time) geothermal heat flux entering the ocean at the bottom and a net flux crossing the air-sea interface. The area averaged flux across the air-sea interface is directed into the ocean for CM4.0 and CM4X-p25, with CM4X-p25 having approximately half the flux as CM4.0. For CM4X-p125, the net air-sea heat flux is directed out of the ocean, and this outward flux closely balances the geothermal heat flux entering the ocean bottom. The result for CM4X-p125 is a global ocean enthalpy (heat content) tendency that is roughly 10 times smaller than either of the two net boundary fluxes.

The top-of-atmosphere heat flux, F_{TOA} , is relatively large in CM4.0, intermediate in CM4X-p25, and nearly zero in CM4X-p125. This result follows from the heat budgets for the ocean and the dominance of the ocean in establishing the global enthalpy balance. Note that the CM4.0 top-of-atmosphere heat flux is close to the CMIP6 median (Irving et al., 2021). The top-of-atmosphere fluxes are consistently higher than the ocean heat content tendencies by about $0.01\text{--}0.03 \text{ W m}^{-2}$. A mismatch between F_{TOA} and dQ/dt is a common feature among climate models and indicative of enthalpy leakage. Indeed, among CMIP6 models the mismatch can reach values of $0.5\text{--}1 \text{ W m}^{-2}$ (Irving et al., 2021).

Coupled climate models face a number of difficulties in conserving enthalpy, both through spurious dissipation, coupling inconsistencies, and missing processes (Lauritzen et al., 2022). In particular, we find in CM4.0 and CM4X a net loss of heat of about 0.17 W m^{-2} from the global climate system, attributable to the lack of temperature tracked by the atmospheric model's water causing a leak of energy through the oceanic precipitation and evaporation. The magnitude of this enthalpy sink is similar to other generations of GFDL models (e.g., see Section 3 in Delworth et al. (2006)). There are limitations of our

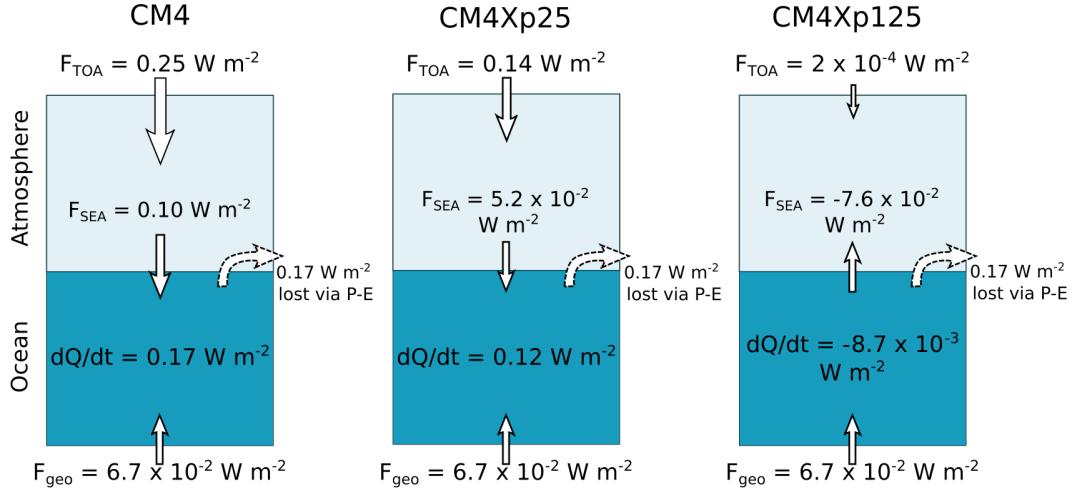


Figure 22. Schematics of the global heat budget for the piControl experiments from CM4.0 (left), CM4X-p25 (center), and CM4X-p125 (right), with all numbers obtained from time averaging over years 151-350, and all areas referring to the global earth surface area rather than ocean surface area. Schematics show the net top-of-atmosphere heat flux, F_{TOA} , air-sea heat flux, F_{SEA} , geothermal heat flux, F_{geo} , and loss of heat from the water cycle, F_{P-E} , all normalized by global surface area. The heat content tendency of the ocean, dQ/dt , is also normalized by global (rather than ocean) surface area. The sign notation is positive for fluxes into the Earth system or fluxes from the atmosphere into the ocean.

heat budget analysis, including the absence of terrestrial terms (including the terrestrial component of the enthalpy sink due to the hydrological cycle), nor have we considered changes in the heat balance from transience in glacier snow or soil. However, the residual of the three heat flux terms and the ocean heat content tendency is

$$F_{TOA} + F_{GEO} + F_{P-E} - dQ/dt \approx -0.07 \text{ W m}^{-2}, \quad (1)$$

which is consistent across all model setups. The similar differences across the models suggests that changes to the atmospheric and terrestrial model components play a secondary role in setting the enthalpy balance differences on climate timescales across the models. Instead, the results suggest that changes in the ocean model vertical enthalpy transport are dominant in setting the global response.

5.7 Ocean heat uptake efficiency and pycnocline depth

In Figure 23 we show the depth-time diagram for the horizontally averaged annual mean Conservative Temperature, now computed for the historical and SSP5-8.5 simulations and with differences computed relative to their respective piControl years 101-350. Warming in the upper ocean starts around year 2000 and penetrates into the deeper ocean throughout the 21st century. The heat uptake patterns are notably similar for the two models, which is consistent with the very close time series of global mean temperatures shown in the right panel of Figure 16.

To quantify the ocean heat uptake depicted in Figure 23, as well as the SST change in Figure 15, we compute the ocean heat uptake efficiency (OHUE). Following the definition in Newsom et al. (2023), we estimate OHUE as the change of global ocean heat uptake (in W m^{-2}) divided by the change of global mean SST (in Kelvin) as computed over years 2090-2100 of the SSP5-8.5 versus years 1850-1890 of the historical simulation.

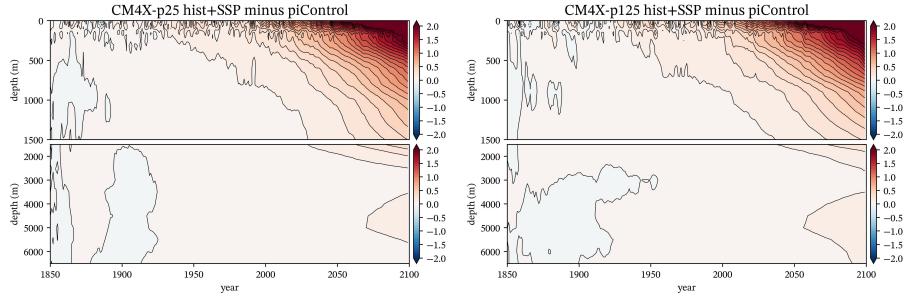


Figure 23. Depth-time drift for the horizontally averaged annual mean Conservative Temperature from CM4X-p25 (left panel) and CM4X-p125 (right). Here we show the historical and SSP5-8.5 simulations (years 1850–2100) relative to years 101–350 of the piControl; i.e., $\theta_{\text{anom}}(z, t) = \sum_{xy} \Delta V [\theta(x, y, z, t) - \theta(x, y, z, t)_{\text{piC}}] / \sum_{xy} \Delta V$, where $\theta(x, y, z, t)_{\text{piC}}$ is from the piControl. Note the split between the upper ocean and deeper ocean.

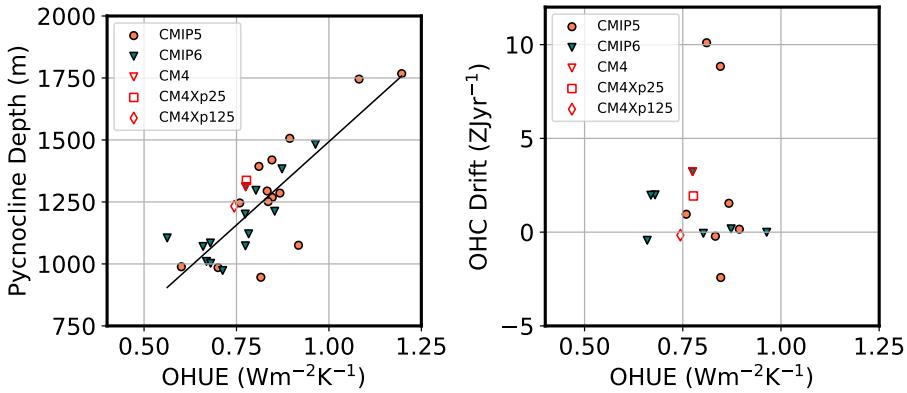


Figure 24. Left panel: ocean heat uptake efficiency (OHUE) versus pycnocline depth in CM4X models, with the values of CMIP5/6 models reproduced from Newsom et al. (2023) and with OHUE and pycnocline depth computed as per Newsom et al. (2023). The pycnocline depths are computed during the 80-year periods corresponding to branching of the historical simulations, so that the two CM4X pycnocline depths are based on piControl years 101–180 whereas the CM4.0 depth is from its piControl years 251–330. We find that the averaging period does not significantly alter results for the pycnocline depth. Right panel: Ocean heat uptake efficiency (OHUE) versus ocean heat content (OHC) drift in CM4X models, with the values of CMIP5/6 models reproduced from Newsom et al. (2023) and Irving et al. (2021). The OHUE is computed as per Newsom et al. (2023) and the OHC drift is computed as per Irving et al. (2021). The OHC drifts are computed as the linear trends of OHC change based on piControl years 151–350 (as in Figure 22). We find that the averaging period does not significantly alter results for the OHC drift. We also show only the subset of CMIP5/6 models analyzed by both Newsom et al. (2023) and Irving et al. (2021).

As seen in Figure 24, relative to the CMIP5/6 intermodel spread found in Newsom et al. (2023), the spread in OHUE is small among CM4X-p25, CM4X-p125, and CM4.0. Consistently, we find the difference in preindustrial pycnocline depth (averaged over model years 101–180) among the CM4X and CM4.0 models is also negligible compared to the CMIP5/6 intermodel spread (Figure 24). This result verifies that OHUE, as well as its

892 relationship with pycnocline depth across CM4X and CMIP5/6 models, is not sensitive
 893 to the simulated global ocean temperature drift and the degree of equilibration of the
 894 preindustrial control run (see right panel of Figure 24).

895 6 Pre-industrial thermal equilibration in a mesoscale dominant regime

896 Thermal equilibration of the piControl experiment is typically a long and resource
 897 intensive process, taking $\mathcal{O}(1000)$ simulation years. We thus ask whether the relatively
 898 rapid (centennial rather than millennial) thermal equilibration of CM4X-p125 is a one-
 899 off result? Or is it the signature of a new regime in climate modeling? Rather than defini-
 900 tive answers, we offer indirect evidence that such behavior is expected as models improve
 901 their physical and numerical fidelity. In so doing, we present some context for the ques-
 902 tion of piControl thermal equilibration and then present a hypothesis to help guide fur-
 903 ther investigations.

904 Following the Irving et al. (2021) study of climate model drift, we acknowledge the
 905 many reasons for piControl drift in the CMIP5/6 climate models. Even so, with CM4X-
 906 p25 and CM4X-p125 having distinct drift behaviors and yet differing solely in their ocean
 907 grid spacing, these two models offer a clean venue to study drift mechanics related to
 908 the differing representation of ocean heat transport.

909 6.1 Drift in the piControl: early climate models

910 During the early decades of climate modeling, it was necessary to augment mod-
 911 eled air-sea fluxes with *a priori* diagnosed and non-interactive fluxes aimed at reducing
 912 the otherwise extremely large model drift. The papers by Sausen et al. (1988) and Manabe
 913 and Stouffer (1988) detail these flux adjustment methods. Without flux adjustments, these
 914 early generation of climate models manifested a sizable drift that made it difficult, if not
 915 impossible, to study climate mechanics since the model drift aliased onto and/or directly
 916 impacted climate variability and change.

917 Climate model drift reduced as models improved their numerics and physics, no-
 918 tably through the representation of poleward ocean heat transport (Weaver & Hughes,
 919 1996). Such advances allowed models to realize sensible climate equilibria without flux
 920 adjustments (e.g., Boville and Gent (1998); Gordon et al. (2000); Delworth et al. (2006)).
 921 Eliminating flux adjustments was a milestone in the history of climate modeling. Even
 922 so, climate models still contain biases and limitations that are fully exposed when re-
 923 moving flux adjustments. So although the current generation of climate models has far
 924 smaller biases than the early generation, they still generally struggle to realize equilibrated
 925 climate states close to estimates for the 1850 climate, and they reach such equilibrated
 926 states only after $\mathcal{O}(1000)$ years (e.g., CM4X-p25, Stouffer (2004), Banks et al. (2006),
 927 and Irving et al. (2021)). For the CMIP5/6 climate models, the spread across climate
 928 model drift is the main cause for model-to-model differences as well as model-to-observation
 929 differences.

930 6.2 Drift in the piControl: state of the science climate models

931 Since the ocean provides the dominant heat reservoir for the Earth's climate sys-
 932 tem, we gauge the degree of climate model thermal equilibration by measuring trends
 933 in ocean temperature and/or the net boundary fluxes. We then claim "success" when
 934 the ocean trends are smaller than a subjective criteria. If computer cost (and energy foot-
 935 print) are of no concern, then we can reach thermal equilibrium by integrating the model
 936 for as long as it takes (Krasting et al., 2018; Rugenstein et al., 2020). Unfortunately, this
 937 brute force approach often leads to an equilibrated climate model with an ocean state
 938 that is far from a reasonable 1850 pre-industrial ocean. Indeed, in many cases, such as
 939 CM4X-p25, the volume mean ocean is warmer in the equilibrated 1850 piControl than

940 the present-day initial conditions. If there is one thing we know about the 1850 ocean,
 941 it is that it had a lower volume mean temperature than the 21st century ocean. We thus
 942 reach the conflicted situation whereby the more equilibrated a climate model, the fur-
 943 ther the ocean deviates from properties corresponding to the real ocean. This situation
 944 presents difficult choices for model development teams, and offers little inspiration for
 945 ocean analysts to study simulated water masses that are corrupted by huge drifts.

946 For a variety of practical reasons that include limitations on computational and hu-
 947 man resources, many models are simply not run to complete thermal equilibrium, so that
 948 the study of historical and future scenarios requires a form of drift removal, such as re-
 949 alized by subtracting the pre-industrial control simulation. Unfortunately, drift removal
 950 makes it difficult to study water mass properties and their relation to the observed ocean,
 951 and it does little to facilitate model-to-model comparisons of interior ocean properties
 952 since drifts across models can be quite different (Séférian et al., 2016; Irving et al., 2021).
 953 Indeed, some drift cannot be removed, such as discussed by Krasting et al. (2024) in the
 954 context of Southern Ocean mode and intermediate water masses and their imprint on
 955 regional sea level.

956 Even for presumably linear signals, there are caveats about the utility of drift re-
 957 moval. For example, Hallberg et al. (2013) questioned the linearity assumption for global
 958 thermosteric sea level projections since ocean density is a nonlinear function of temper-
 959 ature and salinity. As they found, extreme levels of model drift can have a notable role
 960 in sea level projections. A linearity assumption also becomes untenable when investigat-
 961 ing transitions between quasi-equilibrium states, such as hypothesized for the Atlantic
 962 overturning circulation. Indeed, the review paper by Hirschi et al. (2020) suggests that
 963 Atlantic overturning stability in models is dependent on grid spacing, with finer grid mod-
 964 els offering a more realistic depiction of the circulation pathways and, by inference, the
 965 stability properties. Studies of overturning stability thus motivate the use of both fine
 966 grid models and models that are equilibrated.

967 6.3 Time scale for piControl thermal equilibration

968 Consider a climate model with perfect dynamical core, perfect physical parame-
 969 terizations, and perfect numerics. If this model were subject to piControl forcings and
 970 initialized with perfect preindustrial initial conditions, it would presumably remain very
 971 close to its initial condition in perpetuity (ignoring that even 1850 has long-term climate
 972 drift due to centennial scale natural and anthropogenic variability). Instead, suppose this
 973 model is initialized with an exact rendition of present-day ocean temperature and salin-
 974 ity fields. What is the time scale for removing anthropogenic heat from the ocean in a
 975 piControl experiment? Since the bulk of the anthropogenic heat entered the ocean in the
 976 period after 1850, and since it is largely contained in the upper 1000 m of ocean, one hy-
 977 pothesis is that the time scale for its release is also $\mathcal{O}(100)$ years. This hypothesis is tem-
 978 pered by realizing that ocean heat uptake and heat release are not symmetric processes
 979 (Stouffer, 2004; Pudig et al., 2023). Furthermore, there is some heat sequestered in the
 980 deep ocean (Purkey & Johnson, 2010, 2012, 2013; Johnson & Purkey, 2024) through up-
 981 per ocean freshening and corresponding enhanced stratification, rather than the injec-
 982 tion of anthropogenic heat. It is connected to the efficiency by which heat can make its
 983 way to the abyssal ocean in regions not influenced by dense water formation, with that
 984 time affected by vertical mixing, either physically informed or spurious. Given these caveats,
 985 we find that the CM4X-p125 result, which equilibrates to a cooler piControl state than
 986 present-day, supports the hypothesis that the time scale is closer to $\mathcal{O}(100)$ years than
 987 $\mathcal{O}(1000)$ years.

988 During the post-1971 period, IPCC estimates (see Box 3.1 of Rhein et al. (2013))
 989 suggest that the ocean has accumulated roughly 3×10^{23} J = 300 ZJ of anthropogenic
 990 heat, and Zanna et al. (2019) suggest that it has accumulated roughly 400 ZJ since 1870.

991 A global ocean mean boundary heat flux, Q^{heat} , leads to an ocean volume mean temperature trend of roughly (see Appendices A and C of Griffies et al. (2014) for details of equation (2))
 992
 993

$$994 \Delta\Theta/\Delta t = Q^{\text{heat}}/(\rho_o C_p^0 H), \quad (2)$$

995 with $H \approx 3600$ m the ocean volume divided by the ocean surface area, $\rho_0 \approx 1035$ kg m⁻³
 996 the mean ocean density, and $C_p^0 \approx 3990$ J °C⁻¹ kg⁻¹ ocean heat capacity. As a useful
 997 benchmark, note that $Q^{\text{heat}} = 1$ W m⁻² corresponds to an ocean global volume mean
 998 temperature trend of $\Delta\Theta/\Delta t \approx 0.2^\circ\text{C century}^{-1}$. Figure 7 indicates that the CM4X-
 999 p125 simulation cools by roughly 0.07 K during the first 150 years of its piControl sim-
 1000 ulation, which represents an ocean heat release of

$$1001 V \rho_o C_p^0 \Delta\Theta \approx (1.3 \times 10^{18} \text{ m}^3) (1035 \text{ kg m}^{-3}) (4000 \text{ J kg}^{-1} \text{ K}^{-1}) (0.07 \text{ K}) \approx 390 \text{ ZJ}. \quad (3)$$

1002 This number accords with the 400 ZJ estimate from Zanna et al. (2019), in which case
 1003 we infer that the ocean heat released during the CM4X-p125 piControl directly corre-
 1004 sponds to the amount of anthropogenic heat built into the ocean initial conditions. In
 1005 contrast, after 1000 years of simulation, the CM4X-p25 simulation has nearly 1100 ZJ
 1006 *more* heat content (and 0.2 K warmer temperature) than the initial conditions.

1007 Now consider a model with exact present-day initial conditions but flawed dynam-
 1008 ical core and physical parameterizations. The model's piControl equilibrium state will
 1009 generally differ from the correct preindustrial state that has ≈ 400 ZJ less heat content
 1010 than present-day. Depending on model biases, it will tend towards an equilibrium state
 1011 that is warmer (as in CM4X-p25 and ESM2Mb; Krasting et al. (2018)) or even cooler
 1012 than the preindustrial state (as in ESM2G; Krasting et al. (2018)). This re-equilibration
 1013 process is likely to engage the whole ocean, including the deep ocean, and thus have a
 1014 time scale set by the slow diffusive mixing timescale $\tau_{eq} \approx H^2/\kappa_d \approx 5000$ yr (Krasting
 1015 et al., 2018) of unventilated abyssal shadow zones (Holzer et al., 2021).

1016 6.4 The mesoscale dominance hypothesis

1017 Given that CM4X-p25 and CM4X-p125 share all configuration details, except for
 1018 the horizontal ocean grid spacing, we hypothesize that the strength of the ocean mesoscale
 1019 transport accounts for the order of magnitude time scale difference for piControl ther-
 1020 mal equilibration for these two models. Furthermore, we hypothesize that even for mesoscale
 1021 active models, it is critical to have accurate levels of parameterized mixing as well as low
 1022 levels of spurious numerical mixing, thus ensuring that the deep ocean is not engaged
 1023 as part of an erroneous diabatic equilibration process.

1024 Mesoscale eddies and vertical heat transport

1025 As presented in Section 5.3, we invoke a mesoscale eddy mechanism to explain the
 1026 distinct timescales for ocean thermal adjustment in CM4X-p25 and CM4X-p125. Namely,
 1027 Griffies et al. (2015) (along with Gregory (2000), Gnanadesikan et al. (2005), Wolfe et
 1028 al. (2008), Gregory and Tailleux (2011), Delworth et al. (2012), Morrison et al. (2013),
 1029 Hieronymus and Nycander (2013), Zika et al. (2014), D. P. Marshall and Zanna (2014),
 1030 Doddridge et al. (2016), and von Storch et al. (2016)) identified the role of mesoscale ed-
 1031 dyes in regulating vertical heat transport between the upper ocean (roughly the upper
 1032 1000 m) and deeper ocean interior. This regulation is affected by the vertically upward
 1033 transport of positive buoyancy anomalies by eddies, and it holds whether the eddies are
 1034 explicitly represented, as in an eddy-admitting model, or parameterized via a scheme such
 1035 as Gent and McWilliams (1990) and Gent et al. (1995).

1036 For those regions without deep water formation (e.g., the Indian and Pacific basins),
 1037 much of the vertical transport of heat by the mean flow is compensated by an oppos-
 1038 ing heat transport by the mesoscale eddy flow, leaving a residual heat transport that is

1039 smaller than either the mean or eddy heat transport. In so doing, mesoscale eddies re-
 1040 duce the exchange of heat between the upper ocean and the deeper interior ocean; i.e.,
 1041 the upper and deeper oceans become disengaged. As a result, mesoscale eddies reduce
 1042 the role of deep ocean mixing in the process of the piControl thermal equilibration. We
 1043 can see this effect through the distinct temperature drifts found in the Indo-Pacific basin
 1044 (see Figure 20), in which there is only a negligible drift in the deep ocean of CM4X-p125
 1045 whereas the deep ocean in CM4X-p25 has a more sizable drift.

1046 In the intermediate and deep water formation regions of the high latitude Atlantic
 1047 and Southern Ocean, eddies also play a role in determining the exchange between the
 1048 surface and deep ocean, largely through the role of eddy restratification in convection
 1049 regions (J. Marshall & Schott, 1999). Focusing on the Southern Ocean, consider the con-
 1050 ceptual framework from D. P. Marshall and Zanna (2014) as a guide. Namely, their ide-
 1051 alized process model indicates that the bulk of the global ocean heat uptake occurs in
 1052 the Southern Ocean, with that heat uptake mediated by Ekman mechanics and with a
 1053 time scale set by mesoscale eddy processes. In contrast, they find that interior diapyc-
 1054 nal mixing plays a negligible role. To the degree that the mechanics of heat released
 1055 during the piControl is dominated by similar Southern Ocean wind and eddy mechan-
 1056 ics, we conjecture that Southern Ocean ventilation is key to setting the time scale for
 1057 piControl thermal equilibration. That is, we propose the D. P. Marshall and Zanna (2014)
 1058 mechanism as a null hypothesis describing the piControl thermal equilibration time scale.

1059 Restratification by submesoscale eddies in the upper ocean (i.e., mixed layer baro-
 1060 clinic instability), also affects the upper ocean ventilation (Boccaletti et al., 2007). In
 1061 particular, during the CM4X development we found the details of the Fox-Kemper et
 1062 al. (2008, 2011) submesoscale parameterization (see Section A14) to impact on the ver-
 1063 tical heat exchange and thus the thermal drift. Our chosen parameter settings are some-
 1064 what stronger than warranted by process studies. The settings were chosen in develop-
 1065 ing CM4X-p25, with the stronger restratification set to compensate for the somewhat
 1066 weaker restratification from the relatively weak mesoscale eddies. Determining the proper
 1067 interplay between the mesoscale and submesoscale restratification effects remains a topic
 1068 of ongoing research.

1069 *The need to reduce spurious mixing*

1070 The study of Adcroft et al. (2019) showed that thermal drift in an ocean model forced
 1071 by a prescribed atmosphere is directly related to the vertical coordinate, with the hy-
 1072 brid z^*/ρ_{2000} coordinate used in CM4.0 (and in CM4X) leading to significantly less drift
 1073 than z^* used throughout the ocean. These results suggest there are significantly larger
 1074 levels of spurious numerical mixing with the z^* coordinate than the hybrid z^*/ρ_{2000} co-
 1075 ordinate. They are also consistent with the study of Ilicak et al. (2012) who diagnosed,
 1076 using potential energy methods, large spurious mixing in CM2.5 (which uses z^* through-
 1077 out the full ocean depth). The association of a biased warm ocean arising from spuri-
 1078 ous mixing is supported by Krasting et al. (2018) and Hieronymus et al. (2019), with both
 1079 papers using the GFDL-CM2G climate model (which uses an isopycnal coordinate ocean
 1080 component) and finding that the volume-averaged ocean temperature increases with back-
 1081 ground diapycnal diffusivities.

1082 *Summarizing the mesoscale dominance hypothesis*

1083 We hypothesize that there are three ocean model properties necessary to support
 1084 a centennial rather than the millennial time scale for thermal equilibration into an ocean
 1085 that is cooler (with roughly 400 ZJ less heat content than early 21st century) in its 1850
 1086 piControl state: (I) enhanced fidelity of mesoscale features, including transient eddies
 1087 and boundary currents; (II) accurate strength and geography of parameterized numer-
 1088 ical mixing processes; and (III) negligible levels of spurious mixing from numerical dis-

1089 cretization. We refer to ocean models that possess these three properties as *mesoscale*
 1090 *dominant models*. Mesoscale dominant models contrast to those where deep ocean di-
 1091 abatic processes (either parameterized or spurious numerically induced) play a promi-
 1092 nent (and sometimes dominant) role in piControl thermal equilibration. We infer that
 1093 models that are not mesoscale dominant engage their deep ocean circulation during the
 1094 1850 piControl, thus rendering far longer thermal equilibration time scales. These long
 1095 thermal spin-ups also affect long spin-up times for biogeochemical cycles, though bio-
 1096 geochemical spin-ups are also impacted by other processes (Orr et al., 2017; Khatiwala,
 1097 2023, 2024).

1098 CM4X-p125 does not perfectly realize each of the three properties of a mesoscale
 1099 dominant model. In particular, both of the CM4X models have biases in the Southern
 1100 Ocean ventilation, as noted by the overly deep mixed layers in Figure 9 and as further
 1101 explored for CM4.0 in Krasting et al. (2024). The case studies in Part II (Griffies et al.,
 1102 2024) point to further shortcomings that form the topic of ongoing research and devel-
 1103 opment. Even so, we propose that CM4X-p125 approaches mesoscale dominance more
 1104 than CM4X-p25, given the stronger mesoscale eddy activity in CM4X-p125.

1105 6.5 Water mass transformation analysis

1106 As an initial plausibility test of the mesoscale dominance hypothesis, we employ
 1107 water mass transformation budgets (following Drake et al. (2024) and Section 7.4 in Part
 1108 II of Griffies et al. (2024)) to evaluate the extent to which (A) water masses are well-
 1109 equilibrated in CM4X piControl simulations and (B) spurious numerical mixing plays
 1110 a minor role in the water mass budgets. For a water mass $\mathcal{M}_\geq(\sigma_2, t)$ defined as the mass
 1111 of water that is denser than σ_2 at time t , the water mass transformation framework al-
 1112 lows us to attribute the drift, $\partial_t \mathcal{M}_\geq$, to processes that drive it.

1113 Figure 25 shows the piControl drift of dense water masses in CM4X-p25 and CM4X-
 1114 p125, quantified by the normalized water mass change (relative to the initial condition),

$$1115 \Delta \widetilde{\mathcal{M}}_\geq \equiv \widetilde{\mathcal{M}}_\geq(\sigma_2, t) - \widetilde{\mathcal{M}}_\geq(\sigma_2, t = 0) \equiv \frac{\mathcal{M}_\geq(\sigma_2, t)}{\mathcal{M}} - \frac{\mathcal{M}(\sigma_2, t = 0)}{\mathcal{M}}, \quad (4)$$

1116 where $\mathcal{M} \equiv \mathcal{M}_\geq(\min(\sigma_2), t)$ is the total mass of the ocean and tildes denote we have
 1117 normalized the water mass by the total ocean mass. The normalized water mass, $\widetilde{\mathcal{M}}_\geq(\sigma_2, t)$,
 1118 increases monotonically from 0 for the densest water in the ocean to 1 for the lightest
 1119 waters and thus can be thought of as the ocean's cumulative probability density func-
 1120 tion in potential density space (see Figure 26d). $\Delta \widetilde{\mathcal{M}}_\geq$ quantifies how much a given wa-
 1121 ter mass has drifted from the initial condition over time, expressed as a fraction of the
 1122 total ocean mass.

1123 Figure 25a shows that deep water masses ($\sigma_2 > 36 \text{ kg/m}^3$) drift by about -1% per
 1124 century in CM4X-p25 (Figure 25a); because total ocean mass is conserved, this nega-
 1125 tive drift signifies a drift towards lighter densities (consistent with the warming described
 1126 in Section 4). By contrast, water masses that include lighter surface and thermocline wa-
 1127 ters (i.e., $\sigma_2 \leq 36.0 \text{ kg/m}^3$) show few changes after the first 50 years, suggesting that
 1128 they equilibrate more quickly. In CM4X-p125, the deep water masses exhibit smaller wa-
 1129 ter mass changes, with near zero drift after the first 100 years (Figure 25b). The largest
 1130 water mass changes are seen in the Antarctic Bottom Water density class ($\sigma_2 \geq 37 \text{ kg/m}^3$).

1131 To put these piControl water mass drifts in context, we compare them to the anomalous
 1132 water mass changes in the forced historical plus SSP5-8.5 experiments. The forced
 1133 water mass changes are fairly similar in both magnitude, density structure, and timing
 1134 between CM4X-p25 and CM4X-p125 (Figures 25c,d). These forced water mass changes
 1135 are larger and occur more quickly than the piControl water mass changes. In CM4X-
 1136 p125, the forced water mass drifts, $\partial_t \mathcal{M}$, at the end of the SSP5-8.5 experiment are or-
 1137 ders of magnitude larger than the piControl drifts across all density classes. These re-

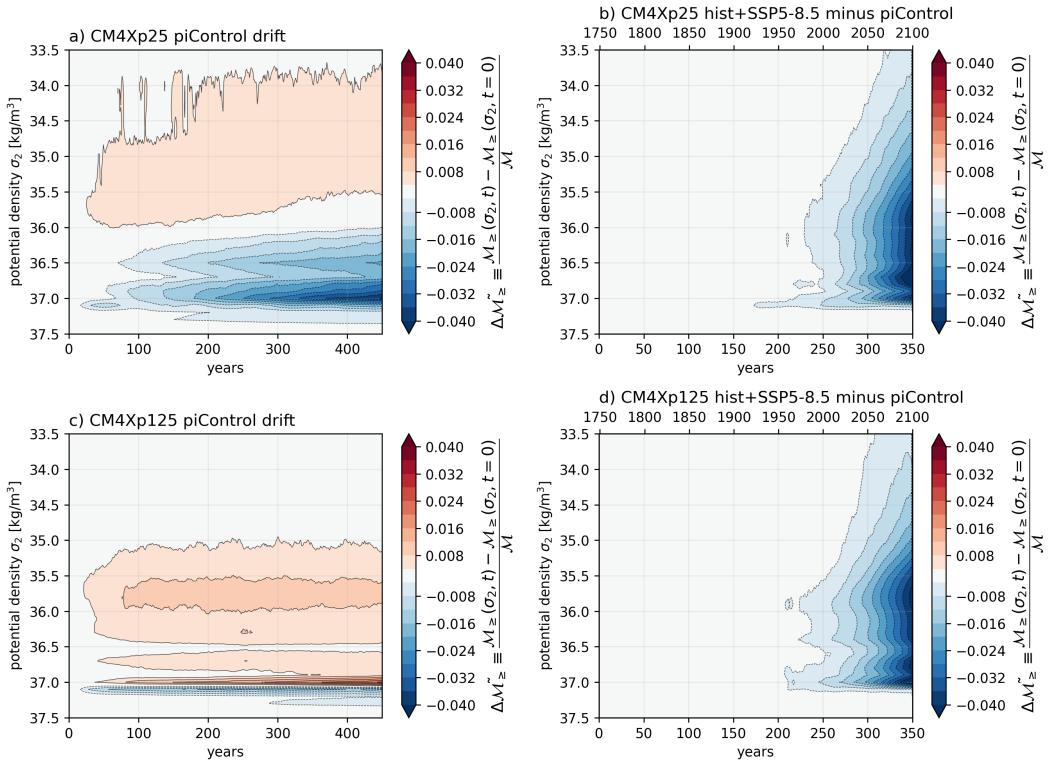


Figure 25. Normalized water mass changes $\Delta\tilde{\mathcal{M}}_{\geq}$ (equation 4, where $\mathcal{M}_{\geq}(\sigma_2, t)$ is the mass of water denser than σ_2 at time t) in CM4X-p25 (a,b) and CM4X-p125 (c,d), expressed as a fraction of the total ocean mass. Panels (a,c) show the changes relative to the initial condition in the piControl experiment. Panels (b,d) show the changes in the forced hist+SSP5-8.5 simulations, relative to the piControl changes. Time is expressed in years after initialization below and historical years above.

sults demonstrate that water masses are indeed well-equilibrated in the CM4X piControl simulations.

To understand how the water mass drifts, $\partial_t \mathcal{M}_{\geq}$, can be so small in the piControl simulations, we now turn to the water mass transformation budget:

$$\partial_t \mathcal{M}_{\geq} = \mathcal{S}_{\geq} + \mathcal{G}, \quad (5)$$

where \mathcal{G} is the total water mass transformation rate (positive when it increases the mass of denser water), which we decompose into the contributions from three processes: boundary buoyancy fluxes ($\mathcal{G}^{(BF)}$; including air-sea, ice-ocean, and geothermal fluxes), parameterized mixing ($\mathcal{G}^{(Mix)}$; representing surface mixed-layer, interior, and bottom boundary layer mixing processes), and spurious numerical mixing ($\mathcal{G}^{(Spur)}$; diagnosed as the residual of the other terms). \mathcal{S}_{\geq} is the water mass tendency due to direct exchange across the ocean's boundary (e.g., evaporation minus precipitation plus rivers and meltwater) and is typically negligible.

Figures 26a,b reveal that water masses of all densities exist in an approximate balance between transformation by boundary fluxes and parameterized mixing. At zeroth order, the formation of dense deep waters ($35 \text{ kg/m}^3 \leq \sigma_2 \leq 36 \text{ kg/m}^3$) and even denser bottom waters ($36.75 \text{ kg/m}^3 \leq \sigma_2 \leq 37.5 \text{ kg/m}^3$) at a rate of 40 Sv at high latitudes is mostly balanced by parameterized mixing-driven transformation. Spurious numeri-

1156 cal transformations due to discretization errors in the advection or Lagrangian remapping
 1157 scheme (see Drake et al. (2024)), however, further modify water masses at rates of
 1158 $\mathcal{O}(5 \text{ Sv})$.

1159 After 440 years of spin-up in the piControl simulations, the water mass transfor-
 1160 mations described above are nearly in balance, with residual water drifts of less than 4 Sv
 1161 in all density classes when averaged over 10 years (Figure 26c). Furthermore, the unbal-
 1162 anced water mass drift seems to be smaller in CM4X-p125 than in CM4X-p25, especially
 1163 at the densities that dominate the global water mass distribution (Figure 26d), provid-
 1164 ing further evidence in support of the mesoscale dominance hypothesis. While analogous
 1165 mass budgets for the CM2.5/CM2.6 hierarchy are not available, other models that use
 1166 the same ocean model code (MOM5) exhibit spurious water mass transformations in ex-
 1167 cess of those driven by parameterized mixing rates (Ilicak et al., 2012; Holmes et al., 2021).
 1168 The relative smallness of spurious numerical mixing in the CM4X water mass budget sup-
 1169 ports our hypothesis that the transition to a hybrid $z^* - \rho_2$ coordinate in CM4X (MOM6
 1170 code) enables the mesoscale dominant regime to emerge.

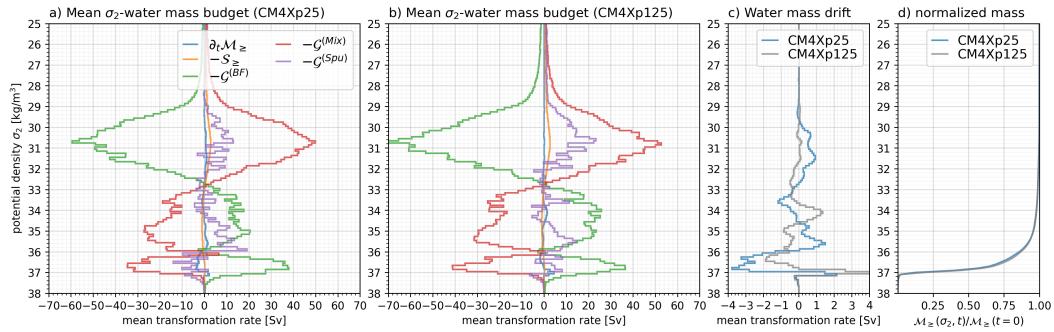


Figure 26. Global water mass budgets (see Section 6 in Part II of Griffies et al. (2024) for details) in potential density (σ_2 ; potential density referenced to 2000 dbar, minus 1000 kg m^{-3}) coordinates for CM4Xp25 (a) and CM4Xp125 (b), averaged over years 441-451 of the piControl. (c) Comparison of the water mass drift $\partial_t \mathcal{M}_\geq$ terms in the CM4X configurations. (d) The normalized water mass distribution, $\tilde{\mathcal{M}}_\geq = \mathcal{M}_\geq(\sigma_2, t)/\mathcal{M}$, defined by the mass of water $\mathcal{M}_\geq(\sigma_2, t)$ that is denser than a given potential density σ_2 at time t divided by the total ocean mass \mathcal{M} . Panel (d) shows that about 85% of the global ocean mass is occupied by deep waters within a narrow potential density range between 36 kg/m^3 and $\sigma_2 = 37.1 \text{ kg/m}^3$.

1171 7 Concluding remarks

1172 As further explored in Part II (Griffies et al., 2024), refined ocean grid spacing im-
 1173 proves many aspects of the ocean and climate system that makes CM4X-p125 more suited
 1174 to a variety of applications. Furthermore, those cases where CM4X-p25 and CM4X-p125
 1175 are quite similar, such as global mean transient climate response and ocean heat uptake
 1176 efficiency, confirm decades of work done with comparatively coarse resolution models.
 1177 That is, for models with relatively modest drift such as CM4X-p25, we can garner a sen-
 1178 sible estimate for many bulk climate response diagnostics as compared to models with
 1179 extremely low drifts such as CM4X-p125. Even so, there are aspects where payoff for the
 1180 enhanced grid resolution is nontrivial, such as the roughly 20% reduction in sea surface
 1181 temperature biases in CM4X-p125, and the order of magnitude reduction in thermal equi-
 1182 libration time for the CM4X-p125 piControl experiment.

Based on prior experience with CM2.5/CM2.6 (Griffies et al., 2015), we expected the more active eddying ocean in CM4X-p125 to play an important role in establishing thermal equilibrium. Even so, we were pleasantly surprised when CM4X-p125 equilibrated to a cooler (by volume mean) ocean after only $\mathcal{O}(100)$ years. Furthermore, the cooler ocean state has roughly 400 ZJ less ocean heat content than the present-day initial conditions, in accord with observational estimates from Zanna et al. (2019). For global-scale studies of future climate, the very small drift in CM4X-p125 allows for a clear connection between the physical ocean response, ocean interior focused model diagnostics such as ocean heat uptake efficiency, and surface/atmosphere diagnostics like the climate feedback parameter or the SST pattern effect (Armour et al., 2013; Andrews et al., 2018). Note, however, that there remain some regional temperature drift within CM4X-p125 that might necessitate drift removal to reveal the role of specific regions in setting future climate.

It is possible that the piControl thermal equilibration found in CM4X-p125 is a one-off fortuitous result, and we have incomplete evidence to argue against that perspective. Even so, evidence based on prior studies and further analysis presented in this paper support our conjecture that CM4X-p125 exemplifies what can be expected with higher fidelity ocean climate models used within a realistic coupled climate model. There certainly is far more to climate model drift than just the ocean physics and numerics. Yet the ocean plays a huge role in the mechanics of climate model drift, and with the CM4X hierarchy proving a fit-for-purpose tool for investigating that role. In this closing section we offer a few remarks about the need to further develop a mechanical/process understanding of piControl thermal equilibration, and then comment on prospects for climate modeling with a mesoscale dominant ocean.

7.1 Mechanisms for piControl thermal equilibration

The piControl drift has traditionally been seen as a necessary, but inconvenient and uninteresting, part of the climate modeling art. We instead suggest that it provides signatures for climate model fidelity, and as such it should be studied (e.g., Banks et al. (2006); Krasting et al. (2024)). Correspondingly, we propose that developing an understanding of piControl drift, whether it be related to spurious numerical processes or a variety of physical processes, is a central task needed to advance climate model integrity and utility. Even if mesoscale dominant models prove difficult to reproduce, the mesoscale dominance hypothesis offers a conceptual framing for thermal drift in piControl simulations. Such framing could, we propose, support the movement towards standard protocols for piControl simulations that allow for a meaningful comparison across models.

Much work is needed to move beyond the descriptive and diagnostic presentation in this paper. One approach could be to map the results from the CM4X suite onto the D. P. Marshall and Zanna (2014) idealized model, thus exploring the genesis and evolution of the water mass adjustment according to a variety of physical ocean processes, with a particular focus on mesoscale eddy and wind effects in the Southern Ocean. Another is to emulate Banks et al. (2006), who explored the piControl drift in their coupled model, pursuing an analysis from deep water formation regions to the top of the atmosphere imbalances. A further approach concerns the temperature-space water mass analysis of Holmes et al. (2019, 2020, 2021) and Deppenmeier et al. (2021), who performed heat budgets within regions bounded by isotherms, thus enabling a focus on diabatic processes associated with the movement of heat and removing the large, and less relevant, signal due to adiabatic processes. We note that water mass transformations are generally simpler to compute and to interpret in temperature space, rather than the more common density space considered in Section 6.5, largely since subtleties of the nonlinear equation of state of seawater are circumvented with a temperature approach.

1233 **7.2 Prospects for eddying climate models**

1234 The common experience of $\mathcal{O}(1000)$ years for piControl thermal equilibration is baked
 1235 into community strategies for climate modeling. Consequently, there is a presumption
 1236 that the explicit representation of mesoscale eddies and associated fine-scale boundary
 1237 currents will long remain computationally out of reach for routine studies of climate dy-
 1238 namics. However, if our experience with CM4X-p125 is reproducible (a big assumption
 1239 at this point in the science), then moving into a mesoscale dominant regime will reduce
 1240 piControl thermal equilibration timescales by roughly one order of magnitude. With mesoscale
 1241 dominant models, the added computational energy and wall-clock time needed for re-
 1242 fined grid spacing are compensated by a significantly reduced thermal equilibration time,
 1243 thus bringing forth the prospect of more common use of a mesoscale active ocean as part
 1244 of realistic climate models.

1245 **Appendix A CM4X model configuration details**

1246 The ocean and sea ice components of CM4X are based on the MOM6 ocean code
 1247 and SIS2 sea ice code as configured in the OM4.0 model detailed in Adcroft et al. (2019).
 1248 In this appendix we summarize those relatively few features of the CM4X ocean that differ
 1249 from OM4.0. We also summarize elements of the atmosphere and land models that
 1250 differ from the CM4.0 configuration of Held et al. (2019).

1251 **A1 Atmosphere model**

1252 As noted in Section 2.2, CM4X makes use of the C192 (nominally 50 km grid spacing)
 1253 version of the AM4 cubed sphere atmosphere model of Zhao (2020). This choice con-
 1254 trasts to the C96 (nominally 100 km grid) version (Zhao et al., 2018a, 2018b) used in
 1255 CM4.0. Our choice to use the C192 is based on the desire to simulate more energetic at-
 1256 mospheric cyclones that are key to the study of sea level extremes, such as those stud-
 1257 ied by Yin et al. (2020). In particular, Zhao (2020, 2022a) showed that the 50 km res-
 1258 olution AM4 (i.e., C192AM4) reasonably simulates the frequency of atmospheric rivers,
 1259 tropical storms, and mesoscale convective systems, as well as their associated precipi-
 1260 tation and extreme precipitation.

1261 In ocean-atmosphere coupled climate models, atmospheric radiative properties and
 1262 their responses to changes in radiative gases, aerosol emissions, and ocean surface tem-
 1263 peratures have a substantial impact on the simulation of SST and ocean volume mean
 1264 temperature. The top of atmosphere (TOA) radiative fluxes are heavily influenced by
 1265 cloud parameterizations, which often involve poorly constrained parameters. We set the
 1266 CM4X cloud parameters using the atmospheric model forced by observed SSTs and sea-
 1267 ice concentrations (SICs), commonly referred to as AMIP (Atmospheric Model Intercom-
 1268 parison Project) simulations. In Table A1 we summarize differences in atmospheric physics
 1269 parameter settings across various GFDL models. Other atmospheric model parameter
 1270 settings (i.e., time-step and dynamical core) in the CM4X atmosphere are identical to
 1271 those in the C192 atmosphere developed by Zhao (2020).

1272 After setting the atmospheric physics parameters, the net TOA radiative flux is
 1273 an emergent property of the simulation. When averaged over the period of 2006-2018,
 1274 the TOA radiative flux is 1.42 W m^{-2} in CM4X AMIP simulations. This value is higher
 1275 than the range of 0.52 W m^{-2} to 1.06 W m^{-2} estimated over 2006-2018, as cited in Ta-
 1276 ble 7.1 of Forster et al. (2021). The higher value was a target of our parameter settings,
 1277 aiming to mitigate the global SST cool biases shown in Figure 14. Even so, this setting
 1278 often comes at the expense of excessive heat uptake in the 1850 control experiment, de-
 1279 pending on ocean processes such as mesoscale eddy transport and parameterized sub-
 1280 mesoscale eddy mixing. This higher TOA radiative balance was also pursued when de-
 1281 veloping CM4.0.

MODEL	vfact	eros_scale_c	eros_scale_t	citation
CM4X	0.875	5e-5	5.75e-5	this paper
C192AM4	0.8	8e-5	8e-5	Zhao (2020)
CM4.0	0.9	4e-5	5e-5	Held et al. (2019)
SPEAR	0.9	4e-5	5e-5	Delworth et al. (2020)
ESM4.1	0.9	5e-5	5e-5	Dunne et al. (2020)

Table A1. Summary of the differences in atmospheric physics parameter setting for various GFDL models. All differences in parameter settings relate to cloud settings. `vfact` is used to set the ice fall velocity, which affects the precipitation efficiency of cold clouds. `eros_scale_c` and `eros_scale_t` are used to set the dissipation time scales of clouds over in the convective and turbulent regimes respectively.

To illustrate the differences in global TOA radiative properties between the atmospheric models used in CM4X and CM4.0, Table A2 provides a comparison of global TOA radiative fluxes simulated by the two models. Both models are forced by the same monthly varying climatological SSTs and SICs averaged over years 1981-2014, with radiative gases and aerosol emissions fixed at present-day (PD, 2010) levels (referred to as 2010Control). The CM4X atmosphere produces roughly 1 W m^{-2} higher net longwave and shortwave radiative fluxes, with the net TOA downward flux being slightly larger. This result indicates that the low drift in CM4X as compared to CM4.0 is not due to differences in their atmospheric models.

Table A2 also provides the total radiative flux perturbation (RFP) derived from the two atmospheric models based on their corresponding present-day (PD, 2010) and preindustrial (PI, 1850) simulations, the aerosol RFP based on PD and PI aerosol emissions, as well as the idealized CESS feedback (Cess et al., 1990) derived from the idealized 2K uniform SST warming and their corresponding 2010Control simulations. These results suggest a broad similarity between the two atmospheric models, with the CM4X atmosphere showing a reduction of aerosol forcing. The reduction of aerosol forcing with increased atmospheric model resolution (C192 versus C96) was discussed in Zhao et al. (2018b, 2018a).

A2 Land model

For CM4X we prescribe a time-invariant (i.e., “static”) map of vegetation that incorporates both natural vegetation and land-use conditions representative of year 1980. Unlike CM4.0, which used dynamic vegetation (i.e., interactive, prognostic), the same static vegetation boundary condition is used in all CM4X simulations. Prescribing the land surface condition removes feedbacks that occur in the Earth system, but also helps to better examine how ocean changes impact the climate within the CM4X configuration. Zhao (2022a) studies climate sensitivity across a suite of GFDL climate models and considers the impact of static versus dynamic vegetation. Additionally, eight of the 11 CMIP6 models analyzed by Arora et al. (2019) use static vegetation (see their Table 2).

A3 Albedos for glaciers

We explored a series of land ice albedo settings motivated by the studies of Delworth et al. (2020), Dunne et al. (2020), and L. Zhang et al. (2021), each of whom identified the importance of near-infrared glacial snow albedos on both the production of Antarctic Bottom Water and the occurrence of intermittent Southern Ocean convection. Our tests confirmed their results, and we chose the brighter set of albedos used in the GFDL-

Global TOA radiative fluxes (W m^{-2})	CM4.0	CM4X
OLR (net LW)	238.44	239.51
SWABS (net SW)	240.16	241.39
NETRAD (net LW+SW)	1.72	1.88
Total RFP between 2010 and 1850	2.71	2.86
Aerosol RFP between 2010 and 1850	-0.76	-0.65
CESS feedback between 2K uniform SST increase & 2010Control	-3.53	-3.54

Table A2. A comparison of global TOA radiative fluxes (W m^{-2}) simulated by the atmospheric models used in CM4X and CM4.0, with both models forced by the same monthly varying climatological SSTs and sea-ice concentrations averaged from 1981–2014, and radiative gases and aerosol emissions fixed at present-day (PD, 2010) levels (referred to as 2010Control). OLR = TOA outgoing longwave radiation; SWABS = TOA net SW radiative flux (absorption, downward positive); NETRAD = TOA net radiative flux (downward positive, i.e., SWABS minus OLR). The table also lists the total radiative flux perturbation (RFP) derived from each model based on present-day (PD, 2010) and preindustrial (PI, 1850) conditions, the aerosol RFP based on PD and PI aerosol emissions, and the idealized CESS feedback (Cess et al., 1990), which is derived based on the idealized 2K uniform SST warming and their corresponding 2010Control simulations. Each climatological simulation was integrated for 30 years.

1316 ESM4.1 configuration of Dunne et al. (2020) (see Table A3). These near-infrared glacial
 1317 snow albedos lie at the upper limit of the observational range (Dunne et al., 2020).

MODEL	F_ISO_WARM_ON_GLACIER	F_ISO_COLD_ON_GLACIER	citation
CM4X	0.77, 0.58	0.92, 0.73	this paper
CM4.0	0.77, 0.43	0.92, 0.68	Held et al. (2019)
ESM4.1	0.77, 0.58	0.92, 0.73	Dunne et al. (2020)
SPEAR	0.77, 0.43	0.92, 0.68	Delworth et al. (2020)

Table A3. Summary of the land glacier albedo settings for a variety of GFDL models, listing the values for (Visible, Near IR) radiation and warm/cold surfaces.

1318 Note that Dunne et al. (2020) states that “we had to increase the near-infrared albedo
 1319 of snow on glaciers to a value of 0.82.” This value of 0.82 is different than the values in
 1320 Table A3 because it refers to the total ratio of shortwave up to shortwave down, which
 1321 is an emergent property of the simulation. Thus, it would be more appropriate for Dunne
 1322 et al. (2020) to say “we found that we had to increase the near-infrared albedo of snow
 1323 on glaciers to obtain an overall shortwave albedo value of 0.82 for these regions...” (John
 1324 Dunne, personal communication).

1325 In Figure A1 we show the time series for the global volume mean ocean temper-
 1326 ature for CM4.0 and CM4X-p25, along with a third configuration with CM4.0 using the
 1327 albedo of snow on glaciers used by CM4X. The nearly parallel behavior of CM4X-p25
 1328 and CM4.0-albedo indicates that the snow on glacier albedo is the key difference, from
 1329 a global enthalpy budget perspective, between CM4.0 and CM4X-p25.

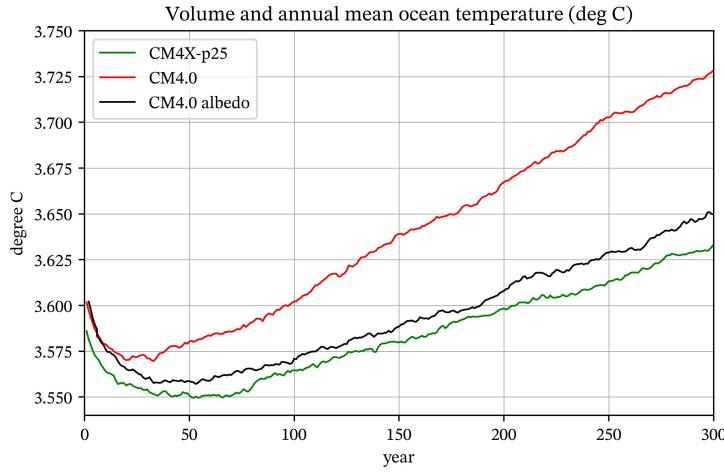


Figure A1. Time series of the global volume mean ocean temperature from CM4.0, CM4.0-albedo, and CM4X-p25. The CM4.0-albedo experiment is identical to CM4.0, with the one exception being the use of the CM4X snow on glacier albedo as per Table A3. The nearly parallel behavior of CM4X-p25 and CM4.0-albedo indicates that the snow on glacier albedo is the key difference, from a global energy budget perspective, between CM4.0 and CM4X-p25.

1330

A4 Albedos for sea ice

1331
1332

In testing the sea ice albedos for the CM4X configuration, we started with those used by Delworth et al. (2020) yet found these to suffer from the following issues:

1333
1334
1335
1336

- Arctic sea ice that was too thin and biased low in summer;
- Antarctic sea ice that was biased low in summer;
- Systematic model drift towards a weakening Antarctic circumpolar current and weakening Antarctic Bottom Water (AABW) transport.

1337
1338
1339
1340
1341
1342

These biases motivated returning the sea ice albedo parameters to the brighter values used in CM4.0 (Held et al., 2019). Doing so increased the sea ice thickness and summer sea ice extent. The change was also motivated by the expectation that cooling the Southern Ocean surface would encourage intermittent oceanic convection and production of Antarctic bottom waters (Delworth et al., 2020; L. Zhang et al., 2021). Table A4 lists the settings for CM4X and compares to those used in other GFDL models.

MODEL	ICE_DELTA_EDD_R_ICE	ICE_DELTA_EDD_R_SNOW	ICE_DELTA_EDD_R_POND
CM4X	1.0	1.0	1.0
CM4.0	1.0	1.0	1.0
ESM4.1	1.5	1.5	1.5
SPEAR	0.0	0.0	0.0

Table A4. Summary of the sea ice albedo settings for a variety of GFDL models. Albedos are in units of observed standard deviations relative to observed mean values from Briegleb and Light (2007).

1343 **A5 Fixing a sea ice bug for Arctic bipolar coordinate fold**

1344 A numerical error was identified in the sea ice along the bipolar grid fold in the Arctic
 1345 Ocean, with this bug present in the OM4.0 and CM4.0 models. This error occurs along
 1346 the half of the bipolar fold that extends from the geographic North Pole to the grid pole
 1347 located in Russia. The error is most visible in zonal and meridional sea ice velocity fields,
 1348 but it also creates artifacts in sea ice thickness and concentration fields, as well as the
 1349 upper ocean dynamical fields. Tests in CM4X-p25 and CM4X-p125 reveal that the bug
 1350 has minimal impact on the global climate. Even so, we corrected the bug for the CM4X
 1351 simulations shown in this paper.

1352 **A6 Horizontal grids for the ocean and sea ice**

1353 For the ocean and sea ice components, we make use of a locally orthogonal coor-
 1354 dinate grid with a nominal grid spacing of 0.125° along the equator for CM4X-p125 and
 1355 0.25° for CM4X-p25. As detailed in Section 2.1.2 of Adcroft et al. (2019), the grid con-
 1356 sists of a Mercator layout between latitudes given in Table A5, which transitions to a
 1357 bipolar northern cap using the methods of Murray (1996) with coordinate singularities
 1358 over Siberia and Canada. In the far south (south of 66.8°S), the grid transitions to a reg-
 1359 ular latitude/longitude grid, and it is capped by a displaced pole south of 78.0°S (see
 1360 Figure 1 of Adcroft et al. (2019)). Use of a southern displaced pole is motivated to op-
 1361 timize the lateral grid spacing in the southernmost parts of domain.

Grid property	CM4X-p25	CM4X-p125
nominal (equatorial) grid spacing	0.25°	0.125°
grid cells ($n_i \times n_j$)	1440×1080	2880×2240
southern edge of Mercator grid	66.8595472°S	$66.8840965847^\circ\text{S}$
northern edge of Mercator grid	64.0589597°N	$64.0316059408^\circ\text{N}$
longitudes of bi-polar singularities	$120^\circ\text{W}, 60^\circ\text{E}$	$120^\circ\text{W}, 60^\circ\text{E}$
southern edge of south spherical grid	78°S	78°S

1362 **Table A5.** Details for the horizontal grids used by the two CM4X ocean and sea ice compo-
 1363 nents. *Resolution* in this paper refers to the number of grid degrees of freedom and as such it is
 1364 an integer number, whereas *grid spacing* refers to the distance between grid cells as measured in
 1365 units of degrees or kilometers.

1362 **A7 Numerical methods for the Coriolis and Magnus accelerations**

1363 There are a variety of numerical methods available in MOM6 for handling the Corio-
 1364 lis plus Magnus accelerations

1365
$$\text{Coriolis} + \text{Magnus} = \frac{f + \zeta}{h} (\hat{\mathbf{z}} \times \mathbf{u} h), \quad (\text{A1})$$

1366 where h is the layer thickness and ζ is the vertical component to the relative vorticity.
 1367 The term $\zeta \hat{\mathbf{z}} \times \mathbf{u}$ is referred to as the *Magnus acceleration* in some treatments, partic-
 1368 ularly the engineering literature, whereas it is commonly referred to as the *nonlinear Cori-*
 1369 *olis acceleration* in the ocean modeling literature. The layer thickness, h , in the numer-
 1370 ator does not exactly cancel h in the denominator. The reason is these two thicknesses
 1371 are computed using distinct horizontal averaging operators to ensure a proper handling
 1372 of the vanishing layer limit where $h \rightarrow 0$.

1373 Table A6 lists the suite of numerical settings found in CM4X as compared to CM4.0
 1374 for the Coriolis and Magnus accelerations, with a summary of these settings given here.

SCHEME	CM4.0	CM4X
NOSLIP	False	False
CORIOLIS_EN_DIS	False	True
CORIOLIS_SCHEME	SADOURNY75_ENSTRO	SADOURNY75_ENSTRO
BOUND_CORIOLIS	True	True
KE_SCHEME	KE_ARAKAWA	KE_ARAKAWA
PV_ADV_SCHEME	PV_ADV_CENTERED	PV_ADV_CENTERED

Table A6. Summary of the various numerical settings associated with the Coriolis and Magnus accelerations as used in CM4.0 and CM4X, with details provided in Section A7.

- For CM4X we set `CORIOLIS_EN_DIS = True`, which means that two estimates of the layer thickness fluxes are used to estimate the Coriolis + Magnus accelerations, and the one that dissipates energy relative to the other one is used. In CM4.0 this parameter was set `False`.
- In CM4.0 and CM4X we set `CORIOLIS_SCHEME = "SADOURNY75_ENSTRO"`, which is based on the enstrophy conserving scheme detailed in Sadourny (1975).
- In CM4.0 and CM4X we set `BOUND_CORIOLIS = True`, which means that the Coriolis terms at u-points are bounded by the four estimates of $(f+\zeta)v$ from the four neighboring v-points, and similarly at v-points. This option would have no effect on the SADOURNY Coriolis scheme if it were possible to use centered difference thickness fluxes.
- In CM4.0 and CM4X we set `KE_SCHEME = "KE_ARAKAWA"`, which determines the discretization for acceleration from the kinetic energy gradient.
- In CM4.0 and CM4X we set `PV_ADV_SCHEME = "PV_ADV_CENTERED"`, which determines the discretization of PV advection as per the centered scheme of Sadourny (1975).

1391 A8 Ocean bottom topography and geothermal heating

1392 For the 0.25° configuration of CM4X-p25, we retained the same bottom topogra-
 1393 phy and horizontal used in OM4.0, as well as the same bottom geothermal heat flux from
 1394 Huw Davies (2013). Here, we present steps used to construct the 0.125° topography for
 1395 CM4X-p125 as shown in Figure A2, along with the geothermal heat flux mapped to this
 1396 bottom topography.

1397 GEBCO and BedMachine data sources

1398 The bottom topography for CM4X-p125 is based on GEBCO (GEBCO Compilation
 1399 Group, 2021). This topography provides global coverage on a 15 arc-second inter-
 1400 val grid. For the region south of 62°S we transitioned to the BedMachine product from
 1401 Morlighem et al. (2017) since BedMachine offers more information about the Southern
 1402 Ocean and goes further inland to capture ice-shelf cavities (planned for future models
 1403 at GFDL but not implemented here). We placed the BedMachine data onto the GEBCO
 1404 grid using linear remapping and then interpolated this merged topography to the CM4X-
 1405 p125 ocean model grid.

1406 Hand tuning of selected straits

1407 For the C-grid of MOM6, flow through openings is available with a single grid tracer
 1408 cell. Hence, the C-grid allows for a more detailed representation of the complex geog-

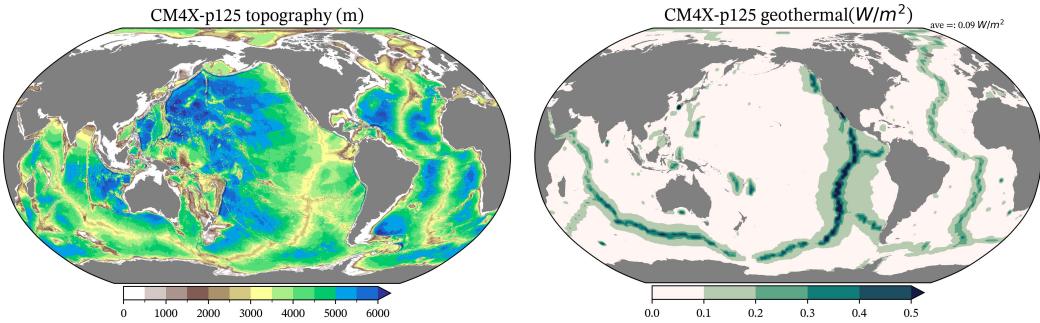


Figure A2. Top panel: ocean bottom depth (in meters) for the 0.125° horizontal grid used in CM4X-p125. For the region south of 62°S we transitioned from the GEBCO (GEBCO Compilation Group, 2021) product used in the north to the BedMachine product from Morlighem et al. (2017). Both CM4X-p25 and CM4X-p125 set their minimum depth at 9.5 m and maximum depth at 6500 m. Bottom panel: geothermal heat flux for CM4X-p125, taken from Huw Davies (2013). The global area mean (ocean area) is 0.09 W m^{-2} , though note that values along the mid-ocean ridges are generally much larger.

raphy of the World Ocean than the B-grid used in earlier MOM versions (e.g., the CM2.6 model of Griffies et al. (2015) and ACCESS-O model of Kiss et al. (2020) use MOM5, which is a B-grid). Even so, a coarsened grid generally admits straits that are artificially closed due to limitations of grid spacing. However, certain openings are particularly important for water masses of the World Ocean, with a primary example being the roughly 10 km wide Strait of Gibraltar connecting the Atlantic and Mediterranean. Conversely, there can be straits that are opened on a grid that are unwarranted, and so need to be closed. These considerations make it necessary to hand-tune special passages and straits to enable flow if artificially closed or to close straits that are best kept closed (e.g., Lake Maracaibo in Venezuela).

The hand-tuning works on a cell-by-cell basis, opening or closing straits according to the size of the grid cell. MOM6 provides the added feature of being able to restrict the openings of channels to a width corresponding to observed values, even if that value differs from the nominal grid cell size. We provide information about the restrictions in a runtime table, with the p125 grid restricting flow for the Dardanelles, Bosphorus, and between Sakhalin and Russia. After the hand-tuning process is complete, we fill inland waters, such as the North American Great Lakes, so that the ocean model only resolves sizable ocean water bodies and excludes lakes.

A9 Vertical coordinates and flow next to topography

MOM6 makes use of the vertical Lagrangian remapping method (see Griffies et al. (2020) for a review). For the vertical grid, we follow the hybrid vertical coordinate of OM4.0 as discussed in Section 2.1.4 of Adcroft et al. (2019), with this vertical coordinate following the approach of Bleck (2002). Here, the upper ocean layer interfaces are aligned with z^* quasi-geopotential surfaces (Adcroft & Campin, 2004), whereas the interior layer interfaces are aligned with potential density referenced to 2000 dbar. Note that the same vertical Lagrangian remapping method is used regardless the vertical coordinate choice. There are 75 coordinate layers for any particular vertical column of ocean, although some of those layers can have vanishing thickness depending on the density.

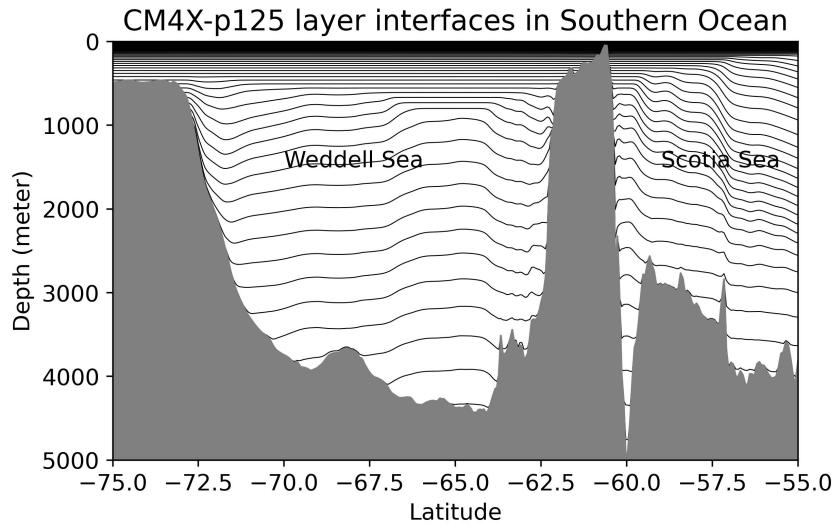


Figure A3. Layer interfaces through the Weddell and Scotia Seas (with South Orkney Islands rising just south of 60S) for the CM4X-p125 configuration as time averaged over years 1980–2009. The horizontal layers in the upper ocean are aligned with z^* whereas the interior layers are aligned with potential density referenced to 2000 dbar. Horizontal coordinate layers in the upper ocean are finely spaced, thus creating the dark black region. Representation of the bottom topography is linear piecewise continuous, with layers feeling a quadratic bottom drag when interacting with the bottom. Note that the strong baroclinicity next to the Antarctic shelf (southerly upward sloping isopycnals) is a signature of the westward flowing Antarctic Slope Current.

In Figure A3 we illustrate vertical coordinate layers in the Southern Ocean for a particular 1-year time mean from CM4X-p125. Along with the coordinate layers, this figure reveals that the solid-earth bottom is represented using a linear piecewise fit to the topography from Figure A2. As in the real ocean, MOM6 does not distinguish sides from bottom. Instead, the flow within a model layer feels a quadratic bottom drag wherever a layer intersects the solid earth (see Section 2.2.4 of Adcroft et al. (2019) for details of the bottom boundary layer scheme, including bottom drag). As discussed by Deremble et al. (2012), such treatment of the interactions of the fluid with solid boundaries is more physically motivated than that rendered by level coordinate models (see Adcroft et al. (1997) as well as Section 6 of Griffies et al. (2000) for more on the limitations of a level coordinate model representation of topography). Realistic flow-topography interactions are essential for capturing boundary waves and the associated sea level patterns along continental shelves, such as the east coast of North America (Wise et al., 2018, 2020; Hughes et al., 2019).

1451 **A10 Parameterization of ocean bottom boundary layer mixing**

The ocean bottom boundary layer parameterization in CM4X is similar to that used in OM4 and CM4 (see Legg et al., 2006), yet with one important change to the mixing efficiency parameter. The parameterization first estimates the turbulent kinetic energy (TKE) production in the bottom boundary layer using information from the dissipation of kinetic energy due to bottom drag. The turbulent vertical diffusivity is then computed by assuming a fraction of the TKE is dissipated and the remaining fraction is available to mix into stable stratification and thus to increase potential energy (following the Osborn-Cox relation, Osborn & Cox, 1972). In both OM4 and CM4, the mixing fraction (pa-

1460 parameter setting `bbl_effic` in MOM6) was set to 20%, which means that 80% of the TKE
 1461 production is dissipated and 20% goes into mixing. However, inspection of the coastal
 1462 waters in these models reveal this setting resulted in far too much bottom boundary layer
 1463 mixing, in some cases vertically homogenizing the water column in shelf regions.

1464 The 20% mixing fraction used in OM4 and CM4 was motivated by studies of tur-
 1465 bulence in stratified fluids without boundaries (Ivey & Imberger, 1991). Given the overly
 1466 strong mixing found with 20% mixing fraction, we chose to reduce it to 1% in CM4X.
 1467 This sizable change is motivated by the mixing fraction of turbulent mixing in large eddy
 1468 simulations of the surface boundary layer (see Reichl & Li, 2019). Further tests in an
 1469 OM4 configuration reveal little sensitivity to the precise value, so long as it is less than
 1470 a few percent.

1471 The reduced bottom boundary layer mixing fraction used in CM4X results in more
 1472 realistic mixed layer depths in coastal regions and more broadly improves circulation in
 1473 regions where dense water masses are formed in shallow regions. In related ongoing stud-
 1474 ies, outside the scope of this paper, we aim to develop a more thorough analysis of the
 1475 sensitivity of the simulations to the bottom boundary layer mixing fraction.

1476 A11 Parameterization of upper ocean boundary layer mixing

1477 The upper ocean boundary layer parameterization is based on Reichl and Hallberg
 1478 (2018), and we include the Reichl and Li (2019) extension that parameterizes effects from
 1479 Langmuir turbulence. The Stokes drift used by the Langmuir turbulence parameteriza-
 1480 tion is estimated following Q. Li et al. (2017). The settings for CM4X are updated rel-
 1481 ative to OM4.0 in order to agree with the recommended settings from Reichl and Li (2019).
 1482 The settings for CM4X were established prior to the recommendations from Reichl et
 1483 al. (2024) to improve upper ocean mixing and its diurnal variability in equatorial regions,
 1484 thus these changes are not included in CM4X. Similarly, the machine learning enhanced
 1485 vertical diffusivity as given in Sane et al. (2023) for the ocean surface boundary layer scheme
 1486 was not used since it was developed after CM4X.

1487 A12 Richardson number for parameterized shear-driven mixing

1488 Shear-driven mixing is parameterized according to the gradient Richardson num-
 1489 ber based scheme of Jackson et al. (2008). The gradient Richardson number, Ri , is the
 1490 ratio of the squared buoyancy frequency, N^2 , to the squared vertical shear of the hor-
 1491 izontal velocity

$$1492 \quad Ri = \frac{N^2}{(\partial_z u)^2 + (\partial_z v)^2} \quad \text{with} \quad N^2 = g(\alpha \partial_z \Theta - \beta \partial_z S), \quad (\text{A2})$$

1493 where g is the gravitational acceleration, Θ is the Conservative Temperature, S is the
 1494 salinity, α is the thermal expansion coefficient, and β is the haline contraction coefficient.
 1495 Ambiguities exist when computing the Richardson number on a discrete grid since trac-
 1496 ers and horizontal velocity components are generally centered at distinct locations. We
 1497 here describe how grid noise can be either generated or suppressed depending on how
 1498 the Richardson number is discretized.

1499 Discrete Richardson number that introduces grid noise

1500 For OM4.0, we chose to interpolate the horizontal C-grid velocity components to
 1501 the A-grid tracer points, with the vertical shear then computed on the A-grid along with
 1502 the buoyancy frequency and Richardson number (left panel of Figure A4). Analogous
 1503 approaches have been used in earlier versions of MOM (based on the B-grid velocity lay-
 1504 out), whereby the Richardson number is computed by either interpolating the tracer to
 1505 the B-grid velocity point or the velocity is interpolated to the tracer point. Unfortunately,

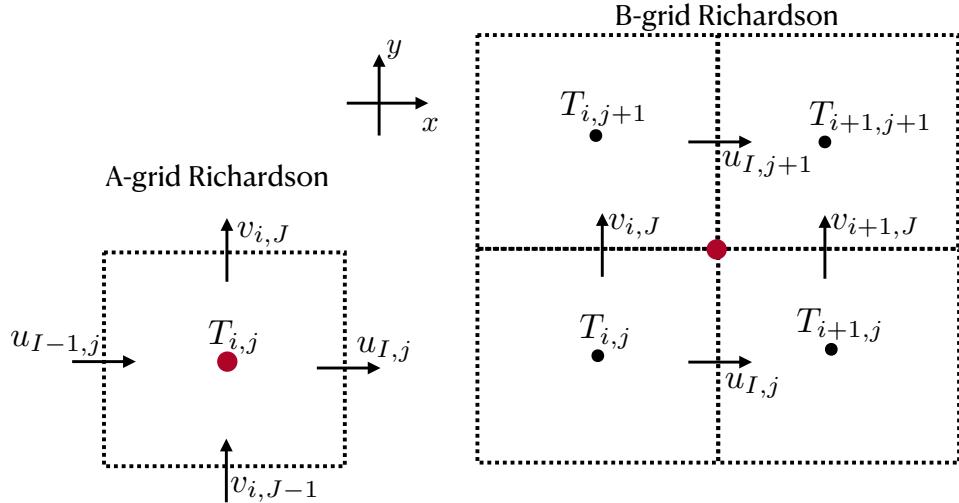


Figure A4. Two methods to compute the gradient Richardson number for the C-grid used in MOM6. Left panel: OM4.0 computes the Richardson number on the A-grid by averaging the horizontal velocity components, $u_{I,j}$ and $v_{i,J}$, to the tracer cell center (red dot at $T_{i,j}$) and then computing the vertical shear at the tracer cell center. Right panel: CM4X computes the Richardson number by averaging both horizontal velocity components as well as the temperature and salinity onto the B-grid corner point (red dot), and then computing the buoyancy frequency and vertical shear, again at the corner point. The A-grid approach can lead to the production of grid noise due to noise in the Richardson number that leads to noise in the mixing coefficients. The B-grid approach has no such noise. Note that in MOM6, I labels a field centered at the east face of tracer cell $T_{i,j}$, and J labels a field centered at the northern face of this cell.

1506 by interpolating just the velocity or just the tracer, one can realize noisy Richardson numbers
1507 that then lead to noisy tracer and velocity fields. That is, a computational grid scale
1508 mode can arise from vertical diffusion schemes based on the Richardson number.

1509 To see how noise can arise, consider a C-grid as in MOM6 with a tracer cell, $T_{i,j}$,
1510 that has zero vertical density stratification and so vanishing N^2 . Assume that this tracer
1511 is surrounded by horizontal C-grid velocities that have nonzero vertical shears, and by
1512 adjacent tracer cells with nonzero vertical density stratification. In this case the discrete
1513 Richardson number at $T_{i,j}$ is zero, which leads to a large eddy diffusivity from the Jackson
1514 et al. (2008) scheme (as well as most other Richardson number based mixing schemes),
1515 thus maintaining the zero vertical tracer stratification at $T_{i,j}$. Yet because the adjacent
1516 tracer cells are assumed to have nonzero density stratification, the Richardson number
1517 interpolated to the $T_{i,j}$ faces, where the horizontal velocity is centered, will be nonzero
1518 and so the vertical shears of the horizontal velocity will remain nonzero. Likewise, the
1519 Richardson number at the tracer cells with nonzero vertical stratification remains nonzero,
1520 in which case the mixing can be small. We are led to a large vertical tracer diffusivity
1521 at $T_{i,j}$ yet with adjacent cells having small tracer diffusivity. Correspondingly, if the ver-
1522 tical stratification has a checkerboard pattern, then this pattern is not damped by the
1523 A-grid Richardson number stencil. Rather, the pattern can be enhanced, thus support-
1524 ing mixing-induced grid noise.

1525 ***Discrete Richardson number that does not introduce grid noise***

1526 Generating grid noise from the vertical diffusion scheme is quite disappointing since
 1527 we generally rely on diffusion to smooth fields rather than generate noise. Hence, for CM4X
 1528 we formulated an alternative discretization aimed at suppressing the computational mode.
 1529 We do so by interpolating the horizontal velocity components, as well as the tempera-
 1530 ture and salinity, to the tracer corner point; i.e., to the B-grid point. The buoyancy fre-
 1531 quency and vertical shear are then computed using the interpolated B-grid fields, as de-
 1532 picted in the right panel of Figure A4. In this manner, we couple across the tracer points
 1533 as well as velocity points, thus suppressing the A-grid computational mode.

1534 Analogous approaches have been used by various codes that use the A-grid Richard-
 1535 son number but perform a horizontal averaging prior to computing the diffusivities. For
 1536 example, earlier versions of MOM made use of this smoothing ever since the implemen-
 1537 tation of the Pacanowski and Philander (1981) shear mixing scheme, which is based on
 1538 the gradient Richardson number. Alternatively, one can average the noisy diffusivities
 1539 prior to using them in the tracer and velocity equations. Although these approaches can
 1540 be of some use for smoothing the noise, we find it more effective to eliminate the com-
 1541 putational mode at its origin rather than relying on downstream smoothing.

1542 Further examination of this modification is provided in (Reichl et al., 2024) with
 1543 a focus on the tropical Pacific. While the new approach helps to mitigate grid scale noise,
 1544 it has little impact on the time mean stratification near the equator.

1545 **A13 Parameterization of local internal tide-induced mixing**

1546 Mixing from breaking internal tides is an important part of the parameterized mix-
 1547 ing of ocean tracers and momentum in CM4.0 and CM4X, with the parameterization
 1548 following A. Melet et al. (2013). The first step of this parameterization, following St. Lau-
 1549 rent et al. (2002), computes the local energy conversion from the astronomically-forced
 1550 barotropic tide into baroclinic internal tides (dimensions of energy per time, or power,
 1551 per unit area)

$$1552 E_{\text{conversion}}(x, y, t) = (\rho_0 \kappa / 2) \langle h^2 \rangle \langle U_{\text{tides}}^2 \rangle N_b, \quad (\text{A3})$$

1553 where $\rho_0 = 1035 \text{ kg m}^{-3}$ is the Boussinesq reference density, $\langle h^2 \rangle$ is the variance of the
 1554 bottom topography (used as a measure of the topographic roughness), $\langle U_{\text{tides}}^2 \rangle$ is the vari-
 1555 ance of the barotropic tides, and N_b is the buoyancy frequency just above the bottom
 1556 boundary layer. The topography variance, tide variance, and bottom buoyancy frequency
 1557 are functions of horizontal position, with N_b also a function of time. While the charac-
 1558 teristic topographic wavenumber, κ (dimension of inverse length), could in principle also
 1559 be provided as a function of horizontal position, we use a spatially-constant value here
 1560 as a means to calibrate the globally-integrated energy conversion rate. Motivated by an
 1561 estimate that 1 TW of deep ocean barotropic-to-baroclinic tidal conversion is required
 1562 to reconcile energy budgets with observations (Egbert & Ray, 2000), we calibrate the
 1563 scheme to render a globally-integrated conversion rate of roughly 1 TW.

1564 We compute the topographic roughness by fitting a series of small planes to each
 1565 local patch of the topography in Figure A2, and then calculating the slope of the plane.
 1566 Figure A5 shows the resulting roughness field for CM4X-p125. For the tide amplitude,
 1567 we use the TPX09 version of the Oregon State University barotropic shallow water tide
 1568 model that uses the methods from Egbert and Erofeeva (2002). This model has a $1/30^\circ$
 1569 grid spacing and we make use of the following eight tidal harmonics: M2, S2, N2, K2,
 1570 K1, O1, P1, Q1. We interpolated the tide amplitudes onto the MOM6 grids, with Fig-
 1571 ure A5 showing the tide amplitude for the p125 grid.

1572 Once the tide amplitude has been interpolated to the model grid, we follow St. Lau-
 1573 rent et al. (2002) and Simmons et al. (2004) in calibrating the global tidal conversion
 1574 rate to be roughly 1 TW. This calibration is performed separately in the p25 and p125

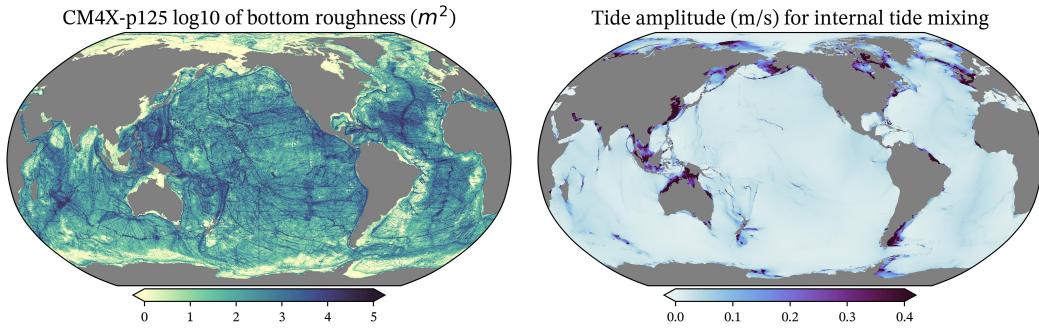


Figure A5. Left panel: Log10 of the bottom roughness (units of squared meters) for the bottom topography for CM4X-p125 from Figure A2. This field is used in the tide-induced mixing parameterization of Section A13. It is notable that straight lines appear from ship data, particularly visible in the high latitudes. Right panel: Amplitude (m/s) of the tides used for the tidal induced mixing in MOM6 for the CM4X-p125 configuration. Note the larger amplitudes on the continental shelves.

configurations by setting values for the characteristic topographic wavenumber, κ (see equation (A3)) responsible for the generation of internal tides. For the piControl simulations, Figure A6a reveals that CM4X-p125 settles onto roughly 1.17 TW globally (0.74 TW in the deep ocean) of energy conversion after 100 years, whereas CM4X-p25 equilibrates to roughly 1.16 TW globally (0.68 TW at depth) after 200 years. Evidently, the two models have their global internal tide mixing parameterizations calibrated to within 1% of each another.

In hindsight, we realized the appropriate observational target was roughly 1 TW of global *deep* ocean tidal conversion (Egbert & Ray, 2000) while the values diagnosed from the CM4X simulations also included conversion in shallow regions. Excluding all regions shallower than 2500 m (using a 2° smoothing of model bathymetry to yield a mask qualitatively similar to that used by Egbert and Ray (2000)), it was discovered that the tidal conversion in the deep ocean was only about 0.74 TW in CM4X-p125 and 0.68 TW in CM4X-p25, suggesting κ should be recalibrated upwards by about 30%.

Figure A6a shows that there is no sizable signal when moving from the piControl to historical. In contrast, the SSP5-8.5 simulation shows a clear signal of enhanced bottom stratification that, in turn, leads to enhanced tidal conversion. Virtually all of this increased conversion is attributable to shallow continental shelves (compare Figure A6a and b), where the bottom stratification responds relatively quickly to enhanced surface warming. This 5% increase in tidal conversion due to global warming is an example of the poorly understood climate-mixing feedback processes that are enabled by increasingly comprehensive and state-dependent mixing parameterizations (A. V. Melet et al., 2022).

To relate this energy input into the internal wave field to irreversible mixing, a constant fraction $q = 30\%$ is first assumed to produce turbulence kinetic energy (primarily through the breaking of high vertical modes) with a vertical structure function that follows Polzin (2009). Only a fraction, $\Gamma \leq 0.2$, (tapered to zero in regions of very weak stratification) of this locally-produced turbulent kinetic energy is used to drive irreversible diapycnal mixing, which is implemented as an enhanced vertical diffusivity. The remaining 70% of the energy input into low mode internal tides is not explicitly accounted for and is instead assumed to be implicit in a latitudinally-varying background diffusivity, k_d .

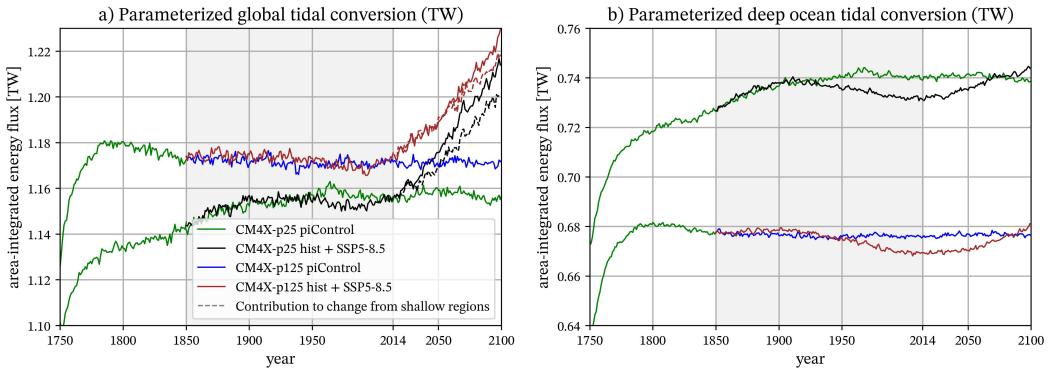


Figure A6. Time series for the energy conversion from the barotropic tide to baroclinic internal tides, as per equation (A3), integrated over (a) the global ocean and (b) regions deeper than 2500 m. The CM4X-p125 piControl simulation equilibrates to roughly 1.17 TW, whereas CM4X-p25 equilibrates to a bit less than 1.16 TW. There is a minor difference in the conversion rates for the historical versus piControl simulations. However, there is a clear signal of enhanced tidal conversion in the SSP5-8.5 simulation, arising primarily from an increase in the area integrated bottom stratification in shallow regions. The gray shaded region is the historical portion of the simulation (1850–2014).

1607

A14 Parameterization of baroclinic eddies in the mixed layer

1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621

We make use of the mixed layer eddy (MLE) restratification scheme from Fox-Kemper et al. (2008) and Fox-Kemper et al. (2011). This scheme provides a parameterization of the mixed layer baroclinic instability as studied by Boccaletti et al. (2007). In general, the upper ocean warms when the MLE restratification is strengthened (heat remains in the upper ocean), and the upper ocean cools when the MLE restratification is weakened (heat can be pumped into the ocean interior). Section 2.2.2 of Adcroft et al. (2019) details the implementation of MLE in OM4.0, and we closely follow that approach for CM4X. The MLE scheme in OM4.0 is set to be (arguably unphysically) strong in order to counteract biases in CM4.0 partly due to under-resolved upper ocean restratification by mesoscale eddies. For CM4X, we reduced the strength of the scheme, anticipating that the eddies in CM4X-p125 would be stronger than CM4.0 or CM4X-p25, thus enabling us to reduce the restratification effects from the submesoscale parameterization. In addition to heat uptake, we monitored the mixed layer depth and sea surface temperature when testing parameter changes.

1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632

There are two parameters in the MLE scheme that we tested in developing CM4X. The first is the frontal length, with a stronger restratification tendency for smaller fronts and weaker for larger fronts. The study by Bodner et al. (2023) updates the Fox-Kemper et al. (2008) scheme by providing a dynamical closure for the front length rather than prescribing it. However, we did not make use of Bodner et al. (2023) given that the CM4X project started a few years prior to the implementation of the updates to MOM6. Hence, the frontal length is a prescribed parameter in OM4.0 and CM4X. The second parameter we tested concerns the timescale for the running time mean filter applied to the mixed layer depth, with details provided in Section 2.2.2 of Adcroft et al. (2019). Lengthening the decay time increases the strength of the restratification since the longer decay allows the MLE scheme to remain enabled for an increased period of time.

1633
1634

We performed non-exhaustive tests in a prototype of CM4X-p25 using a front length of 1000 m and 2000 m. Yet we found relatively little sensitivity to this parameter com-

MODEL	MLE_FRONT_LENGTH	MLE_MLD_DECAY_TIME
CM4.0	500 m	2.592×10^6 s = 30 days
CM4X	500 m	1.728×10^6 s = 20 days

Table A7. Summary of the key MLE parameters as used in CM4.0 and CM4X.

pared to the OM4.0 setting of 500 m, thus prompting us to retain 500 m for CM4X. In contrast, the ocean heat uptake and sea surface temperature are sensitive to the time scale used for the low pass time filter. We weakened the scheme in CM4X by reducing the time scale from the 30 days used in OM4.0 to 20 days in CM4X. This reduction of the MLE strength is partly motivated given that the CM4X-p125 ocean has stronger mesoscale eddies, thus reducing the need for MLE to perform the bulk of upper ocean restratification. Table A7 summarizes the MLE settings used in CM4X and CM4.0.

Appendix B Open Research

Software comprising the model as well as the software used for creating the figures will be placed on Zenodo at the revision stage of this work.

Observation-based datasets used in this paper are cited locally. We are indebted to the many efforts of the various programs providing observational-based data used to help evaluate these simulations, including the following.

- The Argo program provides data that were collected and made freely available by the International Argo Program and the national programs that contribute to it, with access available from

<http://www.argo.ucsd.edu> and <http://argo.jcommops.org>

The Argo Program is part of the Global Ocean Observing System.

- OSNAP data were collected and made freely available by the OSNAP (Overturning in the Subpolar North Atlantic Program) project and all the national programs that contribute to it (www.o-snap.org). The DOI for this data set is

<https://doi.org/10.35090/gatech/70342>

Acknowledgments

This project started in May 2020, during the early stages of the Covid-19 pandemic shutdown. We are grateful to the GFDL computer operations team for keeping the computational resources reliable during the shutdown. We thank the GFDL management for providing the computer resources needed for the development and analysis documented here. We thank John Dunne, Matthew Harrison, and Tony Rosati for very helpful comments on early drafts of this manuscript. Krista Dunne, Sergey Malyshev and Chris Milly kindly provided expertise in helping to update the rivers and lakes for use with the C192 atmosphere coupled to the p125 ocean. The statements, findings, conclusions, and recommendations are those of the author(s) and do not necessarily reflect the views of the National Oceanic and Atmospheric Administration, or the U.S. Department of Commerce.

- The statements, findings, conclusions, and recommendations are those of the author(s) and do not necessarily reflect the views of the National Oceanic and Atmospheric Administration, or the U.S. Department of Commerce.

- A.A. was supported by Award NA18OAR4320123 from the National Oceanic and Atmospheric Administration, U.S. Department of Commerce.
- R.B. was supported under NSF Division of Polar Programs Grant NSF2319828.
- C.Y.C. was supported by Award NA19OAR4310365 from the National Oceanic and Atmospheric Administration, U.S. Department of Commerce.
- H.F.D. was supported by the NOAA Climate and Global Change Postdoctoral Fellowship Program, administered by UCAR's Cooperative Programs for the Advancement of Earth System Science (CPAESS) under Award NA18NWS4620043B.
- H.K. acknowledges the support from Natural Environment Research Council grants NE/T013494/1 and NE/W001543/1.
- M.L. was supported by award NA18OAR4320123 and NA23OAR4320198 from the National Oceanic and Atmospheric Administration, U.S. Department of Commerce.
- G.A.M was supported by NSF (PLR-1425989) and UKRI (MR/W013835/1).
- A.S. was supported by Schmidt Sciences, LLC under the M²LInES project.
- K.E.T acknowledges support from the Southern Ocean Carbon and Climate Observations and Modeling (SOCCOM) Project under NSF Awards PLR-1425989 and OPP-1936222 and 2332379.
- L.Z. was supported by Schmidt Sciences, LLC under the M²LInES project, NSF grant OCE 1912357 and NOAA CVP NA19OAR4310364.
- W.Z. was supported by the National Science Foundation under Grant Number F1240-01(NSF OCE 1912357). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

- Adcroft, A., Anderson, W., Blanton, C., Bushuk, M., Dufour, C. O., Dunne, J. P., ... Zhang, R. (2019). The GFDL global ocean and sea ice model OM4.0: Model description and simulation features. *Journal of Advances in Modeling the Earth System, JAMES*. doi: 10.1029/2019MS001726
- Adcroft, A., & Campin, J.-M. (2004). Rescaled height coordinates for accurate representation of free-surface flows in ocean circulation models. *Ocean Modelling*, 7, 269–284. doi: 10.1016/j.ocemod.2003.09.003
- Adcroft, A., Hill, C., & Marshall, J. (1997). Representation of topography by shaved cells in a height coordinate ocean model. *Monthly Weather Review*, 125, 2293–2315. doi: 10.1175/1520-0493(1997)125<2293:ROTBSC>2.0.CO;2
- Adler, R. F., Sapiano, M. R., Huffman, G. J., Wang, J. J., Gu, G., Bolvin, D., ... Shin, D.-B. (2018). The global precipitation climatology project (GPCP) monthly analysis (new version 2.3) and a review of 2017 global precipitation. *Atmosphere*, 9. doi: 10.3390/atmos9040138
- Andrews, T., Gregory, J. M., Paynter, D., Silvers, L. G., Zhou, C., Mauritsen, T., ... Titchner, H. (2018). Accounting for changing temperature patterns increases historical estimates of climate sensitivity. *Geophysical Research Letters*, 45, 8490–8499. doi: 10.1029/2018GL078887
- Argo. (2023). *Argo float data and metadata from Global Data Assembly Centre (Argo GDAC)* [dataset]. SEANOE. Retrieved from <https://www.seanoe.org/data/00311/42182/> doi: 10.17882/42182
- Armour, K. C., Bitz, C. M., & Roe, G. H. (2013). Time-varying climate sensitivity from regional feedbacks. *Journal of Climate*, 26, 4518–4534. doi: 10.1175/JCLI-D-12-00544.1
- Armour, K. C., Marshall, J., Scott, J. R., Donohoe, A., & Newsom, E. R. (2016). Southern Ocean warming delayed by circumpolar upwelling and equatorward transport. *Nature Geosciences*, 9. doi: 10.1038/NGEO2731
- Arora, V. K., Katavouta, A., Williams, R. G., Jones, C. D., Brovkin, V., Friedling-

- stein, P., ... Ziehn, T. (2019, December). Carbon-concentration and carbon-climate feedbacks in CMIP6 models, and their comparison to CMIP5 models. *Biogeosciences Discussions*, 1–124. Retrieved 2020-06-07, from <https://www.biogeosciences-discuss.net/bg-2019-473/> (Publisher: Copernicus GmbH) doi: <https://doi.org/10.5194/bg-2019-473>
- Banks, H. T., Stark, S., & Keen, A. B. (2006). The adjustment of the coupled climate model HadGEM1 toward equilibrium and the impact on global climate. *Journal of Climate*, 20, 5815–5826. doi: 10.1175/2007JCLI1688.1
- Beadling, R. L. (2023). Global consequences of regional connectivity along the Antarctic margin. *Journal of Geophysical Research: Oceans*, 128. doi: 10.1029/2023JC019908
- Beadling, R. L., Krasting, J. P., Griffies, S. M., Hurlin, W. J., Bronselear, B., Russell, J. L., ... Winton, M. (2022). Importance of the Antarctic Slope Current in the Southern Ocean response to ice sheet melt and wind stress change. *Journal of Geophysical Research - Oceans*, 127. doi: 10.1029/2021jc017608
- Beadling, R. L., Russell, J. L., Stouffer, R. J., Mazloff, M., Talley, L. D., Goodman, P. J., ... Pandde, A. (2020). Representation of Southern Ocean properties across coupled model intercomparison project generations: CMIP3 to CMIP6. *Journal of Climate*, 33, 6555–6581. doi: 10.1175/JCLI-D-19-0970.1
- Bian, C., Jing, Z., Wang, H., Wu, L., Chen, Z., Gan, B., & Yang, H. (2023). Oceanic mesoscale eddies as crucial drivers of global marine heatwaves. *Nature Communications*, 14, 2970. doi: 10.1038/s41467-023-38811-z
- Bleck, R. (2002). An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates. *Ocean Modelling*, 4, 55–88. doi: 10.1016/S1463-5003(01)00012-9
- Boccaletti, G., Ferrari, R., & Fox-Kemper, B. (2007). Mixed layer instabilities and restratification. *Journal of Physical Oceanography*, 35, 1263–1278. doi: 10.1175/JPO3101.1
- Bodner, A., Fox-Kemper, B., Johnson, L., van Rokel, L. P., McWilliams, J. C., Sullivan, P. P., ... Dong, J. (2023). Modifying the mixed layer eddy parameterization to include frontogenesis arrest by boundary layer turbulence. *Journal of Physical Oceanography*, 53, 323–339. doi: 10.1175/JPO-D-21-0297.s1
- Boville, B. A., & Gent, P. R. (1998). The NCAR climate system model, version one. *Journal of Climate*, 11, 1115–1130. doi: 10.1175/1520-0442(1998)011<1115:TNCNSMV>2.0.CO;2
- Briegleb, B. P., & Light, B. (2007). *A delta-Eddington multiple scattering parameterization for solar radiation in the sea ice component of the Community Climate System Model* (Technical Report Nos. TN-4721+STR). Boulder, USA: National Center for Atmospheric Research.
- Burls, N. J., Muir, L., Vincent, E. M., & Fedorov, A. V. (2016). Extra-tropical origin of equatorial Pacific cold bias in climate models with links to cloud albedo. *Climate Dynamics*, 49, 2093–2113. doi: 10.1007/s00382-016-3435-6
- Caesar, L., Rahmstorf, S., Robinson, A., Feulner, G., & Saba, V. (2018). Observed fingerprint of a weakening Atlantic Ocean overturning circulation. *Nature*, 556, 191–196. doi: 10.1038/s41586-018-0006-5
- Cess, R., Potter, G., & Coauthors. (1990). Intercomparison and interpretation of climate feedback processes in 19 atmospheric general circulation models. *Journal of Geophysical Research*, 16601–16615.
- Cessi, P. (2019). The global overturning circulation. *Annual Review of Marine Science*, 11, 249–270. doi: 10.1146/annurev-marine-010318-095241
- Charney, J. G. (1971). Geostrophic turbulence. *Journal of the Atmospheric Sciences*, 28, 1087–1095. doi: 10.1175/1520-0469(1971)028<1087:GT>2.0.CO;2
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., ... Vitart, F. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*

- 1778 *Society*, 137, 553–597. doi: 10.1002/qj.828
- 1779 Delworth, T. L., Broccoli, A. J., Rosati, A., Stouffer, R. J., Balaji, V., Beesley, J. A.,
 1780 ... Zhang, R. (2006). GFDL's CM2 global coupled climate models - Part 1:
 1781 Formulation and simulation characteristics. *Journal of Climate*, 19, 643–674.
 1782 doi: 10.1175/JCLI3629.1
- 1783 Delworth, T. L., Cooke, W. F., Adcroft, A., Bushuk, M., Chen, J.-H., Dunne, K. A.,
 1784 ... others (2020). SPEAR: The next generation GFDL modeling system for
 1785 seasonal to multidecadal prediction and projection. *Journal of Advances in*
 1786 *Modeling Earth Systems*, 12(3), e2019MS001895.
- 1787 Delworth, T. L., Rosati, A., Anderson, W., Adcroft, A. J., Balaji, V., Benson, R., ...
 1788 Zhang, R. (2012). Simulated climate and climate change in the GFDL CM2.5
 1789 high-resolution coupled climate model. *Journal of Climate*, 25, 2755–2781. doi:
 1790 10.1175/JCLI-D-11-00316
- 1791 Deppenmeier, A.-L., Bryan, F. O., Kessler, W. S., & Thompson, L. (2021).
 1792 Modulation of cross-isothermal velocities with ENSO in the Tropical Pa-
 1793 cific Cold Tongue. *Journal of Physical Oceanography*, 51, 1559–1574. doi:
 1794 10.1175/JPO-D-20-0217.1
- 1795 Deremble, B., Hogg, A. M., Berloff, P., & Dewar, W. K. (2012). On the applica-
 1796 tion of no-slip lateral boundary conditions to ‘coarsely’ resolved ocean models.
 1797 *Ocean Modelling*, 39, 411–415. doi: 10.1016/j.ocemod.2011.05.002
- 1798 Doddridge, E. W., Marshall, D. P., & Hogg, A. M. (2016). Eddy Cancellation of
 1799 the Ekman Cell in Subtropical Gyres. *Journal of Physical Oceanography*, 46,
 1800 2995–3010. doi: 10.1175/JPO-D-16-0097.1
- 1801 Drake, H. F., Bailey, S., Dussin, R., Griffies, S. M., Krasting, J. P., MacGilchrist,
 1802 G. A., ... Zika, J. D. (2024, April). *Water mass transformation bud-
 1803 getts in finite-volume generalized vertical coordinate ocean models*. Retrieved
 1804 2024-05-31, from [https://essopenarchive.org/users/242115/
 1805 articles/788549-water-mass-transformation-budgets-in-finite
 1806 -volume-generalized-vertical-coordinate-ocean-models?commit=](https://essopenarchive.org/users/242115/articles/788549-water-mass-transformation-budgets-in-finite-volume-generalized-vertical-coordinate-ocean-models?commit=e6178873ac99d67c98963b54ba60699d57b7703f) doi: 10.22541/essoar
 1807 .171284935.57181910/v1
- 1808 Dunne, J. P., Horowitz, L. W., Adcroft, A., Ginoux, P., Held, I. M., John, J. G., ...
 1809 Zhao, M. (2020). The GFDL Earth system model version 4.1 (GFDL-ESM4.1):
 1810 Model description and simulation characteristics. *Journal of Advances in Mod-
 1811 eling Earth Systems*, 12. doi: 10.1029/2019MS002015
- 1812 Egbert, G. D., & Erofeeva, S. Y. (2002). Efficient inverse modeling of barotropic
 1813 ocean tides. *Journal of Atmospheric and Oceanic Technology*, 19, 183–204.
 1814 doi: 10.1175/1520-0426(2002)019<0183:EIMOBO>2.0.CO;2
- 1815 Egbert, G. D., & Ray, R. D. (2000, June). Significant dissipation of tidal energy
 1816 in the deep ocean inferred from satellite altimeter data. *Nature*, 405(6788),
 1817 775–778. Retrieved 2022-05-12, from <https://www.nature.com/articles/35015531> (Number: 6788 Publisher: Nature Publishing Group) doi:
 1818 10.1038/35015531
- 1819 Eyring, V., Bony, S., Meehl, J. A., Senior, C., Stevens, B., Stouffer, R. J., & Taylor,
 1820 K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase
 1821 6 (CMIP6) experimental design and organisation. *Geoscientific Model Develop-
 1822 ment*, 9, 1937–1958. Retrieved from <http://www.geosci-model-dev.net/9/1937/2016/> doi: 10.5194/gmd-9-1937-2016
- 1823 Forster, P., Storelvmo, T., Armour, K., Collins, W., Dufresne, J.-L., Frame, D., ...
 1824 Zhang, H. (2021). The earth’s energy budget, climate feedbacks, and climate
 1825 sensitivity. In *Climate Change 2021: The Physical Science Basis. Contribution
 1826 of Working Group I to the Sixth Assessment Report of the Intergovernmen-
 1827 tal Panel on Climate Change* (pp. 923–1051). Cambridge, United Kingdom:
 1828 Cambridge University Press. doi: 10.1017/9781009157896.009
- 1829 Fox-Kemper, B., Danabasoglu, G., Ferrari, R., Griffies, S. M., Hallberg, R. W., Hol-

- land, M. M., ... Samuels, B. (2011). Parameterization of mixed layer eddies. III: Global implementation and impact on ocean climate simulations. *Ocean Modelling*, 39, 61–78. doi: 10.1016/j.ocemod.2010.09.002
- Fox-Kemper, B., Ferrari, R., & Hallberg, R. (2008). Parameterization of mixed layer eddies. I: Theory and diagnosis. *Journal of Physical Oceanography*, 38, 1145–1165. doi: 10.1175/2007JPO3792.1
- Fox-Kemper, B., Hewitt, H. T., Xiao, C., Aoalgeirsdóttir, G., Drijfhout, S. S., Edwards, T. L., ... Yu, Y. (2021). Ocean, cryosphere and sea level change. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 1211–1362). Cambridge, United Kingdom: Cambridge University Press. doi: 10.1017/9781009157896.011
- Ganachaud, A., & Wunsch, C. (2000). Improved estimates of global ocean circulation, heat transport and mixing from hydrographic data. *Nature*, 408, 453–456. doi: 10.1038/35044048
- GEBCO Compilation Group. (2021). GEBCO 2021 Grid [dataset]. Retrieved from https://www.bodc.ac.uk/data/published_data_library/catalogue/10.5285/c6612cbe-50b3-0cff-e053-6c86abc09f8f/ doi: 10.5285/c6612cbe-50b3-0cff-e053-6c86abc09f8f
- Gent, P. R., & McWilliams, J. C. (1990). Isopycnal mixing in ocean circulation models. *Journal of Physical Oceanography*, 20, 150–155. doi: 10.1175/1520-0485(1990)020<0150:IMIOCM>2.0.CO;2
- Gent, P. R., Willebrand, J., McDougall, T. J., & McWilliams, J. C. (1995). Parameterizing eddy-induced tracer transports in ocean circulation models. *Journal of Physical Oceanography*, 25, 463–474. doi: 10.1175/1520-0485(1995)025<0463:PEITTI>2.0.CO;2
- Ghimire, B., Williams, C. A., Masek, J., Gao, F., Wang, Z., Schaaf, C., & He, T. (2014). Global albedo change and radiative cooling from anthropogenic land cover change, 1700 to 2005 based on MODIS, land use harmonization, radiative kernels, and reanalysis. *Geophysical Research Letters*, 41, 9087–9096. doi: 10.1002/2014GL061671
- Gleckler, P. J., Taylor, K. E., & Doutriaux, C. (2008). Performance metrics for climate models. *Journal of Geophysical Research*, 113. doi: 10.1029/2007jd008972
- Gnanadesikan, A., Slater, R. D., Swathi, P. S., & Vallis, G. K. (2005). The energetics of ocean heat transport. *Journal of Climate*, 17, 2604–2616. doi: 10.1175/JCLI3436.1
- Gordon, C., Cooper, C., Senior, C., Banks, H., Gregory, J. M., Johns, T. C., ... Woods, R. A. (2000). The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dynamics*, 16, 147–168. doi: 10.1007/s003820050010
- Gregory, J. M. (2000). Vertical heat transports in the ocean and their effect on time-dependent climate change. *Climate Dynamics*, 15, 501–515. doi: 10.1007/s003820000059
- Gregory, J. M., Bouttes, N., Griffies, S. M., Haak, H., Hurlin, W. J., Jungclaus, J., ... Winton, M. (2016). The Flux-Anomaly-Forced Model Intercomparison Project (FAFMIP) contribution to CMIP6: investigation of sea-level and ocean climate change in response to co_2 forcing. *Geoscientific Model Development*, 9, 3993–4017. doi: 10.5194/gmd-9-3993-2016
- Gregory, J. M., Griffies, S. M., Hughes, C. W., Lowe, J. A., Church, J. A., Fukimori, I., ... van de Wal, R. S. W. (2019). Concepts and terminology for sea level-mean, variability and change, both local and global. *Surveys in Geophysics*, 40, 1251–1289. doi: 10.1007/s10712-019-09555-7
- Gregory, J. M., & Tailleux, R. (2011). Kinetic energy analysis of the response of the Atlantic meridional overturning circulation to co_2 -forced climate change. *Cli-*

- 1943 Hirschi, J. J., Barnier, B., Böning, C., Biastoch, A., Blaker, A. T., Coward, A., ...
 1944 Xu, X. (2020). The Atlantic meridional overturning circulation in high-
 1945 resolution models. *Journal of Geophysical Research: Oceans*, 125. doi:
 1946 10.1029/2019JC015522
- 1947 Hofstadter, D. R. (1979). *Gödel, Escher, Bach: An Eternal Golden Braid*. Basic
 1948 Books. (777 pages)
- 1949 Holmes, R. M., Zika, J., & England, M. H. (2019). Diathermal heat transport in a
 1950 global ocean model. *Journal of Physical Oceanography*, 49, 141–161. doi: 10
 1951 .1175/JPO-D-18-0098.1
- 1952 Holmes, R. M., Zika, J., Griffies, S. M., A. McC. Hogg, Kiss, A. E., & England,
 1953 M. H. (2020). The geography of numerical mixing in a suite of global
 1954 ocean models. *Journal of Advances in Modeling Earth Systems*, 13. doi:
 1955 10.1029/2020MS002333
- 1956 Holmes, R. M., Zika, J. D., Griffies, S. M., Hogg, A. M., Kiss, A. E., & England,
 1957 M. H. (2021). The geography of numerical mixing in a suite of global
 1958 ocean models. *Journal of Advances in Modeling Earth Systems*, 13(7),
 1959 e2020MS002333. Retrieved from <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020MS002333> (e2020MS002333 2020MS002333) doi:
 1960 <https://doi.org/10.1029/2020MS002333>
- 1961 Holzer, M., DeVries, T., & de Lavergne, C. (2021). Diffusion controls the ventilation
 1962 of a pacific shadow zone above abyssal overturning. *Nature Communications*,
 1963 12(1), 4348.
- 1964 Hsieh, T., Vecchi, G. A., Yang, W., Held, I. M., & Garner, S. T. (2020). Large-scale
 1965 control on the frequency of tropical cyclones and seeds: a consistent relation-
 1966 ship across a hierarchy of global atmospheric models. *Climate Dynamics*, 55,
 1967 3177–3196. doi: 10.1007/s00382-020-05446-5
- 1968 Huang, B., Liu, C., Banzon, V., Freeman, E., Graham, G., Hankins, B., ... Zhang,
 1969 H.-M. (2020). Improvements of the daily optimum interpolation sea surface
 1970 temperature (DOISST) Version 2.1. *Journal of Climate*, 34, 2923–2939. doi:
 1971 10.1175/JCLI-D-20-0166.1
- 1972 Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore,
 1973 J. H., ... Zhang, H.-M. (2017). *NOAA Extended Reconstructed Sea Surface
 1974 Temperature (ERSST), Version 5* (dataset). U.S. Department of Commerce,
 1975 Washington, D.C.: NOAA National Centers for Environmental Information.
 1976 Retrieved from <https://www.esrl.noaa.gov/psd/> doi: 0.7289/V5T72FNM
- 1977 Hughes, C. W., Fukumori, I., Griffies, S. M., Huthnance, J. M., Minobe, S., Spence,
 1978 P., ... Wise, A. (2019). Sea level and the role of coastal trapped waves in
 1979 mediating the influence of the open ocean on the coast. *Surveys in Geophysics*,
 1980 40, 1467–1492. doi: 10.1007/s10712-019-09535-x
- 1981 Huw Davies, J. (2013). Global map of solid Earth surface heat flow. *Geochemistry,
 1982 Geophysics, Geosystems*, 14, 4608–4622. doi: 10.1002/ggge.20271
- 1983 Hyder, P., Edwards, J. M., Allan, R. P., Hewitt, H. T., Bracegirdle, T. J., Gregory,
 1984 J. M., ... Belcher, S. E. (2018). Critical Southern Ocean climate model biases
 1985 traced to atmospheric model cloud errors. *Nature Communications*, 9. doi:
 1986 10.1038/s41467-018-05634-2
- 1987 Ilicak, M., Adcroft, A. J., Griffies, S. M., & Hallberg, R. W. (2012). Spurious di-
 1988 aneutral mixing and the role of momentum dissipation. *Ocean Modelling*, 45–
 1989 46, 37–58. doi: 10.1016/j.ocemod.2011.10.003
- 1990 Irving, D., Hobbs, W., Church, J., & Zika, J. (2021). A mass and energy conser-
 1991 vation analysis of drift in the CMIP6 ensemble. *Journal of Climate*, 34, 3157–
 1992 3170. doi: 10.1175/JCLI-D-20-0281.1
- 1993 Ivey, G., & Imberger, J. (1991). On the nature of turbulence in a stratified fluid.
 1994 Part I: The energetics of mixing. *Journal of Physical Oceanography*, 21(5),
 1995 650–658.
- 1996 Jackson, L., Hallberg, R. W., & Legg, S. (2008). A parameterization of shear-driven
 1997

- turbulence for ocean climate models. *Journal of Physical Oceanography*, 38, 1033–1053. doi: 10.1175/2007JPO3779.1

Jeevanjee, N., Hassanzadeh, P., Hill, S., & Sheshadri, A. (2017). A perspective on climate model hierarchies. *Journal of Advances in Modeling Earth Systems*, 9, 1760–1771. doi: 10.1002/2017MS001038

Johnson, G. C., & Purkey, S. G. (2024). Refined estimates of global ocean deep and abyssal decadal warming trends. *Geophysical Research Letters*, 51. doi: 10.1029/2024GL111229

Khatiwala, S. (2023). Fast spin-up of geochemical tracers in ocean circulation and climate models. *Journal of Advances in Modeling Earth Systems (JAMES)*, 15. doi: 10.1029/2022MS003447

Khatiwala, S. (2024). Efficient spin-up of Earth System Models using sequence acceleration. *Science Advances*, 10. doi: 10.1126/sciadv.adn2839

Kirtman, B. P., Bitz, C., Bryan, F., Collins, W., Dennis, J., Hearn, N., ... Vertenstein, M. (2012). Impact of ocean model resolution on CCSM climate simulations. *Climate Dynamics*, 39, 1303–1328. doi: 10.1007/s00382-012-1500-3

Kiss, A. E., Hogg, A. M., Hannah, N., Dias, F. B., Brassington, G., Chamberlain, M. A., ... Zhang, X. (2020). ACCESS-OM2: A global ocean-sea ice model at three resolutions. *Geoscientific Model Development*. doi: 10.5194/gmd-13-401-2020

Kjellsson, J., & Zanna, L. (2017). The impact of horizontal resolution on energy transfers in global ocean models. *Fluids*, 2(3), 45. doi: 10.3390/fluids2030045

Krasting, J. P., Griffies, S. M., Tesdal, J.-E., MacGilchrist, G., Beadling, R. L., & Little, C. M. (2024). Steric sea level rise and relationships with model drift and water mass representation in GFDL CM4 and ESM4. *Journal of Climate*, accepted.

Krasting, J. P., Stouffer, R. J., Griffies, S. M., Hallberg, R. W., Malyshev, S. L., Samuels, B. L., & Sentman, L. T. (2018). Role of ocean model formulation in climate response uncertainty. *Journal of Climate*, 22, 9313–9333. doi: 10.1175/JCLI-D-18-0035.1

Lauritzen, P. H., Kevlahan, N. K.-R., Tonizzo, T., Eldred, C., Dubos, T., Gassmann, A., ... Bacmeister, J. T. (2022). Reconciling and improving formulations for thermodynamics and conservation principles in Earth System Models (ESMs). *Journal of Advances in Modeling Earth Systems*, 14. doi: 10.1029/2022MS003117

Lee, J., Gleckler, P. J., Ahn, M.-S., Ordonez, A., Ullrich, P. A., Sperber, K. R., ... Krasting, J. (2023). Objective evaluation of Earth system models: PCMDI Metrics Package (PMP) version 3. *EGUsphere*, 2023, 1–48. Retrieved from <https://egusphere.copernicus.org/preprints/2023/egusphere-2023-2720/> doi: 10.5194/egusphere-2023-2720

Legg, S., Hallberg, R. W., & Girton, J. B. (2006, January). Comparison of entrainment in overflows simulated by z-coordinate, isopycnal and non-hydrostatic models. *Ocean Modelling*, 11(1), 69–97. Retrieved 2018-12-20, from <http://www.sciencedirect.com/science/article/pii/S1463500304001064> doi: 10.1016/j.ocemod.2004.11.006

Lellouche, J.-M., Greiner, E., Bourdallé-Badie, R., Garric, G., Melet, A., Drévillon, M., ... Le Traon, P.-Y. (2021). The Copernicus global 1/12° oceanic and sea ice GLORYS12 reanalysis. *Frontiers in Earth Science*, 9. doi: 10.3389/feart.2021.698876

Li, G., Du, Y., Xu, H., & Ren, B. (2015). An intermodel approach to identify the source of excessive equatorial pacific cold tongue in cmip5 models and uncertainty in observational datasets. *Journal of Climate*, 28, 7630–7640. doi: 10.1175/JCLI-D-15-0168.1

Li, Q., Fox-Kemper, B., Øyvind Breivik, & Webb, A. (2017). Statistical models of global Langmuir mixing. *Ocean Modelling*, 113, 95–114. doi: <https://doi.org/10.1016/j.ocemod.2017.03.001>

- 2053 10.1016/j.ocemod.2017.03.016
- 2054 Li, Q., & Lee, S. (2017). A mechanism of mixed layer formation in the Indo-Western
2055 Pacific Southern Ocean: Preconditioning by an eddy-driven jet-scale over-
2056 turning circulation. *Journal of Physical Oceanography*, *47*, 2755–2772. doi:
2057 10.1175/JPO-D-17-0006.1
- 2058 Locarnini, R. A., Mishonov, A. V., Antonov, J. I., Boyer, T. P., Garcia, H. E., Bara-
2059 nova, O. K., ... Seidov, D. (2018). *World Ocean Atlas 2013, Volume 1: Temperature* (dataset). U.S. Department of Commerce, Washington, D.C.: NOAA/NESDIS.
- 2060 Loeb, N. G., Doelling, D. R., Wang, H., Su, W., Nguyen, C., Corbett, J. G., ...
2061 Seiji, K. (2018). Clouds and the Earth's radiant energy system (CERES)
2062 energy balanced and filled (EBAF) top-of-atmosphere (TOA) edition-4.0
2063 data product. *International Journal of Climatology*, *31*, 895–918. doi:
2064 10.1175/JCLI-D-17-0208.1
- 2065 Maher, P., Gerber, E. P., Medeiros, B., Merlis, T. M., Sherwood, S., Sheshadri, A.,
2066 ... Zurito-Gotor, P. (2019). Model hierarchies for understanding atmospheric
2067 circulation. *Reviews of Geophysics*, *57*, 250–280. doi: 10.1029/2018RG000607
- 2068 Manabe, S., Bryan, K., & Spelman, M. J. (1990). Transient response of a global
2069 ocean–atmosphere model to a doubling of atmospheric carbon dioxide. *Journal
2070 of Physical Oceanography*, *20*, 722–749. doi: 10.1175/1520-0485(1990)020<722:
2071 TROAGO>2.0.CO;2
- 2072 Manabe, S., & Stouffer, R. J. (1988). Two stable equilibria of a coupled ocean-
2073 atmosphere model. *Journal of Climate*, *1*(9), 841–866. doi: 10.1175/1520
2074 -0442(1988)001<0841:TSEOAC>2.0.CO;2
- 2075 Manabe, S., Stouffer, R. J., Spelman, M. J., & Bryan, K. (1991). Transient re-
2076 sponds of a coupled ocean–atmosphere model to gradual changes of atmo-
2077 spheric CO₂. Part 1: Annual mean response. *Journal of Climate*, *4*, 785–818.
2078 doi: 10.1175/1520-0442(1991)004<0785:TROACO>2.0.CO;2
- 2079 Marques, G. M., Loose, N., Yankovsky, E., Steinberg, J. M., Chang, C.-Y., Bhamidi-
2080 pati, N., ... Zanna, L. (2022). Neverworld2: An idealized model hierarchy
2081 to investigate ocean mesoscale eddies across resolutions. *Geoscientific Model
2082 Development*. doi: 10.5194/gmd-15-6567-2022
- 2083 Marshall, D. P., & Zanna, L. (2014). A conceptual model of ocean heat uptake un-
2084 der climate change. *Journal of Climate*, *27*, 8444–8465. doi: 10.1175/JCLI-D
2085 -13-00344.1
- 2086 Marshall, J., & Schott, F. (1999). Open-ocean convection: observations, theory, and
2087 models. *Reviews of Geophysics*, *37*, 1–64. doi: 10.1029/98RG02739
- 2088 McDougall, T. J., Barker, P. M., Holmes, R. M., Pawlowicz, R., Griffies, S. M., &
2089 Durack, P. J. (2021). The interpretation of temperature and salinity vari-
2090 ables in numerical ocean model output, and the calculation of heat fluxes
2091 and heat content. *Geoscientific Model Development*, *14*, 6445–6466. doi:
2092 10.5194/gmd-14-6445-2021
- 2093 Melet, A., Hallberg, R. W., Legg, S., & Polzin, K. (2013). Sensitivity of the Pacific
2094 Ocean state to the vertical distribution of internal-tide driven mixing. *Journal
2095 of Physical Oceanography*, *43*, 602–615. doi: 10.1175/JPO-D-12-055.1
- 2096 Melet, A. V., Hallberg, R., & Marshall, D. P. (2022, January). Chapter 2
2097 - The role of ocean mixing in the climate system. In M. Meredith &
2098 A. Naveira Garabato (Eds.), *Ocean Mixing* (pp. 5–34). Elsevier. Retrieved
2099 2023-01-29, from <https://www.sciencedirect.com/science/article/pii/B9780128215128000098> doi: 10.1016/B978-0-12-821512-8.00009-8
- 2100 Morlighem, M., Williams, C. N., Rignot, E., An, L., Arndt, J. E., Bamber, J. L., ...
2101 Zinglersen, K. B. (2017). Bedmachine v3: Complete bed topography and ocean
2102 bathymetry mapping of greenland from multibeam echo sounding combined
2103 with mass conservation. *Geophysical Research Letters*, *44*, 11051–11061. doi:
2104 10.1002/2017GL074954
- 2105
- 2106
- 2107

- 2108 Morrison, A. K., Saenko, O. A., Hogg, A. M., & Spence, P. (2013). The role of
 2109 vertical eddy flux in southern ocean heat uptake. *Geophysical Research Letters*, 40(20), 5445–5450.
 2110 Retrieved from <http://dx.doi.org/10.1002/2013GL057706> doi: 10.1002/2013GL057706
- 2112 Murray, R. J. (1996). Explicit Generation of Orthogonal Grids for Ocean Mod-
 2113 els. *Journal of Computational Physics*, 126(2), 251–273. Retrieved 2018-
 2114 05-14, from <http://www.sciencedirect.com/science/article/pii/S0021999196901369> doi: 10.1006/jcph.1996.0136
- 2116 Newsom, E., Zanna, L., & Gregory, J. (2023). Background pycnocline depth con-
 2117 strains future ocean heat uptake efficiency. *Geophysical Research Letters*, 50.
 2118 doi: 10.1029/2023GL105673
- 2119 Orr, J. C., Najjar, R. G., Aumont, O., Bopp, L., Bullister, J., Danabasoglu, G., ...
 2120 Yool, A. (2017). Biogeochemical protocols and diagnostics for the CMIP6
 2121 Ocean Model Intercomparison Project (OMIP). *Geoscientific Model Develop-
 2122 ment*, 10, 2169–2199. doi: 10.5194/gmd-10-2169-2017
- 2123 Osborn, T. R., & Cox, C. S. (1972). Oceanic fine structure. *Geophysical Fluid Dy-
 2124 namics*, 3(4), 321–345.
- 2125 Pacanowski, R. C., & Philander, G. (1981). Parameterization of vertical mixing in
 2126 numerical models of the tropical ocean. *Journal of Physical Oceanography*, 11,
 2127 1442–1451. doi: 10.1175/1520-0485(1981)011<1443:POVMIN>2.0.CO;2
- 2128 Polzin, K. L. (2009). An abyssal recipe. *Ocean Modelling*, 30(4), 298–309. doi: 10
 2129 .1016/j.ocemod.2009.07.006
- 2130 Pudig, M. P., Yung, C. K., Zika, J. D., & Holmes, R. M. (2023). Rectified ocean
 2131 heat uptake from oscillatory surface forcing. *Journal of Climate*, 36, 2663–
 2132 2680. doi: 10.1175/JCLI-D-22-0267.1
- 2133 Purkey, S., & Johnson, G. (2010). Warming of global abyssal and deep South-
 2134 ern Ocean waters between the 1990s and 2000s: contributions to global
 2135 heat and sea level rise budgets. *Journal of Climate*, 23, 6336–6351. doi:
 2136 10.1175/2010JCLI3682.1
- 2137 Purkey, S., & Johnson, G. (2012). Global contraction of Antarctic Bottom Water be-
 2138 tween the 1980s and 2000s. *Journal of Climate*, 25, 5830–5844. doi: 10.1175/
 2139 JCLI-D-11-00612.1
- 2140 Purkey, S., & Johnson, G. (2013). Antarctic Bottom Water warming and freshening:
 2141 Contributions to sea level rise, ocean freshwater budgets, and global heat gain.
 2142 *Journal of Climate*, 26, 6105–6122. doi: 10.1175/JCLI-D-12-00834.1
- 2143 Redi, M. H. (1982). Oceanic isopycnal mixing by coordinate rotation. *Journal of
 2144 Physical Oceanography*, 12, 1154–1158. doi: 10.1175/1520-0485(1982)012<1154:
 2145 OIMBCR>2.0.CO;2
- 2146 Reichl, B. G., Adcroft, A., Griffies, S. M., & Hallberg, R. W. (2022). A potential en-
 2147 ergy analysis of ocean surface mixed layers. *Journal of Geophysical Research-
 2148 Oceans*, 127. doi: 10.1029/2021JC018140
- 2149 Reichl, B. G., & Hallberg, R. W. (2018). A simplified energetics based planetary
 2150 boundary layer (ePBL) approach for ocean climate simulations. *Ocean Mod-
 2151 ellling*, 132, 112–129. doi: 10.1016/j.ocemod.2018.10.004
- 2152 Reichl, B. G., & Li, Q. (2019). A parameterization with a constrained potential en-
 2153 ergy conversion rate of vertical mixing due to Langmuir turbulence. *Journal of
 2154 Physical Oceanography*, 49, 2935–2959. doi: 10.1175/JPO-D-18-0258.1
- 2155 Reichl, B. G., Wittenberg, A. T., Griffies, S. M., & Adcroft, A. (2024). Improving
 2156 equatorial upper ocean vertical mixing in the noaa/gfdl om4 model. *Earth
 2157 and Space Science*, 11(10), e2023EA003485. doi: <https://doi.org/10.1029/2023EA003485>
- 2159 Rhein, M., Rintoul, S., Aoki, S., Campos, E., Chambers, D., Feely, R., ... Wang, F.
 2160 (2013). Observations: Ocean. In *Climate change 2013: The physical science
 2161 basis. contribution of working group i to the fifth assessment report of the inter-
 2162 governmental panel on climate change* (pp. 255–315). Cambridge, United

- Kingdom: Cambridge University Press.
- Richter, I. (2015). Climate model biases in the eastern tropical oceans: Causes, impacts and ways forward. *Wiley Interdisciplinary Reviews: Climate Change*, 6(3), 345–358.
- Rugenstein, M., Bloch-Johnson, J., Gregory, J., Andrews, T., Mauritsen, T., Li, C., ... others (2020). Equilibrium climate sensitivity estimated by equilibrating climate models. *Geophysical Research Letters*, 47(4), e2019GL083898.
- Saba, V. S., Griffies, S., Anderson, W., Winton, M., Alexander, M., Delworth, T., ... Zhang, R. (2016). Enhanced warming of the northwest Atlantic Ocean under climate change. *Journal of Geophysical Research–Oceans*, 121, 118–132. doi: 10.1002/2015JC011346
- Sadourny, R. (1975, April). The Dynamics of Finite-Difference Models of the Shallow-Water Equations. *Journal of the Atmospheric Sciences*, 32(4), 680–689. Retrieved 2019-07-28, from [https://journals.ametsoc.org/doi/abs/10.1175/1520-0469\(1975\)032\(0680:TDOFDM\)2.0.CO;2](https://journals.ametsoc.org/doi/abs/10.1175/1520-0469%281975%29032%3C0680%3ATDOFDM%3E2.0.CO%3B2) doi: 10.1175/1520-0469(1975)032(0680:TDOFDM)2.0.CO;2
- Sallée, J.-B., Shuckburgh, E., Bruneau, N., Meijers, A. J. S., Bracegirdle, T. J., Wang, Z., & Roy, T. (2013). Assessment of Southern Ocean water mass circulation and characteristics in CMIP5 models: historical bias and forcing response. *Journal of Geophysical Research: Oceans*, 118. doi: 10.1002/jgrc.20135
- Salmon, R. (1980). Baroclinic instability and geostrophic turbulence. *Geophysical & Astrophysical Fluid Dynamics*, 15(1), 167–211. doi: 10.1080/03091928008241178
- Sane, A., Reichl, B. G., Adcroft, A., & Zanna, L. (2023). Parameterizing vertical mixing coefficients in the ocean surface boundary layer using neural networks. *Journal of Advances in Modeling Earth Systems*, 15(10), e2023MS003890. Retrieved from [\(e2023MS003890 2023MS003890\)](https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2023MS003890) doi: <https://doi.org/10.1029/2023MS003890>
- Sausen, R., Barthel, K., & Hasselmann, K. (1988). Coupled ocean-atmosphere models with flux correction. *Climate Dynamics*, 2, 145–163. doi: 10.1007/BF01053472
- Scott, R. B., & Wang, F. (2005, September). Direct evidence of an oceanic inverse kinetic energy cascade from satellite altimetry. *Journal of Physical Oceanography*, 35(9), 1650–1666. (Publisher: American Meteorological Society Section: Journal of Physical Oceanography) doi: 10.1175/JPO2771.1
- Séférian, R., Gehlen, M., Bopp, L., Resplandy, L., Orr, J. C., Marti, O., ... Romanou, A. (2016). Inconsistent strategies to spin up models in cmip5: implications for ocean biogeochemical model performance assessment. *Geoscientific Model Development*, 9, 1827–1851. doi: 10.5194/gmd-9-1827-2016
- Simmons, H. L., Jayne, S. R., St.Laurent, L. C., & Weaver, A. J. (2004). Tidally driven mixing in a numerical model of the ocean general circulation. *Ocean Modelling*, 6, 245–263. doi: 10.1016/S1463-5003(03)00011-8
- Smith, K. S., & Vallis, G. K. (2001). The scales and equilibration of midocean eddies: freely evolving flow. *Journal of Physical Oceanography*, 31, 554–570. doi: 10.1175/1520-0485(2001)031<0554:TSAEOM>2.0.CO;2
- Solomon, H. (1971). On the representation of isentropic mixing in ocean models. *Journal of Physical Oceanography*, 1, 233–234. doi: 10.1175/1520-0485(1971)001<0233:OTROIM>2.0.CO;2
- St. Laurent, L. C., Simmons, H. L., & Jayne, S. R. (2002). Estimating tidally driven energy in the deep ocean. *Geophysical Research Letters*, 29, 2106–2110. doi: 10.1029/2002GL015
- Stewart, A. L., Chi, X., Solodoch, A., & Hogg, A. M. (2021). High-frequency fluctuations in Antarctic Bottom Water transport driven by Southern Ocean winds.

- 2218 *Geophysical Research Letters*, 49. doi: 10.1029/2021GL094569
- 2219 Stewart, K., Hogg, A., Griffies, S. M., Heerdegen, A., Ward, M., Spence, P., & Eng-
2220 land, M. (2017). Vertical resolution of baroclinic modes in global ocean
2221 models. *Ocean Modelling*, 113, 50–65. doi: 10.1016/j.ocemod.2017.03.012
- 2222 Stouffer, R. J. (2004). Time scales of climate response. *Journal of Climate*, 31, 209–
2223 217. doi: 10.1175/1520-0442(2004)017<0209:TSOCR>2.0.CO;2
- 2224 Stouffer, R. J., Russel, J., & Spelman, M. J. (2006). Importance of oceanic heat up-
2225 take in transient climate change. *Geophysical Research Letters*, 33. doi: 10
2226 .1029/2006GL027242
- 2227 Tesdal, J.-E., MacGilchrist, G. A., Beadling, R. L., Griffies, S. M., Krasting, J. P., &
2228 Durack, P. J. (2023). Revisiting interior water mass responses to surface forc-
2229 ing changes and the subsequent effects on overturning in the Southern Ocean.
2230 *Journal of Geophysical Research - Oceans*, 128. doi: 10.1029/2022JC019105
- 2231 Thomas, M. D., & Federov, A. V. (2017). The eastern subtropical Pacific origin of
2232 the equatorial cold bias in climate models: A Lagrangian perspective. *Journal*
2233 *of Climate*, 30, 5885–5900. doi: 10.1175/JCLI-D-16-0819.1
- 2234 Trenberth, K. E., & Caron, J. M. (2001). Estimates of meridional atmosphere and
2235 ocean heat transports. *Journal of Climate*, 14, 3433–3443. doi: 10.1175/1520
2236 -0442(2001)014<3433:EOMAAO>2.0.CO;2
- 2237 Tsujino, H., Urakawa, S., Nakano, H., Small, R. J., Kim, W. M., Yeager, S. G.,
2238 ... Yamazaki, D. (2018). JRA-55 based surface dataset for driving
2239 ocean-sea-ice models (JRA-55-do). *Ocean Modelling*, 130, 79–139. doi:
2240 10.1016/j.ocemod.2018.07.002
- 2241 Vallis, G. K. (2017). *Atmospheric and oceanic fluid dynamics: Fundamentals and*
2242 *large-scale circulation* (2nd ed.). Cambridge: Cambridge University Press. (946
2243 + xxv pp)
- 2244 von Storch, J.-S., Haak, H., Hertwig, E., & Fast, I. (2016). Vertical heat and salt
2245 fluxes due to resolved and parameterized meso-scale eddies. *Ocean Modelling*,
2246 108, 1–19. doi: 10.1016/j.ocemod.2016.10.001
- 2247 von Schuckmann, K., Minière, A., Gues, F., Cuesta-Valero, F. J., Kirchengast,
2248 G., Adusumilli, S., ... Zemp, M. (2023). Heat stored in the earth system
1960–2020: where does the energy go? *Earth System Science Data*, 15(4),
2249 1675–1709. Retrieved from <https://essd.copernicus.org/articles/15/1675/2023/> doi: 10.5194/essd-15-1675-2023
- 2250 Wang, C., Zhang, L., Lee, S.-K., Wu, L., & Mechoso, C. R. (2014). A global per-
2251 spective on CMIP5 climate model biases. *Nature Climate Change*, 4. doi: 10
2252 .1038/NCLIMATE2118
- 2253 Weaver, A. J., & Hughes, T. M. C. (1996). On the incompatibility of ocean and
2254 atmosphere models and the need for flux adjustments. *Climate Dynamics*, 12,
2255 141–170. doi: 10.1007/BF00211615
- 2256 Welch, P. (1967). The use of fast Fourier transform for the estimation of power
2257 spectra: A method based on time averaging over short, modified peri-
2258 odograms. *IEEE Transactions on Audio and Electroacoustics*, 15, 70–73.
2259 doi: 10.1109/TAU.1967.1161901
- 2260 Wills, R. C. J., Dong, Y., Proistosecu, C., Armour, K. C., & Battisti, D. S. (2022).
2261 Systematic climate model biases in the large-scale patterns of recent sea-
2262 surface temperature and sea-level pressure change. *Geophysical Research*
2263 *Letters*, 49. Retrieved from <https://doi.org/10.1029/2022GL100011> doi:
2264 10.1029/2022GL100011
- 2265 Wise, A., Hughes, C. W., & Polton, J. A. (2018). Bathymetric influence on the
2266 coastal sea level response to ocean gyres at western boundaries. *Journal of*
2267 *Physical Oceanography*, 481, 2949–2964. doi: 10.1175/JPO-D-18-0007.1
- 2268 Wise, A., Hughes, C. W., Polton, J. A., & Huthnance, J. M. (2020). Leaky slope
2269 waves and sea level: Unusual consequences of the beta effect along western
2270 boundaries with bottom topography and dissipation. *Journal of Physical*
2271 *Oceanography*, 33, 103–116. doi: 10.1029/2019JC015470
- 2272

- 2273 *Oceanography*, 50, 217–237. doi: 10.1175/JPO-D-19-0084.1
- 2274 Wittenberg, A. T. (2009). Are historical records sufficient to constrain ENSO simu-
2275 lations? *Geophysical Research Letters*, 36. doi: 10.1029/2009GL038710
- 2276 Wolfe, C. L., Cessi, P., McClean, J. L., & Maltrud, M. E. (2008). Vertical heat
2277 transport in eddying ocean models. *Geophysical Research Letters*, 35(23), n/a–
2278 n/a. Retrieved from <http://dx.doi.org/10.1029/2008GL036138> doi: 10
2279 .1029/2008GL036138
- 2280 Wright, D. G. (1997). An equation of state for use in ocean models: Eckart's for-
2281 mula revisited. *Journal of Atmospheric and Oceanic Technology*, 14, 735–740.
2282 doi: 10.1175/1520-0426(1997)014<735:AEOSFU>2.0.CO;2
- 2283 Yankovsky, E., Zanna, L., & Smith, K. S. (2022). Influences of mesoscale ocean ed-
2284 dies on flow vertical structure in a resolution-based model hierarchy. *Journal of*
2285 *Advances in Modeling Earth Systems*, n/a(n/a). doi: 10.1029/2022MS003203
- 2286 Yin, J., Griffies, S. M., Winton, M., Zhao, M., & Zanna, L. (2020). Response
2287 of storm-related extreme sea level along the us Atlantic Coast to combined
2288 weather and climate forcing. *Journal of Climate*, 33, 3745–3769. doi:
2289 10.1175/JCLI-D-19-0551.1
- 2290 Zanna, L., Khatiwala, S., Gregory, J., Ison, J., & Heimbach, P. (2019). Global recon-
2291 struction of historical ocean heat storage and transport. *Proceedings of the Na-
2292 tional Academy of Science*. doi: 10.1073/pnas.1808838115
- 2293 Zhang, L., Delworth, T. L., Cooke, W., Goosse, H., Bushuk, M., Morioka, Y., &
2294 Yang, X. (2021). The dependence of internal multidecadal variability in the
2295 Southern Ocean on the ocean background mean state. *Journal of Climate*,
2296 34(3), 1061–1080. doi: 0.1175/JCLI-D-20-0049.1
- 2297 Zhang, R., Delworth, T. L., Rosati, A., Anderson, W. G., Dixon, K. W., Lee, H.-
2298 C., & Zeng, F. (2011). Sensitivity of the North Atlantic Ocean circula-
2299 tion to an abrupt change in the Nordic Sea overflow in a high resolution
2300 global coupled climate model. *Journal of Geophysical Research*, 116. doi:
2301 10.1029/2011JC007240
- 2302 Zhang, R., Sutton, R., Danabasoglu, G., Marsh, R., Yeager, S. G., Amrhein, D. E.,
2303 & Little, C. M. (2019). A review of the role of the Atlantic meridional over-
2304 turning circulation in Atlantic multidecadal variability and associated climate
2305 impacts. *Reviews of Geophysics*, 57. doi: 10.1029/2019RG000644
- 2306 Zhao, M. (2020). Simulations of atmospheric rivers, their variability, and response to
2307 global warming using GFDL's new high-resolution general circulation model.
2308 *Journal of Climate*, 33, 10287–10303. doi: 10.1175/JCLI-D-20-0241.1
- 2309 Zhao, M. (2022a). An investigation of the effective climate sensitivity in GFDL's
2310 new climate models CM4.0 and SPEAR. *Journal of Climate*, 35, 5637–5660.
2311 doi: 10.1175/JCLI-D-21-0327.s1
- 2312 Zhao, M. (2022b). A study of AR-, TS-, and MCS-Associated precipitation and
2313 extreme precipitation in present and warmer climates. *Journal of Climate*, 35,
2314 479–497, DOI: 10.1175/JCLI-D-21-0145.1.
- 2315 Zhao, M., Golaz, J.-C., Held, I. M., Guo, H., Balaji, V., Benson, R., ... Xiang, B.
2316 (2018a). The GFDL Global Atmosphere and Land Model AM4.0/LM4.0 - Part
2317 II: Model Description, Sensitivity Studies, and Tuning Strategies. *Journal of*
2318 *Advances in Modeling Earth Systems*, 10. doi: 10.1002/2017MS001209
- 2319 Zhao, M., Golaz, J.-C., Held, I. M., Guo, H., Balaji, V., Benson, R., ... Xiang, B.
2320 (2018b). The GFDL Global Atmosphere and Land Model AM4.0/LM4.0 - Part
2321 I: Simulation Characteristics with Prescribed SSTs. *Journal of Advances in*
2322 *Modeling Earth Systems*, 10. doi: 10.1002/2017MS001208
- 2323 Zika, J. D., Le Sommer, J., Dufour, C. O., Molines, J.-M., Barnier, B., Brasseur, P.,
2324 ... Vivier, F. (2014). Vertical eddy fluxes in the Southern Ocean. *Journal of*
2325 *Physical Oceanography*, 43, 941–955. doi: 10.1175/JPO-D-12-0178.1
- 2326 Zweng, M. M., Reagan, J. R., Antonov, J. I., Locarnini, R. A., Mishonov, A. V.,
2327 Boyer, T. P., ... Biddle, M. M. (2018). *World Ocean Atlas 2013, Volume*

2328 *2: Salinity* (dataset). U.S. Department of Commerce, Washington, D.C.:
2329 NOAA/NESDIS.

The GFDL-CM4X climate model hierarchy, Part I: model description and thermal properties

Stephen M. Griffies^{1,2}, Alistair Adcroft², Rebecca L. Beadling³, Mitchell Bushuk¹, Chiung-Yin Chang², Henri F. Drake⁴, Raphael Dussin¹, Robert W. Hallberg^{1,2}, William J. Hurlin¹, Hemant Khatri⁵, John P. Krasting¹, Matthew Lobo², Graeme A. MacGilchrist⁶, Brandon G. Reichl¹, Aakash Sane², Olga Sergienko², Maike Sonnewald⁷, Jacob M. Steinberg¹, Jan-Erik Tesdal², Matthew Thomas⁸, Katherine E. Turner², Marshall L. Ward¹, Michael Winton¹, Niki Zadeh¹, Laure Zanna⁹, Rong Zhang^{1,2}, Wenda Zhang², Ming Zhao¹

¹NOAA Geophysical Fluid Dynamics Laboratory, Princeton, USA

²Princeton University Atmospheric and Oceanic Sciences Program, Princeton, USA

³Department of Earth and Environmental Science, Temple University, Philadelphia, USA

⁴Department of Earth System Science, University of California, Irvine, USA

⁵Department of Earth, Ocean and Ecological Sciences, University of Liverpool, UK

⁶School of Earth and Environmental Sciences, University of St. Andrews, UK

⁷Computer Science Department, University of California, Davis, USA

⁸Centre for Environment Fisheries and Aquaculture Science, Suffolk, UK

⁹Courant Institute of Mathematical Sciences, New York University, New York, USA

November 23, 2024

Key Points:

- CM4X-p25 and CM4X-p125 are designed to study ocean and sea ice physics, focusing on effects from eddies and boundary currents.
- CM4X-p125 reaches pre-industrial thermal equilibrium in 150 years whereas the coarser CM4X-p25 has yet to equilibrate after 1000 years.
- CM4X-p125's active eddies and negligible spurious mixing render an equilibrated pre-industrial ocean with 400ZJ less heat than present day.

28 **Abstract**

29 We present the GFDL-CM4X (Geophysical Fluid Dynamics Laboratory Climate
 30 Model version 4X) coupled climate model hierarchy. The primary application for CM4X
 31 is to investigate ocean and sea ice physics as part of a realistic coupled Earth climate
 32 model. CM4X utilizes an updated MOM6 (Modular Ocean Model version 6) ocean physics
 33 package relative to CM4.0, and there are two members of the hierarchy: one that uses
 34 a horizontal grid spacing of 0.25° (referred to as CM4X-p25) and the other that uses a
 35 0.125° grid (CM4X-p125). CM4X also refines its atmospheric grid from the nominally
 36 100 km (cubed sphere C96) of CM4.0 to 50 km (C192). Finally, CM4X simplifies the land
 37 model to allow for a more focused study of the role of ocean changes to global mean cli-
 38 mate.

39 CM4X-p125 reaches a global ocean area mean heat flux imbalance of -0.02 W m^{-2}
 40 within $\mathcal{O}(150)$ years in a pre-industrial simulation, and retains that thermally equilibrated
 41 state over the subsequent centuries. This 1850 thermal equilibrium is characterized by
 42 roughly 400 ZJ less ocean heat than present-day, which corresponds to estimates for an-
 43 thropogenic ocean heat uptake between 1850 and present-day. CM4X-p25 approaches
 44 its thermal equilibrium only after more than 1000 years, at which time its ocean has roughly
 45 1100 ZJ *more* heat than its early 21st century ocean initial state. Furthermore, the root-
 46 mean-square sea surface temperature bias for historical simulations is roughly 20% smaller
 47 in CM4X-p125 relative to CM4X-p25 (and CM4.0). We offer the *mesoscale dominance*
 48 *hypothesis* for why CM4X-p125 shows such favorable thermal equilibration properties.

49 **Plain Language Summary**

50 We detail a new climate model hierarchy, CM4X. CM4X has two model configu-
 51 rations, CM4X-p25 and CM4X-p125, that differ only in the ocean/sea ice horizontal grid
 52 spacing. CM4X-p125 outperforms CM4X-p25 for certain climate relevant processes, while
 53 maintaining skill levels seen in previous generations for other results. CM4X-p125 re-
 54 quires about 10 times less time than CM4X-p25 to reach pre-industrial control thermal
 55 equilibration. Also, CM4X-p125 equilibrates to an ocean state with roughly 400 ZJ less
 56 heat content than present-day, consistent with estimates of anthropogenic heat uptake
 57 since 1870, whereas CM4X-p25 equilibrates to a state with roughly 1100 ZJ *more* heat
 58 than present-day. Consequently, the CM4X-p125 ocean state has not drifted far from
 59 observational estimates, so that the ocean water mass structure is more accurate than
 60 models requiring $\mathcal{O}(1000)$ years to equilibrate. We propose the *mesoscale dominance hy-*
61 pothesis to interpret the relatively rapid thermal equilibration of CM4X-p125 to a cooler
 62 pre-industrial state. Such ocean models result from negligible spurious mixing (from nu-
 63 merical truncation errors) along with an active mesoscale transport and realistic par-
 64 ameterization of small scale (diapycnal) mixing. Because of their more rapid thermal equi-
 65 libration, mesoscale dominant ocean models reduce the computational energy footprint
 66 of models that are not mesoscale dominant.

67 **1 Introducing this paper and its thesis**

68 The ocean is the predominant reservoir in the Earth's climate system for enthalpy
 69 (heat), absorbing roughly 90% of the added heat due to anthropogenic increases in green-
 70 house gasses in the atmosphere (Fox-Kemper et al., 2021; von Schuckmann et al., 2023).
 71 Ocean circulation and its attendant physical processes redistribute both pre-industrial
 72 and anthropogenic heat laterally and vertically throughout the World Ocean, with as-
 73 sociated impacts on regional and global climate variations and change.

74 We developed a climate model hierarchy, CM4X, to support research into the ocean's
 75 role in climate. CM4X is comprised of two coupled climate models: CM4X-p25 and CM4X-

76 p125. These two models are identically configured in all aspects, except for their ocean
 77 (and sea ice) horizontal grid spacing and bottom topography. We document the devel-
 78 opment and performance of CM4X in this paper along with its companion in Part II (Griffies
 79 et al., 2024).

80 1.1 Content of this paper

81 We overview the CM4X project in Sections 2 and 3, touching on some of the high-
 82 lights as well as the unresolved problems. We focus in Sections 4 and 5 on the key in-
 83 novation emerging from the CM4X model development. Namely, the CM4X-p125 model
 84 exhibits a relatively rapid $\mathcal{O}(100)$ years thermal equilibration in its pre-industrial control
 85 (piControl) simulation. Furthermore, the global volume integrated ocean heat con-
 86 tent in CM4X-p125 is roughly 400 ZJ less than the present-day initial conditions, which
 87 corresponds to the estimates by Zanna et al. (2019) for the anthropogenic ocean heat
 88 uptake since 1870. In contrast, many other climate models (e.g., CM4.0 and CM4X-p25)
 89 approach a piControl equilibrium only after $\mathcal{O}(1000)$ years and with an ocean heat con-
 90 tent that is higher than present-day. There are relatively few field measurements of the
 91 1850 ocean. Even so, we are virtually certain that the 1850 ocean had less heat content
 92 than the early 21st century ocean. Although the equilibrated heat content is lower in
 93 CM4X-p125, its sea surface temperature is higher than CM4X-p25, which then results
 94 in a reduced bias in the CM4X-p125 historical SST relative to CM4X-p25.

95 We conjecture in Section 6 that the ocean component of CM4X-p125 approximates
 96 a novel class of ocean climate models that are *mesoscale dominant*, with mesoscale dom-
 97 inance described in the next subsection. We conclude the main portion of this paper in
 98 Section 7 by discussing the implications for climate modeling suggested by the CM4X
 99 results. Appendix A then details the CM4X model configuration, focusing on aspects
 100 that are distinct from the OM4.0 ocean/sea ice configuration of Adcroft et al. (2019) and
 101 the CM4.0 climate model of Held et al. (2019). In Part II of this paper (Griffies et al.,
 102 2024) we detail results from a suite of case studies targeting areas of planned and on-
 103 going research with the CM4X hierarchy.

104 1.2 Mesoscale dominance as the paper's central thesis

105 Mesoscale dominant ocean models are characterized by an explicit representation
 106 of vigorous ocean mesoscale features (or with a suitable mesoscale eddy parameteriza-
 107 tion), an accurate parameterization of small scale (diapycnal) mixing, and a low level
 108 of spurious numerical ocean mixing. We hypothesize that these three properties, in com-
 109 bination, form necessary (though not sufficient) conditions for a relatively rapid (cen-
 110 tennial rather than millennial) thermal equilibration of the global climate system under
 111 piControl forcing, and with the piControl equilibrium having less, not more, heat con-
 112 tent than the early 21st century ocean. Stated in a negative manner, we hypothesize that
 113 pi-Control thermal equilibration will require $\mathcal{O}(1000)$ years so long as the ocean model
 114 component offers a poor approximation to mesoscale dominance.

115 We are partly led to propose the mesoscale dominance hypothesis by acknowledg-
 116 ing the history of ocean model developments that have focused on the importance of mesoscale
 117 transport, along with maintaining physically relevant diapycnal mixing and negligible
 118 spurious mixing. Within the context of mesoscale eddy parameterization, Solomon (1971),
 119 Redi (1982), Gent and McWilliams (1990) and Gent et al. (1995) offer a starting point
 120 for this line of research, whereby their formulations stress the need to parameterize the
 121 mesoscale eddy transport while respecting water mass boundaries. The CM4X-p125 re-
 122 sults exemplify this approach through an explicit realization of the mesoscale rather than
 123 via its parameterization, and by employing an ocean model component that pays par-
 124 ticular attention to the numerical needs of reducing spurious mixing between water masses.
 125 Even so, we do not claim that CM4X-p125 perfectly realizes the three characteristics of

126 a mesoscale dominant ocean. Indeed, as documented in Part II of this study (Griffies
 127 et al., 2024), CM4X-p125 has many shortcomings. Rather, we contend that CM4X-p125
 128 approaches these characteristics sufficiently to render a rapid thermal equilibration into
 129 a cooler pre-industrial ocean state, thus offering a taste for what is possible from mesoscale
 130 dominant ocean climate models.

131 We acknowledge the difficulty of quantitatively testing whether a model can be char-
 132 acterized as mesoscale dominant. Indeed, it might be most practical to infer relative scales
 133 of mesoscale dominance based on comparisons across a model hierarchy, such as we do
 134 for CM4X. Developing direct ways of characterizing mesoscale dominance remains a topic
 135 of ongoing research, with some suggestions offered in Section 6 based on water mass anal-
 136 ysis. Even though difficult to test, we propose that the mesoscale dominance hypoth-
 137 esis provides a useful means to frame the question of climate model thermal equilibra-
 138 tion. It also offers targets for model development strategies. In particular, we empha-
 139 size the central role of reducing spurious numerical mixing in developing models capa-
 140 ble of realizing mesoscale dominance. Namely, a strongly energetic ocean mesoscale must
 141 be realized with negligible spurious mixing if the model is to meet the needs of, and the ben-
 142 efits from, mesoscale dominance.

143 Our focus throughout this model description paper concerns the ocean and sea ice.
 144 However, we fully acknowledge that atmospheric and land radiative properties have a
 145 substantial impact on the simulation of sea surface temperature, ocean volume mean tem-
 146 perature, and ocean heat uptake. Radiative properties, in turn, are significantly influ-
 147 enced by cloud parameterizations containing poorly constrained parameters. We doc-
 148 ment the tuning of CM4X cloud parameters in Section A1 of the appendix. Importantly,
 149 by setting all atmospheric model configuration details to be identical for CM4X-p25 and
 150 CM4X-p125, we enable a focus on the ocean’s role in affecting thermal equilibration of
 151 the climate model, as well as other aspects of the simulation.

ACRONYM	MEANING	CITATION OR SECTION
AMIP	Atmospheric Model Intercomparison Project	pcmdi.llnl.gov/mips/amip
AM4	GFDL Atmospheric Model 4.0	Zhao et al. (2018b, 2018a)
CM2-O	GFDL climate model hierarchy 2.0	Delworth et al. (2006), Griffies et al. (2015)
C96	AM4 with cubed-sphere (≈ 100 km)	Zhao et al. (2018b, 2018a)
C192	AM4 with cubed-sphere (≈ 50 km) in CM4X	Zhao (2020)
CMIP6	Coupled Model Intercomparison Project 6	Eyring et al. (2016)
CM4.0	GFDL Climate Model 4.0 (0.25° ocn & C96 atm)	Held et al. (2019)
CM4X	GFDL Climate Model hierarchy	this paper
CM4X-p25	CM4X w/ 0.25° ocn and C192 atm	this paper
CM4X-p125	CM4X with 0.125° ocn and C192 atm	this paper
ESM4.1	GFDL Earth System Model 4.1	Dunne et al. (2020)
GFDL	Geophysical Fluid Dynamics Laboratory	–
MOM6	Modular Ocean Model version 6	Adcroft et al. (2019), Griffies et al. (2020)
OM4.0	GFDL Ocean/Sea-ice Model 4.0 (0.25°)	Adcroft et al. (2019)
OHUE	ocean heat uptake efficiency	Section 5.7
SST	sea surface temperature	Section 4.2
TOA	top of atmosphere	Sections 5.6 and A1

Table 1. Acronyms used in this paper, their meaning, and relevant citation or section.

Acronyms in the upper part of the table refer to model names and configurations, whereas those in the lower portion refer to physical climate properties.

152 2 The CM4X model hierarchy

153 We here introduce the CM4X project and provide reasoning for particular development
 154 strategies and corresponding decisions.

155 2.1 The importance of humility in model development

156 In developing CM4X, we started from the CM4.0 climate model of Held et al. (2019),
 157 along with its OM4.0 ocean/sea ice component from Adcroft et al. (2019). Our initial
 158 development goal was to upgrade the ocean component of CM4.0 to use a finer horizontal
 159 grid for the ocean/sea ice components. In confronting that goal, we chose to revisit
 160 certain earlier choices that then led to further development. We also chose to refine the
 161 atmosphere grid relative to CM4.0 to enhance the representation of atmospheric storms,
 162 which are important not only for studies of high-impact weather and their associated pre-
 163 cipitation (Zhao, 2020, 2022b; Hsieh et al., 2020) but also for simulating ocean extremes
 164 (Yin et al., 2020). A version of CM4.0 with increased atmospheric resolution (C192AM4)
 165 was previously configured for GFDL’s participation in CMIP6 HighResMIP (Haarsma
 166 et al., 2016). Simulations for 1950 control, the historical period, and future projections
 167 up to 2050 were conducted following the CMIP6 HighResMIP specifications, with the
 168 output submitted to CMIP6. However, this coupled model has not been thoroughly ex-
 169 plored or documented.

170 Throughout the development process, we found it crucial to maintain a balance be-
 171 tween what is practical and what is aspirational. Along the way, we experienced the wis-
 172 dom of Hofstadter’s Law (Hofstadter, 1979), which, as paraphrased for the CM4X project,
 173 reads: *Climate model development always takes longer than you expect, even when you*
account for Hofstadter’s Law. Indeed, our original plans in early 2020 for a six-month
 174 project expanded into a multi-year development effort. Part of the extended time frame
 175 arose from expanded visions for the project (i.e., “mission creep”), the entrainment of
 176 more individuals with unique ideas, interests, and scientific needs for an ocean eddying
 177 coupled model, and more simply, from our experience with an application of Murphy’s
 178 Law: *Climate models are Murphy’s Law Machines: anything that can go wrong will go*
wrong.

181 2.2 Balancing the aspirational with the feasible

182 CM4X is designed for research concerning physical ocean and sea ice processes that
 183 manifest in regional and global weather and climate, extending in a seamless fashion from
 184 days (e.g., sea level and temperature extremes) to centuries (e.g., ocean overturning cir-
 185 culation variations and changes). A faithful numerical depiction of such phenomena re-
 186 quires coupled models with relatively fine grid spacing in both the horizontal and ver-
 187 tical. There is growing experience with the payoffs realized by refined ocean grid spac-
 188 ing, in particular for the representation of mesoscale processes and their role in long term
 189 climate such as ocean heat uptake (Griffies et al., 2015). Additionally, refined ocean grid
 190 spacing and subsequent improved topographic representation are essential for the study
 191 of ocean extremes near the coasts. For example, when coupled to an ocean model that
 192 accurately represents the geometry of ocean continental shelves, climate models are key
 193 to the study of coastal ocean heat waves (Bian et al., 2023). The study from Yin et al.
 194 (2020) pointed to the additional benefit of refined atmospheric grid spacing for captur-
 195 ing atmospheric cyclones largely responsible for extremes in coastal sea level. Indeed,
 196 Yin et al. (2020) and Zhao (2020, 2022b) demonstrate the enhanced realism of simulated
 197 tropical and mid-latitude cyclones when moving from a nominally 100 km (C96) version
 198 of the AM4 atmosphere (Zhao et al., 2018a, 2018b) used in CM4.0, to the nominally 50 km
 199 (C192) version of AM4 used by Zhao (2020). These studies, along with those by Hsieh
 200 et al. (2020), motivate us to use the C192 version of AM4 for CM4X. Improved simu-
 201 lations of atmospheric storms would also enable better exploration of the rectified im-

202 pact of small-scale storms on climate, although we do not claim that the model, at its
 203 current resolution, is near convergence.

204 For the ocean component of CM4X, we considered the practical question: *how far*
 205 *can we go in refining the grid spacing while producing a suitable science tool within an*
 206 *acceptable development timeframe?* Centennial-scale climate simulations are necessary
 207 to evaluate the fidelity of inter-annual to inter-decadal modes of coupled climate vari-
 208 ability, yet experience indicates that intellectual patience can greatly dissipate if these
 209 100-year simulations take longer than roughly two months of wall clock time. With this
 210 turnaround in mind, we determined that computer resources available for this project
 211 dictate that 0.125° ocean grid spacing is the finest horizontal grid that we can reason-
 212 ably afford.

213 Furthermore, we remained open to considering advancements across the entire suite
 214 of OM4.0 physical parameterizations documented in Adcroft et al. (2019). We deliber-
 215 ately and systematically advanced the ocean parameterizations (see details in Appendix
 216 A), with tests of changes requiring more than 3×10^3 simulation years in prototypes of
 217 CM4X-p25 and more than 1.7×10^3 simulation years in prototypes of CM4X-p125. This
 218 development process greatly benefited from the roughly 4.5×10^4 coupled model sim-
 219 ulation years used to develop CM4.0 (Adcroft et al., 2019).

220 2.3 Climate model hierarchy based on ocean horizontal grid scale

221 Simulation realism, including a comprehensive suite of physical, chemical, and bi-
 222 ological processes, often runs counter to the needs for unpacking the underlying mech-
 223 anisms rendering scientific insights and robust understanding. In support of understand-
 224 ing, one is motivated to develop models that focus on specific processes and to compro-
 225 mise on comprehensiveness. As emphasized by Held (2005), Jeevanjee et al. (2017), and
 226 Maher et al. (2019), model hierarchies provide a means to span the spectrum between
 227 comprehensive and ideal. Such hierarchies support the science forming the basis for com-
 228 prehensive earth system models, and enhance the science understanding emerging from
 229 model simulations. A model hierarchy can be realized in a variety of manners, for ex-
 230 ample, see Figure 6 in Maher et al. (2019) for atmospheric model hierarchies. The two
 231 CM4X models differ only by their ocean and sea ice grid spacing, along with the rep-
 232 resentation of bottom topography. All other model details are identical. Hence, CM4X
 233 is a hierarchy of scale, with a focus on questions related to the ocean and sea ice.

234 2.4 CM4X design decisions

235 The CM4X hierarchy is inspired by the GFDL CM2-O hierarchy of Griffies et al.
 236 (2015), which consists of three ocean/sea ice models of varying grid spacing that are cou-
 237 pled to the same land model and atmosphere model. The coarsest CM2-O model has a
 238 one degree ocean that uses a mesoscale eddy parameterization. The other two models
 239 have 0.25° and 0.1° ocean grids and use no mesoscale parameterizations (other than bi-
 240 harmonic friction). For the CM4X project, we targeted the role of explicitly represented
 241 mesoscale eddies and boundary currents in the eddy-admitting regime. Inclusion of a one-
 242 degree ocean component to the CM4X hierarchy would necessarily require development
 243 of an eddy parameterization for this model, with such research outside of our scope.

244 Further examples of ocean grid spacing hierarchies include the coupled climate model
 245 of Kirtman et al. (2012), which includes a 1° ocean (with mesoscale eddy parameteri-
 246 zations) and 0.1° ocean (without mesoscale eddy parameterizations). The hierarchy of
 247 Kiss et al. (2020) shares much with CM2-O, yet with a focus on forced ocean/sea ice sim-
 248 ulations using the JRA55-do atmospheric state from Tsujino et al. (2018). The Never-
 249 world2 hierarchy of Marques et al. (2022) consists of idealized stacked shallow water mod-

250 els using MOM6 and is designed to study adiabatic aspects of mesoscale eddies and their
 251 parameterization.

252 *Absence of mesoscale eddy parameterizations in CM4X*

253 We disabled mesoscale eddy parameterizations for CM4X, which also follows the
 254 choice of Adcroft et al. (2019) for OM4.0, as well as in the hierarchies of Griffies et al.
 255 (2015), Kirtman et al. (2012), Kiss et al. (2020), and Marques et al. (2022), for those con-
 256 figurations with 0.25° and finer grid spacing. Disabling mesoscale eddy parameteriza-
 257 tion helps to clarify the physics associated with distinct representations of ocean fluid
 258 mechanics. Even so, this design choice in no way presumes that the ocean mesoscale is
 259 accurately or completely represented by either the 0.25° or 0.125° grids in CM4X (e.g.,
 260 see Figure 1). Indeed, many authors to this paper are actively pursuing research into pa-
 261 rameterizations suited for the eddy active “gray zone” regime of CM4X. This regime re-
 262 mains challenging for the ocean modeling community, similar in vein to the atmospheric
 263 model quest for storm-resolving resolutions, yet it represents the foreseeable future for
 264 GFDL coupled models. The CM4X hierarchy informs this parameterization research.

265 *Component model simplifications to align with CM4X science goals*

266 The Earth system complexity is deliberately limited in the CM4X hierarchy rel-
 267 ative to more comprehensive climate or earth system models. In particular, CM4X does
 268 not include an interactive biogeochemistry component. Furthermore, in CM4X (as for
 269 CM4.0), we use prescribed glaciers and icecaps rather than interactive ice sheets (we choose
 270 distinct albedos relative to the CM4.0 values, as discussed in Appendix A3). Interactive
 271 land ice capabilities are at the cutting-edge of climate modeling science today, and they
 272 are outside of our scope. We made further simplifications to the CM4X land model rel-
 273 ative to CM4.0, with details provided in Appendix A2. In brief, CM4X makes use of static
 274 rather than dynamic land vegetation along with a land use forcing fixed at 1980 values,
 275 along with no CO₂ fertilization. These land model changes compromise CM4X’s abil-
 276 ity to simulate climate in the presence of known forcing agents used by CMIP6 models
 277 such as CM4.0.

278 *CM4X is not a comprehensive CMIP6-class climate model*

279 In some of our analysis, we include results from CM4.0 to benchmark the CM4X
 280 results. In such comparisons, it is important to note that the various configuration dif-
 281 ferences between CM4X and CM4.0 preclude an unambiguous attribution of causes for
 282 simulation differences. The only clean comparison in this paper is between CM4X-p25
 283 and CM4X-p125. Even so, in Appendix A3 we show that from a global mean heat bud-
 284 get perspective, the key difference between CM4.0 and CM4X-p25 concerns the distinct
 285 snow on glacier albedos.

286 Certain of the CM4X design choices, particularly those associated with the land
 287 model, preclude CM4X from being considered a comprehensive CMIP6-class climate model,
 288 whereas CM4.0 directly targeted CMIP6 questions. This limitation is important to ac-
 289 knowledge when comparing CM4.0 and CM4X during the historical simulation since CM4.0
 290 has the CMIP6 land forcings whereas CM4X does not. Simplifying the climate model
 291 by disabling elements of the land model reduced the risk of having the perfect be the en-
 292 emy of the good in our task of developing a fit-for-purpose coupled climate model hier-
 293 archy focused on ocean and sea ice physics.

294 **3 A survey of CM4X results**

295 We here offer a selection of results that highlight facets of the model simulations
 296 using pre-industrial, historical, and future scenario experiments from CMIP6 (Eyring et
 297 al., 2016). Further details for many of these diagnostics are presented in the various case
 298 studies of Part II (Griffies et al., 2024).

299 **3.1 Experiments, initialization, and analysis**

300 We analyze CM4X as realized in the following CMIP6 experiments (Eyring et al.,
 301 2016).

- 302 • **piControl:** Pre-industrial control with radiative forcing fixed at year 1850. This
 303 experiment illustrates how the models drift from their initial conditions, taken from
 304 the early 21st century, and approach thermal equilibrium under pre-industrial forc-
 305 ing.
- 306 • **Historical:** 01January of year 101 from the piControl is used to initialize a histor-
 307 ical simulation that is run from 1850 to 2014. In this historical simulation, we did
 308 not account for temporal evolution in vegetation, land use, or CO₂ fertilization.
- 309 • **SSP5-8.5:** 01January of year 2015 provides the initial condition for the CMIP6 SSP5-
 310 8.5 scenario experiment, which allows us to study how the CM4X models simu-
 311 late climate change through to 2100.

312 The piControl ocean was initialized from the 2013 World Ocean Atlas, which consists
 313 of Locarnini et al. (2018) for temperature and Zweng et al. (2018) for salinity. This anal-
 314 ysis is dominated by late 20th and early 21st century measurements, though with no-
 315 tably poor sampling of the deep ocean. The ocean model makes use of the pre-TEOS10
 316 equation of state from Wright (1997). We interpret the model temperature and salin-
 317 ity according to McDougall et al. (2021), whereby the model's prognostic temperature
 318 field is interpreted as Conservative Temperature. There is no hidden or undocumented
 319 spin-up phase for the ocean temperature and salinity. Rather, all piControl time series
 320 presented in this paper start with ocean initial conditions taken from World Ocean At-
 321 las.

322 For CM4X-p125, we took initial conditions for the land model from CM4X-p25 as
 323 remapped onto the CM4X-p125 land/sea mask grid. Sea ice, iceberg, and calving ini-
 324 tial conditions for CM4X-p125 and CM4X-p25 were taken from prototype configura-
 325 tions that ran for 100 and 70 years, respectively, under 1950 radiative forcing conditions. This
 326 choice avoids “cold starting” the sea ice component, which can introduce large initial salin-
 327 ity and temperature pulses into the high-latitude oceans due to sea ice formation. In both
 328 CM4X-p25 and CM4X-p125, initial conditions for sea ice concentration and sea ice vol-
 329 ume reasonably agree with present-day observations. The initial Pan-Arctic sea ice ex-
 330 tent is $16.6 \times 10^6 \text{ km}^2$ in CM4X-p125 and $14.4 \times 10^6 \text{ km}^2$ in CM4X-p25, and the ini-
 331 tial Pan-Antarctic sea ice extent is $11.4 \times 10^6 \text{ km}^2$ (CM4X-p125) and $7.9 \times 10^6 \text{ km}^2$
 332 (CM4X-p25).

333 For horizontal maps, we present results on the model's native horizontal grid. For
 334 the vertical, we follow the CMIP6 recommendations from Section A4 of Griffies et al.
 335 (2016) by generating results on a 35-level diagnostic z^* grid used in versions of the World
 336 Ocean Atlas (Locarnini et al., 2018). We do so by remapping, on every model time step,
 337 to the z^* grid from the model native grid.

338 **3.2 Climate modeling with an active ocean mesoscale**

339 Hallberg (2013) as well as Griffies and Treguier (2013) noted that the first baro-
 340 clinic Rossby radius, L_d , is a useful measure for gauging the degree by which the ocean

horizontal grid spacing is able to resolve the mesoscale ocean dynamics (see K. Stewart et al. (2017) for complementary statements about vertical grid requirements). The reason is that the most unstable baroclinic waves have a length scale proportional to the deformation radius (e.g., Vallis (2017)). The Rossby radius also determines the scale of coastal boundary waves. Following the methods from Hallberg (2013), we compute L_d as realized in CM4X-p125 (Figure 1). Also in Figure 1, we map the ratio, L_d/Δ , where Δ is the horizontal grid spacing given by $\Delta^2 = (\Delta x)^2 + (\Delta y)^2$. This figure reveals that the 0.125° grid is insufficient to fully resolve the first baroclinic Rossby radius in the middle to high latitudes as well as around the continental shelves, with the 0.25° grid (not shown) satisfying $L_d/\Delta > 2$ in a smaller region encompassing just the tropics.

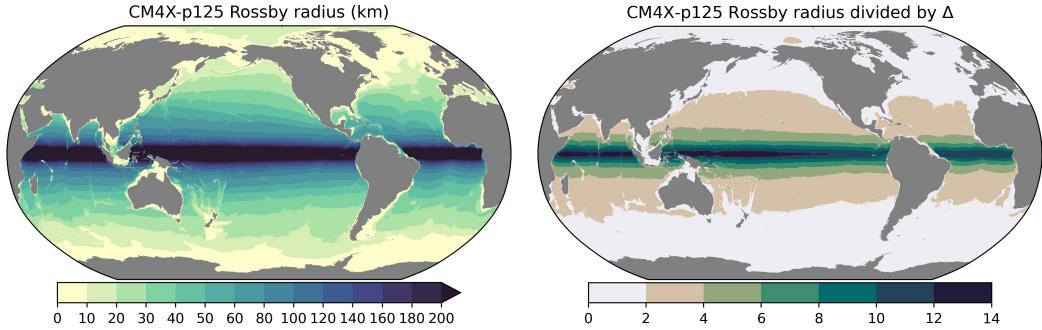


Figure 1. Left panel: first baroclinic Rossby radius, L_d , computed according to Hallberg (2013). The deformation radius is computed as a time average over years 1980-2009 in the historical CM4X-p125 simulation. Right panel: ratio of the first baroclinic Rossby radius to the horizontal grid spacing, Δ , with the grid spacing determined by $\Delta = \sqrt{(\Delta x)^2 + (\Delta y)^2}$. Following Hallberg (2013) and Griffies and Treguier (2013), regions where the ratio is less than two (i.e., the white regions) support a relatively poor representation of mesoscale processes (e.g., baroclinic instability), whose scale is determined by the deformation radius. Roughly 53% of the ocean surface area has $L_d/\Delta \geq 2$ for CM4X-p125, whereas for the same ocean state, CM4X-p25 has only 23% of the ocean area with $L_d/\Delta \geq 2$.

Figure 1 exemplifies the challenge posed by Hallberg (2013). Namely, the need to provide rational and accurate parameterizations of incompletely resolved mesoscale processes in an ocean where those processes have widely varying spatial (and temporal) scales, and as such, where those processes might be adequately resolved in some regions yet incompletely resolved in others. Consequently, there is a need for “scale-aware” physics parameterizations to accommodate the regional dependence of ocean dynamical scales. A further illustration of this challenge is revealed by Figure 2, which shows the magnitude of the gradient in the dynamic sea level as computed for 31 December 2014. According to the Hallberg (2013) criteria, the Rossby radius is unresolved by either model in the high latitudes (i.e., Figure 1). Even so, the dynamic sea level for both CM4X-p25 and CM4X-p125 contains many meandering mesoscale fronts and eddies in the high latitudes. Evidently, the deformation radius provides a scale for baroclinically unstable linear waves, and as such, it is important for the initiation of mesoscale eddy features. Yet, the deformation radius does not necessarily set the scale of statistically equilibrated non-linear mesoscale fronts and eddies. Indeed, mesoscale eddies are here found at scales larger than the deformation radius, which can be interpreted as due to the inverse cascade of kinetic energy occurring in geostrophic turbulence (e.g., Vallis (2017)). A mesoscale eddy parameterization ideally will not overly dampen such eddies, while accounting for those eddies that are missing.

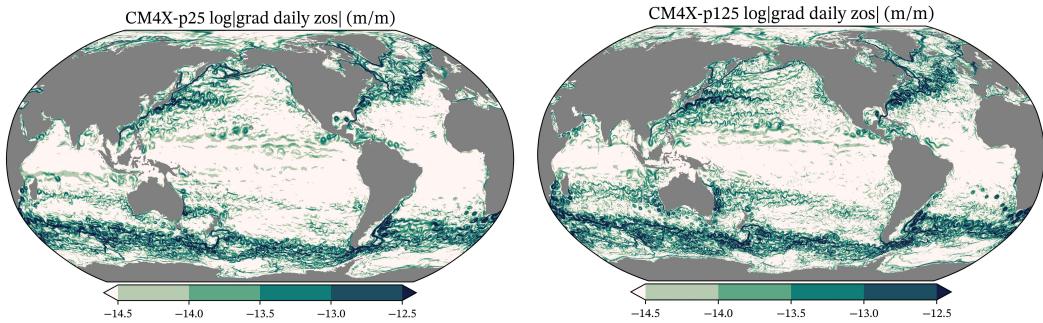


Figure 2. Magnitude of the gradient of the daily mean dynamic sea level for CM4X-p25 (left) and CM4X-p125 (right), as computed from 31 December 2014 in the historical simulations. Meandering mesoscale jets and eddies are represented in these simulations even though the models do not resolve the first baroclinic Rossby radius in most of the high latitudes, in particular within the Southern Ocean (see Figure 1). Note that the dynamic sea level is also known as `zos` within the CMIP community, with details provided in Appendix H of Griffies et al. (2016) and in Section N11 of Gregory et al. (2019).

370 *Eddy activity as revealed by sea level variability*

371 Through the connection between sea level and surface geostrophic currents, sea level
 372 variability provides an effective spatial measure of mesoscale eddy activity in the sur-
 373 face ocean in off-equatorial regions. We thus show in Figure 3 the standard deviation
 374 (root-mean-square) of the daily mean sea level relative to a climatological mean. This
 375 figure reveals the expected increase in variability when moving from CM4X-p25 to CM4X-
 376 p125, given the more active mesoscale eddies admitted by the finer grid spacing. This
 377 increased variability is revealed within the boundary current extensions as well as the
 378 Southern Ocean. For example, the Kuroshio Current in the west Pacific is notably stronger
 379 and extends further eastward in CM4X-p125, making it more consistent with the GLO-
 380 RYS12 analysis product. The Gulf Stream variability also increases in CM4X-p125.

381 Although CM4X-p125 shows more eddy variability than CM4X-p25, it is still far
 382 weaker than the GLORYS12 analysis. Furthermore, there is a notable bias in both CM4X
 383 simulations in which the Gulf Stream exhibits broad and diffuse variability close to the
 384 American coast, which contrasts to GLORYS12 in which variability extends further east-
 385 ward and along a more localized frontal region. We have more to say about deficiencies
 386 of the Gulf Stream in Sections 2 and 6 of Part II (Griffies et al., 2024).

387 *Enhanced role for boundary currents*

388 Along with the inclusion of open ocean transient mesoscale eddies, the CM4X mod-
 389 els capture strong boundary currents, thus allowing for more realistic interactions be-
 390 tween flows along the continental shelves and the open ocean (e.g., for the Antarctic Slope
 391 Current, see Beadling et al. (2022); Beadling (2023); Tesdal et al. (2023)). We illustrate
 392 these features for the depth averaged flow for CM4X-p125 in Figure 4, and then bring
 393 the zonal and meridional flows together in Figure 5 by showing the square root of the
 394 depth averaged kinetic energy.

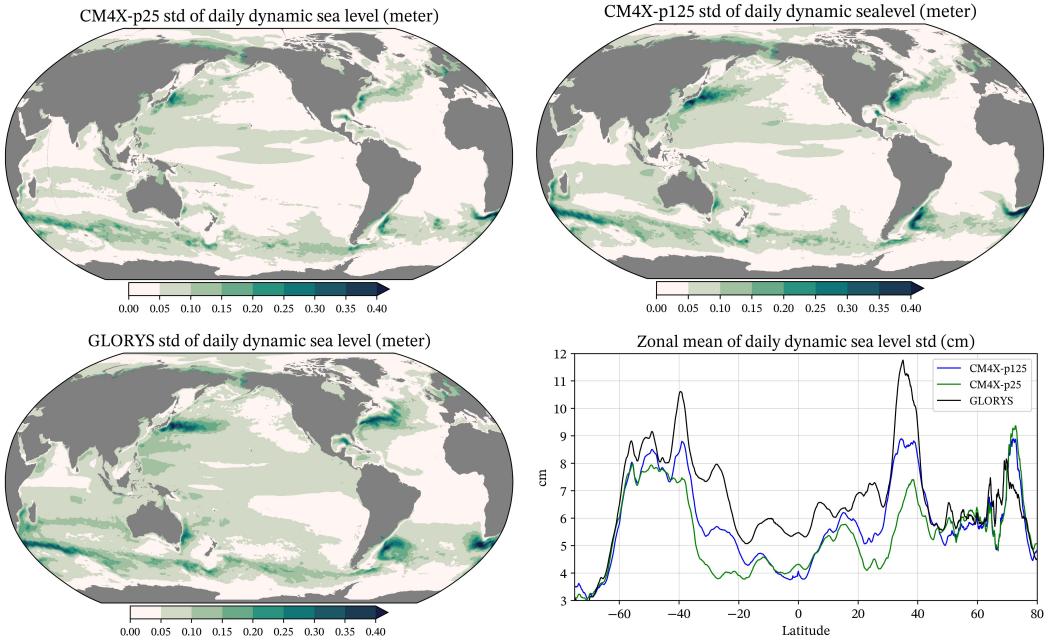


Figure 3. Standard deviation for the daily mean dynamic sea level. The top row shows CM4X-p25 and CM4X-p125, and the bottom row shows the 1/12° GLORYS12 analysis product (Lellouche et al., 2021) as well as the zonal mean for the three. Details for the standard deviation calculation are provided in Section 2 of Part II (Griffies et al., 2024).

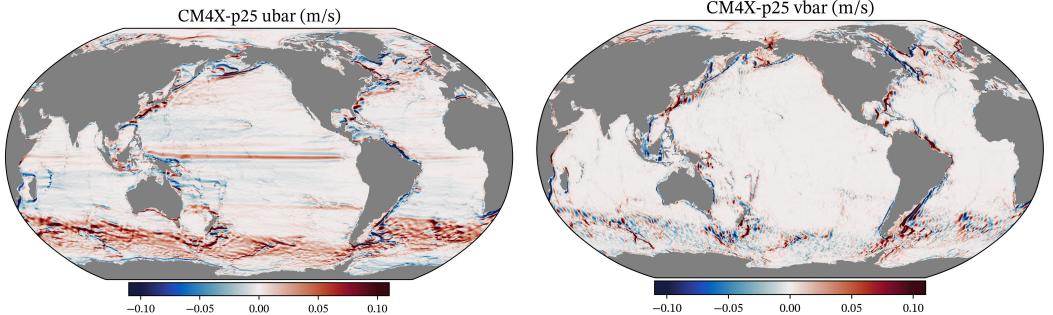


Figure 4. Depth averaged zonal velocity (left) and meridional velocity (right) for CM4X-p125, as computed from the time mean over model years 1980–2009 in the historical simulations. Note the boundary currents as well as the flows at the high latitudes. Additionally, along the equator, we find the zonal flows comprising the equatorial current system.

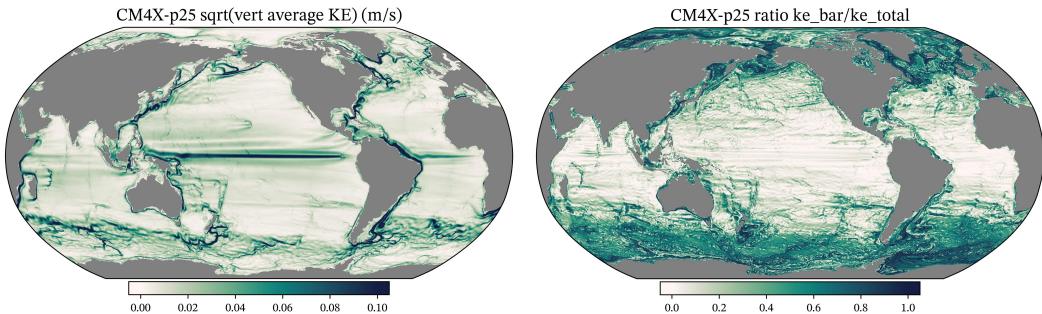
395

Barotropization particularly in the high latitudes

396

As part of the inverse cascade of geostrophic turbulence, we expect mesoscale eddying flows to carry a large portion of the kinetic energy within barotropic motions, and the process of transferring kinetic energy to the barotropic motions is known as “barotropization” (e.g., Charney (1971), Salmon (1980), Smith and Vallis (2001), Scott and Wang (2005), Kjellsson and Zanna (2017), Yankovsky et al. (2022)). Following Yankovsky et al. (2022), in Figure 5 we provide a rudimentary measure of barotropization given by the ratio of kinetic energy contained in the depth-averaged flow (which approximates the barotropic motion) to the depth-averaged kinetic energy. Note how kinetic energy in the lower lat-

404 latitudes is mostly baroclinic, including the equatorial currents. In contrast, high latitude
 405 currents contain a sizable fraction of their kinetic energy in barotropic motions, with this
 406 property related to the relatively weak vertical stratification found particularly in the
 407 high latitude Southern Ocean and North Atlantic. Strong barotropic motions are con-
 408 strained (at least in part) by the need to preserve planetary geostrophic potential vor-
 409 ticity; i.e., to follow contours of constant f/H , where f is the Coriolis parameter and H
 410 is the bottom depth. This property of such flows emphasizes the importance of topographic
 411 representation, as well as for capturing the strong flows that interact with the bottom.



412 **Figure 5.** Left: square root of the depth averaged kinetic energy per mass, $1/2 \sum \Delta z \langle u \rangle \cdot \langle u \rangle / \sum \Delta z$, as computed using the 1980-2009 (in the historical simulations) time mean horizontal velocity for CM4X-p125, $\langle u \rangle$. Right: ratio of the kinetic energy contained in the depth averaged flow, $1/2 \overline{\langle u \rangle} \cdot \overline{\langle u \rangle}$, to the total kinetic energy. Here, we use the depth and time averaged flow as given by $\overline{\langle u \rangle} = \sum \Delta z \langle u \rangle / \sum \Delta z$. Regions where the ratio is large (e.g., Southern Ocean, North Atlantic, boundary currents) are regions where the depth averaged flow plays a leading role in determining the depth averaged kinetic energy.

413 3.3 Northward ocean heat transport

414 In Figure 6 we display the northward ocean heat transport (referenced to 0°C), as
 415 diagnosed from the CM4.0, CM4X-p25, and CM4X-p125 simulations averaged over sim-
 416 ulation years 1980-2009, as well as the reanalysis-based results from Trenberth and Caron
 417 (2001) and ocean *in situ* based analysis of Ganachaud and Wunsch (2000). The model
 418 results are closer to one another than to the reanalysis and *in situ* measures, and gen-
 419 erally show a somewhat reduced magnitude for poleward transport. Furthermore, the
 420 model results are quite close in the northern hemisphere, whereas CM4X-p125 exhibits
 421 somewhat more southward transport away from the tropics in the Indian-Pacific and global
 422 transports. Yet overall we conclude that the CM4.0 and CM4X models are closely aligned
 423 in their poleward heat transports. This result was also found in the CM2-O hierarchy
 424 by Griffies et al. (2015), where it was noted that the structure of the poleward ocean heat
 425 transport is dominated by the atmosphere, whereas the vertical heat transport, and hence
 426 the global balance of heat within the climate system, is dominated by ocean processes
 427 that affect a vertical exchange of heat, such as diapycnal mixing and mesoscale eddy trans-
 port.

428 3.4 Time scales for piControl thermal equilibration

429 The figures shown thus far in this section reveal that both CM4X-p25 and CM4X-
 430 p125 capture a nontrivial amount of mesoscale eddy activity (albeit weaker than the GLO-
 431 RYS12 analysis seen in Figure 3) as well as strong boundary currents, with such assess-
 432 ment relative to non-eddying one-degree models. A key question concerns the degree to

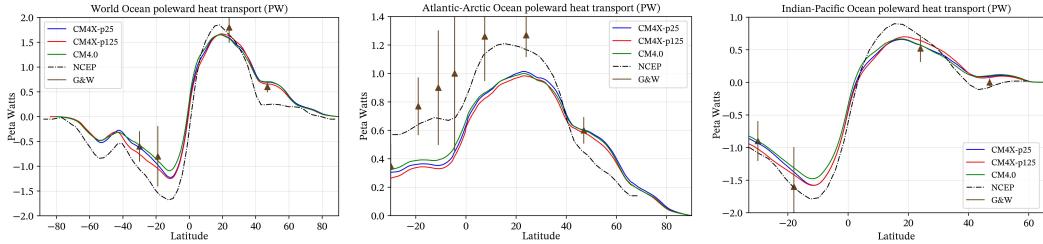


Figure 6. Northward ocean heat transport (referenced to 0°C) for the World Ocean, Atlantic-Arctic Oceans, and Indian-Pacific Oceans. We show results averaged over years 1980-2009 from CM4X-p25, and CM4X-p125, as well as implied ocean heat transports from Trenberth and Caron (2001) using the NCEP reanalysis, and the ocean *in situ* based analysis of Ganachaud and Wunsch (2000). Note the different ranges for the latitude and heat transports for the different panels. The CM4X results are reasonably close to one another, since the poleward heat transport is dominated by the same atmosphere model used in both of the CM4X models.

which these mesoscale features affect the long-term climate and response to anthropogenic warming. Asking this question in a more limited manner, we wish to know whether the strength of ocean eddy activity impacts the time scale for the climate system to reach thermal equilibrium in a piControl simulation. In anticipation of exploring this question in Sections 4, 5 and 6, we here present three time series in Figure 7 from the piControl simulations.

Figure 7 indicates that the area mean SST in the CM4.0 and CM4X models reach a quasi-steady thermal equilibrium after 150 years, though with differing longer term drifts. Note that the SST in CM4X-p25 is cooler than CM4X-p125 and CM4.0. The SST in CM4.0 exhibits a sizable centennial SST fluctuation after year 500, which is related to Southern Ocean super-polynyas in CM4.0 as discussed in Held et al. (2019) and Dunne et al. (2020). As seen in Section 5 of Part II (Griffies et al., 2024), we find no such fluctuations in either CM4X-p25 or CM4X-p125.

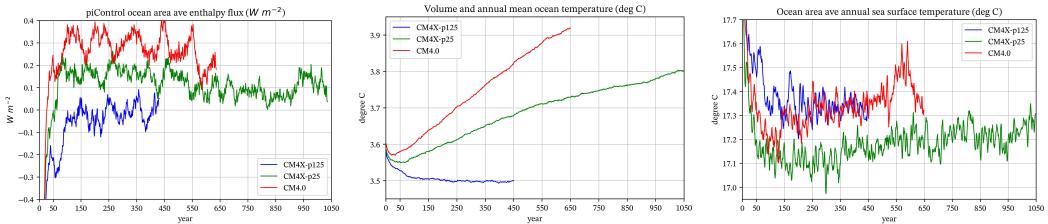


Figure 7. Ocean model diagnostics from the piControl simulations with CM4.0, CM4X-p25, and CM4X-p125. Left panel: net ocean area integrated boundary heat flux entering the ocean, including surface boundary (turbulent fluxes, radiative fluxes, enthalpy associated with mass transfer, and frazil sea ice formation), as well as the bottom geothermal heating. The ocean area provides the normalization for this heat flux, so to get the earth area normalized flux multiply by 0.71. The time series are smoothed with a 30-year rolling mean, hence the missing values near the start and end. Middle panel: time series for the volume mean and annual mean global ocean Conservative Temperature for CM4X-p25, CM4X-p125, and CM4.0. Right panel: time series for the ocean area mean 5-year rolling mean sea surface temperature.

SST is a key ocean field for coupled model simulations due to its interactions with the atmosphere in establishing feedback processes throughout the climate system. But to measure thermal equilibration of the climate system, we must consider the full ocean depth. The reason is the ocean is the major repository for excess enthalpy arising from anthropogenic impacts, thus motivating us to consider the net ocean boundary heat flux and volume mean ocean temperature. (Note that all heat fluxes in this section are reported with an ocean area normalization rather than global area mean.) Figure 7 indicates that the models possess two separate time scales for ocean boundary heat fluxes. During the first few years, there is a negative heat flux (heat leaving the ocean). This initial release of heat is expected since the atmosphere radiative forcing is pre-industrial yet the ocean is initialized with an estimate of the 2013 ocean conditions, with the 2013 ocean warmer than the pre-industrial ocean. Thereafter, the net heat fluxes in both CM4.0 and CM4X-p25 steadily increase, switching from negative to positive after 20 years (CM4.0) and 50 years (CM4X-p25). These models also exhibit centennial scale fluctuations of roughly $\pm 0.1 \text{ W m}^{-2}$ amplitude. The positive heat fluxes entering the ocean in CM4.0 and CM4X-p25 cause the ocean volume mean temperature to exhibit a quasi-linear rise. In contrast, the integrated boundary heat flux in CM4X-p125 increases more gradually over the first 100 years, and thereafter it reaches a near statistical steady state with fluctuations centered around zero. Consequently, the ocean volume mean temperature in CM4X-p125 is nearly steady after around 150 years.

The warming drift in the global volume mean ocean temperature for CM4X-p25 continues until the end of the piControl simulation at year 1050. The trend is generally declining, though with some centennial-scale fluctuations. In particular, during years 751-900, CM4X-p25 has an ocean area mean and time mean boundary heat flux of 0.07 W m^{-2} , whereas for years 901-1050 it is 0.10 W m^{-2} , though it appears to flatten just after year 1000. CM4.0 exhibits a relatively large warming drift, with roughly twice the net imbalance of boundary heat fluxes as found in CM4X-p25. CM4.0 was only run to year 650 due to the presence of Southern Ocean super-polynyas documented in Held et al. (2019) and Dunne et al. (2020). We thus do not know how long it would take CM4.0 to reach thermal equilibrium. In contrast to the long-term warming found in CM4X-p25 and CM4.0, the finer ocean grid spacing in CM4X-p125 leads to far less thermal drift, with roughly -0.02 W m^{-2} net imbalance as averaged over years 100-300 and averaged over the ocean surface area. As a result, CM4X-p125 reaches a nearly steady thermal equilibrium with a cooler ocean after roughly 150 years, and it remains that way for the remaining years of the simulation.

3.5 High latitude overturning and ventilation

Much of our planned research with CM4X concerns the study of physical processes that affect the ocean's meridional overturning circulation (Cessi, 2019; R. Zhang et al., 2019; Hirschi et al., 2020; A. L. Stewart et al., 2021), with those plans guiding the case studies presented in Part II (Griffies et al., 2024). To get a sense of the CM4X overturning circulation, we show the global ocean overturning streamfunction in Figure 8, following the analysis of A. L. Stewart et al. (2021). Both CM4X models display the expected pole-to-pole overturning, whereby the North Atlantic overturning (red clockwise cell) extends into the southern hemisphere and connects to Southern Ocean intermediate waters around 45°S to 50°S . The Atlantic cell is bounded below by the Antarctic Bottom Water cell (blue counter-clockwise cell) that floods the abyssal ocean. Above these cells we find the shallow Ekman-driven overturning cells in the lower latitudes. We have far more to say concerning the Atlantic and Southern Ocean overturning circulation in Sections 5 and 6 of Part II (Griffies et al., 2024).

The mixed layer depth provides a summary diagnostic of processes affecting ocean ventilation. In Figure 9, we show summer and winter climatology of mixed layer depth from the historical and SSP5-8.5 simulations, comparing CM4X-p125 to both Argo and

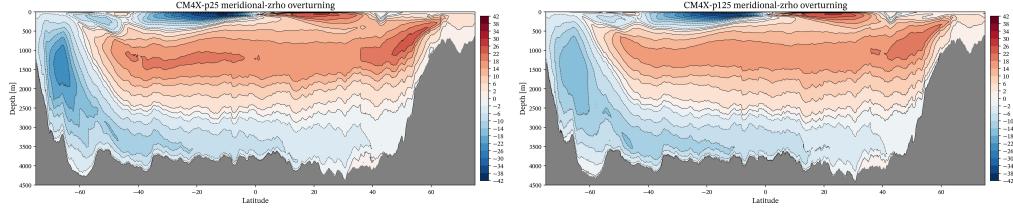


Figure 8. Meridional-density overturning circulation as a function of latitude and potential density referenced to 2000 dbar: $\psi(y, \sigma, t) = -\rho_0 \oint \left[\int_{z'=z_\sigma}^{z'=z_\sigma} v(x, y, z', t) dz' \right] dx$, where $\rho_0 = 1035 \text{ kg m}^{-3}$ is the Boussinesq reference density, σ is the potential density referenced to 2000 dbar, $z_\sigma(x, y, t)$ is the vertical position of the σ isosurface, $z = -H(x, y)$ is the ocean bottom, and v is the meridional velocity. Following A. L. Stewart et al. (2021), we map the circulation onto the time and zonal average depth of the potential density surfaces. Units are mass-Sverdrup (10^9 kg s^{-1}). The left panel shows CM4X-p25 and the right panel CM4X-p125. We make use of 1980-2009 time mean flows and density surfaces to compute this streamfunction. The meridional velocity is mapped to potential density surfaces online using each model time step. Hence, we are here showing the residual-mean overturning circulation.

498 CM4X-p25. Among the regions of notable biases, we see overly deep mixed layer depths
 499 in CM4X-p125 in the mode and intermediate water regions of the Southern Ocean, as
 500 well as in the Labrador Sea (Figure 9D). Biases in the Southern Ocean ventilation for
 501 CM4.0 are the subject of Krasting et al. (2024), with these biases also reflected in the
 502 CM4X simulations. We find a notable reduction in the Labrador Sea bias moving from
 503 CM4X-p25 to CM4X-p125, presumably due to the stronger mesoscale eddies contribut-
 504 ing to more restratification in CM4X-p125 (Figure 9E) (J. Marshall & Schott, 1999). Even
 505 so, the overly deep mixed layer biases in the Southern Ocean worsen in CM4X-p125, pos-
 506 sibly due to the eddy-induced deepening effects in the Indo-West Pacific region described
 507 by Q. Li and Lee (2017). The SSP5-8.5 scenario experiment finds a systematic reduc-
 508 tion in the mixed layer depths throughout most of the high latitude Southern Ocean and
 509 North Atlantic (Figure 9F).

510 3.6 Tropical variability

511 Figure 10 shows the power spectra for the Niño-3 SST index from the historical sim-
 512 ulation (years 1850-2014) with comparison to the NOAA reconstructed product over years
 513 1854-2014 from Huang et al. (2017). Although there is much spread in the spectrum, the
 514 CM4X models show consistently weaker spectral power than the reconstruction across
 515 most time scales. Weak tropical variability is also reflected in the skewness of the sea
 516 surface height (Figure 4 in Part II of Griffies et al. (2024)). We hypothesize that CM4X's
 517 weak tropical variability is related to changes in the ocean mixing parameterizations dis-
 518 cussed in Appendix A. This hypothesis is prompted by noting that the tropical variabil-
 519 ity is stronger (consistent with CM4.0 as documented in Section 5.5 of Held et al. (2019))
 520 when the C192 atmosphere (used here for CM4X) is coupled to the OM4.0 ocean for a
 521 150-year piControl experiment (not shown).

522 Interannual tropical variability can exhibit rather sizable centennial scale modu-
 523 lations (Wittenberg, 2009), so that quasi-stationary power spectra result only after run-
 524 ning multi-centennial scale coupled simulations and/or ensembles. Unfortunately, we could
 525 not afford to run long simulations to examine every separate change made to the ocean
 526 physical parameterizations. Consequently, we are unaware of the precise changes that
 527 led to the weaker variability. Unraveling this story remains an ongoing process. For ex-

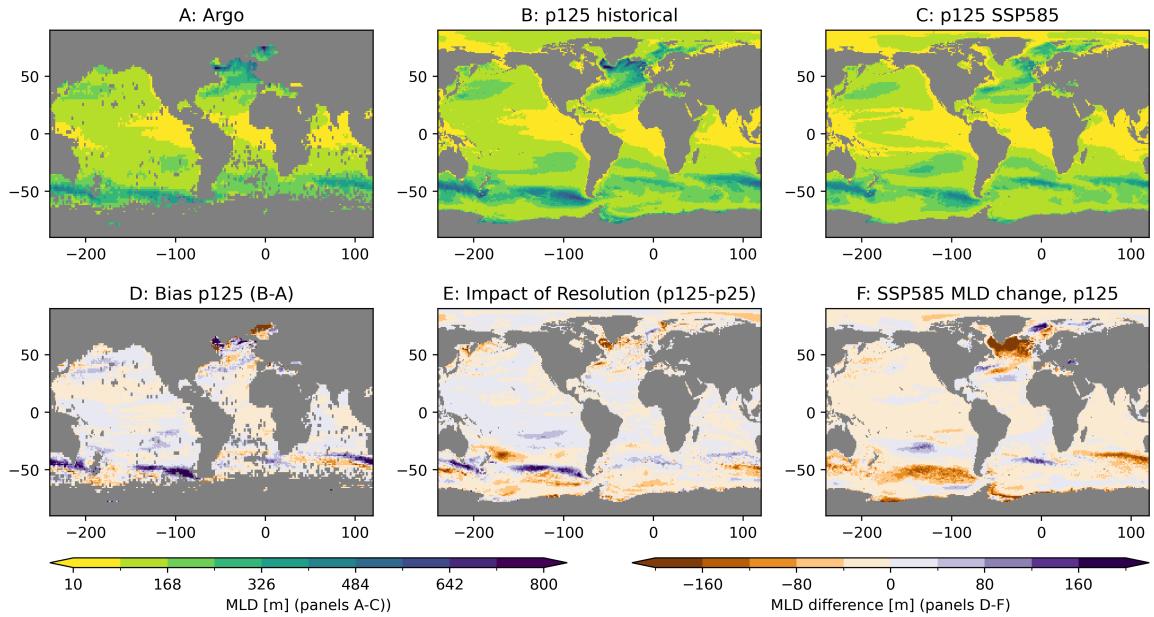


Figure 9. Monthly climatology of the maximum mixed layer depth from Argo (years 2004–2023) (Argo, 2023), as well as years 2000–2014 (historical simulation) and years 2085–2099 (SSP5-8.5 simulation) from CM4X-p25 and CM4X-p125. The mixed layer depth is diagnosed according to the potential energy method from Reichl et al. (2022) (using 2500 J m^{-2} energy criteria). Panel A: estimates from the Argo profiling floats; Panel B: results from CM4X-p125 historical experiment; Panel C: CM4X-p125 SSP5-8.5 simulation. Panel D shows the differences between CM4X-p125 and Argo (Panels B-A), whereas Panel E shows the difference between CM4X-p125 and CM4X-p25. Finally, Panel F shows the impacts from the SSP5-8.5 climate change, showing years 2084–2099 minus years 2000–2014 (Panels C-B) from CM4X-p125. Note that the differences documented in Panels D and E are robust to a longer time average over model years 1955–2014.

528 example, improved representation of upper ocean vertical mixing in this region was not included in CM4X (Reichl et al., 2024).

530 3.7 Coupled climate benchmarks

531 In Figure 11, we present benchmark performance metrics for years 1978–2014 of the
 532 historical simulations in the GFDL earth system model ESM4.1 (Dunne et al., 2020),
 533 along with the climate models CM4.0, CM4X-p25, and CM4X-p125. We compare mod-
 534 els with observations using the PCMDI Metrics Package version 3 (Lee et al., 2023) and
 535 present the results using the “portrait plot” format described in Gleckler et al. (2008).
 536 One caveat for these results is that we only have one ensemble member for each model,
 537 though a single realization is generally reliable for assessing these climatological biases.
 538 As noted by Lee et al. (2023), the GFDL models ESM4.1 and CM4.0 are among the top-
 539 performing CMIP6 models for these atmospheric benchmarks, thus providing a useful
 540 benchmark for the CM4X simulations.

541 Both CM4X-p25 and CM4X-p125 make use of a finer grid for the GFDL-AM4.0
 542 atmosphere relative to CM4.0 (C192 versus C96), which we propose accounts for an over-
 543 all reduction in model biases in both CM4X configurations compared to ESM4.1 and CM4.0.
 544 Furthermore, note how CM4X-p125 exhibits improved performance relative to CM4X-

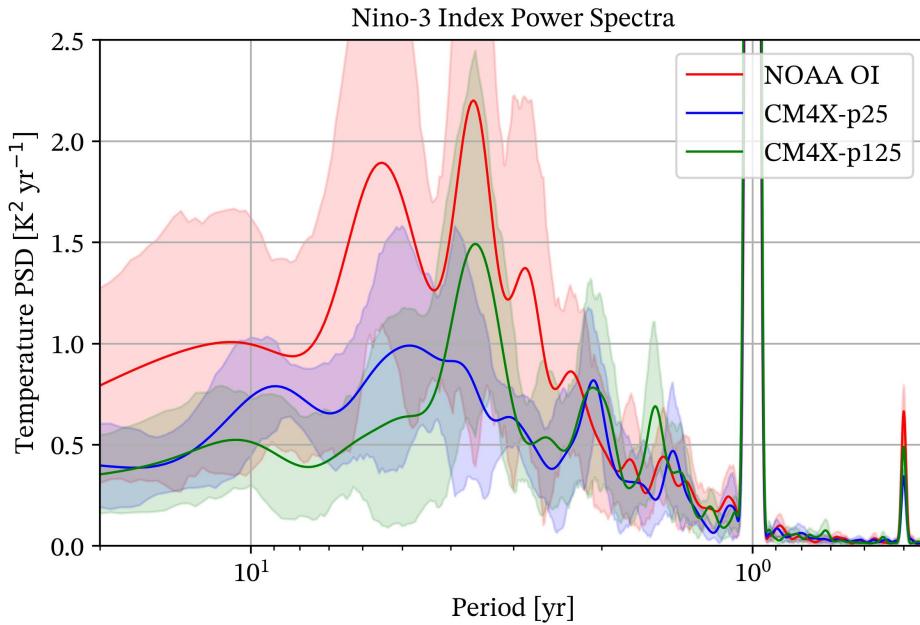


Figure 10. Power spectra for the Niño-3 index, which is computed from the monthly mean sea surface temperature area averaged over the region ($5^{\circ}\text{N} - 5^{\circ}\text{S}$, $150^{\circ}\text{W} - 90^{\circ}\text{W}$) in the Tropical Pacific. We show results from the CM4X simulations over years 1850-2014, as well the reconstructed NOAA temperatures over years 1854-2014 (Huang et al., 2017). The method of computation follows Welch (1967) as realized by the Matlab `pwelch` routine. We use 30-year windows for computing the periodogram and associated confidence intervals.

p25 in the mean mid-tropospheric atmospheric circulation (zg-500) and upper tropospheric temperatures (ta-200). This comparison exemplifies how coupling to the finer resolution ocean in CM4X-p125 can render benefits to the atmospheric simulation. Some seasonal differences are present, particularly in the boundary layer winds and temperatures in JJA. Treatment of the land model differs across ESM4.1, CM4.0, and CM4X (see Sections 2.4 and A2). CM4.0 uses an earlier version of the GFDL Land Model (LM4.0) with dynamic vegetation that responds to a changing climate, whereas ESM4.1 uses an updated version (LM4.1) that has advancements in hydrology and terrestrial vegetation geared towards Earth System Modeling (ESM) applications. CM4X, with its focus on exploring ocean questions, uses a version of LM4.0 with static vegetation (circa 1980). Land surface forcing differences also play a role in the coupled climate model performance (Ghimire et al., 2014; Zhao, 2022a), especially when evaluating model simulations of historical temperature trends. Despite these qualifications, the performance metrics in Figure 11, along with the broader assessment provided by Lee et al. (2023), suggest that CM4X historical simulations perform among the best-in-class for atmospheric metrics within the current generation of coupled climate models.

4 Ocean surface temperature properties

In this section we focus on the surface ocean temperature (SST) and related boundary fluxes.

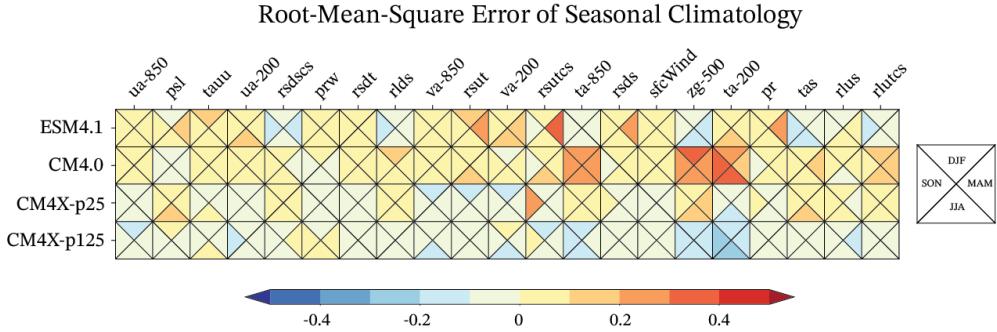


Figure 11. Performance metric portrait plot taken after Gleckler et al. (2008), here focused on atmospheric metrics from the ESM4.1, CM4.0, CM4X-p25, and CM4X-p125 historical simulations over years 1978-2014 of the historical simulations. The root-mean-square-error (RMSE) values are normalized across the models shown here, so we are showing relative error rather than absolute error. As per Table 1 in Lee et al. (2023), we compare the simulations to the precipitation from GPCP-2.3 (Adler et al., 2018); sea level pressure, temperatures, and winds from ERA5 (Hersbach et al., 2020); surface wind stress from ERA-Interim (Dee et al., 2011); and radiation from CERES-EBAF-4.1 (Loeb et al., 2018). Blue shading denotes improved model skill (relative to the models shown) while red shading denotes degraded model skill.

564

4.1 SST variability and turbulent heat fluxes

565
566
567
568
569
570
571
572
573
574
575
576

In Figure 12 we show one measure of transient variability in CM4X-p25 and CM4X-p125 is shown, here mapping the temporal standard deviation of the daily mean SST computed over years 1980-2009. Both models show the characteristic tropical Pacific pattern related to the El Niño-Southern Oscillation variability. In the middle latitudes, variability is largely concentrated near the western boundary currents, such as the Gulf Stream in the North Atlantic and Kuroshio in the North Pacific, with such variability largely associated with transient mesoscale eddies. Additional variations are seen in the Southern Ocean and are associated with mesoscale jets and eddies. In the polar regions, SST variability is muted due to the presence of sea ice and the freezing point lower bound on temperature. For each of the middle and high latitude patterns, CM4X-p125 shows a slight increase in the amplitude of the variability relative to CM4X-p25, along with an extension of the variability eastward away from the western boundaries.

577
578
579
580
581
582
583
584
585
586
587
588
589
590

Following Kirtman et al. (2012), we provide a measure of the impact of ocean variability on the atmosphere by computing the temporal correlations of the monthly SST anomalies with the monthly anomalies of the surface turbulent heat fluxes (sensible plus latent). These correlations are shown in Figure 13 for CM4X. As discussed by Kirtman et al. (2012) (and references therein), a negative correlation reveals regions where the ocean forces the atmosphere. Such regions are here seen to be closely tied to regions with large sea surface temperature variability. The western tropical Pacific is a notable region where the atmosphere forces the ocean, which corresponds to the region of strong tropical atmospheric convection. Kirtman et al. (2012) noted that for coarser models, such as the one-degree class of ocean models that do not admit mesoscale eddy fluctuations, most of the significant correlations are positive, indicating that the ocean is forced by the atmosphere. In contrast, the two CM4X simulations are dominated by correlations indicating that the atmosphere is forced by the ocean. Also observe that the polar regions show a large correlation and yet very little temperature variation. This be-

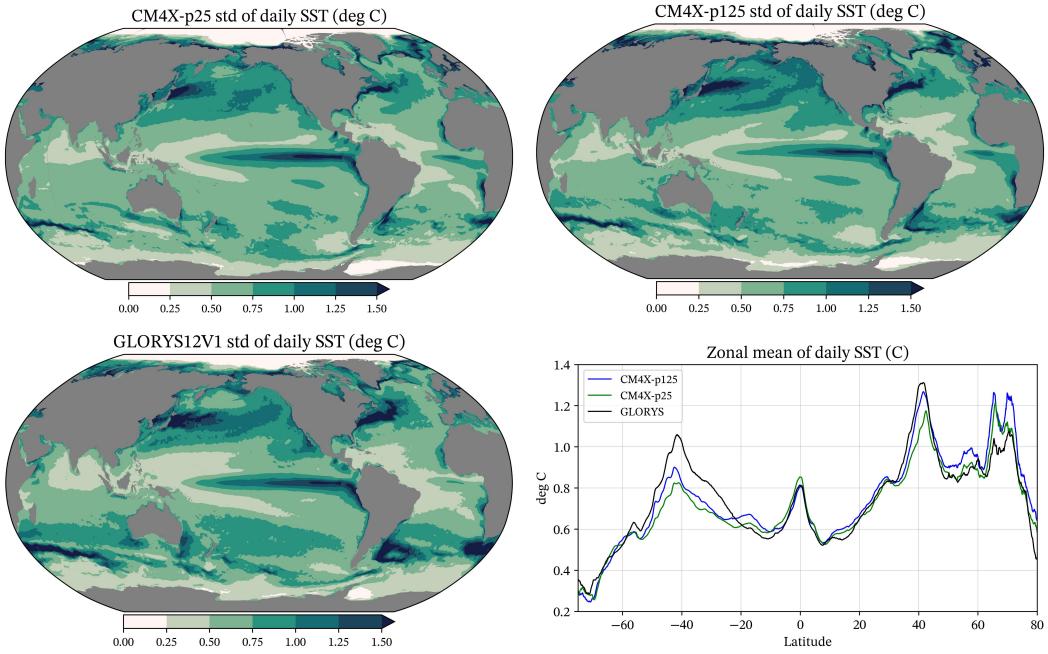


Figure 12. Standard deviation (degree C) of the sea surface temperature in CM4X-p25 and CM4X-p125 as computed over years 1980-2009, as well as the $1/12^\circ$ GLORYS12 analysis product (Lellouche et al., 2021) along with the zonal mean for the three. The standard deviation is computed by taking the mean squared difference of the daily mean SST from the climate mean for that day computed over years 1980-2009.

591 behavior reflects the dominance of sea ice formation/melt that produce latent heat fluxes
 592 yet little temperature changes.

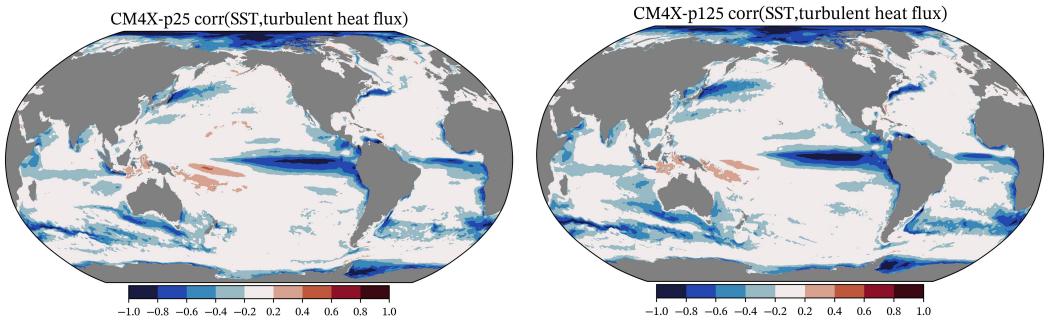


Figure 13. Temporal correlations between the monthly sea surface temperature and the turbulent surface heat fluxes (latent plus sensible). This correlation is computed over years 201-250 from the piControl simulations, with the SST and heat fluxes each computed relative to their climatological mean for each month. The sign convention is such that a positive surface heat flux warms the ocean and negative flux cools. We chose the color bar to emphasize correlations with amplitude larger than 0.2. These correlation maps can be compared to Figure 19 of Kirtman et al. (2012), though note that we chose the opposite sign convention for the heat fluxes.

593 **4.2 Historical sea surface temperature biases**

594 Figure 14 shows the sea surface temperature biases computed from the 30-year time
 595 mean of the simulations over years 1980-2009. There are many common biases across
 596 the models, such as the strong cold bias in the region east of Greenland and the warm
 597 bias in the Labrador Sea. We hypothesize that these North Atlantic biases arise, at least
 598 in part, from relatively poor representation of the mesoscale eddy processes and over-
 599 flow processes (R. Zhang et al., 2011). The eastern boundary regions off California as
 600 well as West Africa south of the equator both show warm biases that generally reflect
 601 biases in the stratocumulus clouds (e.g., see Wang et al. (2014); Richter (2015)). Notably,
 602 however, there is only a modest bias off the west coast of South America in the two CM4X
 603 simulations, which is significantly smaller than those found by Wang et al. (2014) and
 604 Richter (2015) in earlier climate models, and smaller than the CM4.0 bias in this region.
 605 This improvement is presumably due to the finer atmospheric grid in CM4X relative to
 606 CM4.0, though some studies suggest that finer ocean grid spacing can help reduce the
 607 stratocumulus cloud biases by enhancing coastal upwelling (Richter, 2015). We have more
 608 to say about these eastern boundary regions in Section 3 of Griffies et al. (2024).

609 Many of the SST biases in CM4X are broadly reflected in the CM4.0 biases, with
 610 the global root-mean-square bias of 0.92°C in CM4X-p25 close to the 0.90°C found in
 611 CM4.0. However, in nearly all regions in the World Ocean, SST biases in CM4X-p125
 612 are reduced relative to both CM4X-p25 and CM4.0, as reflected by the roughly 20% re-
 613 duction in the global root-mean-square bias (0.73°C for CM4X-p125) and 40% reduc-
 614 tion in the mean bias (-0.51° for CM4X-p25 and -0.29° for CM4X-p125). This bias re-
 615 duction is generally realized by a broadly warmer SST in CM4X-p125, thus reducing,
 616 for example, the broad-scale cool bias in the Southern Hemisphere middle latitudes. Fur-
 617 thermore, it is notable that both CM4X models display only modest biases in the South-
 618 ern Ocean relative to the larger warm bias in CM4.0 and even larger warm biases found
 619 in many other climate models (Sallée et al., 2013; Wang et al., 2014; Hyder et al., 2018;
 620 Beadling et al., 2020).

621 CM4X shows a general improvement with the overly cool eastern equatorial Pa-
 622 cific. This bias has been persistent in models, and is of importance due to its impact on
 623 the Hadley cell (G. Li et al., 2015). The Pacific subtropical gyre bias is also reduced, which
 624 supports a hypothesis from Burls et al. (2016) and Thomas and Federov (2017) that the
 625 equatorial and subtropical biases are connected. Finally, note the reduced bias in CM4X-
 626 p125 found in the western Atlantic near the U.S. coast, whereby the strong warm bias
 627 in CM4X-p25 (also seen in CM4.0) is largely absent in CM4X-p125. This reduced bias
 628 reflects differences in the Gulf Stream as represented in CM4X-p25 and CM4X-p125. We
 629 provide more analysis of the Gulf Stream in Section 6.2 of Part II (Griffies et al., 2024),
 630 where we identify a number of shortcomings in both CM4X models.

631 **4.3 Patterns of SST change under SSP5-8.5**

632 Figure 15 shows the change in SST under the SSP5-8.5 scenario. We compute this
 633 pattern as the difference between the final ten years of the scenario (years 2090-2099)
 634 and the 30-year mean over years 1980-2009 of the historical. There is broad agreement
 635 between CM4.0 and the two CM4X models, showing a general warming pattern in lower
 636 to middle latitude open ocean, with the exception of a muted warming in the southeast
 637 and south central Pacific. A muted southeast and south central Pacific warming has been
 638 seen in historical changes, with this pattern suggested to be related to changes in the
 639 Walker Circulation (Wills et al., 2022). However, a systematic comparison of the histor-
 640 ical SST trends would require multiple realizations for the historical period. The North
 641 Pacific shows general warming, with CM4X-p125 somewhat larger.

642 The Southern Ocean shows muted warming and some patches of cooling, with the
 643 exact positioning of these features likely dependent on decadal modes of climate vari-

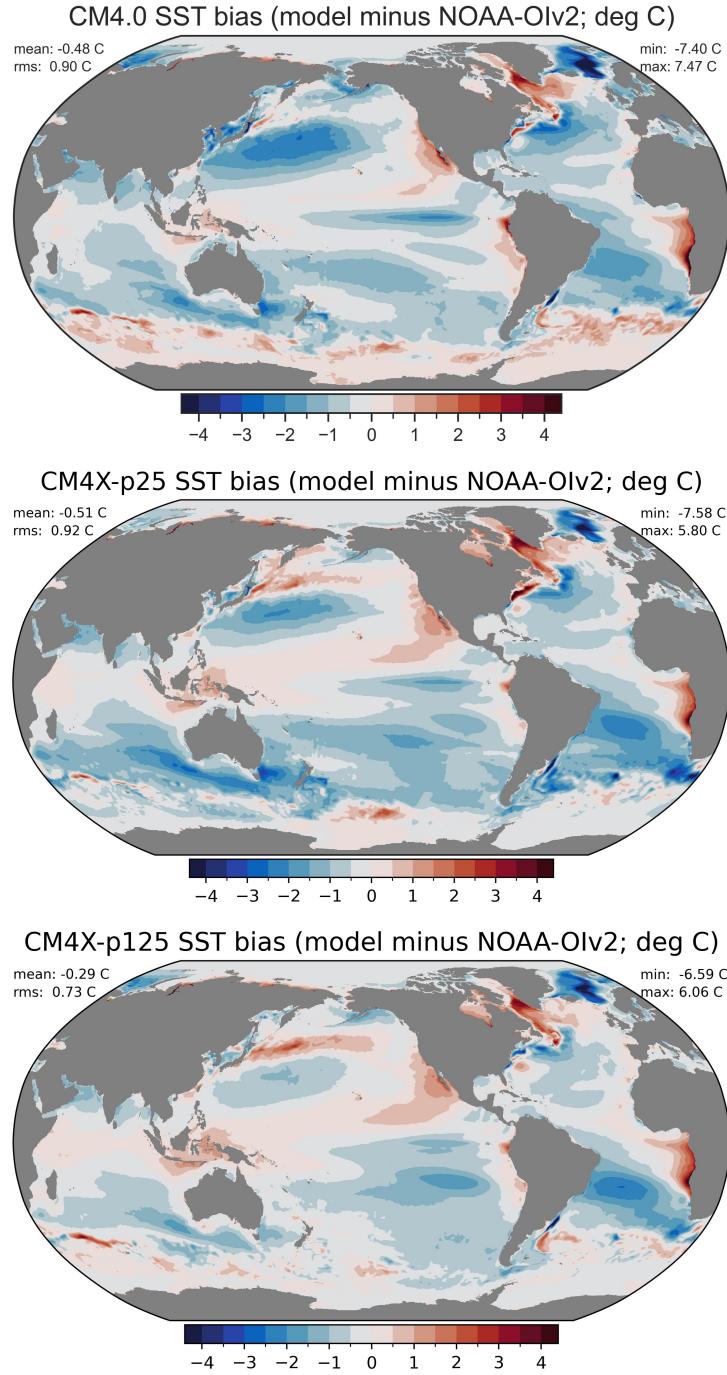


Figure 14. Sea surface temperature bias as defined by the difference between the model time mean over years 1980-2009 versus the NOAA OIv2 observational-based analysis (also averaged over 1980-2009) from Huang et al. (2020). The top panel shows CM4.0, whereas the lower two panels show CM4X-p25 and CM4X-p125. In each case, the NOAA analysis was interpolated to the model grid.

ability. This delayed Southern Ocean warming is related to the role of the deep ocean mixing, in which the Southern Ocean absorbs the bulk of anthropogenic heat and yet

644
645

most of that heat is transported to the north and exported to other ocean basins (Manabe et al., 1990, 1991; Gregory et al., 2016; Armour et al., 2016). The North Atlantic shows a distinctive tripole pattern, with warming next to the east coast of North America, cooling in the subpolar gyre, and warming to the east of Greenland. The cooling is related to a weakening of the Atlantic overturning circulation (Manabe et al., 1990, 1991; R. Zhang et al., 2019), whereas the coastal warming arises from associated changes in the Gulf Stream (Saba et al., 2016; Caesar et al., 2018). Note the larger warming off the American coast in CM4X-p125 relative to CM4X-p25, which is consistent with the CM2.5/CM2.6 results found by Saba et al. (2016), in which refined grid spacing produces both a more accurate simulation of coastal currents and a larger projection for coastal warming. Indeed, we find that CM4X-p125 shows increased SST warming relative to CM4X-p25 around the full North American continent (both Atlantic and Pacific, including the subpolar Pacific).

There is a rather large warming to the east of Greenland. This warming is over the anomalously cold region found in the historical experiment for CM4.0 and the two CM4X models (see Figure 14). This region also has anomalously large positive sea ice biases during the historical simulation (see Figure 8 in Part II (Griffies et al., 2024)). The sea ice melts back during the SSP5-8.5, which accounts for the large SST warming in Figure 15.

5 Ocean spin-up and historical/SSP5-8.5 response

In this section we build on the results from Section 3.4 by focusing on the timescales of equilibration in the piControl simulation. We also present a selection of enthalpy budgets from the historical and SSP5-8.5 simulations. As detailed in Section 3.1, the piControl starts from 21st century ocean observations of temperature and salinity (with zero ocean currents and sea ice spun-up from a previous simulation) and couples to an atmosphere with 1850 pre-industrial radiative forcing. Hence, the piControl experiment is an instantaneous cooling experiment. One may expect the piControl to expel the anthropogenic ocean heat and then equilibrate to a cooler volume mean ocean. Although much of the upper 1000 m of anthropogenic heat might be expected to release within $\mathcal{O}(100)$ years of pre-industrial forcing, it will presumably take far longer for deep ocean heat to be released. Even so, since the bulk of anthropogenic heat is in the upper 1000 m of the ocean, it is plausible that near thermal equilibration can arise on a timescale decoupled from the deep ocean. As shown in this section, this expectation is realized for the CM4X-p125 simulation, which reaches a cooler equilibrium state within 100-150 years, whereas the CM4X-p25 simulation drifts warm and has yet to thermally equilibrate (to the degree of CM4X-p125) even after 1000 years.

5.1 Volume mean Conservative Temperature

In Figure 16, we present the time series for annual mean and ocean volume mean Conservative Temperature from the piControl, historical, and SSP5-8.5 experiments for CM4X. The CM4X models display an initial cooling in their piControl simulations. Yet CM4X-p25 starts to warm after roughly 50 years, with a nearly linear warming trend thereafter. In contrast, CM4X-p125 cools for roughly 100 years and then stays relatively steady afterward for the duration of the experiment. Averaged over years 101-200, the net heat flux entering the ocean (positive flux enters the ocean) is 0.16 W m^{-2} for CM4X-p25. In contrast, for CM4X-p125 the imbalance is a mere -0.02 W m^{-2} , and this imbalance holds until the end of the piControl (we provide a more detailed global heat budget analysis in Section 5.6). Note that we use the ocean surface area to normalize these fluxes. Use of the earth's full surface area reduces the magnitude of the imbalance by roughly 30%.

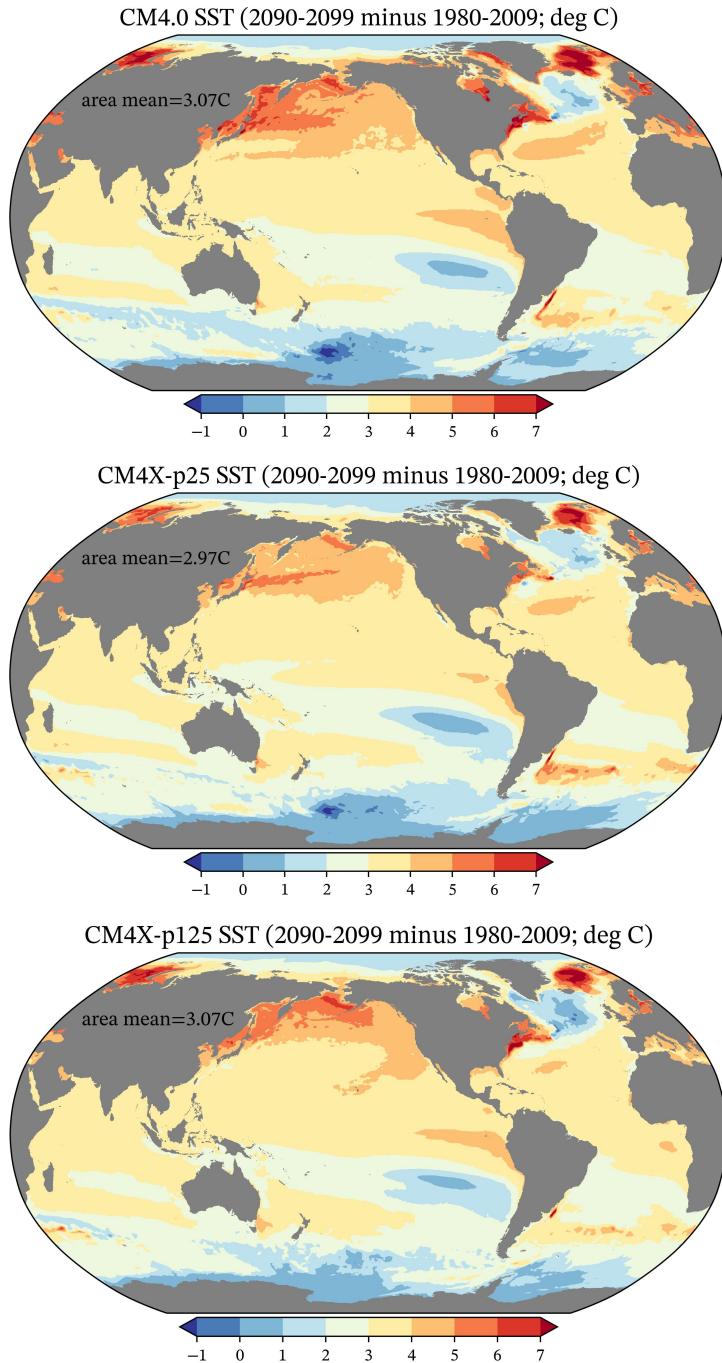


Figure 15. SST change under the SSP5-8.5 scenario, computed as the difference between the ten year mean from years 2090-2099, versus the 30-year mean of the historical 1980-2009. We insert the area SST change average over Asia. The top panel shows CM4.0 and lower panels show CM4X-p25 and CM4X-p125. Note the non-symmetric color scales, with only the dark blues reflecting cooling. Use of the final ten years of the SSP5-8.5 scenario (rather than, say, the final 30 years) enhances the signal from climate change, and it is commonly used for such purpose; e.g., Newsom et al. (2023).

695 The CM4X models include a prescribed geothermal heat flux at the ocean bottom
 696 with a global ocean area mean of 0.094 W m^{-2} (see Figure A2). After approximately
 697 100 years in the piControl simulation, the area integrated geothermal heat flux entering
 698 the ocean bottom in CM4X-p125 is nearly balanced by an equivalent heat flux leaving
 699 the ocean surface, thus leading to a very small residual. In contrast, CM4X-p25 re-
 700 alizes a net positive heating and thus a steady rise in global mean ocean Conservative
 701 Temperature. As a point of comparison, the CM2.5 (0.25° ocean) and CM2.6 (0.1° ocean)
 702 models from the CM2-O hierarchy (Griffies et al., 2015) have a far larger pre-industrial
 703 warming drift of roughly 1.5 W m^{-2} (CM2.5) and 0.5 W m^{-2} (CM2.6) during the sec-
 704 ond century in their piControl simulations.

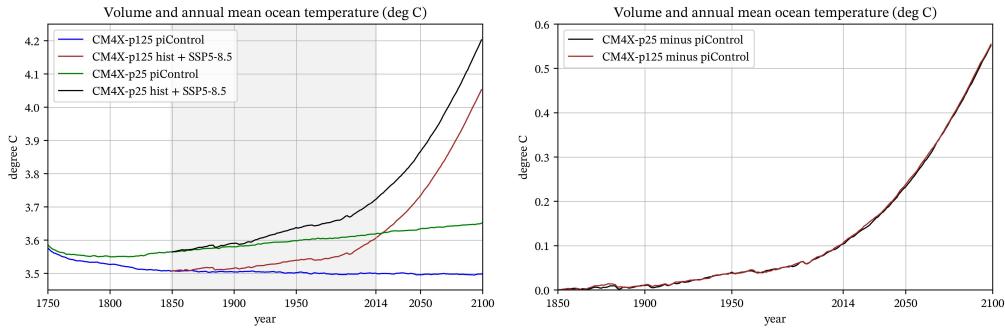


Figure 16. Left panel: time series for the volume mean and annual mean ocean Conservative Temperature for CM4X-p25 and CM4X-p125 when run for the piControl, historical (1850–2014 shaded), and SSP5-8.5 from 2014–2100. The piControl calendar years are nominal, here set to allow for alignment with the historical and SSP5-8.5 experiments. For the piControl experiments, both CM4X models are initialized from the 2013 World Ocean Atlas (Section 3.1), with slight disagreement between CM4X-p125 and CM4X-p25 at the initial time arising from the differences in ocean geometry. There are no "hidden" simulation years prior to those shown here. The historical simulations branch from the piControl from year 101 and run from 1850 to 2014 (gray shaded region). The SSP5-8.5 then starts from the historical at year 2015. Right panel: Historical and SSP5-8.5 time series for CM4X-p25 and CM4X-p125, relative to their respective piControl simulations shown in the left panel. Removing the piControl allows for time series for both models collapse on top of one another.

705 The historical simulations shown in Figure 16 branch at year 101 of the piControl
 706 and then run from year 1850 to 2014. A perfect historical simulation will return the ocean
 707 mean temperature to its 2013 initial condition when the simulation reaches year 2013.
 708 Since CM4X-p25 warms during its piControl, its 1850 initial condition is already at the
 709 2013 temperature, with the result being a simulated 2013 global mean temperature roughly
 710 0.15 K warmer than the 2013 measurements. This result has implications for state-dependent
 711 transient responses, such as those discussed in Stouffer et al. (2006) as well as for sea level
 712 rise (Hallberg et al., 2013). In contrast to CM4X-p25, Figure 16 shows that CM4X-p125
 713 cools during the first 100 years of its piControl, so that its simulated 2013 ocean tem-
 714 perature is only about 0.03 K warmer than 2013 measurements. Based on the further
 715 slight cooling from years 101–150 of the CM4X-p125 piControl, initializing the histor-
 716 ical at year 150 rather than year 100 could have reduced the already small CM4X-p125
 717 overshoot of 0.03 K to an even smaller value.

718 The SSP5-8.5 simulations shown in Figure 16 extend from year 2014 to year 2100.
 719 Removing their respective piControls (right panel of Figure 16), collapses the CM4X-

720 p25 and CM4X-p125 historical and SSP5-8.5 curves on top of one another, thus revealing
 721 the very similar climate sensitivities between the two CM4X models.

722 5.2 Area mean SST

723 In Figure 17, we show the time series for the area-mean SST. For the piControl sim-
 724 ulations, both CM4X models reach a statistically stationary state after roughly 100 years,
 725 with CM4X-p125 approximately 0.25°C warmer than CM4X-p25. The warmer SST in
 726 CM4X-p125 was indicated earlier by the SST bias patterns in Figure 14. For the SSP5-
 727 8.5 climate change simulation, the SST rises by roughly 3°C from 2014 to 2100, with both
 728 models exhibiting a nearly parallel increase. There is a notable cool bias in both CM4X
 729 models found during years 1950-2014. This cool bias is also seen in CM4.0 (see Figure
 730 12 in Held et al. (2019)), and it has been related to aerosol effects by Zhao et al. (2018b).

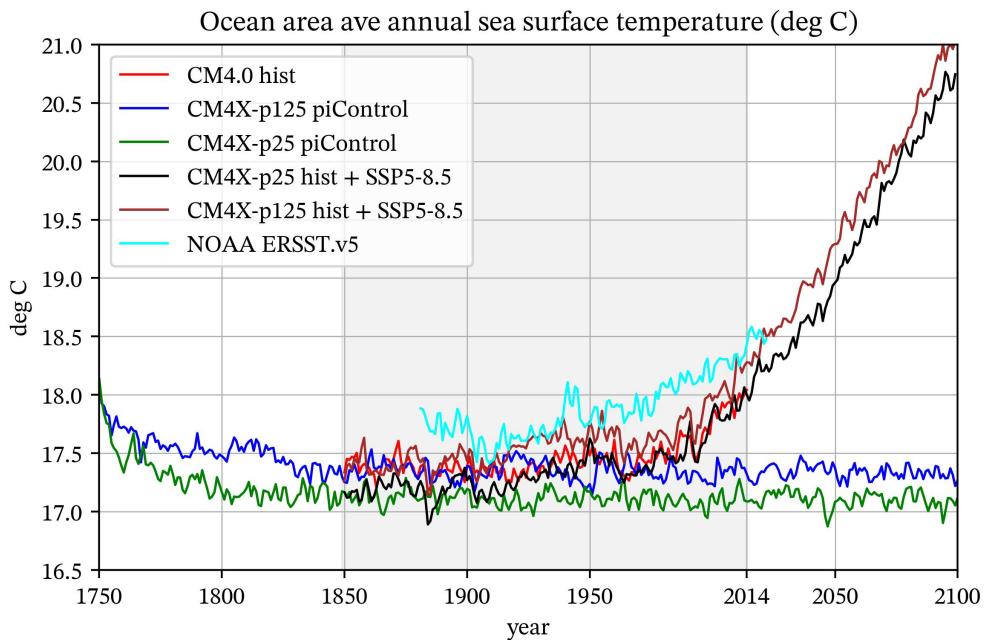


Figure 17. Time series for the ocean area mean annual sea surface temperature as in Figure 16. The historical simulations branch from the piControl and run from year 1850 to 2014 (gray shaded region). The cooling blips in the historical simulation result from volcanic aerosol forcing. The SSP5-8.5 scenario extends from year 2014 to 2100. We also show the NOAA ERSST v5 (Huang et al., 2017) (cyan) starting at year 1880.

731 5.3 Depth-time Conservative Temperature: World Ocean

732 We expose more details about the Conservative Temperature evolution in Figure
 733 18, which shows the depth-time diagram for the global horizontally averaged annual mean
 734 temperature in the piControl simulations. Both models cool in the upper ocean, consis-
 735 tent with the SST in Figure 17 and consistent with the pre-industrial forcing. However,
 736 both also show a warming trend in intermediate depths, and slight cooling again in the
 737 abyssal regions. The upper and intermediate depth trends are enhanced in CM4X-p25
 738 relative to CM4X-p125, thus leading to the cooler surface temperature and warmer in-
 739 termediate temperature in CM4X-p25. As for the global volume mean time series in Fig-

ure 16, it is remarkable how stationary CM4X-p125 is for all depths after roughly 100 years of the piControl.

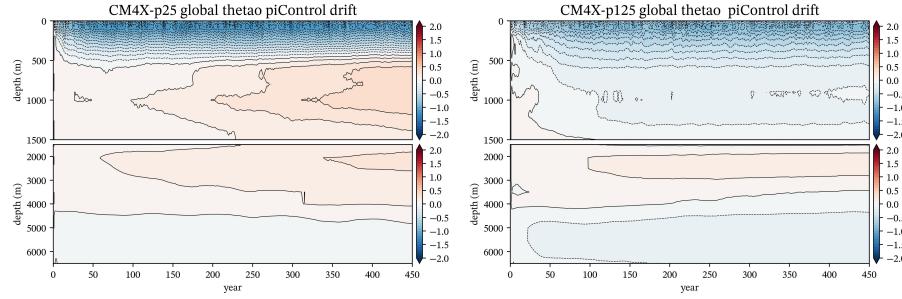


Figure 18. Depth-time drift for the global horizontally averaged annual mean Conservative Temperature. We show temperature evolution relative to the initial year as realized for the piControl simulations, computed as $\theta_{\text{drift}}(z, t) = \sum_{xy} \Delta V [\theta(x, y, z, t) - \theta(x, y, z)_{\text{init}}(x, y, z)] / \sum_{xy} \Delta V$, where $\theta(x, y, z)_{\text{init}}(x, y, z)$ is the annual mean of the first simulation year, and ΔV is the volume of a grid cell. Note the split between the upper ocean and deeper ocean, though with both regions having the same color range, with this range set to correspond to the basin time series shown in Figure 20.

Figure 19 plots centennial scale vertical profiles of the horizontally averaged Conservative Temperature, relative to the initial year. This figure emphasizes the warming found in CM4X-p25 below around 500 m depth and until roughly 4000 m. Furthermore, note how the vertical temperature anomaly in CM4X-p25 grows over the four centuries, partly due to the interior continuing to warm and partly due to the enhanced cooling of the upper ocean after the first century. An upper ocean cooling and interior warming results in a reduction in the surface-to-interior temperature difference. The volume mean ocean warming in CM4X-p25 is a result of a net heat going from the atmosphere into the ocean (see Figure 7 and the heat budget in Section 5.6). The thermal adjustment in the CM4X-p125 piControl is fundamentally distinct from CM4X-p25, in which CM4X-p125 locks into a nearly stationary vertical temperature profile during the second century and afterward, from the surface down to around 3000 m, with cooling down to around 1500 m, warming between 1500 m to 3500 m, and slight cooling below.

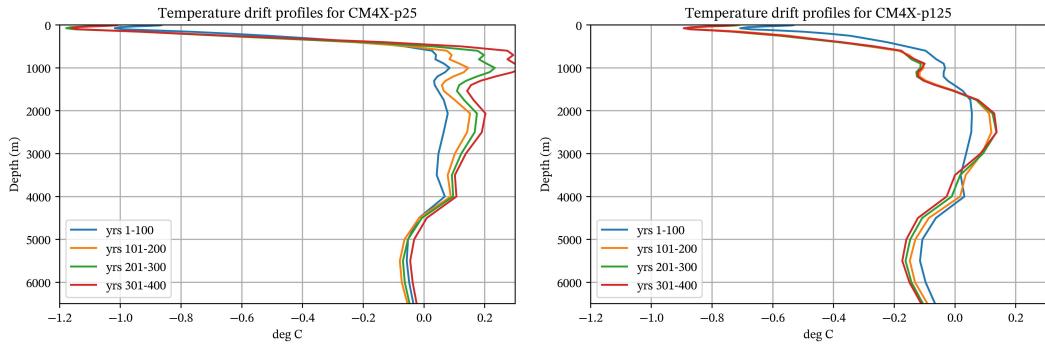


Figure 19. Vertical profiles of horizontally averaged Conservative Temperature relative to the initial year, computed as a time mean for years 1-100, 101-200, 201-300, and 301-400. The left panel shows CM4X-p25 and right panel shows CM4X-p125.

We interpret these results according to the mesoscale eddy effects described by Griffies et al. (2015), whereby mesoscale eddies regulate the drift in water mass properties through their systematic restratification effects (and corresponding shoaling of the pycnocline) and leading to a net upward transport of buoyancy (when integrated horizontally over the globe). Griffies et al. (2015) based this interpretation on simulations with CM2.5 and CM2.6, which are coupled climate models from an earlier generation of GFDL models, with CM2.5 and CM2.6 differing only via their ocean horizontal grid spacing (CM2.5 has a 0.25° grid whereas CM2.6 has a 0.10° grid). As with the CM4X models, the volume mean thermal drift in CM2.6 is significantly smaller than CM2.5.

In an ocean with relatively weak mesoscale eddies, such as CM4X-p25 and CM2.5, the upper ocean boundary forcing is only modestly offset by eddy transport. Surface cooling from the piControl atmosphere reduces the vertical temperature gradient, allowing for more mixing between the relatively warm upper ocean and the cooler interior ocean and accounting for the interior (below 500 m) warm drift. The atmosphere, in response to the relatively cool SSTs in CM4X-p25, enhances the heat going into the ocean surface, which in turn leads to even more heat entering the ocean interior even while the SSTs remain relatively cool.

The story in CM4X-p125 is distinct due to the stronger role of mesoscale eddy transport, whereby stronger eddies act to limit the surface cooling and interior warming after the first century. That is, the surface cooling during the first few decades of the CM4X-p125 piControl is met by an upward transport of ocean heat (presumably dominated by anthropogenically induced heat contained in the initial conditions), thus leading to a cooling of the CM4X-p125 ocean down to around 1500 m. In turn, the eddies partially counteract the destratification affected by the surface cooling. Once the bulk of the heat is released after $\mathcal{O}(100)$ years, the ocean temperature profile remains nearly stationary from the surface down to around 3000 m. This stationarity is matched by a tiny net heat crossing the ocean boundaries, with ocean heat gained at the bottom through geothermal heating (roughly $+0.07 \text{ W m}^{-2}$ averaged over the ocean), and ocean heat lost at the surface of nearly the same magnitude as the geothermal heating (we provide more details of the global heat budget in Section 5.6). In this manner, the CM4X-p125 ocean reaches near thermal equilibrium with the pre-industrial atmosphere after $\mathcal{O}(100)$ years.

5.4 Depth-time Conservative Temperature: basin averages

The depth-time patterns for the World Ocean temperature adjustments seen in Figure 18 arise from an array of water mass adjustments within and between ocean basins. Figure 20 exposes further structure by decomposing the adjustment into basins, and we also include results from CM4.0. This figure reveals that across all basins, CM4X-p125 has a muted adjustment amplitude relative to both CM4X-p25 and CM4.0, and CM4X-p125 shows a quicker equilibration to a quasi-stationary state. Furthermore, the global adjustment of temperature is dominated by the Pacific and, to some degree, the Indian basins (here shown combined), which comprise the bulk of the World Ocean volume. For the Atlantic and Southern Oceans, the surface cooling penetrates to roughly 1000 m, which contrasts with the more surface-focused cooling (upper 100 m) in the Indian and Pacific basins. The deeper penetration for the Atlantic and Southern Oceans reflects the presence of deep ventilation that allows for the release of heat from deeper waters.

CM4X-p25 and CM4.0 show a warming between 500 m-4000 m in the Atlantic/Arctic basins, whereas CM4X-p125 shows a warming starting around 1000 m or deeper. We suspect that the deep subsurface warming pattern, which also penetrates into the Southern Ocean, originates from a strengthened overturning circulation (due to the surface cooling) that brings more of the relatively warm upper ocean waters into the abyss. Finally, the Arctic basin (when shown separately) has a slight cooling in the upper 200m and a

805 larger deep warming signal below 500m in all three models. Ongoing research suggests
806 that this adjustment is related to ocean mixing biases.

807 As emphasized in Section 2.4, the comparison between CM4.0 and CM4X is not
808 clean, since there are many details across the models that differ. Even so, by exhibiting
809 the CM4.0 results in Figure 20, we are able to better gauge the improvements in CM4X-
810 p125 within the ocean interior. One further point of note concerns the hint of a centen-
811 nial time scale variations in the Southern Ocean, with such variations present in the up-
812 per 1000 m across the three models. These variations are also noticeable in the deeper
813 ocean, particularly in CM4.0. We return to such signals in Section 5 of Part II (Griffies
814 et al., 2024). In that discussion we note that this signal corresponds to Southern Ocean
815 polynya activity, with the CM4X signals far smaller and more realistic than the super-
816 polynyas found in CM4.0 and discussed by Held et al. (2019); Dunne et al. (2020); L. Zhang
817 et al. (2021).

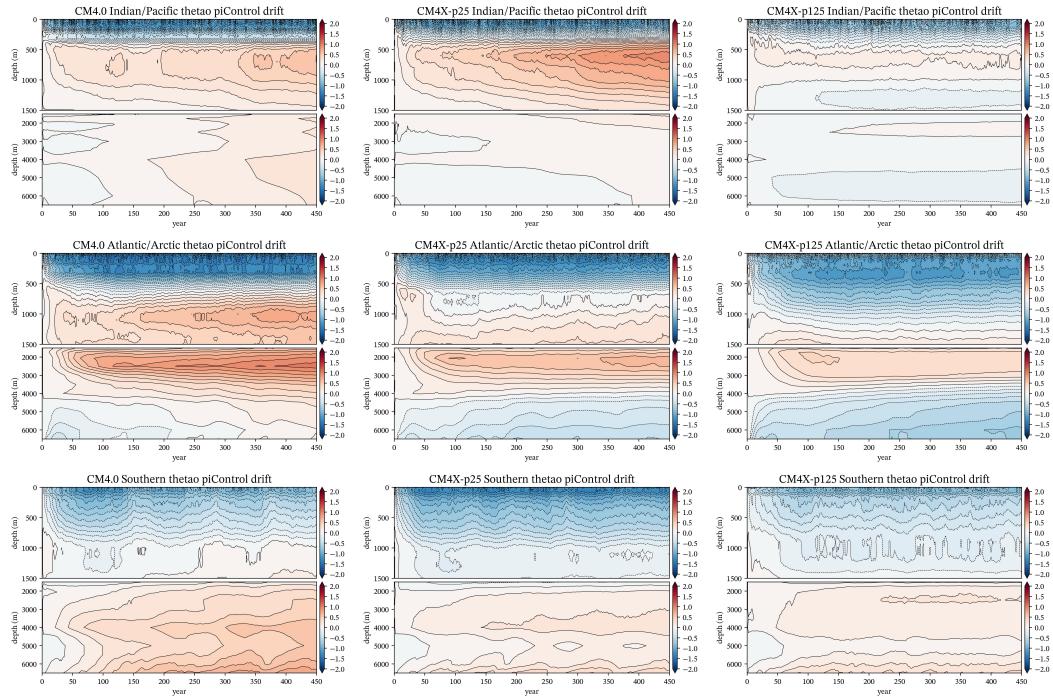


Figure 20. Depth-time adjustment for the horizontally averaged annual mean Conservative Temperature, as in Figure 18, but here decomposed into basins (Indian/Pacific, Atlantic/Arctic, and Southern) and showing CM4.0 (left column), CM4X-p25 (middle column), and CM4X-p125 (right column). Also note the expanded color range relative to Figure 18.

818 5.5 Depth-time salinity: World Ocean

819 The bulk of our study in this paper concerns temperature and heat. To check that
820 our focus is not missing something revealed by salinity, we display in Figure 21 the depth-
821 time adjustments for salinity, here including CM4.0 as well as the two CM4X models.
822 The near surface generally sees an increase in salinity during the first few years, as ex-
823 pected from the increases in sea ice due to the return to pre-industrial forcing in the pi-
824 Control. Below the surface salinity increase, and reaching down to a few hundred me-
825 ters, CM4X-p25 shows an initial freshening that gradually dissipates in favor of a grow-
826 ing salinification in the region between 500 m and 1000 m. In contrast, both CM4.0 and

CM4X-p125 show far less salinity increase in this depth range. In the deeper ocean in each model shows a gradual increase in salinity between 1500 m and 4000 m, and a muted freshening along with some salinification below.

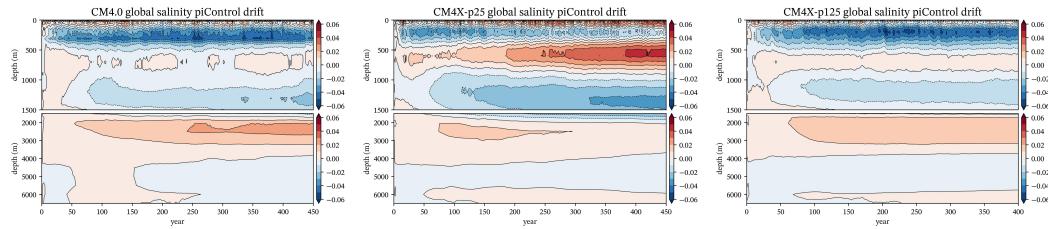


Figure 21. As in Figure 18, yet here for the salinity averaged over the World Ocean rather than temperature, and here including CM4.0 as well as CM4X-p25 and CM4X-p125. Units are g/kg or parts per thousand.

Fleshing out details for the salinity drift patterns requires a detailed analysis that is beyond our scope. The key conclusion we take is that CM4X-p125 shows noticeably less drift than either CM4.0 or CM4X-p25 after the initial 100-150 years. We thus find that the salinity drift patterns are consistent with those seen in temperature, further supporting the conclusion that there is a reduced drift in CM4X-p125 relative to CM4X-p25.

5.6 Global heat (enthalpy) budget

Global diagnostics of the piControl enthalpy budget are shown in Figure 22, with these budgets further illustrating how well CM4X-p125 has thermally equilibrated relative to CM4X-p25 and CM4.0. Thermal equilibrium for the ocean is realized by balancing the net heat flux crossing the ocean boundaries, both through a constant (in time) geothermal heat flux entering the ocean at the bottom and a net flux crossing the air-sea interface. The area averaged flux across the air-sea interface is directed into the ocean for CM4.0 and CM4X-p25, with CM4X-p25 having approximately half the flux as CM4.0. For CM4X-p125, the net air-sea heat flux is directed out of the ocean, and this outward flux closely balances the geothermal heat flux entering the ocean bottom. The result for CM4X-p125 is a global ocean enthalpy (heat content) tendency that is roughly 10 times smaller than either of the two net boundary fluxes.

The top-of-atmosphere heat flux, F_{TOA} , is relatively large in CM4.0, intermediate in CM4X-p25, and nearly zero in CM4X-p125. This result follows from the heat budgets for the ocean and the dominance of the ocean in establishing the global enthalpy balance. Note that the CM4.0 top-of-atmosphere heat flux is close to the CMIP6 median (Irving et al., 2021). The top-of-atmosphere fluxes are consistently higher than the ocean heat content tendencies by about $0.01\text{--}0.03 \text{ W m}^{-2}$. A mismatch between F_{TOA} and dQ/dt is a common feature among climate models and indicative of enthalpy leakage. Indeed, among CMIP6 models the mismatch can reach values of $0.5\text{--}1 \text{ W m}^{-2}$ (Irving et al., 2021).

Coupled climate models face a number of difficulties in conserving enthalpy, both through spurious dissipation, coupling inconsistencies, and missing processes (Lauritzen et al., 2022). In particular, we find in CM4.0 and CM4X a net loss of heat of about 0.17 W m^{-2} from the global climate system, attributable to the lack of temperature tracked by the atmospheric model's water causing a leak of energy through the oceanic precipitation and evaporation. The magnitude of this enthalpy sink is similar to other generations of GFDL models (e.g., see Section 3 in Delworth et al. (2006)). There are limitations of our

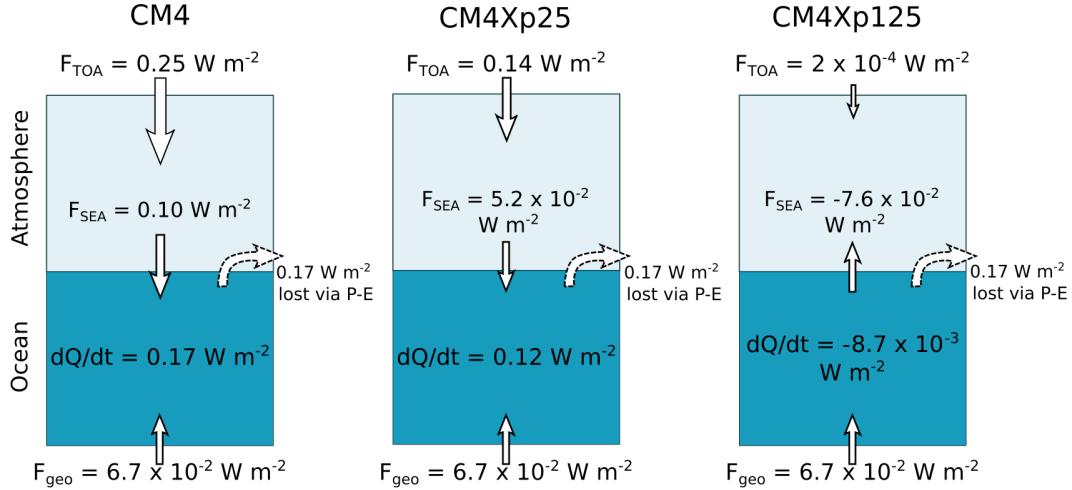


Figure 22. Schematics of the global heat budget for the piControl experiments from CM4.0 (left), CM4X-p25 (center), and CM4X-p125 (right), with all numbers obtained from time averaging over years 151-350, and all areas referring to the global earth surface area rather than ocean surface area. Schematics show the net top-of-atmosphere heat flux, F_{TOA} , air-sea heat flux, F_{SEA} , geothermal heat flux, F_{geo} , and loss of heat from the water cycle, F_{P-E} , all normalized by global surface area. The heat content tendency of the ocean, dQ/dt , is also normalized by global (rather than ocean) surface area. The sign notation is positive for fluxes into the Earth system or fluxes from the atmosphere into the ocean.

heat budget analysis, including the absence of terrestrial terms (including the terrestrial component of the enthalpy sink due to the hydrological cycle), nor have we considered changes in the heat balance from transience in glacier snow or soil. However, the residual of the three heat flux terms and the ocean heat content tendency is

$$F_{TOA} + F_{GEO} + F_{P-E} - dQ/dt \approx -0.07 \text{ W m}^{-2}, \quad (1)$$

which is consistent across all model setups. The similar differences across the models suggests that changes to the atmospheric and terrestrial model components play a secondary role in setting the enthalpy balance differences on climate timescales across the models. Instead, the results suggest that changes in the ocean model vertical enthalpy transport are dominant in setting the global response.

5.7 Ocean heat uptake efficiency and pycnocline depth

In Figure 23 we show the depth-time diagram for the horizontally averaged annual mean Conservative Temperature, now computed for the historical and SSP5-8.5 simulations and with differences computed relative to their respective piControl years 101-350. Warming in the upper ocean starts around year 2000 and penetrates into the deeper ocean throughout the 21st century. The heat uptake patterns are notably similar for the two models, which is consistent with the very close time series of global mean temperatures shown in the right panel of Figure 16.

To quantify the ocean heat uptake depicted in Figure 23, as well as the SST change in Figure 15, we compute the ocean heat uptake efficiency (OHUE). Following the definition in Newsom et al. (2023), we estimate OHUE as the change of global ocean heat uptake (in W m^{-2}) divided by the change of global mean SST (in Kelvin) as computed over years 2090-2100 of the SSP5-8.5 versus years 1850-1890 of the historical simulation.

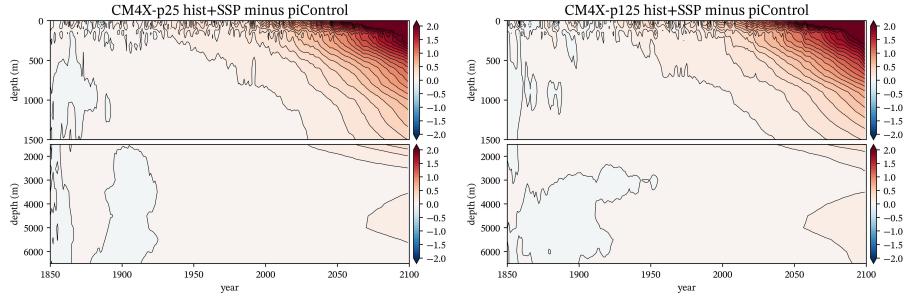


Figure 23. Depth-time drift for the horizontally averaged annual mean Conservative Temperature from CM4X-p25 (left panel) and CM4X-p125 (right). Here we show the historical and SSP5-8.5 simulations (years 1850–2100) relative to years 101–350 of the piControl; i.e., $\theta_{\text{anom}}(z, t) = \sum_{xy} \Delta V [\theta(x, y, z, t) - \theta(x, y, z, t)_{\text{piC}}] / \sum_{xy} \Delta V$, where $\theta(x, y, z, t)_{\text{piC}}$ is from the piControl. Note the split between the upper ocean and deeper ocean.

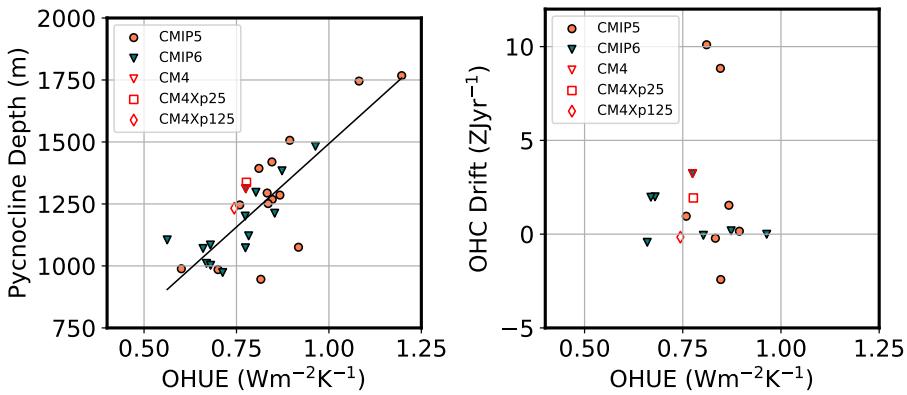


Figure 24. Left panel: ocean heat uptake efficiency (OHUE) versus pycnocline depth in CM4X models, with the values of CMIP5/6 models reproduced from Newsom et al. (2023) and with OHUE and pycnocline depth computed as per Newsom et al. (2023). The pycnocline depths are computed during the 80-year periods corresponding to branching of the historical simulations, so that the two CM4X pycnocline depths are based on piControl years 101–180 whereas the CM4.0 depth is from its piControl years 251–330. We find that the averaging period does not significantly alter results for the pycnocline depth. Right panel: Ocean heat uptake efficiency (OHUE) versus ocean heat content (OHC) drift in CM4X models, with the values of CMIP5/6 models reproduced from Newsom et al. (2023) and Irving et al. (2021). The OHUE is computed as per Newsom et al. (2023) and the OHC drift is computed as per Irving et al. (2021). The OHC drifts are computed as the linear trends of OHC change based on piControl years 151–350 (as in Figure 22). We find that the averaging period does not significantly alter results for the OHC drift. We also show only the subset of CMIP5/6 models analyzed by both Newsom et al. (2023) and Irving et al. (2021).

As seen in Figure 24, relative to the CMIP5/6 intermodel spread found in Newsom et al. (2023), the spread in OHUE is small among CM4X-p25, CM4X-p125, and CM4.0. Consistently, we find the difference in preindustrial pycnocline depth (averaged over model years 101–180) among the CM4X and CM4.0 models is also negligible compared to the CMIP5/6 intermodel spread (Figure 24). This result verifies that OHUE, as well as its

892 relationship with pycnocline depth across CM4X and CMIP5/6 models, is not sensitive
 893 to the simulated global ocean temperature drift and the degree of equilibration of the
 894 preindustrial control run (see right panel of Figure 24).

895 6 Pre-industrial thermal equilibration in a mesoscale dominant regime

896 Thermal equilibration of the piControl experiment is typically a long and resource
 897 intensive process, taking $\mathcal{O}(1000)$ simulation years. We thus ask whether the relatively
 898 rapid (centennial rather than millennial) thermal equilibration of CM4X-p125 is a one-
 899 off result? Or is it the signature of a new regime in climate modeling? Rather than defini-
 900 tive answers, we offer indirect evidence that such behavior is expected as models improve
 901 their physical and numerical fidelity. In so doing, we present some context for the ques-
 902 tion of piControl thermal equilibration and then present a hypothesis to help guide fur-
 903 ther investigations.

904 Following the Irving et al. (2021) study of climate model drift, we acknowledge the
 905 many reasons for piControl drift in the CMIP5/6 climate models. Even so, with CM4X-
 906 p25 and CM4X-p125 having distinct drift behaviors and yet differing solely in their ocean
 907 grid spacing, these two models offer a clean venue to study drift mechanics related to
 908 the differing representation of ocean heat transport.

909 6.1 Drift in the piControl: early climate models

910 During the early decades of climate modeling, it was necessary to augment mod-
 911 eled air-sea fluxes with *a priori* diagnosed and non-interactive fluxes aimed at reducing
 912 the otherwise extremely large model drift. The papers by Sausen et al. (1988) and Manabe
 913 and Stouffer (1988) detail these flux adjustment methods. Without flux adjustments, these
 914 early generation of climate models manifested a sizable drift that made it difficult, if not
 915 impossible, to study climate mechanics since the model drift aliased onto and/or directly
 916 impacted climate variability and change.

917 Climate model drift reduced as models improved their numerics and physics, no-
 918 tably through the representation of poleward ocean heat transport (Weaver & Hughes,
 919 1996). Such advances allowed models to realize sensible climate equilibria without flux
 920 adjustments (e.g., Boville and Gent (1998); Gordon et al. (2000); Delworth et al. (2006)).
 921 Eliminating flux adjustments was a milestone in the history of climate modeling. Even
 922 so, climate models still contain biases and limitations that are fully exposed when re-
 923 moving flux adjustments. So although the current generation of climate models has far
 924 smaller biases than the early generation, they still generally struggle to realize equilibrated
 925 climate states close to estimates for the 1850 climate, and they reach such equilibrated
 926 states only after $\mathcal{O}(1000)$ years (e.g., CM4X-p25, Stouffer (2004), Banks et al. (2006),
 927 and Irving et al. (2021)). For the CMIP5/6 climate models, the spread across climate
 928 model drift is the main cause for model-to-model differences as well as model-to-observation
 929 differences.

930 6.2 Drift in the piControl: state of the science climate models

931 Since the ocean provides the dominant heat reservoir for the Earth's climate sys-
 932 tem, we gauge the degree of climate model thermal equilibration by measuring trends
 933 in ocean temperature and/or the net boundary fluxes. We then claim "success" when
 934 the ocean trends are smaller than a subjective criteria. If computer cost (and energy foot-
 935 print) are of no concern, then we can reach thermal equilibrium by integrating the model
 936 for as long as it takes (Krasting et al., 2018; Rugenstein et al., 2020). Unfortunately, this
 937 brute force approach often leads to an equilibrated climate model with an ocean state
 938 that is far from a reasonable 1850 pre-industrial ocean. Indeed, in many cases, such as
 939 CM4X-p25, the volume mean ocean is warmer in the equilibrated 1850 piControl than

940 the present-day initial conditions. If there is one thing we know about the 1850 ocean,
 941 it is that it had a lower volume mean temperature than the 21st century ocean. We thus
 942 reach the conflicted situation whereby the more equilibrated a climate model, the fur-
 943 ther the ocean deviates from properties corresponding to the real ocean. This situation
 944 presents difficult choices for model development teams, and offers little inspiration for
 945 ocean analysts to study simulated water masses that are corrupted by huge drifts.

946 For a variety of practical reasons that include limitations on computational and hu-
 947 man resources, many models are simply not run to complete thermal equilibrium, so that
 948 the study of historical and future scenarios requires a form of drift removal, such as re-
 949 alized by subtracting the pre-industrial control simulation. Unfortunately, drift removal
 950 makes it difficult to study water mass properties and their relation to the observed ocean,
 951 and it does little to facilitate model-to-model comparisons of interior ocean properties
 952 since drifts across models can be quite different (Séférian et al., 2016; Irving et al., 2021).
 953 Indeed, some drift cannot be removed, such as discussed by Krasting et al. (2024) in the
 954 context of Southern Ocean mode and intermediate water masses and their imprint on
 955 regional sea level.

956 Even for presumably linear signals, there are caveats about the utility of drift re-
 957 moval. For example, Hallberg et al. (2013) questioned the linearity assumption for global
 958 thermosteric sea level projections since ocean density is a nonlinear function of temper-
 959 ature and salinity. As they found, extreme levels of model drift can have a notable role
 960 in sea level projections. A linearity assumption also becomes untenable when investigat-
 961 ing transitions between quasi-equilibrium states, such as hypothesized for the Atlantic
 962 overturning circulation. Indeed, the review paper by Hirschi et al. (2020) suggests that
 963 Atlantic overturning stability in models is dependent on grid spacing, with finer grid mod-
 964 els offering a more realistic depiction of the circulation pathways and, by inference, the
 965 stability properties. Studies of overturning stability thus motivate the use of both fine
 966 grid models and models that are equilibrated.

967 6.3 Time scale for piControl thermal equilibration

968 Consider a climate model with perfect dynamical core, perfect physical parame-
 969 terizations, and perfect numerics. If this model were subject to piControl forcings and
 970 initialized with perfect preindustrial initial conditions, it would presumably remain very
 971 close to its initial condition in perpetuity (ignoring that even 1850 has long-term climate
 972 drift due to centennial scale natural and anthropogenic variability). Instead, suppose this
 973 model is initialized with an exact rendition of present-day ocean temperature and salin-
 974 ity fields. What is the time scale for removing anthropogenic heat from the ocean in a
 975 piControl experiment? Since the bulk of the anthropogenic heat entered the ocean in the
 976 period after 1850, and since it is largely contained in the upper 1000 m of ocean, one hy-
 977 pothesis is that the time scale for its release is also $\mathcal{O}(100)$ years. This hypothesis is tem-
 978 pered by realizing that ocean heat uptake and heat release are not symmetric processes
 979 (Stouffer, 2004; Pudig et al., 2023). Furthermore, there is some heat sequestered in the
 980 deep ocean (Purkey & Johnson, 2010, 2012, 2013; Johnson & Purkey, 2024) through up-
 981 per ocean freshening and corresponding enhanced stratification, rather than the injec-
 982 tion of anthropogenic heat. It is connected to the efficiency by which heat can make its
 983 way to the abyssal ocean in regions not influenced by dense water formation, with that
 984 time affected by vertical mixing, either physically informed or spurious. Given these caveats,
 985 we find that the CM4X-p125 result, which equilibrates to a cooler piControl state than
 986 present-day, supports the hypothesis that the time scale is closer to $\mathcal{O}(100)$ years than
 987 $\mathcal{O}(1000)$ years.

988 During the post-1971 period, IPCC estimates (see Box 3.1 of Rhein et al. (2013))
 989 suggest that the ocean has accumulated roughly 3×10^{23} J = 300 ZJ of anthropogenic
 990 heat, and Zanna et al. (2019) suggest that it has accumulated roughly 400 ZJ since 1870.

991 A global ocean mean boundary heat flux, Q^{heat} , leads to an ocean volume mean temperature trend of roughly (see Appendices A and C of Griffies et al. (2014) for details of equation (2))
 992
 993

$$994 \Delta\Theta/\Delta t = Q^{\text{heat}}/(\rho_o C_p^0 H), \quad (2)$$

995 with $H \approx 3600$ m the ocean volume divided by the ocean surface area, $\rho_0 \approx 1035$ kg m⁻³
 996 the mean ocean density, and $C_p^0 \approx 3990$ J °C⁻¹ kg⁻¹ ocean heat capacity. As a useful
 997 benchmark, note that $Q^{\text{heat}} = 1$ W m⁻² corresponds to an ocean global volume mean
 998 temperature trend of $\Delta\Theta/\Delta t \approx 0.2^\circ\text{C century}^{-1}$. Figure 7 indicates that the CM4X-
 999 p125 simulation cools by roughly 0.07 K during the first 150 years of its piControl sim-
 1000 ulation, which represents an ocean heat release of

$$1001 V \rho_o C_p^0 \Delta\Theta \approx (1.3 \times 10^{18} \text{ m}^3) (1035 \text{ kg m}^{-3}) (4000 \text{ J kg}^{-1} \text{ K}^{-1}) (0.07 \text{ K}) \approx 390 \text{ ZJ}. \quad (3)$$

1002 This number accords with the 400 ZJ estimate from Zanna et al. (2019), in which case
 1003 we infer that the ocean heat released during the CM4X-p125 piControl directly corre-
 1004 sponds to the amount of anthropogenic heat built into the ocean initial conditions. In
 1005 contrast, after 1000 years of simulation, the CM4X-p25 simulation has nearly 1100 ZJ
 1006 *more* heat content (and 0.2 K warmer temperature) than the initial conditions.

1007 Now consider a model with exact present-day initial conditions but flawed dynam-
 1008 ical core and physical parameterizations. The model's piControl equilibrium state will
 1009 generally differ from the correct preindustrial state that has ≈ 400 ZJ less heat content
 1010 than present-day. Depending on model biases, it will tend towards an equilibrium state
 1011 that is warmer (as in CM4X-p25 and ESM2Mb; Krasting et al. (2018)) or even cooler
 1012 than the preindustrial state (as in ESM2G; Krasting et al. (2018)). This re-equilibration
 1013 process is likely to engage the whole ocean, including the deep ocean, and thus have a
 1014 time scale set by the slow diffusive mixing timescale $\tau_{eq} \approx H^2/\kappa_d \approx 5000$ yr (Krasting
 1015 et al., 2018) of unventilated abyssal shadow zones (Holzer et al., 2021).

1016 6.4 The mesoscale dominance hypothesis

1017 Given that CM4X-p25 and CM4X-p125 share all configuration details, except for
 1018 the horizontal ocean grid spacing, we hypothesize that the strength of the ocean mesoscale
 1019 transport accounts for the order of magnitude time scale difference for piControl ther-
 1020 mal equilibration for these two models. Furthermore, we hypothesize that even for mesoscale
 1021 active models, it is critical to have accurate levels of parameterized mixing as well as low
 1022 levels of spurious numerical mixing, thus ensuring that the deep ocean is not engaged
 1023 as part of an erroneous diabatic equilibration process.

1024 Mesoscale eddies and vertical heat transport

1025 As presented in Section 5.3, we invoke a mesoscale eddy mechanism to explain the
 1026 distinct timescales for ocean thermal adjustment in CM4X-p25 and CM4X-p125. Namely,
 1027 Griffies et al. (2015) (along with Gregory (2000), Gnanadesikan et al. (2005), Wolfe et
 1028 al. (2008), Gregory and Tailleux (2011), Delworth et al. (2012), Morrison et al. (2013),
 1029 Hieronymus and Nycander (2013), Zika et al. (2014), D. P. Marshall and Zanna (2014),
 1030 Doddridge et al. (2016), and von Storch et al. (2016)) identified the role of mesoscale ed-
 1031 dyes in regulating vertical heat transport between the upper ocean (roughly the upper
 1032 1000 m) and deeper ocean interior. This regulation is affected by the vertically upward
 1033 transport of positive buoyancy anomalies by eddies, and it holds whether the eddies are
 1034 explicitly represented, as in an eddy-admitting model, or parameterized via a scheme such
 1035 as Gent and McWilliams (1990) and Gent et al. (1995).

1036 For those regions without deep water formation (e.g., the Indian and Pacific basins),
 1037 much of the vertical transport of heat by the mean flow is compensated by an oppos-
 1038 ing heat transport by the mesoscale eddy flow, leaving a residual heat transport that is

1039 smaller than either the mean or eddy heat transport. In so doing, mesoscale eddies re-
 1040 duce the exchange of heat between the upper ocean and the deeper interior ocean; i.e.,
 1041 the upper and deeper oceans become disengaged. As a result, mesoscale eddies reduce
 1042 the role of deep ocean mixing in the process of the piControl thermal equilibration. We
 1043 can see this effect through the distinct temperature drifts found in the Indo-Pacific basin
 1044 (see Figure 20), in which there is only a negligible drift in the deep ocean of CM4X-p125
 1045 whereas the deep ocean in CM4X-p25 has a more sizable drift.

1046 In the intermediate and deep water formation regions of the high latitude Atlantic
 1047 and Southern Ocean, eddies also play a role in determining the exchange between the
 1048 surface and deep ocean, largely through the role of eddy restratification in convection
 1049 regions (J. Marshall & Schott, 1999). Focusing on the Southern Ocean, consider the con-
 1050 ceptual framework from D. P. Marshall and Zanna (2014) as a guide. Namely, their ide-
 1051 alized process model indicates that the bulk of the global ocean heat uptake occurs in
 1052 the Southern Ocean, with that heat uptake mediated by Ekman mechanics and with a
 1053 time scale set by mesoscale eddy processes. In contrast, they find that interior diapyc-
 1054 nal mixing plays a negligible role. To the degree that the mechanics of heat released
 1055 during the piControl is dominated by similar Southern Ocean wind and eddy mechan-
 1056 ics, we conjecture that Southern Ocean ventilation is key to setting the time scale for
 1057 piControl thermal equilibration. That is, we propose the D. P. Marshall and Zanna (2014)
 1058 mechanism as a null hypothesis describing the piControl thermal equilibration time scale.

1059 Restratification by submesoscale eddies in the upper ocean (i.e., mixed layer baro-
 1060 clinic instability), also affects the upper ocean ventilation (Boccaletti et al., 2007). In
 1061 particular, during the CM4X development we found the details of the Fox-Kemper et
 1062 al. (2008, 2011) submesoscale parameterization (see Section A14) to impact on the ver-
 1063 tical heat exchange and thus the thermal drift. Our chosen parameter settings are some-
 1064 what stronger than warranted by process studies. The settings were chosen in develop-
 1065 ing CM4X-p25, with the stronger restratification set to compensate for the somewhat
 1066 weaker restratification from the relatively weak mesoscale eddies. Determining the proper
 1067 interplay between the mesoscale and submesoscale restratification effects remains a topic
 1068 of ongoing research.

1069 *The need to reduce spurious mixing*

1070 The study of Adcroft et al. (2019) showed that thermal drift in an ocean model forced
 1071 by a prescribed atmosphere is directly related to the vertical coordinate, with the hy-
 1072 brid z^*/ρ_{2000} coordinate used in CM4.0 (and in CM4X) leading to significantly less drift
 1073 than z^* used throughout the ocean. These results suggest there are significantly larger
 1074 levels of spurious numerical mixing with the z^* coordinate than the hybrid z^*/ρ_{2000} co-
 1075 ordinate. They are also consistent with the study of Ilicak et al. (2012) who diagnosed,
 1076 using potential energy methods, large spurious mixing in CM2.5 (which uses z^* through-
 1077 out the full ocean depth). The association of a biased warm ocean arising from spuri-
 1078 ous mixing is supported by Krasting et al. (2018) and Hieronymus et al. (2019), with both
 1079 papers using the GFDL-CM2G climate model (which uses an isopycnal coordinate ocean
 1080 component) and finding that the volume-averaged ocean temperature increases with back-
 1081 ground diapycnal diffusivities.

1082 *Summarizing the mesoscale dominance hypothesis*

1083 We hypothesize that there are three ocean model properties necessary to support
 1084 a centennial rather than the millennial time scale for thermal equilibration into an ocean
 1085 that is cooler (with roughly 400 ZJ less heat content than early 21st century) in its 1850
 1086 piControl state: (I) enhanced fidelity of mesoscale features, including transient eddies
 1087 and boundary currents; (II) accurate strength and geography of parameterized numer-
 1088 ical mixing processes; and (III) negligible levels of spurious mixing from numerical dis-

1089 cretization. We refer to ocean models that possess these three properties as *mesoscale*
 1090 *dominant models*. Mesoscale dominant models contrast to those where deep ocean di-
 1091 abatic processes (either parameterized or spurious numerically induced) play a promi-
 1092 nent (and sometimes dominant) role in piControl thermal equilibration. We infer that
 1093 models that are not mesoscale dominant engage their deep ocean circulation during the
 1094 1850 piControl, thus rendering far longer thermal equilibration time scales. These long
 1095 thermal spin-ups also affect long spin-up times for biogeochemical cycles, though bio-
 1096 geochemical spin-ups are also impacted by other processes (Orr et al., 2017; Khatiwala,
 1097 2023, 2024).

1098 CM4X-p125 does not perfectly realize each of the three properties of a mesoscale
 1099 dominant model. In particular, both of the CM4X models have biases in the Southern
 1100 Ocean ventilation, as noted by the overly deep mixed layers in Figure 9 and as further
 1101 explored for CM4.0 in Krasting et al. (2024). The case studies in Part II (Griffies et al.,
 1102 2024) point to further shortcomings that form the topic of ongoing research and devel-
 1103 opment. Even so, we propose that CM4X-p125 approaches mesoscale dominance more
 1104 than CM4X-p25, given the stronger mesoscale eddy activity in CM4X-p125.

1105 6.5 Water mass transformation analysis

1106 As an initial plausibility test of the mesoscale dominance hypothesis, we employ
 1107 water mass transformation budgets (following Drake et al. (2024) and Section 7.4 in Part
 1108 II of Griffies et al. (2024)) to evaluate the extent to which (A) water masses are well-
 1109 equilibrated in CM4X piControl simulations and (B) spurious numerical mixing plays
 1110 a minor role in the water mass budgets. For a water mass $\mathcal{M}_\geq(\sigma_2, t)$ defined as the mass
 1111 of water that is denser than σ_2 at time t , the water mass transformation framework al-
 1112 lows us to attribute the drift, $\partial_t \mathcal{M}_\geq$, to processes that drive it.

1113 Figure 25 shows the piControl drift of dense water masses in CM4X-p25 and CM4X-
 1114 p125, quantified by the normalized water mass change (relative to the initial condition),

$$1115 \Delta \widetilde{\mathcal{M}}_\geq \equiv \widetilde{\mathcal{M}}_\geq(\sigma_2, t) - \widetilde{\mathcal{M}}_\geq(\sigma_2, t = 0) \equiv \frac{\mathcal{M}_\geq(\sigma_2, t)}{\mathcal{M}} - \frac{\mathcal{M}(\sigma_2, t = 0)}{\mathcal{M}}, \quad (4)$$

1116 where $\mathcal{M} \equiv \mathcal{M}_\geq(\min(\sigma_2), t)$ is the total mass of the ocean and tildes denote we have
 1117 normalized the water mass by the total ocean mass. The normalized water mass, $\widetilde{\mathcal{M}}_\geq(\sigma_2, t)$,
 1118 increases monotonically from 0 for the densest water in the ocean to 1 for the lightest
 1119 waters and thus can be thought of as the ocean's cumulative probability density func-
 1120 tion in potential density space (see Figure 26d). $\Delta \widetilde{\mathcal{M}}_\geq$ quantifies how much a given wa-
 1121 ter mass has drifted from the initial condition over time, expressed as a fraction of the
 1122 total ocean mass.

1123 Figure 25a shows that deep water masses ($\sigma_2 > 36 \text{ kg/m}^3$) drift by about -1% per
 1124 century in CM4X-p25 (Figure 25a); because total ocean mass is conserved, this nega-
 1125 tive drift signifies a drift towards lighter densities (consistent with the warming described
 1126 in Section 4). By contrast, water masses that include lighter surface and thermocline wa-
 1127 ters (i.e., $\sigma_2 \leq 36.0 \text{ kg/m}^3$) show few changes after the first 50 years, suggesting that
 1128 they equilibrate more quickly. In CM4X-p125, the deep water masses exhibit smaller wa-
 1129 ter mass changes, with near zero drift after the first 100 years (Figure 25b). The largest
 1130 water mass changes are seen in the Antarctic Bottom Water density class ($\sigma_2 \geq 37 \text{ kg/m}^3$).

1131 To put these piControl water mass drifts in context, we compare them to the anomalous
 1132 water mass changes in the forced historical plus SSP5-8.5 experiments. The forced
 1133 water mass changes are fairly similar in both magnitude, density structure, and timing
 1134 between CM4X-p25 and CM4X-p125 (Figures 25c,d). These forced water mass changes
 1135 are larger and occur more quickly than the piControl water mass changes. In CM4X-
 1136 p125, the forced water mass drifts, $\partial_t \mathcal{M}$, at the end of the SSP5-8.5 experiment are or-
 1137 ders of magnitude larger than the piControl drifts across all density classes. These re-

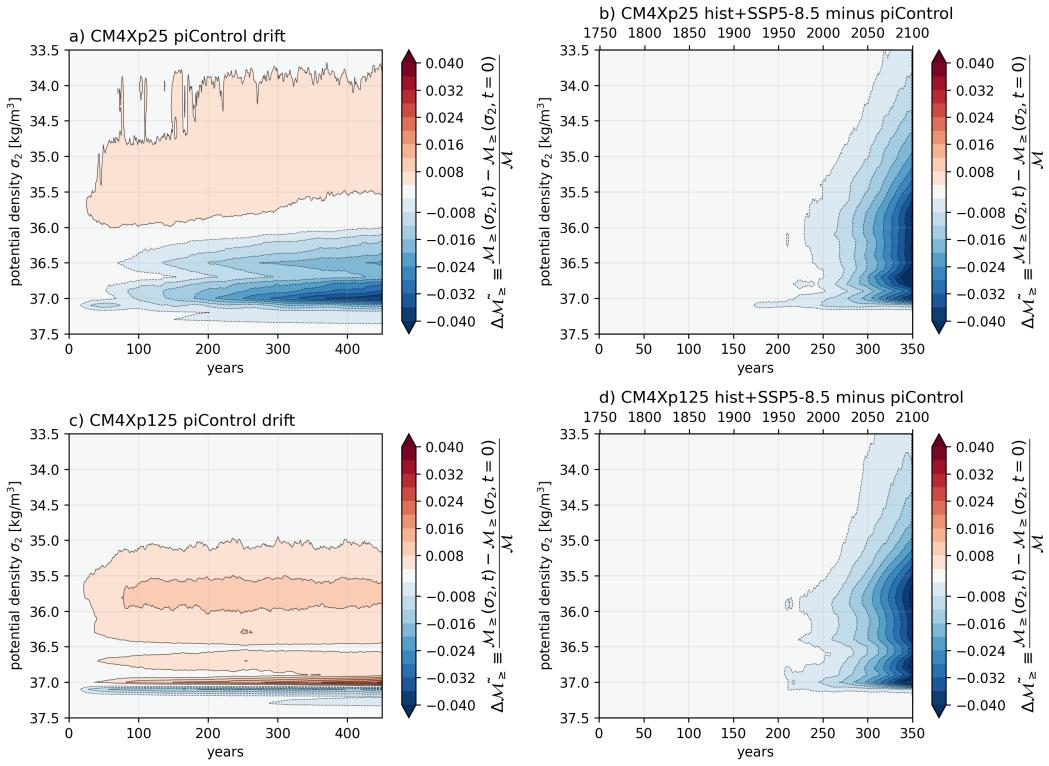


Figure 25. Normalized water mass changes $\Delta\tilde{\mathcal{M}}_{\geq}$ (equation 4, where $\mathcal{M}_{\geq}(\sigma_2, t)$ is the mass of water denser than σ_2 at time t) in CM4X-p25 (a,b) and CM4X-p125 (c,d), expressed as a fraction of the total ocean mass. Panels (a,c) show the changes relative to the initial condition in the piControl experiment. Panels (b,d) show the changes in the forced hist+SSP5-8.5 simulations, relative to the piControl changes. Time is expressed in years after initialization below and historical years above.

sults demonstrate that water masses are indeed well-equilibrated in the CM4X piControl simulations.

To understand how the water mass drifts, $\partial_t \mathcal{M}_{\geq}$, can be so small in the piControl simulations, we now turn to the water mass transformation budget:

$$\partial_t \mathcal{M}_{\geq} = \mathcal{S}_{\geq} + \mathcal{G}, \quad (5)$$

where \mathcal{G} is the total water mass transformation rate (positive when it increases the mass of denser water), which we decompose into the contributions from three processes: boundary buoyancy fluxes ($\mathcal{G}^{(BF)}$; including air-sea, ice-ocean, and geothermal fluxes), parameterized mixing ($\mathcal{G}^{(Mix)}$; representing surface mixed-layer, interior, and bottom boundary layer mixing processes), and spurious numerical mixing ($\mathcal{G}^{(Spur)}$; diagnosed as the residual of the other terms). \mathcal{S}_{\geq} is the water mass tendency due to direct exchange across the ocean's boundary (e.g., evaporation minus precipitation plus rivers and meltwater) and is typically negligible.

Figures 26a,b reveal that water masses of all densities exist in an approximate balance between transformation by boundary fluxes and parameterized mixing. At zeroth order, the formation of dense deep waters ($35 \text{ kg/m}^3 \leq \sigma_2 \leq 36 \text{ kg/m}^3$) and even denser bottom waters ($36.75 \text{ kg/m}^3 \leq \sigma_2 \leq 37.5 \text{ kg/m}^3$) at a rate of 40 Sv at high latitudes is mostly balanced by parameterized mixing-driven transformation. Spurious numeri-

1156 cal transformations due to discretization errors in the advection or Lagrangian remapping
 1157 scheme (see Drake et al. (2024)), however, further modify water masses at rates of
 1158 $\mathcal{O}(5 \text{ Sv})$.

1159 After 440 years of spin-up in the piControl simulations, the water mass transfor-
 1160 mations described above are nearly in balance, with residual water drifts of less than 4 Sv
 1161 in all density classes when averaged over 10 years (Figure 26c). Furthermore, the unbal-
 1162 anced water mass drift seems to be smaller in CM4X-p125 than in CM4X-p25, especially
 1163 at the densities that dominate the global water mass distribution (Figure 26d), provid-
 1164 ing further evidence in support of the mesoscale dominance hypothesis. While analogous
 1165 mass budgets for the CM2.5/CM2.6 hierarchy are not available, other models that use
 1166 the same ocean model code (MOM5) exhibit spurious water mass transformations in ex-
 1167 cess of those driven by parameterized mixing rates (Ilicak et al., 2012; Holmes et al., 2021).
 1168 The relative smallness of spurious numerical mixing in the CM4X water mass budget sup-
 1169 ports our hypothesis that the transition to a hybrid $z^* - \rho_2$ coordinate in CM4X (MOM6
 1170 code) enables the mesoscale dominant regime to emerge.

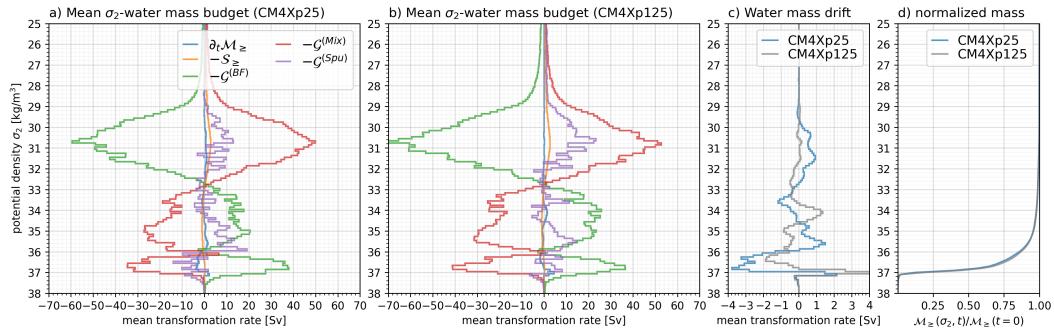


Figure 26. Global water mass budgets (see Section 6 in Part II of Griffies et al. (2024) for details) in potential density (σ_2 ; potential density referenced to 2000 dbar, minus 1000 kg m^{-3}) coordinates for CM4Xp25 (a) and CM4Xp125 (b), averaged over years 441-451 of the piControl. (c) Comparison of the water mass drift $\partial_t \mathcal{M}_\geq$ terms in the CM4X configurations. (d) The normalized water mass distribution, $\tilde{\mathcal{M}}_\geq = \mathcal{M}_\geq(\sigma_2, t)/\mathcal{M}$, defined by the mass of water $\mathcal{M}_\geq(\sigma_2, t)$ that is denser than a given potential density σ_2 at time t divided by the total ocean mass \mathcal{M} . Panel (d) shows that about 85% of the global ocean mass is occupied by deep waters within a narrow potential density range between 36 kg/m^3 and $\sigma_2 = 37.1 \text{ kg/m}^3$.

1171 7 Concluding remarks

1172 As further explored in Part II (Griffies et al., 2024), refined ocean grid spacing im-
 1173 proves many aspects of the ocean and climate system that makes CM4X-p125 more suited
 1174 to a variety of applications. Furthermore, those cases where CM4X-p25 and CM4X-p125
 1175 are quite similar, such as global mean transient climate response and ocean heat uptake
 1176 efficiency, confirm decades of work done with comparatively coarse resolution models.
 1177 That is, for models with relatively modest drift such as CM4X-p25, we can garner a sen-
 1178 sible estimate for many bulk climate response diagnostics as compared to models with
 1179 extremely low drifts such as CM4X-p125. Even so, there are aspects where payoff for the
 1180 enhanced grid resolution is nontrivial, such as the roughly 20% reduction in sea surface
 1181 temperature biases in CM4X-p125, and the order of magnitude reduction in thermal equi-
 1182 libration time for the CM4X-p125 piControl experiment.

Based on prior experience with CM2.5/CM2.6 (Griffies et al., 2015), we expected the more active eddying ocean in CM4X-p125 to play an important role in establishing thermal equilibrium. Even so, we were pleasantly surprised when CM4X-p125 equilibrated to a cooler (by volume mean) ocean after only $\mathcal{O}(100)$ years. Furthermore, the cooler ocean state has roughly 400 ZJ less ocean heat content than the present-day initial conditions, in accord with observational estimates from Zanna et al. (2019). For global-scale studies of future climate, the very small drift in CM4X-p125 allows for a clear connection between the physical ocean response, ocean interior focused model diagnostics such as ocean heat uptake efficiency, and surface/atmosphere diagnostics like the climate feedback parameter or the SST pattern effect (Armour et al., 2013; Andrews et al., 2018). Note, however, that there remain some regional temperature drift within CM4X-p125 that might necessitate drift removal to reveal the role of specific regions in setting future climate.

It is possible that the piControl thermal equilibration found in CM4X-p125 is a one-off fortuitous result, and we have incomplete evidence to argue against that perspective. Even so, evidence based on prior studies and further analysis presented in this paper support our conjecture that CM4X-p125 exemplifies what can be expected with higher fidelity ocean climate models used within a realistic coupled climate model. There certainly is far more to climate model drift than just the ocean physics and numerics. Yet the ocean plays a huge role in the mechanics of climate model drift, and with the CM4X hierarchy proving a fit-for-purpose tool for investigating that role. In this closing section we offer a few remarks about the need to further develop a mechanical/process understanding of piControl thermal equilibration, and then comment on prospects for climate modeling with a mesoscale dominant ocean.

7.1 Mechanisms for piControl thermal equilibration

The piControl drift has traditionally been seen as a necessary, but inconvenient and uninteresting, part of the climate modeling art. We instead suggest that it provides signatures for climate model fidelity, and as such it should be studied (e.g., Banks et al. (2006); Krasting et al. (2024)). Correspondingly, we propose that developing an understanding of piControl drift, whether it be related to spurious numerical processes or a variety of physical processes, is a central task needed to advance climate model integrity and utility. Even if mesoscale dominant models prove difficult to reproduce, the mesoscale dominance hypothesis offers a conceptual framing for thermal drift in piControl simulations. Such framing could, we propose, support the movement towards standard protocols for piControl simulations that allow for a meaningful comparison across models.

Much work is needed to move beyond the descriptive and diagnostic presentation in this paper. One approach could be to map the results from the CM4X suite onto the D. P. Marshall and Zanna (2014) idealized model, thus exploring the genesis and evolution of the water mass adjustment according to a variety of physical ocean processes, with a particular focus on mesoscale eddy and wind effects in the Southern Ocean. Another is to emulate Banks et al. (2006), who explored the piControl drift in their coupled model, pursuing an analysis from deep water formation regions to the top of the atmosphere imbalances. A further approach concerns the temperature-space water mass analysis of Holmes et al. (2019, 2020, 2021) and Deppenmeier et al. (2021), who performed heat budgets within regions bounded by isotherms, thus enabling a focus on diabatic processes associated with the movement of heat and removing the large, and less relevant, signal due to adiabatic processes. We note that water mass transformations are generally simpler to compute and to interpret in temperature space, rather than the more common density space considered in Section 6.5, largely since subtleties of the nonlinear equation of state of seawater are circumvented with a temperature approach.

1233 **7.2 Prospects for eddying climate models**

1234 The common experience of $\mathcal{O}(1000)$ years for piControl thermal equilibration is baked
 1235 into community strategies for climate modeling. Consequently, there is a presumption
 1236 that the explicit representation of mesoscale eddies and associated fine-scale boundary
 1237 currents will long remain computationally out of reach for routine studies of climate dy-
 1238 namics. However, if our experience with CM4X-p125 is reproducible (a big assumption
 1239 at this point in the science), then moving into a mesoscale dominant regime will reduce
 1240 piControl thermal equilibration timescales by roughly one order of magnitude. With mesoscale
 1241 dominant models, the added computational energy and wall-clock time needed for re-
 1242 fined grid spacing are compensated by a significantly reduced thermal equilibration time,
 1243 thus bringing forth the prospect of more common use of a mesoscale active ocean as part
 1244 of realistic climate models.

1245 **Appendix A CM4X model configuration details**

1246 The ocean and sea ice components of CM4X are based on the MOM6 ocean code
 1247 and SIS2 sea ice code as configured in the OM4.0 model detailed in Adcroft et al. (2019).
 1248 In this appendix we summarize those relatively few features of the CM4X ocean that differ
 1249 from OM4.0. We also summarize elements of the atmosphere and land models that
 1250 differ from the CM4.0 configuration of Held et al. (2019).

1251 **A1 Atmosphere model**

1252 As noted in Section 2.2, CM4X makes use of the C192 (nominally 50 km grid spacing)
 1253 version of the AM4 cubed sphere atmosphere model of Zhao (2020). This choice con-
 1254 trasts to the C96 (nominally 100 km grid) version (Zhao et al., 2018a, 2018b) used in
 1255 CM4.0. Our choice to use the C192 is based on the desire to simulate more energetic at-
 1256 mospheric cyclones that are key to the study of sea level extremes, such as those stud-
 1257 ied by Yin et al. (2020). In particular, Zhao (2020, 2022a) showed that the 50 km res-
 1258 olution AM4 (i.e., C192AM4) reasonably simulates the frequency of atmospheric rivers,
 1259 tropical storms, and mesoscale convective systems, as well as their associated precipi-
 1260 tation and extreme precipitation.

1261 In ocean-atmosphere coupled climate models, atmospheric radiative properties and
 1262 their responses to changes in radiative gases, aerosol emissions, and ocean surface tem-
 1263 peratures have a substantial impact on the simulation of SST and ocean volume mean
 1264 temperature. The top of atmosphere (TOA) radiative fluxes are heavily influenced by
 1265 cloud parameterizations, which often involve poorly constrained parameters. We set the
 1266 CM4X cloud parameters using the atmospheric model forced by observed SSTs and sea-
 1267 ice concentrations (SICs), commonly referred to as AMIP (Atmospheric Model Intercom-
 1268 parison Project) simulations. In Table A1 we summarize differences in atmospheric physics
 1269 parameter settings across various GFDL models. Other atmospheric model parameter
 1270 settings (i.e., time-step and dynamical core) in the CM4X atmosphere are identical to
 1271 those in the C192 atmosphere developed by Zhao (2020).

1272 After setting the atmospheric physics parameters, the net TOA radiative flux is
 1273 an emergent property of the simulation. When averaged over the period of 2006-2018,
 1274 the TOA radiative flux is 1.42 W m^{-2} in CM4X AMIP simulations. This value is higher
 1275 than the range of 0.52 W m^{-2} to 1.06 W m^{-2} estimated over 2006-2018, as cited in Ta-
 1276 ble 7.1 of Forster et al. (2021). The higher value was a target of our parameter settings,
 1277 aiming to mitigate the global SST cool biases shown in Figure 14. Even so, this setting
 1278 often comes at the expense of excessive heat uptake in the 1850 control experiment, de-
 1279 pending on ocean processes such as mesoscale eddy transport and parameterized sub-
 1280 mesoscale eddy mixing. This higher TOA radiative balance was also pursued when de-
 1281 veloping CM4.0.

MODEL	vfact	eros_scale_c	eros_scale_t	citation
CM4X	0.875	5e-5	5.75e-5	this paper
C192AM4	0.8	8e-5	8e-5	Zhao (2020)
CM4.0	0.9	4e-5	5e-5	Held et al. (2019)
SPEAR	0.9	4e-5	5e-5	Delworth et al. (2020)
ESM4.1	0.9	5e-5	5e-5	Dunne et al. (2020)

Table A1. Summary of the differences in atmospheric physics parameter setting for various GFDL models. All differences in parameter settings relate to cloud settings. `vfact` is used to set the ice fall velocity, which affects the precipitation efficiency of cold clouds. `eros_scale_c` and `eros_scale_t` are used to set the dissipation time scales of clouds over in the convective and turbulent regimes respectively.

To illustrate the differences in global TOA radiative properties between the atmospheric models used in CM4X and CM4.0, Table A2 provides a comparison of global TOA radiative fluxes simulated by the two models. Both models are forced by the same monthly varying climatological SSTs and SICs averaged over years 1981-2014, with radiative gases and aerosol emissions fixed at present-day (PD, 2010) levels (referred to as 2010Control). The CM4X atmosphere produces roughly 1 W m^{-2} higher net longwave and shortwave radiative fluxes, with the net TOA downward flux being slightly larger. This result indicates that the low drift in CM4X as compared to CM4.0 is not due to differences in their atmospheric models.

Table A2 also provides the total radiative flux perturbation (RFP) derived from the two atmospheric models based on their corresponding present-day (PD, 2010) and preindustrial (PI, 1850) simulations, the aerosol RFP based on PD and PI aerosol emissions, as well as the idealized CESS feedback (Cess et al., 1990) derived from the idealized 2K uniform SST warming and their corresponding 2010Control simulations. These results suggest a broad similarity between the two atmospheric models, with the CM4X atmosphere showing a reduction of aerosol forcing. The reduction of aerosol forcing with increased atmospheric model resolution (C192 versus C96) was discussed in Zhao et al. (2018b, 2018a).

A2 Land model

For CM4X we prescribe a time-invariant (i.e., “static”) map of vegetation that incorporates both natural vegetation and land-use conditions representative of year 1980. Unlike CM4.0, which used dynamic vegetation (i.e., interactive, prognostic), the same static vegetation boundary condition is used in all CM4X simulations. Prescribing the land surface condition removes feedbacks that occur in the Earth system, but also helps to better examine how ocean changes impact the climate within the CM4X configuration. Zhao (2022a) studies climate sensitivity across a suite of GFDL climate models and considers the impact of static versus dynamic vegetation. Additionally, eight of the 11 CMIP6 models analyzed by Arora et al. (2019) use static vegetation (see their Table 2).

A3 Albedos for glaciers

We explored a series of land ice albedo settings motivated by the studies of Delworth et al. (2020), Dunne et al. (2020), and L. Zhang et al. (2021), each of whom identified the importance of near-infrared glacial snow albedos on both the production of Antarctic Bottom Water and the occurrence of intermittent Southern Ocean convection. Our tests confirmed their results, and we chose the brighter set of albedos used in the GFDL-

Global TOA radiative fluxes (W m^{-2})	CM4.0	CM4X
OLR (net LW)	238.44	239.51
SWABS (net SW)	240.16	241.39
NETRAD (net LW+SW)	1.72	1.88
Total RFP between 2010 and 1850	2.71	2.86
Aerosol RFP between 2010 and 1850	-0.76	-0.65
CESS feedback between 2K uniform SST increase & 2010Control	-3.53	-3.54

Table A2. A comparison of global TOA radiative fluxes (W m^{-2}) simulated by the atmospheric models used in CM4X and CM4.0, with both models forced by the same monthly varying climatological SSTs and sea-ice concentrations averaged from 1981–2014, and radiative gases and aerosol emissions fixed at present-day (PD, 2010) levels (referred to as 2010Control). OLR = TOA outgoing longwave radiation; SWABS = TOA net SW radiative flux (absorption, downward positive); NETRAD = TOA net radiative flux (downward positive, i.e., SWABS minus OLR). The table also lists the total radiative flux perturbation (RFP) derived from each model based on present-day (PD, 2010) and preindustrial (PI, 1850) conditions, the aerosol RFP based on PD and PI aerosol emissions, and the idealized CESS feedback (Cess et al., 1990), which is derived based on the idealized 2K uniform SST warming and their corresponding 2010Control simulations. Each climatological simulation was integrated for 30 years.

1316 ESM4.1 configuration of Dunne et al. (2020) (see Table A3). These near-infrared glacial
 1317 snow albedos lie at the upper limit of the observational range (Dunne et al., 2020).

MODEL	F_ISO_WARM_ON_GLACIER	F_ISO_COLD_ON_GLACIER	citation
CM4X	0.77, 0.58	0.92, 0.73	this paper
CM4.0	0.77, 0.43	0.92, 0.68	Held et al. (2019)
ESM4.1	0.77, 0.58	0.92, 0.73	Dunne et al. (2020)
SPEAR	0.77, 0.43	0.92, 0.68	Delworth et al. (2020)

Table A3. Summary of the land glacier albedo settings for a variety of GFDL models, listing the values for (Visible, Near IR) radiation and warm/cold surfaces.

1318 Note that Dunne et al. (2020) states that “we had to increase the near-infrared albedo
 1319 of snow on glaciers to a value of 0.82.” This value of 0.82 is different than the values in
 1320 Table A3 because it refers to the total ratio of shortwave up to shortwave down, which
 1321 is an emergent property of the simulation. Thus, it would be more appropriate for Dunne
 1322 et al. (2020) to say “we found that we had to increase the near-infrared albedo of snow
 1323 on glaciers to obtain an overall shortwave albedo value of 0.82 for these regions...” (John
 1324 Dunne, personal communication).

1325 In Figure A1 we show the time series for the global volume mean ocean temper-
 1326 ature for CM4.0 and CM4X-p25, along with a third configuration with CM4.0 using the
 1327 albedo of snow on glaciers used by CM4X. The nearly parallel behavior of CM4X-p25
 1328 and CM4.0-albedo indicates that the snow on glacier albedo is the key difference, from
 1329 a global enthalpy budget perspective, between CM4.0 and CM4X-p25.

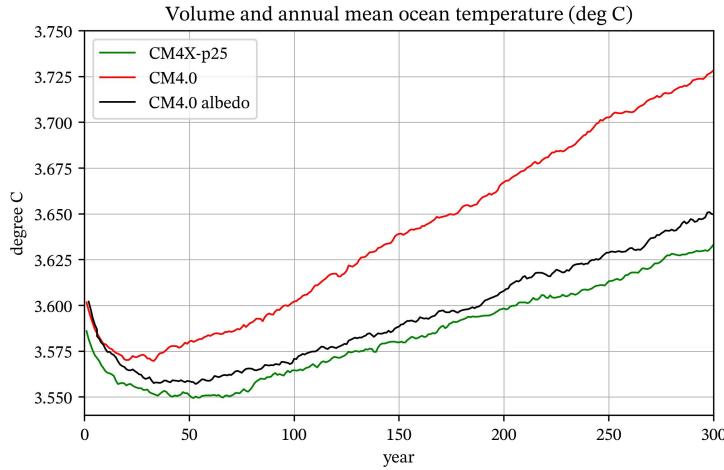


Figure A1. Time series of the global volume mean ocean temperature from CM4.0, CM4.0-albedo, and CM4X-p25. The CM4.0-albedo experiment is identical to CM4.0, with the one exception being the use of the CM4X snow on glacier albedo as per Table A3. The nearly parallel behavior of CM4X-p25 and CM4.0-albedo indicates that the snow on glacier albedo is the key difference, from a global energy budget perspective, between CM4.0 and CM4X-p25.

1330

A4 Albedos for sea ice

1331
1332

In testing the sea ice albedos for the CM4X configuration, we started with those used by Delworth et al. (2020) yet found these to suffer from the following issues:

1333
1334
1335
1336

- Arctic sea ice that was too thin and biased low in summer;
- Antarctic sea ice that was biased low in summer;
- Systematic model drift towards a weakening Antarctic circumpolar current and weakening Antarctic Bottom Water (AABW) transport.

1337
1338
1339
1340
1341
1342

These biases motivated returning the sea ice albedo parameters to the brighter values used in CM4.0 (Held et al., 2019). Doing so increased the sea ice thickness and summer sea ice extent. The change was also motivated by the expectation that cooling the Southern Ocean surface would encourage intermittent oceanic convection and production of Antarctic bottom waters (Delworth et al., 2020; L. Zhang et al., 2021). Table A4 lists the settings for CM4X and compares to those used in other GFDL models.

MODEL	ICE_DELTA_EDD_R_ICE	ICE_DELTA_EDD_R_SNOW	ICE_DELTA_EDD_R_POND
CM4X	1.0	1.0	1.0
CM4.0	1.0	1.0	1.0
ESM4.1	1.5	1.5	1.5
SPEAR	0.0	0.0	0.0

Table A4. Summary of the sea ice albedo settings for a variety of GFDL models. Albedos are in units of observed standard deviations relative to observed mean values from Briegleb and Light (2007).

1343 **A5 Fixing a sea ice bug for Arctic bipolar coordinate fold**

1344 A numerical error was identified in the sea ice along the bipolar grid fold in the Arctic
 1345 Ocean, with this bug present in the OM4.0 and CM4.0 models. This error occurs along
 1346 the half of the bipolar fold that extends from the geographic North Pole to the grid pole
 1347 located in Russia. The error is most visible in zonal and meridional sea ice velocity fields,
 1348 but it also creates artifacts in sea ice thickness and concentration fields, as well as the
 1349 upper ocean dynamical fields. Tests in CM4X-p25 and CM4X-p125 reveal that the bug
 1350 has minimal impact on the global climate. Even so, we corrected the bug for the CM4X
 1351 simulations shown in this paper.

1352 **A6 Horizontal grids for the ocean and sea ice**

1353 For the ocean and sea ice components, we make use of a locally orthogonal coor-
 1354 dinate grid with a nominal grid spacing of 0.125° along the equator for CM4X-p125 and
 1355 0.25° for CM4X-p25. As detailed in Section 2.1.2 of Adcroft et al. (2019), the grid con-
 1356 sists of a Mercator layout between latitudes given in Table A5, which transitions to a
 1357 bipolar northern cap using the methods of Murray (1996) with coordinate singularities
 1358 over Siberia and Canada. In the far south (south of 66.8°S), the grid transitions to a reg-
 1359 ular latitude/longitude grid, and it is capped by a displaced pole south of 78.0°S (see
 1360 Figure 1 of Adcroft et al. (2019)). Use of a southern displaced pole is motivated to op-
 1361 timize the lateral grid spacing in the southernmost parts of domain.

Grid property	CM4X-p25	CM4X-p125
nominal (equatorial) grid spacing	0.25°	0.125°
grid cells ($n_i \times n_j$)	1440×1080	2880×2240
southern edge of Mercator grid	66.8595472°S	$66.8840965847^\circ\text{S}$
northern edge of Mercator grid	64.0589597°N	$64.0316059408^\circ\text{N}$
longitudes of bi-polar singularities	$120^\circ\text{W}, 60^\circ\text{E}$	$120^\circ\text{W}, 60^\circ\text{E}$
southern edge of south spherical grid	78°S	78°S

1362 **Table A5.** Details for the horizontal grids used by the two CM4X ocean and sea ice compo-
 1363 nents. *Resolution* in this paper refers to the number of grid degrees of freedom and as such it is
 1364 an integer number, whereas *grid spacing* refers to the distance between grid cells as measured in
 1365 units of degrees or kilometers.

1362 **A7 Numerical methods for the Coriolis and Magnus accelerations**

1363 There are a variety of numerical methods available in MOM6 for handling the Corio-
 1364 lis plus Magnus accelerations

1365
$$\text{Coriolis} + \text{Magnus} = \frac{f + \zeta}{h} (\hat{\mathbf{z}} \times \mathbf{u} h), \quad (\text{A1})$$

1366 where h is the layer thickness and ζ is the vertical component to the relative vorticity.
 1367 The term $\zeta \hat{\mathbf{z}} \times \mathbf{u}$ is referred to as the *Magnus acceleration* in some treatments, partic-
 1368 ularly the engineering literature, whereas it is commonly referred to as the *nonlinear Cori-*
 1369 *olis acceleration* in the ocean modeling literature. The layer thickness, h , in the numer-
 1370 ator does not exactly cancel h in the denominator. The reason is these two thicknesses
 1371 are computed using distinct horizontal averaging operators to ensure a proper handling
 1372 of the vanishing layer limit where $h \rightarrow 0$.

1373 Table A6 lists the suite of numerical settings found in CM4X as compared to CM4.0
 1374 for the Coriolis and Magnus accelerations, with a summary of these settings given here.

SCHEME	CM4.0	CM4X
NOSLIP	False	False
CORIOLIS_EN_DIS	False	True
CORIOLIS_SCHEME	SADOURNY75_ENSTRO	SADOURNY75_ENSTRO
BOUND_CORIOLIS	True	True
KE_SCHEME	KE_ARAKAWA	KE_ARAKAWA
PV_ADV_SCHEME	PV_ADV_CENTERED	PV_ADV_CENTERED

Table A6. Summary of the various numerical settings associated with the Coriolis and Magnus accelerations as used in CM4.0 and CM4X, with details provided in Section A7.

- For CM4X we set `CORIOLIS_EN_DIS = True`, which means that two estimates of the layer thickness fluxes are used to estimate the Coriolis + Magnus accelerations, and the one that dissipates energy relative to the other one is used. In CM4.0 this parameter was set `False`.
- In CM4.0 and CM4X we set `CORIOLIS_SCHEME = "SADOURNY75_ENSTRO"`, which is based on the enstrophy conserving scheme detailed in Sadourny (1975).
- In CM4.0 and CM4X we set `BOUND_CORIOLIS = True`, which means that the Coriolis terms at u-points are bounded by the four estimates of $(f+\zeta)v$ from the four neighboring v-points, and similarly at v-points. This option would have no effect on the SADOURNY Coriolis scheme if it were possible to use centered difference thickness fluxes.
- In CM4.0 and CM4X we set `KE_SCHEME = "KE_ARAKAWA"`, which determines the discretization for acceleration from the kinetic energy gradient.
- In CM4.0 and CM4X we set `PV_ADV_SCHEME = "PV_ADV_CENTERED"`, which determines the discretization of PV advection as per the centered scheme of Sadourny (1975).

1391 A8 Ocean bottom topography and geothermal heating

1392 For the 0.25° configuration of CM4X-p25, we retained the same bottom topogra-
 1393 phy and horizontal used in OM4.0, as well as the same bottom geothermal heat flux from
 1394 Huw Davies (2013). Here, we present steps used to construct the 0.125° topography for
 1395 CM4X-p125 as shown in Figure A2, along with the geothermal heat flux mapped to this
 1396 bottom topography.

1397 GEBCO and BedMachine data sources

1398 The bottom topography for CM4X-p125 is based on GEBCO (GEBCO Compilation
 1399 Group, 2021). This topography provides global coverage on a 15 arc-second inter-
 1400 val grid. For the region south of 62°S we transitioned to the BedMachine product from
 1401 Morlighem et al. (2017) since BedMachine offers more information about the Southern
 1402 Ocean and goes further inland to capture ice-shelf cavities (planned for future models
 1403 at GFDL but not implemented here). We placed the BedMachine data onto the GEBCO
 1404 grid using linear remapping and then interpolated this merged topography to the CM4X-
 1405 p125 ocean model grid.

1406 Hand tuning of selected straits

1407 For the C-grid of MOM6, flow through openings is available with a single grid tracer
 1408 cell. Hence, the C-grid allows for a more detailed representation of the complex geog-

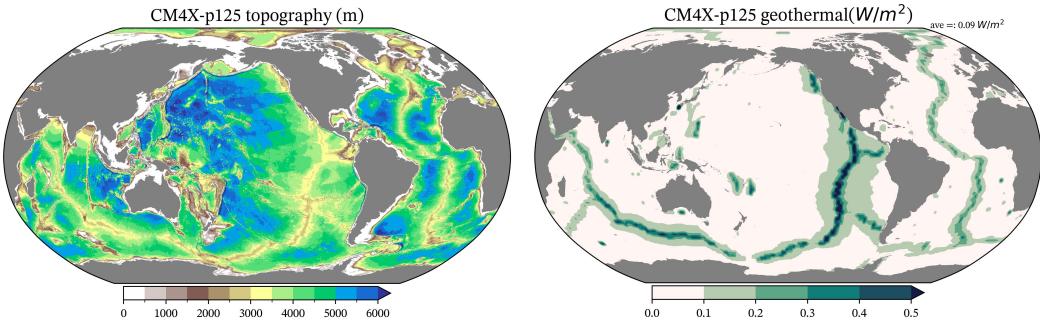


Figure A2. Top panel: ocean bottom depth (in meters) for the 0.125° horizontal grid used in CM4X-p125. For the region south of 62°S we transitioned from the GEBCO (GEBCO Compilation Group, 2021) product used in the north to the BedMachine product from Morlighem et al. (2017). Both CM4X-p25 and CM4X-p125 set their minimum depth at 9.5 m and maximum depth at 6500 m. Bottom panel: geothermal heat flux for CM4X-p125, taken from Huw Davies (2013). The global area mean (ocean area) is 0.09 W m^{-2} , though note that values along the mid-ocean ridges are generally much larger.

raphy of the World Ocean than the B-grid used in earlier MOM versions (e.g., the CM2.6 model of Griffies et al. (2015) and ACCESS-O model of Kiss et al. (2020) use MOM5, which is a B-grid). Even so, a coarsened grid generally admits straits that are artificially closed due to limitations of grid spacing. However, certain openings are particularly important for water masses of the World Ocean, with a primary example being the roughly 10 km wide Strait of Gibraltar connecting the Atlantic and Mediterranean. Conversely, there can be straits that are opened on a grid that are unwarranted, and so need to be closed. These considerations make it necessary to hand-tune special passages and straits to enable flow if artificially closed or to close straits that are best kept closed (e.g., Lake Maracaibo in Venezuela).

The hand-tuning works on a cell-by-cell basis, opening or closing straits according to the size of the grid cell. MOM6 provides the added feature of being able to restrict the openings of channels to a width corresponding to observed values, even if that value differs from the nominal grid cell size. We provide information about the restrictions in a runtime table, with the p125 grid restricting flow for the Dardanelles, Bosphorus, and between Sakhalin and Russia. After the hand-tuning process is complete, we fill inland waters, such as the North American Great Lakes, so that the ocean model only resolves sizable ocean water bodies and excludes lakes.

A9 Vertical coordinates and flow next to topography

MOM6 makes use of the vertical Lagrangian remapping method (see Griffies et al. (2020) for a review). For the vertical grid, we follow the hybrid vertical coordinate of OM4.0 as discussed in Section 2.1.4 of Adcroft et al. (2019), with this vertical coordinate following the approach of Bleck (2002). Here, the upper ocean layer interfaces are aligned with z^* quasi-geopotential surfaces (Adcroft & Campin, 2004), whereas the interior layer interfaces are aligned with potential density referenced to 2000 dbar. Note that the same vertical Lagrangian remapping method is used regardless the vertical coordinate choice. There are 75 coordinate layers for any particular vertical column of ocean, although some of those layers can have vanishing thickness depending on the density.

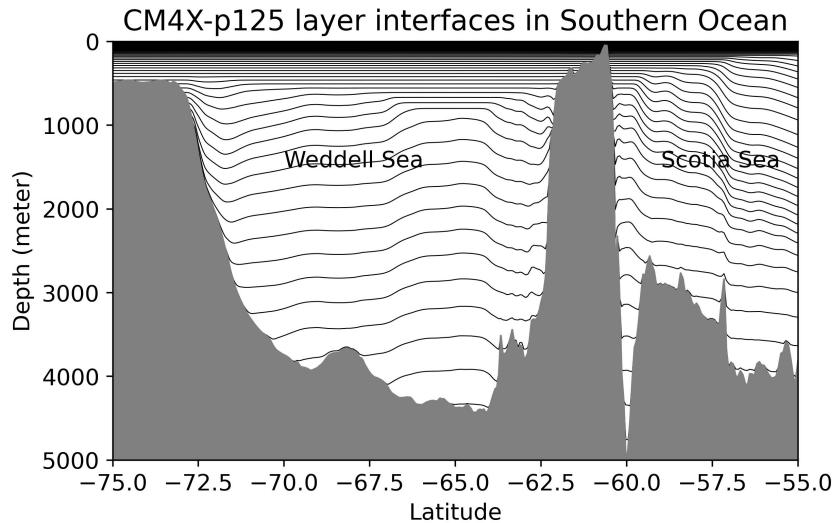


Figure A3. Layer interfaces through the Weddell and Scotia Seas (with South Orkney Islands rising just south of 60S) for the CM4X-p125 configuration as time averaged over years 1980–2009. The horizontal layers in the upper ocean are aligned with z^* whereas the interior layers are aligned with potential density referenced to 2000 dbar. Horizontal coordinate layers in the upper ocean are finely spaced, thus creating the dark black region. Representation of the bottom topography is linear piecewise continuous, with layers feeling a quadratic bottom drag when interacting with the bottom. Note that the strong baroclinicity next to the Antarctic shelf (southerly upward sloping isopycnals) is a signature of the westward flowing Antarctic Slope Current.

In Figure A3 we illustrate vertical coordinate layers in the Southern Ocean for a particular 1-year time mean from CM4X-p125. Along with the coordinate layers, this figure reveals that the solid-earth bottom is represented using a linear piecewise fit to the topography from Figure A2. As in the real ocean, MOM6 does not distinguish sides from bottom. Instead, the flow within a model layer feels a quadratic bottom drag wherever a layer intersects the solid earth (see Section 2.2.4 of Adcroft et al. (2019) for details of the bottom boundary layer scheme, including bottom drag). As discussed by Deremble et al. (2012), such treatment of the interactions of the fluid with solid boundaries is more physically motivated than that rendered by level coordinate models (see Adcroft et al. (1997) as well as Section 6 of Griffies et al. (2000) for more on the limitations of a level coordinate model representation of topography). Realistic flow-topography interactions are essential for capturing boundary waves and the associated sea level patterns along continental shelves, such as the east coast of North America (Wise et al., 2018, 2020; Hughes et al., 2019).

1451 A10 Parameterization of ocean bottom boundary layer mixing

The ocean bottom boundary layer parameterization in CM4X is similar to that used in OM4 and CM4 (see Legg et al., 2006), yet with one important change to the mixing efficiency parameter. The parameterization first estimates the turbulent kinetic energy (TKE) production in the bottom boundary layer using information from the dissipation of kinetic energy due to bottom drag. The turbulent vertical diffusivity is then computed by assuming a fraction of the TKE is dissipated and the remaining fraction is available to mix into stable stratification and thus to increase potential energy (following the Osborn-Cox relation, Osborn & Cox, 1972). In both OM4 and CM4, the mixing fraction (pa-

1460 parameter setting `bbl_effic` in MOM6) was set to 20%, which means that 80% of the TKE
 1461 production is dissipated and 20% goes into mixing. However, inspection of the coastal
 1462 waters in these models reveal this setting resulted in far too much bottom boundary layer
 1463 mixing, in some cases vertically homogenizing the water column in shelf regions.

1464 The 20% mixing fraction used in OM4 and CM4 was motivated by studies of tur-
 1465 bulence in stratified fluids without boundaries (Ivey & Imberger, 1991). Given the overly
 1466 strong mixing found with 20% mixing fraction, we chose to reduce it to 1% in CM4X.
 1467 This sizable change is motivated by the mixing fraction of turbulent mixing in large eddy
 1468 simulations of the surface boundary layer (see Reichl & Li, 2019). Further tests in an
 1469 OM4 configuration reveal little sensitivity to the precise value, so long as it is less than
 1470 a few percent.

1471 The reduced bottom boundary layer mixing fraction used in CM4X results in more
 1472 realistic mixed layer depths in coastal regions and more broadly improves circulation in
 1473 regions where dense water masses are formed in shallow regions. In related ongoing stud-
 1474 ies, outside the scope of this paper, we aim to develop a more thorough analysis of the
 1475 sensitivity of the simulations to the bottom boundary layer mixing fraction.

1476 A11 Parameterization of upper ocean boundary layer mixing

1477 The upper ocean boundary layer parameterization is based on Reichl and Hallberg
 1478 (2018), and we include the Reichl and Li (2019) extension that parameterizes effects from
 1479 Langmuir turbulence. The Stokes drift used by the Langmuir turbulence parameteriza-
 1480 tion is estimated following Q. Li et al. (2017). The settings for CM4X are updated rel-
 1481 ative to OM4.0 in order to agree with the recommended settings from Reichl and Li (2019).
 1482 The settings for CM4X were established prior to the recommendations from Reichl et
 1483 al. (2024) to improve upper ocean mixing and its diurnal variability in equatorial regions,
 1484 thus these changes are not included in CM4X. Similarly, the machine learning enhanced
 1485 vertical diffusivity as given in Sane et al. (2023) for the ocean surface boundary layer scheme
 1486 was not used since it was developed after CM4X.

1487 A12 Richardson number for parameterized shear-driven mixing

1488 Shear-driven mixing is parameterized according to the gradient Richardson num-
 1489 ber based scheme of Jackson et al. (2008). The gradient Richardson number, Ri , is the
 1490 ratio of the squared buoyancy frequency, N^2 , to the squared vertical shear of the hor-
 1491 izontal velocity

$$1492 \quad Ri = \frac{N^2}{(\partial_z u)^2 + (\partial_z v)^2} \quad \text{with} \quad N^2 = g(\alpha \partial_z \Theta - \beta \partial_z S), \quad (\text{A2})$$

1493 where g is the gravitational acceleration, Θ is the Conservative Temperature, S is the
 1494 salinity, α is the thermal expansion coefficient, and β is the haline contraction coefficient.
 1495 Ambiguities exist when computing the Richardson number on a discrete grid since trac-
 1496 ers and horizontal velocity components are generally centered at distinct locations. We
 1497 here describe how grid noise can be either generated or suppressed depending on how
 1498 the Richardson number is discretized.

1499 Discrete Richardson number that introduces grid noise

1500 For OM4.0, we chose to interpolate the horizontal C-grid velocity components to
 1501 the A-grid tracer points, with the vertical shear then computed on the A-grid along with
 1502 the buoyancy frequency and Richardson number (left panel of Figure A4). Analogous
 1503 approaches have been used in earlier versions of MOM (based on the B-grid velocity lay-
 1504 out), whereby the Richardson number is computed by either interpolating the tracer to
 1505 the B-grid velocity point or the velocity is interpolated to the tracer point. Unfortunately,

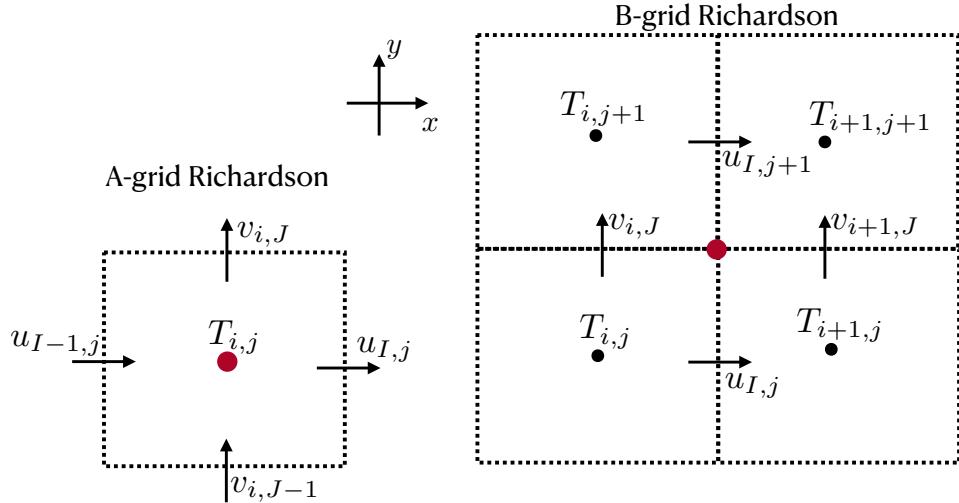


Figure A4. Two methods to compute the gradient Richardson number for the C-grid used in MOM6. Left panel: OM4.0 computes the Richardson number on the A-grid by averaging the horizontal velocity components, $u_{I,j}$ and $v_{i,J}$, to the tracer cell center (red dot at $T_{i,j}$) and then computing the vertical shear at the tracer cell center. Right panel: CM4X computes the Richardson number by averaging both horizontal velocity components as well as the temperature and salinity onto the B-grid corner point (red dot), and then computing the buoyancy frequency and vertical shear, again at the corner point. The A-grid approach can lead to the production of grid noise due to noise in the Richardson number that leads to noise in the mixing coefficients. The B-grid approach has no such noise. Note that in MOM6, I labels a field centered at the east face of tracer cell $T_{i,j}$, and J labels a field centered at the northern face of this cell.

by interpolating just the velocity or just the tracer, one can realize noisy Richardson numbers that then lead to noisy tracer and velocity fields. That is, a computational grid scale mode can arise from vertical diffusion schemes based on the Richardson number.

To see how noise can arise, consider a C-grid as in MOM6 with a tracer cell, $T_{i,j}$, that has zero vertical density stratification and so vanishing N^2 . Assume that this tracer is surrounded by horizontal C-grid velocities that have nonzero vertical shears, and by adjacent tracer cells with nonzero vertical density stratification. In this case the discrete Richardson number at $T_{i,j}$ is zero, which leads to a large eddy diffusivity from the Jackson et al. (2008) scheme (as well as most other Richardson number based mixing schemes), thus maintaining the zero vertical tracer stratification at $T_{i,j}$. Yet because the adjacent tracer cells are assumed to have nonzero density stratification, the Richardson number interpolated to the $T_{i,j}$ faces, where the horizontal velocity is centered, will be nonzero and so the vertical shears of the horizontal velocity will remain nonzero. Likewise, the Richardson number at the tracer cells with nonzero vertical stratification remains nonzero, in which case the mixing can be small. We are led to a large vertical tracer diffusivity at $T_{i,j}$ yet with adjacent cells having small tracer diffusivity. Correspondingly, if the vertical stratification has a checkerboard pattern, then this pattern is not damped by the A-grid Richardson number stencil. Rather, the pattern can be enhanced, thus supporting mixing-induced grid noise.

1525 ***Discrete Richardson number that does not introduce grid noise***

1526 Generating grid noise from the vertical diffusion scheme is quite disappointing since
 1527 we generally rely on diffusion to smooth fields rather than generate noise. Hence, for CM4X
 1528 we formulated an alternative discretization aimed at suppressing the computational mode.
 1529 We do so by interpolating the horizontal velocity components, as well as the tempera-
 1530 ture and salinity, to the tracer corner point; i.e., to the B-grid point. The buoyancy fre-
 1531 quency and vertical shear are then computed using the interpolated B-grid fields, as de-
 1532 picted in the right panel of Figure A4. In this manner, we couple across the tracer points
 1533 as well as velocity points, thus suppressing the A-grid computational mode.

1534 Analogous approaches have been used by various codes that use the A-grid Richard-
 1535 son number but perform a horizontal averaging prior to computing the diffusivities. For
 1536 example, earlier versions of MOM made use of this smoothing ever since the implemen-
 1537 tation of the Pacanowski and Philander (1981) shear mixing scheme, which is based on
 1538 the gradient Richardson number. Alternatively, one can average the noisy diffusivities
 1539 prior to using them in the tracer and velocity equations. Although these approaches can
 1540 be of some use for smoothing the noise, we find it more effective to eliminate the com-
 1541 putational mode at its origin rather than relying on downstream smoothing.

1542 Further examination of this modification is provided in (Reichl et al., 2024) with
 1543 a focus on the tropical Pacific. While the new approach helps to mitigate grid scale noise,
 1544 it has little impact on the time mean stratification near the equator.

1545 **A13 Parameterization of local internal tide-induced mixing**

1546 Mixing from breaking internal tides is an important part of the parameterized mix-
 1547 ing of ocean tracers and momentum in CM4.0 and CM4X, with the parameterization
 1548 following A. Melet et al. (2013). The first step of this parameterization, following St. Lau-
 1549 rent et al. (2002), computes the local energy conversion from the astronomically-forced
 1550 barotropic tide into baroclinic internal tides (dimensions of energy per time, or power,
 1551 per unit area)

$$1552 E_{\text{conversion}}(x, y, t) = (\rho_0 \kappa / 2) \langle h^2 \rangle \langle U_{\text{tides}}^2 \rangle N_b, \quad (\text{A3})$$

1553 where $\rho_0 = 1035 \text{ kg m}^{-3}$ is the Boussinesq reference density, $\langle h^2 \rangle$ is the variance of the
 1554 bottom topography (used as a measure of the topographic roughness), $\langle U_{\text{tides}}^2 \rangle$ is the vari-
 1555 ance of the barotropic tides, and N_b is the buoyancy frequency just above the bottom
 1556 boundary layer. The topography variance, tide variance, and bottom buoyancy frequency
 1557 are functions of horizontal position, with N_b also a function of time. While the charac-
 1558 teristic topographic wavenumber, κ (dimension of inverse length), could in principle also
 1559 be provided as a function of horizontal position, we use a spatially-constant value here
 1560 as a means to calibrate the globally-integrated energy conversion rate. Motivated by an
 1561 estimate that 1 TW of deep ocean barotropic-to-baroclinic tidal conversion is required
 1562 to reconcile energy budgets with observations (Egbert & Ray, 2000), we calibrate the
 1563 scheme to render a globally-integrated conversion rate of roughly 1 TW.

1564 We compute the topographic roughness by fitting a series of small planes to each
 1565 local patch of the topography in Figure A2, and then calculating the slope of the plane.
 1566 Figure A5 shows the resulting roughness field for CM4X-p125. For the tide amplitude,
 1567 we use the TPX09 version of the Oregon State University barotropic shallow water tide
 1568 model that uses the methods from Egbert and Erofeeva (2002). This model has a $1/30^\circ$
 1569 grid spacing and we make use of the following eight tidal harmonics: M2, S2, N2, K2,
 1570 K1, O1, P1, Q1. We interpolated the tide amplitudes onto the MOM6 grids, with Fig-
 1571 ure A5 showing the tide amplitude for the p125 grid.

1572 Once the tide amplitude has been interpolated to the model grid, we follow St. Lau-
 1573 rent et al. (2002) and Simmons et al. (2004) in calibrating the global tidal conversion
 1574 rate to be roughly 1 TW. This calibration is performed separately in the p25 and p125

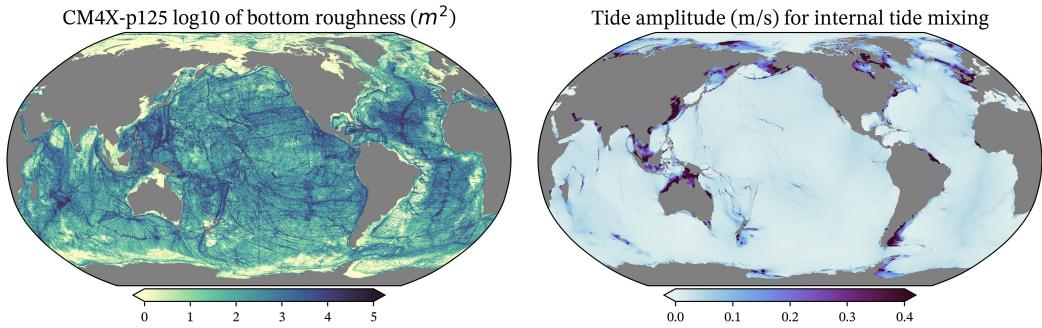


Figure A5. Left panel: Log10 of the bottom roughness (units of squared meters) for the bottom topography for CM4X-p125 from Figure A2. This field is used in the tide-induced mixing parameterization of Section A13. It is notable that straight lines appear from ship data, particularly visible in the high latitudes. Right panel: Amplitude (m/s) of the tides used for the tidal induced mixing in MOM6 for the CM4X-p125 configuration. Note the larger amplitudes on the continental shelves.

configurations by setting values for the characteristic topographic wavenumber, κ (see equation (A3)) responsible for the generation of internal tides. For the piControl simulations, Figure A6a reveals that CM4X-p125 settles onto roughly 1.17 TW globally (0.74 TW in the deep ocean) of energy conversion after 100 years, whereas CM4X-p25 equilibrates to roughly 1.16 TW globally (0.68 TW at depth) after 200 years. Evidently, the two models have their global internal tide mixing parameterizations calibrated to within 1% of each another.

In hindsight, we realized the appropriate observational target was roughly 1 TW of global *deep* ocean tidal conversion (Egbert & Ray, 2000) while the values diagnosed from the CM4X simulations also included conversion in shallow regions. Excluding all regions shallower than 2500 m (using a 2° smoothing of model bathymetry to yield a mask qualitatively similar to that used by Egbert and Ray (2000)), it was discovered that the tidal conversion in the deep ocean was only about 0.74 TW in CM4X-p125 and 0.68 TW in CM4X-p25, suggesting κ should be recalibrated upwards by about 30%.

Figure A6a shows that there is no sizable signal when moving from the piControl to historical. In contrast, the SSP5-8.5 simulation shows a clear signal of enhanced bottom stratification that, in turn, leads to enhanced tidal conversion. Virtually all of this increased conversion is attributable to shallow continental shelves (compare Figure A6a and b), where the bottom stratification responds relatively quickly to enhanced surface warming. This 5% increase in tidal conversion due to global warming is an example of the poorly understood climate-mixing feedback processes that are enabled by increasingly comprehensive and state-dependent mixing parameterizations (A. V. Melet et al., 2022).

To relate this energy input into the internal wave field to irreversible mixing, a constant fraction $q = 30\%$ is first assumed to produce turbulence kinetic energy (primarily through the breaking of high vertical modes) with a vertical structure function that follows Polzin (2009). Only a fraction, $\Gamma \leq 0.2$, (tapered to zero in regions of very weak stratification) of this locally-produced turbulent kinetic energy is used to drive irreversible diapycnal mixing, which is implemented as an enhanced vertical diffusivity. The remaining 70% of the energy input into low mode internal tides is not explicitly accounted for and is instead assumed to be implicit in a latitudinally-varying background diffusivity, k_d .

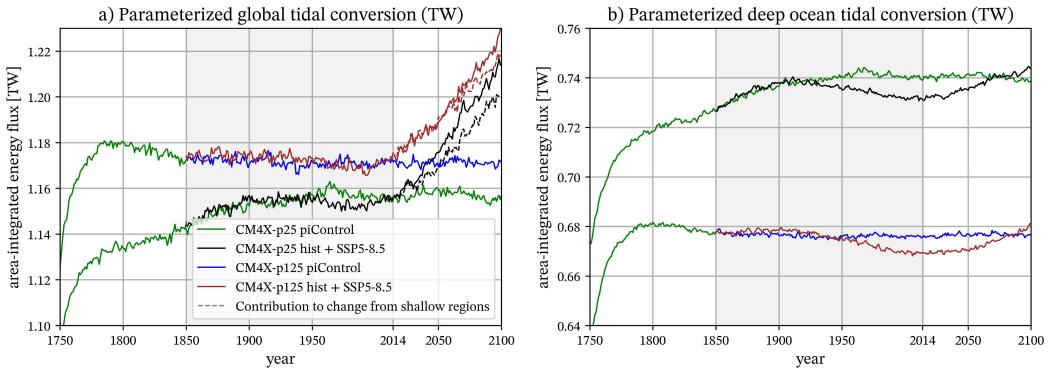


Figure A6. Time series for the energy conversion from the barotropic tide to baroclinic internal tides, as per equation (A3), integrated over (a) the global ocean and (b) regions deeper than 2500 m. The CM4X-p125 piControl simulation equilibrates to roughly 1.17 TW, whereas CM4X-p25 equilibrates to a bit less than 1.16 TW. There is a minor difference in the conversion rates for the historical versus piControl simulations. However, there is a clear signal of enhanced tidal conversion in the SSP5-8.5 simulation, arising primarily from an increase in the area integrated bottom stratification in shallow regions. The gray shaded region is the historical portion of the simulation (1850–2014).

1607

A14 Parameterization of baroclinic eddies in the mixed layer

1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621

We make use of the mixed layer eddy (MLE) restratification scheme from Fox-Kemper et al. (2008) and Fox-Kemper et al. (2011). This scheme provides a parameterization of the mixed layer baroclinic instability as studied by Boccaletti et al. (2007). In general, the upper ocean warms when the MLE restratification is strengthened (heat remains in the upper ocean), and the upper ocean cools when the MLE restratification is weakened (heat can be pumped into the ocean interior). Section 2.2.2 of Adcroft et al. (2019) details the implementation of MLE in OM4.0, and we closely follow that approach for CM4X. The MLE scheme in OM4.0 is set to be (arguably unphysically) strong in order to counteract biases in CM4.0 partly due to under-resolved upper ocean restratification by mesoscale eddies. For CM4X, we reduced the strength of the scheme, anticipating that the eddies in CM4X-p125 would be stronger than CM4.0 or CM4X-p25, thus enabling us to reduce the restratification effects from the submesoscale parameterization. In addition to heat uptake, we monitored the mixed layer depth and sea surface temperature when testing parameter changes.

1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632

There are two parameters in the MLE scheme that we tested in developing CM4X. The first is the frontal length, with a stronger restratification tendency for smaller fronts and weaker for larger fronts. The study by Bodner et al. (2023) updates the Fox-Kemper et al. (2008) scheme by providing a dynamical closure for the front length rather than prescribing it. However, we did not make use of Bodner et al. (2023) given that the CM4X project started a few years prior to the implementation of the updates to MOM6. Hence, the frontal length is a prescribed parameter in OM4.0 and CM4X. The second parameter we tested concerns the timescale for the running time mean filter applied to the mixed layer depth, with details provided in Section 2.2.2 of Adcroft et al. (2019). Lengthening the decay time increases the strength of the restratification since the longer decay allows the MLE scheme to remain enabled for an increased period of time.

1633
1634

We performed non-exhaustive tests in a prototype of CM4X-p25 using a front length of 1000 m and 2000 m. Yet we found relatively little sensitivity to this parameter com-

MODEL	MLE_FRONT_LENGTH	MLE_MLD_DECAY_TIME
CM4.0	500 m	2.592×10^6 s = 30 days
CM4X	500 m	1.728×10^6 s = 20 days

Table A7. Summary of the key MLE parameters as used in CM4.0 and CM4X.

pared to the OM4.0 setting of 500 m, thus prompting us to retain 500 m for CM4X. In contrast, the ocean heat uptake and sea surface temperature are sensitive to the time scale used for the low pass time filter. We weakened the scheme in CM4X by reducing the time scale from the 30 days used in OM4.0 to 20 days in CM4X. This reduction of the MLE strength is partly motivated given that the CM4X-p125 ocean has stronger mesoscale eddies, thus reducing the need for MLE to perform the bulk of upper ocean restratification. Table A7 summarizes the MLE settings used in CM4X and CM4.0.

Appendix B Open Research

Software comprising the model as well as the software used for creating the figures will be placed on Zenodo at the revision stage of this work.

Observation-based datasets used in this paper are cited locally. We are indebted to the many efforts of the various programs providing observational-based data used to help evaluate these simulations, including the following.

- The Argo program provides data that were collected and made freely available by the International Argo Program and the national programs that contribute to it, with access available from

<http://www.argo.ucsd.edu> and <http://argo.jcommops.org>

The Argo Program is part of the Global Ocean Observing System.

- OSNAP data were collected and made freely available by the OSNAP (Overturning in the Subpolar North Atlantic Program) project and all the national programs that contribute to it (www.o-snap.org). The DOI for this data set is

<https://doi.org/10.35090/gatech/70342>

Acknowledgments

This project started in May 2020, during the early stages of the Covid-19 pandemic shutdown. We are grateful to the GFDL computer operations team for keeping the computational resources reliable during the shutdown. We thank the GFDL management for providing the computer resources needed for the development and analysis documented here. We thank John Dunne, Matthew Harrison, and Tony Rosati for very helpful comments on early drafts of this manuscript. Krista Dunne, Sergey Malyshev and Chris Milly kindly provided expertise in helping to update the rivers and lakes for use with the C192 atmosphere coupled to the p125 ocean. The statements, findings, conclusions, and recommendations are those of the author(s) and do not necessarily reflect the views of the National Oceanic and Atmospheric Administration, or the U.S. Department of Commerce.

- The statements, findings, conclusions, and recommendations are those of the author(s) and do not necessarily reflect the views of the National Oceanic and Atmospheric Administration, or the U.S. Department of Commerce.

- A.A. was supported by Award NA18OAR4320123 from the National Oceanic and Atmospheric Administration, U.S. Department of Commerce.
- R.B. was supported under NSF Division of Polar Programs Grant NSF2319828.
- C.Y.C. was supported by Award NA19OAR4310365 from the National Oceanic and Atmospheric Administration, U.S. Department of Commerce.
- H.F.D. was supported by the NOAA Climate and Global Change Postdoctoral Fellowship Program, administered by UCAR's Cooperative Programs for the Advancement of Earth System Science (CPAESS) under Award NA18NWS4620043B.
- H.K. acknowledges the support from Natural Environment Research Council grants NE/T013494/1 and NE/W001543/1.
- M.L. was supported by award NA18OAR4320123 and NA23OAR4320198 from the National Oceanic and Atmospheric Administration, U.S. Department of Commerce.
- G.A.M was supported by NSF (PLR-1425989) and UKRI (MR/W013835/1).
- A.S. was supported by Schmidt Sciences, LLC under the M²LInES project.
- K.E.T acknowledges support from the Southern Ocean Carbon and Climate Observations and Modeling (SOCCOM) Project under NSF Awards PLR-1425989 and OPP-1936222 and 2332379.
- L.Z. was supported by Schmidt Sciences, LLC under the M²LInES project, NSF grant OCE 1912357 and NOAA CVP NA19OAR4310364.
- W.Z. was supported by the National Science Foundation under Grant Number F1240-01(NSF OCE 1912357). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

- Adcroft, A., Anderson, W., Blanton, C., Bushuk, M., Dufour, C. O., Dunne, J. P., ... Zhang, R. (2019). The GFDL global ocean and sea ice model OM4.0: Model description and simulation features. *Journal of Advances in Modeling the Earth System, JAMES*. doi: 10.1029/2019MS001726
- Adcroft, A., & Campin, J.-M. (2004). Rescaled height coordinates for accurate representation of free-surface flows in ocean circulation models. *Ocean Modelling*, 7, 269–284. doi: 10.1016/j.ocemod.2003.09.003
- Adcroft, A., Hill, C., & Marshall, J. (1997). Representation of topography by shaved cells in a height coordinate ocean model. *Monthly Weather Review*, 125, 2293–2315. doi: 10.1175/1520-0493(1997)125<2293:ROTBSC>2.0.CO;2
- Adler, R. F., Sapiano, M. R., Huffman, G. J., Wang, J. J., Gu, G., Bolvin, D., ... Shin, D.-B. (2018). The global precipitation climatology project (GPCP) monthly analysis (new version 2.3) and a review of 2017 global precipitation. *Atmosphere*, 9. doi: 10.3390/atmos9040138
- Andrews, T., Gregory, J. M., Paynter, D., Silvers, L. G., Zhou, C., Mauritsen, T., ... Titchner, H. (2018). Accounting for changing temperature patterns increases historical estimates of climate sensitivity. *Geophysical Research Letters*, 45, 8490–8499. doi: 10.1029/2018GL078887
- Argo. (2023). *Argo float data and metadata from Global Data Assembly Centre (Argo GDAC)* [dataset]. SEANOE. Retrieved from <https://www.seanoe.org/data/00311/42182/> doi: 10.17882/42182
- Armour, K. C., Bitz, C. M., & Roe, G. H. (2013). Time-varying climate sensitivity from regional feedbacks. *Journal of Climate*, 26, 4518–4534. doi: 10.1175/JCLI-D-12-00544.1
- Armour, K. C., Marshall, J., Scott, J. R., Donohoe, A., & Newsom, E. R. (2016). Southern Ocean warming delayed by circumpolar upwelling and equatorward transport. *Nature Geosciences*, 9. doi: 10.1038/NGEO2731
- Arora, V. K., Katavouta, A., Williams, R. G., Jones, C. D., Brovkin, V., Friedling-

- stein, P., ... Ziehn, T. (2019, December). Carbon-concentration and carbon-climate feedbacks in CMIP6 models, and their comparison to CMIP5 models. *Biogeosciences Discussions*, 1–124. Retrieved 2020-06-07, from <https://www.biogeosciences-discuss.net/bg-2019-473/> (Publisher: Copernicus GmbH) doi: <https://doi.org/10.5194/bg-2019-473>
- Banks, H. T., Stark, S., & Keen, A. B. (2006). The adjustment of the coupled climate model HadGEM1 toward equilibrium and the impact on global climate. *Journal of Climate*, 20, 5815–5826. doi: 10.1175/2007JCLI1688.1
- Beadling, R. L. (2023). Global consequences of regional connectivity along the Antarctic margin. *Journal of Geophysical Research: Oceans*, 128. doi: 10.1029/2023JC019908
- Beadling, R. L., Krasting, J. P., Griffies, S. M., Hurlin, W. J., Bronselear, B., Russell, J. L., ... Winton, M. (2022). Importance of the Antarctic Slope Current in the Southern Ocean response to ice sheet melt and wind stress change. *Journal of Geophysical Research - Oceans*, 127. doi: 10.1029/2021jc017608
- Beadling, R. L., Russell, J. L., Stouffer, R. J., Mazloff, M., Talley, L. D., Goodman, P. J., ... Pandde, A. (2020). Representation of Southern Ocean properties across coupled model intercomparison project generations: CMIP3 to CMIP6. *Journal of Climate*, 33, 6555–6581. doi: 10.1175/JCLI-D-19-0970.1
- Bian, C., Jing, Z., Wang, H., Wu, L., Chen, Z., Gan, B., & Yang, H. (2023). Oceanic mesoscale eddies as crucial drivers of global marine heatwaves. *Nature Communications*, 14, 2970. doi: 10.1038/s41467-023-38811-z
- Bleck, R. (2002). An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates. *Ocean Modelling*, 4, 55–88. doi: 10.1016/S1463-5003(01)00012-9
- Boccaletti, G., Ferrari, R., & Fox-Kemper, B. (2007). Mixed layer instabilities and restratification. *Journal of Physical Oceanography*, 35, 1263–1278. doi: 10.1175/JPO3101.1
- Bodner, A., Fox-Kemper, B., Johnson, L., van Rokel, L. P., McWilliams, J. C., Sullivan, P. P., ... Dong, J. (2023). Modifying the mixed layer eddy parameterization to include frontogenesis arrest by boundary layer turbulence. *Journal of Physical Oceanography*, 53, 323–339. doi: 10.1175/JPO-D-21-0297.s1
- Boville, B. A., & Gent, P. R. (1998). The NCAR climate system model, version one. *Journal of Climate*, 11, 1115–1130. doi: 10.1175/1520-0442(1998)011<1115:TNCNSM>2.0.CO;2
- Briegleb, B. P., & Light, B. (2007). *A delta-Eddington multiple scattering parameterization for solar radiation in the sea ice component of the Community Climate System Model* (Technical Report Nos. TN-4721+STR). Boulder, USA: National Center for Atmospheric Research.
- Burls, N. J., Muir, L., Vincent, E. M., & Fedorov, A. V. (2016). Extra-tropical origin of equatorial Pacific cold bias in climate models with links to cloud albedo. *Climate Dynamics*, 49, 2093–2113. doi: 10.1007/s00382-016-3435-6
- Caesar, L., Rahmstorf, S., Robinson, A., Feulner, G., & Saba, V. (2018). Observed fingerprint of a weakening Atlantic Ocean overturning circulation. *Nature*, 556, 191–196. doi: 10.1038/s41586-018-0006-5
- Cess, R., Potter, G., & Coauthors. (1990). Intercomparison and interpretation of climate feedback processes in 19 atmospheric general circulation models. *Journal of Geophysical Research*, 16601–16615.
- Cessi, P. (2019). The global overturning circulation. *Annual Review of Marine Science*, 11, 249–270. doi: 10.1146/annurev-marine-010318-095241
- Charney, J. G. (1971). Geostrophic turbulence. *Journal of the Atmospheric Sciences*, 28, 1087–1095. doi: 10.1175/1520-0469(1971)028<1087:GT>2.0.CO;2
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., ... Vitart, F. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*

- 1778 *Society*, 137, 553–597. doi: 10.1002/qj.828
- 1779 Delworth, T. L., Broccoli, A. J., Rosati, A., Stouffer, R. J., Balaji, V., Beesley, J. A.,
 1780 ... Zhang, R. (2006). GFDL's CM2 global coupled climate models - Part 1:
 1781 Formulation and simulation characteristics. *Journal of Climate*, 19, 643–674.
 1782 doi: 10.1175/JCLI3629.1
- 1783 Delworth, T. L., Cooke, W. F., Adcroft, A., Bushuk, M., Chen, J.-H., Dunne, K. A.,
 1784 ... others (2020). SPEAR: The next generation GFDL modeling system for
 1785 seasonal to multidecadal prediction and projection. *Journal of Advances in*
 1786 *Modeling Earth Systems*, 12(3), e2019MS001895.
- 1787 Delworth, T. L., Rosati, A., Anderson, W., Adcroft, A. J., Balaji, V., Benson, R., ...
 1788 Zhang, R. (2012). Simulated climate and climate change in the GFDL CM2.5
 1789 high-resolution coupled climate model. *Journal of Climate*, 25, 2755–2781. doi:
 1790 10.1175/JCLI-D-11-00316
- 1791 Deppenmeier, A.-L., Bryan, F. O., Kessler, W. S., & Thompson, L. (2021).
 1792 Modulation of cross-isothermal velocities with ENSO in the Tropical Pa-
 1793 cific Cold Tongue. *Journal of Physical Oceanography*, 51, 1559–1574. doi:
 1794 10.1175/JPO-D-20-0217.1
- 1795 Deremble, B., Hogg, A. M., Berloff, P., & Dewar, W. K. (2012). On the applica-
 1796 tion of no-slip lateral boundary conditions to ‘coarsely’ resolved ocean models.
 1797 *Ocean Modelling*, 39, 411–415. doi: 10.1016/j.ocemod.2011.05.002
- 1798 Doddridge, E. W., Marshall, D. P., & Hogg, A. M. (2016). Eddy Cancellation of
 1799 the Ekman Cell in Subtropical Gyres. *Journal of Physical Oceanography*, 46,
 1800 2995–3010. doi: 10.1175/JPO-D-16-0097.1
- 1801 Drake, H. F., Bailey, S., Dussin, R., Griffies, S. M., Krasting, J. P., MacGilchrist,
 1802 G. A., ... Zika, J. D. (2024, April). *Water mass transformation bud-
 1803 getts in finite-volume generalized vertical coordinate ocean models*. Retrieved
 1804 2024-05-31, from [https://essopenarchive.org/users/242115/
 1805 articles/788549-water-mass-transformation-budgets-in-finite
 1806 -volume-generalized-vertical-coordinate-ocean-models?commit=](https://essopenarchive.org/users/242115/articles/788549-water-mass-transformation-budgets-in-finite-volume-generalized-vertical-coordinate-ocean-models?commit=e6178873ac99d67c98963b54ba60699d57b7703f) doi: 10.22541/essoar
 1807 .171284935.57181910/v1
- 1808 Dunne, J. P., Horowitz, L. W., Adcroft, A., Ginoux, P., Held, I. M., John, J. G., ...
 1809 Zhao, M. (2020). The GFDL Earth system model version 4.1 (GFDL-ESM4.1):
 1810 Model description and simulation characteristics. *Journal of Advances in Mod-
 1811 eling Earth Systems*, 12. doi: 10.1029/2019MS002015
- 1812 Egbert, G. D., & Erofeeva, S. Y. (2002). Efficient inverse modeling of barotropic
 1813 ocean tides. *Journal of Atmospheric and Oceanic Technology*, 19, 183–204.
 1814 doi: 10.1175/1520-0426(2002)019<0183:EIMOBO>2.0.CO;2
- 1815 Egbert, G. D., & Ray, R. D. (2000, June). Significant dissipation of tidal energy
 1816 in the deep ocean inferred from satellite altimeter data. *Nature*, 405(6788),
 1817 775–778. Retrieved 2022-05-12, from <https://www.nature.com/articles/35015531> (Number: 6788 Publisher: Nature Publishing Group) doi:
 1818 10.1038/35015531
- 1819 Eyring, V., Bony, S., Meehl, J. A., Senior, C., Stevens, B., Stouffer, R. J., & Taylor,
 1820 K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase
 1821 6 (CMIP6) experimental design and organisation. *Geoscientific Model Develop-
 1822 ment*, 9, 1937–1958. Retrieved from <http://www.geosci-model-dev.net/9/1937/2016/> doi: 10.5194/gmd-9-1937-2016
- 1823 Forster, P., Storelvmo, T., Armour, K., Collins, W., Dufresne, J.-L., Frame, D., ...
 1824 Zhang, H. (2021). The earth’s energy budget, climate feedbacks, and climate
 1825 sensitivity. In *Climate Change 2021: The Physical Science Basis. Contribution
 1826 of Working Group I to the Sixth Assessment Report of the Intergovernmen-
 1827 tal Panel on Climate Change* (pp. 923–1051). Cambridge, United Kingdom:
 1828 Cambridge University Press. doi: 10.1017/9781009157896.009
- 1829 Fox-Kemper, B., Danabasoglu, G., Ferrari, R., Griffies, S. M., Hallberg, R. W., Hol-

- land, M. M., ... Samuels, B. (2011). Parameterization of mixed layer eddies. III: Global implementation and impact on ocean climate simulations. *Ocean Modelling*, 39, 61–78. doi: 10.1016/j.ocemod.2010.09.002
- Fox-Kemper, B., Ferrari, R., & Hallberg, R. (2008). Parameterization of mixed layer eddies. I: Theory and diagnosis. *Journal of Physical Oceanography*, 38, 1145–1165. doi: 10.1175/2007JPO3792.1
- Fox-Kemper, B., Hewitt, H. T., Xiao, C., Aoalgeirsdóttir, G., Drijfhout, S. S., Edwards, T. L., ... Yu, Y. (2021). Ocean, cryosphere and sea level change. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 1211–1362). Cambridge, United Kingdom: Cambridge University Press. doi: 10.1017/9781009157896.011
- Ganachaud, A., & Wunsch, C. (2000). Improved estimates of global ocean circulation, heat transport and mixing from hydrographic data. *Nature*, 408, 453–456. doi: 10.1038/35044048
- GEBCO Compilation Group. (2021). GEBCO 2021 Grid [dataset]. Retrieved from https://www.bodc.ac.uk/data/published_data_library/catalogue/10.5285/c6612cbe-50b3-0cff-e053-6c86abc09f8f/ doi: 10.5285/c6612cbe-50b3-0cff-e053-6c86abc09f8f
- Gent, P. R., & McWilliams, J. C. (1990). Isopycnal mixing in ocean circulation models. *Journal of Physical Oceanography*, 20, 150–155. doi: 10.1175/1520-0485(1990)020<0150:IMIOCM>2.0.CO;2
- Gent, P. R., Willebrand, J., McDougall, T. J., & McWilliams, J. C. (1995). Parameterizing eddy-induced tracer transports in ocean circulation models. *Journal of Physical Oceanography*, 25, 463–474. doi: 10.1175/1520-0485(1995)025<0463:PEITTI>2.0.CO;2
- Ghimire, B., Williams, C. A., Masek, J., Gao, F., Wang, Z., Schaaf, C., & He, T. (2014). Global albedo change and radiative cooling from anthropogenic land cover change, 1700 to 2005 based on MODIS, land use harmonization, radiative kernels, and reanalysis. *Geophysical Research Letters*, 41, 9087–9096. doi: 10.1002/2014GL061671
- Gleckler, P. J., Taylor, K. E., & Doutriaux, C. (2008). Performance metrics for climate models. *Journal of Geophysical Research*, 113. doi: 10.1029/2007jd008972
- Gnanadesikan, A., Slater, R. D., Swathi, P. S., & Vallis, G. K. (2005). The energetics of ocean heat transport. *Journal of Climate*, 17, 2604–2616. doi: 10.1175/JCLI3436.1
- Gordon, C., Cooper, C., Senior, C., Banks, H., Gregory, J. M., Johns, T. C., ... Woods, R. A. (2000). The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dynamics*, 16, 147–168. doi: 10.1007/s003820050010
- Gregory, J. M. (2000). Vertical heat transports in the ocean and their effect on time-dependent climate change. *Climate Dynamics*, 15, 501–515. doi: 10.1007/s003820000059
- Gregory, J. M., Bouttes, N., Griffies, S. M., Haak, H., Hurlin, W. J., Jungclaus, J., ... Winton, M. (2016). The Flux-Anomaly-Forced Model Intercomparison Project (FAFMIP) contribution to CMIP6: investigation of sea-level and ocean climate change in response to co_2 forcing. *Geoscientific Model Development*, 9, 3993–4017. doi: 10.5194/gmd-9-3993-2016
- Gregory, J. M., Griffies, S. M., Hughes, C. W., Lowe, J. A., Church, J. A., Fukimori, I., ... van de Wal, R. S. W. (2019). Concepts and terminology for sea level-mean, variability and change, both local and global. *Surveys in Geophysics*, 40, 1251–1289. doi: 10.1007/s10712-019-09555-7
- Gregory, J. M., & Tailleux, R. (2011). Kinetic energy analysis of the response of the Atlantic meridional overturning circulation to co_2 -forced climate change. *Cli-*

- mate Dynamics*(37), 893–914. doi: 10.1007/s00382-010-0847-6

Griffies, S. M., Adcroft, A., Albert, M. S., Beadling, R. L., Bushuk, M., Chang, C.-Y., ... Zika, J. D. (2024). The GFDL-CM4X climate model hierarchy, Part II: case studies. *Journal of Advances in Modeling Earth Systems*.

Griffies, S. M., Adcroft, A., & Hallberg, R. W. (2020). A primer on the vertical Lagrangian remap method in ocean models based on finite volume generalized vertical coordinates. *Journal of Advances in Modeling Earth Systems*, 12. doi: 10.1029/2019MS001954

Griffies, S. M., Böning, C. W., Bryan, F. O., Chassignet, E. P., Gerdes, R., Hasumi, H., ... Webb, D. (2000). Developments in ocean climate modelling. *Ocean Modelling*, 2, 123–192. doi: 10.1016/S1463-5003(00)00014-7

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

Griffies, S. M., & Treguier, A.-M. (2013). Ocean models and ocean modeling. In G. Siedler, S. M. Griffies, J. Gould, & J. Church (Eds.), *Ocean circulation and climate, 2nd edition: A 21st century perspective* (Vol. 103, pp. 521–552). Academic Press. doi: 10.1016/B978-0-12-391851-2.00020-9

Griffies, S. M., Winton, M., Anderson, W. G., Benson, R., Delworth, T. L., Dufour, C., ... Zhang, R. (2015). Impacts on ocean heat from transient mesoscale eddies in a hierarchy of climate models. *Journal of Climate*, 28, 952–977. doi: 10.1175/JCLI-D-14-00353.1

Griffies, S. M., Yin, J., Durack, P. J., Goddard, P., Bates, S., Behrens, E., ... Zhang, Z. (2014). An assessment of global and regional sea level for years 1993–2007 in a suite of interannual CORE-II simulations. *Ocean Modelling*, 78, 35–89. doi: 10.1016/j.ocemod.2014.03.004

Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., ... J.-S. von Storch (2016). High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6. *Geoscientific Model Development*, 9, 4185–4208. Retrieved from <http://www.geosci-model-dev.net/9/4185/2016/> doi: 10.5194/gmd-9-4185-2016

Hallberg, R. W. (2013). Using a resolution function to regulate parameterizations of oceanic mesoscale eddy effects. *Ocean Modelling*, 72, 92–103. doi: 10.1016/j.ocemod.2013.08.007

Hallberg, R. W., Adcroft, A. J., Dunne, J. P., Krasting, J. P., & Stouffer, R. (2013). Sensitivity of twenty-first-century global-mean steric sea level rise to ocean model formulation. *Journal of Climate*, 26, 2947–2956. doi: 10.1175/JCLI-D-12-00506.1

Held, I. M. (2005). Gap between simulation and understanding in climate modeling. *Bulletin of the American Meteorological Society*, 86, 1609–1614. doi: 10.1175/BAMS-86-11-1609

Held, I. M., Guo, H., Adcroft, A., Dunne, J. P., Horowitz, L. W., Krasting, J., ... Zadeh, N. (2019). Structure and performance of GFDL's CM4.0 climate model. *Journal of Advances in Modeling the Earth System*, 11, 3691–3726. doi: 10.1029/2019MS001829

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., ... Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146, 1999–2049. doi: 10.1002/qj.3803

Hieronymus, M., & Nycander, J. (2013). The budgets of heat and salinity in NEMO. *Ocean Modelling*, 67, 28–38. doi: 10.1016/j.ocemod.2013.03.006

Hieronymus, M., Nycander, J., Nilsson, J., Döös, K., & Hallberg, R. (2019). Oceanic overturning and heat transport: The role of background diffusivity. *Journal of Climate*, 32(3), 701–716.

- 1943 Hirschi, J. J., Barnier, B., Böning, C., Biastoch, A., Blaker, A. T., Coward, A., ...
 1944 Xu, X. (2020). The Atlantic meridional overturning circulation in high-
 1945 resolution models. *Journal of Geophysical Research: Oceans*, 125. doi:
 1946 10.1029/2019JC015522
- 1947 Hofstadter, D. R. (1979). *Gödel, Escher, Bach: An Eternal Golden Braid*. Basic
 1948 Books. (777 pages)
- 1949 Holmes, R. M., Zika, J., & England, M. H. (2019). Diathermal heat transport in a
 1950 global ocean model. *Journal of Physical Oceanography*, 49, 141–161. doi: 10
 1951 .1175/JPO-D-18-0098.1
- 1952 Holmes, R. M., Zika, J., Griffies, S. M., A. McC. Hogg, Kiss, A. E., & England,
 1953 M. H. (2020). The geography of numerical mixing in a suite of global
 1954 ocean models. *Journal of Advances in Modeling Earth Systems*, 13. doi:
 1955 10.1029/2020MS002333
- 1956 Holmes, R. M., Zika, J. D., Griffies, S. M., Hogg, A. M., Kiss, A. E., & England,
 1957 M. H. (2021). The geography of numerical mixing in a suite of global
 1958 ocean models. *Journal of Advances in Modeling Earth Systems*, 13(7),
 1959 e2020MS002333. Retrieved from <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020MS002333> (e2020MS002333 2020MS002333) doi:
 1960 <https://doi.org/10.1029/2020MS002333>
- 1961 Holzer, M., DeVries, T., & de Lavergne, C. (2021). Diffusion controls the ventilation
 1962 of a pacific shadow zone above abyssal overturning. *Nature Communications*,
 1963 12(1), 4348.
- 1964 Hsieh, T., Vecchi, G. A., Yang, W., Held, I. M., & Garner, S. T. (2020). Large-scale
 1965 control on the frequency of tropical cyclones and seeds: a consistent relation-
 1966 ship across a hierarchy of global atmospheric models. *Climate Dynamics*, 55,
 1967 3177–3196. doi: 10.1007/s00382-020-05446-5
- 1968 Huang, B., Liu, C., Banzon, V., Freeman, E., Graham, G., Hankins, B., ... Zhang,
 1969 H.-M. (2020). Improvements of the daily optimum interpolation sea surface
 1970 temperature (DOISST) Version 2.1. *Journal of Climate*, 34, 2923–2939. doi:
 1971 10.1175/JCLI-D-20-0166.1
- 1972 Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore,
 1973 J. H., ... Zhang, H.-M. (2017). *NOAA Extended Reconstructed Sea Surface
 1974 Temperature (ERSST), Version 5* (dataset). U.S. Department of Commerce,
 1975 Washington, D.C.: NOAA National Centers for Environmental Information.
 1976 Retrieved from <https://www.esrl.noaa.gov/psd/> doi: 0.7289/V5T72FNM
- 1977 Hughes, C. W., Fukumori, I., Griffies, S. M., Huthnance, J. M., Minobe, S., Spence,
 1978 P., ... Wise, A. (2019). Sea level and the role of coastal trapped waves in
 1979 mediating the influence of the open ocean on the coast. *Surveys in Geophysics*,
 1980 40, 1467–1492. doi: 10.1007/s10712-019-09535-x
- 1981 Huw Davies, J. (2013). Global map of solid Earth surface heat flow. *Geochemistry,
 1982 Geophysics, Geosystems*, 14, 4608–4622. doi: 10.1002/ggge.20271
- 1983 Hyder, P., Edwards, J. M., Allan, R. P., Hewitt, H. T., Bracegirdle, T. J., Gregory,
 1984 J. M., ... Belcher, S. E. (2018). Critical Southern Ocean climate model biases
 1985 traced to atmospheric model cloud errors. *Nature Communications*, 9. doi:
 1986 10.1038/s41467-018-05634-2
- 1987 Ilicak, M., Adcroft, A. J., Griffies, S. M., & Hallberg, R. W. (2012). Spurious di-
 1988 aneutral mixing and the role of momentum dissipation. *Ocean Modelling*, 45–
 1989 46, 37–58. doi: 10.1016/j.ocemod.2011.10.003
- 1990 Irving, D., Hobbs, W., Church, J., & Zika, J. (2021). A mass and energy conser-
 1991 vation analysis of drift in the CMIP6 ensemble. *Journal of Climate*, 34, 3157–
 1992 3170. doi: 10.1175/JCLI-D-20-0281.1
- 1993 Ivey, G., & Imberger, J. (1991). On the nature of turbulence in a stratified fluid.
 1994 Part I: The energetics of mixing. *Journal of Physical Oceanography*, 21(5),
 1995 650–658.
- 1996 Jackson, L., Hallberg, R. W., & Legg, S. (2008). A parameterization of shear-driven
 1997

- turbulence for ocean climate models. *Journal of Physical Oceanography*, 38, 1033–1053. doi: 10.1175/2007JPO3779.1

Jeevanjee, N., Hassanzadeh, P., Hill, S., & Sheshadri, A. (2017). A perspective on climate model hierarchies. *Journal of Advances in Modeling Earth Systems*, 9, 1760–1771. doi: 10.1002/2017MS001038

Johnson, G. C., & Purkey, S. G. (2024). Refined estimates of global ocean deep and abyssal decadal warming trends. *Geophysical Research Letters*, 51. doi: 10.1029/2024GL111229

Khatiwala, S. (2023). Fast spin-up of geochemical tracers in ocean circulation and climate models. *Journal of Advances in Modeling Earth Systems (JAMES)*, 15. doi: 10.1029/2022MS003447

Khatiwala, S. (2024). Efficient spin-up of Earth System Models using sequence acceleration. *Science Advances*, 10. doi: 10.1126/sciadv.adn2839

Kirtman, B. P., Bitz, C., Bryan, F., Collins, W., Dennis, J., Hearn, N., ... Vertenstein, M. (2012). Impact of ocean model resolution on CCSM climate simulations. *Climate Dynamics*, 39, 1303–1328. doi: 10.1007/s00382-012-1500-3

Kiss, A. E., Hogg, A. M., Hannah, N., Dias, F. B., Brassington, G., Chamberlain, M. A., ... Zhang, X. (2020). ACCESS-OM2: A global ocean-sea ice model at three resolutions. *Geoscientific Model Development*. doi: 10.5194/gmd-13-401-2020

Kjellsson, J., & Zanna, L. (2017). The impact of horizontal resolution on energy transfers in global ocean models. *Fluids*, 2(3), 45. doi: 10.3390/fluids2030045

Krasting, J. P., Griffies, S. M., Tesdal, J.-E., MacGilchrist, G., Beadling, R. L., & Little, C. M. (2024). Steric sea level rise and relationships with model drift and water mass representation in GFDL CM4 and ESM4. *Journal of Climate*, accepted.

Krasting, J. P., Stouffer, R. J., Griffies, S. M., Hallberg, R. W., Malyshev, S. L., Samuels, B. L., & Sentman, L. T. (2018). Role of ocean model formulation in climate response uncertainty. *Journal of Climate*, 22, 9313–9333. doi: 10.1175/JCLI-D-18-0035.1

Lauritzen, P. H., Kevlahan, N. K.-R., Tonizzo, T., Eldred, C., Dubos, T., Gassmann, A., ... Bacmeister, J. T. (2022). Reconciling and improving formulations for thermodynamics and conservation principles in Earth System Models (ESMs). *Journal of Advances in Modeling Earth Systems*, 14. doi: 10.1029/2022MS003117

Lee, J., Gleckler, P. J., Ahn, M.-S., Ordonez, A., Ullrich, P. A., Sperber, K. R., ... Krasting, J. (2023). Objective evaluation of Earth system models: PCMDI Metrics Package (PMP) version 3. *EGUsphere*, 2023, 1–48. Retrieved from <https://egusphere.copernicus.org/preprints/2023/egusphere-2023-2720/> doi: 10.5194/egusphere-2023-2720

Legg, S., Hallberg, R. W., & Girton, J. B. (2006, January). Comparison of entrainment in overflows simulated by z-coordinate, isopycnal and non-hydrostatic models. *Ocean Modelling*, 11(1), 69–97. Retrieved 2018-12-20, from <http://www.sciencedirect.com/science/article/pii/S1463500304001064> doi: 10.1016/j.ocemod.2004.11.006

Lellouche, J.-M., Greiner, E., Bourdallé-Badie, R., Garric, G., Melet, A., Drévillon, M., ... Le Traon, P.-Y. (2021). The Copernicus global 1/12° oceanic and sea ice GLORYS12 reanalysis. *Frontiers in Earth Science*, 9. doi: 10.3389/feart.2021.698876

Li, G., Du, Y., Xu, H., & Ren, B. (2015). An intermodel approach to identify the source of excessive equatorial pacific cold tongue in cmip5 models and uncertainty in observational datasets. *Journal of Climate*, 28, 7630–7640. doi: 10.1175/JCLI-D-15-0168.1

Li, Q., Fox-Kemper, B., Øyvind Breivik, & Webb, A. (2017). Statistical models of global Langmuir mixing. *Ocean Modelling*, 113, 95–114. doi: <https://doi.org/10.1016/j.ocemod.2017.03.001>

- 2053 10.1016/j.ocemod.2017.03.016
- 2054 Li, Q., & Lee, S. (2017). A mechanism of mixed layer formation in the Indo-Western
2055 Pacific Southern Ocean: Preconditioning by an eddy-driven jet-scale over-
2056 turning circulation. *Journal of Physical Oceanography*, *47*, 2755–2772. doi:
2057 10.1175/JPO-D-17-0006.1
- 2058 Locarnini, R. A., Mishonov, A. V., Antonov, J. I., Boyer, T. P., Garcia, H. E., Bara-
2059 nova, O. K., ... Seidov, D. (2018). *World Ocean Atlas 2013, Volume 1: Temperature* (dataset). U.S. Department of Commerce, Washington, D.C.: NOAA/NESDIS.
- 2060 Loeb, N. G., Doelling, D. R., Wang, H., Su, W., Nguyen, C., Corbett, J. G., ...
2061 Seiji, K. (2018). Clouds and the Earth's radiant energy system (CERES)
2062 energy balanced and filled (EBAF) top-of-atmosphere (TOA) edition-4.0
2063 data product. *International Journal of Climatology*, *31*, 895–918. doi:
2064 10.1175/JCLI-D-17-0208.1
- 2065 Maher, P., Gerber, E. P., Medeiros, B., Merlis, T. M., Sherwood, S., Sheshadri, A.,
2066 ... Zurito-Gotor, P. (2019). Model hierarchies for understanding atmospheric
2067 circulation. *Reviews of Geophysics*, *57*, 250–280. doi: 10.1029/2018RG000607
- 2068 Manabe, S., Bryan, K., & Spelman, M. J. (1990). Transient response of a global
2069 ocean–atmosphere model to a doubling of atmospheric carbon dioxide. *Journal
2070 of Physical Oceanography*, *20*, 722–749. doi: 10.1175/1520-0485(1990)020<722:
2071 TROAGO>2.0.CO;2
- 2072 Manabe, S., & Stouffer, R. J. (1988). Two stable equilibria of a coupled ocean-
2073 atmosphere model. *Journal of Climate*, *1*(9), 841–866. doi: 10.1175/1520
2074 -0442(1988)001<0841:TSEOAC>2.0.CO;2
- 2075 Manabe, S., Stouffer, R. J., Spelman, M. J., & Bryan, K. (1991). Transient re-
2076 sponds of a coupled ocean–atmosphere model to gradual changes of atmo-
2077 spheric CO₂. Part 1: Annual mean response. *Journal of Climate*, *4*, 785–818.
2078 doi: 10.1175/1520-0442(1991)004<0785:TROACO>2.0.CO;2
- 2079 Marques, G. M., Loose, N., Yankovsky, E., Steinberg, J. M., Chang, C.-Y., Bhamidi-
2080 pati, N., ... Zanna, L. (2022). Neverworld2: An idealized model hierarchy
2081 to investigate ocean mesoscale eddies across resolutions. *Geoscientific Model
2082 Development*. doi: 10.5194/gmd-15-6567-2022
- 2083 Marshall, D. P., & Zanna, L. (2014). A conceptual model of ocean heat uptake un-
2084 der climate change. *Journal of Climate*, *27*, 8444–8465. doi: 10.1175/JCLI-D
2085 -13-00344.1
- 2086 Marshall, J., & Schott, F. (1999). Open-ocean convection: observations, theory, and
2087 models. *Reviews of Geophysics*, *37*, 1–64. doi: 10.1029/98RG02739
- 2088 McDougall, T. J., Barker, P. M., Holmes, R. M., Pawlowicz, R., Griffies, S. M., &
2089 Durack, P. J. (2021). The interpretation of temperature and salinity vari-
2090 ables in numerical ocean model output, and the calculation of heat fluxes
2091 and heat content. *Geoscientific Model Development*, *14*, 6445–6466. doi:
2092 10.5194/gmd-14-6445-2021
- 2093 Melet, A., Hallberg, R. W., Legg, S., & Polzin, K. (2013). Sensitivity of the Pacific
2094 Ocean state to the vertical distribution of internal-tide driven mixing. *Journal
2095 of Physical Oceanography*, *43*, 602–615. doi: 10.1175/JPO-D-12-055.1
- 2096 Melet, A. V., Hallberg, R., & Marshall, D. P. (2022, January). Chapter 2
2097 - The role of ocean mixing in the climate system. In M. Meredith &
2098 A. Naveira Garabato (Eds.), *Ocean Mixing* (pp. 5–34). Elsevier. Retrieved
2099 2023-01-29, from <https://www.sciencedirect.com/science/article/pii/B9780128215128000098> doi: 10.1016/B978-0-12-821512-8.00009-8
- 2100 Morlighem, M., Williams, C. N., Rignot, E., An, L., Arndt, J. E., Bamber, J. L., ...
2101 Zinglersen, K. B. (2017). Bedmachine v3: Complete bed topography and ocean
2102 bathymetry mapping of greenland from multibeam echo sounding combined
2103 with mass conservation. *Geophysical Research Letters*, *44*, 11051–11061. doi:
2104 10.1002/2017GL074954

- 2108 Morrison, A. K., Saenko, O. A., Hogg, A. M., & Spence, P. (2013). The role of
 2109 vertical eddy flux in southern ocean heat uptake. *Geophysical Research Letters*, 40(20), 5445–5450.
 2110 Retrieved from <http://dx.doi.org/10.1002/2013GL057706> doi: 10.1002/2013GL057706
- 2112 Murray, R. J. (1996). Explicit Generation of Orthogonal Grids for Ocean Mod-
 2113 els. *Journal of Computational Physics*, 126(2), 251–273. Retrieved 2018-
 2114 05-14, from <http://www.sciencedirect.com/science/article/pii/S0021999196901369> doi: 10.1006/jcph.1996.0136
- 2116 Newsom, E., Zanna, L., & Gregory, J. (2023). Background pycnocline depth con-
 2117 strains future ocean heat uptake efficiency. *Geophysical Research Letters*, 50.
 2118 doi: 10.1029/2023GL105673
- 2119 Orr, J. C., Najjar, R. G., Aumont, O., Bopp, L., Bullister, J., Danabasoglu, G., ...
 2120 Yool, A. (2017). Biogeochemical protocols and diagnostics for the CMIP6
 2121 Ocean Model Intercomparison Project (OMIP). *Geoscientific Model Develop-
 2122 ment*, 10, 2169–2199. doi: 10.5194/gmd-10-2169-2017
- 2123 Osborn, T. R., & Cox, C. S. (1972). Oceanic fine structure. *Geophysical Fluid Dy-
 2124 namics*, 3(4), 321–345.
- 2125 Pacanowski, R. C., & Philander, G. (1981). Parameterization of vertical mixing in
 2126 numerical models of the tropical ocean. *Journal of Physical Oceanography*, 11,
 2127 1442–1451. doi: 10.1175/1520-0485(1981)011<1443:POVMIN>2.0.CO;2
- 2128 Polzin, K. L. (2009). An abyssal recipe. *Ocean Modelling*, 30(4), 298–309. doi: 10
 2129 .1016/j.ocemod.2009.07.006
- 2130 Pudig, M. P., Yung, C. K., Zika, J. D., & Holmes, R. M. (2023). Rectified ocean
 2131 heat uptake from oscillatory surface forcing. *Journal of Climate*, 36, 2663–
 2132 2680. doi: 10.1175/JCLI-D-22-0267.1
- 2133 Purkey, S., & Johnson, G. (2010). Warming of global abyssal and deep South-
 2134 ern Ocean waters between the 1990s and 2000s: contributions to global
 2135 heat and sea level rise budgets. *Journal of Climate*, 23, 6336–6351. doi:
 2136 10.1175/2010JCLI3682.1
- 2137 Purkey, S., & Johnson, G. (2012). Global contraction of Antarctic Bottom Water be-
 2138 tween the 1980s and 2000s. *Journal of Climate*, 25, 5830–5844. doi: 10.1175/
 2139 JCLI-D-11-00612.1
- 2140 Purkey, S., & Johnson, G. (2013). Antarctic Bottom Water warming and freshening:
 2141 Contributions to sea level rise, ocean freshwater budgets, and global heat gain.
 2142 *Journal of Climate*, 26, 6105–6122. doi: 10.1175/JCLI-D-12-00834.1
- 2143 Redi, M. H. (1982). Oceanic isopycnal mixing by coordinate rotation. *Journal of
 2144 Physical Oceanography*, 12, 1154–1158. doi: 10.1175/1520-0485(1982)012<1154:
 2145 OIMBCR>2.0.CO;2
- 2146 Reichl, B. G., Adcroft, A., Griffies, S. M., & Hallberg, R. W. (2022). A potential en-
 2147 ergy analysis of ocean surface mixed layers. *Journal of Geophysical Research-
 2148 Oceans*, 127. doi: 10.1029/2021JC018140
- 2149 Reichl, B. G., & Hallberg, R. W. (2018). A simplified energetics based planetary
 2150 boundary layer (ePBL) approach for ocean climate simulations. *Ocean Mod-
 2151 ellling*, 132, 112–129. doi: 10.1016/j.ocemod.2018.10.004
- 2152 Reichl, B. G., & Li, Q. (2019). A parameterization with a constrained potential en-
 2153 ergy conversion rate of vertical mixing due to Langmuir turbulence. *Journal of
 2154 Physical Oceanography*, 49, 2935–2959. doi: 10.1175/JPO-D-18-0258.1
- 2155 Reichl, B. G., Wittenberg, A. T., Griffies, S. M., & Adcroft, A. (2024). Improving
 2156 equatorial upper ocean vertical mixing in the noaa/gfdl om4 model. *Earth
 2157 and Space Science*, 11(10), e2023EA003485. doi: <https://doi.org/10.1029/2023EA003485>
- 2159 Rhein, M., Rintoul, S., Aoki, S., Campos, E., Chambers, D., Feely, R., ... Wang, F.
 2160 (2013). Observations: Ocean. In *Climate change 2013: The physical science
 2161 basis. contribution of working group i to the fifth assessment report of the inter-
 2162 governmental panel on climate change* (pp. 255–315). Cambridge, United

- Kingdom: Cambridge University Press.
- Richter, I. (2015). Climate model biases in the eastern tropical oceans: Causes, impacts and ways forward. *Wiley Interdisciplinary Reviews: Climate Change*, 6(3), 345–358.
- Rugenstein, M., Bloch-Johnson, J., Gregory, J., Andrews, T., Mauritsen, T., Li, C., ... others (2020). Equilibrium climate sensitivity estimated by equilibrating climate models. *Geophysical Research Letters*, 47(4), e2019GL083898.
- Saba, V. S., Griffies, S., Anderson, W., Winton, M., Alexander, M., Delworth, T., ... Zhang, R. (2016). Enhanced warming of the northwest Atlantic Ocean under climate change. *Journal of Geophysical Research–Oceans*, 121, 118–132. doi: 10.1002/2015JC011346
- Sadourny, R. (1975, April). The Dynamics of Finite-Difference Models of the Shallow-Water Equations. *Journal of the Atmospheric Sciences*, 32(4), 680–689. Retrieved 2019-07-28, from [https://journals.ametsoc.org/doi/abs/10.1175/1520-0469\(1975\)032\(0680:TDOFDM\)2.0.CO;2](https://journals.ametsoc.org/doi/abs/10.1175/1520-0469%281975%29032%3C0680%3ATDOFDM%3E2.0.CO%3B2) doi: 10.1175/1520-0469(1975)032(0680:TDOFDM)2.0.CO;2
- Sallée, J.-B., Shuckburgh, E., Bruneau, N., Meijers, A. J. S., Bracegirdle, T. J., Wang, Z., & Roy, T. (2013). Assessment of Southern Ocean water mass circulation and characteristics in CMIP5 models: historical bias and forcing response. *Journal of Geophysical Research: Oceans*, 118. doi: 10.1002/jgrc.20135
- Salmon, R. (1980). Baroclinic instability and geostrophic turbulence. *Geophysical & Astrophysical Fluid Dynamics*, 15(1), 167–211. doi: 10.1080/03091928008241178
- Sane, A., Reichl, B. G., Adcroft, A., & Zanna, L. (2023). Parameterizing vertical mixing coefficients in the ocean surface boundary layer using neural networks. *Journal of Advances in Modeling Earth Systems*, 15(10), e2023MS003890. Retrieved from [\(e2023MS003890 2023MS003890\)](https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2023MS003890) doi: <https://doi.org/10.1029/2023MS003890>
- Sausen, R., Barthel, K., & Hasselmann, K. (1988). Coupled ocean-atmosphere models with flux correction. *Climate Dynamics*, 2, 145–163. doi: 10.1007/BF01053472
- Scott, R. B., & Wang, F. (2005, September). Direct evidence of an oceanic inverse kinetic energy cascade from satellite altimetry. *Journal of Physical Oceanography*, 35(9), 1650–1666. (Publisher: American Meteorological Society Section: Journal of Physical Oceanography) doi: 10.1175/JPO2771.1
- Séférian, R., Gehlen, M., Bopp, L., Resplandy, L., Orr, J. C., Marti, O., ... Romanou, A. (2016). Inconsistent strategies to spin up models in cmip5: implications for ocean biogeochemical model performance assessment. *Geoscientific Model Development*, 9, 1827–1851. doi: 10.5194/gmd-9-1827-2016
- Simmons, H. L., Jayne, S. R., St.Laurent, L. C., & Weaver, A. J. (2004). Tidally driven mixing in a numerical model of the ocean general circulation. *Ocean Modelling*, 6, 245–263. doi: 10.1016/S1463-5003(03)00011-8
- Smith, K. S., & Vallis, G. K. (2001). The scales and equilibration of midocean eddies: freely evolving flow. *Journal of Physical Oceanography*, 31, 554–570. doi: 10.1175/1520-0485(2001)031<0554:TSAEOM>2.0.CO;2
- Solomon, H. (1971). On the representation of isentropic mixing in ocean models. *Journal of Physical Oceanography*, 1, 233–234. doi: 10.1175/1520-0485(1971)001<0233:OTROIM>2.0.CO;2
- St. Laurent, L. C., Simmons, H. L., & Jayne, S. R. (2002). Estimating tidally driven energy in the deep ocean. *Geophysical Research Letters*, 29, 2106–2110. doi: 10.1029/2002GL015
- Stewart, A. L., Chi, X., Solodoch, A., & Hogg, A. M. (2021). High-frequency fluctuations in Antarctic Bottom Water transport driven by Southern Ocean winds.

- 2218 *Geophysical Research Letters*, 49. doi: 10.1029/2021GL094569
- 2219 Stewart, K., Hogg, A., Griffies, S. M., Heerdegen, A., Ward, M., Spence, P., & Eng-
2220 land, M. (2017). Vertical resolution of baroclinic modes in global ocean
2221 models. *Ocean Modelling*, 113, 50–65. doi: 10.1016/j.ocemod.2017.03.012
- 2222 Stouffer, R. J. (2004). Time scales of climate response. *Journal of Climate*, 31, 209–
2223 217. doi: 10.1175/1520-0442(2004)017<0209:TSOCR>2.0.CO;2
- 2224 Stouffer, R. J., Russel, J., & Spelman, M. J. (2006). Importance of oceanic heat up-
2225 take in transient climate change. *Geophysical Research Letters*, 33. doi: 10
2226 .1029/2006GL027242
- 2227 Tesdal, J.-E., MacGilchrist, G. A., Beadling, R. L., Griffies, S. M., Krasting, J. P., &
2228 Durack, P. J. (2023). Revisiting interior water mass responses to surface forc-
2229 ing changes and the subsequent effects on overturning in the Southern Ocean.
2230 *Journal of Geophysical Research - Oceans*, 128. doi: 10.1029/2022JC019105
- 2231 Thomas, M. D., & Federov, A. V. (2017). The eastern subtropical Pacific origin of
2232 the equatorial cold bias in climate models: A Lagrangian perspective. *Journal*
2233 *of Climate*, 30, 5885–5900. doi: 10.1175/JCLI-D-16-0819.1
- 2234 Trenberth, K. E., & Caron, J. M. (2001). Estimates of meridional atmosphere and
2235 ocean heat transports. *Journal of Climate*, 14, 3433–3443. doi: 10.1175/1520
2236 -0442(2001)014<3433:EOMAAO>2.0.CO;2
- 2237 Tsujino, H., Urakawa, S., Nakano, H., Small, R. J., Kim, W. M., Yeager, S. G.,
2238 ... Yamazaki, D. (2018). JRA-55 based surface dataset for driving
2239 ocean-sea-ice models (JRA-55-do). *Ocean Modelling*, 130, 79–139. doi:
2240 10.1016/j.ocemod.2018.07.002
- 2241 Vallis, G. K. (2017). *Atmospheric and oceanic fluid dynamics: Fundamentals and*
2242 *large-scale circulation* (2nd ed.). Cambridge: Cambridge University Press. (946
2243 + xxv pp)
- 2244 von Storch, J.-S., Haak, H., Hertwig, E., & Fast, I. (2016). Vertical heat and salt
2245 fluxes due to resolved and parameterized meso-scale eddies. *Ocean Modelling*,
2246 108, 1–19. doi: 10.1016/j.ocemod.2016.10.001
- 2247 von Schuckmann, K., Minière, A., Gues, F., Cuesta-Valero, F. J., Kirchengast,
2248 G., Adusumilli, S., ... Zemp, M. (2023). Heat stored in the earth system
1960–2020: where does the energy go? *Earth System Science Data*, 15(4),
2249 1675–1709. Retrieved from <https://essd.copernicus.org/articles/15/1675/2023/> doi: 10.5194/essd-15-1675-2023
- 2250 Wang, C., Zhang, L., Lee, S.-K., Wu, L., & Mechoso, C. R. (2014). A global per-
2251 spective on CMIP5 climate model biases. *Nature Climate Change*, 4. doi: 10
2252 .1038/NCLIMATE2118
- 2253 Weaver, A. J., & Hughes, T. M. C. (1996). On the incompatibility of ocean and
2254 atmosphere models and the need for flux adjustments. *Climate Dynamics*, 12,
2255 141–170. doi: 10.1007/BF00211615
- 2256 Welch, P. (1967). The use of fast Fourier transform for the estimation of power
2257 spectra: A method based on time averaging over short, modified peri-
2258 odograms. *IEEE Transactions on Audio and Electroacoustics*, 15, 70–73.
2259 doi: 10.1109/TAU.1967.1161901
- 2260 Wills, R. C. J., Dong, Y., Proistosecu, C., Armour, K. C., & Battisti, D. S. (2022).
2261 Systematic climate model biases in the large-scale patterns of recent sea-
2262 surface temperature and sea-level pressure change. *Geophysical Research*
2263 *Letters*, 49. Retrieved from <https://doi.org/10.1029/2022GL100011> doi:
2264 10.1029/2022GL100011
- 2265 Wise, A., Hughes, C. W., & Polton, J. A. (2018). Bathymetric influence on the
2266 coastal sea level response to ocean gyres at western boundaries. *Journal of*
2267 *Physical Oceanography*, 481, 2949–2964. doi: 10.1175/JPO-D-18-0007.1
- 2268 Wise, A., Hughes, C. W., Polton, J. A., & Huthnance, J. M. (2020). Leaky slope
2269 waves and sea level: Unusual consequences of the beta effect along western
2270 boundaries with bottom topography and dissipation. *Journal of Physical*
2271 *Oceanography*, 33, 103–115. doi: 10.1029/2019JC015470
- 2272

- 2273 *Oceanography*, 50, 217–237. doi: 10.1175/JPO-D-19-0084.1
- 2274 Wittenberg, A. T. (2009). Are historical records sufficient to constrain ENSO simu-
2275 lations? *Geophysical Research Letters*, 36. doi: 10.1029/2009GL038710
- 2276 Wolfe, C. L., Cessi, P., McClean, J. L., & Maltrud, M. E. (2008). Vertical heat
2277 transport in eddying ocean models. *Geophysical Research Letters*, 35(23), n/a–
2278 n/a. Retrieved from <http://dx.doi.org/10.1029/2008GL036138> doi: 10
2279 .1029/2008GL036138
- 2280 Wright, D. G. (1997). An equation of state for use in ocean models: Eckart's for-
2281 mula revisited. *Journal of Atmospheric and Oceanic Technology*, 14, 735–740.
2282 doi: 10.1175/1520-0426(1997)014<735:AEOSFU>2.0.CO;2
- 2283 Yankovsky, E., Zanna, L., & Smith, K. S. (2022). Influences of mesoscale ocean ed-
2284 dies on flow vertical structure in a resolution-based model hierarchy. *Journal of*
2285 *Advances in Modeling Earth Systems*, n/a(n/a). doi: 10.1029/2022MS003203
- 2286 Yin, J., Griffies, S. M., Winton, M., Zhao, M., & Zanna, L. (2020). Response
2287 of storm-related extreme sea level along the us Atlantic Coast to combined
2288 weather and climate forcing. *Journal of Climate*, 33, 3745–3769. doi:
2289 10.1175/JCLI-D-19-0551.1
- 2290 Zanna, L., Khatiwala, S., Gregory, J., Ison, J., & Heimbach, P. (2019). Global recon-
2291 struction of historical ocean heat storage and transport. *Proceedings of the Na-
2292 tional Academy of Science*. doi: 10.1073/pnas.1808838115
- 2293 Zhang, L., Delworth, T. L., Cooke, W., Goosse, H., Bushuk, M., Morioka, Y., &
2294 Yang, X. (2021). The dependence of internal multidecadal variability in the
2295 Southern Ocean on the ocean background mean state. *Journal of Climate*,
2296 34(3), 1061–1080. doi: 0.1175/JCLI-D-20-0049.1
- 2297 Zhang, R., Delworth, T. L., Rosati, A., Anderson, W. G., Dixon, K. W., Lee, H.-
2298 C., & Zeng, F. (2011). Sensitivity of the North Atlantic Ocean circula-
2299 tion to an abrupt change in the Nordic Sea overflow in a high resolution
2300 global coupled climate model. *Journal of Geophysical Research*, 116. doi:
2301 10.1029/2011JC007240
- 2302 Zhang, R., Sutton, R., Danabasoglu, G., Marsh, R., Yeager, S. G., Amrhein, D. E.,
2303 & Little, C. M. (2019). A review of the role of the Atlantic meridional over-
2304 turning circulation in Atlantic multidecadal variability and associated climate
2305 impacts. *Reviews of Geophysics*, 57. doi: 10.1029/2019RG000644
- 2306 Zhao, M. (2020). Simulations of atmospheric rivers, their variability, and response to
2307 global warming using GFDL's new high-resolution general circulation model.
2308 *Journal of Climate*, 33, 10287–10303. doi: 10.1175/JCLI-D-20-0241.1
- 2309 Zhao, M. (2022a). An investigation of the effective climate sensitivity in GFDL's
2310 new climate models CM4.0 and SPEAR. *Journal of Climate*, 35, 5637–5660.
2311 doi: 10.1175/JCLI-D-21-0327.s1
- 2312 Zhao, M. (2022b). A study of AR-, TS-, and MCS-Associated precipitation and
2313 extreme precipitation in present and warmer climates. *Journal of Climate*, 35,
2314 479–497, DOI: 10.1175/JCLI-D-21-0145.1.
- 2315 Zhao, M., Golaz, J.-C., Held, I. M., Guo, H., Balaji, V., Benson, R., ... Xiang, B.
2316 (2018a). The GFDL Global Atmosphere and Land Model AM4.0/LM4.0 - Part
2317 II: Model Description, Sensitivity Studies, and Tuning Strategies. *Journal of*
2318 *Advances in Modeling Earth Systems*, 10. doi: 10.1002/2017MS001209
- 2319 Zhao, M., Golaz, J.-C., Held, I. M., Guo, H., Balaji, V., Benson, R., ... Xiang, B.
2320 (2018b). The GFDL Global Atmosphere and Land Model AM4.0/LM4.0 - Part
2321 I: Simulation Characteristics with Prescribed SSTs. *Journal of Advances in*
2322 *Modeling Earth Systems*, 10. doi: 10.1002/2017MS001208
- 2323 Zika, J. D., Le Sommer, J., Dufour, C. O., Molines, J.-M., Barnier, B., Brasseur, P.,
2324 ... Vivier, F. (2014). Vertical eddy fluxes in the Southern Ocean. *Journal of*
2325 *Physical Oceanography*, 43, 941–955. doi: 10.1175/JPO-D-12-0178.1
- 2326 Zweng, M. M., Reagan, J. R., Antonov, J. I., Locarnini, R. A., Mishonov, A. V.,
2327 Boyer, T. P., ... Biddle, M. M. (2018). *World Ocean Atlas 2013, Volume*

2328 *2: Salinity* (dataset). U.S. Department of Commerce, Washington, D.C.:
2329 NOAA/NESDIS.