JAMES | Journal of Advances in Modeling Earth Systems®

RESEARCH ARTICLE
10.1029/2024MS004861

The GFDL-CM4X Climate Model Hierarchy, Part I: Model Description and Thermal Properties

Stephen M. Griffies[1,2], Alistair Adcroft[2], Rebecca L. Beadling[3], Mitchell Bushuk[1], Chiung-Yin Chang[2], Henri F. Drake[4], Raphael Dussin[1], Robert W. Hallberg[1,2], William J. Hurlin[1], Hemant Khatri[5], John P. Krasting[1], Matthew Lobo[2], Graeme A. MacGilchrist[6], Brandon G. Reichl[1], Aakash Sane[2], Olga Sergienko[2], Maike Sonnewald[7], Jacob M. Steinberg[1], Jan-Erik Tesdal[2], Matthew Thomas[8], Katherine E. Turner[9], Marshall L. Ward[1], Michael Winton[1], Niki Zadeh[1], Laure Zanna[10], Rong Zhang[1,2], Wenda Zhang[2], and Ming Zhao[1]

[1]NOAA Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA, [2]Princeton University Atmospheric and Oceanic Sciences Program, Princeton, NJ, USA, [3]Department of Earth and Environmental Science, Temple University, Philadelphia, PA, USA, [4]Department of Earth System Science, University of California, Irvine, CA, USA, [5]Department of Earth, Ocean and Ecological Sciences, University of Liverpool, Liverpool, UK, [6]School of Earth and Environmental Sciences, University of St. Andrews, St. Andrews, UK, [7]Computer Science Department, University of California, Davis, CA, USA, [8]Centre for Environment Fisheries and Aquaculture Science, Suffolk, UK, [9]Department of Geosciences, University of Arizona, Tucson, AZ, USA, [10]Courant Institute of Mathematical Sciences, New York University, New York, NY, USA

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

Correspondence to:
S. M. Griffies,
Stephen.M.Griffies@gmail.com

Citation:
Griffies, S. M., Adcroft, A., Beadling, R. L., Bushuk, M., Chang, C.-Y., Drake, H. F., et al. (2025). The GFDL-CM4X climate model hierarchy, Part I: Model description and thermal properties. *Journal of Advances in Modeling Earth Systems*, *17*, e2024MS004861. https://doi.org/10.1029/2024MS004861

Received 23 NOV 2024
Accepted 4 SEP 2025

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 CM4.0–50 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 W 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 ZJ less ocean heat than present-day, which corresponds to estimates for anthropogenic ocean heat uptake between 1870 and present-day. CM4X-p25 approaches its thermal equilibrium only after more than 1000 years, at which time its ocean has roughly 1100 ZJ *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 *mesoscale dominance hypothesis* for why CM4X-p125 shows such favorable thermal equilibration properties.

Plain Language Summary

We detail a new climate model hierarchy, CM4X. CM4X has two model configurations, CM4X-p25 and CM4X-p125, that differ only in the ocean/sea ice horizontal grid spacing. CM4X-p125 outperforms CM4X-p25 for certain climate processes, while maintaining skill levels seen in previous generations for other results. CM4X-p125 requires about 10 times less time than CM4X-p25 to reach pre-industrial control thermal equilibration. Also, CM4X-p125 equilibrates to an ocean state with roughly 400 ZJ less heat content than present-day, consistent with estimates of anthropogenic heat uptake since 1870, whereas CM4X-p25 equilibrates to a state with roughly 1100 ZJ more heat than present-day. Consequently, the CM4X-p125 ocean state has not drifted far from observational estimates. We propose the *mesoscale dominance hypothesis* to interpret the relatively rapid thermal equilibration of CM4X-p125 to a cooler and more realistic pre-industrial state. Such ocean models result from negligible spurious mixing (from numerical truncation errors) along with an active mesoscale transport and realistic parameterization of small scale (diapycnal) mixing. Noting the preliminary nature of our results, and with caveats detailed in this paper, we suggest that the more rapid thermal equilibration possible from mesoscale dominant ocean models greatly reduces the computational energy footprint of models that are not mesoscale dominant.

1. Introducing This Paper and Its Thesis

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

We developed a climate model hierarchy, CM4X, to support research into the ocean's role in climate. CM4X is comprised of two coupled climate models: CM4X-p25 and CM4X-p125. These two models are identically configured in all aspects, except for their ocean (and sea ice) horizontal grid spacing and bottom topography. We document the development and performance of CM4X in this paper along with its companion in Part II (Griffies et al., 2025a).

1.1. Content of This Paper

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

We conjecture in Section 6 that the ocean component of CM4X-p125 approximates a novel class of ocean climate models that are *mesoscale dominant*, with mesoscale dominance described in the next subsection. We conclude the main portion of this paper in Section 7 by discussing the implications for climate modeling suggested by the CM4X results. Appendix A then details the CM4X model configuration, focusing on aspects that are distinct from the OM4.0 ocean/sea ice configuration of Adcroft et al. (2019) and the CM4.0 climate model of Held et al. (2019). In Part II of this paper (Griffies et al., 2025a) we detail results from a suite of case studies targeting areas of planned and ongoing research with the CM4X hierarchy.

1.2. Thermally Equilibrated piControl Simulations

We consider a thermally equilibrated piControl simulation to have a centennial scale top of the atmosphere radiative imbalance that is at least an order of magnitude smaller than estimates for the present-day radiative imbalance. Making use of the estimates summarized in Forster et al. (2021), the top of the atmosphere imbalance in a thermally equilibrated piControl simulation is no larger in magnitude than roughly 0.05 W m^{-2} to 0.1 W m^{-2}. Such imbalances are commonly realized by climate models only after $\mathcal{O}(1000)$ years. We question the physical relevance of the $\mathcal{O}(1000)$ timescale by presenting a model, CM4X-p125, that reaches thermal equilibration after $\mathcal{O}(100)$ years, with its equilibrated ocean state having less heat content than its initial state. The $\mathcal{O}(100)$ year timescale of CM4X-p125 corresponds to the time over which most of the anthropogenic heat has penetrated the upper portion of the ocean, thus providing a physically motivated timescale for piControl thermal equilibration.

1.3. Mesoscale Dominance as the Paper's Central Thesis

Mesoscale dominant ocean models are characterized by the following three mutually dependent physical and numerical properties: (a) an explicit representation of vigorous ocean mesoscale features (or with a suitable mesoscale eddy parameterization) that enables the stirring and restratification effects from mesoscale eddies; (b) an accurate parameterization of small scale processes that enable mixing to transform water mass distributions (both diapycnal and isopycnal mixing processes); (c) a low level of spurious numerical mixing that allows for the

Table 1
Acronyms Used in This Paper, Their Meaning, and Relevant Citation or Section

Acronym	Meaning	Citation or section
AMIP	Atmospheric Model Intercomparison Project	pcmdi.llnl.gov/mips/amip
AM4	GFDL Atmospheric Model 4.0	Zhao et al. (2018a, 2018b)
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. (2018a, 2018b)
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

Note. 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.

physical aspects of the mesoscale and small scale to be realized without aliasing into one another. We hypothesize that these three properties, in combination, form necessary (though not sufficient) conditions for a relatively rapid (centennial rather than millennial) thermal equilibration of the global climate system under piControl forcing, and with the piControl equilibrium ocean having less, not more, heat content than the early 21st century initial condition. Stated in a negative manner, we hypothesize that piControl thermal equilibration will require $\mathcal{O}(1000)$ years so long as the ocean model component offers a poor approximation to mesoscale dominance.

We are partly led to propose the mesoscale dominance hypothesis by acknowledging the history of ocean model developments that have focused on the importance of mesoscale transport, along with maintaining physically relevant diapycnal mixing and negligible spurious mixing. Within the context of mesoscale eddy parameterization, Solomon (1971), Redi (1982), Gent and McWilliams (1990) and Gent et al. (1995) offer a starting point for this line of research, whereby their formulations stress the need to parameterize the mesoscale eddy transport while respecting water mass boundaries. The CM4X-p125 results exemplify this approach through an explicit realization of the mesoscale (i.e., a mesoscale active model) rather than via its parameterization. It also does so by employing an ocean model component that pays particular attention to the numerical needs of reducing spurious mixing between water masses (e.g., Adcroft et al. (2019) and Griffies et al. (2020)), and to the physical needs of parameterizing diapycnal mixing processes, such as breaking internal gravity waves (e.g., MacKinnon et al. (2013) and MacKinnon et al. (2017)). Even so, we do not claim that CM4X-p125 perfectly realizes the three characteristics of a mesoscale dominant ocean. Indeed, as documented in Part II of this study (Griffies et al., 2025a), CM4X-p125 has many shortcomings. Rather, we contend that CM4X-p125 approaches these characteristics sufficiently to render a rapid thermal equilibration into a cooler pre-industrial ocean state, thus offering a taste for what is possible from mesoscale dominant ocean climate models.

We acknowledge the difficulty of quantitatively testing whether a model can be characterized as mesoscale dominant. Indeed, it might be most practical to infer relative scales of mesoscale dominance based on comparisons across a model hierarchy, such as we do for CM4X. Developing direct ways of characterizing mesoscale dominance remains a topic of ongoing research, with some suggestions offered in Section 6 based on water mass analysis. Even though difficult to test, we propose that the mesoscale dominance hypothesis provides a useful means to frame the question of climate model thermal equilibration. It also offers targets for model development

strategies. In particular, we emphasize the central role of reducing spurious numerical mixing in developing models capable of realizing mesoscale dominance. Namely, a strongly energetic ocean mesoscale must be realized with negligible spurious mixing if the model is to meet the needs of, and the benefits from, mesoscale dominance.

Our focus throughout this model description paper concerns the ocean and sea ice. However, we fully acknowledge that atmospheric and land radiative properties have a substantial impact on the simulation of SST, ocean volume mean temperature, and ocean heat uptake. Radiative properties, in turn, are significantly influenced by cloud parameterizations containing poorly constrained parameters. We document the tuning of CM4X cloud parameters in Section A1 of the appendix. Importantly, by setting all atmospheric model configuration details to be identical for CM4X-p25 and CM4X-p125, we enable a focus on the ocean's role in affecting thermal equilibration of the climate model, as well as other aspects of the simulation.

2. The CM4X Model Hierarchy

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

2.1. The Importance of Humility in Model Development

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

Throughout the development process, we found it crucial to maintain a balance between what is practical and what is aspirational. Along the way, we experienced the wisdom of Hofstadter's Law (Hofstadter, 1979), which, as paraphrased for the CM4X project, 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 6-month project expanded into a multi-year development effort. Part of the extended time frame arose from expanded visions for the project (i.e., “mission creep”), the entrainment of more individuals with unique ideas, interests, and scientific needs for an ocean eddying coupled model, and more simply, from our experience with an application of Murphy's Law: *Climate models are Murphy's Law Machines: anything that can go wrong will go wrong.*

2.2. Balancing the Aspirational With the Feasible

CM4X is designed for research concerning physical ocean and sea ice processes that manifest in regional and global weather and climate, extending in a seamless fashion from days (e.g., sea level and temperature extremes) to centuries (e.g., ocean overturning circulation variations and changes). A faithful numerical depiction of such phenomena requires coupled models with relatively fine grid spacing in both the horizontal and vertical. There is growing experience with the payoffs realized by refined ocean grid spacing, in particular for the representation of mesoscale processes and their role in long term climate such as ocean heat uptake (Griffies et al., 2015). In brief, transient ocean mesoscale eddies are generated by baroclinic instability of unstable waves (e.g., Vallis (2017)). This process affects a transfer from available potential energy to kinetic energy, and in so doing it restratifies the ocean by flattening isopycnals and thus throttling vertical transfer of buoyancy (which is generally dominated by temperature) within the ocean interior.

Additionally, refined ocean grid spacing and subsequent improved topographic representation are essential for the study of ocean extremes near the coasts. For example, when coupled to an ocean model that accurately represents the geometry of ocean continental shelves, climate models are key to the study of coastal ocean heat waves (Bian

et al., 2023). The study from Yin et al. (2020) pointed to the additional benefit of refined atmospheric grid spacing for capturing atmospheric cyclones largely responsible for extremes in coastal sea level. Indeed, Yin et al. (2020) and Zhao (2020, 2022b) demonstrate the enhanced realism of simulated tropical and mid-latitude cyclones when moving from a nominally 100 km (C96) version of the AM4 atmosphere (Zhao et al., 2018a, 2018b) used in CM4.0, to the nominally 50 km (C192) version of AM4 used by Zhao (2020). These studies, along with those by Hsieh et al. (2020), motivate us to use the C192 version of AM4 for CM4X. Improved simulations of atmospheric storms would also enable better exploration of the rectified impact of small-scale storms on climate, in particular tropical storms, although we do not claim that the model at its current resolution is near convergence.

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

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

2.3. Climate Model Hierarchy Based on Ocean Horizontal Grid Scale

Simulation realism, including a comprehensive suite of physical, chemical, and biological processes, often runs counter to the needs for unpacking the underlying mechanisms rendering scientific insights and robust understanding. In support of understanding, one is motivated to develop models that focus on specific processes and to compromise on comprehensiveness. As emphasized by Held (2005), Jeevanjee et al. (2017), and Maher et al. (2019), model hierarchies provide a means to span the spectrum between comprehensive and ideal. Such hierarchies support the science forming the basis for comprehensive earth system models, and enhance the scientific understanding that emerges from model simulations. A model hierarchy can be realized in a variety of manners, for example, see Figure 6 in Maher et al. (2019) for atmospheric model hierarchies. The two CM4X models differ only by their ocean and sea ice grid spacing, along with the representation of bottom topography. All other model details are identical. Hence, CM4X is a hierarchy of scale, with a focus on questions related to the ocean and sea ice.

2.4. CM4X Design Decisions

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

Further examples of ocean horizontal grid spacing hierarchies include the coupled climate model of Kirtman et al. (2012), which includes a 1° ocean (with mesoscale eddy parameterizations) and 0.1° ocean (without mesoscale eddy parameterizations). Roberts et al. (2019) document a hierarchy of coupled models from the Hadley Centre used as part of the HighResMIP project (Haarsma et al., 2016). The hierarchy of Kiss et al. (2020) shares much with CM2-O, yet with a focus on forced ocean/sea ice simulations using the JRA55-do atmospheric state from Tsujino et al. (2018). The Neverworld2 hierarchy of Marques et al. (2022) consists of idealized stacked

shallow water models using MOM6 and is designed to study adiabatic aspects of mesoscale eddies and their parameterization.

2.4.1. Concerning the Vertical Coordinates

In addition to horizontal grid spacing, Lindzen and Fox-Rabinovitz (1989) pointed to the importance of vertical grid spacing to ensure an accurate representation of the vertical structure of those quasi-geostrophic eddies admitted by the horizontal grid. K. Stewart et al. (2017) extended this analysis to include the specific needs of resolving vertical structure over the global ocean with its widely varying bottom topography. Xu et al. (2023) then pointed to the benefits of using an isopycnal vertical coordinate given that it naturally refines its vertical grid spacing in regions of strong water mass property gradients, such as occurs in mesoscale eddies. Following Adcroft et al. (2019), we employ 75 hybrid vertical coordinates for the CM4X ocean, which transition from geopotential-like vertical coordinates in the upper boundary region to isopycnal coordinates in the ocean interior. Hence, the representation of mesoscale eddies in the CM4X ocean benefits from the properties of isopycnal models as identified by Xu et al. (2023). In Appendix A9 we provide more details of the vertical coordinates used in the CM4X ocean.

2.4.2. Absence of Mesoscale Eddy Parameterzations in CM4X

We disabled mesoscale eddy parameterizations for CM4X, which also follows the choice of Adcroft et al. (2019) for OM4.0, as well as in the hierarchies of Griffies et al. (2015), Kirtman et al. (2012), Kiss et al. (2020), and Marques et al. (2022), for those configurations with 0.25° and finer grid spacing. Disabling mesoscale eddy parameterization helps to identify the physics associated with distinct representations of ocean fluid mechanics in a manner that does not also require us to account for the role of an eddy parameterization. This design choice is based on simplicity of analysis, clarity of the sensitivity to grid changes, and, we hope, to enhanced robustness of the results across models. However, this choice in no way presumes that the ocean mesoscale is accurately or completely represented by either the 0.25° or 0.125° grids in CM4X (e.g., see Figure 1). Indeed, many authors to this paper are actively pursuing research into parameterizations suited for the eddy active "gray zone" regime of CM4X. This regime remains challenging for the ocean modeling community, similar in vein to the atmospheric modeler's quest for storm-resolving resolutions, yet it represents the foreseeable future for GFDL coupled models. The CM4X hierarchy informs this parameterization research, in particular by emphasizing the role of ocean mesoscale eddies in the thermal equilibration of piControl simulations.

2.4.3. Component Model Simplifications to Align With CM4X Science Goals

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

2.4.4. CM4X Is Not a Comprehensive CMIP6-Class Climate Model

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

Certain of the CM4X design choices, particularly those associated with the land model, preclude CM4X from being considered a comprehensive CMIP6-class climate model, whereas CM4.0 directly targeted CMIP6 questions. This limitation is important to acknowledge when comparing CM4.0 and CM4X during the historical simulation since CM4.0 has the CMIP6 land forcings whereas CM4X does not. Simplifying the climate model by

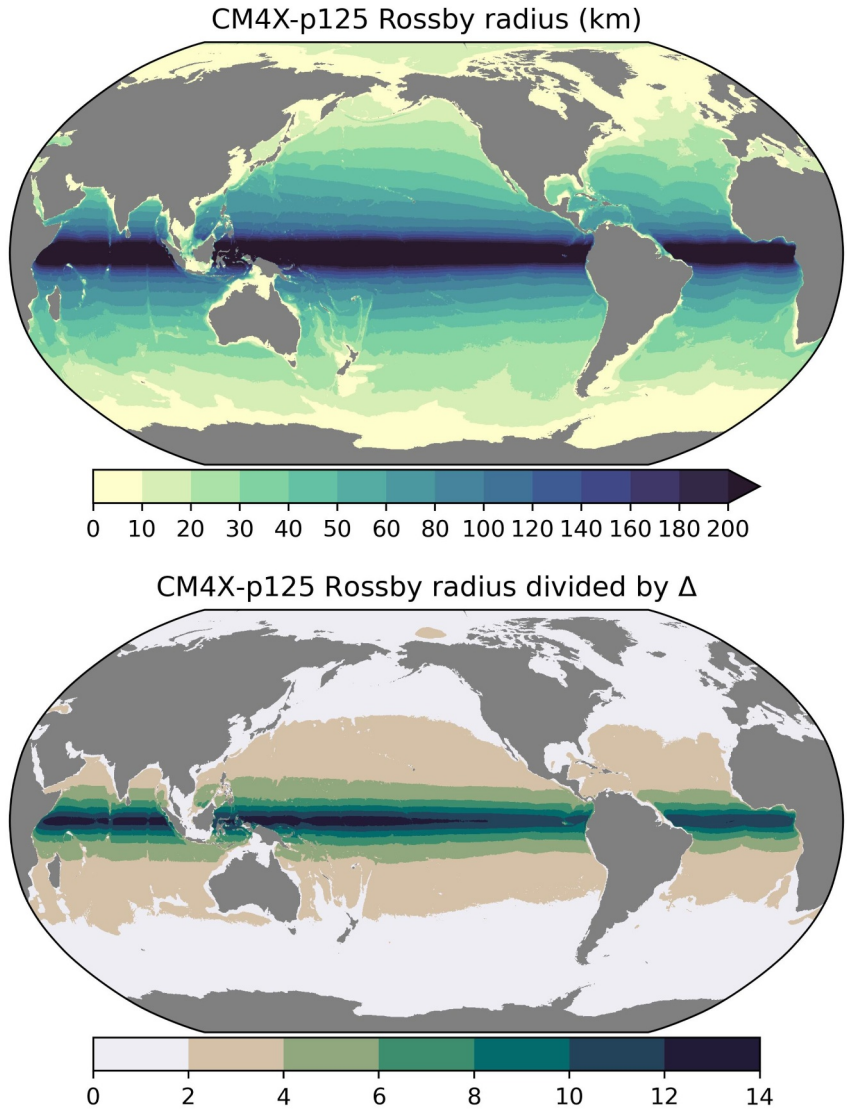

Figure 1. First panel: first baroclinic Rossby radius, L_d, computed according to Hallberg (2013). The deformation radius is computed as a time average over years 1980–2009 in the historical CM4X-p125 simulation. Second panel: ratio of the first baroclinic Rossby radius to the horizontal grid spacing, 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 $L_d/\Delta \geq 2$.

disabling elements of the land model reduced the risk of having the perfect be the enemy of the good in our task of developing a fit-for-purpose coupled climate model hierarchy focused on ocean and sea ice physics.

3. A Survey of CM4X Results

We here offer a selection of results that highlight facets of the model simulations using pre-industrial, historical, and future scenario experiments from CMIP6 (Eyring et al., 2016). Further details for many of these diagnostics are presented in the various case studies of Part II (Griffies et al., 2025a).

3.1. Experiments, Initialization, and Analysis

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

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

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

For CM4X-p125, we took initial conditions for the land model from CM4X-p25 as remapped onto the CM4X-p125 land/sea mask grid. Sea ice, iceberg, and calving initial conditions for CM4X-p125 and CM4X-p25 were taken from prototype configurations that ran for 100 and 70 years, respectively, under 1950 radiative forcing conditions. This choice avoids "cold starting" the sea ice component, which can introduce large initial salinity and temperature pulses into the high-latitude oceans due to sea ice formation. In both CM4X-p25 and CM4X-p125, initial conditions for sea-ice concentration (SIC) and sea ice volume reasonably agree with present-day observations. The initial Pan-Arctic sea ice extent is 16.6×10^6 km^2 in CM4X-p125 and 14.4×10^6 km^2 in CM4X-p25, and the initial Pan-Antarctic sea ice extent is 11.4×10^6 km^2 (CM4X-p125) and 7.9×10^6 km^2 (CM4X-p25). The corresponding present-day observed values are approximately 15×10^6 km^2 and 6×10^6 km^2 in the Arctic and Antarctic, respectively. Also note that the sea ice state loses memory of its initial conditions over a 2–3 years timescale (Blanchard-Wrigglesworth et al., 2011; Zampieri et al., 2021), thus for long pre-industrial control simulations this level of initial agreement with observations is acceptable.

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

3.2. Climate Modeling With an Active Ocean Mesoscale

Hallberg (2013) as well as Griffies and Treguier (2013) noted that the first baroclinic 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 Lindzen and Fox-Rabinovitz (1989), K. Stewart et al. (2017), and Xu et al. (2023) 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 (including deep water formation regions) 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 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 daily mean dynamic sea level. According to the Hallberg (2013) criteria, the Rossby radius is unresolved by either model in the high latitudes

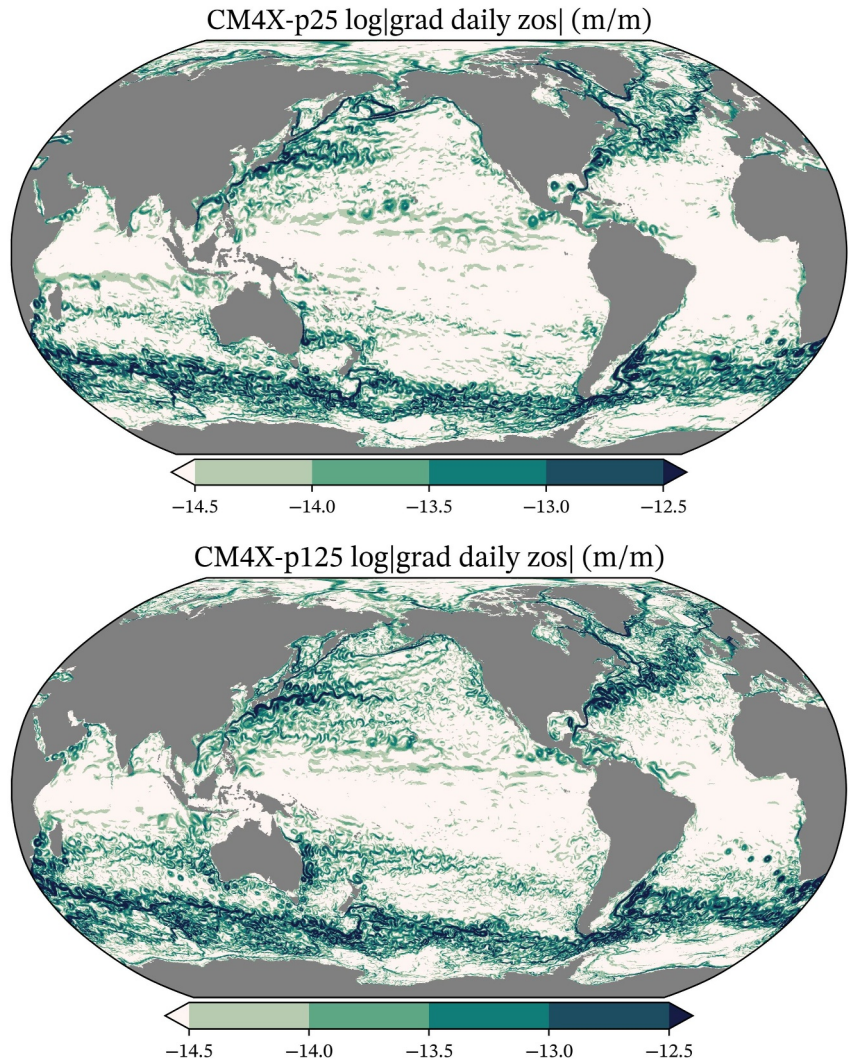

Figure 2. Magnitude of the gradient of the daily mean dynamic sea level (| ∇ (zos)|) for CM4X-p25 (first panel) and CM4X-p125 (second panel), as computed from 31 December 2014 in the historical simulations. Meandering mesoscale jets and eddies are represented in these simulations even though the models 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).

(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 nonlinear 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.

3.2.1. Eddy Activity as Revealed by Sea Level Variability

Through the connection between sea level and surface geostrophic currents, sea level variability provides an effective spatial measure of mesoscale eddy activity in the surface ocean in off-equatorial regions. We show in Figure 3 the standard deviation (root-mean-square) of the daily mean sea level relative to a daily climatological

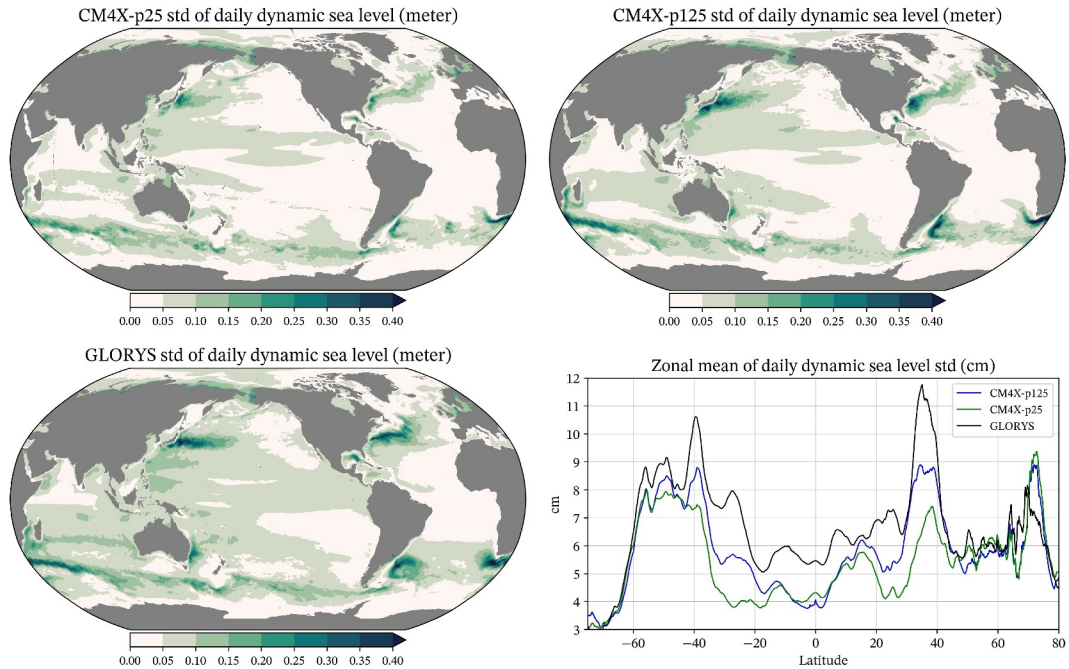

Figure 3. Standard deviation for the daily mean dynamic sea level computed over years 1995–2014, showing CM4X-p25, CM4X-p125, and the 1/12° GLORYS12 analysis product (Lellouche et al., 2021), along with the zonal mean for the three. The standard deviations are computed relative to a daily climatological mean, with full details of the standard deviation calculation provided in Section 2 of Part II (Griffies et al., 2025a).

mean. This figure reveals the expected increase in variability when moving from CM4X-p25 to CM4X-p125, given the more active mesoscale eddies admitted by the finer grid spacing. This increased variability is revealed within the boundary current extensions as well as the Southern Ocean. For example, the Kuroshio Current in the west Pacific is notably stronger and extends further eastward in CM4X-p125, making it more consistent with the GLORYS12 analysis product. The Gulf Stream variability also increases in CM4X-p125.

Although CM4X-p125 shows more eddy variability than CM4X-p25, it is still far weaker than the GLORYS12 analysis product from Lellouche et al. (2021). Furthermore, there is a notable bias in both CM4X simulations in which the Gulf Stream exhibits broad and diffuse variability close to the American coast, which contrasts to GLORYS12 in which variability extends further eastward and along a more localized frontal region. We have more to say about deficiencies of the Gulf Stream in Sections 2 and 6 of Part II (Griffies et al., 2025a).

3.2.2. Enhanced Role for Boundary Currents

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

3.2.3. Barotropization Particularly in the High Latitudes

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), Arbic and Flierl (2004), 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 latitudes is mostly baroclinic, including the

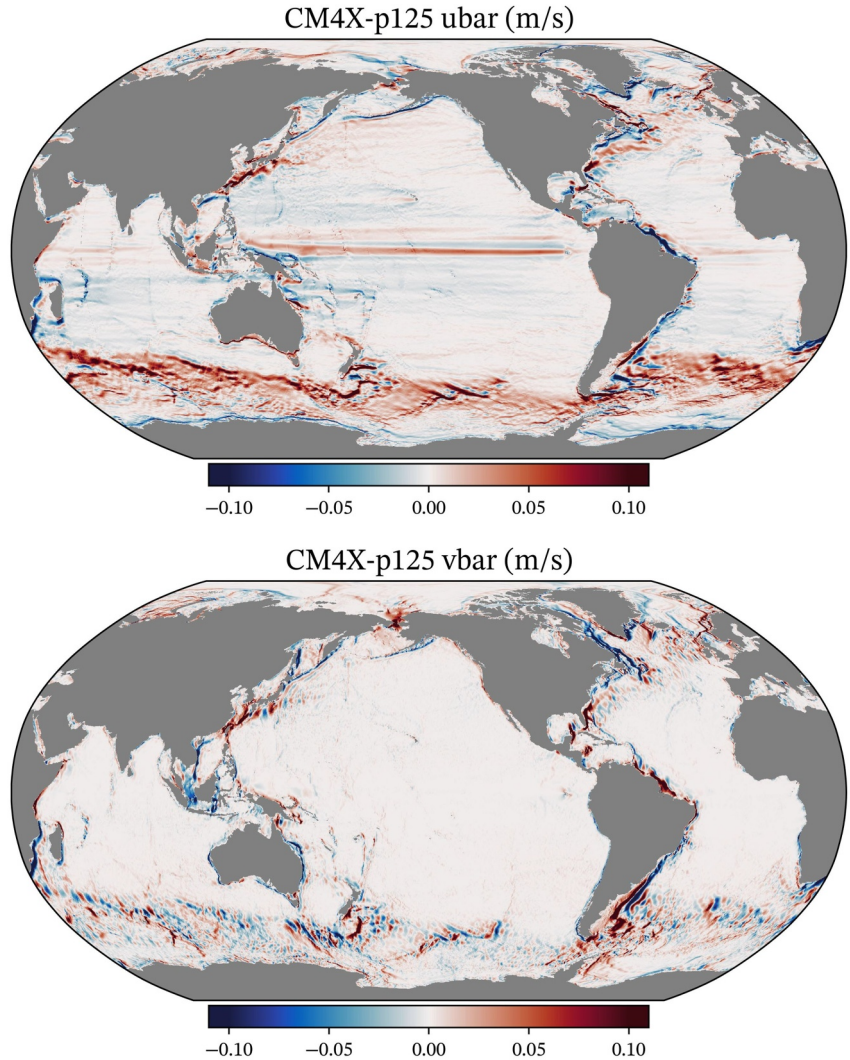

Figure 4. Depth averaged zonal velocity and meridional velocity for CM4X-p125, as computed from the time mean over model years 1980–2009 in the historical simulations. Note the boundary currents as well as the flows at the high latitudes. Additionally, along the equator, we find the zonal flows comprising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

equatorial currents. In contrast, high latitude currents contain a sizable fraction of their kinetic energy in barotropic motions, with this property related to the relatively weak vertical stratification found particularly in the high latitude Southern Ocean and North Atlantic. Strong barotropic motions are constrained (at least in part) by the need to preserve planetary geostrophic potential vorticity; that is, to follow contours of constant f/H, where f is the Coriolis parameter and H is the bottom depth. This property of such flows emphasizes the importance of topographic representation, as well as for capturing the strong flows that interact with the bottom. In Appendix A8 and A9 we provide details for the representation of the ocean bottom in the CM4X models.

3.3. Northward Ocean Heat Transport

In Figure 6 we display the northward ocean heat transport (referenced to 0°C), as diagnosed from the CM4.0, CM4X-p25, and CM4X-p125 simulations averaged over simulation years 1980–2009, as well as the reanalysis-based results from Trenberth and Caron (2001) and Trenberth et al. (2019) (with updates to 2023), along with the ocean in situ based analysis of Ganachaud and Wunsch (2000). The model results are closer to one another than to the reanalysis and in situ measures, and generally show a somewhat reduced magnitude for poleward transport. Furthermore, the model results are quite close in the northern hemisphere, whereas CM4X-p125 exhibits somewhat more southward transport away from the tropics in the Indian-Pacific and global transports,

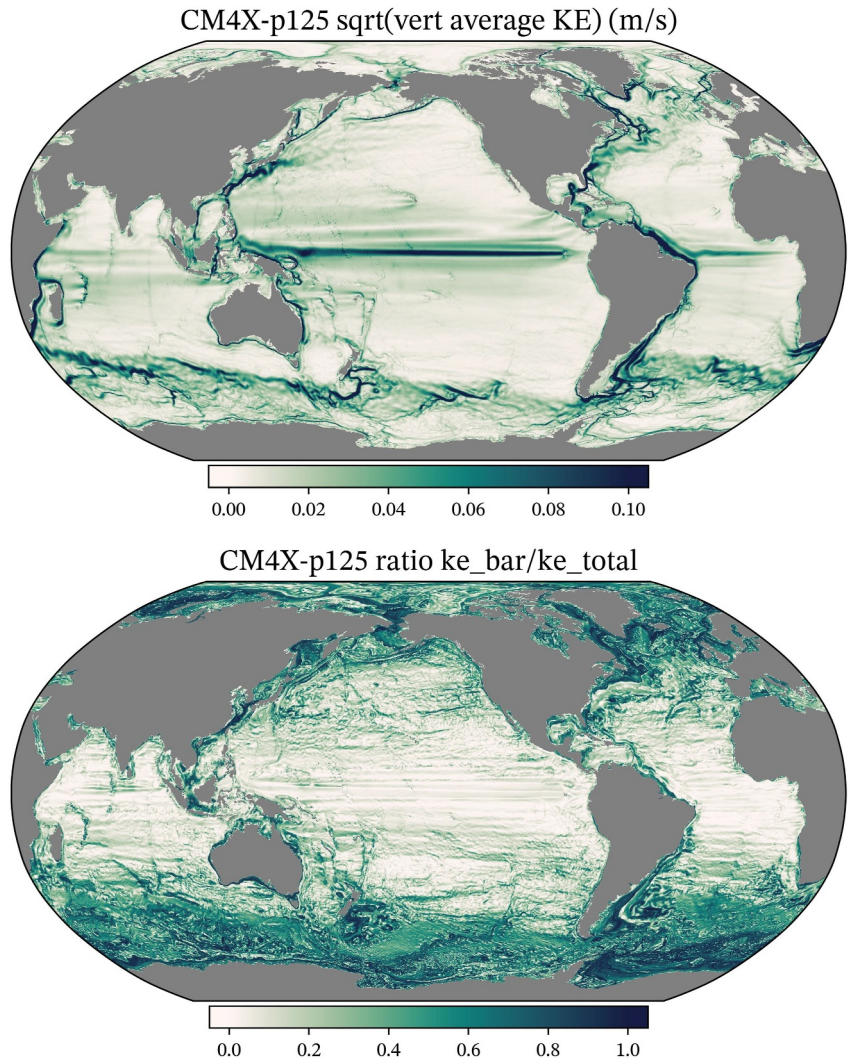

Figure 5. First panel: square root of the depth averaged kinetic energy per mass, $1/2 \sum \Delta z \langle \boldsymbol{u} \rangle \cdot \langle \boldsymbol{u} \rangle / \sum \Delta z$, as computed using the 1980–2009 (in the historical simulations) time mean horizontal velocity for CM4X-p125, $\langle \boldsymbol{u} \rangle$. Second panel: ratio of the kinetic energy contained in the depth averaged flow, $1/2\, \overline{\langle \boldsymbol{u} \rangle} \cdot \overline{\langle \boldsymbol{u} \rangle}$, to the total kinetic energy. Here, we use the depth and time averaged flow as given by $\overline{\langle \boldsymbol{u} \rangle} = \sum \Delta z \langle \boldsymbol{u} \rangle / \sum \Delta z$. 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.

presumably due to its stronger mesoscale eddies. Yet overall we conclude that the CM4.0 and CM4X models are closely aligned in their poleward heat transports. This result was also found in the CM2-O hierarchy by Griffies et al. (2015), where it was noted that the structure of the poleward ocean heat transport is dominated by the atmosphere, whereas the vertical heat transport, and hence the global balance of heat within the climate system, is dominated by ocean processes that affect a vertical exchange of heat, such as diapycnal mixing and mesoscale eddy transport.

3.4. Time Scales for piControl Thermal Equilibration

The figures shown thus far in this section reveal that both CM4X-p25 and CM4X-p125 capture a nontrivial amount of mesoscale eddy activity (albeit weaker than the GLORYS12 analysis seen in Figure 3) as well as strong boundary currents, with this assessment relative to non-eddying one-degree models. A key question concerns the degree to 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

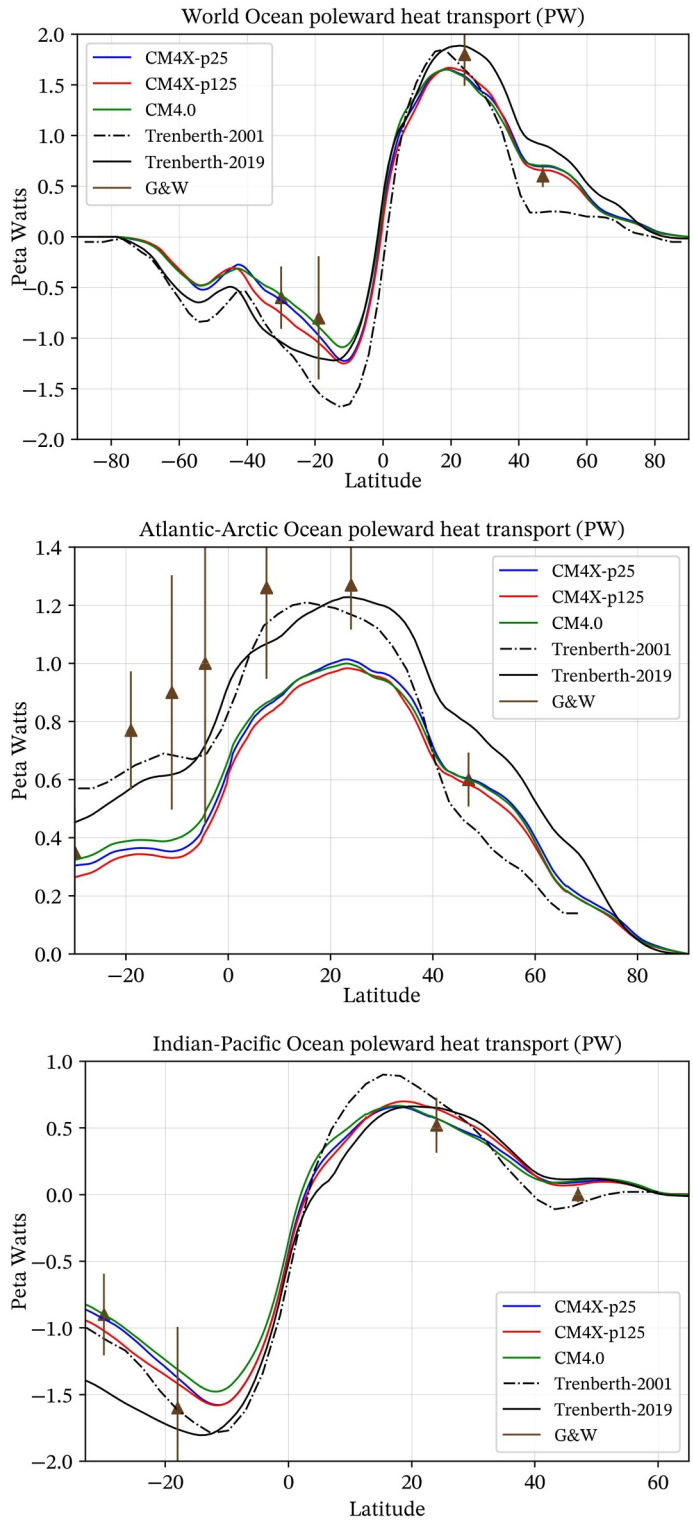

Figure 6. Northward ocean heat transport (referenced to 0°C) for the World Ocean, Atlantic-Arctic Oceans, and Indian-Pacific Oceans. We show results averaged over years 1980–2009 from CM4X-p25, and CM4X-p125. We compare these model results to the implied ocean heat transports from Trenberth and Caron (2001) and Trenberth et al. (2019) using reanalysis products, 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.

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 7c 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., 2025a), we find no such fluctuations in either CM4X-p25 or CM4X-p125.

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. We do so since 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 ± 0.1 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.

We offer the mesoscale dominance hypothesis in Section 6 as a means to frame the results shown in Figure 7. This hypothesis proposes that ocean models with an active mesoscale eddy transport, an accurate parameterization of small scale mixing, and a negligible degree of spurious numerical mixing are together necessary (though not sufficient) to realize centennial scale thermal equilibration in piControl simulations, and with that equilibration being to a cooler piControl ocean state.

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; Hirschi et al., 2020; A. L. Stewart et al., 2021; R. Zhang et al., 2019), with those plans guiding the case studies presented in Part II (Griffies et al., 2025a). 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

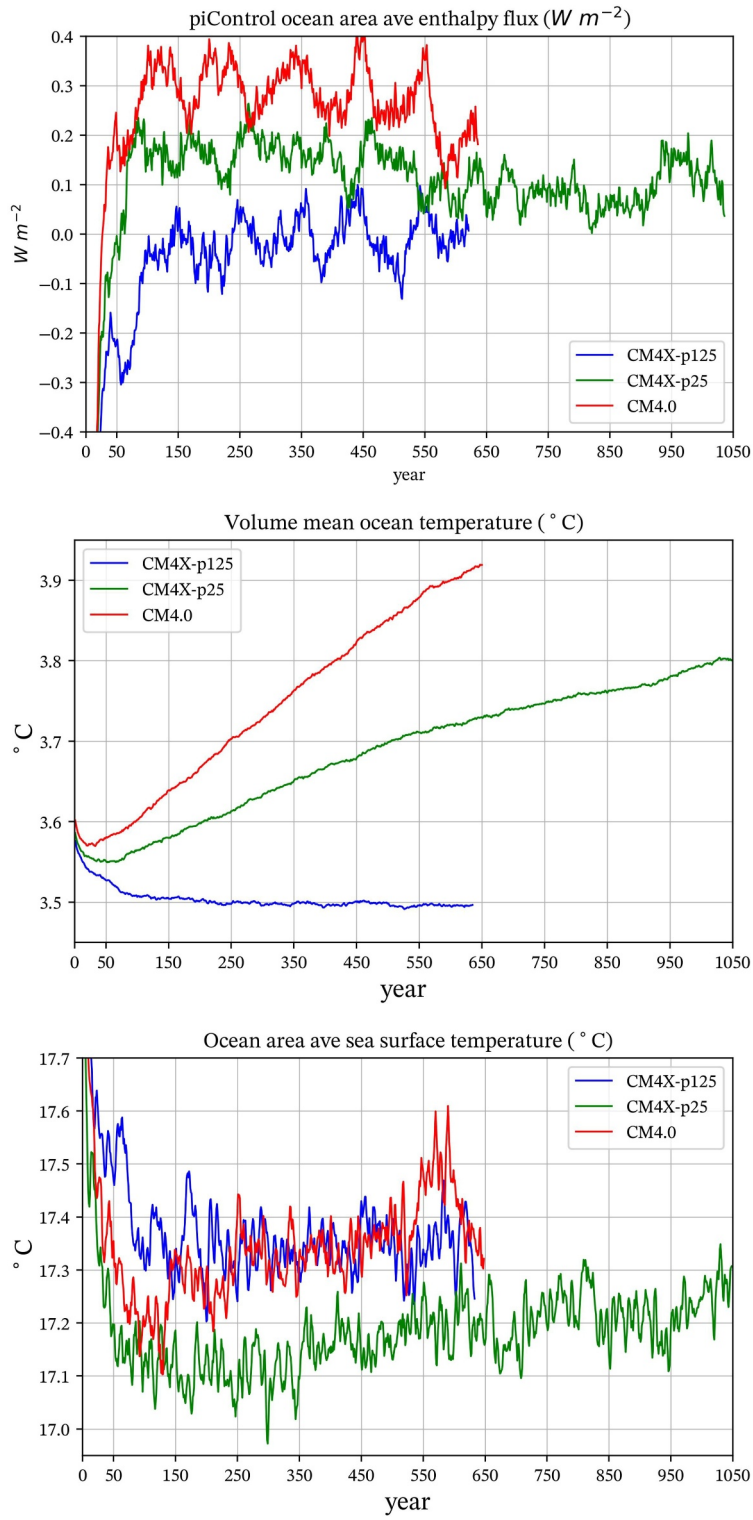

Figure 7. Ocean model diagnostics from the piControl simulations with CM4.0, CM4X-p25, and CM4X-p125. First panel: net ocean area integrated boundary heat flux entering the ocean, including surface boundary (turbulent fluxes, radiative fluxes, enthalpy associated with mass transfer, and frazil sea ice 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. Second panel: time series for the volume mean and annual mean global ocean Conservative Temperature for CM4X-p25, CM4X-p125, and CM4.0. Third panel: time series for the ocean area mean 5-year rolling mean sea surface temperature (temperature in the top model grid cell).

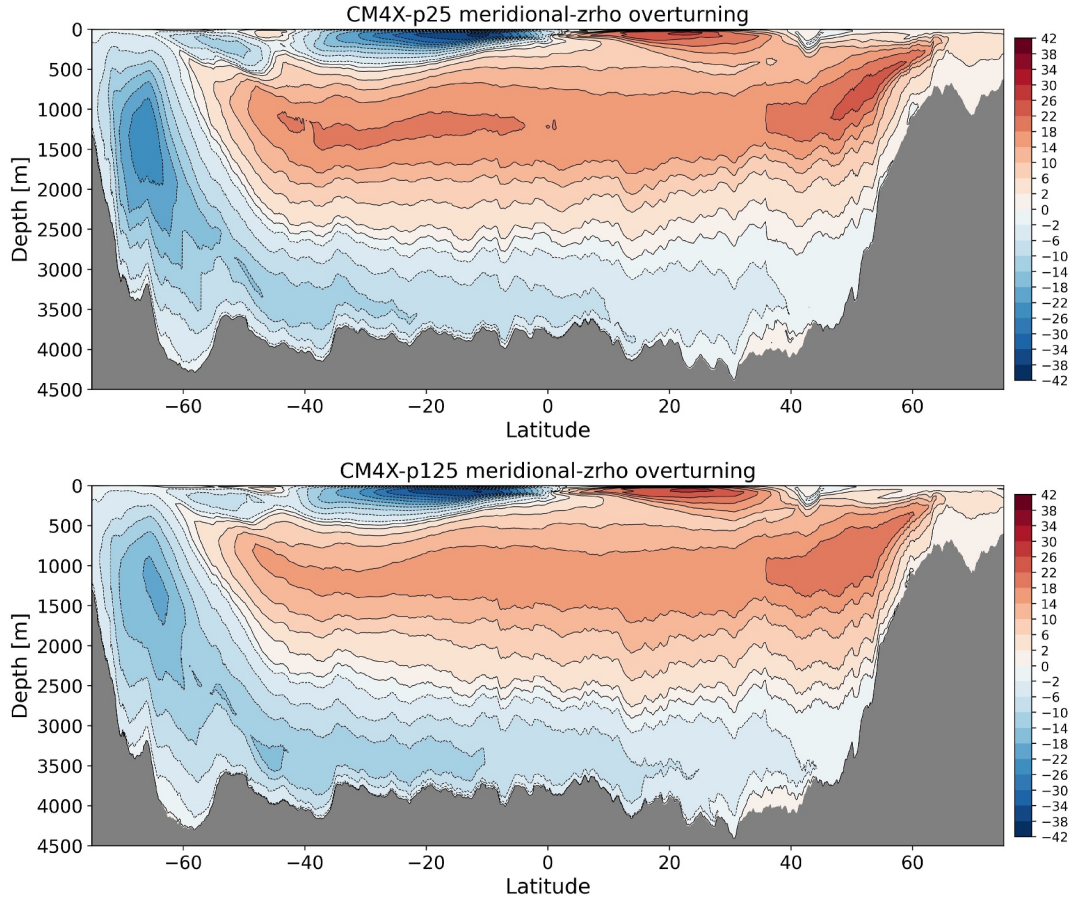

Figure 8. Meridional-density overturning circulation from CMX-p25 and CM4X-p125 as a function of latitude and potential density referenced to 2000 dbar: $\psi(y,\sigma,t) = -\rho_0 \oint \left[\int_{z'=-H}^{z'=z_\sigma} v(x,y,z',t)\,\mathrm{d}z'\right]\mathrm{d}x$, where $\rho_0 = 1035$ kg m^{-3} is 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 $\left(10^9 \text{ kg s}^{-1}\right)$. 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.

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. It is notable that the somewhat weaker overturning in CM4X-p125 is contrary to the expectation based on Hirschi et al. (2020), who find a general deepening and strengthening of the overturning with refined grid spacing.

Furthermore, Yeager et al. (2021) find that the AMOC is weaker and less dense in a high resolution coupled model than that simulated in a low resolution counterpart, and the high resolution results compare better with observations. The stronger and denser AMOC in the low resolution model is related to its overestimated Labrador Sea deep convection compared to the improved Labrador Sea deep convection in the high resolution model. CM4X-p125 also has an improved and reduced Labrador Sea deep convection compared to that simulated in CM4X-p25, which may contribute to the weaker AMOC in CM4X-p125. The different results across different models reflect the complexity of the overturning, which involves a huge suite of processes that makes it difficult to expose a single mechanism. We have more to say concerning the Atlantic and Southern Ocean overturning circulation in Sections 5 and 6 of Part II (Griffies et al., 2025a).

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 CM4X-p25. Among the regions of notable biases, we see overly deep

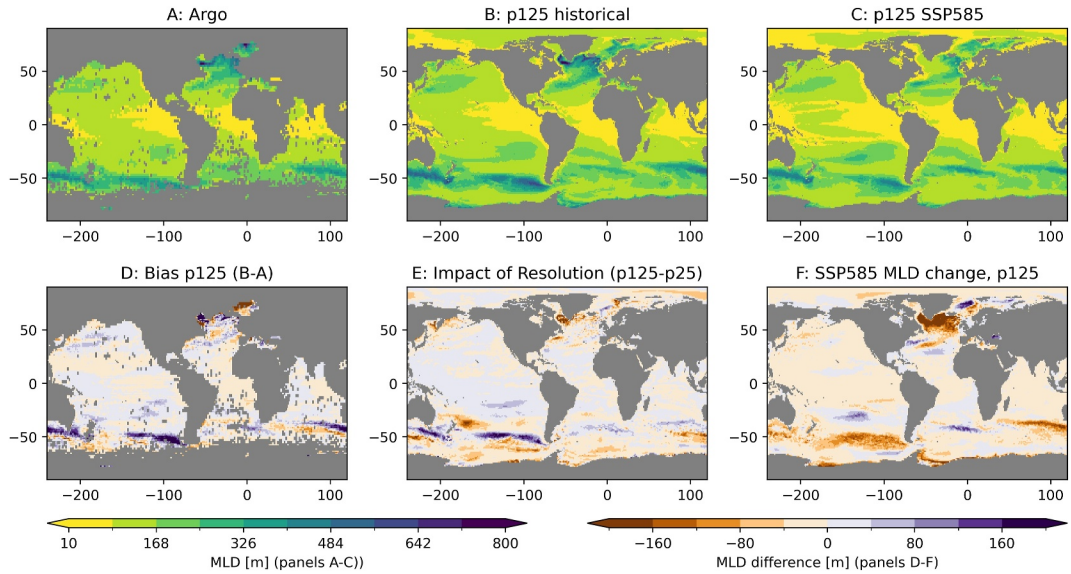

Figure 9. Maximum value of the monthly mean mixed layer depth climatology from Argo (years 2004–2023) (Argo, 2023), as well as years 2000–2014 (historical simulation) and years 2085–2099 (SSP5-8.5 simulation) from CM4X-p25 and CM4X-p125. The mixed layer depth is diagnosed according to the potential energy 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, e) are robust to a longer time average over model years 1955–2014.

mixed layer depths in CM4X-p125 in the mode and intermediate water regions of the Southern Ocean, as well as in the Labrador Sea (Figure 9d). Biases in the Southern Ocean ventilation for CM4.0 are the subject of Krasting et al. (2024), with these biases also reflected in the CM4X simulations. We find a notable reduction in the Labrador Sea bias moving from CM4X-p25 to CM4X-p125, presumably due to the stronger mesoscale eddies contributing to more restratification in CM4X-p125 (Figure 9e) (J. Marshall & Schott, 1999). Even so, the overly deep mixed layer biases in the Southern Ocean worsen in CM4X-p125, possibly due to the eddy-induced deepening effects in the Indo-West Pacific region described by Q. Li and Lee (2017). The SSP5-8.5 scenario experiment finds a systematic reduction in the mixed layer depths throughout most of the high latitude Southern Ocean and North Atlantic (Figure 9f).

3.6. Tropical Variability

Figure 10 shows the power spectra for the Niño-3 SST index from the historical simulation (years 1850–2014) with comparison to the NOAA reconstructed product over years 1854–2014 from Huang et al. (2017). Although there is much spread in the spectrum, the CM4X models show consistently weaker spectral power than the reconstruction across most time scales. Weak tropical variability is also reflected in the skewness of the sea surface height (Figure 3 in Part II of Griffies et al. (2025a)).

Interannual tropical variability can exhibit rather sizable centennial scale modulations (Wittenberg, 2009), so that quasi-stationary power spectra result only after running multi-centennial scale coupled simulations and/or ensembles. Even so, we consider the variability in CM4X to be notably weaker than CM4.0, and we hypothesize that this weak variability is related to changes in the ocean mixing parameterizations discussed in Appendix A. This hypothesis is prompted by noting that the tropical variability is stronger (consistent with CM4.0 as documented in Section 5.5 of Held et al. (2019)) when the C192 atm (used here for CM4X) is coupled to the OM4.0 ocean for a 150-year piControl experiment (not shown). Unfortunately, we could not afford to run long simulations to examine every separate change made to the ocean physical parameterizations. Consequently, we are unaware of the precise changes that led to the weaker variability. Unraveling this story remains an ongoing process.

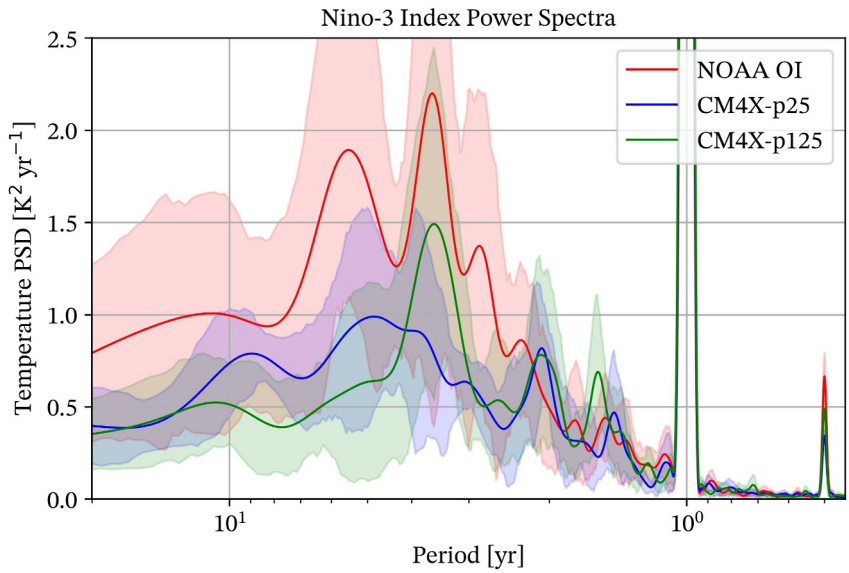

Figure 10. Power spectra for the Niño-3 index, which is computed from the monthly mean sea surface temperature area averaged over the region (5 °N − 5 °S, 150 °W − 90 °W) in the Tropical Pacific. We show results from the CM4X simulations over years 1850–2014, as well the reconstructed NOAA temperatures 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 shaded confidence intervals.

3.7. Coupled Climate Benchmarks

In Figure 11, we present benchmark performance metrics for years 1978–2014 of the historical simulations in the GFDL earth system model ESM4.1 (Dunne et al., 2020), along with the climate models CM4.0, CM4X-p25, and CM4X-p125. We compare models with observations using the PCMDI Metrics Package version 3 (Lee

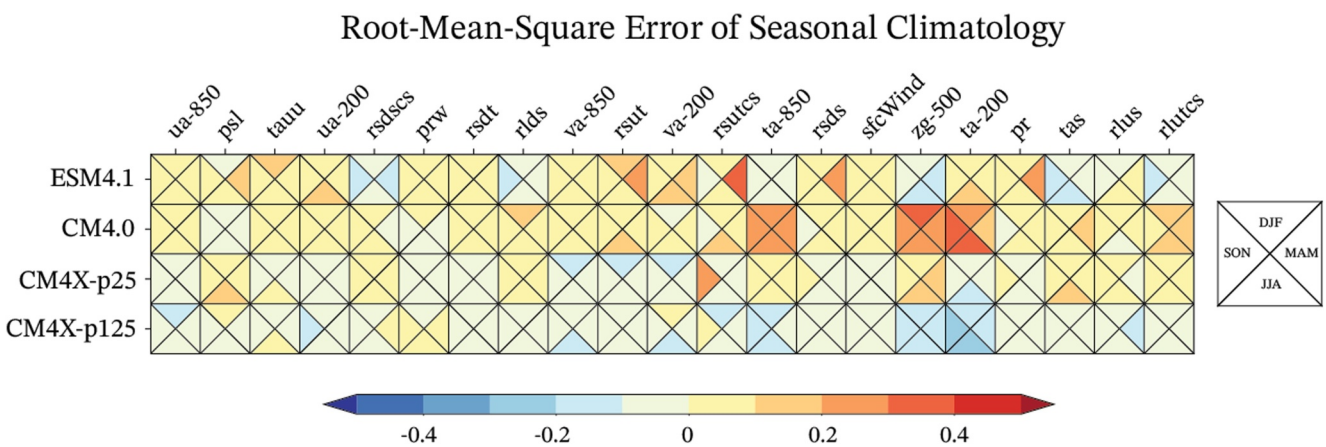

Figure 11. Performance metric portrait plot taken after Gleckler et al. (2008), here focused on atmospheric metrics from the ESM4.1, CM4.0, CM4X-p25, and CM4X-p125 historical simulations over years 1978–2014 of the historical simulations. The root-mean-square-error (RMSE) values are normalized across the models shown here, so we are showing relative error rather than absolute error. As per Table 1 in Lee et al. (2024), 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. That is, we show four models in this portrait plot with biases relative to the reanalysis products as detailed in (Lee et al., 2024) and colors relative to the four models shown. Variables labeled with 200, 500, and 850 are defined at pressure levels 200 hPac, 500 hPac, and 850 hPac. Further names refer to the following: ua = zonal wind, psl = sea level pressure, tauu = surface zonal wind stress, rsdscs = surface downwelling shortwave radiation under clear-sky conditions, prw = precipatable water, rsdt = top of the atmosphere incident shortwave radiation, rlds = surface downwelling longwave radiation, va = meridional wind, rsut = upwelling shortwave at top of the atmopshere, rsutcs = clear sky rsut, ta = air temperature, rsds = surface downwelling shortwave radiation, sfc Wind = surface wind speed, zg = geopotential height, pr = precipitation rate, tas = 2 m air temperature, rlus = surface upwelling longwave radiation, rlutcs = top of atmosphere longwave radiation under clear sky conditions.

et al., 2024) and present the results using the "portrait plot" format described in Gleckler et al. (2008). One caveat for these results is that we only have one ensemble member for each model, though a single realization is generally reliable for assessing these climatological biases. As noted by Lee et al. (2024), the GFDL models ESM4.1 and CM4.0 are among the top-performing CMIP6 models for these atmospheric benchmarks, thus providing a useful benchmark for the CM4X simulations.

Both CM4X-p25 and CM4X-p125 make use of a finer grid for the GFDL-AM4.0 atm relative to CM4.0 (C192 vs. C96), which we propose accounts for an overall reduction in model biases in both CM4X configurations compared to ESM4.1 and CM4.0. Furthermore, note how CM4X-p125 exhibits improved performance relative to CM4X-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 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. (2024), 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 SST and related boundary fluxes.

4.1. SST Variability and Turbulent Heat Fluxes

In Figure 12 we show one measure of transient variability in CM4X-p25 and CM4X-p125, 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.

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 SST 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 in Figure 13 are dominated by correlations indicating that the atmosphere is forced by the ocean. Furthermore, the patterns are quite similar for the two CM4X models, which we expect since they both exhibit similar patterns of SST variability (Figure 12) and they use the same atmosphere model. Also observe that the polar regions show a large correlation and yet very little temperature variation. This behavior reflects the dominance of sea ice formation/melt that produce latent heat fluxes yet little temperature changes.

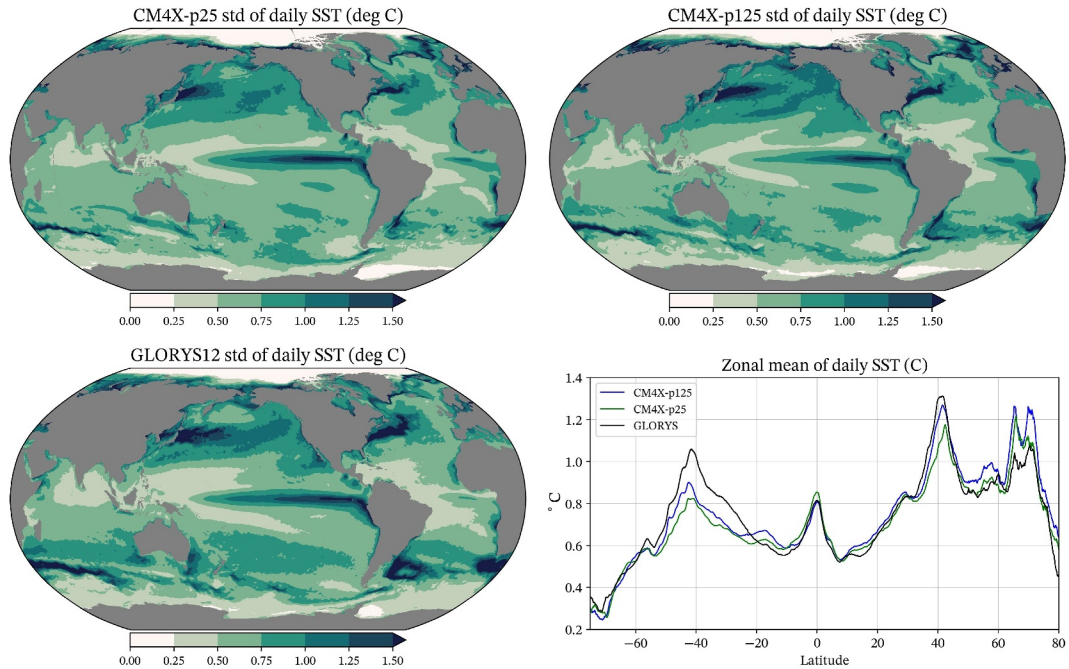

Figure 12. Standard deviation (°C) of the sea surface temperature (SST) in CM4X-p25 and CM4X-p125 as computed over years 1980–2009, as well as the 1/12° GLORYS12 analysis product (Lellouche et al., 2021) along with the zonal mean for the three. The standard deviation is computed by taking the mean squared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

4.2. Historical Sea Surface Temperature Biases

Figure 14 shows the SST biases computed from the 30-year time mean of the simulations over years 1980–2009. There are many common biases across the models, such as the strong cold bias in the region east of Greenland and the warm bias in the Labrador Sea. We hypothesize that these North Atlantic biases arise, at least in part, from relatively poor representation of the mesoscale eddy processes and overflow processes (R. Zhang et al., 2011). The eastern boundary regions off California as well as West Africa south of the equator both show warm biases that generally reflect biases in the stratocumulus clouds (e.g., see Wang et al. (2014), Richter (2015)). Notably, however, there is only a modest bias off the west coast of South America in the two CM4X simulations, which is significantly smaller than those found by Wang et al. (2014) and Richter (2015) in earlier climate models, and smaller than the CM4.0 bias in this region. This improvement is presumably due to the finer atmospheric grid in CM4X relative to CM4.0, though some studies suggest that finer ocean grid spacing can help reduce the stratocumulus cloud biases by enhancing coastal upwelling (Richter, 2015). We have more to say about these eastern boundary regions in Section 3 of Part II (Griffies et al., 2025a).

Many of the SST biases in CM4X are broadly reflected in the CM4.0 biases, with the global root-mean-square bias of 0.92°C in CM4X-p25 close to the 0.90°C found in CM4.0. However, in nearly all regions in the World Ocean, SST biases in CM4X-p125 are reduced relative to both CM4X-p25 and CM4.0, as reflected by the roughly 20% reduction in the global root-mean-square bias (0.73°C for CM4X-p125) and 40% reduction in the mean bias (−0.51° for CM4X-p25 and −0.29° for CM4X-p125). This bias reduction is generally realized by a broadly warmer SST in CM4X-p125, thus reducing, relative to CM4X-p25, the broad-scale cool bias in the Southern Hemisphere middle latitudes. Furthermore, it is notable that both CM4X models display only modest biases in the Southern Ocean relative to the larger warm bias in CM4.0 and even larger warm biases found in many other climate models (Beadling et al., 2020; Hyder et al., 2018; Sallée et al., 2013; Wang et al., 2014).

CM4X shows a general improvement with the overly cool eastern equatorial Pacific. This bias has been persistent in models, and is important due to its impact on the Hadley cell (G. Li et al., 2015). The Pacific subtropical gyre bias is also reduced, which supports a hypothesis from Burls et al. (2016) and Thomas and Federov (2017) that the equatorial and subtropical biases are connected. Finally, note the reduced bias in CM4X-p125 found in the

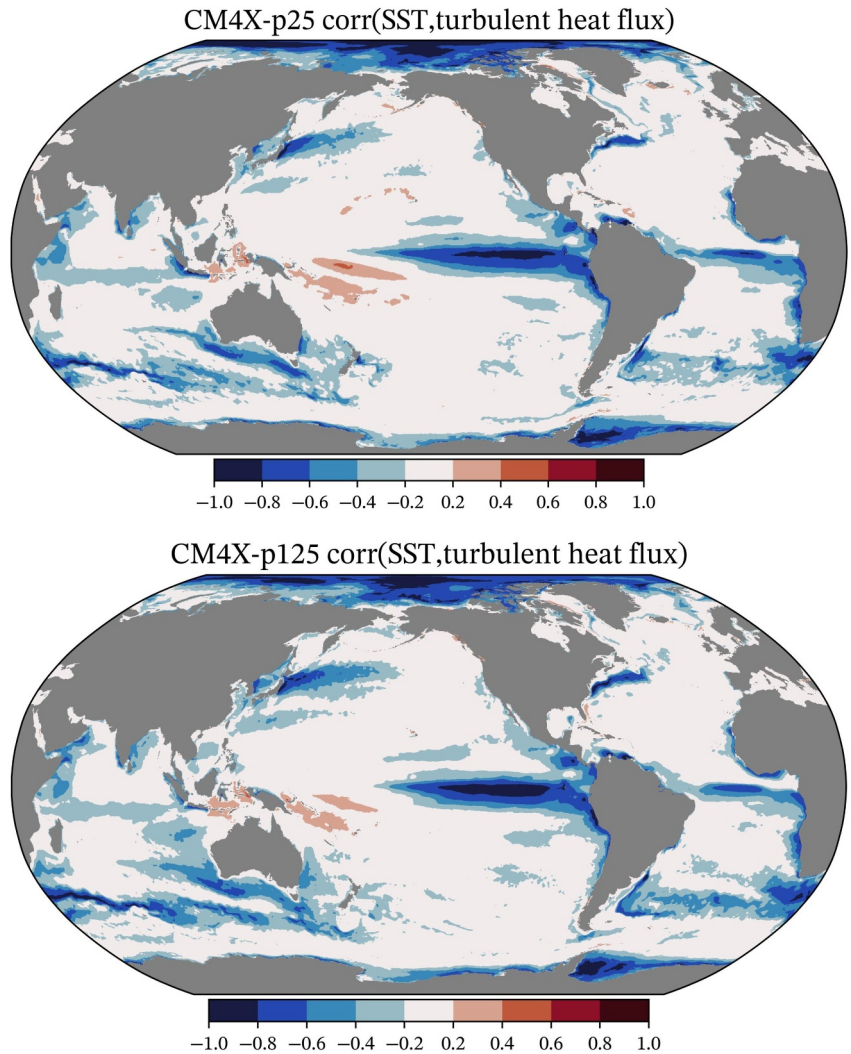

Figure 13. Temporal correlations between the monthly sea surface temperature (SST) and the turbulent surface heat fluxes (latent plus sensible). This correlation is computed over years 201–250 from the piControl simulations for CM4X-p25 and CM4X-p125, with the SST and heat fluxes each computed relative 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.

western Atlantic near the U.S. coast, whereby the strong warm bias in CM4X-p25 (also seen in CM4.0) has changed to a modest cool bias in CM4X-p125. This reduced bias reflects differences in the Gulf Stream as represented in CM4X-p25 and CM4X-p125. We provide more analysis of the Gulf Stream in Section 6.2 of Part II (Griffies et al., 2025a), where we identify a number of shortcomings in both CM4X models.

4.3. Patterns of SST Change Under SSP5-8.5

Figure 15 shows the change in SST under the SSP5-8.5 scenario. We compute this pattern as the difference between the final 10 years of the scenario (years 2090–2099) and the 30-year mean over years 1980–2009 of the historical. There is broad agreement between CM4.0 and the two CM4X models, showing a general warming pattern in lower to middle latitude open ocean. A notable exception is the muted warming in the southeast and south central Pacific, which has been seen in historical changes and with this pattern suggested to be related to changes in the Walker Circulation (Wills et al., 2022). However, a systematic comparison of the historical SST trends would require multiple realizations for the historical period. The North Pacific shows general warming, with CM4X-p125 somewhat larger.

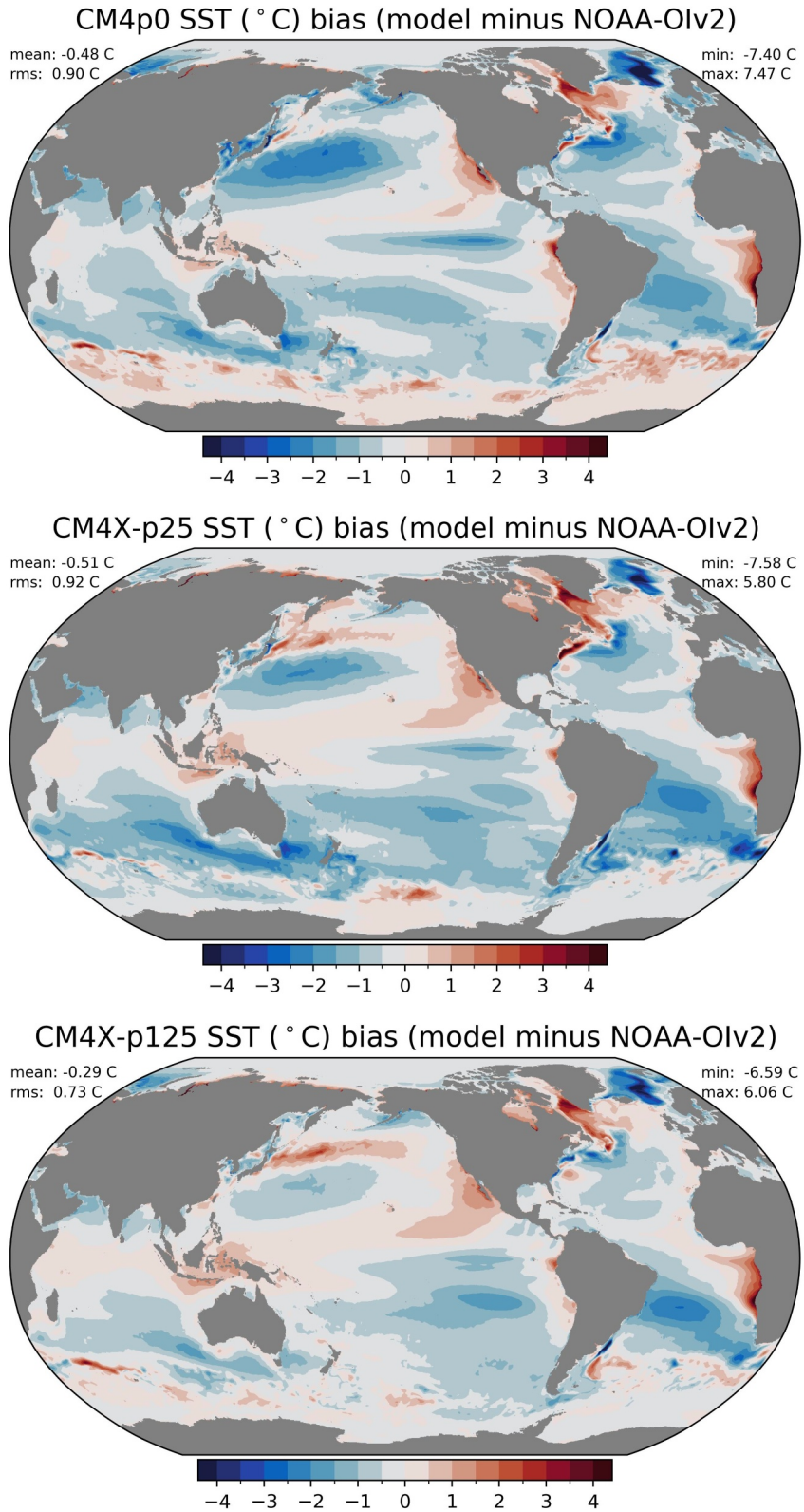

Figure 14. Sea surface temperature bias as defined by the difference between the model time mean over years 1980–2009 versus the NOAA OIv2 observational-based analysis (also averaged over 1980–2009) from Huang et al. (2020). The top panel shows CM4.0, whereas the lower two panels show CM4X-p25 and CM4X-p125.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

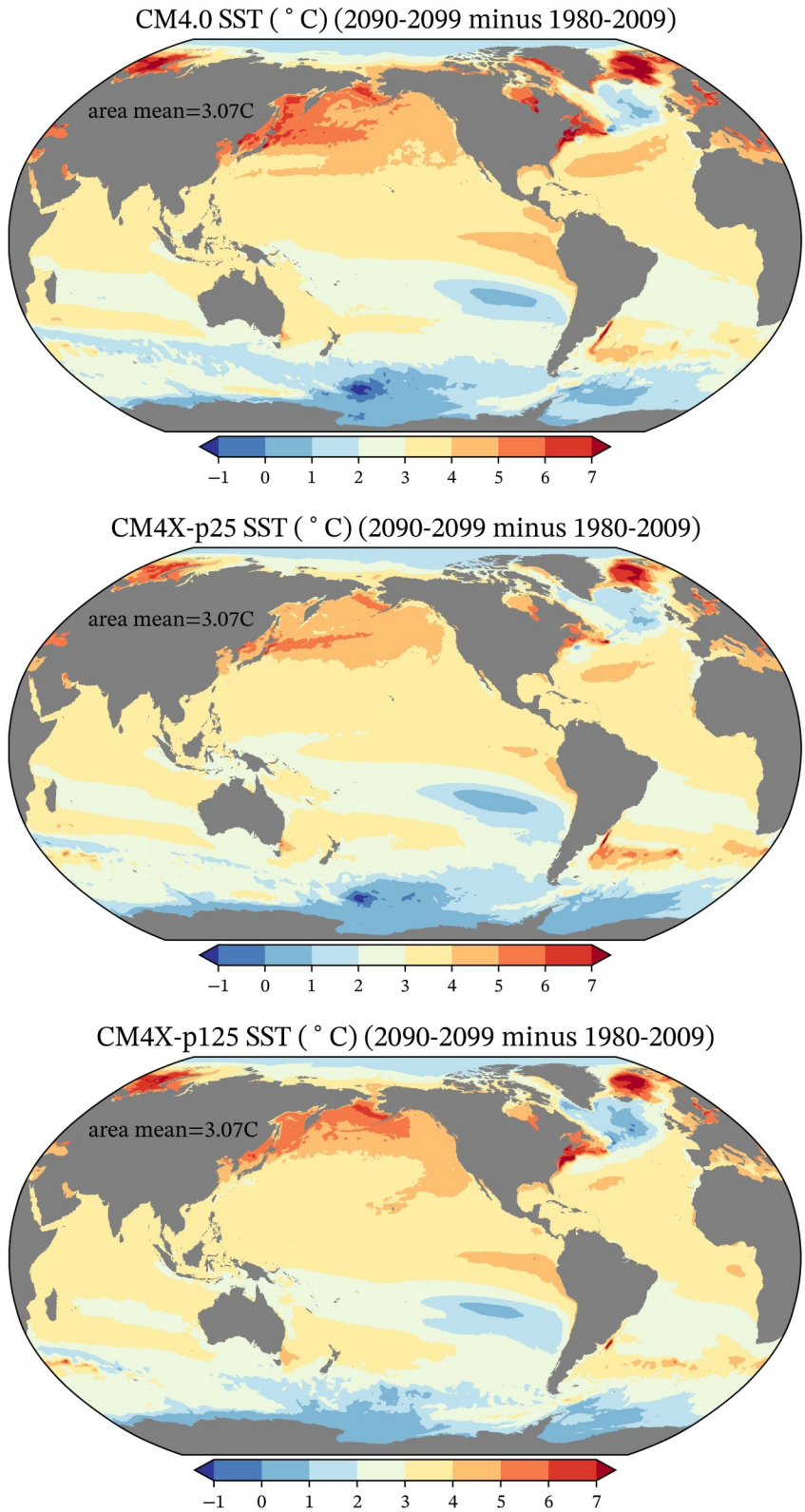

Figure 15. Sea surface temperature (SST) change under the SSP5-8.5 scenario, computed as the difference between the 10 year mean from years 2090–2099, versus the 30-year mean of the historical 1980–2009. Note the area average SST change is listed. The top panel shows CM4.0 and lower panels show CM4X-p25 and CM4X-p125. Note the non-symmetric color scales, with only the dark blue colors reflecting cooling. Use of the final 10 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; for example, Newsom et al. (2023).

The Southern Ocean shows muted warming and some patches of cooling, with the exact positioning of these features likely dependent on decadal modes of climate variability. 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 most of that heat is transported to the north and exported to other ocean basins (Armour et al., 2016; Gregory et al., 2016; Manabe et al., 1990, 1991). 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; Manabe et al., 1991; R. Zhang et al., 2019), whereas the coastal warming arises from associated changes in the Gulf Stream (Caesar et al., 2018; Saba et al., 2016). 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., 2025a)). 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 1,000 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 1,000 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%.

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

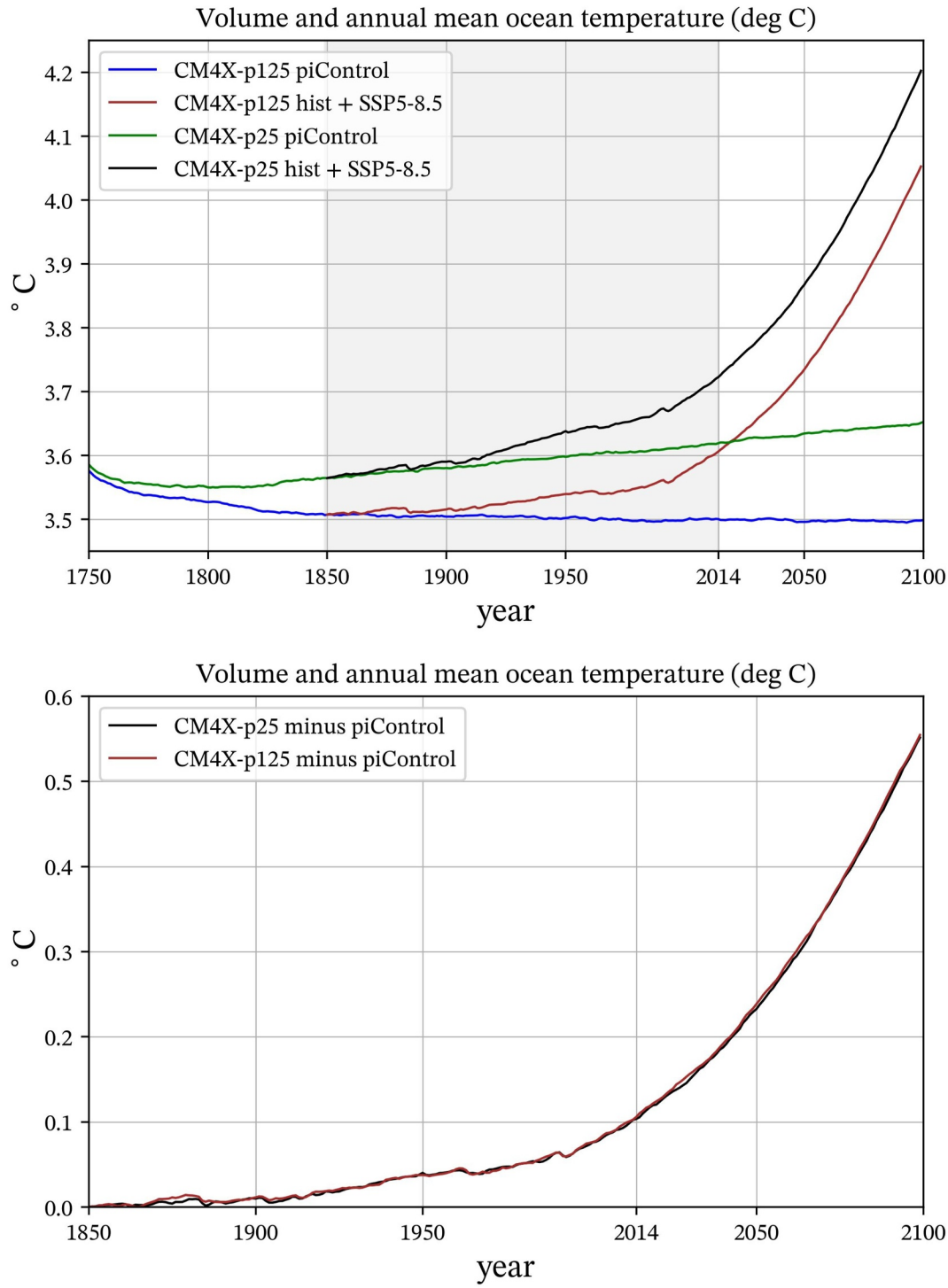

Figure 16. First panel: time series for the volume mean and annual mean ocean Conservative Temperature for CM4X-p25 and CM4X-p125 when run for the piControl, historical (1850–2014 shaded), and SSP5-8.5 from 2014 to 2100. The piControl calendar years are nominal, here set to allow for alignment with the 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. Second 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 find the time series for both models collapsing onto one another.

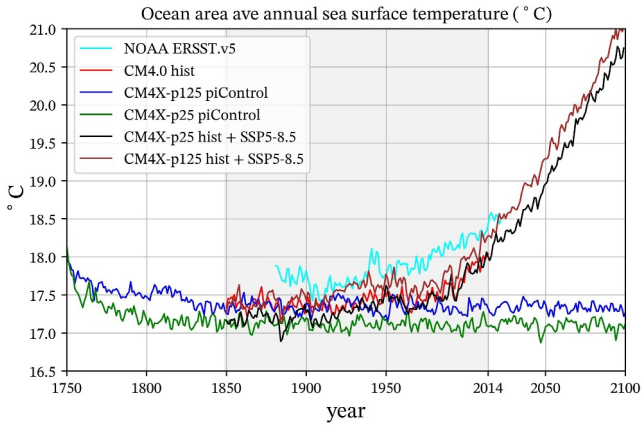

Figure 17. Time series for the ocean area mean annual sea surface temperature. The historical simulations branch from the piControl and run from year 1850 to 2014 (gray shaded region). The cooling blips in the historical simulation result from volcanic aerosol forcing. The SSP5-8.5 scenario extends from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

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

The SSP5-8.5 simulations shown in Figure 16 extend from year 2014 to year 2100. Removing their respective piControls (right panel of Figure 16), collapses the CM4X-p25 and CM4X-p125 historical and SSP5-8.5 curves on top of one another, thus revealing the very similar transient climate sensitivities between the two CM4X models.

5.2. Area Mean SST

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

5.3. Depth-Time Conservative Temperature: World Ocean

We expose more details about the Conservative Temperature evolution in Figure 18, which shows the depth-time diagram for the global horizontally averaged annual mean temperature in the piControl simulations. Both models cool in the upper ocean, consistent with the SST in Figure 17 and consistent with the pre-industrial forcing. However, both also show a warming trend in intermediate depths, and slight cooling again in the abyssal regions. The upper and intermediate depth trends are enhanced in CM4X-p25 relative to CM4X-p125, thus leading to the cooler surface temperature and warmer intermediate temperature in CM4X-p25. As for the global volume mean time series in Figure 16, it is remarkable how stationary CM4X-p125 is for all depths after roughly 100 years of the piControl.

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 4,000 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 3,000 m, with cooling down to around 1,500 m, warming between 1,500 m and 3,500 m, and slight cooling below.

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

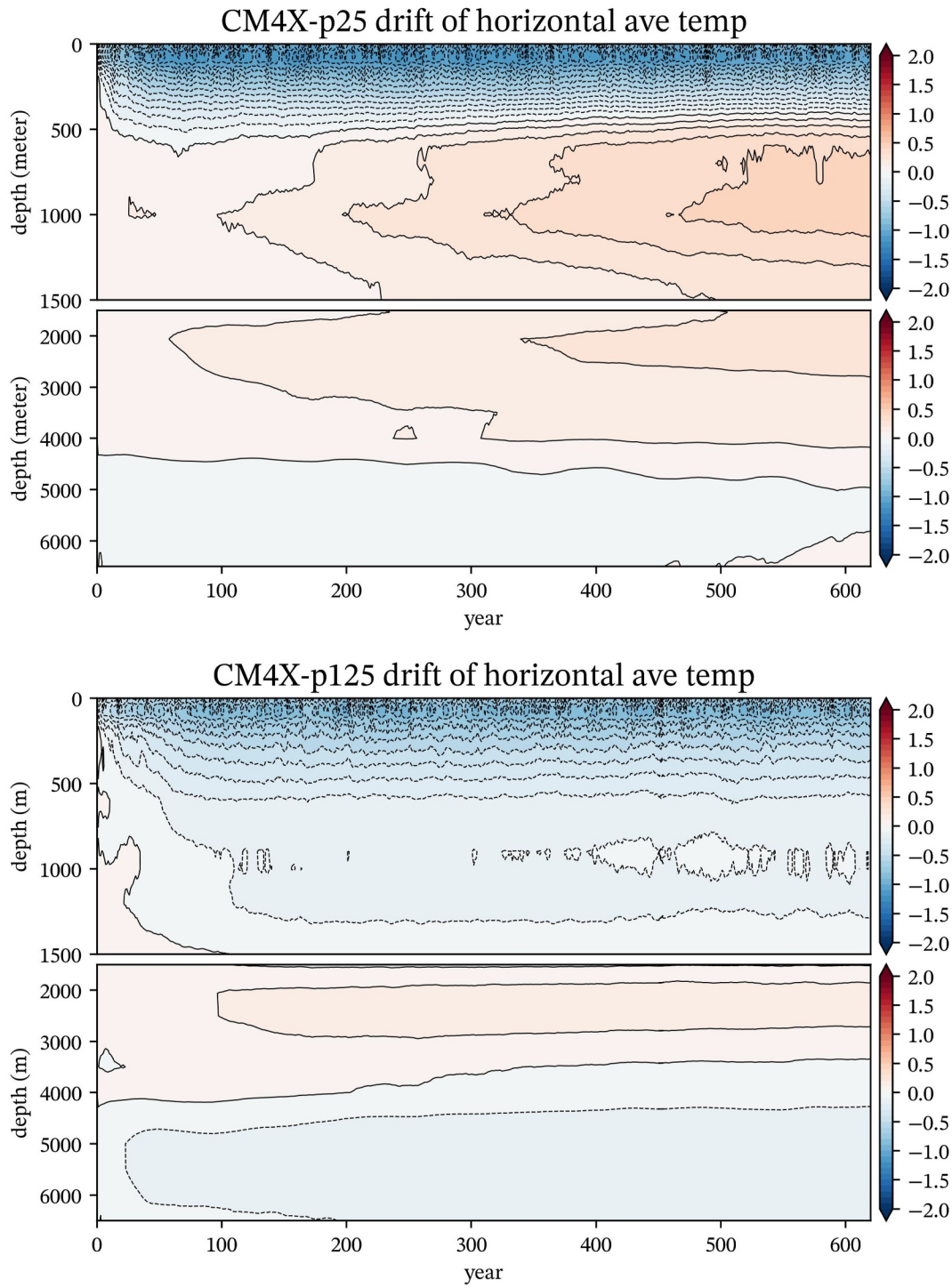

Figure 18. Depth-time drift for the global horizontally averaged annual mean Conservative Temperature. We show temperature evolution relative to the initial year as realized for the piControl simulations, computed as $\Theta_{\text{drift}}(z,t) = \sum_{xy} \Delta V \left[\Theta(x,y,z,t) - \Theta(x,y,z)_{\text{init}}\right] / \sum_{xy} \Delta V$, where $\Theta(x,y,z)_{\text{init}}$ 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.

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

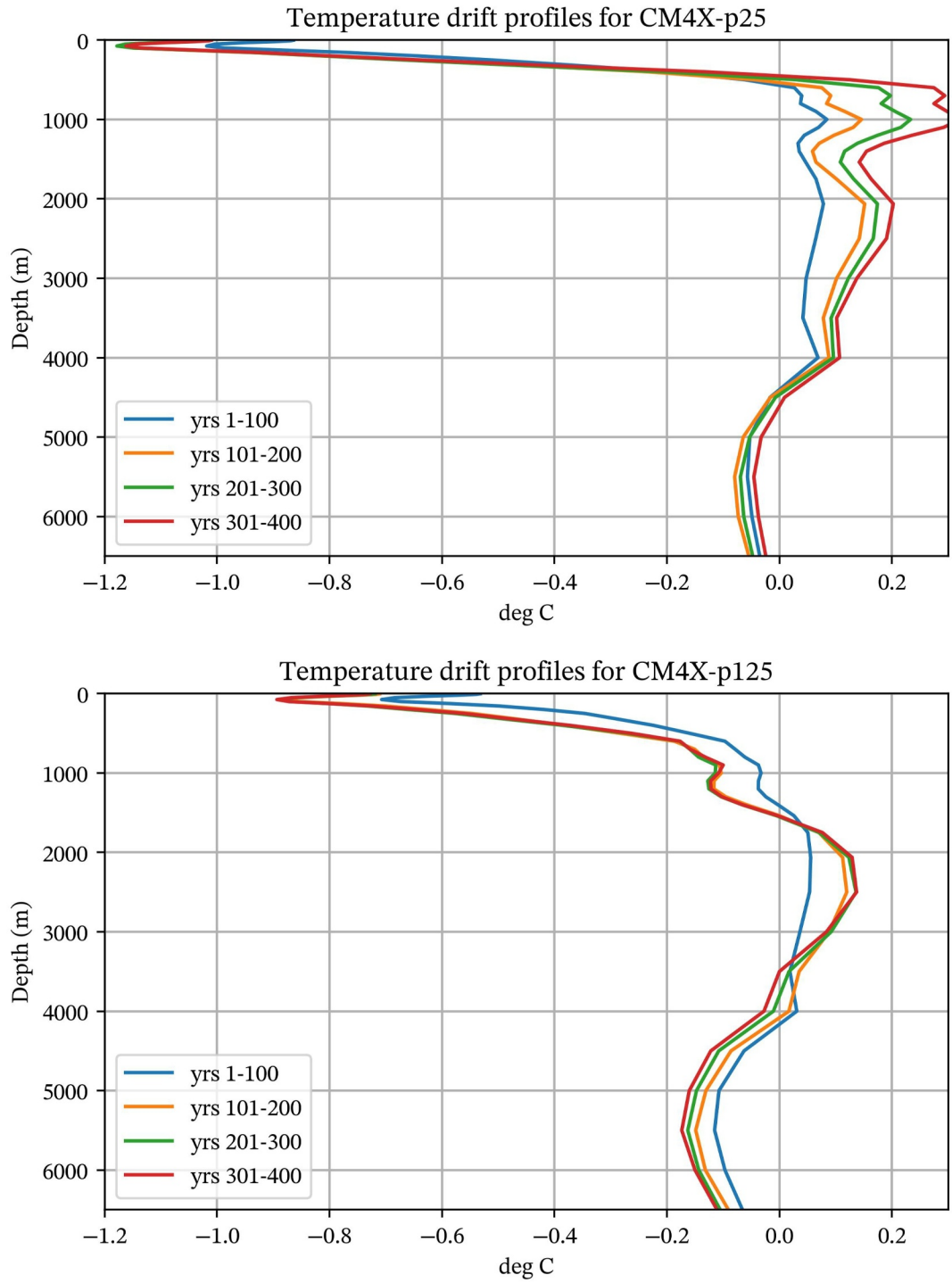

Figure 19. Vertical profiles of horizontally averaged Conservative Temperature relative to the initial year, computed as a time mean for years 1–100, 101–200, 201–300, and 301–400, for CM4X-p25 and CM4X-p125.

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

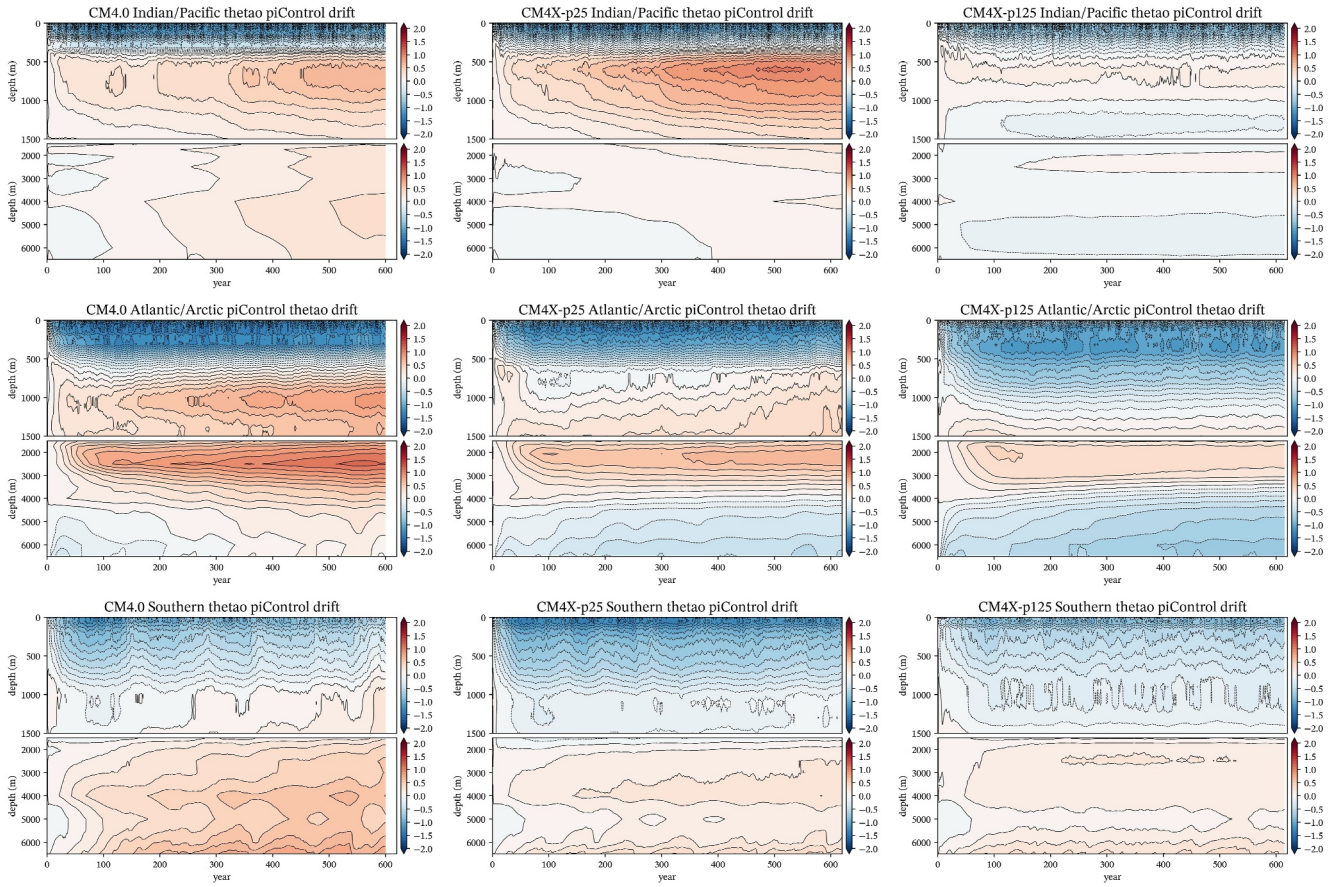

Figure 20. Depth-time drift for the horizontally averaged annual mean Conservative Temperature, as in Figure 18, but here decomposed into basins (Indian/Pacific, Atlantic/Arctic, and Southern) and showing CM4.0 (left column), CM4X-p25 (middle column), and CM4X-p125 (right column).

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 1,500 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, 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

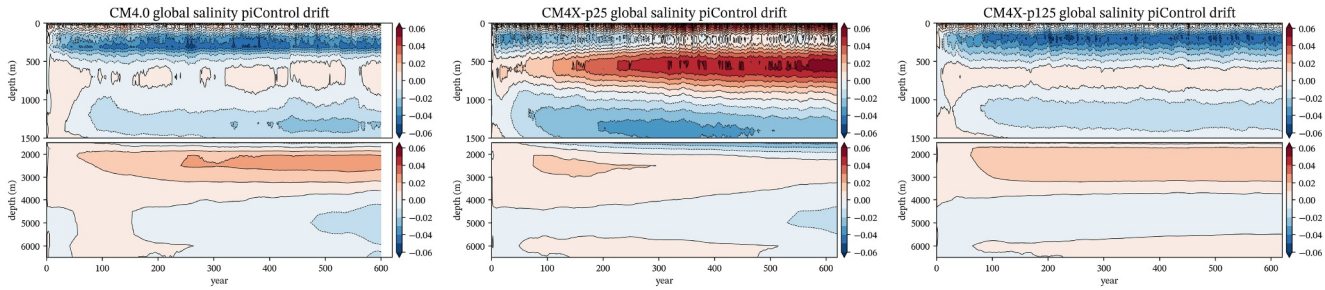

Figure 21. As in Figure 18, yet here for the salinity averaged over the World Ocean for CM4.0, CM4X-p25, and CM4X-p125. Units are g/kg or parts per thousand.

Ocean volume. For the Atlantic and Southern Oceans, the surface cooling penetrates to roughly 1,000 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. Note the larger cooling drift in the deep Atlantic-Arctic basin for CM4X-p125 relative to CM4X-p25.

CM4X-p25 and CM4.0 show a warming between 500 and 4,000 m in the Atlantic/Arctic basins, whereas CM4X-p125 shows a warming starting around 1,000 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 200 m and a larger deep warming signal below 500 m in all three models. Ongoing research suggests that this adjustment is related to biases in the parameterized ocean vertical diffusion.

As emphasized in Section 2.4, the comparison between CM4.0 and CM4X is not clean, since there are many details across the models that differ. Even so, by exhibiting the CM4.0 results in Figure 20 we are able to better gauge the improvements in CM4X-p125 within the ocean interior. One further point of note concerns the hint of a centennial time scale variations in the Southern Ocean, with such variations present in the upper 1,000 m across the three models. These variations are also noticeable in the deeper ocean, particularly in CM4.0. We return to such signals in Section 5 of Part II (Griffies et al., 2025a). In that discussion we note that this signal corresponds to Southern Ocean polynya activity, with the CM4X signals far smaller and more realistic than the super-polynyas found in CM4.0 and discussed by Held et al. (2019), Dunne et al. (2020), and L. Zhang et al. (2021).

5.5. Depth-Time Salinity: World Ocean

The bulk of our study in this paper concerns temperature and heat. To check that our focus is not missing something revealed by salinity, we display in Figure 21 the depth-time adjustments for salinity, here including CM4.0 as well as the two CM4X models. The near surface generally sees an increase in salinity during the first few years, as expected from the increases in sea ice due to the return to pre-industrial forcing in the piControl. Below the surface salinity increase, and reaching down to a few hundred meters, CM4X-p25 shows an initial freshening that gradually dissipates in favor of a growing salinification in the region between 500 and 1,000 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.

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)

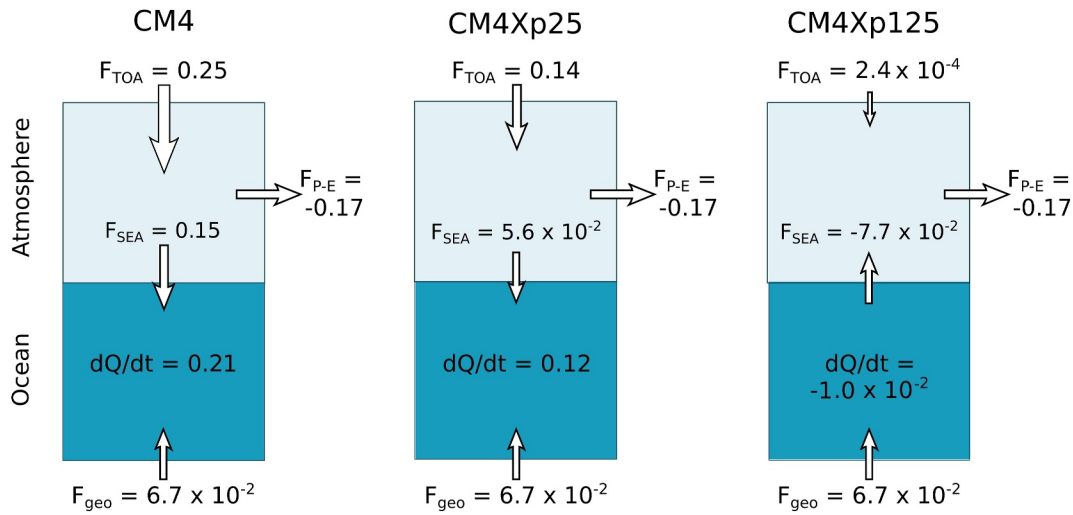

Figure 22. Schematics of the global heat budget for the piControl experiments from CM4.0 (left), CM4X-p25 (center), and CM4X-p125 (right). All numbers are in units W m^{-2} and are obtained from time averaging over years 151–350, and all areas referring to the global earth surface area rather than ocean 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, $F_{\mathrm{P-E}}$, all normalized by global surface area. The heat content tendency of the ocean, $\mathrm{d}Q/\mathrm{d}t$, 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.

geothermal heat flux F_{geo} entering the ocean at the bottom and a net flux F_{SEA} 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 $\mathrm{d}Q/\mathrm{d}t$ that is roughly 10 times smaller than either of the two net boundary fluxes.

The thermal equilibration of the ocean is not a sufficient condition for the thermal equilibration of the entire climate system; although the thermal ocean adjustment operates on long timescales relative to the atmosphere and terrestrial system, other heat sinks and sources may exist. Considering the additional changes to the atmosphere and land models between CM4.0 and CM4X, we expand our consideration to the top-of-atmosphere heat flux. The top-of-atmosphere heat flux, F_{TOA}, is relatively large in CM4.0 with a net heat gain of 0.25 W m^{-2}, intermediate in CM4X-p25 with a net heat gain of 0.14 W m^{-2}, and nearly zero in CM4X-p125. The progressive decrease in F_{TOA} follows from the oceanic heat budgets and is consistent with the ocean establishing the global enthalpy balance on multicentennial timescales. We note that the CM4.0 top-of-atmosphere heat flux is around the CMIP6 median (Irving et al., 2021).

Additionally, the top-of-atmosphere flux magnitudes in CM4.0 and CM4X-p25 are consistently higher than those of the OHC tendencies. A mismatch between F_{TOA} and $\mathrm{d}Q/\mathrm{d}t$ is a common feature among climate models, even for those without drift. The residual is indicative of enthalpy leakage through a combination of spurious dissipation, coupling inconsistencies, and missing processes (Lauritzen et al., 2022). Among CMIP6 models this leakage can reach 0.5 W m^{-2} to 1 W m^{-2} (Irving et al., 2021). Heat flux diagnostics within CM4.0 and CM4X indicate that there is a net loss of heat from the climate system of −0.17 W m^{-2}. This heat sink is associated with the hydrological cycle, as the atmospheric model does not track the sensible heat of water vapor. The magnitude of this enthalpy sink is similar to other generations of GFDL climate models (e.g., Section 3 in Delworth et al. (2006)).

Our ocean-centric enthalpy budget analysis has its limitations, notably by missing terrestrial terms such as the terrestrial enthalpy sink from the water cycle and changes in glacier snow or soil. While changes to the terrestrial setup preclude a full budget comparison, we can estimate the magnitude of the missing non-ocean enthalpy tendencies $\mathrm{d}Q_{\mathrm{non-ocean}}/\mathrm{d}t$ and non-ocean sinks F_{other}:

$$(dQ_{\text{nonocean}}/dt - F_{\text{other}}) + dQ/dt = F_{\text{TOA}} + F_{\text{GEO}} + F_{\text{P-E}}. \quad (1)$$

In CM4.0 the sum of the non-ocean residuals $dQ_{\text{non-ocean}}/dt - F_{\text{other}}$ is -0.06 W m^{-2}, while for CM4Xp25 it is -0.08 W m^{-2} and for CM4Xp125 it is -0.09 W m^{-2}. Thus, the improvements in the top-of-atmosphere energy imbalance are set mainly by the improvements in the ocean model drift through its vertical enthalpy transport.

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 Figure 16.

Along with the similar SST change discussed in Section 4.3 (Figure 15), the comparable ocean heat uptake in CM4X-p25 and CM4X-p125 (Figure 23) further suggests these two models exhibit similar ocean heat uptake efficiency (OHUE). Following Newsom et al. (2023), we estimate OHUE as the change in global ocean heat uptake (in W m^{-2}) divided by the change in global mean SST, calculated over the period 2090–2100 of the SSP5-8.5 scenario relative to 1850–1890 of the historical simulation. As shown in Figure 24, the resulting OHUE for CM4X-p25 and CM4X-p125 are similar, and relatively close to the CM4.0 OHUE as compared to the broader intermodel spread in CMIP5/6 models. Newsom et al. (2023) find that the intermodel spread in OHUE correlates well with the spread in preindustrial pycnocline depth. This correlation points to the importance of physical processes that influence pycnocline depth for affecting OHUE. For CM4X-p25, CM4X-p125, and CM4.0, we find their pycnocline depths to be relatively close (first panel of Figure 24).

The relative insensitivity of the OHUE and pynocline depth in the CM4X models contrasts to their distinct degree of piControl thermal equilibration (e.g., see the planetary heat budget in Figure 22). Perhaps the processes controlling piControl drift are not identical to those affecting OHUE. To further this conjecture, we show in the second panel of Figure 24 the piControl ocean temperature drift versus the OHUE. This figure provides no evidence of a direct relation between piControl drift and OHUE across the CMIP models, thus suggesting the inability to infer a climate model's response to anthropogenic warming from its degree of piControl equilibration.

5.8. Comments on the SSP5-8.5 Comparison

The SSP5-8.5 experiment for CM4X-p25 and CM4X-p125 reveal very close responses in their bulk properties, such as the volume mean temperature in Figure 16 and heat uptake in Figure 23. Evidently, CM4X-p25 and CM4X-p125 are indistinguishable for purposes of climate sensitivity as revealed by bulk properties. We suggest that this close relation arises since the two climate models use the same atmospheric and land components. Even so, the slight differences in SST patterns in Figure 15 and OHUE in Figure 24 might be statistically significant, given that these patterns are more closely dependent on ocean process representations. Yet to quantify the statistical significance of these pattern differences requires ensembles that are unavailable to us.

6. Pre-Industrial Thermal Equilibration in a Mesoscale Dominant Regime

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

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

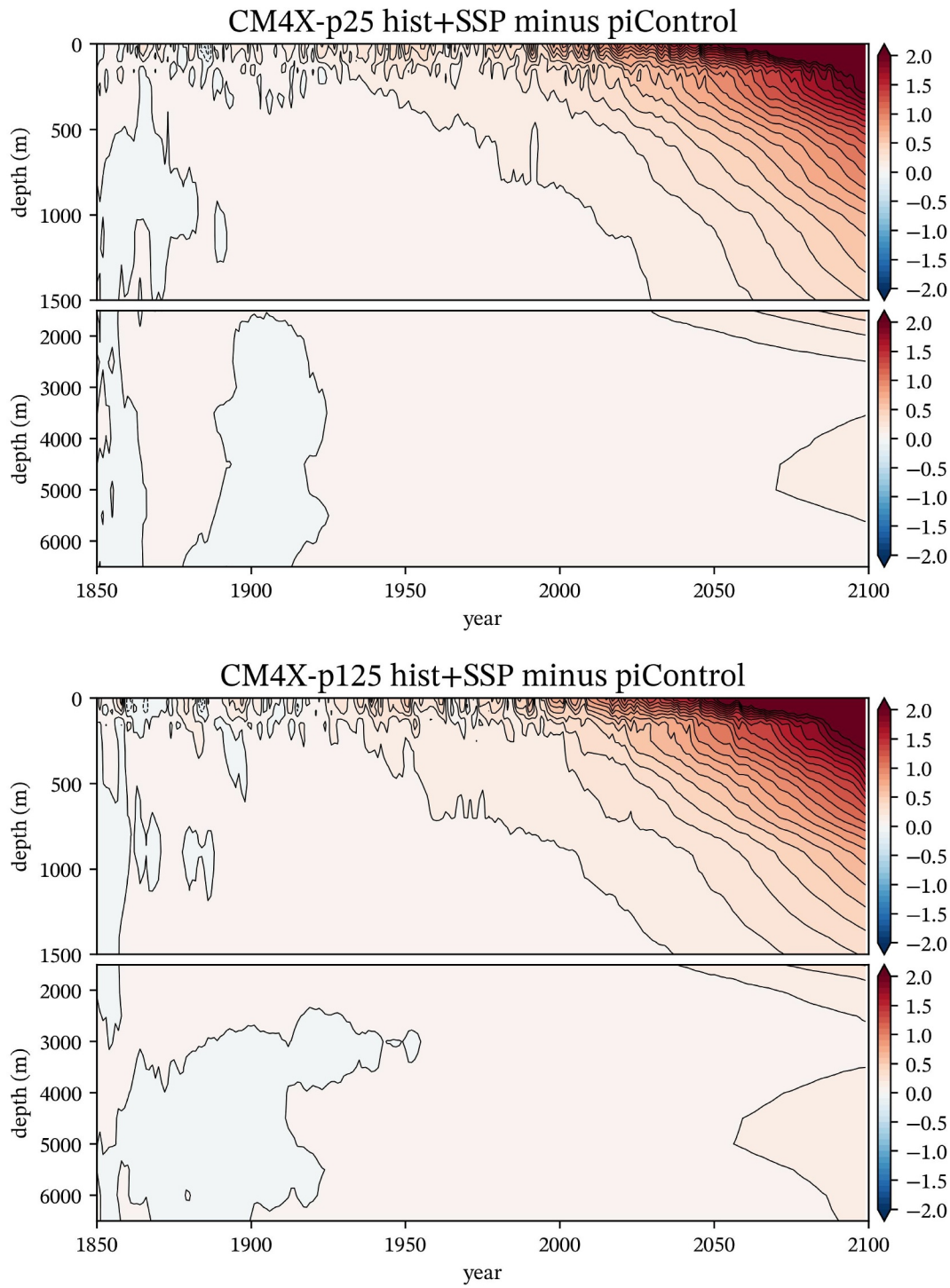

Figure 23. Depth-time drift for the horizontally averaged annual mean Conservative Temperature from CM4X-p25 and CM4X-p125. Here we show the historical and SSP5-8.5 simulations (years 1850–2100) relative to years 101–350 of the piControl; that is, $\Theta_{\text{anom}}(z,t) = \sum_{xy}\Delta V\left[\Theta(x,y,z,t)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

6.1. Drift in the piControl: Early Climate Models

During the early decades of climate modeling, it was necessary to augment modeled air-sea fluxes with a priori diagnosed and non-interactive fluxes aimed at reducing the otherwise extremely large model drift. The papers by

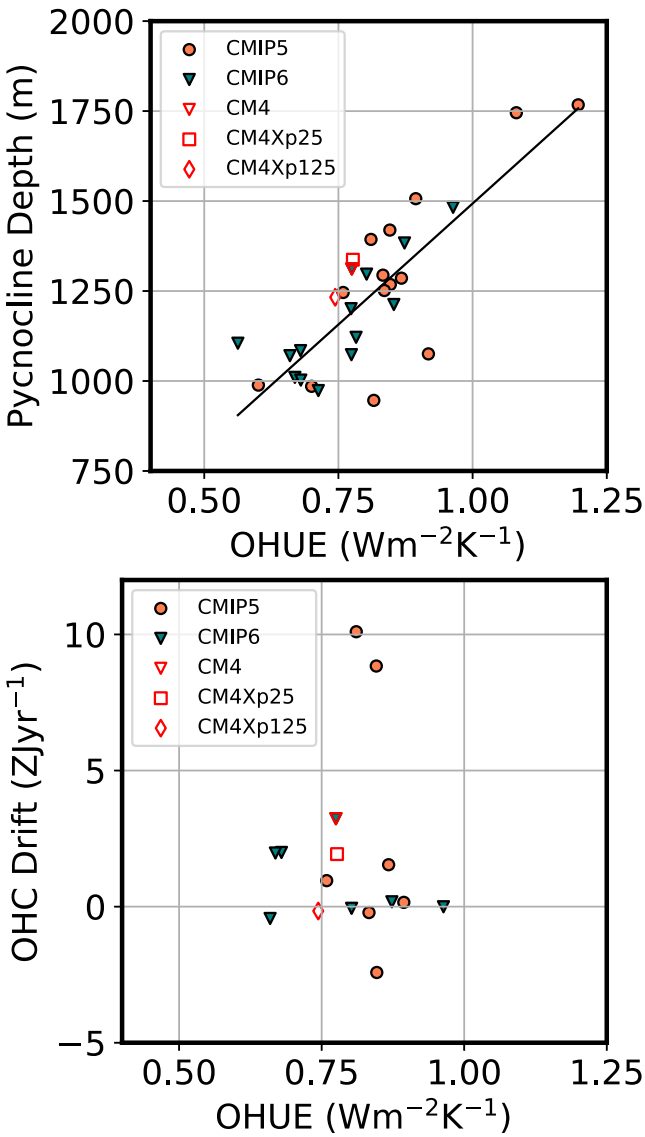

Figure 24. First panel: ocean heat uptake efficiency (OHUE) versus pycnocline depth in CM4X models, with the values of CMIP5/6 models reproduced from Newsom et al. (2023) and with OHUE and pycnocline depth computed as per Newsom et al. (2023). The pycnocline depths are computed during the 80-year periods 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. Second panel: 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).

Sausen et al. (1988) and Manabe and Stouffer (1988) detail these flux adjustment methods. Without flux adjustments, these early generation of climate models manifested a sizable drift that made it difficult, if not impossible, to study climate mechanics since the model drift aliased onto and/or directly impacted climate variability and change.

Climate model drift reduced as models improved their numerics and physics, notably through the representation of poleward ocean heat transport (Weaver & Hughes, 1996). Such advances allowed models to realize sensible climate equilibria without flux adjustments (e.g., Boville and Gent (1998), Gordon et al. (2000), Delworth et al. (2006)). Eliminating flux adjustments was a milestone in the history of climate modeling. Even so, climate models still contain biases and limitations that are fully exposed when removing flux adjustments. So although the current generation of climate models has far smaller biases than the early generation, they still generally struggle to realize thermally equilibrated climate states close to estimates for the 1850 climate, and they reach such equilibrated states only after $\mathcal{O}(1000)$ years (e.g., CM4X-p25, Stouffer (2004), Banks et al. (2006), and Irving et al. (2021)). For the CMIP5/6 climate models, the spread across climate model drift is the main cause for model-to-model differences as well as model-to-observation differences.

6.2. Drift in the piControl: State of the Science Climate Models

Since the ocean provides the dominant heat reservoir for the Earth's climate system, we gauge the degree of climate model thermal equilibration by measuring trends in ocean temperature and/or the net boundary fluxes. We then claim "success" when the ocean trends are smaller than a subjective criteria. If computer cost (and energy footprint) are of no concern, then we can reach thermal equilibrium by integrating the model for as long as it takes (Krasting et al., 2018; Rugenstein et al., 2020). Unfortunately, this brute force approach often leads to an equilibrated climate model with an ocean state that is far from a reasonable 1850 pre-industrial ocean. Indeed, in many cases, such as CM4X-p25, the volume mean ocean is warmer in the equilibrated 1850 piControl than the present-day initial conditions. If there is one thing we know about the 1850 ocean, it is that it had a lower volume mean temperature than the 21st century ocean. We thus reach the conflicted situation whereby the more equilibrated a climate model, the further the ocean deviates from properties corresponding to the real ocean. This situation presents difficult choices for model development teams, and offers little inspiration for ocean analysts to study simulated water masses that are corrupted by huge drifts.

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

Even for presumably linear signals, there are caveats about the utility of drift removal. For example, Hallberg et al. (2013) questioned the linearity assumption for global thermosteric sea level projections since ocean density

is a nonlinear function of temperature and salinity. As they found, extreme levels of model drift can have a notable role in sea level projections. A linearity assumption also becomes untenable when investigating transitions between quasi-equilibrium states, such as hypothesized for the Atlantic overturning circulation. Indeed, the review paper by Hirschi et al. (2020) suggests that Atlantic overturning stability in models is dependent on grid spacing, with finer grid models offering a more realistic depiction of the circulation pathways and, by inference, the stability properties. Studies of overturning stability thus motivate the use of both fine grid models and models that are equilibrated.

6.3. Time Scale for piControl Thermal Equilibration

Consider a climate model with perfect dynamical core, perfect physical parameterizations, and perfect numerics. If this model were subject to piControl forcings and initialized with perfect preindustrial initial conditions, it would presumably remain very close to its initial condition in perpetuity (ignoring that even 1850 has long-term climate drift due to centennial scale natural and anthropogenic variability–see Gebbie and Huybers (2019)). Instead, suppose this model is initialized with an exact rendition of present-day ocean temperature and salinity fields. What is the time scale for removing anthropogenic heat from the ocean in a piControl experiment? Since the bulk of the anthropogenic heat entered the ocean in the period after 1850, and since it is largely contained in the upper 1000 m of ocean, one hypothesis is that the time scale for its release is also $\mathcal{O}(100)$ years. This hypothesis is tempered by realizing that ocean heat uptake and heat release are not symmetric processes (Pudig et al., 2023; Stouffer, 2004). Furthermore, there is some heat sequestered in the deep ocean (Johnson & Purkey, 2024; Purkey & Johnson, 2010, 2012, 2013) through upper ocean freshening and corresponding enhanced stratification, rather than the injection of anthropogenic heat. It is connected to the efficiency by which heat can make its way to the abyssal ocean in regions not influenced by dense water formation, with that time affected by vertical mixing, either physically informed or spurious. Given these caveats, we find that the CM4X-p125 result, which equilibrates to a cooler piControl state than present-day, supports the hypothesis that the time scale is closer to $\mathcal{O}(100)$ years than $\mathcal{O}(1000)$ years.

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

$$\Delta\Theta/\Delta t = Q^{\text{heat}}/(\rho_o\, C^0_{\text{p}}\, H), \tag{2}$$

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

$$V\rho_o\, C^0_{\text{p}}\, \Delta\Theta \approx \left(1.3 \times 10^{18}\ \text{m}^3\right)\left(1035\ \text{kg m}^{-3}\right)\left(4000\ \text{J kg}^{-1}\ \text{K}^{-1}\right)(0.07\ \text{K}) \approx 390\ \text{ZJ}. \tag{3}$$

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

Now consider a model with exact present-day initial conditions but flawed dynamical core and physical parameterizations. The model's piControl equilibrium state will generally differ from the correct preindustrial state that has ≈400 ZJ less heat content than present-day. Depending on model biases, it will tend towards an equilibrium state that is warmer (as in CM4X-p25 and ESM2Mb; Krasting et al. (2018)) or even cooler than the preindustrial state (as in ESM2G; Krasting et al. (2018)). This re-equilibration process is likely to engage the whole ocean, including the deep ocean, and thus have a time scale set by the slow diffusive mixing timescale

$\tau_{eq} \approx H^2/\kappa_d \approx 5000$ yr (Krasting et al., 2018) of unventilated shadow zones in the deep Pacific Ocean (Holzer et al., 2021).

6.4. The Mesoscale Dominance Hypothesis

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

6.4.1. Mesoscale Eddies and Vertical Heat Transport

As presented in Section 5.3, we invoke a mesoscale eddy mechanism to explain the distinct timescales for ocean thermal adjustment in CM4X-p25 and CM4X-p125. Namely, Griffies et al. (2015) (along with Gregory (2000), Gnanadesikan et al. (2005), Wolfe et al. (2008), Gregory and Tailleux (2011), Delworth et al. (2012), Morrison et al. (2013), Hieronymus and Nycander (2013), Zika et al. (2014), D. P. Marshall and Zanna (2014), Doddridge et al. (2016), and von Storch et al. (2016)) identified the role of mesoscale eddies in regulating vertical heat transport between the upper ocean (roughly the upper 1,000 m) and deeper ocean interior. This regulation is affected by the vertically upward transport of positive buoyancy anomalies by eddies, and it holds whether the eddies are explicitly represented, as in an eddy-admitting model, or parameterized via a scheme such as Gent and McWilliams (1990) and Gent et al. (1995).

For those regions without deep water formation (e.g., the Indian and Pacific basins), much of the vertical transport of heat by the mean flow is compensated by an opposing heat transport by the mesoscale eddy flow, leaving a residual heat transport that is smaller than either the mean or eddy heat transport. In so doing, mesoscale eddies reduce the exchange of heat between the upper ocean and the deeper interior ocean; that is, the upper and deeper oceans become disengaged. As a result, mesoscale eddies reduce the role of deep ocean mixing in the process of the piControl thermal equilibration. We can see this effect through the distinct temperature drifts found in the Indo-Pacific basin (see Figure 20), in which there is only a negligible drift in the deep ocean of CM4X-p125 whereas the deep ocean in CM4X-p25 has a more sizable drift.

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

Restratification by submsesoscale eddies in the upper ocean (i.e., mixed layer baroclinic instability), also affects the upper ocean ventilation (Boccaletti et al., 2007). In particular, during the CM4X development we found the details of the Fox-Kemper et al. (2008); Fox-Kemper et al. (2011) submesoscale parameterization (see Section A14) to impact on the vertical heat exchange and thus the thermal drift. Our chosen parameter settings are somewhat stronger than warranted by process studies. The settings were chosen in developing CM4X-p25, with the stronger restratification set to compensate for the somewhat weaker restratification from the relatively weak mesoscale eddies. Determining the proper interplay between the mesoscale and submesoscale restratification effects remains a topic of ongoing research.

6.4.2. The Need to Reduce Spurious Mixing

The study of Adcroft et al. (2019) showed that thermal drift in an ocean model forced by a prescribed atmosphere is directly related to the vertical coordinate, with the hybrid z^*/ρ_{2000} coordinate used in CM4.0 (and in CM4X) leading to significantly less drift than z^* used throughout the ocean. These results suggest there are significantly larger levels of spurious numerical mixing with the z^* coordinate than the hybrid z^*/ρ_{2000} coordinate. They are also consistent with the study of Ilicak et al. (2012) who diagnosed, using potential energy methods, large spurious mixing in CM2.5 (which uses z^* throughout the full ocean depth). The association of a biased warm ocean arising from spurious mixing is supported by Krasting et al. (2018) and Hieronymus et al. (2019), with both papers using the GFDL-CM2G climate model (which uses an isopycnal coordinate ocean component) and finding that the volume-averaged ocean temperature increases with an increase in background diapycnal tracer diffusivity.

6.4.3. Summarizing the Mesoscale Dominance Hypothesis

We hypothesize that there are three ocean model properties necessary to support a centennial rather than the millennial time scale for piControl thermal equilibration into an ocean that is cooler (with roughly 400 ZJ less heat content than early 21st century) in its 1850 piControl state: (a) enhanced fidelity of mesoscale features, including transient eddies and boundary currents; (b) accurate strength and geography of parameterized numerical mixing processes; and (c) negligible levels of spurious mixing from numerical discretization. We refer to ocean models that possess these three properties as *mesoscale dominant models*. Mesoscale dominant models contrast to those where deep ocean diabatic processes (either parameterized or spurious numerically induced) play a prominent (and sometimes dominant) role in piControl thermal equilibration. We infer that models that are not mesoscale dominant engage their deep ocean circulation during the 1850 piControl, thus rendering far longer thermal equilibration time scales. These long thermal spin-ups also affect long spin-up times for biogeochemical cycles, though biogeochemical spin-ups are also impacted by other processes (Khatiwala, 2023, 2024; Orr et al., 2017).

CM4X-p125 does not perfectly realize each of the three properties of a mesoscale dominant model. In particular, both of the CM4X models have biases in the Southern Ocean ventilation, as noted by the overly deep mixed layers in Figure 9 and as further explored for CM4.0 in Krasting et al. (2024). The case studies in Part II (Griffies et al., 2025a) point to further shortcomings that form the topic of ongoing research and development. Even so, we propose that CM4X-p125 approaches mesoscale dominance more than CM4X-p25, given the stronger mesoscale eddy activity in CM4X-p125. Furthermore, we emphasize that deep mixing below the pynocline exists in the ocean, and the CM4X models employ parameterizations of breaking gravity waves such as those reviewed in (MacKinnon et al., 2013, 2017). Yet in the presence of negligible levels of spurious mixing and active mesoscale eddy transport, the deep mixing in CM4X-p125 is presumably in a near steady balance with advection so that there is only a very small net trend.

6.5. Water Mass Transformation Analysis

As an initial plausibility test of the mesoscale dominance hypothesis, we employ water mass transformation budgets (following Drake et al. (2025) and Section 7.4 in Part II of Griffies et al. (2025a); see review by Groeskamp et al. (2019)) to evaluate the extent to which (a) water masses are well-equilibrated in CM4X piControl simulations and (b) spurious numerical mixing plays a lesser role than parameterized mixing in the water mass budgets. For a water mass $\mathcal{M}_{\geq}(\sigma_2, t)$ defined as the mass of water that is denser than σ_2 at time t, the water mass transformation framework allows us to attribute the drift, $\partial_t \mathcal{M}_{\geq}$, to processes that drive it.

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

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

where $\mathcal{M} \equiv \mathcal{M}_{\geq}(\min(\sigma_2), t)$ is the total mass of the ocean and tildes denote we have normalized the water mass by the total ocean mass. The normalized water mass, $\widetilde{\mathcal{M}}_{\geq}(\sigma_2, t)$, increases monotonically from 0 for the densest water in the ocean to 1 for the lightest waters and thus can be thought of as the ocean's cumulative probability

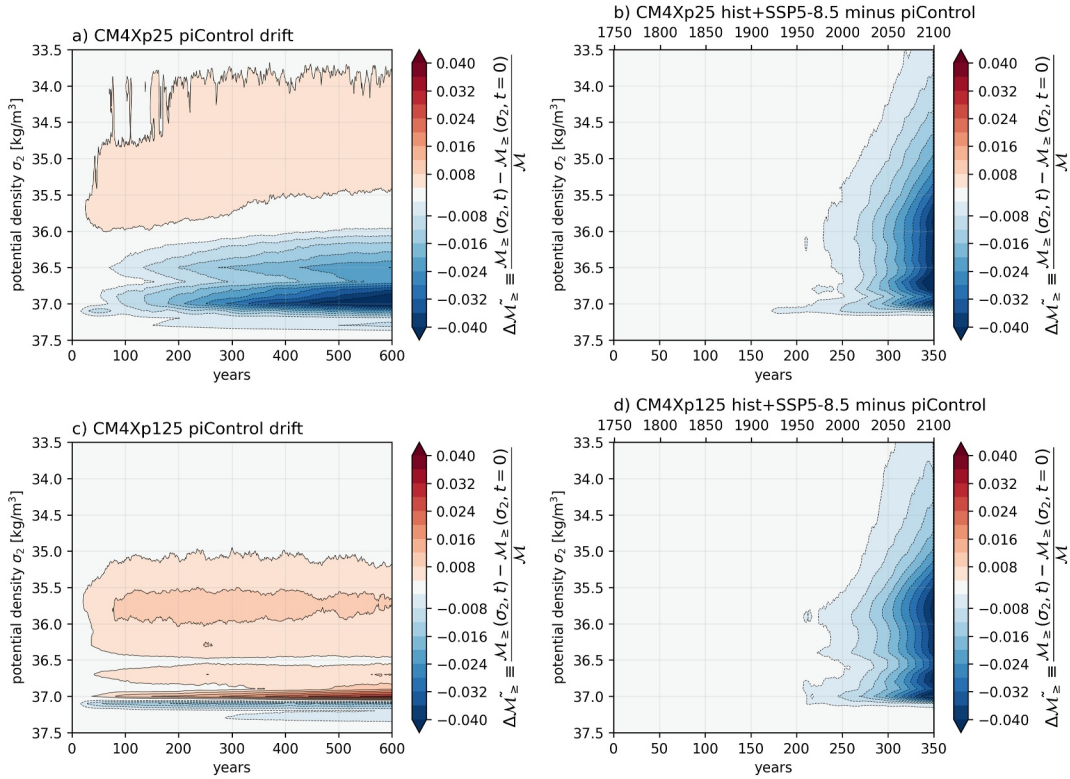

Figure 25. Normalized water mass changes $\Delta\tilde{\mathcal{M}}_{\geq}(\sigma_2, t)$ (Equation 4, where $\mathcal{M}_{\geq}(\sigma_2, t)$ is the mass of water denser than σ_2 at time t) in CM4X-p25 (a, b) and CM4X-p125 (c, d), expressed as a fraction of the total ocean mass. Panels (a, c) show the 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.

density function in potential density space (see Figure 26d). $\Delta\widetilde{\mathcal{M}}_{\geq}$ quantifies how much a given water mass has drifted from the initial condition over time, expressed as a fraction of the total ocean mass.

Figure 25a shows that deep water masses $\left(\sigma_2 > 36\,\mathrm{kg/m^3}\right)$ drift by about -1% per century in CM4X-p25 (Figure 25a). Neglecting exchanges with sea ice and land, the total ocean mass (or volume for the Boussinesq ocean used here) is conserved, so that this negative drift signifies a drift towards lighter densities (consistent with the warming described in Section 4). By contrast, water masses that include lighter surface and thermocline waters (i.e., $\sigma_2 \leq 36.0\,\mathrm{kg/m^3}$) show few changes after the first 50 years, suggesting that they equilibrate more quickly. In CM4X-p125, the deep water masses exhibit vanishingly small drifts after the first 100 years, except for a narrow dipole pattern centered around $\sigma_2 = 37\,\mathrm{kg/m^3}$ that corresponds to a contraction in the density range of Antarctic Bottom Waters (Figure 25c).

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

$$\partial_t \mathcal{M}_{\geq} = \mathcal{S}_{\geq} + \mathcal{G}, \tag{5}$$

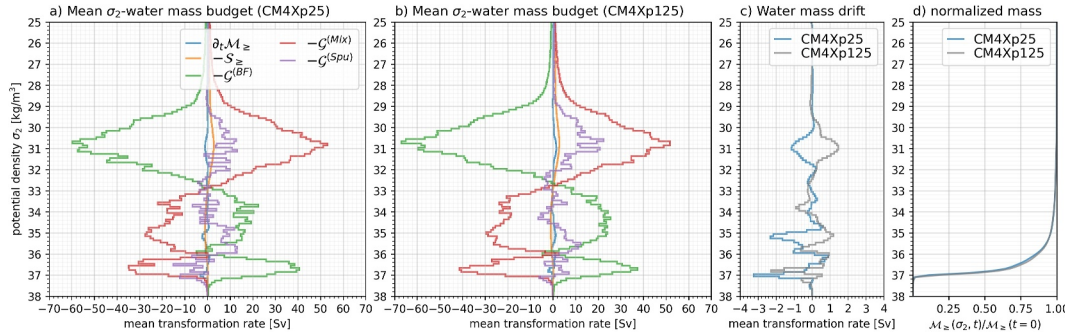

Figure 26. Global water mass budgets (see Section 6 in Part II of Griffies et al. (2025a) for details) in potential density ($\sigma_2 = \rho_{2000} - 1000$: potential density referenced to 2000 dbar, minus 1000 kg m^{-3}) coordinates for CM4Xp25 (a) and CM4Xp125 (b), averaged over years 591–600 of the piControl. (c) Comparison of the water mass drift $\partial_t \mathcal{M}_{\geq}$ terms in the CM4X configurations. (d) The normalized water mass distribution, $\widetilde{\mathcal{M}_{\geq}} = \mathcal{M}_{\geq}(\sigma_2, t)/\mathcal{M}$, 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 $\sigma_2 = 37.1$ kg/m^3.

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 ($\mathcal{G}^{(\mathrm{BF})}$; including air-sea, ice-ocean, and geothermal fluxes of heat, salt, and freshwater), parameterized mixing ($\mathcal{G}^{(\mathrm{Mix})}$; representing surface mixed-layer, interior, and bottom boundary layer mixing processes), and spurious numerical mixing ($\mathcal{G}^{(Spu)}$; diagnosed as the residual of the other terms). $S_{\geq}$ 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 and 26b 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 $\left(35\ \mathrm{kg/m^3} \leq \sigma_2 \leq 36\ \mathrm{kg/m^3}\right)$ and even denser bottom waters $\left(36.75\ \mathrm{kg/m^3} \leq \sigma_2 \leq 37.5\ \mathrm{kg/m^3}\right)$ at a rate of 40 Sv at high latitudes is mostly balanced by parameterized mixing-driven transformation. Spurious numerical transformations due to discretization errors in the advection or Lagrangian remapping scheme (see Drake et al. (2025)), however, further modify water masses at rates of up to $\pm\mathcal{O}(5\ \mathrm{Sv})$. The magnitude of spurious numerical mixing is about 40% of parameterized mixing for lighter water masses but generally less than 10% for the dense waters involved in the meridional overturning circulation. We hypothesize that this difference is due to the hybrid vertical coordinate transitioning from z^* at the surface and at high latitudes to ρ_{2000} in the global deep ocean.

After 590 years of spin-up in the piControl simulations, the water mass transformations described above are nearly in balance, with residual water drifts of less than 4 Sv in all density classes when averaged over 10 years (Figure 26c). Furthermore, both spurious numerical mixing and unbalanced water mass drifts tend to be slightly smaller in CM4X-p125 than in CM4X-p25, especially for the dense waters with $\sigma_2 > 32.5\ \mathrm{kg/m^3}$ that represent 99% of the global water mass distribution (Figure 26d), providing further evidence in support of the mesoscale dominance hypothesis. While analogous mass budgets for the CM2.5/CM2.6 hierarchy are not available, other models that use the same ocean model code (MOM5) exhibit spurious water mass transformations in excess of those driven by parameterized mixing rates (Holmes et al., 2021; Ilicak et al., 2012). The relative smallness of spurious numerical mixing in the CM4X water mass budget supports our hypothesis that the transition to a hybrid $z^* - \rho_{2000}$ coordinate in CM4X (MOM6 code) enables the mesoscale dominant regime to emerge.

7. Concluding Remarks

As further explored in Part II (Griffies et al., 2025a), refined ocean grid spacing improves many aspects of the ocean and climate system that makes CM4X-p125 more suited to a variety of applications. Furthermore, those cases where CM4X-p25 and CM4X-p125 are quite similar, such as global mean transient climate response and OHUE, confirm decades of work done with comparatively coarse resolution models. That is, for models with relatively modest drift such as CM4X-p25, we can garner a sensible estimate for many bulk climate response diagnostics as compared to models with extremely low drifts such as CM4X-p125. Even so, there are aspects

where payoff for the enhanced grid resolution is nontrivial, such as the roughly 20% reduction in SST biases in CM4X-p125, and the order of magnitude reduction in thermal equilibration 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 OHC 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 OHUE, and surface/atmosphere diagnostics like the climate feedback parameter or the SST pattern effect (Andrews et al., 2018; Armour et al., 2013). 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), Holmes et al. (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.

7.2. Prospects for Eddying Climate Models

The common experience of $\mathcal{O}(1000)$ years for piControl thermal equilibration is baked into community strategies for climate modeling. Consequently, there is a presumption that the explicit representation of mesoscale eddies and associated fine-scale boundary currents will long remain computationally out of reach for routine studies of climate dynamics. However, if our experience with CM4X-p125 is reproducible (a big assumption at this point in the science), then moving into a mesoscale dominant regime will reduce piControl thermal equilibration timescales by roughly one order of magnitude. With mesoscale dominant models, the added computational energy and

wall-clock time needed for refined grid spacing are compensated by a significantly reduced thermal equilibration time, thus bringing forth the prospect of more common use of a mesoscale active ocean as part of realistic climate models used to study present and future climate.

Our focus on thermal equilibration for piControl simulations identified $\mathcal{O}(100)$ years as the physically relevant timescale, which contrasts to the $\mathcal{O}(1000)$ years found by most climate models. However, paleoclimate studies generally involve a full-depth ocean state that is far from the present day ocean. Hence, for paleoclimate studies using models initialized with a present-day ocean state, it is necessary to directly simulate a thermal equilibration of the full ocean and to thus pay the price of running for $\mathcal{O}(1000)$ years. Additionally, multi-century piControl simulations are needed to assess natural climate variability modes to provide a baseline for the impacts of anthropogenic changes.

Appendix A: CM4X Model Configuration Details

The ocean and sea ice components of CM4X are based on the MOM6 ocean code and SIS2 sea ice code as configured in the OM4.0 model detailed in Adcroft et al. (2019). In this appendix we summarize those relatively few features of the CM4X ocean that differ from OM4.0. We also summarize elements of the atmosphere and land models that differ from the CM4.0 configuration of Held et al. (2019). Note that the component models are coupled via the GFDL exchange grid coupler documented in (Balaji et al., 2006).

CM4X was run on NOAA's Gaea C5 computer. In the absence of model diagnostics, CM4X-p25 realized 7.5 simulated years per day using 864 cores for the atmosphere/land and 4671 cores for the ocean/ice. CM4X-p125 realized 3.7 years per day using 864 cores for the atmosphere and 8603 cores for the ocean/ice.

A1. Atmosphere Model

As noted in Section 2.2, CM4X makes use of the C192 (nominally 50 km grid spacing) version of the AM4 cubed sphere atmosphere model of Zhao (2020). This choice contrasts to the C96 (nominally 100 km grid) version (Zhao et al., 2018a, 2018b) used in CM4.0. Our choice to use the C192 is based on the desire to simulate more energetic atmospheric cyclones that are key to the study of sea level extremes, such as those studied by Yin et al. (2020). In particular, Zhao (2020, 2022a) showed that the 50 km AM4 (i.e., C192AM4) reasonably simulates the frequency of atmospheric rivers, tropical storms, and mesoscale convective systems, as well as their associated precipitation and extreme precipitation.

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

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

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 when not coupled to an interactive ocean or sea ice model. 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

Table A1
Summary of the Differences in Atmospheric Physics Parameter Setting for Various GFDL Models

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)

Note. 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.

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 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 vs. C96) was discussed in Zhao et al. (2018a, 2018b).

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.

Table A2
A Comparison of Global Top of Atmosphere 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 to 2014, and Radiative Gases and Aerosol Emissions Fixed at Present-Day (PD, 2010) Levels (Referred to as 2010 Control)*

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

Note. 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.

Table A3
Summary of the Land Glacier Albedo Settings for a Variety of GFDL Models

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)

Note. The parameters F_ISO_WARM_ON_GLACIER and F_ISO_COLD_ON_GLACIER provide the albedo values for warm and cold glacier surfaces, respectively. There are two albedo values provided, corresponding to Visible and Near Infrared radiation, respectively.

Including dynamic vegetation increases climate sensitivity, resulting in roughly 0.4 K increase in temperature in response to doubling of CO_2. Zhao et al. (2018b) find that the majority of this increase is due to an increase in effective radiative forcing, rather than due to feedback. 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-ESM4.1 configuration of Dunne et al. (2020) (see Table A3). These near-infrared glacial snow albedos lie at the upper limit of the observational range (Dunne et al., 2020).

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

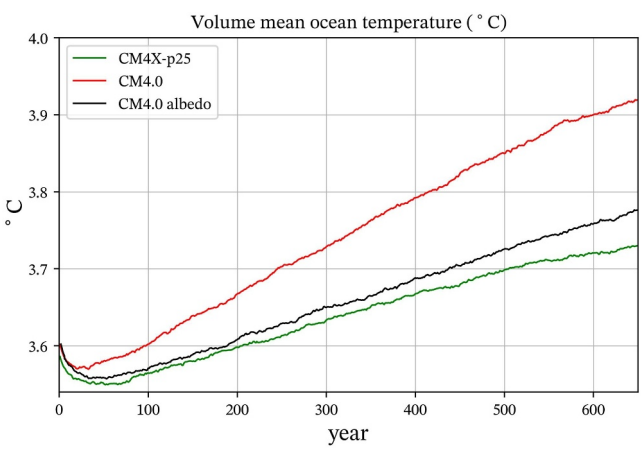

Figure A1. Time series of the global volume mean ocean temperature from CM4.0, CM4.0-albedo, and CM4X-p25. The CM4.0-albedo experiment is identical to CM4.0, with the one exception being the use of the CM4X snow on glacier albebo as per Table A3. The nearly parallel behavior of CM4X-p25 and CM4.0-albedo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

In Figure A1 we show the time series for the global volume mean ocean temperature for CM4.0 and CM4X-p25, along with a third configuration with CM4.0 using the albedo of snow on glaciers used by CM4X. 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 enthalpy budget perspective, between CM4.0 and CM4X-p25. The impact of the glacier albedo change is dominated by a large response over the Antarctic continent, characterized by a decrease in absorbed shortwave radiation at both the surface and top-of-atmosphere, which leads to a decrease in global ocean heat uptake (Delworth et al., 2020).

A4. Albedos for Sea Ice

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:

- 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.

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

Table A4
Summary of the Sea Ice Albedo Settings for a Variety of 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

Note. The sea ice albedo settings are in units of observed standard deviations relative to observed mean values following the Delta-Eddington formulation of Briegleb and Light (2007). For example, a value of 1.0 corresponds to an albedo value that is 1 standard deviation above the observed central estimate. The parameters ICE_DELTA_EDD_R_ICE, ICE_-DELTA_EDD_R_SNOW, ICE_DELTA_EDD_R_POND provide the albedo standard deviation values for bare sea ice, snow-covered sea ice, and ponded sea ice, respectively.

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.

A5. Fixing a Sea Ice Bug for Arctic Bipolar Coordinate Fold

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

A6. Horizontal Grids for the Ocean and Sea Ice

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

Table A5
Details for the Horizontal Grids Used by the Two CM4X Ocean and Sea Ice Components

Grid property	CM4X-p25	CM4X-p125
Nominal (equatorial) grid spacing	0.25°	0.125°
Grid cells (ni × nj)	1440 × 1080	2880 × 2240
Southern edge of Mercator grid	66.8595472°S	66.8840965847°S
Northern edge of Mercator grid	64.0589597°N	64.0316059408°N
Longitudes of bi-polar singularities	120°W, 60°E	120°W, 60°E
Southern edge of south spherical grid	78°S	78°S

Note. Resolution in this paper refers to the number of grid degrees of freedom and as such it is an integer number, whereas *grid spacing* refers to the distance between grid cells as measured in units of degrees or kilometers.

A7. Numerical Methods for the Coriolis and Magnus Accelerations

There are a variety of numerical methods available in MOM6 for handling the Coriolis plus Magnus accelerations

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

where h is the layer thickness and ζ is the vertical component to the relative vorticity. The term $\zeta\hat{\boldsymbol{z}} \times \boldsymbol{u}$ is referred to as the *Magnus acceleration* in some treatments, particularly the engineering literature, whereas it is commonly referred to as the *nonlinear Coriolis acceleration* in the ocean modeling literature. The layer thickness, h, in the numerator does not exactly cancel h in the denominator. The reason is these two thicknesses are computed using distinct horizontal averaging operators to ensure a proper handling of the vanishing layer limit where $h \to 0$.

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

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

- For CM4X we set NOSLIP = False, which means that the solid boundaries feel a drag law as discussed in Appendix A9, rather than the no-slip condition commonly applied in B-grid models (MOM6 uses the C-grid).
- 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).

A8. Ocean Bottom Topography and Geothermal Heating

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

A8.1. GEBCO and BedMachine Data Sources

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

A8.2. Hand Tuning of Selected Straits

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

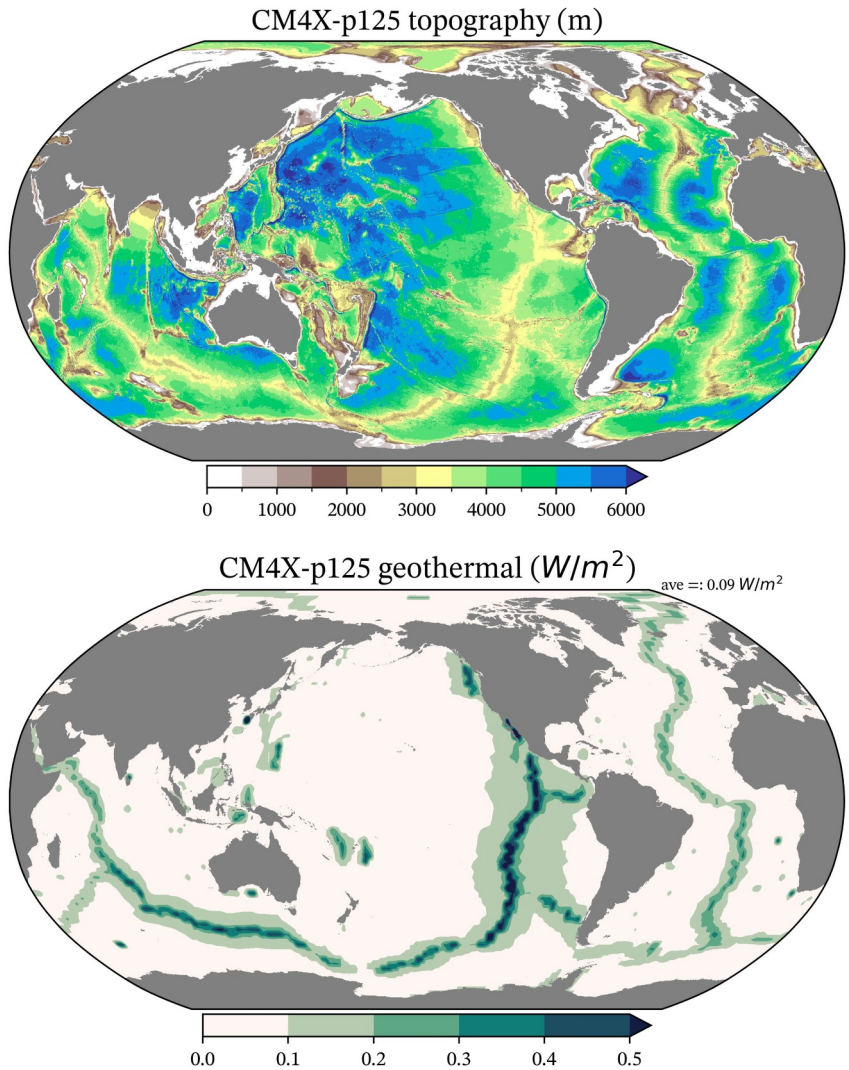

Figure A2. First panel: ocean bottom depth (in meters) for the 0.125° horizontal grid used in CM4X-p125. For the region south of 62°S we transitioned from the GEBCO (GEBCO Compilation Group, 2021) product used in the north to the BedMachine product from Morlighem et al. (2017). Both CM4X-p25 and CM4X-p125 set their minimum depth at 9.5 m and maximum depth at 6500 m. Second panel: geothermal heat flux for CM4X-p125, taken from Huw Davies (2013). The global area mean (ocean area) is 0.094 W m^{-2}.

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 straights that are best kept closed (e.g., Lake Maracaibo in Venezuela).

The hand-tuning works on a cell-by-cell basis, opening or closing straights according to the size of the grid cell. MOM6 provides the added feature of being able to horizontally 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, Bosporus, 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.

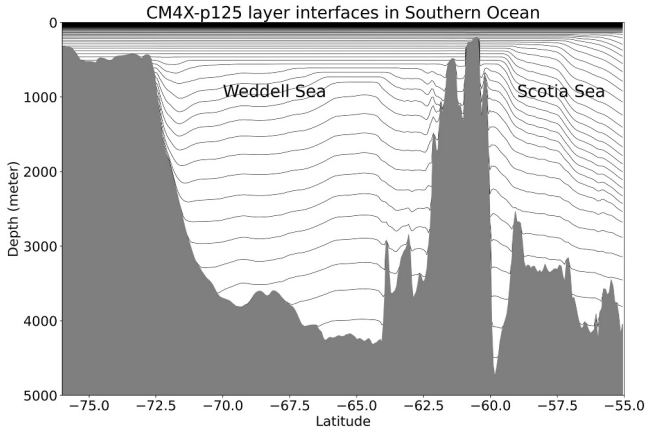

Figure A3. Layer interfaces through the Weddell and Scotia Seas (with South Orkney Islands rising just south of 60S) for the CM4X-p125 configuration as time averaged over years 1980–2009. The horizontal layers in the upper ocean are aligned with z^* whereas the interior layers are aligned with potential density, ρ_{2000}, 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 continental slope (southerly upward sloping isopycnals) is a signature of the westward flowing Antarctic Slope Current.

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, ρ_{2000}, referenced to 2000 dbar. Note that the same vertical Lagrangian remapping method is used regardless of 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.

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 (Hughes et al., 2019; Wise et al., 2018, 2020).

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

The 20% mixing fraction used in OM4 and CM4 was motivated by studies of turbulence in stratified fluids without boundaries (Ivey & Imberger, 1991). Given the overly strong mixing found with 20% mixing fraction, we chose to reduce it to 1% in CM4X. The value of 1% is physically motivated to be consistent with the mixing fraction of turbulent mixing in large eddy simulations of the surface boundary layer (see Reichl & Li, 2019, their Figure 2). Further tests in an OM4 configuration reveal little sensitivity to the precise value, so long as it is less than a few percent.

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

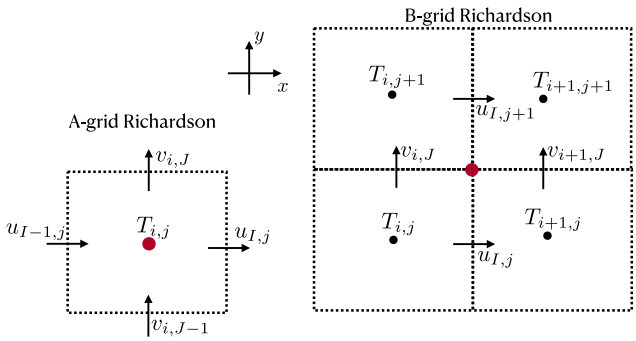

Figure A4. Two methods to compute the gradient Richardson number for the C-grid used in MOM6. Left panel: OM4.0 computes the Richardson number on the A-grid by averaging the horizontal velocity components, $u_{I,j}$ and $v_{i,J}$, to the tracer cell center (red dot at $T_{i,j}$) and then computing the 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.

A11. Parameterization of Upper Ocean Boundary Layer Mixing

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

A12. Richardson Number for Parameterized Shear-Driven Mixing

Shear-driven mixing is parameterized according to the gradient Richardson number based scheme of Jackson et al. (2008). The gradient Richardson number, Ri, is the ratio of the squared buoyancy frequency, N^2, to the squared vertical shear of the horizontal velocity

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

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

A12.1. Discrete Richardson Number that Introduces Grid Noise

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

A12.2. Discrete Richardson Number That Does Not Introduce Grid Noise

Generating grid noise from the vertical diffusion scheme is quite disappointing since we generally rely on diffusion to smooth fields rather than generate noise. Hence, for CM4X we formulated an alternative discretization aimed at suppressing the computational mode. We do so by interpolating the horizontal velocity components, as well as the temperature and salinity, to the tracer corner point; that is, to the B-grid point. The buoyancy frequency and vertical shear are then computed using the interpolated B-grid fields, as depicted in the right panel of Figure A4. In this manner, we couple across the tracer points as well as velocity points, thus suppressing the A-grid computational mode.

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

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

A13. Parameterization of Local Internal Tide-Induced Mixing

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

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

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

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

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

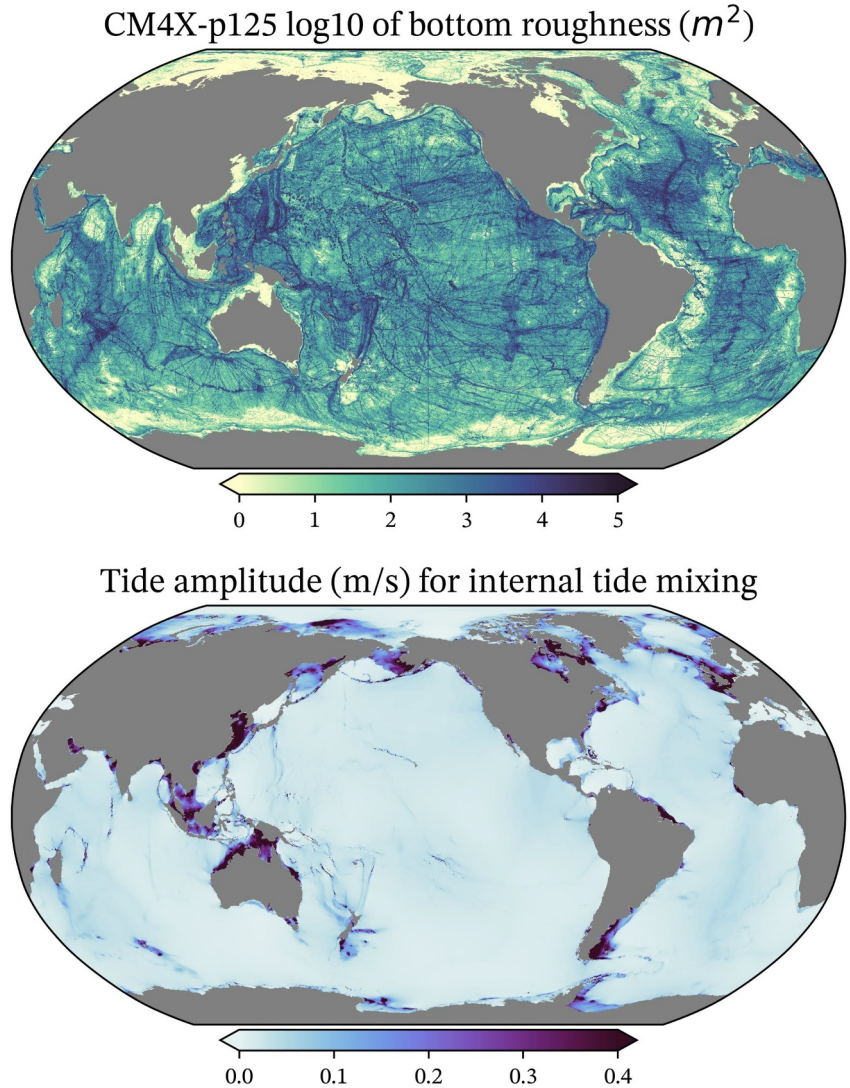

Figure A5. First panel: Log10 of the bottom roughness (units of squared meters) for the bottom topography for CM4X-p125 from Figure A2. This field is used in the tide-induced mixing parameterization of Section A13. It is notable that straight lines appear from ship data, particularly visible in the high latitudes. Second 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.

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 Figures A6a and A6b), 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).

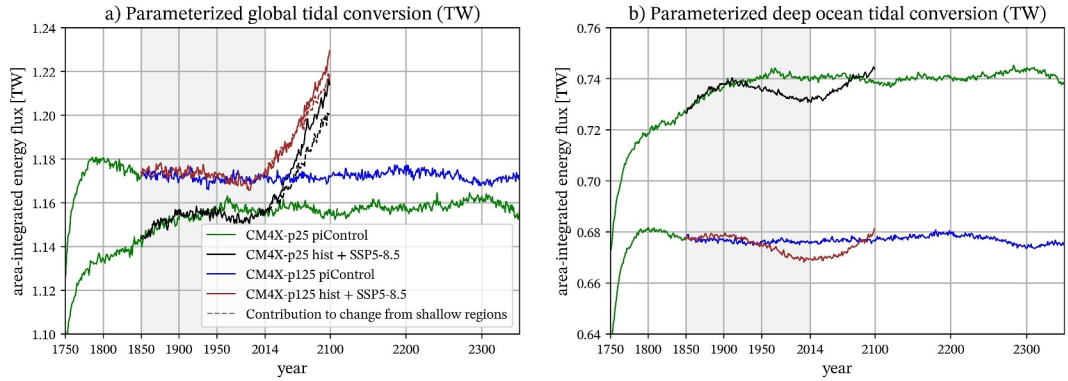

Figure A6. Time series for the energy conversion from the barotropic tide to baroclinic internal tides, as per Equation A3, integrated over (a) the global ocean and (b) regions deeper than 2500 m. The CM4X-p125 piControl simulation equilibrates to roughly 1.17 TW, whereas CM4X-p25 equilibrates to a bit 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).

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 TKE 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.

A14. Parameterization of Baroclinic Eddies in the Mixed Layer

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 SST when testing parameter changes.

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.

We performed non-exhaustive tests in a prototype of CM4X-p25 using a front length of 1,000 m and 2,000 m. Yet we found relatively little sensitivity of the overturning to this parameter compared to the OM4.0 setting of 500 m, thus

Table A7
Summary of the Key MLE Parameters as Used in CM4.0 and CM4X

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

prompting us to retain 500 m for CM4X. In contrast, the ocean heat uptake and SST 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 OM4.0–20 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.

Data Availability Statement

Software comprising the model as well as the software for creating the figures are available from Griffies et al. (2025b). Observation-based data sets 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.
- Data from the RAPID AMOC observing project is funded by the UK Natural Environment Research Council, U.S. National Science Foundation (NSF) and with support from NOAA. The data is freely available from https://rapid.ac.uk/, and we made use of the release documented in Moat et al. (2025).
- Data from the full OSNAP (Overturning in the Subpolar North Atlantic Program) array (Fu et al., 2023) were downloaded from https://www.o-snap.org/. OSNAP data are collected and made freely available by the OSNAP 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 throughout this time. We thank the GFDL management for providing computer resources needed for the development and analysis documented here. Krista Dunne, Sergey Malyshev and Chris Milly kindly provided expertise in helping to update the rivers and lakes for use with the C192 atm coupled to the p125 ocean. We thank Gokhan Danabasoglu and John Fasullo at NCAR for providing the updated Trenberth et al. (2019) heat transports used in Figure 22. We thank John Dunne, Matthew Harrison, Tony Rosati, and Jianjun Yin for very helpful comments on early drafts of this manuscript. We thank Gokhan Danabasoglu, Baylor Fox-Kemper, Malte Jansen, Nathaniel Tarshish, Eli Tziperman, and Carl Wunsch for comments on this work as presented during seminars. We thank the editors (Tapio Schneider and Ian Grooms) and reviewers (Anonymous, Gokhan Danabasoglu, and Peter Korn) for their constructive and valuable reviews. Each of us has been supported either directly or indirectly by NOAA's Oceanic and Atmospheric Research to work at the Geophysical Fluid Dynamics Laboratory and/or at Princeton University through the Cooperative Institute for Modeling the Earth System (CIMES). We offer profound gratitude to NOAA for generously and reliably supporting earth system science research and education at GFDL, CIMES, and the Princeton University Atmospheric and Oceanic Sciences Program. 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

References

Adcroft, A., Anderson, W., Blanton, C., Bushuk, M., Dufour, C. O., Dunne, J. P., et al. (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*, *11*(10), 3167–3211. https://doi.org/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*(3–4), 269–284. https://doi.org/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. https://doi.org/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., et al. (2018). The global precipitation climatology project (GPCP) monthly analysis (new version 2.3) and a review of 2017 global precipitation. *Atmosphere*, *9*(4), 138. https://doi.org/10.3390/atmos9040138

Andrews, T., Gregory, J. M., Paynter, D., Silvers, L. G., Zhou, C., Mauritsen, T., et al. (2018). Accounting for changing temperature patterns increases historical estimates of climate sensitivity. *Geophysical Research Letters*, *45*(16), 8490–8499. https://doi.org/10.1029/2018GL078887

Arbic, B., & Flierl, G. (2004). Baroclinically unstable geostrophic turbulence in the limits of strong and weak bottom ekman friction: Application to midocean eddies. *Journal of Physical Oceanography*, 2257–2273. https://doi.org/10.1175/1520-0485(2004)034⟨2257:BUGTIT⟩2.0.CO;2

Argo. (2023). Argo float data and metadata from Global Data Assembly Centre (Argo GDAC) [Dataset]. *SEANOE*. https://doi.org/10.17882/42182

Armour, K. C., Bitz, C. M., & Roe, G. H. (2013). Time-varying climate sensitivity from regional feedbacks. *Journal of Climate*, *26*(13), 4518–4534. https://doi.org/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 Geoscience*, *9*(7), 549–554. https://doi.org/10.1038/NGEO2731

Arora, V. K., Katavouta, A., Williams, R. G., Jones, C. D., Brovkin, V., Friedlingstein, P., et al. (2019). Carbon-concentration and carbon-climate feedbacks in CMIP6 models, and their comparison to CMIP5 models. *Biogeosciences Discussions*, 1–124. https://doi.org/10.5194/bg-2019-473

Balaji, V., Anderson, J. L., Held, I. M., Winton, M., Durachta, J. W., Malyshev, S., & Stouffer, R. J. (2006). The exchange grid: A mechanism for data exchange between earth system components on independent grids. *Computational Fluid Dynamics 2005: Theory and Applications*, 179–186. https://doi.org/10.1016/B978-044452206-1/50021-5

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*(23), 5815–5826. https://doi.org/10.1175/2007JCLI1688.1

Beadling, R. L. (2023). Global consequences of regional connectivity along the Antarctic margin. *Journal of Geophysical Research: Oceans*, *128*(7), e2023JC019908. https://doi.org/10.1029/2023JC019908

Beadling, R. L., Krasting, J. P., Griffies, S. M., Hurlin, W. J., Bronselaer, B., Russell, J. L., et al. (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*(5), e2021JC017608. https://doi.org/10.1029/2021jc017608

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

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*(1), 2970. https://doi.org/10.1038/s41467-023-38811-z

Blanchard-Wrigglesworth, E., Bitz, C., & Holland, M. (2011). Influence of initial conditions and climate forcing on predicting Arctic sea ice. *Geophysical Research Letters*, *38*(18). https://doi.org/10.1029/2011GL048807

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 M^{2}LInES 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 M^{2}LInES 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.

Bleck, R. (2002). An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates. *Ocean Modelling*, *4*(1), 55–88. https://doi.org/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*(9), 1263–1278. https://doi.org/10.1175/JPO3101.1

Bodner, A., Fox-Kemper, B., Johnson, L., van Rokel, L. P., McWilliams, J. C., Sullivan, P. P., et al. (2023). Modifying the mixed layer eddy parameterization to include frontogenesis arrest by boundary layer turbulence. *Journal of Physical Oceanography*, *53*(1), 323–339. https://doi.org/10.1175/jpo-d-21-0297.1

Boville, B. A., & Gent, P. R. (1998). The NCAR climate system model, version one. *Journal of Climate*, *11*, 1115–1130. https://doi.org/10.1175/1520-0442(1998)011⟨1115:TNCSMV⟩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)*. 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*(5–6), 2093–2113. https://doi.org/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*(7700), 191–196. https://doi.org/10.1038/s41586-018-0006-5

Cess, R., Potter, G., Blanchet, J. P., Boer, G. J., Genio, A. D. D., andV. Dymnikov, M. D., et al. (1990). Intercomparison and interpretation of climate feedback processes in 19 atmospheric general circulation models. *Journal of Geophysical Research*, *95*(D10), 16601–16615. https://doi.org/10.1029/JD095iD10p16601

Cessi, P. (2019). The global overturning circulation. *Annual Review of Marine Science*, *11*(1), 249–270. https://doi.org/10.1146/annurev-marine-010318-095241

Charney, J. G. (1971). Geostrophic turbulence. *Journal of the Atmospheric Sciences*, *28*(6), 1087–1095. https://doi.org/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., et al. (2011). The ERA-interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597. https://doi.org/10.1002/qj.828

Delworth, T. L., Broccoli, A. J., Rosati, A., Stouffer, R. J., Balaji, V., Beesley, J. A., et al. (2006). GFDL's CM2 global coupled climate models - Part 1: Formulation and simulation characteristics. *Journal of Climate*, *19*(5), 643–674. https://doi.org/10.1175/JCLI3629.1

Delworth, T. L., Cooke, W. F., Adcroft, A., Bushuk, M., Chen, J.-H., Dunne, K. A., et al. (2020). SPEAR: The next generation GFDL modeling system for seasonal to multidecadal prediction and projection. *Journal of Advances in Modeling Earth Systems*, *12*(3), e2019MS001895. https://doi.org/10.1029/2019MS001895

Delworth, T. L., Rosati, A., Anderson, W., Adcroft, A. J., Balaji, V., Benson, R., et al. (2012). Simulated climate and climate change in the GFDL CM2.5 high-resolution coupled climate model. *Journal of Climate*, *25*(8), 2755–2781. https://doi.org/10.1175/jcli-d-11-00316.1

Deppenmeier, A.-L., Bryan, F. O., Kessler, W. S., & Thompson, L. (2021). Modulation of cross-isothermal velocities with ENSO in the tropical Pacific Cold Tongue. *Journal of Physical Oceangraphy*, *51*(5), 1559–1574. https://doi.org/10.1175/JPO-D-20-0217.1

Deremble, B., Hogg, A. M., Berloff, P., & Dewar, W. K. (2012). On the application of no-slip lateral boundary conditions to 'coarsely' resolved ocean models. *Ocean Modelling*, *39*(3–4), 411–415. https://doi.org/10.1016/j.ocemod.2011.05.002

Doddridge, E. W., Marshall, D. P., & Hogg, A. M. (2016). Eddy cancellation of the Ekman cell in subtropical gyres. *Journal of Physical Oceanography*, *46*(10), 2995–3010. https://doi.org/10.1175/JPO-D-16-0097.1

Drake, H. F., Bailey, S., Dussin, R., Griffies, S. M., Krasting, J. P., MacGilchrist, G. A., et al. (2025). Water mass transformation budgets in finite-volume generalized vertical coordinate ocean models. *Journal of Advances in Modeling Earth Systems*, *17*(3), 17. https://doi.org/10.1029/2024MS004383

Dunne, J. P., Horowitz, L. W., Adcroft, A., Ginoux, P., Held, I. M., John, J. G., et al. (2020). The GFDL Earth system model version 4.1 (GFDL-ESM4.1): Model description and simulation characteristics. *Journal of Advances in Modeling Earth Systems*, *12*(11), e2019MS002015. https://doi.org/10.1029/2019MS002015

Egbert, G. D., & Erofeeva, S. Y. (2002). Efficient inverse modeling of barotropic ocean tides. *Journal of Atmospheric and Oceanic Technology*, *19*, 183–204. https://doi.org/10.1175/1520-0426(2002)019⟨0183:EIMOBO⟩2.0.CO;2

Egbert, G. D., & Ray, R. D. (2000). Significant dissipation of tidal energy in the deep ocean inferred from satellite altimeter data. *Nature*, *405*(6788), 775–778. https://doi.org/10.1038/35015531

Eyring, V., Bony, S., Meehl, J. A., Senior, C., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organisation. *Geoscientific Model Development*, *9*(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Forster, P., Storelvmo, T., Armour, K., Collins, W., Dufresne, J.-L., Frame, D., et al. (2021). The Earth's energy budget, climate feedbacks, and climate sensitivity. In *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. 923–1054). Cambridge University Press. https://doi.org/10.1017/9781009157896.009

Fox-Kemper, B., Danabasoglu, G., Ferrari, R., Griffies, S. M., Hallberg, R. W., Holland, M. M., et al. (2011). Parameterization of mixed layer eddies. III: Global implementation and impact on ocean climate simulations. *Ocean Modelling*, *39*(1–2), 61–78. https://doi.org/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*(6), 1145–1165. https://doi.org/10.1175/2007JPO3792.1

Fox-Kemper, B., Hewitt, H. T., Xiao, C., Aoalgeirsdóttir, G., Drijfhout, S. S., Edwards, T. L., et al. (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 University Press. https://doi.org/10.1017/9781009157896.011

Fu, Y., Lozier, M. S., Biló, T. C., Bower, A. S., Cunningham, S. A., Cyr, F., et al. (2023). Seasonality of the meridional overturning circulation in the subpolar North Atlantic [Dataset]. *Communications Earth & Environment*, *4*(1). https://doi.org/10.1038/s43247-023-00848-9.181.

Ganachaud, A., & Wunsch, C. (2000). Improved estimates of global ocean circulation, heat transport and mixing from hydrographic data. *Nature*, *408*(6811), 453–456. https://doi.org/10.1038/35044048

Gebbie, G., & Huybers, P. (2019). The little ice age and 20th-century deep Pacific cooling. *Science*, *363*(6422), 70–74. https://doi.org/10.1126/science.aar8413

GEBCO Compilation Group. (2021). GEBCO 2021 grid [Dataset]. https://doi.org/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. https://doi.org/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. https://doi.org/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*(24), 9087–9096. https://doi.org/10.1002/2014GL061671

Gleckler, P. J., Taylor, K. E., & Doutriaux, C. (2008). Performance metrics for climate models. *Journal of Geophysical Research*, *113*(D6). https://doi.org/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*(14), 2604–2616. https://doi.org/10.1175/JCLI3436.1

Gordon, C., Cooper, C., Senior, C., Banks, H., Gregory, J. M., Johns, T. C., et al. (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*(2–3), 147–168. https://doi.org/10.1007/s003820050010

Gregory, J. M. (2000). Vertical heat transports in the ocean and their effect on time-dependent climate change. *Climate Dynamics*, *15*(7), 501–515. https://doi.org/10.1007/s003820000059

Gregory, J. M., Bouttes, N., Griffies, S. M., Haak, H., Hurlin, W. J., Jungclaus, J., et al. (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*(11), 3993–4017. https://doi.org/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. https://doi.org/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. *Climate Dynamics*, *37*(37), 893–914. https://doi.org/10.1007/s00382-010-0847-6

Griffies, S. M., Adcroft, A., Beadling, R. L., Bushuk, M., Chang, C.-Y., Drake, H. F., et al. (2025a). The GFDL-CM4X climate model hierarchy, Part II: Case studies. *Journal of Advances in Modeling Earth Systems*. https://doi.org/10.1029/2024MS004862

Griffies, S. M., Adcroft, A., Beadling, R. L., Bushuk, M., Chang, C.-Y., Drake, H. F., et al. (2025b). Supporting simulated data and code for: The GFDL-CM4X climate model hierarchy, Part I and Part II [Dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.15383537

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*(10), e2019MS001954. https://doi.org/10.1029/2019MS001954

Griffies, S. M., Böning, C. W., Bryan, F. O., Chassignet, E. P., Gerdes, R., Hasumi, H., et al. (2000). Developments in ocean climate modelling. *Ocean Modelling*, *2*(3–4), 123–192. https://doi.org/10.1016/S1463-5003(00)00014-7

Griffies, S. M., Danabasoglu, G., Durack, P. J., Adcroft, A. J., Balaji, V., Böning, C. W., et al. (2016). OMIP contribution to CMIP6: Experimental and diagnostic protocol for the physical component of the ocean model intercomparison project. *Geoscientific Model Development*, *9*, 3231–3296. https://doi.org/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. https://doi.org/10.1016/B978-0-12-391851-2.00020-9

Griffies, S. M., Winton, M., Anderson, W. G., Benson, R., Delworth, T. L., Dufour, C., et al. (2015). Impacts on ocean heat from transient mesoscale eddies in a hierarchy of climate models. *Journal of Climate*, *28*(3), 952–977. https://doi.org/10.1175/JCLI-D-14-00353.1

Griffies, S. M., Yin, J., Durack, P. J., Goddard, P., Bates, S., Behrens, E., et al. (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. https://doi.org/10.1016/j.ocemod.2014.03.004

Groeskamp, S., Griffies, S. M., Iudicone, D., Marsh, R., Nurser, A. J. G., & Zika, J. D. (2019). The water mass transformation framework for ocean physics and biogeochemistry. *Annual Review of Marine Science*, *11*, 1–35. https://doi.org/10.1146/annurev-marine-010318-095421

Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., et al. (2016). High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6. *Geoscientific Model Development*, *9*(11), 4185–4208. https://doi.org/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. https://doi.org/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*(9), 2947–2956. https://doi.org/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*(11), 1609–1614. https://doi.org/10.1175/BAMS-86-11-1609

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

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hieronymus, M., & Nycander, J. (2013). The budgets of heat and salinity in NEMO. *Ocean Modelling*, *67*, 28–38. https://doi.org/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. https://doi.org/10.1175/JCLI-D-18-0438.1

Hirschi, J. J., Barnier, B., Böning, C., Biastoch, A., Blaker, A. T., Coward, A., et al. (2020). The Atlantic meridional overturning circulation in high-resolution models. *Journal of Geophysical Research: Oceans*, *125*(4), e2019JC015522. https://doi.org/10.1029/2019JC015522

Hofstadter, D. R. (1979). *Gödel, Escher, Bach: An Eternal Golden Braid* (p. 777). Basic Books.

Holmes, R. M., Zika, J., & England, M. H. (2019). Diathermal heat transport in a global ocean model. *Journal of Physical Oceanography*, *49*(1), 141–161. https://doi.org/10.1175/JPO-D-18-0098.1

Holmes, R. M., Zika, J., Griffies, S. M., Hogg, A. M. C., Kiss, A. E., & England, M. H. (2020). The geography of numerical mixing in a suite of global ocean models. *Journal of Advances in Modeling Earth Systems*, *13*(7), e2020MS002333. https://doi.org/10.1029/2020MS002333

Holmes, R. M., Zika, J. D., Griffies, S. M., Hogg, A. M., Kiss, A. E., & England, M. H. (2021). The geography of numerical mixing in a suite of global ocean models. *Journal of Advances in Modeling Earth Systems*, *13*(7), e2020MS002333. https://doi.org/10.1029/2020MS002333

Holzer, M., DeVries, T., & de Lavergne, C. (2021). Diffusion controls the ventilation of a pacific shadow zone above abyssal overturning. *Nature Communications*, *12*(1), 4348. https://doi.org/10.1038/s41467-021-24648-x

Hsieh, T., Vecchi, G. A., Yang, W., Held, I. M., & Garner, S. T. (2020). Large-scale control on the frequency of tropical cyclones and seeds: A consistent relationship across a hierarchy of global atmospheric models. *Climate Dynamics*, *55*(11–12), 3177–3196. https://doi.org/10.1007/s00382-020-05446-5

Huang, B., Liu, C., Banzon, V., Freeman, E., Graham, G., Hankins, B., et al. (2020). Improvements of the daily optimum interpolation sea surface temperature (DOISST) Version 2.1. *Journal of Climate*, *34*(8), 2923–2939. https://doi.org/10.1175/JCLI-D-20-0166.1

Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., et al. (2017). NOAA extended reconstructed sea surface temperature (ERSST), version 5 [Dataset]. *U.S. Department of Commerce, Washington, D.C.: NOAA National Centers for Environmental Information*. https://doi.org/10.7289/V5T72FNM

Hughes, C. W., Fukumori, I., Griffies, S. M., Huthnance, J. M., Minobe, S., Spence, P., et al. (2019). Sea level and the role of coastal trapped waves in mediating the influence of the open ocean on the coast. *Surveys in Geophysics*, *40*(6), 1467–1492. https://doi.org/10.1007/s10712-019-09535-x

Huw Davies, J. (2013). Global map of solid Earth surface heat flow. *Geochemistry, Geophysics, Geosystems*, *14*(10), 4608–4622. https://doi.org/10.1002/ggge.20271

Hyder, P., Edwards, J. M., Allan, R. P., Hewitt, H. T., Bracegirdle, T. J., Gregory, J. M., et al. (2018). Critical Southern Ocean climate model biases traced to atmospheric model cloud errors. *Nature Communications*, *9*(1), 3625. https://doi.org/10.1038/s41467-018-05634-2

Ilicak, M., Adcroft, A. J., Griffies, S. M., & Hallberg, R. W. (2012). Spurious dianeutral mixing and the role of momentum dissipation. *Ocean Modelling*, *45*(46), 37–58. https://doi.org/10.1016/j.ocemod.2011.10.003

Irving, D., Hobbs, W., Church, J., & Zika, J. (2021). A mass and energy conservation analysis of drift in the CMIP6 ensemble. *Journal of Climate*, *34*(8), 3157–3170. https://doi.org/10.1175/JCLI-D-20-0281.1

Ivey, G., & Imberger, J. (1991). On the nature of turbulence in a stratified fluid. Part I: The energetics of mixing. *Journal of Physical Oceanography*, *21*(5), 650–658. https://doi.org/10.1175/1520-0485(1991)021⟨0650:OTNOTI⟩2.0.CO;2

Jackson, L., Hallberg, R. W., & Legg, S. (2008). A parameterization of shear-driven turbulence for ocean climate models. *Journal of Physical Oceanography*, *38*(5), 1033–1053. https://doi.org/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*(4), 1760–1771. https://doi.org/10.1002/2017MS001038

Johnson, G. C., & Purkey, S. G. (2024). Refined estimatesof global ocean deep and abyssal decadal warming trends. *Geophysical Research Letters*, *51*(18), e2024GL111229. https://doi.org/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*, *15*(2), 15. https://doi.org/10.1029/2022MS003447

Khatiwala, S. (2024). Efficient spin-up of Earth System Models using sequence acceleration. *Science Advances*, *10*(18), eadn2839. https://doi.org/10.1126/sciadv.adn2839

Kirtman, B. P., Bitz, C., Bryan, F., Collins, W., Dennis, J., Hearn, N., et al. (2012). Impact of ocean model resolution on CCSM climate simulations. *Climate Dynamics*, *39*(6), 1303–1328. https://doi.org/10.1007/s00382-012-1500-3

Kiss, A. E., Hogg, A. M., Hannah, N., Dias, F. B., Brassington, G., Chamberlain, M. A., et al. (2020). ACCESS-OM2: A global ocean-sea ice model at three resolutions. *Geoscientific Model Development*, *13*(2), 401–442. https://doi.org/10.5194/gmd-13-401-2020

Kjellsson, J., & Zanna, L. (2017). The impact of horizontal resolution on energy transfers in global ocean models. *Fluid*, *2*(3), 45. https://doi.org/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*, *37*(24), 6563–6583. https://doi.org/10.1175/JCLI-D-23-0591.1

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. https://doi.org/10.1175/JCLI-D-18-0035.1

Lauritzen, P. H., Kevlahan, N. K.-R., Toniazzo, T., Eldred, C., Dubos, T., Gassmann, A., et al. (2022). Reconciling and improving formulations for thermodynamics and conservation principles in Earth System Models (ESMs). *Journal of Advances in Modeling Earth Systems*, *14*(9), e2022MS003117. https://doi.org/10.1029/2022MS003117

Lee, J., Gleckler, P. J., Ahn, M.-S., Ordonez, A., Ullrich, P. A., Sperber, K. R., et al. (2024). Objective evaluation of earth system models: PCMDI metrics package (PMP) version 3. *Geoscientific Model Development*, *17*(9), 3919–3948. https://doi.org/10.5194/gmd-17-3919-2024

Legg, S., Hallberg, R. W., & Girton, J. B. (2006). Comparison of entrainment in overflows simulated by z-coordinate, isopycnal and non-hydrostatic models. *Ocean Modelling*, *11*(1), 69–97. https://doi.org/10.1016/j.ocemod.2004.11.006

Lellouche, J.-M., Greiner, E., Bourdallé-Badie, R., Garric, G., Melet, A., Drévillon, M., et al. (2021). The Copernicus global 1/12 oceanic and sea ice GLORYS12 reanalysis. *Frontiers in Earth Science*, *9*, 698876. https://doi.org/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*(19), 7630–7640. https://doi.org/10.1175/JCLI-D-15-0168.1

Li, Q., Fox-Kemper, B., Øyvind, B., & Webb, A. (2017). Statistical models of global Langmuir mixing. *Ocean Modelling*, *113*, 95–114. https://doi.org/10.1016/j.ocemod.2017.03.016

Li, Q., & Lee, S. (2017). A mechanism of mixed layer formation in the Indo-Western Pacific Southern Ocean: Preconditioning by an eddy-driven jet-scale overturning circulation. *Journal of Physical Oceanography*, *47*(11), 2755–2772. https://doi.org/10.1175/JPO-D-17-0006.1

Lindzen, R. S., & Fox-Rabinovitz, M. (1989). Consistent horizontal and vertical resolution. *Monthly Weather Review*, *117*, 2575–2583. https://doi.org/10.1175/1520-0493(1989)117⟨2575:CVAHR⟩2.0.CO;2

Locarnini, R. A., Mishonov, A. V., Antonov, J. I., Boyer, T. P., Garcia, H. E., Baranova, O. K., et al. (2018). World Ocean Atlas 2013, volume 1: Temperature [Dataset]. *U.S. Department of Commerce, Washington, D.C.: NOAA/NESDIS*. https://doi.org/10.7289/V55X26VD

Loeb, N. G., Doelling, D. R., Wang, H., Su, W., Nguyen, C., Corbett, J. G., et al. (2018). Clouds and the Earth's radiant energy system (CERES) energy balanced and filled (EBAF) top-of-atmosphere (TOA) edition-4.0 data product. *International Journal of Climatology*, *31*(2), 895–918. https://doi.org/10.1175/JCLI-D-17-0208.1

MacKinnon, J. A., St. Louis, L., & Naveira Garabato, A. C. (2013). Diapycnal mixing processes in the ocean interior. In G. Siedler, S. M. Griffies, J. Gould, & J. Church (Eds.), *Ocean circulation and climate, 2nd edition: A 21st century perspective* (Vol. 103, pp. 159–183). Academic Press. https://doi.org/10.1016/B978-0-12-391851-2.00007-6

MacKinnon, J. A., Zhao, Z., Whalen, C. B., Waterhouse, A. F., Trossman, D. S., Sun, O. M., et al. (2017). Climate process team on internal-wave driven ocean mixing. *Bulletin of the American Meteorological Society*, *98*(11), 2429–2454. https://doi.org/10.1175/BAMS-D-16-0030.1

Maher, P., Gerber, E. P., Medeiros, B., Merlis, T. M., Sherwood, S., Sheshadri, A., et al. (2019). Model hierarchies for understanding atmospheric circulation. *Reviews of Geophysics*, *57*(2), 250–280. https://doi.org/10.1029/2018RG000607

Manabe, S., Bryan, K., & Spelman, M. J. (1990). Transient response of a global ocean–atmosphere model to a doubling of atmospheric carbon dioxide. *Journal of Physical Oceanography*, *20*, 722–749. https://doi.org/10.1175/1520-0485(1990)020⟨0722:TROAGO⟩2.0.CO;2

Manabe, S., & Stouffer, R. J. (1988). Two stable equilibria of a coupled ocean-atmosphere model. *Journal of Climate*, *1*(9), 841–866. https://doi.org/10.1175/1520-0442(1988)001⟨0841:TSEOAC⟩2.0.CO;2

Manabe, S., Stouffer, R. J., Spelman, M. J., & Bryan, K. (1991). Transient responses of a coupled ocean–atmosphere model to gradual changes of atmospheric CO2. Part 1: Annual mean response. *Journal of Climate*, *4*, 785–818. https://doi.org/10.1175/1520-0442(1991)004⟨0785:TROACO⟩2.0.CO;2

Marques, G. M., Loose, N., Yankovsky, E., Steinberg, J. M., Chang, C.-Y., Bhamidipati, N., et al. (2022). Neverworld2: An idealized model hierarchy to investigate ocean mesoscale eddies across resolutions. *Geoscientific Model Development*, *15*(17), 6567–6579. https://doi.org/10.5194/gmd-15-6567-2022

Marshall, D. P., & Zanna, L. (2014). A conceptual model of ocean heat uptake under climate change. *Journal of Climate*, *27*(22), 8444–8465. https://doi.org/10.1175/JCLI-D-13-00344.1

Marshall, J., & Schott, F. (1999). Open-ocean convection: Observations, theory, and models. *Reviews of Geophysics*, *37*, 1–64. https://doi.org/10.1029/98RG02739

McDougall, T. J., Barker, P. M., Holmes, R. M., Pawlowicz, R., Griffies, S. M., & Durack, P. J. (2021). The interpretation of temperature and salinity variables in numerical ocean model output, and the calculation of heat fluxes and heat content. *Geoscientific Model Development*, *14*(10), 6445–6466. https://doi.org/10.5194/gmd-14-6445-2021

Melet, A., Hallberg, R. W., Legg, S., & Polzin, K. (2013). Sensitivity of the Pacific Ocean state to the vertical distribution of internal-tide driven mixing. *Journal of Physical Oceanography*, *43*(3), 602–615. https://doi.org/10.1175/JPO-D-12-055.1

Melet, A. V., Hallberg, R., & Marshall, D. P. (2022). Chapter 2 - The role of ocean mixing in the climate system. In M. Meredith & A. Naveira Garabato (Eds.), *Ocean mixing* (pp. 5–34). Elsevier. https://doi.org/10.1016/B978-0-12-821512-8.00009-8

Moat, B. I., Smeed, D. A., Rayner, D., Johns, W. E., Smith, R., Volkov, D., et al. (2025). Atlantic meridional overturning circulation observed by the RAPID-MOCHA-WBTS (RAPID-Meridional overturning circulation and heatflux array-western boundary time series) array at 26N from 2004 to 2023 (v2023.1a) [Dataset]. *British Oceanographic Data Centre - Natural Environment Research Council, UK*. https://doi.org/10.5285/33826d6e-801c-b0a7-e063-7086abc0b9db

Morlighem, M., Williams, C. N., Rignot, E., An, L., Arndt, J. E., Bamber, J. L., et al. (2017). Bedmachine v3: Complete bed topography and ocean bathymetry mapping of Greenland from multibeam echo sounding combined with mass conservation. *Geophysical Research Letters*, *44*(21), 11051–11061. https://doi.org/10.1002/2017GL074954

Morrison, A. K., Saenko, O. A., Hogg, A. M., & Spence, P. (2013). The role of vertical eddy flux in southern ocean heat uptake. *Geophysical Research Letters*, *40*(20), 5445–5450. https://doi.org/10.1002/2013GL057706

Murray, R. J. (1996). Explicit generation of orthogonal grids for ocean models. *Journal of Computational Physics*, *126*(2), 251–273. https://doi.org/10.1006/jcph.1996.0136

Newsom, E., Zanna, L., & Gregory, J. (2023). Background pycnocline depth constrains future ocean heat uptake efficiency. *Geophysical Research Letters*, *50*(22), e2023GL105673. https://doi.org/10.1029/2023GL105673

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

Osborn, T. R., & Cox, C. S. (1972). Oceanic fine structure. *Geophysical Fluid Dynamics*, *3*(4), 321–345. https://doi.org/10.1080/03091927208236085

Pacanowski, R. C., & Philander, G. (1981). Parameterization of vertical mixing in numerical models of the tropical ocean. *Journal of Physical Oceanography*, *11*, 1442–1451. https://doi.org/10.1175/1520-0485(1981)011⟨1443:POVMIN⟩2.0.CO;2

Polzin, K. L. (2009). An abyssal recipe. *Ocean Modelling*, *30*(4), 298–309. https://doi.org/10.1016/j.ocemod.2009.07.006

Pudig, M. P., Yung, C. K., Zika, J. D., & Holmes, R. M. (2023). Rectified ocean heat uptake from oscillatory surface forcing. *Journal of Climate*, *36*(8), 2663–2680. https://doi.org/10.1175/JCLI-D-22-0267.1

Purkey, S., & Johnson, G. (2010). Warming of global abyssal and deep Southern Ocean waters between the 1990s and 2000s: Contributions to global heat and sea level rise budgets. *Journal of Climate*, *23*, 6336–6351. https://doi.org/10.1175/2010JCLI3682.1

Purkey, S., & Johnson, G. (2012). Global contraction of Antarctic bottom water between the 1980s and 2000s. *Journal of Climate*, *25*(17), 5830–5844. https://doi.org/10.1175/JCLI-D-11-00612.1

Purkey, S., & Johnson, G. (2013). Antarctic Bottom Water warming and freshening: Contributions to sea level rise, ocean freshwater budgets, and global heat gain. *Journal of Climate*, *26*(16), 6105–6122. https://doi.org/10.1175/JCLI-D-12-00834.1

Redi, M. H. (1982). Oceanic isopycnal mixing by coordinate rotation. *Journal of Physical Oceanography*, *12*(10), 1154–1158. https://doi.org/10.1175/1520-0485(1982)012⟨1154:OIMBCR⟩2.0.CO;2

Reichl, B. G., Adcroft, A., Griffies, S. M., & Hallberg, R. W. (2022). A potential energy analysis of ocean surface mixed layers. *Journal of Geophysical Research-Oceans*, *127*(7), e2021JC018140. https://doi.org/10.1029/2021JC018140

Reichl, B. G., & Hallberg, R. W. (2018). A simplified energetics based planetary boundary layer (ePBL) approach for ocean climate simulations. *Ocean Modelling*, *132*, 112–129. https://doi.org/10.1016/j.ocemod.2018.10.004

Reichl, B. G., & Li, Q. (2019). A parameterization with a constrained potential energy conversion rate of vertical mixing due to Langmuir turbulence. *Journal of Physical Oceanography*, *49*(11), 2935–2959. https://doi.org/10.1175/JPO-D-18-0258.1

Reichl, B. G., Wittenberg, A. T., Griffies, S. M., & Adcroft, A. (2024). Improving equatorial upper ocean vertical mixing in the noaa/gfdl om4 model. *Earth and Space Science*, *11*(10), e2023EA003485. https://doi.org/10.1029/2023EA003485

Rhein, M., Rintoul, S., Aoki, S., Campos, E., Chambers, D., Feely, R., et al. (2013). Observations: Ocean. In *Climate change 2013: The physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change* (pp. 255–315). University Press. https://doi.org/10.1017/CBO9781107415324.010

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. https://doi.org/10.1002/wcc.338

Roberts, M. J., Baker, A., Blockley, E. W., Calvert, D., Coward, A., Hewitt, H. T., et al. (2019). Description of the resolution hierarchy of the global coupled HadGEM3-GC3.1 model as used in CMIP6 HighResMIP experiments. *Geoscientific Model Development*, *12*, 4999–5028. https://doi.org/10.5194/gmd-12-4999-2019

Rugenstein, M., Bloch-Johnson, J., Gregory, J., Andrews, T., Mauritsen, T., Li, C., et al. (2020). Equilibrium climate sensitivity estimated by equilibrating climate models. *Geophysical Research Letters*, *47*(4), e2019GL083898. https://doi.org/10.1029/2019GL083898

Saba, V. S., Griffies, S., Anderson, W., Winton, M., Alexander, M., Delworth, T., et al. (2016). Enhanced warming of the northwest Atlantic Ocean under climate change. *Journal of Geophysical Research–Oceans*, *121*(1), 118–132. https://doi.org/10.1002/2015JC011346

Sadourny, R. (1975). The dynamics of finite-difference models of the shallow-water equations. *Journal of the Atmospheric Sciences*, *32*(4), 680–689. https://doi.org/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*(4), 1830–1844. https://doi.org/10.1002/jgrc.20135

Salmon, R. (1980). Baroclinic instability and geostrophic turbulence. *Geophysical & Astrophysical Fluid Dynamics*, *15*(1), 167–211. https://doi.org/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. https://doi.org/10.1029/2023MS003890

Sausen, R., Barthel, K., & Hasselmann, K. (1988). Coupled ocean-atmosphere models with flux correction. *Climate Dynamics*, *2*(3), 145–163. https://doi.org/10.1007/BF01053472

Scott, R. B., & Wang, F. (2005). Direct evidence of an oceanic inverse kinetic energy cascade from satellite altimetry. *Journal of Physical Oceanography*, *35*(9), 1650–1666. https://doi.org/10.1175/JPO2771.1

Séférian, R., Gehlen, M., Bopp, L., Resplandy, L., Orr, J. C., Marti, O., et al. (2016). Inconsistent strategies to spin up models in CMIP5: Implications for ocean biogeochemical model performance assessment. *Geoscientific Model Development*, *9*(5), 1827–1851. https://doi.org/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*(3–4), 245–263. https://doi.org/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. https://doi.org/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*(3), 233–234. https://doi.org/10.1175/1520-0485(1971)001⟨0233:OTROIM⟩2.0.CO;2

Stewart, A. L., Chi, X., Solodoch, A., & Hogg, A. M. (2021). High-frequency fluctuations in Antarctic bottom water transport driven by Southern Ocean winds. *Geophysical Research Letters*, *49*(17), e2021GL094569. https://doi.org/10.1029/2021GL094569

Stewart, K., Hogg, A., Griffies, S. M., Heerdegen, A., Ward, M., Spence, P., & England, M. (2017). Vertical resolution of baroclinic modes in global ocean models. *Ocean Modelling*, *113*, 50–65. https://doi.org/10.1016/j.ocemod.2017.03.012

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. https://doi.org/10.1029/2002GL015

Stouffer, R. J. (2004). Time scales of climate response. *Journal of Climate*, *31*(1), 209–217. https://doi.org/10.1175/1520-0442(2004)017⟨0209:TSOCR⟩2.0.CO;2

Stouffer, R. J., Russel, J., & Spelman, M. J. (2006). Importance of oceanic heat uptake in transient climate change. *Geophysical Research Letters*, *33*(17). https://doi.org/10.1029/2006GL027242

Tesdal, J.-E., MacGilchrist, G. A., Beadling, R. L., Griffies, S. M., Krasting, J. P., & Durack, P. J. (2023). Revisiting interior water mass responses to surface forcing changes and the subsequent effects on overturning in the Southern Ocean. *Journal of Geophysical Research - Oceans*, *128*(3), e2022JC019105. https://doi.org/10.1029/2022JC019105

Thomas, M. D., & Federov, A. V. (2017). The eastern subtropical Pacific origin of the equatorial cold bias in climate models: A Lagrangian perspective. *Journal of Climate*, *30*(15), 5885–5900. https://doi.org/10.1175/JCLI-D-16-0819.1

Trenberth, K. E., & Caron, J. M. (2001). Estimates of meridional atmosphere and ocean heat transports. *Journal of Climate*, *14*, 3433–3443. https://doi.org/10.1175/1520-0442(2001)014⟨3433:EOMAAO⟩2.0.CO;2

Trenberth, K. E., Zhang, Y., Fasullo, J. T., & Cheng, L. (2019). Observation-based estimates of global and basin ocean meridional heat transport time series. *Journal of Climate*, *32*(14), 4567–4583. https://doi.org/10.1175/JCLI-D-18-0872.1

Tsujino, H., Urakawa, S., Nakano, H., Small, R. J., Kim, W. M., Yeager, S. G., et al. (2018). JRA-55 based surface dataset for driving ocean-sea-ice models (JRA-55-do). *Ocean Modelling*, *130*, 79–139. https://doi.org/10.1016/j.ocemod.2018.07.002

Vallis, G. K. (2017). *Atmospheric and oceanic fluid dynamics: Fundamentals and large-scale circulation* (2nd ed.). Cambridge University Press. (946 + xxv pp).

von Schuckmann, K., Minière, A., Gues, F., Cuesta-Valero, F. J., Kirchengast, G., Adusumilli, S., et al. (2023). Heat stored in the earth system 1960–2020: Where does the energy go? *Earth System Science Data*, *15*(4), 1675–1709. https://doi.org/10.5194/essd-15-1675-2023

von Storch, J.-S., Haak, H., Hertwig, E., & Fast, I. (2016). Vertical heat and salt fluxes due to resolved and parameterized meso-scale eddies. *Ocean Modelling*, *108*, 1–19. https://doi.org/10.1016/j.ocemod.2016.10.001

Wang, C., Zhang, L., Lee, S.-K., Wu, L., & Mechoso, C. R. (2014). A global perspective on CMIP5 climate model biases. *Nature Climate Change*, *4*(3), 201–205. https://doi.org/10.1038/NCLIMATE2118

Weaver, A. J., & Hughes, T. M. C. (1996). On the incompatibility of ocean and atmosphere models and the need for flux adjustments. *Climate Dynamics*, *12*(3), 141–170. https://doi.org/10.1007/BF00211615

Welch, P. (1967). The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, *15*(2), 70–73. https://doi.org/10.1109/TAU.1967.1161901

Wills, R. C. J., Dong, Y., Proistosecu, C., Armour, K. C., & Battisti, D. S. (2022). Systematic climate model biases in the large-scale patterns of recent sea-surface temperature and sea-level pressure change. *Geophysical Research Letters*, *49*(17), e2022GL100011. https://doi.org/10.1029/2022GL100011

Wise, A., Hughes, C. W., & Polton, J. A. (2018). Bathymetric influence on the coastal sea level response to ocean gyres at western boundaries. *Journal of Physical Oceanography*, *481*(12), 2949–2964. https://doi.org/10.1175/JPO-D-18-0007.1

Wise, A., Hughes, C. W., Polton, J. A., & Huthnance, J. M. (2020). Leaky slope waves and sea level: Unusual consequences of the beta effect along western boundaries with bottom topography and dissipation. *Journal of Physical Oceanography*, *50*(1), 217–237. https://doi.org/10.1175/JPO-D-19-0084.1

Wittenberg, A. T. (2009). Are historical records sufficient to constrain ENSO simulations? *Geophysical Research Letters*, *36*(12). https://doi.org/10.1029/2009GL038710

Wolfe, C. L., Cessi, P., McClean, J. L., & Maltrud, M. E. (2008). Vertical heat transport in eddying ocean models. *Geophysical Research Letters*, *35*(23). https://doi.org/10.1029/2008GL036138

Wright, D. G. (1997). An equation of state for use in ocean models: Eckart's formula revisited. *Journal of Atmospheric and Oceanic Technology*, *14*(3), 735–740. https://doi.org/10.1175/1520-0426(1997)014⟨0735:AEOSFU⟩2.0.CO;2

Xu, X., Chassignet, E., & Wallcraft, A. (2023). Impact of vertical resolution on representing baroclinic modes and water mass distribution in the North Atlantic circulation. *Ocean Modelling*, *186*, 102261. https://doi.org/10.1016/j.ocemod.2023.102261

Yankovsky, E., Zanna, L., & Smith, K. S. (2022). Influences of mesoscale ocean eddies on flow vertical structure in a resolution-based model hierarchy. *Journal of Advances in Modeling Earth Systems*, *14*(11), e2022MS003203. https://doi.org/10.1029/2022MS003203

Yeager, S., Castruccio, F., Chang, P., Danabasoglu, G., Maroon, E., Small, J., et al. (2021). An outsized role for the Labrador Sea in the multidecadal variability of the Atlantic overturning circulation. *Science Advances*, *7*(41), eabh3592. https://doi.org/10.1126/sciadv.abh3592

Yin, J., Griffies, S. M., Winton, M., Zhao, M., & Zanna, L. (2020). Response of storm-related extreme sea level along the us Atlantic Coast to combined weather and climate forcing. *Journal of Climate*, *33*(9), 3745–3769. https://doi.org/10.1175/JCLI-D-19-0551.1

Zampieri, L., Kauker, F., Fröhle, J., Sumata, H., Hunke, E. C., & Goessling, H. F. (2021). Impact of sea-ice model complexity on the performance of an unstructured-mesh sea-ice/ocean model under different atmospheric forcings. *Journal of Advances in Modeling Earth Systems*, *13*(5), e2020MS002438. https://doi.org/10.1029/2020MS002438

Zanna, L., Khatiwala, S., Gregory, J., Ison, J., & Heimbach, P. (2019). Global reconstruction of historical ocean heat storage and transport. *Proceedings of the National Academy of Sciences*, *116*(4), 1126–1131. https://doi.org/10.1073/pnas.1808838115

Zhang, L., Delworth, T. L., Cooke, W., Goosse, H., Bushuk, M., Morioka, Y., & Yang, X. (2021). The dependence of internal multidecadal variability in the Southern Ocean on the ocean background mean state. *Journal of Climate*, *34*(3), 1061–1080. https://doi.org/10.1175/JCLI-D-20-0049.1

Zhang, R., Delworth, T. L., Rosati, A., Anderson, W. G., Dixon, K. W., Lee, H.-C., & Zeng, F. (2011). Sensitivity of the North Atlantic Ocean circulation to an abrupt change in the Nordic Sea overflow in a high resolution global coupled climate model. *Journal of Geophysical Research*, *116*(C12), C12024. https://doi.org/10.1029/2011JC007240

Zhang, R., Sutton, R., Danabasoglu, G., Marsh, R., Yeager, S. G., Amrhein, D. E., & Little, C. M. (2019). A review of the role of the Atlantic meridional overturning circulation in Atlantic multidecadal variability and associated climate impacts. *Reviews of Geophysics*, *57*(2), 316–375. https://doi.org/10.1029/2019RG000644

Zhao, M. (2020). Simulations of atmospheric rivers, their variability, and response to global warming using GFDL's new high-resolution general circulation model. *Journal of Climate*, *33*(23), 10287–10303. https://doi.org/10.1175/JCLI-D-20-0241.1

Zhao, M. (2022a). An investigation of the effective climate sensitivity in GFDL's new climate models CM4.0 and SPEAR. *Journal of Climate*, *35*(17), 5637–5660. https://doi.org/10.1175/jcli-d-21-0327.1

Zhao, M. (2022b). A study of AR-TS-and MCS-Associated precipitation and extreme precipitation in present and warmer climates. *Journal of Climate*, *35*(2), 479–497. https://doi.org/10.1175/JCLI-D-21-0145.1

Zhao, M., Golaz, J.-C., Held, I. M., Guo, H., Balaji, V., Benson, R., et al. (2018a). The GFDL global atmosphere and land model AM4.0/LM4.0 - Part II: Model description, sensitivity studies, and tuning strategies. *Journal of Advances in Modeling Earth Systems*, *10*(3), 735–769. https://doi.org/10.1002/2017MS001209

Zhao, M., Golaz, J.-C., Held, I. M., Guo, H., Balaji, V., Benson, R., et al. (2018b). The GFDL global atmosphere and land model AM4.0/LM4.0 - Part I: Simulation characteristics with prescribed SSTs. *Journal of Advances in Modeling Earth Systems*, *10*(3), 691–734. https://doi.org/10.1002/2017MS001208

Zika, J. D., Le Sommer, J., Dufour, C. O., Molines, J.-M., Barnier, B., Brasseur, P., et al. (2014). Vertical eddy fluxes in the Southern Ocean. *Journal of Physical Oceanography*, *43*(5), 941–955. https://doi.org/10.1175/JPO-D-12-0178.1

Zweng, M. M., Reagan, J. R., Antonov, J. I., Locarnini, R. A., Mishonov, A. V., Boyer, T. P., et al. (2018). World Ocean Atlas 2013, volume 2: Salinity [Dataset]. *U.S. Department of Commerce, Washington, D.C.: NOAA/NESDIS*. https://doi.org/10.7289/V5251G4D