Geophysical Research Letters

RESEARCH LETTER
10.1029/2025GL117040

Evaluation of Quasi-Equilibrium Criteria for Coupled Climate Model Simulations

Agatha M. de Boer[1,2], Srinath Krishnan[1,2,3], Natalie J. Burls[4], David K. Hutchinson[5], and Martin Renoult[1,2]

[1]Department of Geological Sciences, Stockholm University, Stockholm, Sweden, [2]Bolin Centre for Climate Research, Stockholm University, Stockholm, Sweden, [3]Now at: CICERO Center for International Climate Research, Oslo, Norway, [4]Department of Atmospheric, Oceanic and Earth Sciences, George Mason University, Fairfax, VA, USA, [5]Climate Change Research Centre, University of New South Wales, Sydney, NSW, Australia

Abstract We evaluate five commonly-applied criteria to validate that a climate model is in so-called "quasi-equilibrium," using a suite of five simulations with CO_2 concentrations between 1× and 16× Pre-Industrial values. We find that major changes in ocean circulation can occur after common thermal equilibrium criteria are reached, such as a small Top of Atmosphere radiative flux imbalance, or weak trends in surface air temperature, sea surface temperature, and deep ocean temperature. Ocean circulation change, in turn, impact high-latitude SAT, sea ice, and the Inter-tropical Convergence Zone position. For future modeling studies and intercomparison projects aiming for an ocean in quasi-equilibrium, we suggest that time series of key meridional overturning circulation (MOC) metrics in the Atlantic, Pacific, and Southern Ocean are saved, and that MOC trends are less than 1 Sv/1000 years, and DOT trends less than 0.1°C/century for the final 1000 years of the simulations.

Plain Language Summary Climate models are time consuming and expensive to run. Therefore, simulations of different climate states should ideally not be longer than the minimum length needed to draw reliable conclusions. However, there is currently no consensus or general guidance about what such a minimum length should be. Here, we tested five common ways scientists check whether a climate model has reached a state that is close enough to the expected final answer such that the conclusions will not change if the simulation is extended. We find that even when air temperature or sea surface temperature seem to have reached their final values, the deep ocean circulation can still shift significantly. These changes in ocean circulation can then affect temperatures near the poles, sea ice coverage, and the location of the tropical rain belt. We suggest refined methods to verify that the final state of the ocean circulation in a simulation has been reached.

Key Points:

- Low top of atmosphere radiative flux imbalance or weak trends in atmosphere or deep-ocean temperatures are not sufficient indicators of ocean stability
- Switches in the ocean's circulation state impact high-latitude surface temperature, sea ice, and the position of the inter-tropical convergence zone
- Simulation-length time series of key ocean overturning metrics is needed to verify the equilibration state of the ocean circulation

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

A. M. de boer,
agatha.deboer@geo.su.se

Citation:

de Boer, A. M., Krishnan, S., Burls, N. J., Hutchinson, D. K., & Renoult, M. (2025). Evaluation of quasi-equilibrium criteria for coupled climate model simulations. *Geophysical Research Letters*, *52*, e2025GL117040. https://doi.org/10.1029/2025GL117040

Received 21 MAY 2025
Accepted 8 OCT 2025

Author Contributions:

Conceptualization: Agatha M. de Boer
Data curation: Srinath Krishnan, Natalie J. Burls
Formal analysis: Srinath Krishnan
Funding acquisition: Agatha M. de Boer
Investigation: Agatha M. de Boer, Srinath Krishnan, Natalie J. Burls, David K. Hutchinson, Martin Renoult
Methodology: Agatha M. de Boer, Natalie J. Burls, David K. Hutchinson, Martin Renoult

1. Introduction

Coupled climate models simulate land, ocean, and atmosphere systems to replicate past climates or test sensitivities to variable forcing, boundary conditions or model parameterizations (De Boer et al., 2008; Hutchinson et al., 2018; Klinger et al., 2003). Climate simulations often have an initial rapid response to adjust to new boundary conditions (e.g., a quadrupling of atmospheric CO_2), followed by a longer adjustment that can take hundreds to thousands of years (Bonan et al., 2022; Curtis & Fedorov, 2024a; Rugenstein et al., 2016). While the real ocean and atmosphere are never in equilibrium, because their external environment continuously change faster than they can adapt, their model counterparts with fixed forcing and boundary conditions (i.e., non-transient) should theoretically be able to reach a state of equilibrium. We define such (full) equilibrium here as the point when the statistics (i.e., mean state and variability) of the physical variables are invariant to longer simulation length. Atmospheric equilibrium is often assumed if there is no visual or statistical trend remaining in the global mean SAT (Armstrong et al., 2016) while, similarly, the ocean is argued to be equilibrated when trends are zero in one or more of the following metrics; global mean ocean temperature, SST, upper ocean temperature, DOT, deep ocean salinity, ideal age tracer or the maximum of the ocean's MOC (Fischer & Jungclaus, 2010; Galbraith & de Lavergne, 2019; Goldner et al., 2014; Hewitt & Mitchell, 1998; Hutchinson et al., 2018; Klockmann et al., 2018; Liu et al., 2002; Liu et al., 2003; Zhu et al., 2015). Alternatively, the net heat fluxes across the system components, such as the residual flux at the TOA, are required to be zero (Gregory et al., 2004). For many studies, such as high-resolution modeling or when several sensitivity experiments are required, it may be

Project administration: Agatha M. de Boer
Resources: Natalie J. Burls
Supervision: Agatha M. de Boer
Validation: Srinath Krishnan
Visualization: Agatha M. de Boer, Srinath Krishnan
Writing – original draft: Agatha M. de Boer, Srinath Krishnan
Writing – review & editing: Agatha M. de Boer, Srinath Krishnan, Natalie J. Burls, David K. Hutchinson, Martin Renoult

optimal to terminate simulations earlier, at so-called quasi-equilibrium, which we formally define here as the point when longer simulation time is not expected to alter the conclusions anymore.

"Quasi-equilibrium" can differ for different studies, depending on what aspects of the system is being investigated. Usually, trends in relevant system variables (e.g., SAT for atmospheric studies or DOT for ocean studies) or the final TOA flux imbalance are argued to be sufficiently small for robust results. However, despite the obvious incentive to minimize model integration time without losing fidelity, there appears to be no benchmark study to suggest what an appropriate cut-off value should be for these criteria, with a wide range of measures commonly applied. Climate model intercomparison projects (CMIP) such as CMIP5 and CMIP6 generally leave it to the modeling groups to decide which criteria to use and merely provide a minimum simulation length (Eyring et al., 2016; Haywood et al., 2016; Lunt et al., 2017) and/or encourage participants to save output variables that allow for the application of various equilibrium tests (Kageyama et al., 2017; Otto-Bliesner et al., 2017). This leads to modeling groups running their simulations to different degrees of equilibrium, thus complicating the interpretation of their results, especially with regard to ocean circulation. For example, simulations that were too short in the Last Glacial Maximum contributions to the Paleoclimate Intercomparison Project 3 (PMIP3), led to model-model and model-data discrepancies in the ocean circulation and Antarctic sea ice (Marzocchi & Jansen, 2017). The Eocene-PMIP design protocol addresses this to some extent by suggesting minimum values for TOA flux and trends in SAT (Lunt et al., 2017), but the impact of these thresholds have not been tested. As new PMIP projects are currently in development, for example, for the Eocene and Miocene, it is timely to revisit the utility of these thresholds and suggest refinements if needed.

Here we aim to make progress by comparing and evaluating five common criteria used to test for model quasi-equilibrium in a suite of five climate simulations with increasing levels of CO_2 in the Community Earth System Model (CESM version 1.0.4; Shields et al., 2012). This suite of simulations was chosen for the quasi-equilibrium criteria evaluation because it spans a large range of climate states and exhibit different types of time-dependent behavior, including equilibration, oscillatory behavior, large abrupt changes in deep water formation, and gradual warming trends. We neglect here biogeochemistry, ice-sheet, and vegetation dynamics and focus on coupled atmosphere-ocean-sea-ice models. The ocean is the slowest component to adjust, so the focus is on the deep ocean circulation, and its feedback on faster system components, such as SAT, sea ice, and the position of the Inter-Tropical Convergence Zone (ITCZ).

2. Model and Simulations

The simulations were performed using CESM version 1.0.4, developed by the National Center for Atmospheric Research. The atmospheric and land surface components are the Community Atmospheric Model 4 (CAM4) and Community Land Model version 4 (CLM4), with a spectral truncation of T31. The ocean model component is the Parallel Ocean Program version 2 (POP2; Danabasoglu et al., 2012; Smith, 2010), which has a resolution that varies from 3° near the poles to 1° at the equator. The sea-ice component is based on the Community Ice Code version 4 (CICE4; Holland et al., 2012; Hunke et al., 2008). We present four simulations that have atmospheric CO_2 concentrations abruptly set at 2, 4, 8, and 16 times the base Pre-Industrial (PI) value, named $2xCO_2$, $4xCO_2$, $8xCO_2$ and $16xCO_2$, respectively and compare them to a PI simulation named PI-cntrl. These simulations were first presented in Fedorov et al. (2015) and Burls and Fedorov (2017) as 800 years-long simulations, with the exception of the abrupt $4xCO_2$ simulation which was run for 3000 years. All simulations were initiated from PI initial conditions. For this study the $2xCO_2$, $4xCO_2$, $8xCO_2$ and $16xCO_2$ simulations are 5000 years long and the PI-cntrl simulation is 3800 years long.

The ocean circulation state in the simulations is evaluated through maxima and minima of the MOC (see e.g. Naik et al., 2025 for a standard definition of the MOC, Figure 1). Specifically, the maximum in Atlantic MOC (i.e., AMOC) is calculated between 45°N and 68°N, the maximum in Pacific MOC (i.e., PMOC) between 25°N and 58°N, and the Southern Ocean MOC (i.e., SOMOC) is the absolute value of the minimum of the global streamfunction south of 60°S. The maximum global MOC (GMOC) is evaluated north of 40°S. All streamfunction maxima and minima are computed below 500 m to avoid the wind-driven surface circulation.

The simulations experience large and abrupt changes in deep ocean circulation, or instability events, which we define here as an instant where the overturning streamfunction in the Atlantic, Pacific, or Southern Ocean undergoes a change of >50% when comparing the 300 years prior to the 300 years after the event. The events are indicated by the vertical lines in the figures and labeled according to the experiment number # (where # indicates

10.1029/2025GL117040

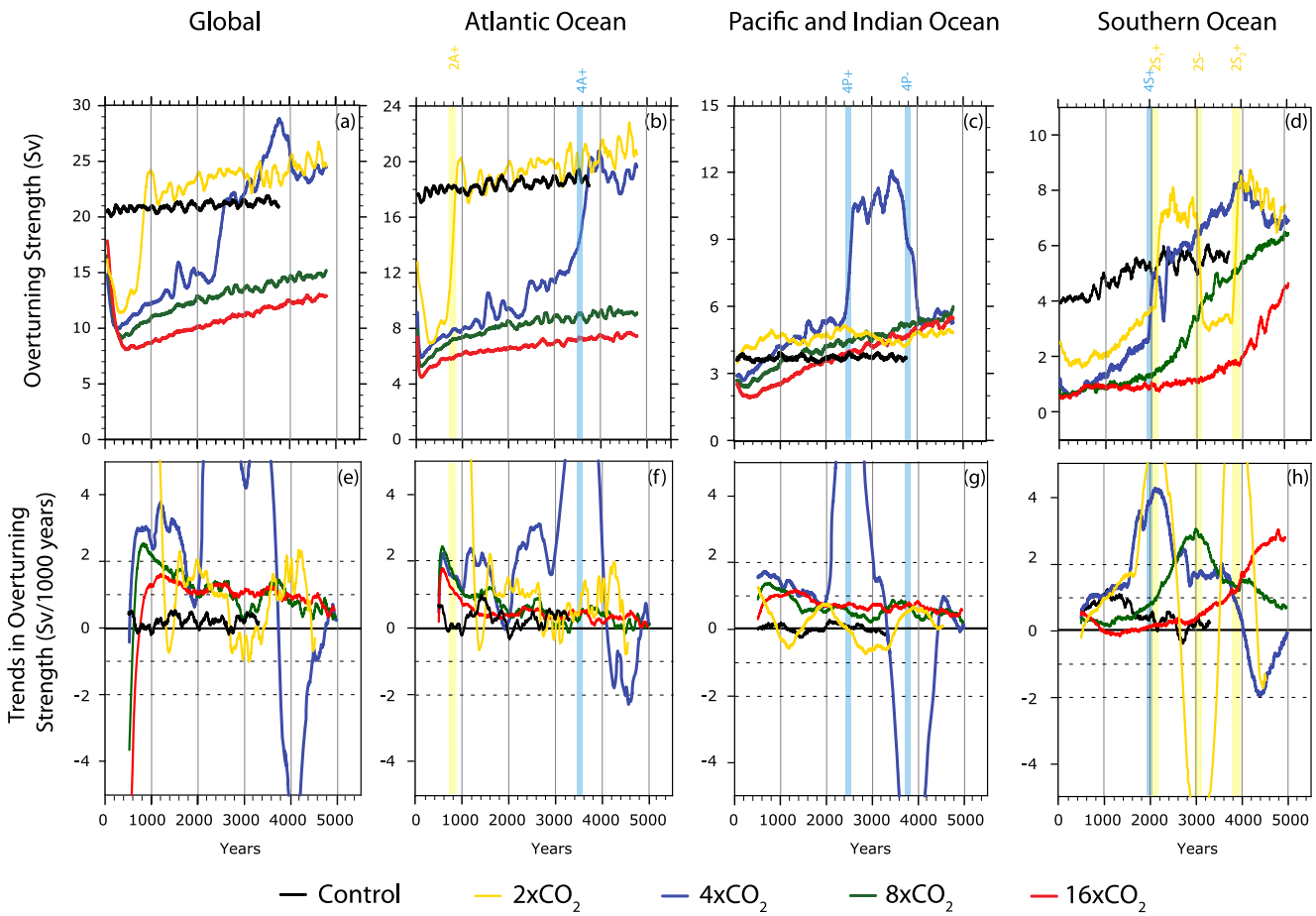

Figure 1. Top row: Time series of overturning circulation streamfunction maxima and minima for (a) Global, (b) Atlantic Ocean, (c) Indo-Pacific Ocean, (d) and Southern Ocean for the PI-cntrl (black), 2xCO_2 (yellow), 4xCO_2 (blue), 8xCO_2 (green), and 16xCO_2 (red) simulations, applying 100-year average smoothing (1 Sv = $10^6/m^3$). Bottom row: Trends over the last 1000 years, calculated using a least squares linear regression, in the (e) Global, (f) Atlantic, (g) Into-Pacific, and (h) Southern Ocean streamfunctions. Vertical colored bars indicate a rapid change in overturning and are labeled according to the experiment (i.e., CO_2 multiplication factor), basin of change, and sign of change, as described in the text.

the multiple of PI CO_2), the basin of change (A for Atlantic, P for Pacific, and S for Southern Ocean) and the sign of the MOC change (+ for increase and − for decrease). Using this definition, we identify eight events, namely 2A+, 4A+, 4P+, 4P−, 2S_1+, 2S−, 2S_2+, and 4S+. There are two 2S+ events, with subscripts 1 and 2 to distinguish them. The three events in the Southern Ocean in the 2xCO_2 simulation appear to be part of a ∼2000-year long oscillation, but the run is too short to confirm that. The instability events in the Pacific in the 4xCO_2 simulations are the focus of a separate study of abrupt PMOC change (Curtis & Fedorov, 2024b). The 8xCO_2 and 16xCO_2 simulations do not feature any instability events according to the definition here, but their SOMOC streamfunction maxima are still increasing steadily at the end of the simulation.

3. Evaluation of Selected Criteria in the Five Simulations

In this study, five commonly-applied criteria for model quasi-equilibrium are evaluated, namely, the TOA radiative balance and trends in global mean SAT, SST, DOT, and ocean ideal age tracer. The first four are measures of thermal equilibrium of various parts of the ocean-atmosphere system, where thermal equilibrium here refers to a zero net transfer of heat at the system boundary, or a constant mean temperature in its interior. The ocean ideal age tracer is a measure of how much time has passed since a parcel has been at the surface. Thus the trend in the age tracer is a measure of how equilibrated the ocean circulation is (Baatsen et al., 2018; Goldner et al., 2014).

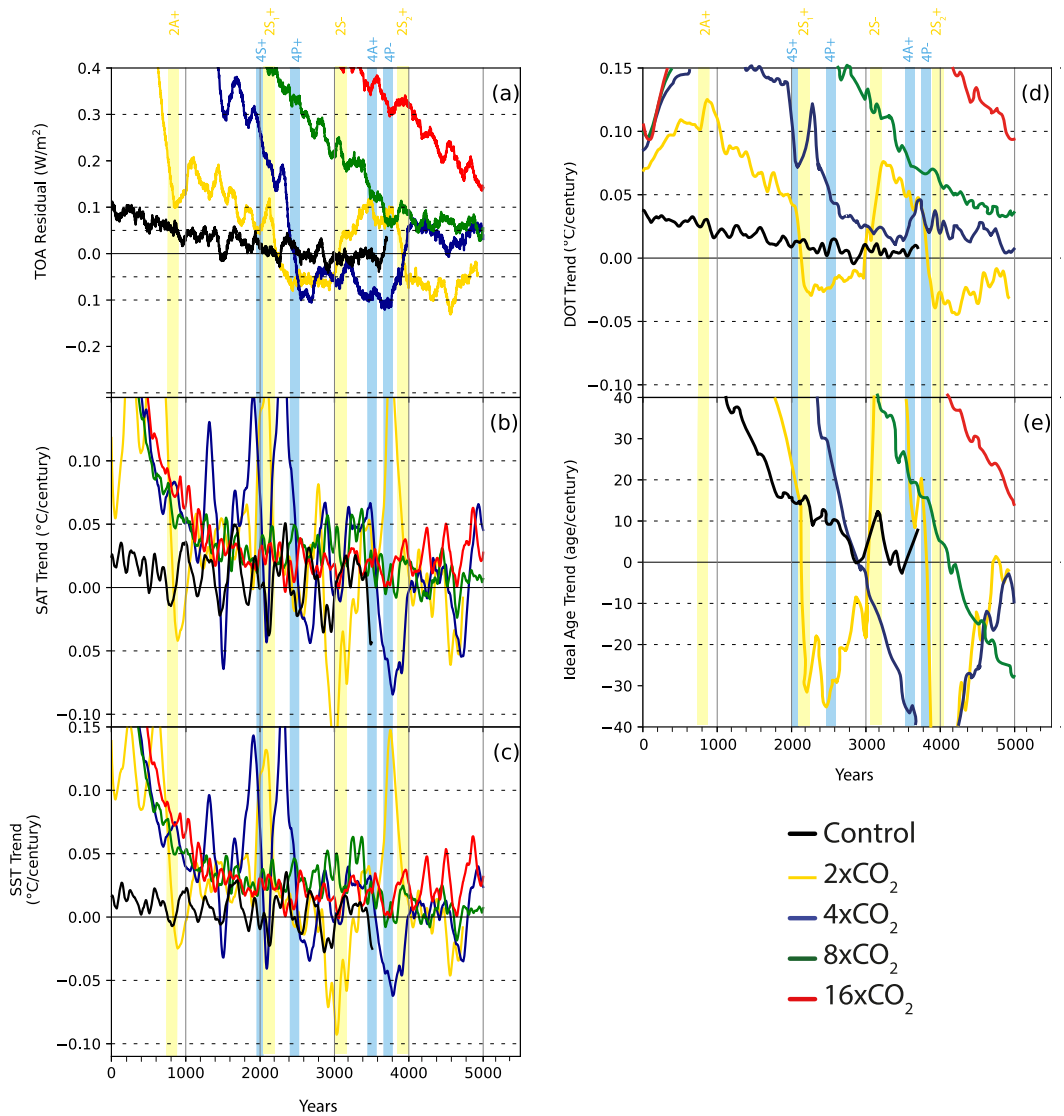

Figure 2. Five commonly applied equilibrium criteria, namely (a) Top of Atmosphere (TOA) radiative balance, (b) trend in global mean surface air temperature (SAT), (c) trend in sea surface temperature (SST), (d) trend in deep ocean temperature below 2,000 m (DOT), and (e) trend in ideal age tracer below 2,000 m. All trends are calculated using a least squares linear regression from the 100 years of annual data surrounding the time point, and then smoothed with a 100-year running mean. The horizontal dashes lines indicate previously used or newly proposed threshold values (see text).

A zero net TOA (or top-of-model in practice) radiative flux indicates that the Earth has reached a radiative balance and is in thermal equilibrium. A residual value of between 0.2 and 0.7 W/m^2 is commonly considered sufficient for quasi-equilibrium in coupled ocean-atmosphere studies (Dunne et al., 2012; Haywood et al., 2005; Hourdin et al., 2013; Lunt et al., 2017; Rugenstein et al., 2019). For example, the LongRunMIP project requires a simulation length of at least 1000 years in their design protocol, which is the time on average when the residual TOA flux reaches 0.5 W/m^2 in abrupt quadrupling-of-CO_2 CMIP experiments (Rugenstein et al., 2019). Similarly, in the design protocol for DeepMIP-Eocene, a minimum residual TOA flux balance of 0.3 W/m^2 is suggested (Lunt et al., 2017). Our simulations indicate that these orders of TOA flux thresholds do not guarantee an equilibrium. The 2xCO_2 experiment, the TOA flux reaches as low as 0.1 W/m^2 at ~1700 years, yet there are three instability events in the Southern Ocean after that point (Figures 1 and 2a). These events may be part of a stable ~2000-year-long coupled ocean-atmosphere oscillation, though the simulation is too short to confirm that and thus verify whether it has equilibrated. However, the criterion is certainly not sufficient in the 4xCO_2 simulation,

where the TOA flux reaches 0.1 W/m^2 after ∼2400 years, yet the simulation has a sharp increase in its PMOC at ∼2500 years (4P+) and an increase in its AMOC (4A+) and decrease in its PMOC between 3500 and 4000 years (4P−), respectively (Figures 1 and 2a). The 8xCO_2 TOA flux reaches 0.1 W/m^2 at ∼3600 years but at this point the SOMOC is still increasing by more than 1 Sv/1000 years (Figures 1 and 2a). The 16xCO_2 TOA flux never reaches the threshold of 0.1 W/m^2. The PI-cntrl TOA flux starts at ∼0.1 W/m^2 and drops to ∼0 W/m^2 around 2000 years, after which it appears to be in full equilibrium, with trends in MOCs and thermal criteria visually indistinguishable from zero. In summary, the TOA flux threshold of 0.1 W/m^2 is not sufficient and we discuss alternatives in the concluding section.

For climate studies, such as those focusing on equilibrium climate sensitivity, the trend in global mean SAT is often taken as an indicator of the degree of equilibrium. In some cases, no quantitative values are mentioned and the authors just state that the SAT has stabilized (e.g., Armstrong et al., 2016). In other studies, the trends in SAT have been calculated and quasi-equilibrium is assumed when these are under a threshold value of, for example, 0.02°C/century (Cao et al., 2019), 0.05°C/century (Braconnot et al., 2007) or 0.1°C/century (Dunne et al., 2012). An analogous metric that is sometimes applied is the global SST, for which trends have been required to fall, for example, under 0.02°C/century (Shi & Lohmann, 2016) and 0.1°C/century (Lunt et al., 2017). The trends in SAT and SST are qualitatively and quantitatively similar (Figures 2b and 2c) and are discussed here together. In all simulations SAT and SST trends drop below the threshold of 0.05°C/century by 1200 years, and as discussed, they are not in quasi-equilibrium at that point yet. It is problematic to just lower the threshold to, say 0.02°C/century, because internal variability pushes the trend beyond this boundary even for the PI-cntrl simulation after 2000 years. More aggressive smoothing or longer term trends may help remove these short term effects. However, the 16xCO_2 simulation has small long term mean trends of order 0.02°C/century in SAT and SST between 2000 and 4000 years (Figures 2b and 2c), at which point it is clearly not in thermal equilibrium when its TOA flux is considered (Figure 2a), suggesting that this is not a reliable criterion for the equilibrium of the larger atmosphere-ocean system.

Deep ocean equilibrium is frequently diagnosed through trends in DOT, with applied thresholds for quasi-equilibrium typically of the order of 0.1°C/century (Hewitt et al., 2003; Hutchinson et al., 2018; Ladant et al., 2018) or 0.06°C/century (Farnsworth et al., 2019), though lower values have also been reported, for example, 0.007°C/century (Renssen et al., 2007). What exactly constitutes the deep ocean varies as well between studies; it is sometimes evaluated at a specific depth such as 2,200 m (Fischer & Jungclaus, 2010), 2,700 m (Farnsworth et al., 2019) or 4,000 m (Hutchinson et al., 2018), or over a range of depths such as below 2,000 m (Baatsen et al., 2018; Sherriff-Tadano et al., 2018) or over the global ocean (Renssen et al., 2007). Here we calculate trends of the mean ocean temperature below 2,000 m, which include deep and abyssal waters of the ocean (For comparison, trends calculated at individual depths of 2,000 m and 3,600 m are shown in Figure S1 in Supporting Information S1.) The threshold of 0.1°C/century is not sufficient to guarantee equilibrium, especially in the 4xCO_2 simulation that has three large instability events after this threshold is reached (Figure 2d). Problematically for this threshold, the latter two of these events, 4A+ and 4P−, occur after the DOT had dropped as low as ∼0.01°C/century for the preceding 1000 years.

The final diagnostic is the trend in the ocean ideal age tracer, which is a measure of changes in deep ocean ventilation. While the MOC streamfunction represents the zonally integrated meridional flow, the age tracer indicates how long a water parcel has been away from the surface. It starts at zero at the beginning of a simulation and initially grows due to the interior "source term" of aging. In equilibrium, aging is balanced by young surface waters mixed down to the deep ocean through convection. Especially in the Southern Ocean, the age tracer can capture circulation changes, such as vertical mixing, that are not captured by the SOMOC (De Boer et al., 2008). Zero or weak trends in the age tracer have been used to argue for model equilibrium, but without specific thresholds (Baatsen et al., 2018; Goldner et al., 2014). As expected, here the age tracer trends are initially positive, and drops quickly when deepwater formation kicks in (Figure 2e). This is particularly clear in basin specific trends in the age tracer where large negative trends co-inside with the inception of deep water formation (e.g., negative spikes at 2A+ and 4A+ in Figure S2n of Supporting Information S1). In our otherwise equilibrated PI-cntrl, the remaining trend is less than 15 aging years/century for the last 1500 years of the simulation. The 2xCO_2 simulation's trend oscillates around zero with an amplitude of ∼60 aging years/century, while the other simulations only reach trends below 15 aging years/century for a short period. Only the PI-cntrl simulation has such low trends in all basins for more than 1000 years (Figure S2 in Supporting Information S1), suggesting it holds promise as a criterion for quasi-equilibrium.

10.1029/2025GL117040

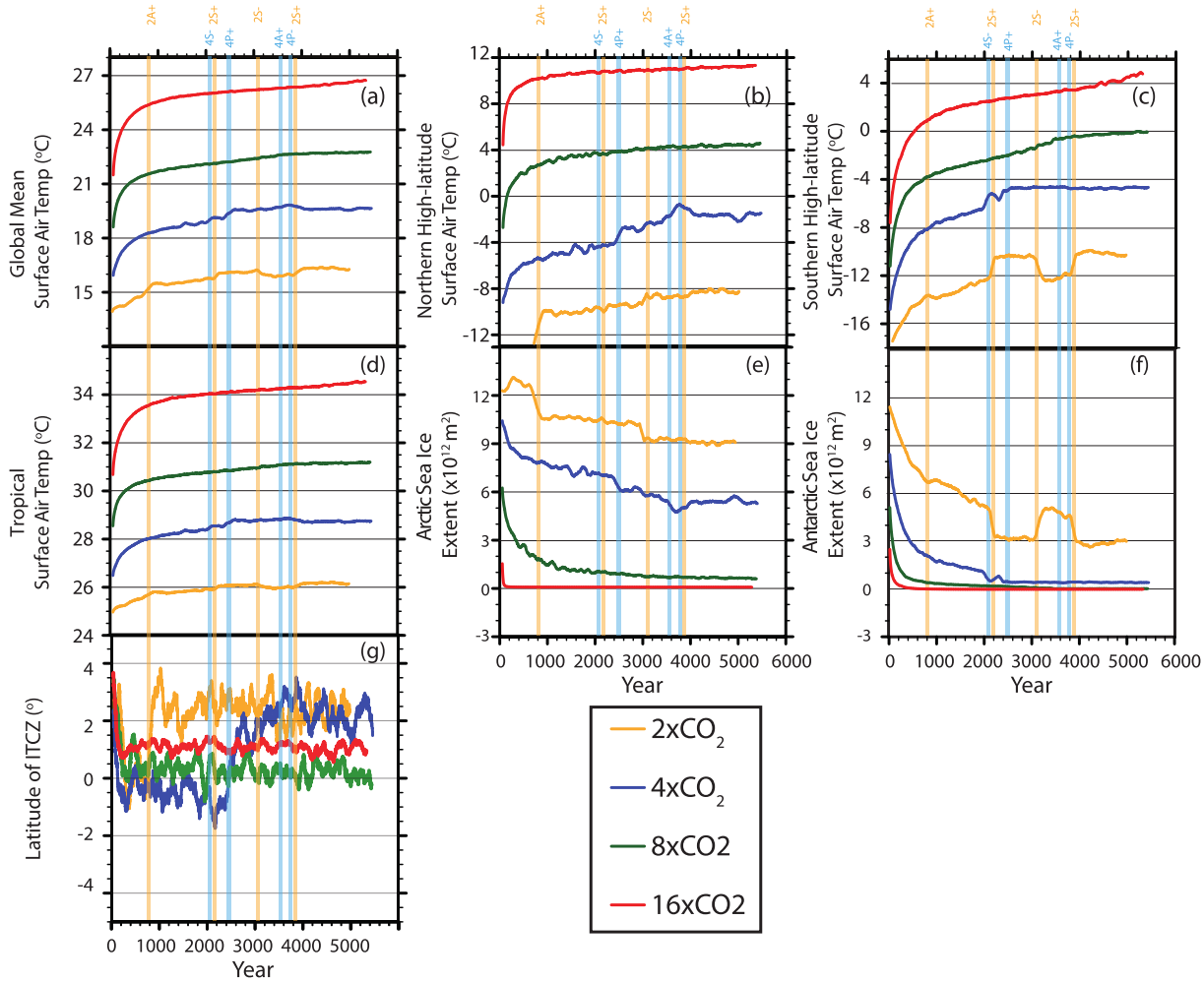

Figure 3. Impact of instability events on key variables. Time series of (a) SAT, (b) Northern Hemisphere polar SAT (60°N–90°N), (c) Southern Hemisphere polar SAT (60°S–90°S), (d) tropical SAT (30°S–30°N), (e) Arctic sea ice extent, (f) Antarctic sea ice extent, and (g) position of ITCZ. All variables are smoothed by a 30-year running mean filter, except the ITCZ, which is smoothed by a 100-year running mean filter. Vertical bars indicate instability events as explained in the text.

4. Impact of Deep Ocean Circulation Changes

Here, we illustrate the impact of changes in deep ocean circulation on climate variables, to inform for which applications it is relevant to continue a simulation until the deep ocean is in a steady state. The overturning has a noticeable but weak impact on the global SAT (Figure 3a). An increase in the PMOC of ~5 Sv is associated with a ~0.5°C increase in SAT in the 4xCO_2 simulation (Figure 3a, 4P+ bar). The impact of the AMOC increase on SAT is less obvious in this simulation, because it overlaps with a decrease in the PMOC (Figure 3a, 4A+, 4P− bars). In the 2xCO_2 simulation, an increase in the SOMOC of ~4 Sv leads to a ~0.5°C increase in SAT (Figures 1 and 3a, 4S+, 2S+ bars) with a decrease in the SOMOC leading to a similar decrease in SAT (Figure 3a, 2S− bar). Not surprisingly, the impact of overturning on SAT is more pronounced at the high latitudes (poleward of 60°) in the hemisphere where the convection changes. The northern high latitudes warm at the 2A+, 4P+, and 4A+ events (Figure 3b). But it is the GMOC that corresponds more closely to high northern latitude SAT and Arctic sea ice extent. In the 4xCO_2 simulation, the Arctic sea ice extent reaches a minimum at 3800 years when the GMOC peaks (Figure 1a). It is a time when the AMOC increases and the PMOC decreases and neither peaks individually. Instability events based on the SOMOC are co-incident with Southern Hemisphere polar temperature jumps, with a ~4 Sv SOMOC change corresponding to ~2°C temperature change of the same sign (Figures 1 and 3c). Along with this SAT change, the Antarctic sea ice extent reduces from 5 to 3 × 10^{12} m^2 (Figure 3f).

The mean tropical SAT (30°S–30°N) is not sensitive to overturning changes (Figure 3d) but the regional tropical climate is. Specifically, the GMOC impacts the latitude of the Intertropical Convergence Zone (ITCZ) which is here defined as the latitude of maximum zonally-averaged precipitation (Figure 3g). During the 4P+ event, the PMOC increases by ∼5 Sv with a concomitant northward shift in the ITCZ of ∼3° latitude. The ITCZ also shifts northward by ∼3° latitude at the 2A+ event at ∼800 years when the AMOC increases from ∼8 Sv to ∼19 Sv. For the $8xCO_2$ and the $16xCO_2$ simulations, the position of the ITCZ does not change during the simulation, consistent with no large change in northern overturning. The ocean circulation changes do not impact either amplitude or variability of the ENSO index (evaluated but not shown).

5. Discussion and Suggestions

Five common criteria for verifying that a coupled climate model is in quasi-equilibrium have been assessed in four simulations with an abrupt CO_2 increase from the PI, and the corresponding PI simulation. Problematically, the ocean circulation still undergoes significant changes after low thresholds in the thermal equilibrium criteria are reached, such as TOA flux of 0.1 W/m^2, SAT/SST trends of 0.05°C/century, and DOT of 0.05°C/century. One way forward is of course to reduce the thresholds tested here even further to, for example, TOA flux threshold of 0.05 W/m^2 or DOT of 0.02°C/century. However, there are several problems with this approach, especially if only one of these criteria is used: (a) The time series of the criteria has variability that pushes the criteria beyond the threshold temporarily. If the trends are calculated only at one snapshot over the last 100 years of the simulation, which is often the case, it may not represent the longer term trend of the criteria in the simulation. (b) If a simulation is statistically stable, but has internal variability with very low frequency, the trends may be more often outside the thresholds than within (e.g., in the $2xCO_2$ simulation). (c) In contrast, a simulation can have low values for TOA flux and DOT trends for as long as 500 years (e.g., the $4xCO_2$ run), suggesting that it is in quasi-equilibrium, before its circulation state suddenly switches. (d) Some models do not conserve energy, so that even a long simulation that shows no drift in any aspect of climate may still have a non-zero TOA energy imbalance (Delworth et al., 2006; Rugenstein et al., 2019). This so-called energy leakage is usually much larger than 0.05 W/m^2 and depends on the model and the background climate state (Lunt et al., 2021; Renoult et al., 2023). Thus, even if a threshold of 0.05 W/m^2 was found to work, it would not be generally applicable across models and applications.

The challenges with these individual thermal quasi-equilibrium criteria motivate us to find a fresh approach. For the sake of providing guidance to the community, especially in light of newly planned PMIP design protocols, we propose here a method that is based on the criteria presented here (Figures 1 and 2). We fully acknowledge that other methods may also work and that these criteria need to be further tested. The first suggestion is that modeling groups store and report 10-year averages of the TOA flux, the DOT (and age tracer if available) at ∼3,500 km, the maximum AMOC and PMOC (between 30°N and 60°N and below 500 m), and the minimum SOMOC (south of 60°S) for the full length of the simulation. The DOT and age tracer at ∼3,600 m are similar in magnitude and variability to the mean below 2,000 m (Figures 2d, 2e and Figure S1 in Supporting Information S1), but simpler to analyze. The time series of these variables (as opposed to only final 100-year data) is useful because it illustrates the frequency of variability in the system and whether, for example, the trends are accelerating or decelerating.

To verify model quasi-equilibrium, we suggest a two-step approach. The first step is to check for first-order thermal quasi-equilibrium and the suggested criteria is that the final 1000-year average of the TOA flux is <0.1 W/m^2 (if the model conserves energy) and/or the final 1000-year DOT trend is less than 0.1°C/century. These are not overly restrictive as the purpose is only to check that the majority of the thermal forcing is equilibrated. The long period over which to calculate the trend will remove the impact of most internal variability. The second step is to calculate 1000-year trends of the MOC streamfunctions (as in Figures 1e–1h). Here, we suggest that the simulations are continued until the MOC trends are less than 1 Sv/1000 years for 1000 years (or oscillating around 0 Sv/1000 years in case of low frequency variability such as in the $2xCO_2$ simulation). Applying these criteria to our simulations, we find that the PI-cntrl satisfies this after 2500 years. The time series of MOCs suggest that the $2xCO_2$ is equilibrated but there is a ∼2000-year oscillation; running for at least another 2000-year cycle would verify whether this oscillation is robust. The $4xCO_2$ and $8xCO_2$ MOC streamfunctions appear to level off by the end of the 5000 years simulation, and final trends are below the threshold, but not yet for 1000 years. The $16xCO_2$ run has just reached the 0.1 W/m^2 threshold at 5000 years but its SOMOC is sharply trending upwards (Figures 1d and 1h). Ideally the models are run for 2000 years after the criteria is reached, but modelers may not always have that opportunity. An alternative threshold could be that the global and/or basin-

specific age tracer trends be less than 15 age-years/century for the final 1000 years, but this needs to be verified in longer simulations.

While the suggested criteria are stricter than what have been the suggested thresholds in past CMIP protocol papers, the models should reach them faster than the 5000-year simulation length here. The abrupt CO_2 simulations all start from much colder than equilibrium initial conditions, with surface warming applied to a relatively cold abyssal ocean, a strong stratification is established initially that delays the ventilation and thermal adjustment of the deep ocean, especially in the $8xCO_2$ and $16xCO_2$ simulations. Simulations initialized from a state that is closer to the final state, or warmer than the final state should adjust much faster (Hutchinson et al., 2018; Zhang et al., 2022). Also for many studies, for example those investigating the impact of parameterizations on cloud properties, a short integration time may suffice. However, some fast-adjusting system components, such as sea ice, are impacted by slower changes in deep ocean circulation. In our simulations, a switch on of Northern Hemispheric overturning (i.e., a GMOC, composed of either AMOC or PMOC) increases Northern high latitude SAT by 1–2°C and reduces Arctic sea ice extent by ∼15%. It also leads to a northward shift of the ITCZ by ∼3° latitude. Similarly, an intensification of the SOMOC is associated with Southern high latitude warming and a reduction in Antarctic sea ice extent (∼40% in the $2xCO_2$ simulation). Our study suggests a leading role for the Southern Ocean and sea-ice feedbacks in ocean equilibration. The relatively cool simulation of $2xCO_2$ has global millennial scale oscillations with rapid onset and termination that originate in the Southern Ocean. Rapid changes in SOMOC as well as strong correlations between the SOMOC and Southern Ocean SAT, occur only when sea ice is present, suggesting a pivotal role for sea ice feedbacks in driving these oscillations.

PMIP comparison papers with a focus on the deep ocean circulation, for example, Naik et al. (2025) on the Miocene ocean circulation and Zhang et al. (2022) on the Eocene ocean circulation, have reported the end-of-simulation TOA flux values or DOT trends in order to indicate model equilibrium. While these are the best estimates of the circulation at those times from models, it remains uncertain at this point how robust their conclusions are. With the newly suggested criteria here, and time series of MOC metrics, a new generation of CMIP simulations will become available with more robust results and greatly improve our ability to assess the ocean's overturning state in different time periods and climates. This next generation of MIPs would also be a good opportunity to formally compare and verify this new approach to quasi-equilibrium across models. Finally, despite the many improvements that increased model resolution brings (Hirschi et al., 2020), the long simulation length needed to equilibrate the deep ocean circulation makes the routine use of them for equilibrium studies unfeasible at this time.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

The model data presented in this study is available for download from the Bolin Centre database (Krishnan et al., 2025). The source code and boundary conditions for the CESM model used to generate our results can be accessed at http://www.cesm.ucar.edu/models/cesm1.0.

Acknowledgments

AdB acknowledges support from Swedish Research Council Grants 2022-06725 and 2020-04791. NB received support from NSF Grants AGS-1844380 and OCN-2402414 and DH received support from Australian Research Council Grant DE220100279. Analysis and data handling were enabled by resources from the National Academic Infrastructure for Supercomputing in Sweden (NAISS), partially funded by Swedish Research Council Grant 2022-06725. High-performance computing support from the Cheyenne system (https://doi.org/10.5065/D6RX99HX) was provided by National Center for Atmospheric Research (NCAR), and sponsored by the National Science Foundation.

References

Armstrong, E., Valdes, P., House, J., & Singarayer, J. (2016). The role of CO_2 and dynamic vegetation on the impact of temperate land-use change in the HadCM3 coupled climate model. *Earth Interactions*, *20*(10), 1–20. https://doi.org/10.1175/ei-d-15-0036.1

Baatsen, M., vonder Heydt, A. S., Huber, M., Kliphuis, M. A., Bijl, P. K., Sluijs, A., & Dijkstra, H. A. (2018). Equilibrium state and sensitivity of the simulated middle-to-late Eocene climate. *Climate of the Past Discussions*, *2018*, 1–49. https://doi.org/10.5194/cp-2018-43

Bonan, D. B., Thompson, A. F., Newsom, E. R., Sun, S., & Rugenstein, M. (2022). Transient and equilibrium responses of the Atlantic overturning circulation to warming in coupled climate models: The role of temperature and salinity. *Journal of Climate*, *35*(15), 5173–5193. https://doi.org/10.1175/jcli-d-21-0912.1

Braconnot, P., Otto-Bliesner, B., Harrison, S., Joussaume, S., Peterchmitt, J. Y., Abe-Ouchi, A., et al. (2007). Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum – Part 1: Experiments and large-scale features. *Climate of the Past*, *3*(2), 261–277. https://doi.org/10.5194/cp-3-261-2007

Burls, N. J., & Fedorov, A. V. (2017). Wetter subtropics in a warmer world: Contrasting past and future hydrological cycles. *Proceedings of the National Academy of Sciences*, *114*(49), 12888–12893. https://doi.org/10.1073/pnas.1703421114

Cao, J., Wang, B., & Ma, L. (2019). Attribution of global monsoon response to the last glacial maximum forcings. *Journal of Climate*, *32*(19), 6589–6605. https://doi.org/10.1175/jcli-d-18-0871.1

Curtis, P. E., & Fedorov, A. V. (2024a). Collapse and slow recovery of the Atlantic Meridional Overturning Circulation (AMOC) under abrupt greenhouse gas forcing. *Climate Dynamics*, *62*(7), 5949–5970. https://doi.org/10.1007/s00382-024-07185-3

Curtis, P. E., & Fedorov, A. V. (2024b). Spontaneous activation of the Pacific Meridional Overturning Circulation (PMOC) in long-term ocean response to greenhouse forcing. *Journal of Climate*, *37*(5), 1551–1565. https://doi.org/10.1175/jcli-d-23-0393.1

Danabasoglu, G., Bates, S. C., Briegleb, B. P., Jayne, S. R., Jochum, M., Large, W. G., et al. (2012). The CCSM4 ocean component. *Journal of Climate*, *25*(5), 1361–1389. https://doi.org/10.1175/jcli-d-11-00091.1

De Boer, A. M., Toggweiler, J. R., & Sigman, D. M. (2008). Atlantic dominance of the meridional overturning circulation. *Journal of Physical Oceanography*, *38*(2), 435–450. https://doi.org/10.1175/2007jpo3731.1

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 I: Formulation and simulation characteristics. *Journal of Climate*, *19*(5), 643–674. https://doi.org/10.1175/jcli3629.1

Dunne, J. P., John, J. G., Adcroft, A. J., Griffies, S. M., Hallberg, R. W., Shevliakova, E., et al. (2012). GFDL's ESM2 global coupled climate–carbon earth system models. Part I: Physical formulation and baseline simulation characteristics. *Journal of climate*, *25*(19), 6646–6665. https://doi.org/10.1175/jcli-d-11-00560.1

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

Farnsworth, A., Lunt, D. J., O'Brien, C. L., Foster, G. L., Inglis, G. N., Markwick, P., et al. (2019). Climate sensitivity on geological timescales controlled by nonlinear feedbacks and ocean circulation. *Geophysical Research Letters*, *46*(16), 9880–9889. https://doi.org/10.1029/2019gl083574

Fedorov, A. V., Burls, N. J., Lawrence, K. T., & Peterson, L. C. (2015). Tightly linked zonal and meridional sea surface temperature gradients over the past five million years. *Nature Geoscience*, *8*(12), 975–980. https://doi.org/10.1038/ngeo2577

Fischer, N., & Jungclaus, J. H. (2010). Effects of orbital forcing on atmosphere and ocean heat transports in Holocene and Eemian climate simulations with a comprehensive Earth system model. *Climate of the Past*, *6*(2), 155–168. https://doi.org/10.5194/cp-6-155-2010

Galbraith, E., & de Lavergne, C. (2019). Response of a comprehensive climate model to a broad range of external forcings: Relevance for deep ocean ventilation and the development of late Cenozoic ice ages. *Climate Dynamics*, *52*(1), 653–679. https://doi.org/10.1007/s00382-018-4157-8

Goldner, A., Herold, N., & Huber, M. (2014). Antarctic glaciation caused ocean circulation changes at the Eocene–Oligocene transition. *Nature*, *511*(2), 574–578. https://doi.org/10.1038/nature13597

Gregory, J. M., Ingram, W. J., Palmer, M. A., Jones, G. S., Stott, P. A., Thorpe, R. B., et al. (2004). A new method for diagnosing radiative forcing and climate sensitivity. *Geophysical Research Letters*, *31*(3). https://doi.org/10.1029/2003gl018747

Haywood, A. M., Dekens, P., Ravelo, A. C., & Williams, M. (2005). Warmer tropics during the mid-Pliocene? Evidence from alkenone paleothermometry and a fully coupled ocean-atmosphere GCM. *Geochemistry, Geophysics, Geosystems*, *6*(3). https://doi.org/10.1029/2004gc000799

Haywood, A. M., Dowsett, H. J., Dolan, A. M., Rowley, D., Abe-Ouchi, A., Otto-Bliesner, B., et al. (2016). The Pliocene Model Intercomparison Project (PlioMIP) Phase 2: Scientific objectives and experimental design. *Climate of the Past*, *12*(3), 663–675. https://doi.org/10.5194/cp-12-663-2016

Hewitt, C. D., & Mitchell, J. F. B. (1998). A fully coupled GCM simulation of the climate of the mid-Holocene. *Geophysical Research Letters*, *25*(3), 361–364. https://doi.org/10.1029/97gl03721

Hewitt, C. D., Stouffer, R. J., Broccoli, A. J., Mitchell, J. F. B., & Valdes, P. J. (2003). The effect of ocean dynamics in a coupled GCM simulation of the Last Glacial Maximum. *Climate Dynamics*, *20*(2–3), 203–218. https://doi.org/10.1007/s00382-002-0272-6

Hirschi, J. J. M., 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

Holland, M. M., Bailey, D. A., Briegleb, B. P., Light, B., & Hunke, E. (2012). Improved sea ice shortwave radiation physics in CCSM4: The impact of melt ponds and aerosols on Arctic sea ice. *Journal of Climate*, *25*(5), 1413–1430. https://doi.org/10.1175/jcli-d-11-00078.1

Hourdin, F., Foujols, M.-A., Codron, F., Guemas, V., Dufresne, J.-L., Bony, S., et al. (2013). Impact of the LMDZ atmospheric grid configuration on the climate and sensitivity of the IPSL-CM5A coupled model. *Climate Dynamics*, *40*(9), 2167–2192. https://doi.org/10.1007/s00382-012-1411-3

Hunke, E., Lipscomb, W., Turner, A., Jeffery, N., & Elliott, S. (2008). *The Los Alamos sea ice model documentation and software user's manual, Version 4.0*. Los Alamos National Laboratory.

Hutchinson, D. K., de Boer, A. M., Coxall, H. K., Caballero, R., Nilsson, J., & Baatsen, M. (2018). Climate sensitivity and meridional overturning circulation in the late Eocene using GFDL CM2.1. *Climate of the Past*, *14*(6), 789–810. https://doi.org/10.5194/cp-14-789-2018

Kageyama, M., Albani, S., Braconnot, P., Harrison, S. P., Hopcroft, P. O., Ivanovic, R. F., et al. (2017). The PMIP4 contribution to CMIP6–Part 4: Scientific objectives and experimental design of the PMIP4-CMIP6 Last Glacial Maximum experiments and PMIP4 sensitivity experiments. *Geoscientific Model Development*, *10*(11), 4035–4055. https://doi.org/10.5194/gmd-10-4035-2017

Klinger, B. A., Drijfhout, S., Marotzke, J., & Scott, J. R. (2003). Sensitivity of basinwide meridional overturning to diapycnal diffusion and remote wind forcing in an idealized Atlantic-Southern Ocean geometry. *Journal of Physical Oceanography*, *33*(1), 249–266. https://doi.org/10.1175/1520-0485(2003)033<0249:sobmot>2.0.co;2

Klockmann, M., Mikolajewicz, U., & Marotzke, J. (2018). Two AMOC States in response to decreasing greenhouse gas concentrations in the coupled climate model MPI-ESM. *Journal of Climate*, *31*(19), 7969–7984. https://doi.org/10.1175/jcli-d-17-0859.1

Krishnan, S., de Boer, A. M., & Burls, N. (2025). Data from the CESM v1.0.4 model for evaluation of quasi-equilibrium criteria in coupled climate model simulations [Dataset]. *Bolin Centre Database*. https://doi.org/10.17043/krishnan-2025-cesm-1

Ladant, J.-B., Donnadieu, Y., Bopp, L., Lear, C. H., & Wilson, P. A. (2018). Meridional contrasts in productivity changes driven by the opening of Drake Passage. *Paleoceanography and Paleoclimatology*, *33*(3), 302–317. https://doi.org/10.1002/2017pa003211

Liu, Z., Otto-Bliesner, B., Kutzbach, J., Li, L., & Shields, C. (2003). Coupled climate simulation of the evolution of global monsoons in the Holocene. *Journal of Climate*, *16*(15), 2472–2490. https://doi.org/10.1175/1520-0442(2003)016<2472:ccsote>2.0.co;2

Liu, Z., Shin, S.-I., Otto-Bliesner, B., Kutzbach, J. E., Brady, E. C., & Lee, D. (2002). Tropical cooling at the last glacial maximum and extratropical ocean ventilation[1]. *Geophysical Research Letters*, *29*(10), 48-41–44. https://doi.org/10.1029/2001gl013938

Lunt, D. J., Bragg, F., Chan, W. L., Hutchinson, D. K., Ladant, J. B., Morozova, P., et al. (2021). DeepMIP: Model intercomparison of early Eocene climatic optimum (EECO) large-scale climate features and comparison with proxy data. *Climate of the Past*, *17*(1), 203–227. https://doi.org/10.5194/cp-17-203-2021

Lunt, D. J., Huber, M., Anagnostou, E., Baatsen, M. L. J., Caballero, R., DeConto, R., et al. (2017). The DeepMIP contribution to PMIP4: Experimental design for model simulations of the EECO, PETM, and pre-PETM (version 1.0). *Geoscientific Model Development*, *10*(2), 889–901. https://doi.org/10.5194/gmd-10-889-2017

Marzocchi, A., & Jansen, M. F. (2017). Connecting Antarctic sea ice to deep-ocean circulation in modern and glacial climate simulations. *Geophysical Research Letters*, *44*(12), 6286–6295. https://doi.org/10.1002/2017gl073936

Naik, T. J., de Boer, A. M., Coxall, H. K., Burls, N. J., Bradshaw, C. D., Donnadieu, Y., et al. (2025). Ocean meridional overturning circulation during the early and middle Miocene. *Paleoceanography and Paleoclimatology*, *40*(4), e2024PA005055. https://doi.org/10.1029/2024pa005055

Otto-Bliesner, B. L., Braconnot, P., Harrison, S. P., Lunt, D. J., Abe-Ouchi, A., Albani, S., et al. (2017). The PMIP4 contribution to CMIP6 – Part 2: Two interglacials, scientific objective and experimental design for Holocene and Last Interglacial simulations. *Geoscientific Model Development*, *10*(11), 3979–4003. https://doi.org/10.5194/gmd-10-3979-2017

Renoult, M., Sagoo, N., Zhu, J., & Mauritsen, T. (2023). Causes of the weak emergent constraint on climate sensitivity at the Last Glacial Maximum. *Climate of the Past*, *19*(2), 323–356. https://doi.org/10.5194/cp-19-323-2023

Renssen, H., Goosse, H., & Fichefet, T. (2007). Simulation of Holocene cooling events in a coupled climate model. *Quaternary Science Reviews*, *26*(15), 2019–2029. https://doi.org/10.1016/j.quascirev.2007.07.011

Rugenstein, M., Bloch-Johnson, J., Abe-Ouchi, A., Andrews, T., Beyerle, U., Cao, L., et al. (2019). LongRunMIP: Motivation and design for a large collection of millennial-length AOGCM simulations. *Bulletin of the American Meteorological Society*, *100*(12), 2551–2570. https://doi.org/10.1175/bams-d-19-0068.1

Rugenstein, M. A. A., Sedláček, J., & Knutti, R. (2016). Nonlinearities in patterns of long-term ocean warming. *Geophysical Research Letters*, *43*(7), 3380–3388. https://doi.org/10.1002/2016gl068041

Sherriff-Tadano, S., Abe-Ouchi, A., Yoshimori, M., Oka, A., & Chan, W.-L. (2018). Influence of glacial ice sheets on the Atlantic meridional overturning circulation through surface wind change. *Climate Dynamics*, *50*(7), 2881–2903. https://doi.org/10.1007/s00382-017-3780-0

Shi, X., & Lohmann, G. (2016). Simulated response of the mid-Holocene Atlantic meridional overturning circulation in ECHAM6-FESOM/MPIOM. *Journal of Geophysical Research: Oceans*, *121*(8), 6444–6469. https://doi.org/10.1002/2015jc011584

Shields, C., Bailey, D., Danabasoglu, G., Jochum, M., Neale, R., Levis, S., & Park, S. (2012). Low-resolution CCSM4. *Journal of Climate*, *25*(12), 3993–4014. https://doi.org/10.1175/jcli-d-11-00260.1

Smith, R. D. (2010). The Parallel Ocean Program (POP) reference manual: Ocean component of the community climate system model (CCSM). *Rep. LAUR-01853*, *141*, 1–141.

Zhang, Y., de Boer, A. M., Lunt, D. J., Hutchinson, D. K., Ross, P., vande Flierdt, T., et al. (2022). Early Eocene Ocean meridional overturning circulation: The roles of atmospheric forcing and Strait geometry. *Paleoceanography and Paleoclimatology*, *37*(3), e2021PA004329. https://doi.org/10.1029/2021pa004329

Zhu, J., Liu, Z., Zhang, J., & Liu, W. (2015). AMOC response to global warming: Dependence on the background climate and response timescale. *Climate Dynamics*, *44*(11), 3449–3468. https://doi.org/10.1007/s00382-014-2165-x