

# Geophysical Research Letters®



## RESEARCH LETTER

10.1029/2023GL107320

### Key Points:

- Models with disproportionate Atlantic Meridional Overturning Circulation (AMOC) declines under abrupt CO<sub>2</sub> doubling experience less warming
- AMOC decline modulates transient warming by influencing *both* low cloud feedbacks and the global effective heat capacity
- Therefore, abrupt quadrupling of pre-industrial CO<sub>2</sub> simulations do not necessarily capture the behavior of lesser warming scenarios

### Supporting Information:

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

### Correspondence to:

A. N. Poletti,  
[apolet@uw.edu](mailto:apolet@uw.edu)

### Citation:

Poletti, A. N., Frierson, D. M. W., & Armour, K. C. (2024). Why do CO<sub>2</sub> quadrupling simulations warm more than twice as much as CO<sub>2</sub> doubling simulations in CMIP6? *Geophysical Research Letters*, 51, e2023GL107320. <https://doi.org/10.1029/2023GL107320>

Received 14 NOV 2023

Accepted 11 FEB 2024

## Why Do CO<sub>2</sub> Quadrupling Simulations Warm More Than Twice as Much as CO<sub>2</sub> Doubling Simulations in CMIP6?

A. N. Poletti<sup>1</sup> , D. M. W. Frierson<sup>1</sup>, and K. C. Armour<sup>1,2</sup> 

<sup>1</sup>Department of Atmospheric Sciences, University of Washington, Seattle, WA, USA, <sup>2</sup>School of Oceanography, University of Washington, Seattle, WA, USA

**Abstract** We compare abrupt CO<sub>2</sub>-quadrupling (abrupt-4xCO<sub>2</sub>) and -doubling (abrupt-2xCO<sub>2</sub>) simulations across 10 CMIP6 models. Two models (CESM2 and MRI-ESM2-0) warm substantially more than twice as much under abrupt-4xCO<sub>2</sub> than abrupt-2xCO<sub>2</sub>, which cannot be explained by the non-logarithmic scaling of CO<sub>2</sub> forcing. Using an energy balance model, we show that increased warming rates within these two models are driven by both less-negative radiative feedbacks and smaller global effective heat capacity under abrupt-4xCO<sub>2</sub>. These differences are caused by a decrease in low cloud cover *and* shallower ocean heat storage, respectively; both are linked to smaller fractional declines in the Atlantic Meridional Overturning Circulation (AMOC) under abrupt-4xCO<sub>2</sub> (relative to abrupt-2xCO<sub>2</sub>). On a global scale, higher climate sensitivity under larger forcing can be explained by a feedback-temperature dependence; however, we find that forcing-dependent spatial warming patterns due to AMOC decline are an important physical mechanism which reduces warming in a way that is not captured by a global-mean framework.

**Plain Language Summary** Climate model simulations with abruptly quadrupled CO<sub>2</sub> concentrations are often used to characterize the climate response of global climate models to greenhouse gas forcing and inform predictions of the future. In this study, we ask how relevant CO<sub>2</sub>-quadrupling experiments are to more modest warming scenarios by comparing them to CO<sub>2</sub>-doubling experiments. It is usually assumed that CO<sub>2</sub>-quadrupling will produce about twice as much warming than CO<sub>2</sub>-doubling, but this is not the case in some models. Two global climate models in particular warm much more under CO<sub>2</sub>-quadrupling than twice the warming under CO<sub>2</sub>-doubling. By fitting each climate model to a simpler energy balance model, we find that these two climate models are strongly impacted by the effective heat capacity (which can be impacted by ocean circulation changes) in comparison to other climate models where radiative effects (such as effects from clouds) are more impactful. Changes in the strength of an Atlantic Ocean circulation are important for both radiative effects and heat storage. Therefore, the behavior of our planet at 1.5 or 2 times CO<sub>2</sub> cannot necessarily be estimated with quadrupling CO<sub>2</sub> experiments.

## 1. Introduction

Understanding Earth's climate response to greenhouse gas forcing is essential for climate adaptation and mitigation. A key metric of the global climate response which is correlated with transient warming in climate models (Grose et al., 2018; Sherwood et al., 2020) is the effective climate sensitivity (EffCS). EffCS is often calculated by extrapolating to equilibrium the regression between the global top-of-atmosphere (TOA) radiation anomaly and the global near-surface air temperature anomaly of abrupt-quadrupling CO<sub>2</sub> experiments, often referred to as a “Gregory regression” (Gregory et al., 2004). However, quadrupling CO<sub>2</sub> concentrations from pre-industrial greatly exceeds present radiative forcing change; thus it is an open question as to whether abrupt-4xCO<sub>2</sub> experiments properly represent changes at lower warming levels.

Warming within abrupt-4xCO<sub>2</sub> simulations has been found to be larger than that expected from abrupt-doubling CO<sub>2</sub> experiments, even when dividing the abrupt-4xCO<sub>2</sub> response by a factor of two to account for its forcing being approximately twice that of abrupt-2xCO<sub>2</sub> (Bloch-Johnson et al., 2021; Good et al., 2015; Gregory et al., 2015; Meraner et al., 2013; Mitevski et al., 2021). This fractionally larger warming at higher forcing levels can arise from higher EffCS due to more-positive cloud (Bloch-Johnson et al., 2021) or water-vapor feedbacks (Meraner et al., 2013), and/or the effective radiative forcing (ERF) scaling non-logarithmically with CO<sub>2</sub> concentration (Byrne & Goldblatt, 2013; Etminan et al., 2016; Mitevski et al., 2022). Ocean dynamics also impact

© 2024. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

surface warming (Kostov et al., 2014; Winton et al., 2013), and these circulations can change non-linearly with forcing (Drijfhout, 2015; Rugenstein et al., 2016).

In this paper, we will compare how and why the transient response to abrupt changes of CO<sub>2</sub> varies across forcing levels using an ensemble of 10 general circulation models (GCMs) from the Coupled Model Intercomparison Project Phase 6 (CMIP6; Eyring et al., 2016) for which experiments of both abrupt-doubling (abrupt-2xCO<sub>2</sub>) and abrupt-quadrupling (abrupt-4xCO<sub>2</sub>) from pre-industrial CO<sub>2</sub> levels are available. We focus on two models in particular, CESM2 and MRI-ESM2-0, which show the largest differences in climate response at higher forcing.

## 2. Materials and Methods

### 2.1. Abrupt-2xCO<sub>2</sub> and Abrupt-4xCO<sub>2</sub> Simulations

This study uses annually averaged CMIP6 data (Eyring et al., 2016; Tables S1–S3 in Supporting Information S1). The ensemble mean refers to a multimodel mean excluding CESM2 and MRI-ESM2-0, such that the two models of interest can be compared to the rest of the ensemble. Occasionally, we use the entire 10 model ensemble mean in statistical significance testing. When performing an ensemble mean across zonal or global structure, each model is interpolated to a common latitude-longitude grid (taken from CanESM5). We calculate anomalies by subtracting the mean of the final 100 years of the pre-industrial control (piControl) from the abrupt-4xCO<sub>2</sub> or abrupt-2xCO<sub>2</sub> data, since we did not find significant drift in the last 100 years of the piControl surface air temperature. Since our analysis constitutes comparisons of a model with itself at a different forcings, time-dependent drift will not impact results as long as the scenarios have the same branch point (see Table S1 in Supporting Information S1).

When comparing warming at different CO<sub>2</sub> levels, we must choose a factor ( $\mu = F_{2x}/F_{4x}$ , where F is the ERF of either abrupt-2xCO<sub>2</sub> or abrupt-4xCO<sub>2</sub>) to “scale” them by to account for their different forcing magnitudes. If CO<sub>2</sub> radiative forcing scaled logarithmically with CO<sub>2</sub> concentration (Myhre et al., 1998; Ramaswamy et al., 2018), this would mean that abrupt-4xCO<sub>2</sub> forcing would be exactly twice the value of abrupt-2xCO<sub>2</sub> forcing (e.g.,  $\mu = 0.5$ ). However, recent research shows that CO<sub>2</sub>-forcing scales slightly non-logarithmically with CO<sub>2</sub> concentration (Byrne & Goldblatt, 2013; Etminan et al., 2016). On average across CMIP6 models, a factor of 0.476 is needed to equalize forcing between abrupt-4xCO<sub>2</sub> and abrupt-2xCO<sub>2</sub> (IPCC AR6, 2021; Meinshausen et al., 2020).

Using the mean ERF (estimated as the y-intercept of a Gregory regression over years 1–150) across our 10 models, we found the mean scaling factor is 0.488 (Table 1c), which is similar to the aforementioned IPCC AR6 value. Compared to ERF values calculated using Fixed-SST abrupt-4xCO<sub>2</sub> runs (Smith et al., 2020), our abrupt-4xCO<sub>2</sub> ERF values are  $-0.064 \pm 0.33$  W/m<sup>2</sup> different. However, Fixed-SST runs do not exist for abrupt-2xCO<sub>2</sub>, so we cannot determine the precision of our scaling factors.

We have performed this analysis using several scaling factors including  $\mu = 0.5$ , the IPCC scaling value ( $\mu = 0.476$ ) and the model-specific ratio for the years 1–150 (an ensemble mean value of  $\mu = 0.488$ ). Our conclusions are robust across these scaling factors. Hereafter, we will use the model-specific scaling factor calculated over the years 1–150 in Table 1c, and we will refer to results from abrupt-4xCO<sub>2</sub> experiments multiplied by the model-specific scaling factor as “scaled” simulations.

### 2.2. Transient Warming and Effective Equilibrium Climate Sensitivity

The time series of globally averaged annual surface air temperature anomaly shows that the transient warming in abrupt-2xCO<sub>2</sub> experiments is smaller compared to the scaled warming in abrupt-4xCO<sub>2</sub> experiments in CMIP6 models (Figure 1a). To quantify warming differences across forcings, we consider a percent difference defined as:

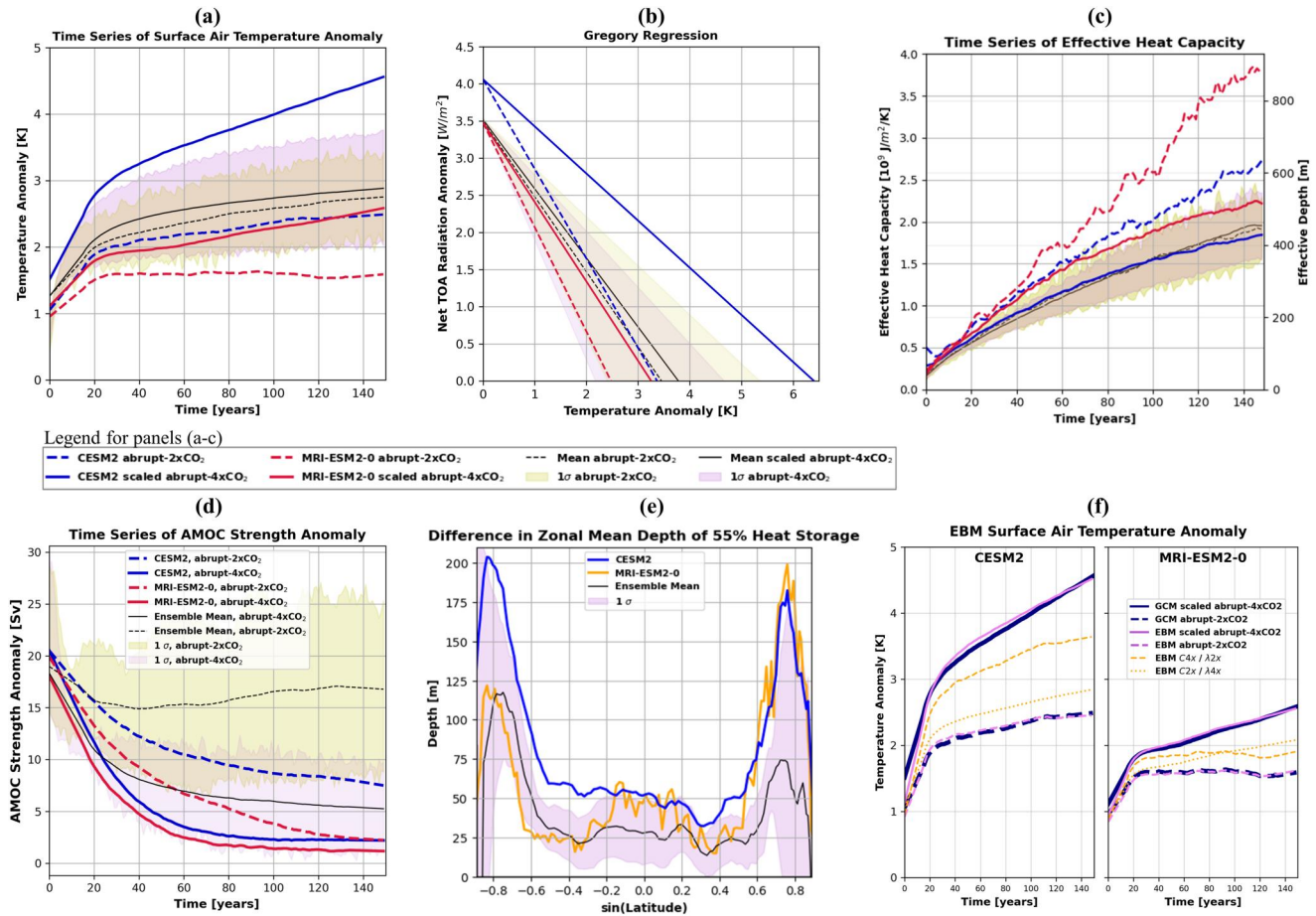
$$= \frac{\mu T_{4x} - T_{2x}}{\mu T_{4x}} \times 100\% \quad (1)$$

where  $T_{4x}$  and  $T_{2x}$  are the final 20-year average temperatures in the scaled abrupt-4xCO<sub>2</sub> and abrupt-2xCO<sub>2</sub> runs respectively. When we compare the final 20-year average of the temperature anomaly, CESM2 and MRI-ESM2-0 have 44.8% and 37.9% less warming in abrupt-2xCO<sub>2</sub> than scaled abrupt-4xCO<sub>2</sub>, while the ensemble average difference is 4.5% (Figure 1a).

**Table 1**  
*Comparison of Warming, ERF, and Feedbacks in Abrupt-2xCO<sub>2</sub> and Abrupt-4xCO<sub>2</sub> Simulations*

Model	(a) Transient response			(b) Estimated equilibrium response			(c) Effective radiative forcing			(d) Radiative feedback parameter		
	dT/dt <sub>2xCO<sub>2</sub></sub> [10 <sup>-3</sup> K/yr]	dT/dt <sub>4xCO<sub>2</sub></sub> [10 <sup>-3</sup> K/yr]	Difference [%]	EffCS <sub>2xCO<sub>2</sub></sub> [K]	EffCS <sub>4xCO<sub>2</sub></sub> [K]	Difference [%]	ERF <sub>2xCO<sub>2</sub></sub> [W/m <sup>2</sup> ]	ERF <sub>4xCO<sub>2</sub></sub> [W/m <sup>2</sup> ]	Ratio [unitless]	λ <sub>2xCO<sub>2</sub></sub> [W/m <sup>2</sup> /K]	λ <sub>4xCO<sub>2</sub></sub> [W/m <sup>2</sup> /K]	Difference [%]
CanESM5	6.51	7.84	17%	4.67	5.35	12.7%	3.53	7.46	0.473	-0.76	-0.66	-15.2%
<b>CESM2</b>	<b>4.17</b>	<b>12.30</b>	<b>66%</b>	<b>3.37</b>	<b>6.40</b>	<b>47.3%</b>	<b>4.06</b>	<b>6.65</b>	<b>0.611</b>	<b>-1.21</b>	<b>-0.63</b>	<b>-92.1%</b>
CNRM-CM6-1	7.78	5.15	-51%	4.20	4.46	5.8%	3.27	7.11	0.460	-0.78	-0.73	-6.8%
GISS-E2-1-G	4.43	2.54	74%	2.63	2.62	-0.4%	3.78	7.81	0.484	-1.43	-1.44	0.7%
GISS-E2-2-G	2.71	0.44	-515.9%	2.49	2.36	-5.5%	3.74	7.59	0.493	-1.50	-1.59	5.7%
GISS-E2-1-H	3.90	4.18	6.7%	2.96	3.20	7.5%	3.76	7.28	0.516	-1.27	-1.17	-8.5%
IPSL-CM6A-LR	6.02	6.72	11%	3.97	4.80	17.3%	3.69	7.06	0.523	-0.93	-0.77	-20.8%
HadGEM3-GC31-LL	8.79	8.58	-2.4%	4.88	5.36	9.0%	3.34	6.89	0.485	-0.68	-0.62	-9.7%
MIROC6	2.79	1.84	-52%	2.18	2.41	9.5%	3.52	7.43	0.474	-1.62	-1.46	-11.0%
<b>MRI-ESM2-0</b>	<b>-0.28*</b>	<b>5.62</b>	<b>105%</b>	<b>2.48</b>	<b>3.25</b>	<b>23.7%</b>	<b>3.48</b>	<b>6.70</b>	<b>0.519</b>	<b>-1.40</b>	<b>-1.07</b>	<b>-30.8%</b>
<i>Ensemble Mean</i>	<i>5.37 ± 2.27</i>	<i>4.66 ± 2.94</i>	<i>-21 ± 37%</i>	<i>3.50 ± 1.06</i>	<i>3.82 ± 1.31</i>	<i>6.99 ± 7.19</i>	<i>3.58 ± 0.20</i>	<i>7.33 ± 0.30</i>	<i>0.488 ± 0.02</i>	<i>-1.12 ± 0.38</i>	<i>-1.06 ± 0.40</i>	<i>-8.20 ± 8.36</i>

*Note.* (a) The slope of the temperature anomaly against time for years 21 to 150, (b) the effective climate sensitivity, (c) the effective radiative forcing estimated from years 1 to 150, and (d) the net radiative feedback parameter estimated from years 1 to 150 for abrupt-2xCO<sub>2</sub>. Each statistic includes the (a, b, d) scaled abrupt-4xCO<sub>2</sub> value or (c) the unscaled abrupt-4xCO<sub>2</sub> value. (a, b, d) the percent difference (Equation 1), or (c) the ratio of abrupt-2xCO<sub>2</sub> and abrupt-4xCO<sub>2</sub>. The ensemble mean percent difference, ratio, and their respective standard deviations are calculated as the mean of each model's percent difference/ratio and the standard deviation of that column. All of the linear regressions have  $p < 0.05$  (\*) except for the column (a) MRI-ESM2-0 abrupt-2xCO<sub>2</sub>, which has a  $p$ -value of  $p = 0.0503$ . Using a  $t$ -test, we find that the difference in the radiative feedback parameter between forcings is significant ( $p < 0.05$ ) for CanESM5, CESM2, IPSL-CM6A-LR, HadGEM3-GC31-LL, and MRI-ESM2-0 (Table S4 in Supporting Information S1). Our models of interest, CESM2 and MRI-ESM2-0 are bolded.



**Figure 1.** (a) Surface temperature anomaly against time, (b) the net TOA radiative flux against surface temperature anomaly (see Equation 2), (c) the effective heat capacity against time for abrupt-2xCO<sub>2</sub> and scaled abrupt-4xCO<sub>2</sub>. The legend beneath the first row corresponds with panels (a)–(c). (d) The AMOC strength anomaly against time. (e) The difference in the zonal-mean depth of 55% heat storage between abrupt-2xCO<sub>2</sub> and abrupt-4xCO<sub>2</sub> (abrupt-2xCO<sub>2</sub> minus abrupt-4xCO<sub>2</sub>) after the first 20 years. (f) The EBM surface temperature anomaly against time. (a)–(f) We use annual mean data which is then (a, c, d, f) filtered using a Savitzky-Golay with a 31-year filter (temperature) or a 5-year (heat capacity and AMOC strength) filter to enhance readability. The abrupt-4xCO<sub>2</sub> data is scaled in (a–c, f) and unscaled in (d).

The substantially (44.8% and 37.9%) smaller abrupt-2xCO<sub>2</sub> warming compared to the scaled abrupt-4xCO<sub>2</sub> warming in MRI-ESM2-0 and CESM2 (respectively) cannot be explained by non-logarithmic CO<sub>2</sub>-forcing in these models because we scaled the simulations using model-specific ERF ratios (Table 1c). Additionally, CESM2 and MRI-ESM2-0 have the largest warming rate difference between abrupt-2xCO<sub>2</sub> and abrupt-4xCO<sub>2</sub>, as in, the most suppressed warming rate at abrupt-2xCO<sub>2</sub> compared to abrupt-4xCO<sub>2</sub> (Table 1a). The rate of change in temperature over time (the slope of a linear regression of annual mean temperature over time) for MRI-ESM2-0 is negative for years 21–150 (the period after the initial rapid warming; Figure 1a). In other words, MRI-ESM2-0 has stalled in its warming over the final century of the 2xCO<sub>2</sub> run (but not in its abrupt-4xCO<sub>2</sub> run). In contrast, GISS-E2-2-G shows the reverse behavior wherein the rate of warming at abrupt-4xCO<sub>2</sub> is near-zero.

We see a similar pattern when comparing the EffCS across forcings. We perform a Gregory regression assuming the standard model of global energy balance, defined as:

$$R_{\text{TOA}}(t) = F + \lambda T(t), \quad (2)$$

where  $R$  is the TOA radiative anomaly,  $T$  is the near-surface air temperature anomaly,  $F$  is the ERF, and  $\lambda$  is the radiative feedback parameter. If the forcing is a doubling of CO<sub>2</sub> ( $F_{2x}$ ) or scaled to a doubling of CO<sub>2</sub> (e.g.,  $\mu F_{4x}$ , where  $\mu$  is the model specific scaling factor), the EffCS can be estimated as the  $x$ -intercept, or the temperature when the excess TOA radiative flux reaches zero:

$$\begin{aligned}\text{EffCS}_{2x} &= T(R=0) = \frac{-F_{2x}}{\lambda_{2x}} \\ \text{EffCS}_{4x} &= T(R=0) = \frac{-\mu F_{4x}}{\lambda_{4x}},\end{aligned}\quad (3)$$

where  $\lambda_{2x}$  and  $\lambda_{4x}$  are the net radiative feedbacks for abrupt-2xCO<sub>2</sub> and abrupt-4xCO<sub>2</sub>, respectively. The abrupt-2xCO<sub>2</sub> EffCS is smaller than the abrupt-4xCO<sub>2</sub> EffCS in nearly all CMIP6 models studied here (Table 1b); however, CESM2 and MRI-ESM2-0 have the largest difference (47.3% and 23.7%, respectively) owing to large differences in their radiative feedback parameters (Table 1d). Interestingly, GISS-E2-2-G has only a −5.5% difference in EffCS and a 5.7% difference in the radiative feedback parameter between abrupt-2xCO<sub>2</sub> and abrupt-4xCO<sub>2</sub> (Table 1b, d).

While previous studies (e.g., Bloch-Johnson et al., 2021) have focused on the dependence of transient warming on differences in radiative feedbacks (and thus EffCS) between warming levels, here we also consider the global effective heat capacity. The role of effective heat capacity can be seen by extending the linear energy balance model (Equation 2) to include time dependence (Donohoe et al., 2014):

$$\frac{d}{dt}(C(t) * T(t)) = F + \lambda T(t) = R_{\text{TOA}}(t) \quad (4)$$

where  $C$  is the heat capacity of the Earth system and  $t$  is time. In our energy balance equation, the time evolution of temperature is controlled by three variables: the heat capacity  $C$  (in units of J/m<sup>2</sup>/K), the ERF (W/m<sup>2</sup>), and the radiative feedback parameter (W/m<sup>2</sup>/K). We imagine a model in which warming signals penetrate downward through the ocean over a depth that may depend on climate state. Heat absorbed by the Earth system corresponds with either an increase in surface air temperature or an increase in the effective depth over which a temperature anomaly is distributed (an increase in effective heat capacity). We can calculate the heat capacity using the following equation:

$$C(t) = \frac{\int_0^t R_{\text{TOA}}(t') dt'}{T(t')} \quad (5)$$

In what follows, we will compare CESM2 and MRI-ESM2-0 to the remaining eight-model ensemble mean to explore the following research questions:

1. What are the roles of radiative feedbacks and effective heat capacity in driving reduced warming in abrupt-2xCO<sub>2</sub> relative to the scaled abrupt-4xCO<sub>2</sub>?
2. What mechanisms cause differences in radiative feedbacks and effective heat capacity between abrupt-2xCO<sub>2</sub> and scaled abrupt-4xCO<sub>2</sub> scenarios?

### 3. Results

#### 3.1. Radiative Feedbacks and Effective Heat Capacity Across Forcings

##### 3.1.1. Net Radiative Feedback Across Forcings

A more-negative radiative feedback parameter (and consequently, lower EffCS) in abrupt-2xCO<sub>2</sub> than in abrupt-4xCO<sub>2</sub> contributes to reduced warming in abrupt-2xCO<sub>2</sub> compared to scaled abrupt-4xCO<sub>2</sub>. In Table 1d, we see that the radiative feedback parameter (calculated over 150 years) across the mean ensemble is more negative in abrupt-2xCO<sub>2</sub> (−1.12 ± 0.38 W/m<sup>2</sup>/K) compared to abrupt-4xCO<sub>2</sub> (−1.06 ± 0.40 W/m<sup>2</sup>/K) with an ensemble-mean percent difference of −8.20% ± 8.36. GISS-E2-1-G and GISS-E2-2-G are the only models where the radiative feedback is more positive in abrupt-2xCO<sub>2</sub>. CESM2 and MRI-ESM2-0 are once again outliers, in that the radiative feedback parameter has the largest difference across forcings, with a percent difference of −92.1% and −30.8%, respectively (Table 1d; Figure 1b; Table S4 in Supporting Information S1).



### 3.1.2. The Effective Heat Capacity Across Forcings

The time series of heat capacity for CESM2 and MRI-ESM2-0 also differs from the ensemble mean (Figure 1c). The ensemble mean heat capacity is nearly identical across forcings; the final 20-year average effective heat capacity is  $0.050 \pm 0.28$  GJ/m<sup>2</sup>/K larger at abrupt-4xCO<sub>2</sub> than abrupt-2xCO<sub>2</sub> (increasing from 1.85 GJ/m<sup>2</sup>/K at abrupt-2xCO<sub>2</sub> and 1.90 GJ/m<sup>2</sup>/K at abrupt-4xCO<sub>2</sub>), or, equivalently, the effective depth of heat storage is  $11 \pm 65$  m deeper in abrupt-4xCO<sub>2</sub> than abrupt-2xCO<sub>2</sub> (Table 2a). However, in CESM2 and MRI-ESM2-0, the abrupt-2xCO<sub>2</sub> heat capacity is substantially larger than that of abrupt-4xCO<sub>2</sub> at all times (Figure 1c). The final heat capacity in CESM2 and MRI-ESM2-0 is 0.8 and 1.5 GJ/m<sup>2</sup>/K larger in abrupt-2xCO<sub>2</sub> than in abrupt-4xCO<sub>2</sub>, nearly 190 and 350 m deeper, respectively. In contrast, GISS-E2-2-G has a larger effective heat capacity at abrupt-4xCO<sub>2</sub> than abrupt-2xCO<sub>2</sub> and subsequently a deeper effective depth of heat storage (155 m deeper; Table 2a); this may be related to the suppressed abrupt-4xCO<sub>2</sub> warming rate in GISS-E2-2-G.

### 3.1.3. The Relative Importance of the Net Radiative Feedback and the Effective Heat Capacity

To explore the relative importance of feedbacks and heat capacity in setting the differences in global warming rate between forcing levels, we consider the energy balance model (Equation 4) in the following form:

$$\frac{dT}{dt} = \frac{F + \lambda T - T \frac{dC}{dt}}{C} \quad (6)$$

We use values of the radiative feedback parameter, the effective radiative forcing, and the effective heat capacity calculated over 150 years (Table S6 in Supporting Information S1) to recreate a global-mean time series for each model. To fit the temperature time series, we use a 1-year time step, and heat capacity data filtered using a Savitzky–Golay filter, a method of locally estimated scatterplot smoothing (LOESS; Chen et al., 2004; Savitzky & Golay, 1964), 1-degree polynomial fit with a window of 5 years to reduce the impact of natural variability. The EBM fits the global-mean annual temperature data well with an average error of  $\pm 0.08$  and  $\pm 0.14$  K for the abrupt-4xCO<sub>2</sub> and abrupt-2xCO<sub>2</sub> ensemble mean data, respectively (Table S6, Figure S1 in Supporting Information S1).

We run two EBM experiments to calculate the relative importance of the radiative feedback parameter and the effective heat capacity: (a) the radiative feedback parameter from abrupt-2xCO<sub>2</sub> and the effective heat capacity from abrupt-4xCO<sub>2</sub> (referred to as the “ $\lambda_{2x}/C_{4x}$ ” experiment), and (b) the radiative feedback from abrupt-4xCO<sub>2</sub> and the effective heat capacity from abrupt-2xCO<sub>2</sub> (referred to as the “ $\lambda_{4x}/C_{2x}$ ” experiment). The EBM time serieses are in Figure 1f. Thus, the relative importance of the radiative feedback is,

$$I_\lambda = \frac{\mu T_{4x} - T_{\lambda_{2x}/C_{4x}}}{\mu T_{4x} - T_{2x}} \times 100\%, \quad (7)$$

and the relative importance of the effective heat capacity is,

$$I_c = \frac{\mu T_{4x} - T_{\lambda_{4x}/C_{2x}}}{\mu T_{4x} - T_{2x}} \times 100\%, \quad (8)$$

with a residual of

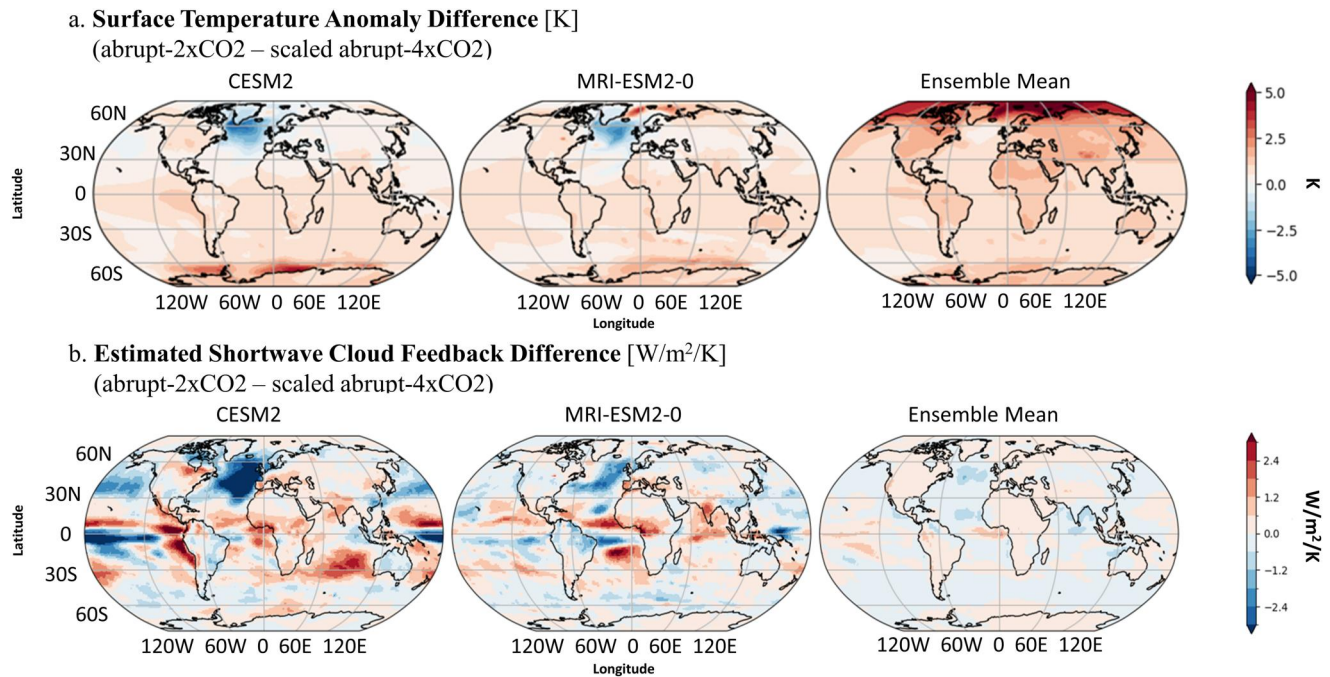
$$I_r = \left( 1 - \frac{\mu T_{4x} - T_{\lambda_{2x}/C_{4x}}}{\mu T_{4x} - T_{2x}} - \frac{\mu T_{4x} - T_{\lambda_{4x}/C_{2x}}}{\mu T_{4x} - T_{2x}} \right) \times 100\% \quad (9)$$

To estimate each relative importance, we use the final 20-year average of the global-mean temperature time series (Table 2b). Five models have a relative importance of effective heat capacity larger than the ensemble average: CESM2, GISS-E2-1-G, GISS-E2-2-G, HadGEM3-GC31-LL, and MRI-ESM2-0 (Table 2b). For CESM2, the radiative feedback parameter accounts for 80.8% and the effective heat capacity accounts for 42.2% of the difference between scaled abrupt-4xCO<sub>2</sub> and abrupt-2xCO<sub>2</sub> runs (with a residual of  $-23.0\%$ ). In MRI-ESM2, the radiative feedback parameter accounts for 49.5% and the effective heat capacity accounts for 69.7 of that difference (with a residual of  $-17.3\%$ ). Interestingly GISS-E2-1-G, GISS-E2-2-G, and HadGEM3-GC31-LL also

**Table 2**  
Comparison of Heat Capacity, AMOC Decline, and Depth of Heat Storage in Abrupt-2xCO<sub>2</sub> and Abrupt-4xCO<sub>2</sub> Simulations

Model	(a) Heat capacity			(b) Relative importance			(c) AMOC decline			(d) Mean NH depth of 55% heat storage		
	C <sub>2xCO<sub>2</sub></sub> [GJ/m <sup>2</sup> /K] (m)	C <sub>4xCO<sub>2</sub></sub> [GJ/m <sup>2</sup> /K] (m)	Difference [GJ/m <sup>2</sup> /K] (m)	I <sub>h</sub> [%]	I <sub>c</sub> [%]	Residual [%]	ΔΨ <sub>2xCO<sub>2</sub></sub> [Sv]	ΔΨ <sub>4xCO<sub>2</sub></sub> [Sv]	Ratio [%]	Z <sub>2xCO<sub>2</sub></sub> [m]	Z <sub>4xCO<sub>2</sub></sub> [m]	Difference [m]
CanESM5	1.54 (357)	1.55 (360)	-0.01 (-2.3)	101.89	-2.96	1.07	-3.03	-7.98	38%	587	505	82
<b>CESM2</b>	<b>2.61 (606)</b>	<b>1.79 (415)</b>	<b>0.82 (190)</b>	<b>80.76</b>	<b>42.24</b>	<b>-23.00</b>	-15.24	<b>-20.87</b>	<b>73%</b>	<b>697</b>	<b>574</b>	<b>123</b>
CNRM-CM6-1	1.93 (448)	1.89 (438)	0.04 (9.28)	297.69	-199.97	2.28	-10.22	-16.29	62%	655	603	52
GISS-E2-1-G	2.26 (524)	2.49 (578)	-0.23 (-53.4)	111.02	<b>87.98</b>	1.00	+0.87	-19.26	-4.5%	958	839	119
GISS-E2-2-G	1.43 (332)	2.10 (487)	-0.67 (-155)	30.07	63.69	6.24	-2.06	-16.03	-13%	673	790	-117
GISS-E2-1-H	1.91 (443)	1.88 (435)	0.03 (6.96)	94.48	6.93	-1.41	-	-	-	683	659	24
HadGEM3-GC31-LL	1.74 (403)	1.57 (364)	0.17 (39.4)	65.60	39.15	-4.75	-	-17.28	-	-	-	-
IPSL-CM6A-LR	1.38 (320)	1.33 (309)	0.05 (11.6)	108.08	-7.76	-0.32	-	-	-	-	528	-
MIROC6	2.64 (612)	2.42 (562)	0.22 (51.1)	113.96	-11.93	-2.03	-	-15.67	-	-	-	-
<b>MRI-ESM2-0</b>	<b>3.71 (861)</b>	<b>2.18 (506)</b>	<b>1.53 (355)</b>	<b>49.50</b>	<b>69.74</b>	<b>-19.24</b>	-19.16	<b>-20.27</b>	<b>95%</b>	<b>919</b>	<b>796</b>	<b>122</b>
Ensemble Mean	1.85 ± 0.43 (430 ± 100)	1.90 ± 0.42 (441 ± 98.0)	-0.050 ± 0.28 (-11.6 ± 65.1)	102.9 ± 87.26	-3.11 ± 87.42	0.26 ± 3.26	-	-	-	-	-	-

*Note.* The final 20 years average of (a) the effective heat capacity, (b) the relative importance of the radiative feedbacks (Equation 7) and the effective heat capacity (Equation 8), (c) the change in AMOC decline from pre-industrial, and (d) the global-mean depth of 55% heat storage for (a, c, d) abrupt-2xCO<sub>2</sub>, abrupt-4xCO<sub>2</sub>, (a, d) the difference between abrupt-2xCO<sub>2</sub> and abrupt-4xCO<sub>2</sub>, (b) the residual relative importance (Equation 9), and (c) the ratio of the abrupt-2xCO<sub>2</sub> and the unscalled abrupt-4xCO<sub>2</sub>, represented as a percent. See Table S6 in Supporting Information S1 for the average fit error of EBM results. GISS-E2-1-G does not have piControl streamfunction data available, so the first 10 and final 20 years are used. The ensemble mean is only included for analysis where all 10 model data are available. Our models of interest, CESM2 and MRI-ESM2-0 are bolded.



**Figure 2.** Differences (a) in surface temperature anomaly in K, and (b) the APRP shortwave cloud feedback in W/m<sup>2</sup>/K (a), (b) between abrupt-2xCO<sub>2</sub> and scaled abrupt-4xCO<sub>2</sub> for the years 1–150 for CESM2 (left), MRI-ESM2-0 (middle) and the ensemble mean excluding CESM2 and MRI-ESM2-0 (right).

show a large role for heat capacity changes, with the radiative feedback parameter accounting for 11.0%, 30.7%, and 65.6% and the effective heat capacity accounts for 88.0%, 63.7%, and 39.2% of that difference (with a residual of 1.7%, 6.2%, and –4.8%), respectively; although these three models show little difference between the abrupt-2xCO<sub>2</sub> and scaled abrupt-4xCO<sub>2</sub> EffCS (–0.01, 0.13, and 0.48 K difference in EffCS; Table 1a, b). Contrastingly, for the ensemble, the radiative feedback parameter accounts for 102.9% and the effective heat capacity only accounts for –3.1% of the difference in warming between scaled abrupt-4xCO<sub>2</sub> and abrupt-2xCO<sub>2</sub> (with a residual of 0.3%).

The EBM has a nonlinear dependence on the parameters  $\lambda$  and  $C$ . Thus, we expect that models with large differences in parameters (the radiative feedback parameter and effective heat capacity) across forcings to have large residuals, as is the case for CESM2 and MRI-ESM2-0. In summary: 1. For CESM2, a more-negative radiative feedback parameter is the dominant mechanism and a larger effective heat capacity is a significant mechanism (compared to the ensemble) in the slowed warming in abrupt-2xCO<sub>2</sub> simulations, and 2. For MRI-ESM2-0, the effective heat capacity is the dominant mechanism causing less warming in abrupt-2xCO<sub>2</sub> simulations.

### 3.2. Mechanisms of Forcing Dependent Radiative Feedbacks and Heat Capacity

#### 3.2.1. Mechanisms of Radiative Feedback Changes

The phenomenon of EffCS increasing under larger forcing is often characterized by a global-mean second-order feedback temperature dependence (e.g., Bloch-Johnson et al., 2021). However, we know that the spatial pattern of warming also has an impact on sensitivity (e.g., Armour et al., 2013; Dong et al., 2020; Rose et al., 2014; Zhou et al., 2017). If the spatial pattern of warming changes with forcing level (Good et al., 2015; Lin et al., 2019; Mitevski et al., 2023), then a global-mean framework would not be sufficient to understand changes in sensitivity with forcing. Thus, we analyze the spatial structure of the global temperature anomaly (Figure 2a) and radiative feedbacks (Figure 2b).

Our decomposition of the global-mean Gregory regression finds that CESM2 and MRI-ESM2-0 exhibit the largest difference in the shortwave cloud feedback across forcings, increasing 0.23 and 0.19 W/m<sup>2</sup>/K from abrupt-2xCO<sub>2</sub> to abrupt-4xCO<sub>2</sub>, which accounts for 36.8% and 121.2% of the total feedback difference across forcings, respectively (Table S5 in Supporting Information S1; Andrews et al., 2012). Therefore, we focus on the shortwave



cloud budget hereafter. Both models also see an increase in the longwave clear sky feedback with forcing (Table S5 in Supporting Information S1), consistent with previously proposed mechanisms for a positive second-order temperature dependence (Bloch-Johnson et al., 2021).

We estimate the spatial structure of the shortwave cloud feedback using the approximate partial radiative perturbation (APRP) method (Taylor et al., 2007) wherein we divide the changes in the spatial energy budget calculated from APRP by the global mean temperature anomaly for the years 1–150. We find that CESM2 and MRI-ESM2-0 have a strongly negative shortwave cloud feedback in the North Atlantic, south of Greenland, in abrupt-2xCO<sub>2</sub> likely due to an increase in low cloud cover (Figure 2b; Ceppi et al., 2016, 2017). In contrast, the ensemble shortwave cloud feedback is more negative in abrupt-4xCO<sub>2</sub> than abrupt-2xCO<sub>2</sub>. Additionally, CESM2 and MRI-ESM2-0 stand out as having cooling in the North Atlantic indicative of the North Atlantic Warming Hole (NAWH) in abrupt-2xCO<sub>2</sub> and substantially less warming in the same location in abrupt-4xCO<sub>2</sub> (Figure 2a). The enhanced NAWH low cloud cover (Figures 2a and 2b; Jackson et al., 2015; Zhang et al., 2010) under abrupt-2xCO<sub>2</sub> for CESM2 and MRI-ESM2-0 is consistent with Atlantic Meridional Overturning Circulation (AMOC) decline. AMOC decline has a marked impact on global energy transport (Drijfhout, 2015; Kostov et al., 2014; Liu et al., 2020; Mecking & Drijfhout, 2023; Poletti et al., 2024); particularly, AMOC decline impacts the spatial pattern of warming by reducing the northward transport of heat, leading to colder SSTs in the subpolar North Atlantic, increased estimated inversion strength, and enhanced low cloud cover at the NAWH (Drijfhout, 2015; Drijfhout et al., 2012; Hu et al., 2020; Lin et al., 2019; Mitevski et al., 2023; Rugenstein et al., 2013; Trossman et al., 2016), just as we see in CESM2 and MRI-ESM2-0 in abrupt-2xCO<sub>2</sub>. The APRP and spatial feedback results for clear sky, albedo, and net effects are in supporting Figures S3 and S4 in Supporting Information S1.

### 3.2.2. Ocean Dynamics and the Atlantic Meridional Overturning Circulation

Differences in AMOC decline across models can be linked to differences in hemispheric warming (Lin et al., 2019) and in ocean heat storage and uptake (Frolicher et al., 2015) across models. Given the impact of North Atlantic Ocean dynamics on the spatial pattern of warming, we further investigate the AMOC strength anomaly and the effective depth of heat storage.

Lin et al. (2023) found that models with a stronger climatological AMOC experience larger declines at abrupt-4xCO<sub>2</sub> and in 1% ramping experiments. However, AMOC decline may not scale with forcing; as in some models may approach a nearly complete AMOC shutdown under abrupt-2xCO<sub>2</sub>, leaving little room for further decrease in abrupt-4xCO<sub>2</sub>. Since the impact that AMOC decline has on transient global warming scales with the magnitude of the decline (Hu et al., 2020), a disproportionate AMOC decline in lower forcing scenarios would introduce a disproportionate decrease in global warming rate in abrupt-2xCO<sub>2</sub> compared to abrupt-4xCO<sub>2</sub> due to a lower limit in AMOC strength (i.e., if AMOC nearly collapses under abrupt-2xCO<sub>2</sub>, it cannot decline any more than that under abrupt-4xCO<sub>2</sub>).

We define the AMOC strength ( $\Psi$ ) as the maximum streamfunction anomaly value north of 30°N and below 500 m (Lin et al., 2019) and the AMOC decline as the difference between the piControl and final 20-year average AMOC strength. In abrupt-4xCO<sub>2</sub>, the ensemble average AMOC strength decreases while in abrupt-2xCO<sub>2</sub>, the ensemble average AMOC strength decreases and then partially recovers (Figure 1d). The abrupt-2xCO<sub>2</sub> ensemble average has larger differences across models, particularly during the decades of AMOC recovery.

CESM2 and MRI-ESM2-0 have the most disproportionate AMOC declines in abrupt-2xCO<sub>2</sub> compared to abrupt-4xCO<sub>2</sub> (Figure 1d). CESM2 and MRI-ESM2-0 decline 72% and 95% of the abrupt-4xCO<sub>2</sub> value under abrupt-2xCO<sub>2</sub> forcing (Table 2c). CNRM-CM6A-LR also shows a larger AMOC decline in abrupt-2xCO<sub>2</sub> than abrupt-4xCO<sub>2</sub> with a ratio of 62% (Table 2d) but without large changes in EffCS (Table 1b) or effective heat capacity (Table 2a) across forcings. Since we only have a subset of the total ensemble in our AMOC analysis, we cannot compare the AMOC decline ratios to an ensemble mean. Therefore, it's possible that CNRM-CM6A-LR has a typical abrupt-2xCO<sub>2</sub> AMOC decline (when compared to the theoretical ensemble mean) which is not large enough to cause the nonlinearities seen in CESM2 and MRI-ESM2-0. In contrast, abrupt-2xCO<sub>2</sub> AMOC decline in CanESM5 is 38% of the abrupt-4xCO<sub>2</sub> value. GISS-E2-1-G is the only model to experience AMOC recovery to its initial strength in abrupt-2xCO<sub>2</sub> (as in, the AMOC strength declines and then strengthens). Mitevski et al. (2021) shows that GISS-E2-1-G has a disproportionately large AMOC decline and smaller EffCS at abrupt-3xCO<sub>2</sub>, suggesting a large degree of model dependence in the dynamics identified here.

We can use an analysis of column integrated heat storage to find the depth over which most of the anomalous heat is stored:

$$\text{Percent of Heat Storage (z)} = \frac{\int_0^z \rho C_p T(t, z', \lambda, \theta) dz'}{\int_0^{z_B} \rho C_p T(t, z', \lambda, \theta) dz'} \times 100\% \quad (10)$$

We can solve for the depth  $z$  at which some percentage of heat storage has accumulated. In comparing the heat capacity time series and the global depth of heat storage time series, we found that the depth of 55% heat storage was most representative of the heat capacity (see Supporting Information S1); however, these results are robust between 50% and 80% (Figure S5 in Supporting Information S1).

We first consider the difference in zonal structure of the depth of heat storage between abrupt-2xCO<sub>2</sub> and abrupt-4xCO<sub>2</sub>, after the first 20 years (Figure 1e; Figure S6 in Supporting Information S1). Figure 1e only includes models with available ocean potential temperature data at both forcings which are: CanESM5, CESM2, CNRM-CM6-1, GISS-E2-1-G, GISS-E2-2-G, GISS-E2-1-H, and MRI-ESM2-0 (Table S3 in Supporting Information S1). While the Southern Ocean is known to dominate ocean heat storage (Frolicher et al., 2015; Winton et al., 2013), we focus our analysis on the depth of heat storage since it is correlated with the effective heat capacity (Donohoe et al., 2014). We find that the difference in depth of ocean heat storage from abrupt-4xCO<sub>2</sub> to abrupt-2xCO<sub>2</sub> is roughly equivalent in the Southern and Northern oceans.

All of the models show larger peaks in depth of heat storage in the Southern Ocean and the North Atlantic/North Pacific in abrupt-2xCO<sub>2</sub> simulations. CESM2 has enhanced depth of heat storage in abrupt-2xCO<sub>2</sub> at all latitudes, compared to the ensemble mean, but especially in the North Atlantic/North Pacific and the Southern Ocean. However, the difference in depth of heat storage for MRI-ESM2-0 is substantially larger than the ensemble mean *only* in the North Atlantic/North Pacific. Thus, AMOC decline may act as a mechanism of enhanced depth of heat storage through surface cooling in the subpolar North Atlantic (Kostov et al., 2014; Winton et al., 2013).

Given the North Hemisphere signal in the zonal structure, we next consider the final 20-year average depth of heat storage in the Northern Hemisphere (Table 2d; Figure S7 in Supporting Information S1). CESM2, MRI-ESM2-0, and GISS-E2-1-G have the deepest heat storage in abrupt-2xCO<sub>2</sub> (but not necessarily at abrupt-4xCO<sub>2</sub>) and the largest difference between abrupt-2xCO<sub>2</sub> and abrupt-4xCO<sub>2</sub>. Thus, the three models with the largest effective heat capacity (Figure 1c), the largest importance of heat capacity in determining transient warming, and the deepest abrupt-2xCO<sub>2</sub> ocean heat storage are also the three models with disproportionate AMOC declines (either at abrupt-2xCO<sub>2</sub> or abrupt-3xCO<sub>2</sub>; Table 2; Mitevski et al., 2021). Possibly, ocean stratification, more broadly, scales nonlinearly with forcing; however, the exact mechanisms linking AMOC decline to changes in global depth of heat storage and effective heat capacity deserve further study.

#### 4. Conclusions

We began this study with the observation that the rate of transient warming in CESM2 and MRI-ESM2-0 are significantly reduced in abrupt-2xCO<sub>2</sub> compared to scaled abrupt-4xCO<sub>2</sub> simulations, even when accounting for the non-logarithmic scaling of CO<sub>2</sub>-forcing with concentration. Previous literature attributes differences in warming rates across forcing levels to differences in radiative feedbacks and thus EffCS (e.g., Meraner et al., 2013). These differences are often characterized in terms of a global-mean second-order feedback-temperature dependence (e.g., Bloch-Johnson et al., 2021). However, we find that both radiative feedbacks *and* the global effective heat capacity are necessary in explaining warming differences between abrupt-2xCO<sub>2</sub> and abrupt-4xCO<sub>2</sub> simulations of most models. Additionally, radiative feedbacks and the heat capacity appear to be linked by spatial warming pattern changes possibly caused by nonlinearities in AMOC decline with forcing. Thus, a global-mean framework would be insufficient for explaining higher model sensitivity at higher forcings.

For CESM2 and MRI-ESM2, the more-negative radiative feedback parameter in abrupt-2xCO<sub>2</sub> is in part caused by reduced shortwave cloud feedbacks from disproportionately large AMOC decline. The larger effective heat capacity in abrupt-2xCO<sub>2</sub> is reflected in enhanced depth of ocean heat storage, particularly in the Southern Ocean and Atlantic Ocean. Therefore, a large effective heat capacity and enhanced depth of Northern Hemisphere heat storage are correlated with models that exhibit large AMOC declines at abrupt-2xCO<sub>2</sub>. This indicates a potential

mechanism where AMOC decline can slow global warming through the depth of ocean heat storage and enhanced low cloud cover. This mechanism is relevant at any forcing, not just abrupt-2xCO<sub>2</sub>. For instance, GISS-E2-1-G experiences disproportionate AMOC decline at abrupt-3xCO<sub>2</sub> (Mitevski et al., 2021) rather than abrupt-2xCO<sub>2</sub>. Thus, in other GCMs it may be other forcings levels which experience suppressed warming due to AMOC nonlinearity.

This study is constrained first by 150-year runs (while ocean interactions can occur on centennial to millennial timescales; Ackerman et al., 2020; Cheng et al., 2013; Weaver et al., 2012), second by the number of available abrupt-2xCO<sub>2</sub> data sets, and third by the lack of Fixed-SST runs for abrupt-2xCO<sub>2</sub> scenarios to more accurately estimate the ratio of the radiative forcings.

Ultimately, nonlinearities in the magnitude of AMOC decline across forcings are a possible cause of reduced warming in CESM2 and MRI-ESM2-0 under an abrupt-doubling of CO<sub>2</sub>, through both radiative feedbacks and the effective heat capacity. The phenomenon where AMOC decline is proportionally larger at lower forcings (Drijfhout, 2015; Good et al., 2015; Mitevski et al., 2021; Rugenstein et al., 2016) is well documented. The magnitude of AMOC decline is highly variable across models (Cheng et al., 2013; Frolicher et al., 2015; Lin et al., 2023) and introduces significant variability in the global temperature pattern (Lin et al., 2019). In some GCMs, the AMOC declines by well over 50% of its initial strength under abrupt-doubling of CO<sub>2</sub>, which makes it impossible to see a proportional decline under abrupt-quadrupling unless the circulation is to reverse. Thus, forcing-dependent changes in ocean dynamics can dramatically impact the rate of transient warming across GCMs while introducing significant variability across models.

Abrupt-4xCO<sub>2</sub> simulations are still used definitionally to calculate effective climate sensitivity (EffCS), and these EffCS values have been argued to govern future warming independent of the scenario used (e.g., Sherwood et al., 2020; IPCC AR6). We've demonstrated that this is a poor assumption within a subset of CMIP6 models for which AMOC decline with forcing is nonlinear with forcing level. Second, we have analyzed the mechanisms by which a nonlinear AMOC decline with forcing level causes warming under 2xCO<sub>2</sub> to be substantially less than warming under 4xCO<sub>2</sub> (after normalizing by ERF differences). Here, we find links between AMOC decline, spatial patterns of SST changes, and radiative feedbacks—similar to several previous analyses (e.g., Drijfhout, 2015; Drijfhout et al., 2012; Lin et al., 2019; Trossman et al., 2016) but at odds with the common assumption that feedback differences between 2xCO<sub>2</sub> and 4xCO<sub>2</sub> scale with global-mean warming (Bloch-Johnson et al., 2021).

In summary, our analysis reveals that AMOC nonlinearity with CO<sub>2</sub> forcing level affects global-mean warming in large part through changes in the effective heat capacity, rather than feedback changes alone. When AMOC changes are substantially different under lower forcing, abrupt-quadrupling experiments alone will not constrain effective equilibrium climate sensitivity. As we exceed 1.5 times pre-industrial CO<sub>2</sub> concentrations (UK Met Office, 2021), simulations of climate and ocean dynamics across different forcings become increasingly relevant for projecting global warming in the 21st century.

## Data Availability Statement

The GCM data used for this study can be found in the CMIP6 data repository at <https://esgf-node.llnl.gov/search/cmip6/>. The data and code used to create the figures and tables is available at <https://doi.org/10.5281/zenodo.7829225> (Poletti, 2023).

## Acknowledgments

Poletti was funded through the UW Provost Fellowship, UW Future Rivers Fellowship, and the National Science Foundation (Grant AGS-1665247). Frierson was funded through the same National Science Foundation Grant AGS-1665247. Armour was funded by the National Science Foundation (Grant AGS-1752796), the National Oceanic and Atmospheric Administration MAPP Program (Award NA20OAR4310391), and an Alfred P. Sloan Research Fellowship (Grant FG-2020-13568). We thank Aaron Levine for the APRP code.

## References

- Ackerman, L., Danek, C., Gierz, P., & Lohmann, G. (2020). AMOC recovery in a multicentennial scenario using a coupled atmosphere-ocean-ice sheet model. *Geophysical Research Letters*, 47(16). <https://doi.org/10.1029/2019GL086810>
- Andrews, T., Gregory, J. M., Webb, M. J., & Taylor, K. E. (2012). Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models. *Geophysical Research Letters*, 39(9). <https://doi.org/10.1029/2012GL051607>
- 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>
- Bloch-Johnson, J., Rugenstein, M., Stolpe, M. B., Rohrschneider, T., Zheng, Y., & Gregory, J. M. (2021). Climate sensitivity increases under higher CO<sub>2</sub> levels due to feedback temperature dependence. *Geophysical Research Letters*, 48(4), e2020GL089074. <https://doi.org/10.1029/2020GL089074>
- Byrne, B., & Goldblatt, C. (2013). Radiative forcing at high concentrations of well-mixed greenhouse gases. *Geophysical Research Letters*, 41(1), 152–160. <https://doi.org/10.1002/2013GL058456>
- Ceppi, P., Briant, F., Zelinka, M. D., & Hartmann, D. L. (2017). Cloud feedback mechanisms and their representation in global climate models. *WIREs Climate Change*, 8(4), e465. <https://doi.org/10.1002/wcc.465>

- Ceppi, P., McCoy, D. T., & Hartmann, D. L. (2016). Observational evidence for a negative shortwave cloud feedback in middle to high latitudes. *Geophysical Research Letters*, 43(3), 1331–1339. <https://doi.org/10.1002/2015GL067499>
- Chen, J., Jonsson, P., Tamura, M., Gu, Z., Matsushita, B., & Eklundh, L. (2004). A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter. *Remote Sensing of Environment*, 91(3–4), 332–344. <https://doi.org/10.1016/j.rse.2004.03.014>
- Cheng, W., Chiang, J. C. H., & Zhang, D. (2013). Atlantic meridional overturning circulation (AMOC) in CMIP5 models: RCP and historical simulations. *Journal of Climate*, 26(18), 7187–7197. <https://doi.org/10.1175/JCLI-D-12-00496.1>
- Dong, Y., Armour, K. C., Zelinka, M. D., Proistosescu, C., Battisti, D. S., Zhou, C., & Andrews, T. (2020). Intermodel spread in the pattern effect and its contribution to climate sensitivity in CMIP5 and CMIP6 models. *Journal of Climate*, 33(18), 7755–7775. <https://doi.org/10.1175/JCLI-D-19-1011.1>
- Donohoe, A., Armour, K. C., Pendergrass, A. G., & Battisti, D. S. (2014). Shortwave and longwave radiative contributions to global warming under increasing CO<sub>2</sub>. *PNAS*, 111(47), 16700–16705. <https://doi.org/10.1073/pnas.1412190111>
- Drijfhout, S. (2015). Competition between global warming and an abrupt collapse of the AMOC in Earth's energy imbalance. *Scientific Reports*, 5(1), 14877. <https://doi.org/10.1038/srep14877>
- Drijfhout, S., van Oldenborgh, G. J., & Cimadoribus, A. (2012). Is a decline of AMOC causing the warming hole above the North Atlantic in observed and modeled warming patterns? *Journal of Climate*, 25(24), 8373–8379. <https://doi.org/10.1175/JCLI-D-12-00490.1>
- Etiman, M., Myhre, G., Highwood, E. J., & Shine, K. P. (2016). Radiative forcing of carbon dioxide, methane, and nitrous oxide: A significant revision of the methane radiative forcing. *Geophysical Research Letters*, 43(24). <https://doi.org/10.1002/2016GL071930>
- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model inter-comparison project phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, 9(5), 1937–1958. <https://doi.org/10.5194/gmd-9-1937-2016>
- Frolicher, T. L., Sarmiento, J. L., Paynter, D. J., Dunne, J. P., Krasting, J. P., & Winton, M. (2015). Dominance of the Southern Ocean in the anthropogenic carbon and heat uptake in CMIP5 models. *Journal of Climate*, 28(2), 862–886. <https://doi.org/10.1175/JCLI-D-14-00117.1>
- Good, P., Lowe, J. A., Andrews, T., Wiltshire, A., Chadwick, R., Ridley, J. K., et al. (2015). Nonlinear regional warming with increasing CO<sub>2</sub> concentrations. *Nature Climate Change*, 5(2), 138–142. <https://doi.org/10.1038/nclimate2498>
- Gregory, J. M., Andrews, T., & Good, P. (2015). The inconstancy of the transient climate response parameter under increasing CO<sub>2</sub>. *Royal Society*, 373(2054), 20140417. <https://doi.org/10.1098/rsta.2014.0417>
- 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>
- Grose, M. R., Gregory, J., Colman, R., & Andrews, T. (2018). What climate sensitivity index is most useful for projections? *Geophysical Research Letters*, 45(3), 1559–1566. <https://doi.org/10.1002/2017GL075742>
- Hu, A., Van Roekel, L., Weijer, W., Garuba, O. A., Cheng, W., & Nadiga, B. T. (2020). Role of AMOC in transient climate response to greenhouse gas forcing in two coupled models. *Journal of Climate*, 33(14), 5845–5859. <https://doi.org/10.1175/JCLI-D-19-1027.1>
- IPCC. (2021). *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.
- Jackson, L. C., Kahana, R., Graham, T., Ringer, M. A., Woollings, T., Mecking, J. V., & Wood, R. A. (2015). Global and European climate impacts of a slowdown of the AMOC in a high resolution GCM. *Climate Dynamics*, 45(11–12), 3299–3316. <https://doi.org/10.1007/s00382-015-2540-2>
- Kostov, Y., Armour, K. C., & Marshall, J. (2014). Impact of the Atlantic meridional overturning circulation on ocean heat storage and transient climate change. *Geophysical Research Letters*, 41(6), 2108–2116. <https://doi.org/10.1002/2013GL058998>
- Lin, Y.-J., Hwang, Y.-T., Ceppi, P., & Gregory, J. M. (2019). Uncertainty in the evolution of climate feedback traced to the strength of the Atlantic meridional overturning circulation. *Geophysical Research Letters*, 46(21), 12331–12339. <https://doi.org/10.1029/2019GL083084>
- Lin, Y.-J., Rose, B. E. J., & Hwang, Y.-T. (2023). Mean state AMOC affects AMOC weakening through subsurface warming in Labrador Sea. *Journal of Climate*, 36(12), 3895–3915. <https://doi.org/10.1175/JCLI-D-22-0464.1>
- Liu, W., Fedorov, A. V., Xie, S.-P., & Hu, S. (2020). Climate impacts of a weakened Atlantic meridional overturning circulation in a warming climate. *Science Advances*, 6(26). <https://doi.org/10.1126/sciadv.aaz4876>
- Mecking, J. V., & Drijfhout, S. S. (2023). The decrease in ocean heat transport in response to global warming. *Nature Climate Change*, 13(11), 1229–1236. <https://doi.org/10.1038/s41558-023-01829-8>
- Meinshausen, M., Nicholls, Z. R. J., Lewis, J., Gidden, M. J., Vogel, E., Freund, M., et al. (2020). The shared socio-economic pathway (SSP) greenhouse gas concentrations and their extensions to 2500. *Geoscientific Model Development*, 13(8), 3571–3605. <https://doi.org/10.5194/gmd-13-3571-2020>
- Meraner, K., Mauritsen, T., & Aiko, V. (2013). Robust increase in equilibrium climate sensitivity under global warming. *Geophysical Research Letters*, 40(22), 5944–5948. <https://doi.org/10.1002/2013GL058118>
- Met Office. (2021). *Mauna Loa carbon dioxide forecast for 2021*. UK Met Office. Retrieved from <https://www.metoffice.gov.uk/research/climate/seasonal-to-decadal/long-range/forecasts/co2-forecast-for-2021>
- Mitevski, I., Dong, Y., Polvani, L. M., Rugenstein, M., & Orbe, C. (2023). Non-monotonic feedback dependence under abrupt CO<sub>2</sub> forcing due to a North Atlantic pattern effect. *Geophysical Research Letters*, 50(14). <https://doi.org/10.1029/2023GL103617>
- Mitevski, I., Orbe, C., Chemke, R., Nazarenko, L., & Polvani, L. M. (2021). Non-Monotonic response of the climate system to abrupt CO<sub>2</sub> forcing. *Geophysical Research Letters*, 48(6). <https://doi.org/10.1029/2020GL090861>
- Mitevski, I., Polvani, L. M., & Orbe, C. (2022). Asymmetric warming/cooling response to CO<sub>2</sub> increase/decrease mainly due to non-logarithmic forcing, not feedbacks. *Geophysical Research Letters*, 49(5). <https://doi.org/10.1029/2021GL097133>
- Myhre, G., Highwood, E. J., Shine, K. P., & Stordal, F. (1998). New estimates of radiative forcing due to well mixed greenhouse gases. *Geophysical Research Letters*, 25(14), 2715–2718. <https://doi.org/10.1029/98GL01908>
- Poletti, A. N. (2023). CMIP6 doubling vs. Quadrupling data and analysis [Dataset]. Department of Atmospheric Science, University of Washington. <https://doi.org/10.5281/zenodo.7829225>
- Poletti, A. N., Frierson, D. M. W., Aueronson, T., Nikumbh, A., Carroll, R., Henshaw, W., & Scheff, J. (2024). Atmosphere and ocean energy transport in extreme warming scenarios. *PLOS Climate*, 3(2), e0000343. <https://doi.org/10.1371/journal.pclm.0000343>
- Ramaswamy, V., Collins, W., Haywood, J., Lean, J., Mahowald, N., Myhre, G., et al. (2018). Radiative forcing of climate: The historical evolution of the radiative forcing concept, the forcing agents and their quantification, and applications. *Journal of Climate*, 59(1), 14.1–14.101. <https://doi.org/10.1175/AMSMONOGRAPH-D-19-0001.1>

- Rose, B. E. J., Armour, K. C., Battisti, D. S., Feldl, N., & Koll, D. D. B. (2014). The dependence of transient climate sensitivity and radiative feedbacks on the spatial pattern of ocean heat uptake. *Geophysical Research Letters*, 41(3), 1071–1078. <https://doi.org/10.1002/2013GL058955>
- 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>
- Rugenstein, M. A. A., Winton, M., Stouffer, R. J., Griffies, S. M., & Hallberg, R. (2013). Northern high-latitude heat budget decomposition and transient warming. *Journal of Climate*, 26(2), 609–621. <https://doi.org/10.1175/JCLI-D-11-00695.1>
- Savitzky, A., & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(1964), 1627–1639. <https://doi.org/10.1021/ac60214a047>
- Sherwood, S. C., Webb, M. J., Annan, J. D., Armour, K. C., Forster, P. M., Hargreaves, J. C., et al. (2020). An assessment of Earth's climate sensitivity using multiple lines of evidence. *Reviews of Geophysics*, 58(4), e2019RG000678. <https://doi.org/10.1029/2019RG000678>
- Smith, C. J., Kramer, R. J., Myhre, G., Alterskjær, K., Collins, W., Sima, A., et al. (2020). Effective radiative forcing and adjustments in CMIP6 models. *Atmospheric Chemistry and Physics*, 20(16), 9591–9618. <https://doi.org/10.5194/acp-20-9591-2020>
- Taylor, K. E., Crucifix, M., Braconnot, P., Hewitt, C. D., Doutriaux, C., Broccoli, A. J., et al. (2007). Estimating shortwave radiative forcing and response in climate models. *Journal of Climate*, 20(11), 2530–2543. <https://doi.org/10.1175/JCLI4143.1>
- Trossman, D. S., Palter, J. B., Merlis, T. M., Huang, Y., & Xia, Y. (2016). Large-scale ocean circulation-cloud interactions reduce the pace of transient climate change. *Geophysical Research Letters*, 43(8), 3935–3943. <https://doi.org/10.1002/2016GL067931>
- Weaver, A. J., Sedlacek, J., Eby, M., Alexander, K., Cressin, E., Fichefet, T., et al. (2012). Stability of the Atlantic meridional overturning circulation: A model intercomparison. *Geophysical Research Letters*, 39(20). <https://doi.org/10.1029/2012GL053763>
- Winton, M., Griffies, S. M., Samuels, B. L., Sarmiento, J. L., & Frölicher, T. L. (2013). Connecting changing ocean circulation with changing climate. *Journal of Climate*, 26(7), 2268–2278. <https://doi.org/10.1175/JCLI-D-12-00296.1>
- Zhang, R., Kang, S. M., & Held, I. M. (2010). Sensitivity of climate change induced by the weakening of the Atlantic meridional overturning circulation to cloud feedback. *Journal of Climate*, 23(2), 378–389. <https://doi.org/10.1175/2009JCLI13118.1>
- Zhou, C., Zelinka, M. D., & Klein, S. A. (2017). Analyzing the dependence of global cloud feedback on the spatial pattern of sea surface temperature change with a Green's function approach. *Journal of Advances in Modeling Earth Systems*, 9(5), 2174–2189. <https://doi.org/10.1002/2017MS001096>

## References From the Supporting Information

- Andrews, T., Gregory, J. M., & Webb, M. J. (2015). The dependence of radiative forcing and feedback on evolving patterns of surface temperature change in climate models. *Journal of Climate*, 28(4), 1630–1648. <https://doi.org/10.1175/JCLI-D-14-00545.1>
- Leach, N. J., Jenkins, S., Nicholls, Z., Smith, C. J., Lynch, J., Cain, M., et al. (2020). FaIRv2.0.0: A generalized impulse response model for climate uncertainty and future scenario exploration. *Geoscientific Model Development*, 14(5), 3007–3036. <https://doi.org/10.5194/gmd-14-3007-2021>