On the Control of North-Hemispheric Feedbacks by AMOC, Evidence from CMIP and Slab-Ocean Modeling

Kai-Uwe Eiselt[a] and Rune Grand Graversen[a,b]

[a] *Department of Physics and Technology, University of Tromsø, Tromsø, Norway*

[b] *Norwegian Meteorological Institute, Tromsø, Norway*

Corresponding author: Kai-Uwe Eiselt, kai-uwe.eiselt@uit.no

ABSTRACT: The climate sensitivity of the Earth and the radiative climate feedback both change over time due to a so-called "pattern effect", i.e., changing patterns of surface warming. This is suggested by numerical climate model experiments. The Atlantic Meridional Overturning Circulation (AMOC) influences surface warming patterns as it redistributes energy latitudinally. Thus, this ocean circulation may play an important role for climate-feedback change over time. In this study, two groups of members of the Coupled Model Intercomparison Project (CMIP) phases 5 and 6 abrupt4xCO2 experiment are distinguished: one group showing weak and the other strong feedback change over time. It is found that both groups differ significantly in the AMOC response to 4xCO_2. Therefore, experiments with a slab-ocean model (SOM) with quadrupling of the CO_2 concentration are performed where the AMOC change is mimicked by changing the ocean heat transport. It is found that in the Northern Hemisphere extra-tropics the CMIP model group differences can be qualitatively reproduced by the SOM experiments, indicating that the AMOC plays an important role in setting the surface warming pattern. However, in the tropics and especially in the Southern Hemisphere other explanations are necessary.

1. Introduction

Equilibrium climate sensitivity (ECS), i.e., the magnitude of the warming of the Earth's climate system in response to a given forcing, is a widely used metric for describing global climate change. Another important parameter is the top-of-atmosphere (TOA) radiative response to a given warming, which is called the climate feedback and is also widely used. These are metrics intended to describe the change of the complex climate system by a single number and thus their ability for interpretation is highly limited. To gain more insight into the climate system's response to a forcing, and to better understand climate sensitivity and feedback, individual physical processes important for the climate are often investigated. The processes most typically considered are the change of surface albedo (SA), the change of water-vapour (WV) concentration, the change of the temperature lapse rate (LR), and the change of cloud properties. All these processes cause specific feedbacks that influence the magnitude of climate sensitivity. The SA feedback is mostly due to the melting of sea ice and a reduction of snow cover which decreases the surface albedo of the Earth and thus causes a positive feedback, raising the climate sensitivity (e.g., Hall 2004; Winton 2006; Graversen and Wang 2009). The WV feedback is also positive and arises due to the fact that at higher temperature the atmosphere can hold more water vapour which inhibits more thermal radiation from escaping to space and thus enhances the warming (e.g., Held and Soden 2000; Manabe and Wetherald 1967). The LR feedback arises due to different warming at different altitudes in the atmosphere. If an atmospheric column warms more aloft than the surface below, this increases the Earth's cooling efficiency and thus constitutes a negative feedback. In contrast, a positive feedback results if the lower atmospheric layers warm more than aloft. Due to strong convection in the tropics the LR feedback is often negative there, and due to stable atmospheric stratification at high latitudes it is positive in those regions (Manabe and Wetherald 1975; Graversen et al. 2014). Its global average is thought to be negative (e.g., Soden and Held 2006). Finally, the nature of the cloud feedback is manifold and it remains unclear whether it is positive or negative (e.g., Zelinka et al. 2016; Bjordal et al. 2020; Mülmenstädt et al. 2021), although a recent extensive review of the literature argues for a moderately positive feedback (Sherwood et al. 2020). Even based on the latest generation of global climate models, the cloud feedback remains the mechanism associated with the largest inter-model variance (Zelinka et al. 2020). To understand and constrain cloud feedback, typically it is subdivided into different categories based on different cloud regimes and physical processes. Two types of cloud feedback are robustly assessed to be positive in the review by Sherwood et al. (2020): the high-cloud altitude and the tropical marine low-cloud feedback. The only negative cloud feedback component according to Sherwood et al. (2020) is the tropical anvil cloud area feedback, although its magnitude is highly uncertain and may be zero. Based on climate models, magnitudes of the different cloud feedback types generally fall within the uncertainty range of the expert assessment in Sherwood et al. (2020), although a large variance across models as well as outliers remain (Zelinka et al. 2022). Indeed, Zelinka et al. (2022) find that increased skill in simulating mean-state cloud properties does not lead to the cloud feedback being in better agreement with expert judgement.

In spite of ever increasing research efforts it remains difficult to constrain the estimated values of climate feedback and ECS (Arrhenius 1896; Charney et al. 1979; Sherwood et al. 2020). Typically, the ECS is derived by running a numerical climate model experiment starting in a quasi-equilibrium (pre-industrial) state but with instantaneous doubling or quadrupling of the CO_2 concentration. Running a fully-coupled Earth-System Model (ESM) to a new equilibrium takes thousands of simulation years and thus requires immense computational resources (e.g., Paynter et al. 2018; Rugenstein et al. 2020). Hence, such experiments are usually only run for a few hundred years, and the relationship between surface-air temperature (SAT) and TOA radiative-flux imbalance is extrapolated to a new equilibrium state, which provides an estimate for ECS, typically referred to as *effective* climate sensitivity (e.g., Sherwood et al. 2020). This procedure is known as the Gregory method (Gregory et al. 2004). It was already noted by Gregory et al. (2004) that the relationship between SAT and TOA imbalance appears non-linear and thus estimates derived with the Gregory method from short (i.e., $\ll 1000$ years) model simulations may diverge from the ECS as derived from running the model to equilibration. The time-dependence of the effective climate sensitivity has since become a topic of much research attention (see e.g., Eiselt and Graversen 2022, for a brief literature overview).

Recently, the so-called *pattern effect* (Stevens et al. 2016) has emerged as the most prominent explanation for the time-dependence of climate sensitivity. That is, sea-surface temperature (SST) warming patterns in response to a forcing of the climate system change over time to favour regions of different cooling efficiency causing the global cooling efficiency and thus climate sensitivity to change over time (Andrews and Webb 2018; Ceppi and Gregory 2017; Dong et al. 2019, 2020).

However, it is still unclear whether the pattern effect is consistent in terms of a general pattern across climate models (i.e., do models show similar spatial patterns of warming, and similar warming-pattern changes?). It is also unclear how a given surface-warming pattern is triggered due to a forcing. Dong et al. (2019, 2020) point to the Indo-Pacfic Warm Pool (IPWP) as one region that appears particularly important for the pattern effect. The IPWP is characterised by strong convection and weak stratification implying that the region cools efficiently. Using a Green's function approach, Dong et al. (2020) predict the change of climate feedback over time across members of the Coupled Model Intercomparison Project (CMIP) phases 5 and 6 with some skill. In support of these findings, Eiselt and Graversen (2022) show that the change of the surface warming of the IPWP relative to the change of the global mean warming is robustly negatively correlated with the change over time of climate feedback across CMIP5 and CMIP6 members. However, some uncertainties remain and, as Dong et al. (2020) note, the reconstruction of the change of feedback over time is less reliable for CMIP6 members than for those of CMIP5.

A further contribution to the pattern effect may emerge from a differential polar (especially Arctic) and extra-polar warming development (Lin et al. 2019; Bellomo et al. 2021; Mitevski et al. 2021; Eiselt and Graversen 2022). In the climatology, as less solar radiation per unit area hits the surface at high latitudes than at low latitudes, an excess of energy in the tropics and a deficit of energy at the poles are induced, which is compensated by a poleward energy transport. A considerable part of this energy transport is accomplished by the ocean, and a major part of this ocean heat transport in the Northern Hemisphere is due to the Atlantic Meridional Overturning Circulation (AMOC; e.g., Buckley and Marshall 2016). The AMOC is thought to originate from the formation of dense surface water in the Arctic that sinks causing warm surface water from the tropics to move northward. As the climate system warms in response to a forcing, the latitudinal energy transport may change and thus influence the surface-warming difference between the Arctic and the tropics.

The Arctic atmosphere exhibits a strongly stable stratification and thus a low cooling efficiency as compared to lower latitudes, since a warming at the surface is not easily spread to higher altitudes and thus "trapped" close to the surface. Hence, a relatively stronger warming over time in the Arctic than over the rest of the globe would over time lead to a less negative climate feedback and higher climate sensitivity. Eiselt and Graversen (2022) find evidence for this effect, since the change of climate feedback over time correlates positively with the change over time of warming in the Arctic relative to the global mean warming.

Because of its transport of warm water poleward, the AMOC exhibits to some degree control over the surface warming in the Arctic and especially in the North Atlantic. Members of CMIP5 and 6 show a large spread in both the pre-industrial control (piControl) AMOC and in the response of the AMOC to an abrupt quadrupling of the CO_2 concentration (abrupt4xCO2; Lin et al. 2019; Bellomo et al. 2021). Indeed, Bellomo et al. (2021) find that models with a large AMOC decline in response to abrupt CO_2 quadrupling exhibit a distinct lack of warming in the North Atlantic (50-70°N, 80°W-10°E; known as the North Atlantic Warming Hole), while this is not the case for models with small AMOC decline. Lin et al. (2019) find that models that quickly slow down their AMOC in response to abrupt CO_2 quadrupling start recovering the AMOC in later years, while models with a moderate initial AMOC slowdown show little or no recovery. The former group of models (called "high" in Lin et al. 2019) shows a shift in warming from low latitudes to the Arctic over time, and these models weaken their climate feedback more over time than does the latter group of models (called "low" in Lin et al. 2019), which is consistent with the aspects regrading Arctic atmospheric stability mentioned above. As changes of atmospheric stability directly affect the atmospheric temperature lapse rate (Ceppi and Gregory 2017; Andrews and Webb 2018), Eiselt and Graversen (2022) investigated the change of the LR feedback over time and its impact on the total feedback in the CMIP5 and 6 abrupt4xCO2 experiments. It was found that the change of these feedbacks strongly correlates across models. Hence, it was found that the LR feedback change mostly dominates the total feedback change, except for a few models for which the cloud feedback change appears more important. To further analyse the causes of differences in feedback change between models, two model groups were extracted based on the change of LR feedback over time (G1: weak LR feedback change, G2: strong LR feedback change) and compared in terms of surface warming and individual climate feedbacks. Further investigation reveals that the development of the AMOC in G1 and G2 corresponds remarkably well to the development of the groups named "low" and "high", respectively, in Lin et al. (2019) (see especially their Fig. 1), although their division was based on AMOC strength, and the groups are comprised of different members (except for one member, NorESM1-M).

In this study we extend the analysis of Eiselt and Graversen (2022) and show that certain differences between G2 and G1 may be related to their difference in AMOC development. However, in fully-coupled climate model experiments it is difficult to establish causality. Thus, we employ a slab-ocean model (SOM) to mimic the AMOC-related changes in ocean heat transport and investigate their effects. In a SOM, the ocean-model component is simplified to a mixed-layer slab where the energy balance is computed based on a lateral and steady ocean heat flux (called Q-flux), including an annual cycle, and heat exchange with the atmosphere in form of radiation and sensible and latent turbulent fluxes. A change in the AMOC can be mimicked in a SOM by changing the Q-flux. We institute such a Q-flux change roughly corresponding to the difference in AMOC between G2 and G1 (see appendix A) while also abruptly raising the CO_2 concentration by a factor of four (as in the CMIP abrupt4xCO2 experiments). Then we investigate the response in terms of SAT as well as TOA radiative fluxes due to individual climate feedbacks and qualitatively compare to the difference in response between G2 and G1. We note that since in a SOM the ocean is inactive, with our experiments we cannot establish how changes in the climate system induced by the AMOC change (e.g., sea-ice melt or surface and atmospheric warming) may feed back on and alter the AMOC itself (Liu et al. 2019; Todd et al. 2020; Dai 2022).

Singh et al. (2022) perform similar experiments to those presented here, by changing the meridional ocean heat transport in a slab-ocean model while increasing the atmospheric CO_2 concentration. In contrast to our experiments they change the *zonally integrated* heat transport, thus neglecting possible pattern effects arising from regional Q-flux changes. The general significance of the pattern effect has been elaborated above. Additionally, and more specifically with respect to Q-flux change, Lin et al. (2021) show that the regional location of the Q-flux change is highly important. Hence, Singh et al. (2022) do not find a North Atlantic Warming Hole in their experiments with reduced ocean heat transport, which, however, is prevalent in fully-coupled model experiments that exhibit AMOC decline (Bellomo et al. 2021), and which is observed in our experiments (not shown). This indicates that important *regional* climate changes and impacts may be ignored if the regional pattern of the heat transport change is neglected. Furthermore, Singh et al. (2022) do not consider a recovery of the ocean heat transport after its initial decline, which is observed in some of the coupled models and which may affect the change of climate feedback over time (e.g., Lin et al. 2019). In the present study, additional experiments are performed where AMOC recovery is mimicked and compared to those without AMOC recovery. Finally, Singh et al. (2022) use an older version (i.e., CESM1) of the same model used in the present study (i.e., CESM2) that underwent substantial changes (e.g., in the cloud parametrisation) and exhibits considerable differences in terms of feedbacks and climate sensitivity (Gettleman et al. 2019).

The paper is structured as follows: Section 2 gives an overview of the CMIP data as well as the SOM experiment design and section 3 describes briefly the radiative kernel method employed to derive radiative fluxes. In section 4 we first describe the differences between the CMIP model groups G2 and G1 in terms of AMOC and then proceed to present and dicuss the results of our SOM experiments and compare them with CMIP model groups G1 and G2. In section 5, some further discussion and concluding remarks are offered.

2. Models and experiments

a. CMIP experiments

The CMIP data used in this study is taken from the abrupt4xCO2 and piControl experiments from the CMIP5 (Taylor et al. 2009) and CMIP6 (Eyring et al. 2016) archives, and the members from both archives are treated equally. Consistent with previous studies, anomalies were calculated as abrupt4xCO2 minus a 21-year running mean over the piControl for all variables used (e.g., Caldwell et al. 2016; Zelinka et al. 2020). Two groups of models were extracted from these experiments, one with small (G1) and one with large (G2) lapse-rate feedback change over time, where change over time is defined as the feedback over the years 21-150 of the simulation minus that over the years 1-20 (see also section 1). A further motivation and description of the methodology employed in choosing the members of G1 and G2 can be found in Eiselt and Graversen (2022) and the group members are listed in Tables S4 and S5, respectively, in their online supplemental material.

b. CESM2-SOM and experiment design

In order to mimic the AMOC change, the Community Earth System Model Version 2.1.3 (CESM2; Danabasoglu et al. 2020) is employed in the slab-ocean configuration (hereafter called CESM2-SOM). In this set-up, the Community Atmosphere Model Version 6 (CAM6; Danabasoglu et al. 2020) on a ~2° resolution, the Community Land Model Version 5 (CLM5; Lawrence et al. 2019), the Model for Scale Adaptive River Transport (MOSART; Li et al. 2013), and the Los

Alamos Sea Ice Model Version 5.1.2 (CICE5; Hunke et al. 2015) are coupled to a dynamically inactive ocean component (the slab ocean). The ocean component consists of an isothermal mixed layer with prescribed horizontal energy transport in the form of the so-called Q-flux. The Q-flux is derived using the climatology of mixed-layer depth, SSTs, and ocean heat uptake (Bitz et al. 2012) over the last 80 years of a 100-year fully-coupled CESM2 pre-industrial-state control simulation using the Parallel Ocean Programme Version 2 (POP2; Smith et al. 2010; Danabasoglu et al. 2020) as the dynamically active ocean component. This 100-year simulation was continued at our server from a 300-year pre-industrial control simulation conducted at the National Center for Atmospheric Research (NCAR) servers which is publicly available from the NCAR website.

Since in a SOM the Q-flux is prescribed, there is no dynamical ocean response and thus no AMOC change due to a greenhouse-gas forcing. However, the effect of a change in AMOC can be mimicked in a SOM by changing the Q-fluxes, thus simulating a change in ocean heat transport. Hereby the AMOC change can be separately investigated. As the AMOC transports warm water from the tropics into the North Atlantic, a decline in AMOC leads to less ocean heat uptake (more ocean heat release) in the tropics and more ocean heat uptake (less ocean heat release) in the North Atlantic. The strong influence of the AMOC on the North Atlantic surface heat flux is demonstrated e.g. in the fresh-water hosing experiments of Jackson et al. (2015) and the flux-anomaly experiments of Todd et al. (2020). A change in ocean heat transport associated with the AMOC change can be implemented in a SOM by decreasing the Q-flux in the tropical Atlantic, and increasing it in the northern Atlantic. The regions that are most affected by the AMOC are different across the CMIP models. Hence, we perform several experiments with different northern boundaries of AMOC-change impact region, with northern boundaries of 70, 75, and 80°N, respectively. However, while the impact on sea ice, temperature, and radiative fluxes is somewhat larger if the boundary is situated further north, this does not qualitatively change the results and conclusions. The results shown here are for the impact region being 50-80°N and 75°W-25°E. The boundaries chosen for the region in the tropical Atlantic are 15°S-15°N and 75°W-25°E. In all experiments the Q-flux in the tropics is chosen to exactly compensate the Q-flux change in the North Atlantic region, implying that the global-mean Q-flux change is zero and no net forcing is introduced.

To determine appropriate values for the Q-flux change, we employ an order-of-magnitude estimation of the energy-transport change associated with the AMOC-change difference between G2 and G1 (appendix A). To test the dependence of the results on the magnitude of the Q-flux change, several simulations with different magnitudes are performed. The global as well as the regional surface-temperature response increases non-linearly with the Q-flux change (see Fig. S1 in the online supplemental material). This may be due to non-linear aspects of the feedback changes triggered by the Q-flux change. For instance, the Q-flux changes non-linearly impact the Arctic sea-ice melt (see section 4 and Fig. S1). The non-linear nature of the response to the Q-flux change complicates a quantitative analysis and we thus emphasise again that our comparison of CESM2-SOM experiments with the fully-coupled models is qualitative.

To mimic that the AMOC does not instantly assume its final value in response to the abrupt CO_2 forcing, the Q-flux is changed linearly over the course of 12 months (the Q-flux is changed per month) and then held constant. Further experiments were performed where the Q-flux is changed linearly over the course of 60 and 120 months but the results appear qualitatively independent of the timing of AMOC decline and thus are not shown. Finally, an experiment was performed where the Q-flux is changed linearly over the first 12 months and then the change is reversed, although more slowly, with the Q-flux obtaining its original value in year 30, i.e., mimicking full AMOC recovery.

The experiments with mimicked AMOC change are in the following referred to as *dQ*. In all these experiments the atmospheric CO_2 concentration is abruptly quadrupled to 1138.8 ppm. To investigate the impact of the mimicked AMOC change a control simulation is performed where only the CO_2 concentration is abruptly quadrupled but no Q-flux change is implemented (hereafter called *no-dQ*). To account for the effect of internal variability, three ensembles of both dQ and no-dQ experiments were run, starting from different years in the pre-industrial-state control simulation.

3. Methods

a. Radiative kernel method

To estimate the radiative fluxes due to individual climate feedback processes we employ the radiative kernel method (Soden et al. 2008; Shell et al. 2008). In this method it is assumed that the total TOA radiative flux change (N) can by partitioned into contributions from independent climate variables and that the TOA radiative flux change in response to a small change of a climate variable is linear. The radiative kernels for a specific climate variable are generated by perturbing this variable by a given amount in a climate model and then executing only the radiation code. As an example the SA kernel can thus be derived as

$$N(a+\delta a,T,w,c)-N(a,T,w,c)=N(\delta a)=\frac{\partial N}{\partial a}(a,T,w,c)\delta a\equiv K^{a}\delta a, \tag{1}$$

where a is the SA, δa the SA perturbation (typically 1%), T the temperature, w the WV mixing ratio, c a set of cloud properties, and K^a is the SA kernel. In a climate model experiment, the TOA radiative flux change due to a given climate feedback process is obtained by multiplying the associated kernel with the climate variable in question. Radiative kernels derived from one climate model are assumed to be applicable across climate models as radiative transfer schemes across climate models are well tested and fairly similar (Soden et al. 2008). Due to the strongly non-linear effects of clouds on radiation, standard cloud radiative kernels are inappropriate. However, the cloud radiative effect, calculated from model output all-sky minus clear-sky fluxes, can be adjusted by the cloud masking of the other feedbacks to obtain the cloud feedback (Soden et al. 2008).

In the present study the radiative kernels provided by Shell et al. (2008) are used to calculate TOA radiative fluxes for both CMIP and CESM2-SOM simulations (for more details on the choice of radiative kernels see Eiselt and Graversen 2022). For consistency, all radiative fluxes are positive downward.

b. AMOC index, sea-ice area, and SEB

The AMOC index is calculated as the maximum of the meridional overturning stream function (based on the variables named *msftmyz* or *msftyyz* in CMIP5 and *msftmz* or *msftyz* in CMIP6) north of 30°N in the Atlantic basin below 500 m depth. Note that the necessary model output was not available for all G1 and G2 members (Tables S1 and S2 in the online supplemental material).

The sea-ice area is calculated by multiplying the sea-ice concentration with the ocean-grid-cell area and summing separately over the northern and Southern Hemispheres.

The surface energy balance (SEB) is calculated from the model output surface fluxes, including net surface long-wave and short-wave fluxes, surface latent heat flux, and surface sensible heat flux.

4. Results

We now first describe the differences between G2 and G1 with respect to AMOC in their piControl state and in their response to an abrupt quadrupling of the CO_2 concentration. We hypothesise that the difference between the two model groups when it comes to the response in SEB, temperature, and atmospheric stability in the mid-latitude North Atlantic (MLNA; 40-60°N, 10-60°W) is strongly influenced by the AMOC (e.g., Yeager et al. 2012, 2015; Jackson et al. 2015; Todd et al. 2020). However, it should be noted that the change of the SEB in fully-coupled atmosphere-ocean models may not only depend on the AMOC but is influenced by atmospheric changes as well (Todd et al. 2020). We conclude the section with a discussion of the differences in global and regional mean SAT and TOA radiative fluxes between G2 and G1 and we qualitatively compare these differences with those between our CESM2-SOM experiments without a Q-flux change (no-dQ) and with a Q-flux change applied (dQ; see section 2b).

a. Differences in AMOC and mid-latitude North Atlantic SEB

The piControl state AMOC is considerably stronger in G2 than in G1 (on average by 8.77 Sv; the difference is 53 % of the G1 mean and 35 % of the G2 mean), indicating that there is a larger equator-to-pole energy transport by the AMOC in G2 than in G1 (Fig. 1a). Consistent with the hypothesis stated above, there is more piControl ocean heat release (OHR; negative SEB) in the MLNA region in G2 than in G1 (Fig. 2a). In response to the abrupt 4xCO_2 forcing, the AMOC and OHR in the MLNA region decline in both groups, but this happens quicker in G2 than in G1 (see Figs. 1b-d as well as 2b-d). In G2, AMOC and OHR rapidly decline for about 15 years after the CO_2 forcing and then exhibit a slower decline until about year 50, after which they start a slight recovery that continues over the remainder of the simulation (Figs. 1b and 2b). Conversely, AMOC and OHR in G1 decline more slowly than in G2 over the first about 50 years and then remain constant for the rest of the simulation. After around year 25, the OHR in G1 even becomes negative. Notably, the total *change* of the AMOC at around year 50 is somewhat larger in G2 than in G1 (-12.11 Sv and -9.67 Sv, respectively), but due to the recovery in G2, it is smaller in G2 than in G1 at the end of the simulation (Fig. 1cd). A similar development obtains for the OHR change (Fig. 2cd). We conclude that the development of the SEB in G1 and G2 in the MLNA region is strongly influenced by the AMOC. For a summary of the values of piControl average AMOC, as well as AMOC early (years 1-15) and late (years 51-150) trends for the individual members of G1 and G2 see Table S1 and S2, respectively.

We note that while the model groups G1 and G2 were chosen based on their change in lapse-rate feedback over time (Eiselt and Graversen 2022), they exhibit remarkably similar AMOC development as the model groups “low” and “high”, respectively, defined by Lin et al. (2019) based on the magnitude of the AMOC decline in CMIP5 abrupt4xCO2 experiments (compare their Fig. 1 to our Fig. 1). Consistently, Lin et al. (2019) find that the models with weaker AMOC (“low”) exhibit less (especially lapse-rate) feedback change over time than the models with stronger AMOC (“high”).

b. Temperature and atmospheric stability in the mid-latitude North Atlantic

The model groups G1 and G2 exhibit distinct differences in their development of temperature and lower tropospheric stability (LTS, Klein and Hartmann 1993) over the MLNA region in the abrupt4xCO2 experiment (Fig. 3). Surface temperature in the MLNA in the piControl simulation is significantly higher in G2 than in G1 (Fig. S3a). This holds for both SST and SAT, but we focus here on SAT as this variable is normally used in the calculation of LTS. The temperature at the 700 hPa level is also higher in G2 than in G1, but the difference at this level is not significant ($p > 0.05$; Fig. S3d). Accordingly, the LTS is higher in G1 than in G2, but not significantly ($p > 0.05$; Fig. 3a). These differences between G2 and G1 are consistent with those in AMOC. As described in section 4a, G2 has a significantly stronger piControl AMOC than G1, which explains the significantly higher G2 surface temperatures in the MLNA region as a stronger AMOC advects more warm water to this region from the south. The temperature at 700 hPa is partly driven by atmospheric advection and hence non-local factors and is thus less sensitive to the AMOC-induced surface warming. Hence, it is also in agreement with the AMOC impact that the difference between G2 and G1 of the 700 hPa temperature is smaller than that of the SAT.

In response to the abrupt 4xCO_2 forcing, the SAT in the MLNA region quickly increases in both G1 and G2 (Fig. S3bc). However, in G2 the SAT plateaus after a few years, while it increases further in G1. Subsequently, in G2 the increase of the SAT resumes so that in the later years the SAT exhibits a stronger trend in G2 than in G1. This is consistent with the explanation due to AMOC: In G2 the AMOC initially slows down more than in G1 and thus imposes a stronger

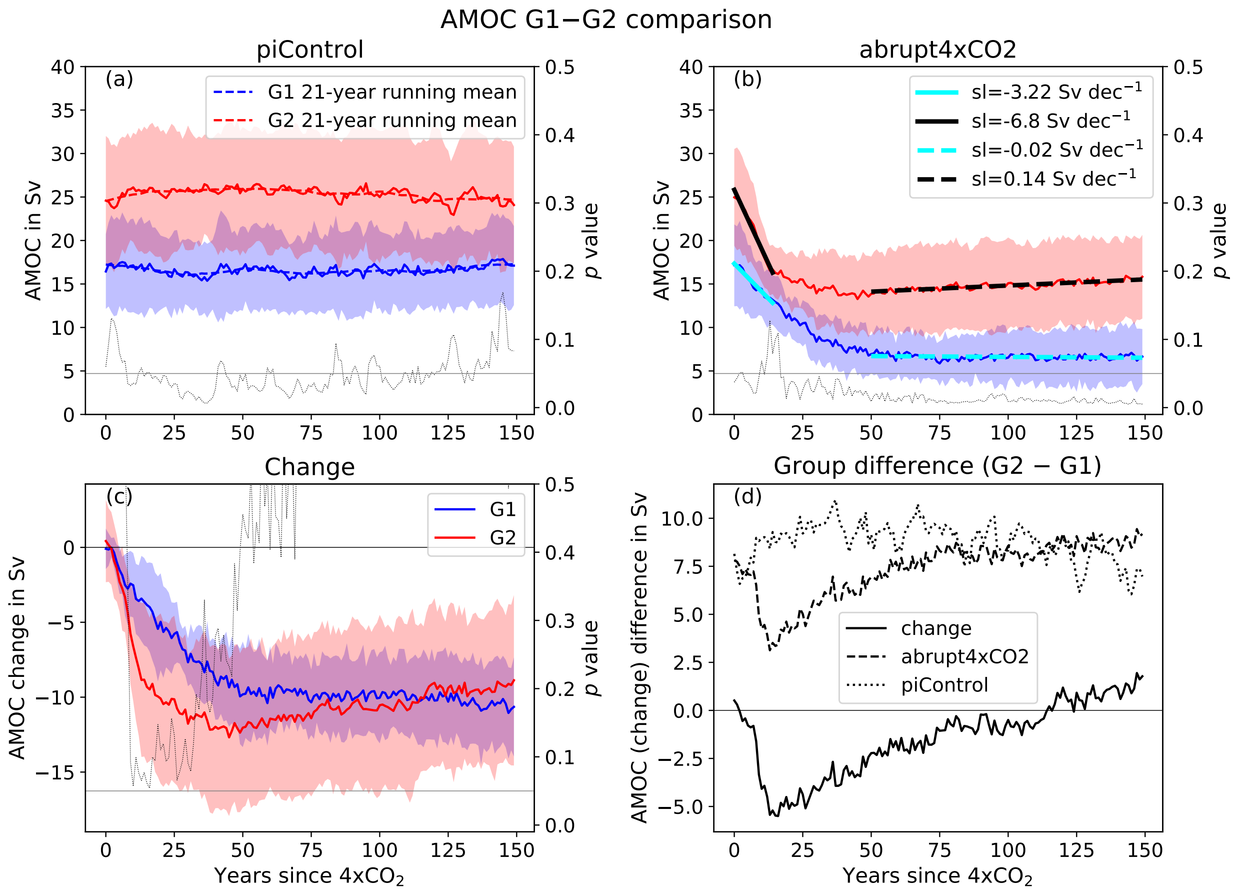

FIG. 1. Atlantic Meridional Overturning Circulation (AMOC) index for G1 and G2 in piControl (a; both annual and 21-year running mean; the period corresponds to the abrupt4xCO2 simulation), in abrupt4xCO2 (b), and the difference between abrupt4xCO2 and a 21-year running mean of the piControl simulation (c), and the G2-minus-G1 difference (d). The blue and red solid lines show the G1 and G2 means, respectively, and the shading indicates the ± 1-sigma spread of models. In panels a-c the thin, black, dotted line shows the *p* value of a two-sided Welch's *t*-test for the difference in group mean and the gray horizontal line indicates a *p* value threshold of 0.05. Panel b includes linear trends for the early (years 1-15; solid) and late (years 51-150, dashed) periods. For details of the AMOC index calculation see section 3b.

surface cooling influence in the MLNA region. However, as the AMOC starts recovering in G2 and not in G1 in the later simulation years, it accelerates the local SAT warming in G2 only. In contrast, for the 700 hPa temperature no plateauing is observed in G2 (Fig. 3ef). Thus, the 700 hPa temperature is more similar in the two model groups than is the SAT, although G1 again exhibits

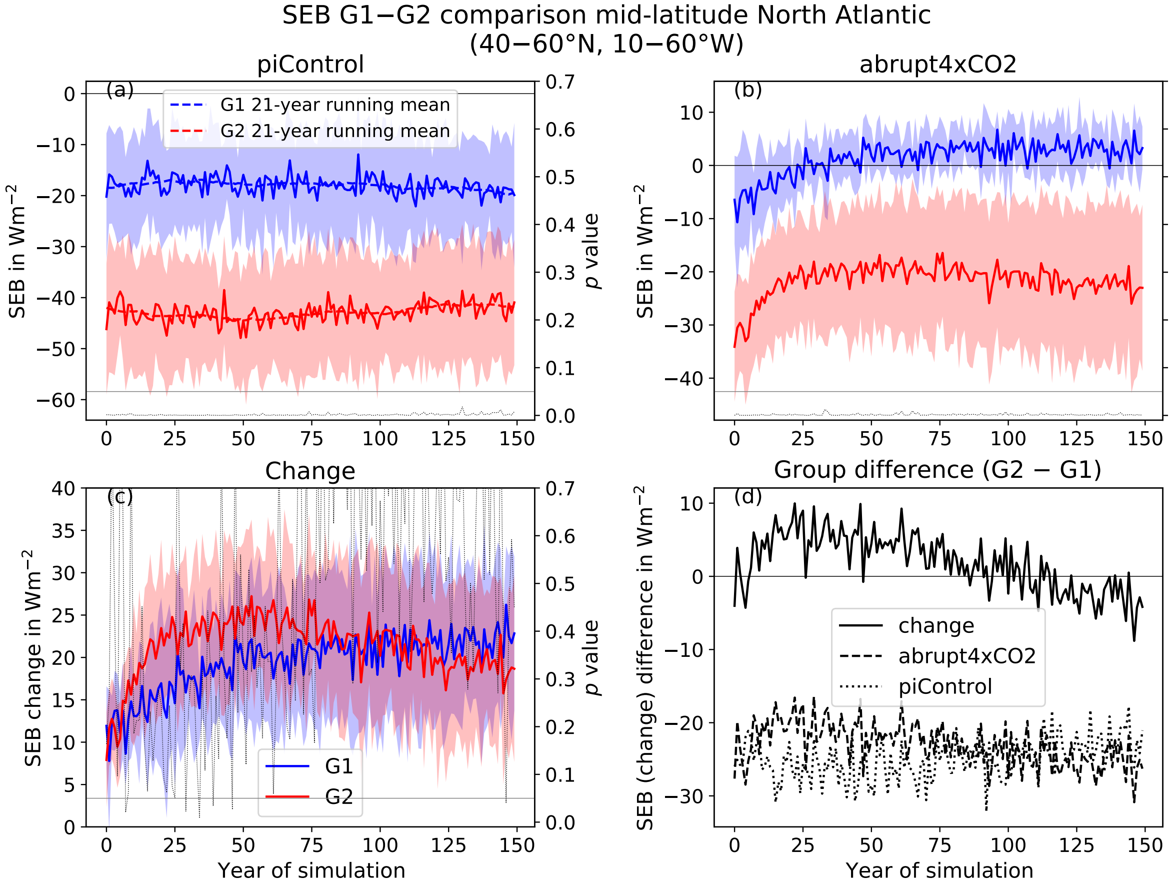

FIG. 2. As Fig. 1 but for surface energy balance (SEB; units are positive down) averaged over the mid-latitude North Atlantic (40-60°N, 10-60°W). Note that the members of G1 and G2 for which AMOC was not available were *not* excluded here; however, excluding these members does not qualitatively impact the results (see Fig. S2 in the online supplemental material).

more warming. Consistent with the discussion above, this indicates that the 700 hPa temperature is partly driven by non-local factors and to a lesser extent by the AMOC.

The change in LTS (Fig. 3bc) depends on the difference in changes in temperature at the two levels. Due to the plateauing of the SAT that is only observed in G2, the LTS initially increases more in G2 than in G1. However in the later years of the simulation, due to the stronger increase of the SAT in G2 than in G1, the LTS increase stagnates in G2 while it continues in G1. The differences in the changes in atmospheric stability in the MLNA region thus seem to be mostly determined by differences in surface changes, and the AMOC appears to be a driving factor. Indeed, the differences in LTS in this region are qualitatively similar to the global mean LTS differences (Eiselt and Graversen 2022, especially their Fig. 6e). This is consistent with the findings of Lin et al. (2019) reporting that the change over time of atmospheric stability correlates with the magnitude of AMOC decline, and this points to the importance of the AMOC for global atmospheric stability and hence the lapse-rate feedback and the Earth's cooling efficiency (Lin et al. 2019; Ceppi and Gregory 2017; Eiselt and Graversen 2022).

Furthermore, these results are consistent with Bellomo et al. (2021) who divide the CMIP abrupt4xCO2 simulations into two groups according to the magnitude of AMOC decline. Congruent with the differences between G2 and G1 presented here and in Eiselt and Graversen (2022), Bellomo et al. (2021) find that models with strong AMOC decline exhibit reduced warming, especially in the North Atlantic (the North Atlantic Warming Hole), compared to those with weak AMOC decline.

We note that Singh et al. (2022) perform similar experiments to the ones presented here employing CESM1 in the SOM configuration. They do not find a North Atlantic Warming Hole in their experiments (see especially their Fig. 6). However, they change the *zonally averaged* ocean heat transport. That is, they do not prescribe a distinct Q-flux pattern but change the Q-flux zonally uniformly. This may indicate the importance of the *pattern* of ocean heat transport for SST patterns and thus the global climate feedback (see also Lin et al. 2019; Rugenstein et al. 2016; Mitevski et al. 2021; Lin et al. 2021).

c. Regional SAT and TOA radiative fluxes

In the following the differences between G2 and G1 in global and regional SATs as well as in radiative fluxes are discussed. The CMIP results are compared to CESM2-SOM experiments which have been designed to mimic AMOC change (see section 2b). As the implementation of the AMOC change in the CESM2-SOM experiments effectively constitutes a redistribution of energy across latitudes, we show zonal means over different latitude ranges. Furthermore, as the simulated AMOC change mostly affects the Northern Hemisphere, both hemispheres are discussed separately. Specifically, the regions (R) considered are R>75°, R>60°, R>30°, R<30° (i.e., 0-30°), and R>0°, where in the following ° is replaced by N in case of the Northern Hemisphere (NH) and by S in case of the Southern Hemisphere (SH).

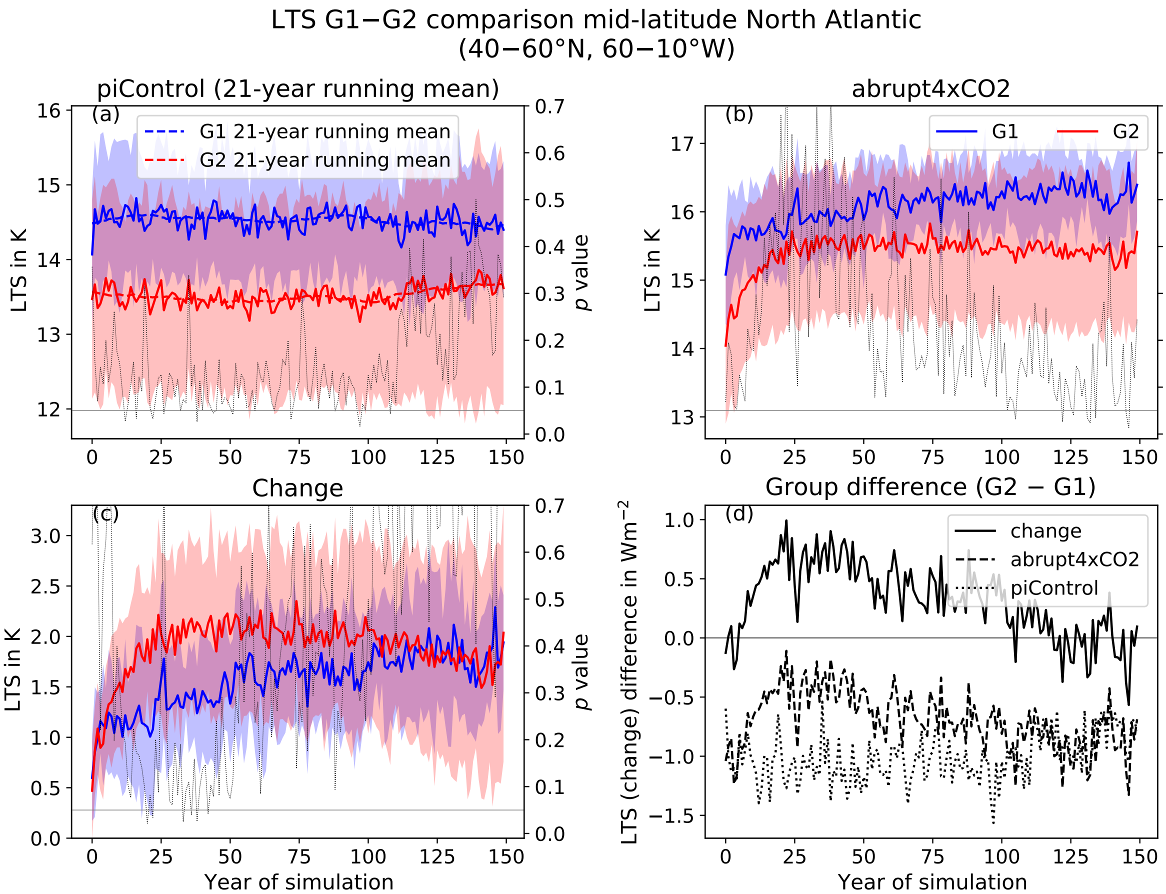

FIG. 3. As Fig. 1 but for lower tropospheric stability (LTS) averaged over the mid-latitude North Atlantic (40-60°N, 10-60°W). The LTS was calculated as the difference between the 700 hPa and surface potential temperature (see Fig. S3 for similar plots of the surface and 700 hPa temperature).

1) SURFACE-AIR TEMPERATURE

Figure 4ab shows the G2-minus-G1 difference in SAT for the above-defined regions as well as the entire Earth. The global mean initially becomes increasingly negative and stays constant after about 50 years of simulation. While differences in R>0N and R>0S are similar, the differences are larger further poleward with the most extreme differences in the Arctic (R>75N). It is remarkable that in the regions of the NH (Fig. 4a) the differences between G2 and G1 initially strongly increase in magnitude (especially in R>75N and R>60N), but after about year 40, they start decreasing and continue to do so throughout the remainder of the 150-year simulation. In the SH (Fig. 4b), the development is different since the initial increase of the difference is slower, but continues throughout the whole simulation. These differences in development appear strongly affected by the differences in sea-ice development (e.g., Eiselt and Graversen 2022; Graversen et al. 2014; Dai et al. 2019; Jenkins and Dai 2021), which in turn may be affected by the development of the AMOC. That is, the fast weakening of the AMOC in G2 described in section 4a decelerates the NH warming in response to the CO_2 forcing and thus inhibits Arctic sea-ice loss. Conversely, the slower AMOC weakening in G1 causes Arctic sea ice to melt faster than in G2. Note that Arctic sea-ice loss as a feedback loop likely affects the AMOC (Levermann et al. 2007; Liu et al. 2019), but this causality cannot be investigated with a SOM. Melting sea ice exposes the underlying warmer ocean and thus enables the release of heat into the atmosphere which increases SAT. Furthermore, sea-ice loss decreases the albedo and hence leads to more absorbed solar radiation, increasing SAT further. Finally, the stable stratification in the Arctic confines the warming to the surface and thus inhibits efficient cooling, again contributing to an increase in SAT. In G1 the Arctic sea ice initially declines much faster than in G2 (Fig. 5a), consistent with the strong negative initial G2-minus-G1 difference in SAT. After some time, most of the Arctic sea ice in G1 has melted and hence the sea-ice decline slows down, accompanied by a smaller pace of the SAT rise in R>75N. Conversely in G2, more sea ice remains, which continues to melt in concert with the steady rise of SAT in R>75N. This causes the difference in SAT between G2 and G1 to decline in the later years of the CMIP simulations (especially in R>75N and R>60N). In Antarctica, the sea ice initially also declines faster in G1 than in G2, but it declines generally slower than in the Arctic and the rate of decline is more similar in G1 and G2 (Fig. 5b). Consistently, the SAT differences in R>75S and R>60S between G2 and G1 are generally smaller (except in the last simulation years) than in R>75N and R>60N, respectively, and they remain almost constant after about year 50 in the SH.

Figure 4cd presents the SAT differences between the dQ and no-dQ CESM2-SOM experiments averaged over the above-defined regions. In general, they appear qualitatively similar to those based on the CMIP groups, especially in the global mean and in the NH. That is, the further north, the larger the magnitude of the difference between the dQ and the no-dQ experiment, and the differences in R>0N are distinctly larger than in R>0S. Notably however, while the difference between G2 and G1 in R>75N and R>60N starts declining in magnitude after the initial increase, the difference between dQ and no-dQ in the same regions remains almost constant. This indicates that the changes implemented in CESM2-SOM to mimic differences in AMOC change are not

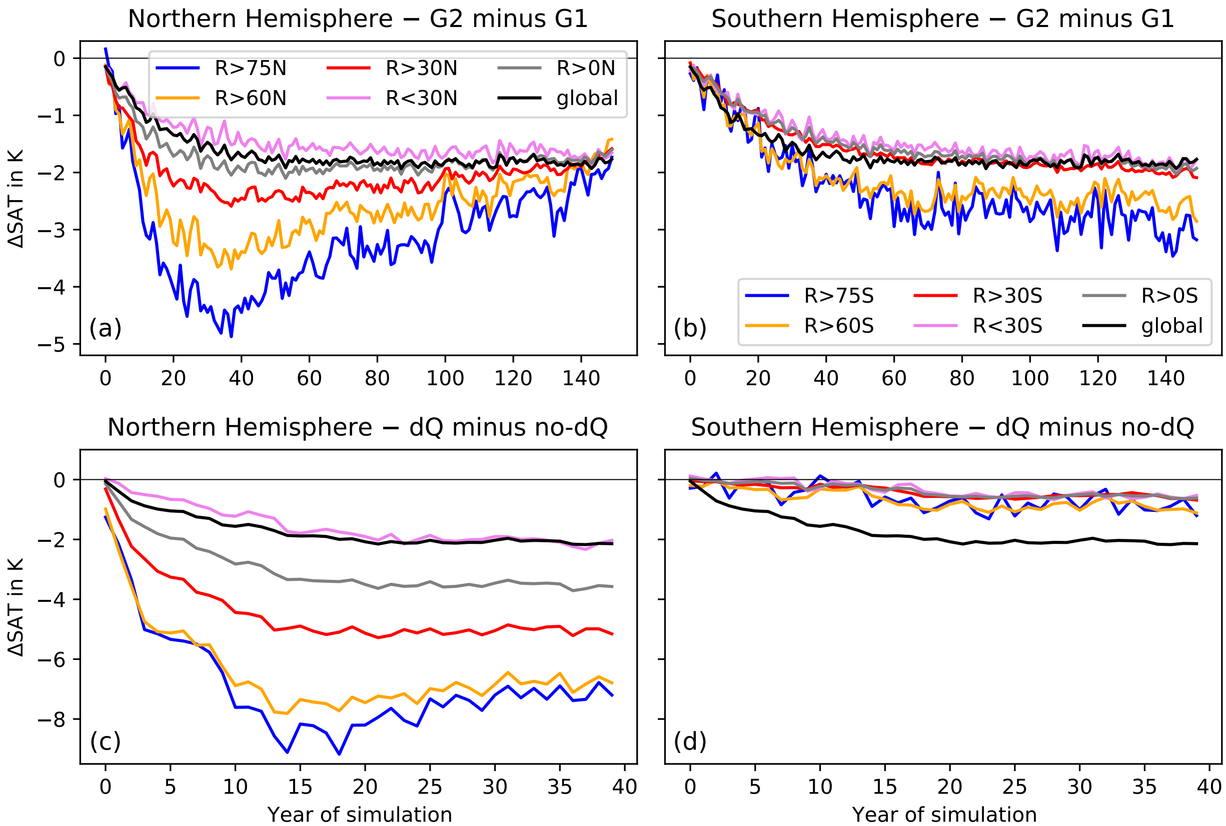

FIG. 4. Difference in surface-air temperature between G2 and G1 (panels a and b) as well as between the CESM2-SOM experiments dQ and no-dQ (panels c and d). Shown are zonal averages over latitude regions R>75°, R>60°, R>30°, R<30° (i.e., 0-30°), and R>0° for both hemispheres. In the legend ° is replaced by N in case of the Northern Hemisphere (NH) and by S in case of the Southern Hemisphere (SH). Panels a and c depict the values for the NH, panels b and d for the SH. Fig. S4 in the online supplemental material shows the p values of the differences based on a two-sided Welch's t-test.

sufficient to fully explain the development of the SAT differences between G2 and G1 in the NH. Recall from section 4a that the AMOC in G2 starts recovering after its initial decline which is not the case in G1. Hence, during the later years of the simulation relatively more energy is transported from the tropics poleward to the north in G2 and it seems plausible that this is partly responsible for the decline in the SAT difference in R>30N, R>60N, and R>75N. Indeed, in additional experiments with CESM2-SOM where the energy transport recovers after the initial 1-year decline over the course of years 2-30 of the simulation, the magnitude of the difference between SATs in R>30N and R>60N to the no-dQ experiment also decreases after the initial increase (see Fig. S5). This suggests that a possible AMOC recovery after its initial decline may play an important role for the transient climate sensitivity (see also Lin et al. 2019).

The SAT development is consistent with the effect of sea-ice melt mentioned above. Accordingly, Fig. 5c shows that Arctic sea ice declines much faster in no-dQ than in dQ and melts almost completely after about 15 years, while in dQ it declines initially slower but continues to melt over the whole 40-year simulation period engendering similar effects on SAT as explained above based on CMIP models (for maps of Arctic sea-ice extent of the CMIP and CESM2-SOM experiments see Figs. S7 and S8, respectively). However, after about year 25, the Arctic sea-ice area remains almost constant, consistent with the constant SAT difference between dQ and no-dQ in R>30N, R>60N, and R>75N. Thus, a consistent conjecture for the differences in temperature development in the Arctic, and more generally in the NH extra-tropics, is that a stronger decline in AMOC in G2 moderates the sea-ice loss compared to G1 which modifies various local feedbacks associated with sea ice (as described above). Together, these feedbacks and the decreased northward energy transport due to AMOC decline lead to a slower warming of the NH in G2 than in G1 (see also Mitevski et al. 2021). As expected, in the CESM2-SOM experiments where the mimicked AMOC recovers after the initial decline, the Arctic sea-ice loss is only initially slowed down but then continues similarly to the no-dQ experiment and it is completely lost after about 25 years of simulation (Fig. S6a).

Antarctic sea ice develops similar in dQ and no-dQ (Fig. 5d) and thus exhibits no differential effect on SAT. This is expected since the main impact of the mimicked AMOC change should concentrate in the NH. Consistently, dQ and no-dQ are found to be more similar in terms of SAT development in the SH than in the NH (Fig. 4cd). Indeed, differences in Antarctic sea-ice loss between the CESM2-SOM experiments are even smaller than those between G2 and G1, indicating that other factors than AMOC affect the Antarctic sea-ice development in fully coupled models. In agreement with the effects of sea-ice melt described above, the G2-minus-G1 differences in SH extra-tropical SATs (R>30S, R>60S, and R>75S) are larger than the differences between the CESM2-SOM experiments (compare Fig. 4b and d). We note that the development of Antarctic sea ice appears unaffected by AMOC recovery (Fig. S6b).

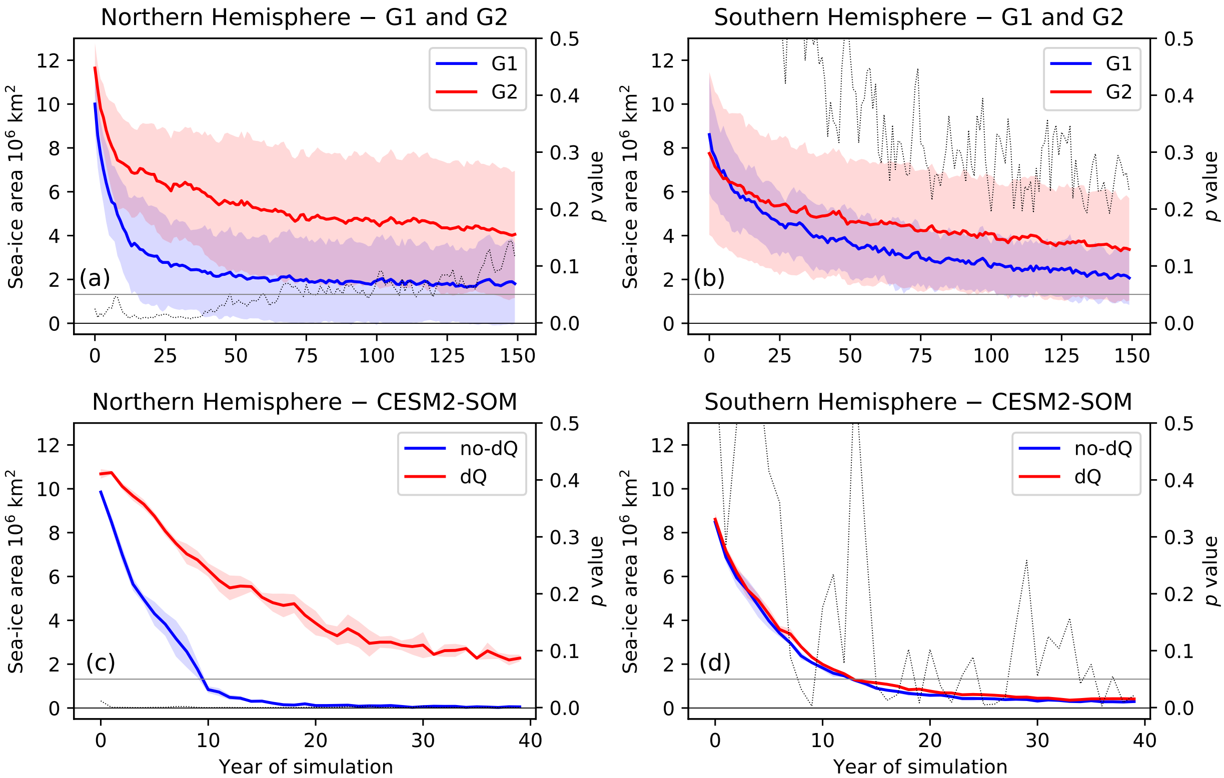

Fig. 5. Sea-ice area development in G1 and G2 (panels a and b) as well as in no-dQ and dQ (panels c and d). Shown is the sea-ice area integrated over the Northern Hemisphere (panels a and c) and over the Southern Hemisphere (panels b and d). The shading denotes the ±1-sigma spread across group members in (a) and (b) and across ensemble members in (c) and (d). See text for details. No sea-ice data were available for one member of G2 (BCC-CSM2-MR) so this model is excluded in (a) and (b). The thin, black, dotted line shows the p value of a two-sided Welch's t-test for the difference in group/ensemble mean, and the gray horizontal line indicates a p value threshold of 0.05. Note that in panel (c) the p value is almost invisible as it is close to zero.

2) Radiative fluxes

Eiselt and Graversen (2022) showed that G1 and G2 exhibit distinct differences in regional feedback change with the Arctic being the region with the largest differences. Here we discuss differences in global and regional radiative fluxes at TOA induced by these feedbacks. We show that the TOA radiative flux differences between G2 and G1 in the NH extra-tropics are qualitatively reproducible by CESM2-SOM experiments with mimicked AMOC change while differences in the tropics and in the SH require other explanations.

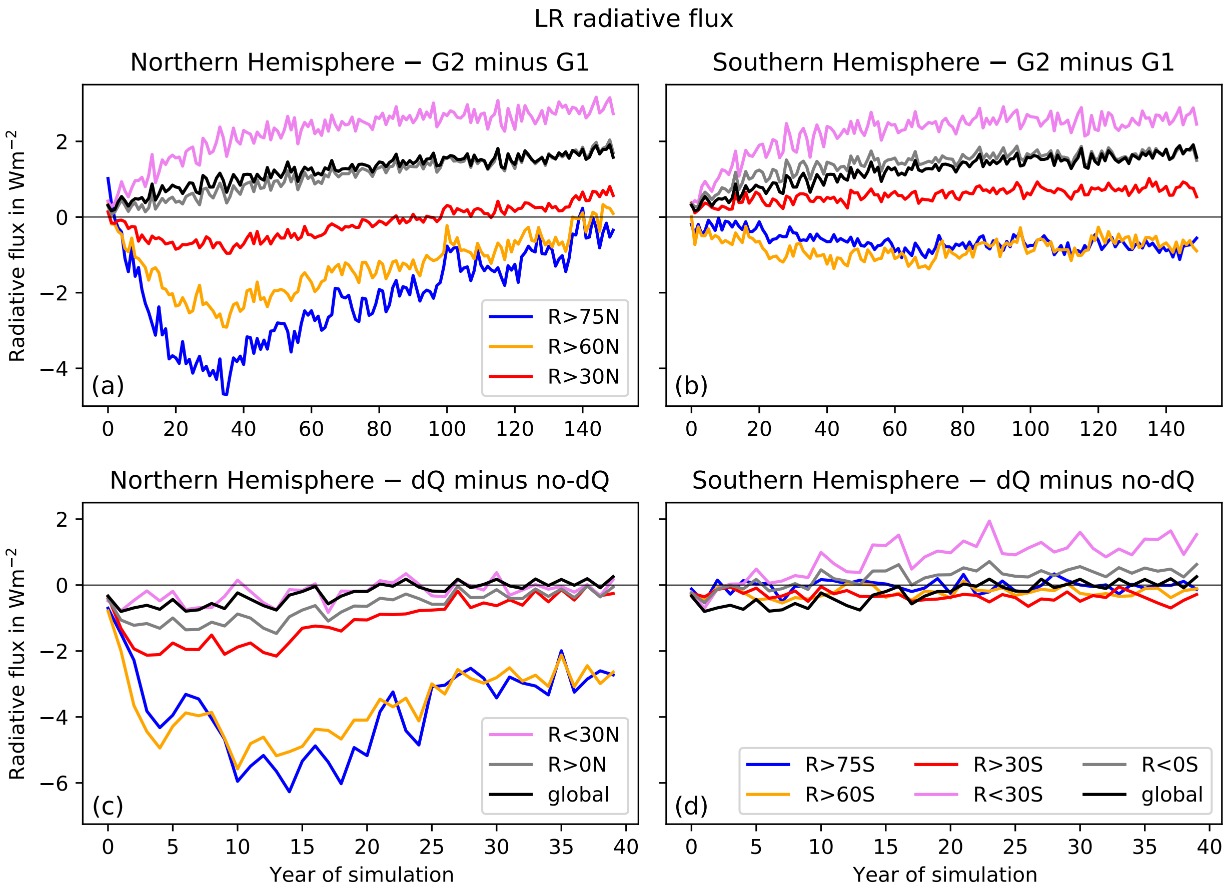

FIG. 6. Same as Fig. 4 but for lapse-rate feedback-induced top-of-atmosphere radiative-flux differences.

The G2-minus-G1 differences in LR and SA feedback-induced TOA radiative fluxes (Figs. 6ab and 7ab, respectively) in R>75N and R>60N, as well as R>75S and R>60S show similar patterns as the SAT differences, consistent with earlier findings that LR feedback in stably stratified regions is mostly determined by surface temperatures (e.g., Jenkins and Dai 2021), and with the above-explained connection between SAT and sea ice. In R>75N and R>60N, the differences in both SA and LR fluxes exhibit an initial increase in magnitude, but after about year 40 they decrease. In R>75S and R>60S, the change of the difference is initially slower but continues to slowly increase in magnitude over the whole CMIP simulation period. In the tropics (R<30N and R<30S), the differences in SA feedback-induced TOA radiative flux are negligible. However, the differences in LR feedback-induced TOA radiative flux are positive in both R<30N and R<30S and exhibit similar development in these regions. Furthermore, in R>30N the difference in LR TOA radiative

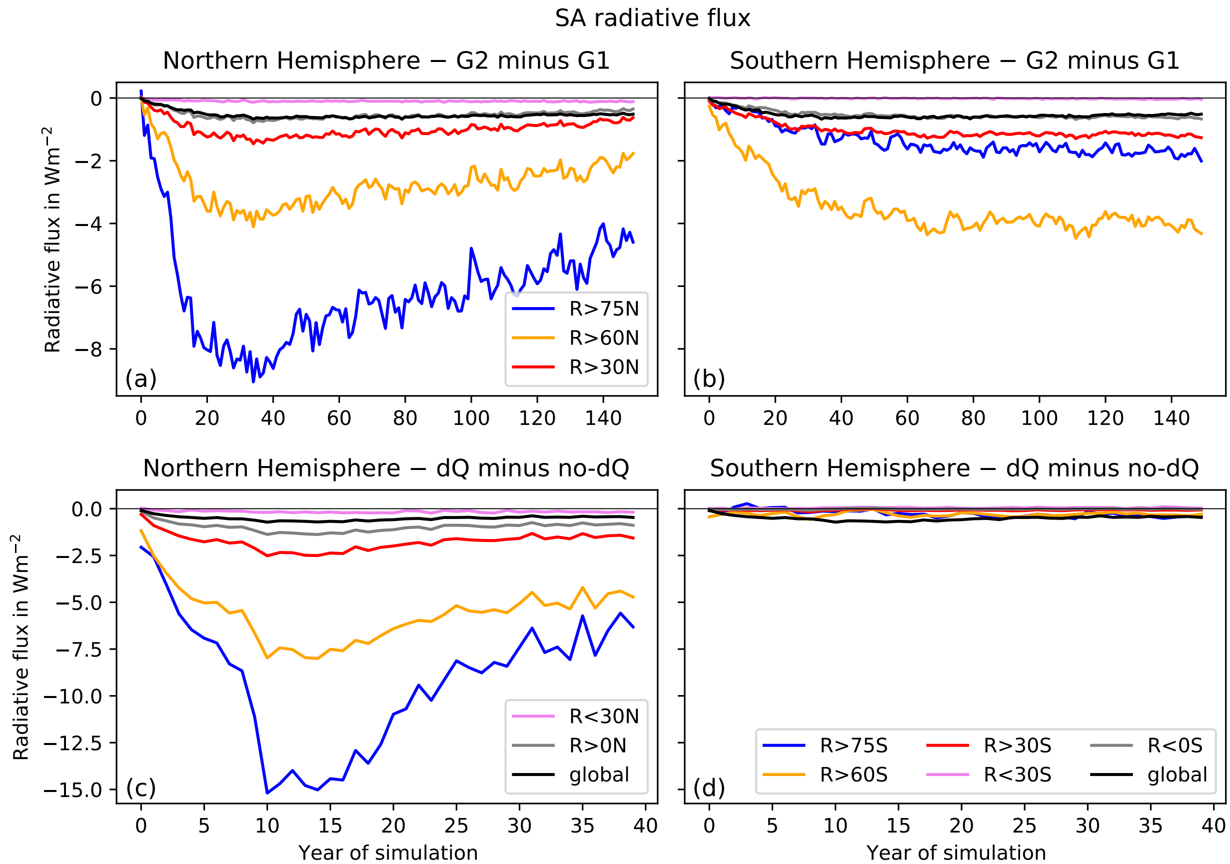

FIG. 7. As Fig. 4 but for surface-albedo feedback-induced top-of-atmosphere radiative-flux differences.

flux is initially negative and then becomes increasingly positive while it is always positve in R>30S and increases only slightly over the course of the simulation.

The differences in LR and SA feedback-induced fluxes between the CESM2-SOM experiments (dQ minus no-dQ; Figs. 6cd and 7cd, respectively) are qualitatively similar to the G2-minus-G1 differences in LR and SA TOA fluxes in the NH extra-tropics (i.e., R>30N, R>60N, R>75N). As for the G2-minus-G1 difference, there is a decrease of the difference between the dQ and no-dQ experiments from about year 15 after the initial fast increase. However, this decrease appears to stop around year 25. Since the LR and especially the SA TOA radiative fluxes are strongly influenced by sea ice, this is consistent with the Arctic sea-ice area development described above: The initially stronger sea-ice decline in no-dQ contributes to stronger positive SA and LR TOA fluxes. However, at around year 15 the Arctic sea ice has mostly melted in no-dQ while it continues to decline in the dQ experiment and thus the the SA and LR TOA fluxes increase more strongly in dQ than in no-dQ, causing the magnitude of their difference to decrease. After around year 25 the Arctic sea-ice area in dQ remains almost constant, consistent with the almost constant SA and LR TOA flux differences.

In our additional CESM2-SOM experiment with full AMOC recovery, the development of the difference in LR radiative flux in the NH extra-tropics appears qualitatively more similar to the G2-minus-G1 difference since both show about full recovery (Fig. S9). However, for the SA radiative flux, the recovery after the initial increase is much stronger than for the CMIP model difference (Fig. S10), which is consistent with the development of Arctic sea ice in the experiment explained above (Fig. S6a). While differences in time scales between a SOM and fully-coupled models are difficult to interpret, this may indicate that the AMOC as implemented in our CESM2-SOM experiments affects sea ice more strongly than the AMOC in fully-coupled models. In fully-coupled models with dynamical ocean components, the ocean itself may change in response to an AMOC change, for instance so that the cooling effect of the AMOC in the North Atlantic is spread to adjacent ocean areas. Such an effect is suppressed in the SOM experiments.

Since the G2-minus-G1 difference in Antarctic sea ice is not reproduced by the CESM2-SOM experiments (Fig. 5bd), it is consistent that the difference in Antarctic SA flux is not reproduced either. The differences in LR flux in the tropics (R<30N and R<30S) are discussed in more detail in section 4c3.

For the WV feedback-induced TOA radiative fluxes the G2-minus-G1 differences are generally similar across the NH and the SH and always negative (Fig. 8ab). However, similar to LR and SA fluxes, the differences in the NH regions increase initially faster in magnitude and then remain either constant (R<30N) or start slowly decreasing (R>30N, R>60N, R>75N), while in the SH regions the differences increase initially more slowly in magnitude but continue to decrease for the whole simulation. Notably, in both hemispheres the WV flux differences are larger in magnitude in the tropics than in the extra-tropical regions. This is also true for the dQ minus no-dQ differences (Fig. 8cd). Generally, for the NH regions, the G2-minus-G1 differences in WV flux are qualitatively reproduced by the dQ minus no-dQ differences, but this is not the case in the SH. In R>30S, R>60S, and R>75S there is little or no difference in WV TOA radiative flux between dQ and no-dQ, but in R<30S the difference is *positive* which is qualitatively different from the G2-minus-G1 difference. This again indicates that other factors than AMOC change are needed to explain

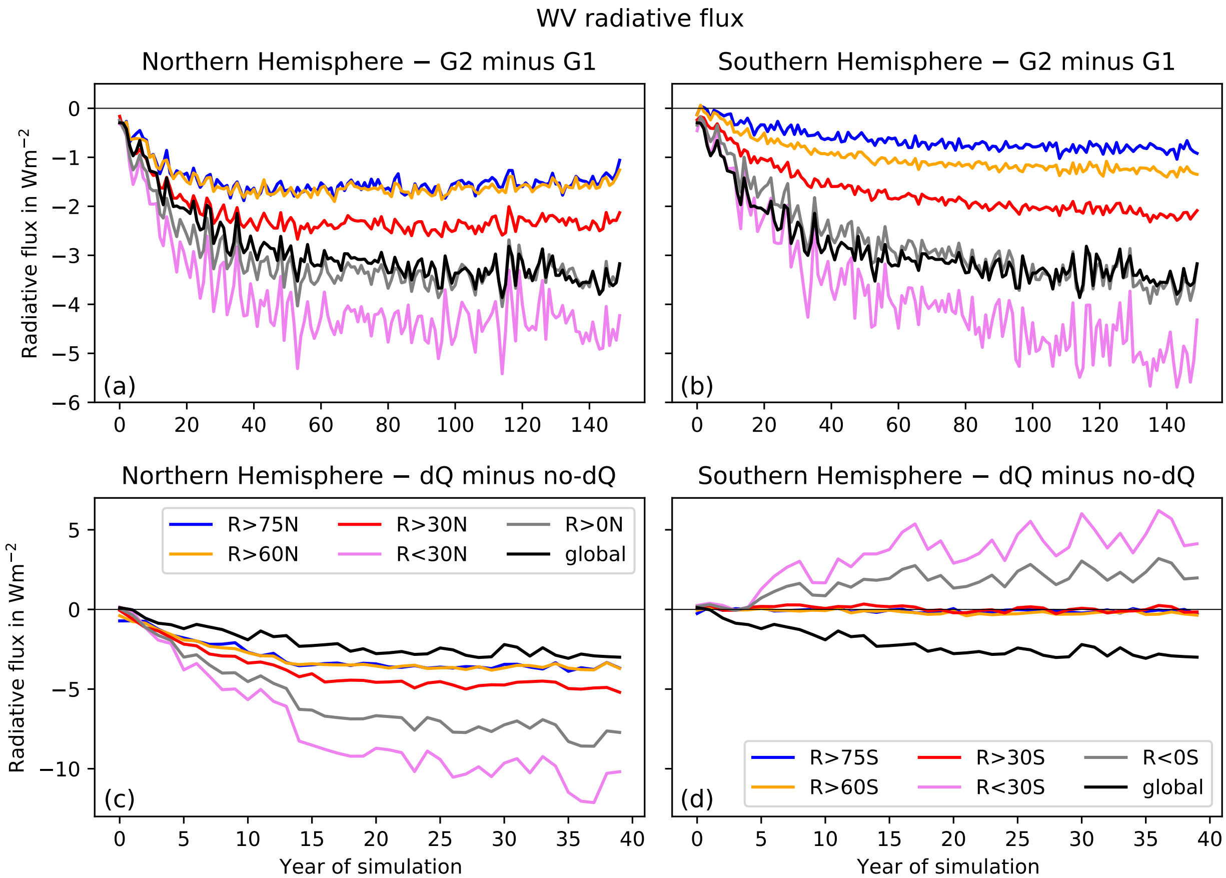

FIG. 8. As Fig. 4 but for water-vapour feedback-induced top-of-atmosphere radiative-flux differences.

differences between CMIP simulations in the SH and the tropics. The differences in WV radiative flux between the CMIP and the CESM2-SOM simulations in the tropics are further discussed in section 4c3.

Clouds and thus cloud feedback are notoriously model-dependent and hence the comparison of the SOM experiments based on a single model with groups of multiple fully-coupled models may be less instructive than for other feedbacks. Furthermore, cloud radiative flux changes can be caused by multiple, compensating factors (changes of, e.g., cloud area, cloud droplet size, cloud phase, cloud height etc.) whereby these changes are more difficult to interpret.

The long-wave cloud (CLW) radiative-flux differences between G2 and G1 are generally smaller than those of the other radiative fluxes (Fig. 9ab). The differences between the CESM2-SOM experiments in the extra-tropics are small but in the tropics they are much larger and thus qualitatively different from the differences between the CMIP model groups. Indeed, the CLW flux differences

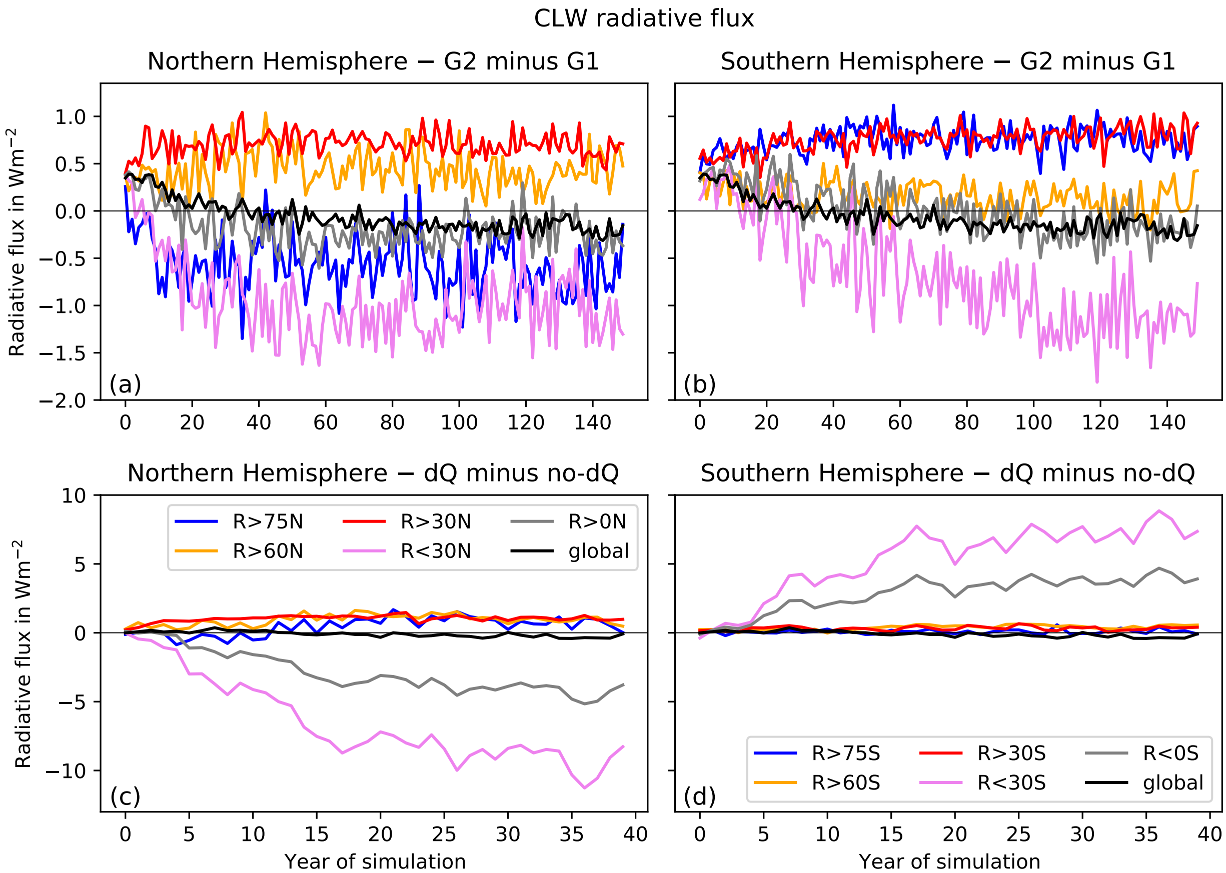

FIG. 9. As Fig. 4 but for long-wave cloud feedback-induced top-of-atmosphere radiative-flux differences.

between the CESM2-SOM in the tropics are similar to the WV flux differences. The short-wave cloud (CSW) radiative-flux differences between G2 and G1 are generally larger than those of the CLW fluxes and negative in all regions, except for R>75S where they are close to zero (Fig. 10). These differences appear to be qualitatively reproduced in the CESM2-SOM experiments, except in the NH tropics (R<30N), where the difference is positive. Notably, in the CESM2-SOM experiments, the CLW and CSW flux differences in the tropics are of opposite sign and thus compensate each other to some degree which is not the case for the G2-minus-G1 difference. The differences in the tropics are discussed in more detail in section 4c3.

Generally it may be concluded that for the NH extra-tropics, the TOA radiative fluxes and their development are qualitatively reproduced, hereby providing supporting evidence that the AMOC change is important in explaining differences in NH climate reponse between G2 and G1. However,

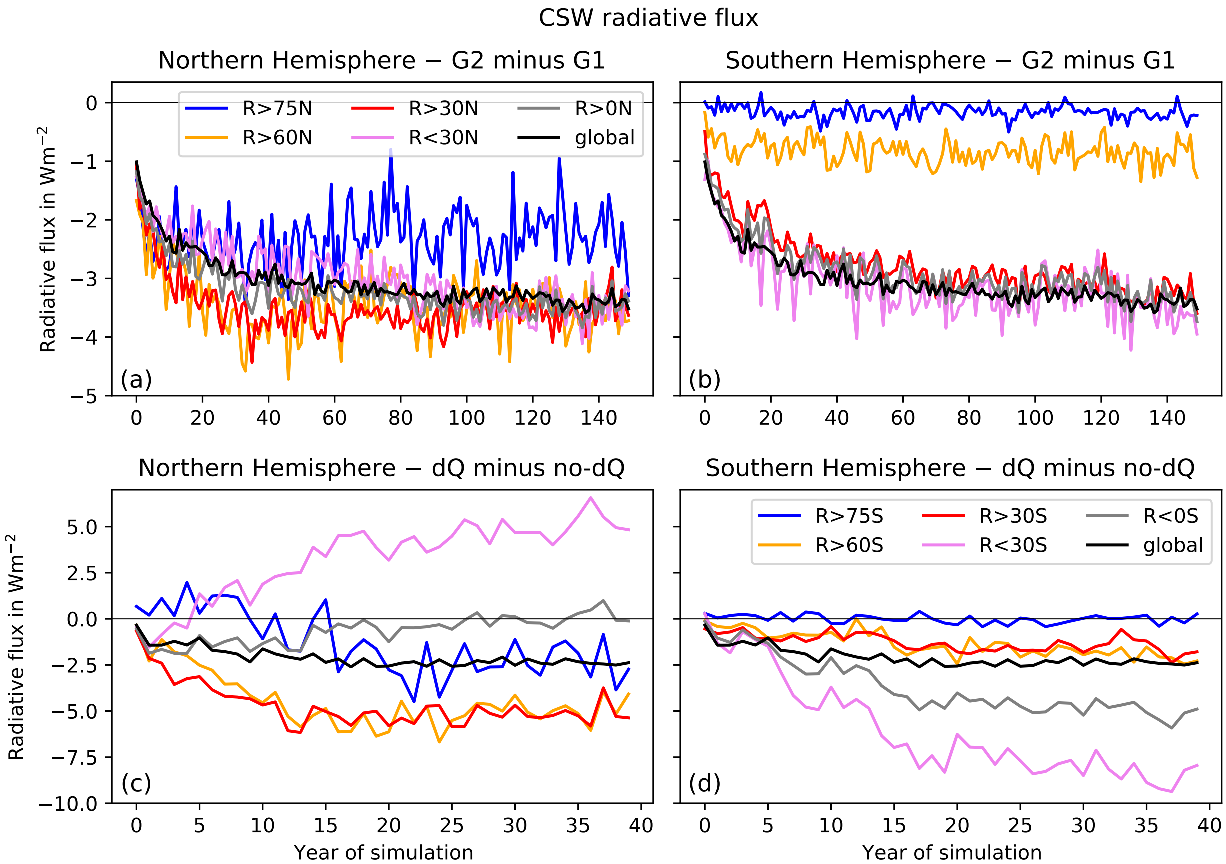

FIG. 10. As Fig. 4 but for short-wave cloud feedback-induced top-of-atmosphere radiative-flux differences.

differences in the tropics are not captured by the SOM experiments. In the following, the dQ and no-dQ experiments are further analysed to explain the difference.

3) Differences between CESM2-SOM and CMIP experiments

The differences between dQ and no-dQ in the tropics and the opposing effects seen in the northern tropics (R<30N) and the southern tropics (R<30S) are consistent with a concomitant change in the Hadley cell and the difference in warming between the hemispheres. As the mimicked AMOC change effectively constitutes a decline in northward oceanic heat transport in the NH, it is expected that this ocean heat-transport reduction is at least partly compensated by an increase in northward heat transport in the atmosphere (known as Bjerknes compensation, Bjerknes 1964), and thus a stronger NH Hadley cell. This has been documented in previous studies where ocean heat transport is changed in a SOM (e.g., Singh et al. 2022).

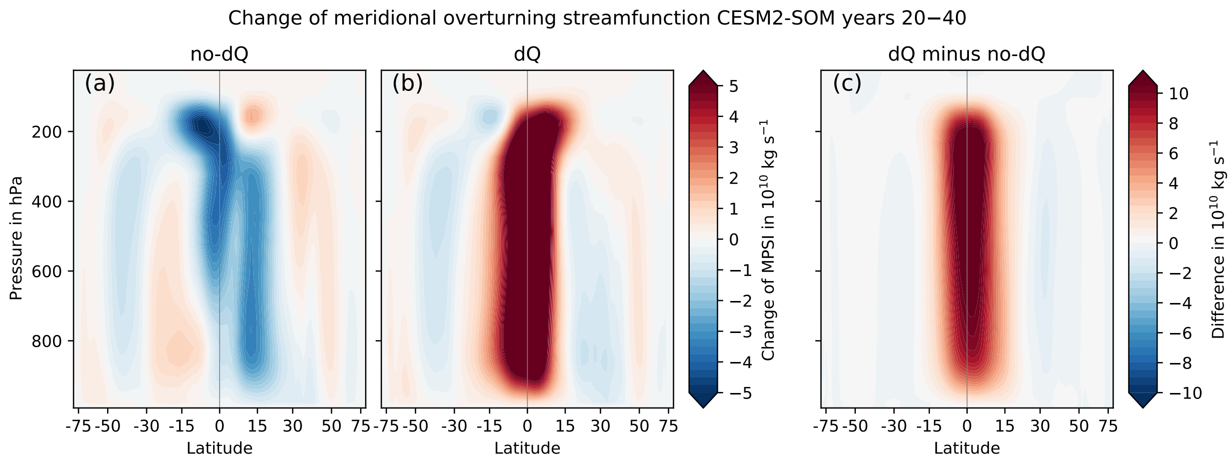

FIG. 11. Change of the atmospheric meridional overturning streamfunction for no-dQ (a) and dQ (b) and their difference (c). The streamfunctions averaged over years 20-40 of the simulations.

In the no-dQ case, the northern Hadley cell weakens over the course of the simulation while the southern Hadley cell strengthens and moves slightly northward so that the ascending branch in the annual mean becomes situated slightly north of the equator (Figs. 11a, S11a, and S12). In the dQ case, the opposite development obtains and the northern Hadley cell strengthens and shifts slightly southward so that in the annual mean the ascending branch is now situated south of the equator (Figs. 11b, S11b, and S12). As a consequence, there is more ascending motion in the R<30S and more descending motion in R<30N in dQ relative to no-dQ (Fig. 11c; see Figs. S13 and S14 for vertical cross sections of the vertical velocity), increasing and decreasing, respectively, atmospheric humidity and cloud. The descending of the dry tropopause air together with a generally colder NH (Fig. S15) leads to a lower specific humidity in R<30N across the vertical profile in dQ than in no-dQ and induces a smaller TOA radiative flux associated with WV. The opposite is the case in R<30S, although to a lesser degree (see Figs. S16 and 8cd). Furthermore, the increase in descent in R<30N and ascent in R<30S implies that the cloud content is reduced in R<30N but increased in R<30S (Fig. S17) so that in the former region more long-wave radiation can escape to space and more short-wave radiation can reach the surface, while in the latter the opposite is the case. Thus, the dQ minus no-dQ difference in CLW flux is negative in R<30N, but positive in R<30S (Fig. 9cd), while the opposite obtains for the CSW flux difference (Fig. 10cd). Finally, comparing the difference in temperature lapse rates between dQ and no-dQ in the tropics, it is found that the difference in R<30N is positive at lower and negative at higher levels, while the difference in the R<30S is always negative (Fig. S18). This is consistent with the LR feedback induced flux difference being close to zero in the R<30N and positive in R<30S (Fig. 6cd). However, the governing processes for the difference in lapse rate are less clear. We speculate that due to the larger cloud amount in the R<30S in the dQ case, the emissivity is stronger and thus the atmosphere cools more efficiently. Note that in the radiative kernel method applied here, changes in emissivity and the resulting higher cooling efficiency are included in the CLW radiative fluxes but not in the LR radiative fluxes. However, this effect is difficult to diagnose as it is masked by the effect of clouds blocking outgoing long-wave radiation from the surface. A further but smaller effect may be that due to more deep convective cloud less solar radiation is absorbed in the atmosphere, also leading to a relative cooling in R<30S. These results are broadly consistent with the findings of Singh et al. (2022) for their case with decreased ocean heat transport.

The differences between no-dQ and dQ as regards humidity, temperature, and clouds are to some extent consistent with the differences between G2 and G1, although the differences between the CESM2-SOM experiments seem generally more extreme than those between G2 and G1. In terms of the Hadley cell, the difference between dQ and no-dQ (Fig. 11c) is qualitatively similar to the G2-minus-G1 difference (Fig. S19c), although the latter is weaker. Comparing the change of Hadley circulation separately for G1 and G2 (Fig. S19ab) as well as dQ and no-dQ (Fig. 11ab) it is clear that while G1 and no-dQ are similar, the response of the Hadley circulation in dQ is considerably stronger than in G2. This indicates that the mimicked AMOC change in the SOM experiment is more effective at latitudinally redistributing energy than the AMOC in the CMIP model experiments or that other processes important for the difference between G2 and G1 are compensating the AMOC response in the tropics. The difference in humidity between G2 and G1 is latitudinally more symmetric around the equator than between the CESM2-SOM experiments. However, the G2-minus-G1 difference is still more negative in the NH than in the SH around year 40 (Fig. S20). As the NH (and especially the Arctic) in G2 warms relatively more over the later years of the abrupt4xCO2 simulation than in G1, the humidity difference profile becomes almost latitudinally symmetric (Fig. S21). This is not the case in the CESM2-SOM simulations as the difference remains latitudinally asymmetric for the whole simulation. However, if we let the AMOC recover in the CESM2-SOM simulation, the difference between humidity profiles becomes almost symmetric by the end of the 40-year simulation after being asymmetric in the first years (not shown).

We conclude that by our simple AMOC implementation, we can qualitatively reproduce differences in regional SATs and TOA radiative fluxes for the NH seen in our CMIP model groups G2 and G1. However, parts of the model response in the CMIP experiments are not reproduced in the CESM2-SOM experiments, especially in the SH as well as partly in the tropics. This indicates that the differences between G2 and G1 are not fully explained by differences in AMOC development.

5. Dicussion and Conclusion

Eiselt and Graversen (2022) distinguished two climate model groups based on the magnitude of climate feedback change over time. Here we show that they differ significantly in terms of their response of the Atlantic Meridional Overturning Circulation (AMOC) to the CO_2 quadrupling. The influence of the AMOC on climate feedback is investigated employing a slab-ocean model (SOM) where, in addition to the abrupt quadrupling of the CO_2 concentration, a change in the ocean heat transport (Q-flux in the SOM) is prescribed to mimic the difference in the AMOC evolution between G2 and G1. It is found that the differences between surface-albedo, lapse-rate, and water-vapour feedback-induced TOA radiative fluxes in the Northern Hemisphere between SOM experiments with and without prescribed Q-flux change are qualitatively similar to those between G2 and G1. Furthermore, differences in Arctic sea-ice decline and in the development of the Hadley circulation are qualitatively similar. However, unexplained differences remain, especially in the tropics and in the Southern Hemisphere, indicating that the AMOC change alone is insufficient to explain the change of climate feedback over time in response to the CO_2 forcing.

An important process that is not accounted for in the experiments conducted for this study is Antarctic Bottom Water (AABW) formation which appears to affect the climate system in ways similar to the AMOC but which has received much less attention (He et al. 2017). The lack of a representation of a change in AABW formation in our SOM experiments may at least partly explain the fact that the differences between G2 and G1 in the Southern Hemisphere are not qualitatively reproduced by the SOM experiments. However, based on an AABW formation index similar to He et al. (2017), no significant differences between G2 and G1 are found in both the piControl and the abrupt4xCO2 experiment (Fig. S22). Notably though, the differences in surface temperature and

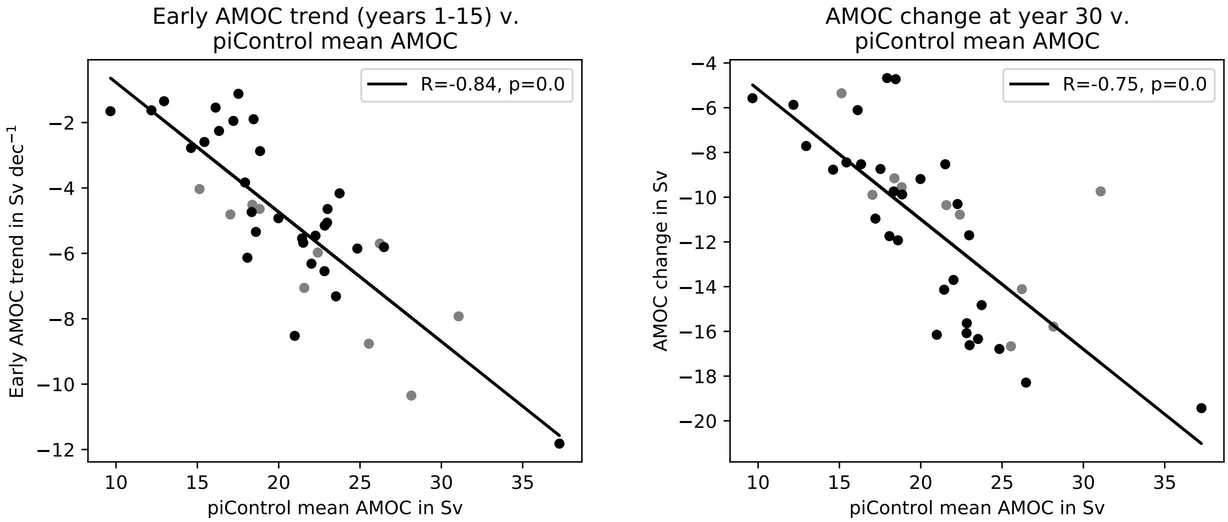

FIG. 12. Pre-industrial control mean AMOC v. AMOC trend over years 1 to 15 (left) and change of AMOC averaged over years 28-32 (right). Members of CMIP5 are depicted in gray, members of CMIP6 in black.

radiative fluxes between G2 and G1 in the Southern Hemisphere are smaller and do not exhibit a similar distinct development as those in the Northern Hemisphere (i.e., fast-paced increase of the differences followed by a slower decrease; e.g., Fig. 4). Thus, it may be more difficult to robustly connect changes in the ocean heat transport in the Southern Hemisphere to changes in other climate variables in the group comparison study.

Another important feature in the G2−G1 comparison that is not taken into account in our slab-ocean model set-up is the difference between their pre-industrial states. As pointed out in section 4a and shown in Fig. 1a, G2 exhibits a significantly stronger pre-industrial AMOC than G1. Furthermore, the Arctic sea-ice extent is larger and the Northern Hemispheric surface temperature lower in G2 than in G1 (not shown). However, these differences are not statistically significant ($p > 0.05$). The fact that the model group with the larger pre-industrial AMOC exhibits the larger AMOC change may generally correspond with the notion of "capacity to change" introduced by Kajtar et al. (2021). That is, since the AMOC is expected to decline in response to global warming (IPCC 2021), in models with a stronger base-state AMOC it has a larger capacity to decline, resulting in larger climate impacts. In support of this, we find significant correlations between pre-industrial AMOC strength and both the pace and strength of AMOC decline (Fig. 12; see also Gregory et al. 2005; Bellomo et al. 2021; Lin et al. 2019; He et al. 2017). This indicates that if a strong base-state

AMOC exists, it will also generally decline significantly in response to a sufficiently large forcing (see Mitevski et al. 2021), engendering the effects described in the present study. However, the physical processes that cause the variation of the base-state AMOC and its importance for the AMOC decline under forcing conditions are a topic of ongoing research. Some studies highlight the impact of ocean-model resolution on AMOC strength, although these studies are often based on few models implying that the confidence in their conclusions remains low (Winton et al. 2014; Jackson et al. 2020; Roberts et al. 2020). Model resolution is out of the scope of our study, but we report that G1 and G2 differ only slightly in terms of ocean-model resolution, and hence model resolution is unimportant for our conclusions. Other studies find a connection between AMOC decline and the base-state Arctic sea-ice extent (Levermann et al. 2007; Lin et al. 2023): A stronger base-state AMOC is accompanied by less base-state sea-ice cover in the Labrador Sea. Under warming in response to a CO_2 forcing, this causes a larger decline in turbulent surface heat fluxes and thus leads to a stronger AMOC decline. We cannot corroborate this mechanism for our CMIP model groups, as G2 with the significantly stronger base-state AMOC compared to G1 shows only a slightly and statistically non-significantly smaller sea-ice cover in the Labrador Sea (Fig. S23). Recently, Lin et al. (2023) found that the base-state stratification in the Labrador Sea influences the AMOC decline. A stronger AMOC causes the CO_2-induced surface warming to efficiently sink to deeper layers, thus leading to a positive buoyancy anomaly at depth that weakens the AMOC. Jackson et al. (2020), on the other hand, elucidate the importance of the spatial structure of the AMOC, again highlighting the role of the Labrador Sea. Under a warming due to CO_2, models with more deep water formation (DWF) in the Labrador Sea experience a decline in DWF as the atmosphere warms and the ocean cools (due to AMOC decline). In the Labrador Sea the region of DWF does not move further north and thus DWF is continuously reduced, leading to a strong decline in AMOC. Conversely, in models that have more DWF in the Greenland-Iceland-Norway Sea, DWF can move further north, hereby maintaining a stronger AMOC. Jackson et al. (2020) also find that the spatial structure of the AMOC may be connected to model resolution. However, as indicated above, this result is based on relatively few ocean models.

As found by Lin et al. (2019) and confirmed by the results from the SOM experiments in the present study, the *recovery* of AMOC may also have an important impact on climate feedback and sensitivity. However, correlations across fully-coupled models are ambiguous. The total feedback change is only weakly correlated with the late AMOC trend (years 51−150; R=0.2, p=0.21; Fig. S24, left) but the lapse-rate feedback change exhibits considerable correlation with the late AMOC trend (R=0.5, p=0.001; Fig. S24, right; this increases to R=0.65 if three outliers are excluded). This suggests that on the one hand the AMOC is specifically important for the lapse-rate feedback and on the other hand that other processes than the AMOC are important for the change of climate feedback as well, e.g., Southern Ocean cloud feedback (Bjordal et al. 2020; Zelinka et al. 2020) and surface-temperature development in the Indo-Pacific Warm Pool (Dong et al. 2019, 2020). It may be noted that while the G2−G1 comparison suggests that the models with a stronger base-state AMOC and a larger AMOC decline also exhibit a larger AMOC recovery in later years (see also Lin et al. 2019), the across model correlation of both early AMOC trend and pre-industrial AMOC with late AMOC trend are weak (R=-0.1 and R=0.08, respectively; see Fig. S25).

In line with a number of recent studies, our findings point to the importance of the change of the AMOC (He et al. 2017; Lin et al. 2019; Bellomo et al. 2021; Mitevski et al. 2021) and the change of ocean heat transport in general (Rugenstein et al. 2016; Singh et al. 2022) for the change of climate feedback and sensitivity, consistent with the notion of the pattern effect (Stevens et al. 2016). Furthermore, studies such as He et al. (2017), Jackson et al. (2020), and Lin et al. (2023) in particular, and Kajtar et al. (2021) more generally, together with the findings presented here indicate that to correctly model the AMOC response to a greenhouse-gas forcing, it is important to correctly represent the base state of the climate system in general and the ocean circulation and AMOC in particular. Thus, more research into the real-world base-state AMOC is needed to more confidently gauge the influence of the AMOC on ongoing and future climate change. Moreover, as indicated by the importance of a possible AMOC recovery, the drivers of AMOC need to be better understood in order to confidently predict the longer-term response of the AMOC to a greenhouse-gas forcing.

Acknowledgments. The authors want to thank Patrick Stoll for helpful discussions and comments. We acknowledge the World Climate Research Programme's (WCRP) Working Group on Coupled Modelling as well as the Earth System Grid Federation (ESGF) for making available and archiving the model output in the CMIP5 and CMIP6 archives. The data downloaded and generated during this research is stored at the Nird storage facilities provided by the Norwegian e-infrastructure for research and education, UNINETT Sigma2, under the project NS9063K. The CESM2 simulations were performed on the FRAM supercomputer at the University of Tromsø (UiT) provided by UNINETT Sigma2, under the project NN9345k. The work is part of the project *UiT – Climate Initiative, Ice-ocean-atmosphere interactions in the Arctic – from the past to the future*, funded by the Faculty of Science and Technology, University of Tromsø. We thank two anonymous reviewers and the editor, Stephen Yeager, for their help in improving the work.

Data availability statement. The original CMIP model output is available on the WCRP's CMIP5 (`https://esgf-node.llnl.gov/search/cmip5/`) and CMIP6 archives (`https://esgf-node.llnl.gov/search/cmip6/`). The procedure for generating the climate feedback radiative fluxes using the radiative kernels is described in Eiselt and Graversen (2022). The proprietary CESM numerical model simulations presented in this study are too large to archive or to transfer. Instead, we provide all the information needed to replicate the simulations; we used model version 2.1.3, freely available at `https://www.cesm.ucar.edu/models/cesm2`. Pre- and postprocessing code for the CMIP and CESM2-SOM data is available at `https://doi.org/10.5281/zenodo.7950682`. The scripts for increasing the CO_2 concentration and for changing the Q-flux in CESM2-SOM are available at `https://doi.org/10.5281/zenodo.7937804`.

APPENDIX A

Estimation of Q-flux change due to AMOC change

Here we show a brief derivation of our order-of-magnitude estimation of the Q-flux change due to the change of AMOC. Our method is similar to that presented in Buckley and Marshall (2016) but simpler and more *ad hoc* since we employ surface temperature instead of ocean potential temperature of upper and lower branch of the AMOC. However, since the values derived by our simpler method (see below and Tables A1 and S3) are comparable to those given in Buckley and

Marshall (2016), we are confident that our method is suitable for order-of-magnitude estimates as applied here.

We begin by defining a meridional streamfunction Ψ for the zonally integrated volume transport in the Atlantic sector such that:

$$v = -\frac{\partial \Psi}{\partial z} \tag{A1}$$

where v is zonally integrated meridional velocity in the Atlantic basin. Following Buckley and Marshall (2016), the heat transport, E, by the AMOC can be expressed as

$$E = -\rho_0 c_p \int_{-H}^{0} \frac{\partial \Psi}{\partial z} \theta \mathrm{d}z, \tag{A2}$$

where ρ_0 is the reference level density, c_p is the specific heat capacity of water at constant pressure, H is the depth of the Atlantic basin, and θ is potential temperature (with the ocean surface as a reference level). Assuming that the stream function vanishes at the the top and bottom of the Atlantic basin, i.e., the vertically integrated mass flux is zero, and that the upper and lower branches have spatially uniform potential temperature, denoted as θ_s and θ_b respectively, eq. A2 yields:

$$E = \rho_0 c_p \Psi_m \Delta\theta, \tag{A3}$$

where Ψ_m is the maximum of Ψ at the interface between the upper and lower branch of AMOC at about 1000 m depth, and $\Delta\theta = \theta_s - \theta_b$ is the difference between the potential temperature in the upper and lower branch of the AMOC. Assuming now that the poleward upper-branch water originates from surface water in the south, and the lower branch return-flow water from sinking surface water in the north, the heat transport of the AMOC becomes:

$$E = \rho_0 c_p \Psi_m \Delta T, \tag{A4}$$

where ΔT can be roughly captured by the surface-temperature difference between the tropical and North Atlantic.

Both a change in the meridional surface-temperature difference ($\delta\Delta T$) and in the strength of the AMOC ($\delta\Psi_m$) can lead to an energy transport change (δE):

$$E + \delta E = \rho_0 c_p (\Psi_m + \delta\Psi_m)(\Delta T + \delta\Delta T). \tag{A5}$$

It follows that the change in energy transport δE can be expressed as:

$$\delta E = \rho_0 c_p (\Psi_m \delta\Delta T + \delta\Psi_m \Delta T + \delta\Psi_m \delta\Delta T). \tag{A6}$$

Substituting the values for G1 and G2 on the right-hand side we obtain that the difference between G2 and G1 in terms of change of northward energy transport is about -0.5 PW (averaged over years 13-17 of abrupt4xCO2). See Table A1 for the values of the individual terms of eq. A6 and Table S3 for the values of the individual parameters.

In the dQ experiments this energy is added as Q-flux to an area over the North Atlantic as defined in section 2b and equates to about 50 Wm^{-2} (for a sensitivity analysis of the choice of these settings see Fig. S1 in the online supplemental material). To balance this flux and keep the global mean Q-flux change at zero to not introduce a global net forcing, the Q-flux change implemented in the tropic region (see section section 2b) is about -25 Wm^{-2}.

term	G1	G2
$\Psi_m \delta\Delta T$	-135.94	-76.5
$\delta\Psi_m \Delta T$	-361.23	-918.82
$\delta\Psi_m \delta\Delta T$	30.21	26.56

TABLE A1. Values for the terms in eq. A6 in 10^{12} W (terms multiplied by $\rho_0 c_p$). The values are derived from differences between G2 and G1 averaged over the years 13 to 17 in the abrupt4xCO2 experiment.

References

Andrews, T., and M. J. Webb, 2018: The dependence of global cloud and lapse rate feedback on the spatial structure of tropical Pacific warming. *J. Climate*, **31**, 641–654, https://doi.org/10.1175/JCLI-D-17-0087.1.

Arrhenius, S., 1896: On the influence of carbonic acid in the air upon the temperature of the ground. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*,

5, 237–276, https://doi.org/10.1080/14786449608620846.

Bellomo, K., M. Angeloni, S. Corti, and J. v. Hardenberg, 2021: Future climate change shaped by inter-model differences in Atlantic meridional overturning circulation response. *Nat. Commun.*, **12**, 3659, https://doi.org/10.1038/s41467-021-24015-w.

Bitz, C. M., K. M. Shell, P. R. Gent, D. A. Bailey, G. Danabasoglu, K. C. Armour, M. M. Holland, and J. T. Kiehl, 2012: Climate sensitivity of the Community Climate System Model, Version 4. *J. Climate*, **25**, 3053–3070, https://doi.org/10.1175/JCLI-D-11-00290.1.

Bjerknes, J., 1964: Atlantic air-sea interaction. *Advances in Geophysics*, **10**, 1–82, https://doi.org/10.1016/S0065-2687(08)60005-9.

Bjordal, J., T. Storelvmo, K. Alterskjær, and T. Karlsen, 2020: Equilibrium climate sensitivity above 5 °C plausible due to state-dependent cloud feedback. *Nat. Geosci.*, **13**, 718–721, https://doi.org/10.1038/s41561-020-00649-1.

Buckley, M. W., and J. Marshall, 2016: Observations, inferences, and mechanisms of Atlantic Meridional Overturning Circulation variability: A review. *Rev. Geophys.*, **54**, 5–63, https://doi.org/10.1002/2015RG000493.

Caldwell, P. M., M. D. Zelinka, K. E. Taylor, and K. Marvel, 2016: Quantifying the sources of intermodel spread in equilibrium climate sensitivity. *J. Climate*, **25**, 513–524, https://doi.org/10.1175/JCLI-D-15-0352.1.

Ceppi, P., and J. M. Gregory, 2017: Relationship of tropospheric stability to climate sensitvity and Earth's observed radiation budget. *Proc. Natl. Acad. Sci. (USA)*, **114**, 13 126–13 131, https://doi.org/10.1073/pnas.1714308114.

Charney, J. G., and Coauthors, 1979: Carbon dioxide and climate: a scientific assessment. Report of an ad hoc study group on carbon dioxide and climate, National Academy of Sciences.

Dai, A., 2022: Arctic amplification is the main cause of the Atlantic meridional overturning circulation weakening und large CO_2 increases. *Climate Dyn.*, **58**, 3243–3259, https://doi.org/10.1007/s00382-021-06096-x.

Dai, A., D. Luo, M. Song, and J. Liu, 2019: Arctic amplification is caused by sea-ice loss under increasing CO_2. *Nat. Commun.*, **10**, 121, https://doi.org/10.1038/s41467-018-07954-9.

Danabasoglu, G., and Coauthors, 2020: The Community Earth System Model Version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, **12**, e2019MS001916, https://doi.org/10.1029/2019MS001916.

Dong, Y., K. C. Armour, M. D. Zelinka, C. Proistosescu, D. S. Battisti, C. Zhou, and T. Andrews, 2020: Intermodel spread in the pattern effect and its contribution of climate sensitivity in CMIP5 and CMIP6 models. *J. Climate*, **33**, 7755–7775, https://doi.org/10.1175/JCLI-D-19-1011.1.

Dong, Y., C. Proistosescu, K. C. Armour, and D. S. Battisti, 2019: Attributing historical and future evolution of radiative feedbacks to regional warming patterns using a Green's function approach: The preeminence of the Western Pacific. *J. Climate*, **32**, 5471–5491, https://doi.org/10.1175/JCLI-D-18-0843.1.

Eiselt, K.-U., and R. G. Graversen, 2022: Change in climate sensitivity and its dependence on lapse-rate feedback in $4{\times}CO_2$ climate mode experiments. *J. Climate*, **35**, 2919–2932, https://doi.org/10.1175/JCLI-D-21-0623.1.

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

Gettleman, A., and Coauthors, 2019: High climate sensitivity in the Community Earth System Model Version 2 (CESM2). *Geophys. Res. Lett.*, **46**, 8329–8337, https://doi.org/10.1029/2019GL083978.

Graversen, R. G., P. L. Langen, and T. Mauritsen, 2014: Polar amplification in CCSM4: Contributions from the lapse rate and surface albedo feedbacks. *J. Climate*, **27**, 4433–4450, https://doi.org/10.1175/JCLI-D-13-00551.1.

Graversen, R. G., and M. Wang, 2009: Polar amplification in a coupled climate model with locked albedo. *Climate Dyn.*, **33**, 629–643, https://doi.org/10.1007/s00382-009-0535-6.

Gregory, J. M., and Coauthors, 2004: A new method for diagnosing radiative forcing and climate sensitivity. *Geophys. Res. Lett.*, **31**, L03205, https://doi.org/10.1029/2003GL018747.

Gregory, J. M., and Coauthors, 2005: A model intercomparison of changes in the atlantic thermohaline circulation in response to increasing atmospheric co2 concentration. *Geophys. Res. Lett.*, **32**, L12703, https://doi.org/10.1029/2005GL023209.

Hall, A., 2004: The role of surface albedo feedback in climate. *J. Climate*, **17**, 1550–1568, https://doi.org/10.1175/1520-0442(2004)017<1550:TROSAF>2.0.CO;2.

He, J., M. Winton, G. Vecchi, L. Jia, and M. Rugenstein, 2017: Transient climate sensitivity depends on base climate ocean circulation. *J. Climate*, **30**, 1493–1504, https://doi.org/10.1175/JCLI-D-16-0581.1.

Held, I., and B. J. Soden, 2000: Water vapor feedback and global warming. *Annu. Rev. Energy Environ.*, **25**, 441–475, https://doi.org/10.1146/annurev.energy.25.1.441.

Hunke, E. C., W. H. Lipscomb, A. K. Turner, N. Jeffery, and S. Elliott, 2015: CICE: The Los Alamos Sea Ice Model. Documentation and Software User's Manual. Version 5.1. Tech. Rep. LA-CC-06-012, T-3 Fluid Dynamics Group, Los Alamos National Laboratory.

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*, Vol. In Press. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, https://doi.org/10.1017/9781009157896.

Jackson, L. C., R. Kahana, T. Graham, M. A. Ringer, T. Woollings, J. Mecking, and R. A. Wood, 2015: Global and european climate impacts of a slowdown of the AMOC in a high resolution gcm. *Climate Dyn.*, **45**, 3299–3316, https://doi.org/10.1007/s00382-015-2540-2.

Jackson, L. C., and Coauthors, 2020: Impact of ocean resolution and mean state on the rate of AMOC weakening. *Climate Dyn.*, **55**, 1711–1732, https://doi.org/10.1007/s00382-020-05345-9.

Jenkins, A., and A. Dai, 2021: The impact of sea-ice loss on Arctic climate feedbacks and their role for Arctic amplification. *Geophys. Res. Lett.*, **48**, e2021GL094599, https://doi.org/10.1029/2021GL094599.

Kajtar, J. B., A. Santoso, M. Collins, A. S. Taschetto, M. H. England, and L. M. Frankcombe, 2021: CMIP5 intermodel relationships in the baseline Southern Ocean climate system and with future projections. *Earth's Future*, **9**, e2020EF001873, https://doi.org/10.1029/2020EF001873.

Klein, S. A., and D. L. Hartmann, 1993: The seasonal cycle of low stratiform clouds. *J. Climate*, **6**, 1587–1606, https://doi.org/10.1175/1520-0442(1993)006<1587:TSCOLS>2.0.CO;2.

Lawrence, D. M., and Coauthors, 2019: The Community Land Model Version 5: Description of new features, benchmarking, and impact of forcing uncertainty. *Journal of Advances in Modeling Earth Systems*, **11**, 4245–4287, https://doi.org/10.1029/2018MS001583.

Levermann, A., J. Mignot, S. Nawrath, and S. Rahmstorf, 2007: The role of northern sea ice cover for the weakening of the thermohaline circulation und global warming. *J. Climate*, **20**, 4160–4171, https://doi.org/10.1175/JCLI4232.1.

Li, H. Y., M. S. Wigmosta, H. Wu, M. Y. Huang, Y. H. Ke, A. M. Coleman, and L. R. Leung, 2013: A physically based runoff routing model for land surface and Earth system models. *Journal of Hydrometeorology*, **14**, 808–228, https://doi.org/10.1175/Jhm-D-12-015.1.

Lin, Y.-J., Y.-T. Hwang, P. Ceppi, and J. M. Gregory, 2019: Uncertainty in the evolution of climate feedback traced to strength of the Atlantic meridional overturning circulation. *Geophys. Res. Lett.*, **46**, 12 331–12 339, https://doi.org/10.1029/2019GL083084.

Lin, Y.-J., Y.-T. Hwang, J. Lu, F. Liu, and B. E. J. Rose, 2021: The dominant contribution of southern ocean heat uptake to time-evolving radiative feedback in cesm. *Geophys. Res. Lett.*, **48**, e2021GL093302, https://doi.org/10.1029/2021GL093302.

Lin, Y.-J., B. E. J. Rose, and Y.-T. Hwang, 2023: Mean state AMOC affects AMOC weakening through subsurface warming in the Labrador Sea. *J. Climate*, **32**, 3895–3915, https://doi.org/10.1175/JCLI-D-22-0464.1.

Liu, W., A. Fedorov, and F. Sévellec, 2019: The mechanisms of the Atlantic meridional overturning circulation slowdown induced by Arctic sea ice decline. *J. Climate*, **32**, 977–996, https://doi.org/10.1175/JCLI-D-18-0231.1.

Manabe, S., and W. T. Wetherald, 1967: Thermal equilibrium of the atmopsphere with a given distribution of relative humidity. *J. Atmos. Sci.*, **24**, 241–259, https://doi.org/10.1175/1520-0469(1967)024<0241:TEOTAW>2.0.CO;2.

Manabe, S., and W. T. Wetherald, 1975: The effects of doubling of the CO_2 concentration on the climate of a general circulation model. *J. Atmos. Sci.*, **32**, 3–15, https://doi.org/10.1175/1520-0469(1975)032<0003:TEODTC>2.0.CO;2.

Mitevski, I., C. Orbe, R. Chemke, L. Nazarenko, and L. M. Polvani, 2021: Non-monotonic response of the climate system to abrupt CO_2 forcing. *Geophys. Res. Lett.*, **48**, e2020GL090861, https://doi.org/10.1029/2020GL090861.

Mülmenstädt, J., and Coauthors, 2021: An underestimated negative cloud feedback from cloud lifetime changes. *Nat. Clim. Change*, **11**, 508–513, https://doi.org/10.1038/s41558-021-01038-1.

Paynter, D., T. L. Frölicher, L. W. Horowitz, and L. G. Silvers, 2018: Equilibrium climate sensitivity obtained from multimillennial runs of two GFDL climate models. *J. Geophys. Res. Atmos.*, **123**, 1921–1941, https://doi.org/10.1002/2017JD027885.

Roberts, M. J., and Coauthors, 2020: Sensitivity of the Atlantic Meridional Overturning Circulation to Model Resolution in CMIP6 HighResMIP simulations and implications for future changes. *Journal of Advances in Modeling Earth Systems*, **12**, e2019MS002014, https://doi.org/10.1029/2019MS002014.

Rugenstein, M. A. A., K. Caldeira, and R. Knutti, 2016: Dependence of global radiative feedbacks on evolving patterns of surface heat fluxes. *Geophys. Res. Lett.*, **43**, 9877–9885, https://doi.org/10.1002/2016GL070907.

Rugenstein, M. A. A., and Coauthors, 2020: Equilibrium climate sensitivity estimated by equilibrating climate models. *Geophys. Res. Lett.*, **47**, e2019GL083898, https://doi.org/10.1029/2019GL083898.

Shell, K. M., J. T. Kiehl, and C. A. Shields, 2008: Using the radiative kernel technique to calculate climate feedbacks in NCAR's Community Atmospheric Model. *J. Climate*, **21**, 2269–2282, https://doi.org/10.1175/2007JCLI2044.1.

Sherwood, S. C., and Coauthors, 2020: An assessment of Earth's climate sensitivity using multiple lines of evidence. *Rev. Geophys.*, **58**, e2019RG000678, https://doi.org/10.1029/2019RG000678.

Singh, H., N. Feldl, J. E. Kay, and A. L. Morrison, 2022: Climate sensitivity is sensitive to changes in ocean heat transport. *J. Climate*, **35**, 2653–2674, https://doi.org/10.1175/JCLI-D-21-0674.1.

Smith, R., P. Jones, B. Briegleb, F. Bryan, G., and D. J. Dennis, 2010: The Parallel Oean Program (POP) reference manual, Ocean component of the Community Climate System Model (CCSM). Tech. Rep. LAUR-10-01853, LANL Tech. Report.

Soden, B. J., and I. M. Held, 2006: An assessment of climate feedbacks in coupled ocean-atmosphere models. *J. Climate*, **19**, 3354–3360, https://doi.org/10.1175/JCLI3799.1.

Soden, B. J., I. M. Held, R. Colman, K. M. Shell, J. T. Kiehl, and C. A. Shields, 2008: Quantifying climate feedbacks using radiative kernels. *J. Climate*, **21**, 3504–3520, https://doi.org/10.1175/2007JCLI2110.1.

Stevens, B., S. C. Sherwood, S. Bony, and M. J. Webb, 2016: Prospects for narrowing bounds on Earth's climate sensitivity. *Earth's Future*, **4**, 512–522, https://doi.org/10.1002/2016EF000376.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2009: A summary of the CMIP5 experiment design. Tech. rep. URL https://pcmdi.llnl.gov/mips/cmip5/docs/Taylor_CMIP5_design.pdf?id=98.

Todd, A., and Coauthors, 2020: Ocean-only FAFMIP: Understanding regional patterns of ocean heat content and dynamic sea level change. *Journal of Advances in Modeling Earth Systems*, **12**, e2019MS002027, https://doi.org/10.1029/2019MS002027.

Winton, M., 2006: Surface albedo feedback estimates from the AR4 climate models. *J. Climate*, **19**, 359–365, https://doi.org/10.1175/JCLI3624.1.

Winton, M., W. G. Anderson, T. L. Delworth, S. M. Griffies, W. J. Hurlin, and A. Rosati, 2014: Has coarse ocean resolution biased simulations of transient climate sensitivity? *Geophys. Res. Lett.*, **41**, 8522—-8529, https://doi.org/10.1002/2014GL061523.

Yeager, S., A. Karspeck, and G. Danabasoglu, 2015: Predicted slowdown in the rate of atlantic sea ice loss. *Geophys. Res. Lett.*, **42**, 10 704–10 713, https://doi.org/10.1002/2015GL065364.

Yeager, S., A. Karspeck, G. Danabasoglu, J. Tribbia, and H. Teng, 2012: A decadal prediction case study: late twentieth-century North Atlantic ocean heat content. *J. Climate*, **25**, 5173–5189, https://doi.org/10.1175/JCLI-D-11-00595.1.

Zelinka, M. D., S. A. Klein, Y. Qin, and T. A. Meyers, 2022: Evaluating climate models' cloud feedbacks against expert judgement. *J. Geophys. Res. Atmos.*, **127**, e2021JD35198, https://doi.org/10.1029/2021JD035198.

Zelinka, M. D., T. A. Meyers, D. T. McCoy, S. Po-Chedley, P. M. Caldwell, P. Ceppi, S. A. Klein, and K. E. Taylor, 2020: Causes of higher climate sensitivity in CMIP6 models. *Geophys. Res. Lett.*, **47**, e2019GL085782, https://doi.org/10.1029/2019GL085782.

Zelinka, M. D., C. Zhou, and S. A. Klein, 2016: Insights from a refined decomposition of cloud feedbacks. *Geophys. Res. Lett.*, **43**, 9259–9269, https://doi.org/10.1002/2016GL069917.