UiT The Arctic University of Norway

Faculty of Science and Technology
Department of Physics and Technology

On the time-dependence of climate sensitivity

Kai-Uwe Eiselt

A dissertation for the degree of Philosophiae Doctor

September 2023

Dedicated to the Norwegian taxpayer.

"Ich bin der Ansicht, dass die Menschen, die das große Glück haben, so viel studieren zu dürfen, und sogar für das noch extra bezahlt zu werden, die *Verpflichtung* haben, so zu schreiben, dass jeder, der sich dafür interessiert es verstehen kann."
–Karl Popper, *Das Prinzip der Kritik in der Offenen Gesellschaft*

"I have often thought that the nature of science would be better understood if we called theories 'misconceptions' from the outset, instead of only after we have discovered their successors."
–David Deutsch, *The Beginning of Infinity*

"However, science does not lead anywhere. It can illuminate various courses of action and quantify the risks and tradeoffs. But science cannot make choices for us."
–Judith Curry, *Climate Uncertainty and Risk*

Abstract

Climate sensitivity is defined as the global-mean surface-temperature response to a forcing of the climate system, typically due to a doubling of the CO_2 concentration. It is one of the most important metrics in climate science, and is used at the climate science–policy interface to inform decision-making on mitigation and adaptation. Typically, climate sensitivity is estimated in numerical global climate model (GCM) experiments.

A remarkable result of these GCM experiments was that the climate sensitivity *changes over time*. However, a general explanation remained initially elusive. About ten years ago, the so-called *pattern effect* emerged and has since gained prominence as a general hypothesis to explain this phenomenon. The mechanism of the pattern effect works as follows. In response to a forcing, surface-warming patterns change to favour or disfavour geographical regions of different atmospheric stability, thus activating different *climate feedbacks*, and hence changing climate sensitivity. In most GCMs the climate sensitivity increases over time due to a shift of the surface warming to regions of stronger atmospheric stability, impacting especially the cloud and lapse-rate feedbacks.

The aim of this thesis is to study the time-dependence of climate sensitivity. Aspects of the pattern effect and its impact on climate feedbacks in GCM experiments are investigated. The most important feedbacks in the climate system are induced by changes in the surface albedo, the water-vapour concentration, the temperature lapse rate, and cloud properties. It is found that in fully-coupled GCM experiments the change over time of climate sensitivity is mostly due to changes of the lapse-rate and cloud feedbacks. A main cause for the change in the lapse-rate feedback is differential latitudinal warming, connected to changes in the Atlantic meridional overturning circulation (AMOC) and sea ice. Conversely, cloud feedback changes are primarily induced by shifting warming patterns in the Southern Ocean and the tropics. While dependent on the cloud parametrisation, both effects appear qualitatively consistent across different GCMs, with potentially significant impacts on climate sensitivity.

Historical surface-warming trends, particularly in the Southern Ocean and the East Pacific, are ill-understood and not reproduced in historical GCM simulations. Thus, to increase confidence in estimates of climate sensitivity and its change over time, future research should address the driving mechanisms of changing surface-warming patterns.

Acknowledgements

Time passes, things end. Gestern, heute, morgen. One of the challenges of ageing is coming to terms with the fact that life consists of finite trajectories, and, in the end, is itself a finite trajectory. It is difficult to believe that these four years are now coming to a close, although, in retrospect, two of them seem to contract in a strange way.

Only one name marks the title page, but many people have contributed to the development of this thesis. First and foremost Rune Graversen, my supervisor. Thank you, Rune, for accepting me into the programme, for always being there when I had questions, and for making the publications, the travels, and, in the end, the thesis possible. I really enjoyed discussing the climate system and its physical processes with you, as well as the brooding over the best formulations in our articles.

For the last three years my office mates were Sajidah and Eirik. Thanks Eirik, for the serenity and for inviting me over on the national holiday. Thanks Sajidah, for the many technical and non-technical conversations about your and my work and about the other aspects of PhD-candidate-life. I am also grateful to all the other people in the group for creating a good atmosphere around the office. Some others who are no longer in the group but whom I also wanted to thank are: Sindre, for always having the Norwegian “det blir bra” attitude; Tuomas, for helping me find the volleyball-connection in Tromsø; Patrick, for the many conversations and the interest in and suggestions for my work; and Hege-Beate, for the interesting discussions and for helping me with ideas I was not sure about.

From time to time it is a wise choice for a doctoral student to go on a holiday to take his mind off all the ideas, questions, confusions. I am privileged to have the greatest group of friends to do exactly that. I hope the Denmark holidays will remain a constant also in the future. Special thanks and a tip of the hat go out to Daniel, for being a source of common sense in an increasingly mad-seeming world. Thanks for tolerating my countless long and convoluted messages and for always being *based* in reality. Immer eine Ehre.

Mein letzter Dank geht an meine Geschwister, Eltern und Großeltern. Schwierig vorzustellen, wie es ist, den Bruder, Sohn, Enkel in die große weite Welt davonziehen zu sehen. Ich danke euch für die immerwährende Unterstützung und dass ihr mich immer wieder in der alten Heimat willkommen heißt.

Contents

List of Figures

List of Tables

List of Abbreviations

AA Arctic amplification

AMO Atlantic Multidecadal Oscillation

AMOC Atlantic meridional overturning circulation

AR4 Fourth Assessment Report

AR5 Fifth Assessment Report

AR6 Sixth Assessment Report

CCCma Centre for Climate Modelling and Analysis

CESM Community Earth System Model

CMIP Coupled Model Intercomparison Project

CRE cloud radiative effect

CSIRO Commonwealth Scientific and Industrial Research Organisation

CSLT clear-sky linearity test

EBM energy balance model

EC emergent constraint

ECS equilibrium climate sensitivity

EffCS effective climate sensitivity

EP East Pacific

ERF effective radiative forcing

FAR First Assessment Report

FAT fixed anvil temperature

GSAT global-mean surface-air temperature

GCM global climate model

GFMIP Green's Function Model Intercomparison Project

GHG greenhouse gas

HadCM Hadley Centre Model

IAM Integrated Assessment Model

IPCC Intergovernmental Panel on Climate Change

IPWP Indo-Pacific warm pool

LES large eddy simulation

LGM Last Glacial Maximum

LongRunMIP Long Run Model Intercomparison Project

LR lapse rate

mPWP mid-Pliocene Warm Period

PDF probability density function

PDO Pacific Decadal Oscillation

PETM Paleocene-Eocene Thermal Maximum

PRP partial radiative perturbation

RCP Representative Carbon Pathways

SA surface albedo

SAR Second Assessment Report

SAT surface-air temperature

SOM slab-ocean model

SSP Shared Socioeconomic Pathways

SST sea-surface temperature

TAR Third Assessment Report

TOA top-of-the-atmosphere

UNFCCC United Nations Framework Convention on Climate Change

WCRP World Climate Research Programme

WP West Pacific

WV water vapour

1 Introduction

The capacity of humans to have considerable impact on the Earth's climate system via greenhouse gas (GHG) emissions, such as CO_2, was recognised long ago (e.g., Arrhenius 1896). Especially in recent decades, the issue of climate change and humanity's role in it, together with the potentially large impact and risk it poses, has lead to efforts to establish international cooperation for generating knowledge on the state of the problem and on this basis to implement political action. To this end, in 1988 the Intergovernmental Panel on Climate Change (IPCC)[1] was formed. Starting in 1990 (IPCC 1990), the IPCC has over the last three decades released multiple assessment reports synthesising the state of the knowledge on climate change and the anthropogenic impact, the most recent one being the Sith Assessment Report (AR6, IPCC 2021). Furthermore, several political agreements and treaties on "climate action" such as the United Nations Framework Convention on Climate Change (UNFCCC)[2], the Kyoto Protocol (United Nations 1998), and the Paris Climate Agreement (United Nations 2015), have been signed by multiple countries, intended to implement internationally coordinated political action to reduce anthropogenic GHG emissions and humanity's impact on the climate in general. Climate change has thus become a topic of large impact and interest. This has driven ever-increasing research efforts to understand the climate system, as documented in the IPCC reports, also spurred by technological progress making possible more and more

1. `https://www.ipcc.ch/`, visited 27.06.2023
2. `https://unfccc.int/`, visited 07.08.2023

sophisticated observational instruments (e.g., satellites and radiosondes) as well as increasingly complex climate modelling.

The monumental growth of computational power in recent decades has lead to the possibility of a more and more realistic representation of the climate (or Earth) system in numerical global climate models (GCMs)[3], now often referred to as Earth system models (ESMs), owing to their increased sophistication and inclusion of biogeochemical processes. For clarity, such models will be referred to only as GCMs in this thesis. Fully-coupled GCMs include individual numerical model components of different realms of the climate system, such as the atmosphere, the ocean, the sea ice, and land. These components are connected via a coupler that translates the individual model component output into input for the other components. Hence, GCMs are highly complex systems. One important aspect of GCMs is that they discretise the climate system into vertical and horizontal grids. Due to current computational limitations, the horizontal scale of one grid-cell is typically $O(100)$ km. Since many important processes in the climate system—especially processes related to clouds—occur on much smaller scales, so-called *parametrisations* need to be introduced in a GCM to represent these processes. Parametrisations are empirical relationships that can be used to infer sub-grid scale processes (e.g., cloud formation) from climate variables predicted at scale by the GCM. An example for a very simple cloud parametrisation is to make cloud formation a function of relative humidity: Clouds form when the relative humidity in a grid-cell exceeds a certain threshold (e.g., Parker 2009). Parametrisations in current GCMs are much more sophisticated, indeed taking up most of the computational time of the model (e.g., Palmer and Stevens 2019). However, despite the ever-increasing sophistication, particularly the cloud parametrisation contributes to substantial uncertainty in GCM results (e.g., Zelinka et al. 2020; Paper III, Eiselt and Graversen 2023b; see also section 2.2.4).
Many climate modelling centres around the world have developed their own GCMs, generating a multitude of different numerical climate models and experiments, although at the cost of clarity regarding the results. Under the scope of the World Climate Research Program (WCRP)[4] this has lead to international efforts to standardise model experiments and establish comparability, culminating in the Coupled Model Intercomparison Project (CMIP). Accompanying the development of new climate model releases, several versions of CMIP have been published, the latest of which are CMIP5 (Taylor et al. 2009) and CMIP6 (Eyring et al. 2016). The data are publically available on internet archives[5]

3. Note that the abbreviation "GCM" is also used for general circulation model. However, this typically only refers to models of the atmosphere, while GCM is here used to refer to global climate models irrespective of coupling to an ocean model.
4. `https://www.wcrp-climate.org`, visited 07.08.2023
5. CMIP5: `https://esgf-node.llnl.gov/search/cmip5/`, visited 27.06.2023
CMIP6: `https://esgf-node.llnl.gov/search/cmip6/`, visited 27.06.2023

and were used to inform the IPCC's most recent assessment reports. GCMs are an important tool to study the climate system and its change, and were the principle research tool employed in the work for this thesis.

The main target of the international agreements on curbing emissions is to reduce human-caused global warming. The amount of global-mean surface warming caused by a change of the CO_2 concentration is referred to as *climate sensitivity*. The climate senstivity is effectively an attempt at reducing the complexity of the climate system to one single parameter, thus making it easy to formulate policy targets such as a limitation of global warming to 2 °C or 1.5 °C (IPCC 2018)[6]. The magnitude of the climate sensitivity determines the CO_2-concentration change in the atmosphere that causes such a temperature change (e.g., IPCC 2021, p. 1009), and thus it also determines the emissions target that must be set. Hence, climate sensitivity has become a parameter of great interest, both scientifically and politically, and much depends on its accurate estimation.

Outline

Climate sensitivity, its estimation in GCMs, and issues with its determination are the main focus of this thesis. It is structured as follows. The remainder of chapter 1 begins by introducing the forcing-feedback-sensitivity framework which is used to study climate sensitivity (section 1.1). Then, a short overview of the history of published estimates of climate sensitivity is given (section 1.2), followed by a description of the specific methods for estimating climate sensitivity (section 1.3). Based on the knowledge on the methodology of the estimation, section 1.3 concludes with introducing the problem of the *time-dependence* of climate sensitivity, defining the overarching topic of this thesis. Chapter 2 concerns the physical climate feedbacks acting in the climate system, first describing how they are estimated in GCM experiments (section 2.1) and then explaining the individual feedback processes (2.2). Drawing on the first two chapters, chapter 3 presents the currently most prominent explanation of the time-dependence of climate sensitivity: the so-called "pattern effect". Summaries of the three papers written for this thesis are given in chapter 4. Finally, chapter 5 offers conclusions from the presented research and gives an outlook as well as some recommendations for future research. The papers written for this thesis are attached in chapters 6, 7, and 8.

6. It must be noted that these targets appear to be at least somewhat arbitrary and ill-defined with respect to their baseline (Knutti et al. 2016; for a critical review of the history of these targets see Lüning and Vahrenholt 2017).

1.1 The Forcing-Feedback-Sensitivity Framework

In climate models, climate change is typically studied with forcing experiments. To understand these experiments it is expedient to introduce the *forcing-feedback-sensitivity framework*. Energy in the climate system mostly originates from the sun in the form of short-wave radiation that is intercepted by the Earth. As the solar radiation varies only little on the timescales important here (i.e., the recent century) it can be treated as a constant (S_0)[7]. Because of the Earth's non-zero planetary albedo α some of the solar radiation is directly reflected back to space, and thus effectively does not enter the climate system. The remaining radiation is absorbed as heat into the climate system, increasing its temperature T. The radiation emitted by the Earth can be derived from this temperature via the Stefan-Boltzmann law. A radiative equilibrium is obtained when the net radiation from the sun coming in at the top of the atmosphere (TOA) is equal to the outgoing radiation at TOA from the Earth and can be expressed as:

$$\frac{S_0 - \alpha}{4} = \epsilon \sigma T^4, \tag{1.1}$$

where σ is the Stefan-Boltzmann constant and ϵ is the emissivity, i.e., the degree to which the Earth deviates from a black body ($0 < \epsilon < 1$, with $\epsilon = 1$ representing a perfect black body). With the factor of one fourth on the left-hand side it is taken into account that the solar radiation is intercepted by the cross-section of the Earth while the Earth emits radiation across its spherical surface. Eq. 1.1 can be solved for T to obtain the equilibrium temperature. Notably, this corresponds to the *emission* temperature of the Earth and not its surface temperature. The latter is influenced by the Earth's atmosphere and the so-called greenhouse effect, and thus generally much higher than the emission temperature. An estimate of the surface temperature can be obtained by, e.g., adding an atmospheric layer to the model, again assuming equilibrium conditions and solving for the surface temperature.

A disequilibrium or imbalance of incoming and outgoing radiation at TOA leads to a change in the climate system and is typically referred to as a *forcing*. A forcing constitutes a change to the climate system due to an external factor such as, e.g., a change in the solar radiation, volcanic eruptions, or anthropogenic GHG emissions. Compared to the absolute values of incoming and outgoing radiation, current values of the forcing due to these factors are small and the response of the climate system to a small forcing as also assumed to be small. Typically, in a simple physical model describing the response of the climate

7. There appears to be some debate on this subject (e.g., Connolly et al. 2021; Stefani 2021; see also IPCC 2021, p. 297). However, since it is common practice in GCM historical experiments to implement only small solar-forcing change (Matthes et al. 2017) and for the purposes of clarity and simplicity, this introduction does not go further into detail here.

system to a forcing, the parameter chosen to represent the climate system is the global-mean surface-air temperature (GSAT) change, ΔT_s. A linear expression for the TOA net radiation N in terms of ΔT_s can be derived from a first-order Taylor expansion of $N(\Delta T_s)$ around $\Delta T_s = 0$:

$$N = N(\Delta T_s = 0) + \frac{\partial N}{\partial \Delta T_s}\Delta T_s = F + \lambda \Delta T_s, \tag{1.2}$$

where $N(\Delta T_s = 0)$ corresponds to the TOA imbalance before any climate system response, i.e., the forcing F, and $\frac{\partial N}{\partial \Delta T_s}$ describes how the TOA radiation changes with ΔT_s and is called climate *feedback* λ. Note that in a stable climate system λ is negative, i.e., a temperature increase reduces the TOA imbalance.

The forcing F can either be positive (e.g., GHG concentration increase) or negative (e.g., volcanic eruption) and causes the climate system to respond by adjusting its GSAT. At some point this adjustment leads to a new equilibrium state, i.e., $N = 0$. Given this condition, equation 1.2 can be solved for ΔT_s, thus obtaining the GSAT change at equilibrium. This is typically called equilibrium climate *sensitivity* (ECS):

$$\text{ECS} = -\frac{F}{\lambda}. \tag{1.3}$$

Note that in the common definition of ECS, ice sheet and vegetation are assumed to be constant (Sherwood et al. 2020, p. 4); in the case of non-constant ice sheets the metric is called *Earth system sensitivity* (IPCC 2021, p. 2223). Typically, ECS is estimated in GCM experiments where the CO_2 concentration is instantaneously doubled or quadrupled ("abrupt-forcing experiments") in a pre-industrial control-state climate that is supposed to be in radiative equilibirium[8] (e.g., Charney et al. 1979). Most commonly nowadays, ECS is estimated by running experiments with quadrupled CO_2 concentration and dividing the resulting response by two (Taylor et al. 2009; Eyring et al. 2016). This is a valid approach under the following assumptions. (1) The radiative forcing due to CO_2 scales with the logarithm of the CO_2 concentration. This is a sufficiently accurate approximation (e.g., Romps et al. 2022)[9], and hence:

$$F_{4x} = F_{4x}\frac{F_{2x}}{F_{2x}} = F_{2x}\frac{F_{4x}}{F_{2x}} = F_{2x}\frac{\log 4}{\log 2} = F_{2x}\frac{\log 2^2}{\log 2} = 2F_{2x}, \tag{1.4}$$

8. Note that this is not always the case in the GCMs used for these experiments, which has potential implications for the reliability of the ECS estimates (Sanderson and Rugenstein 2022; see section 5.2). It may even be that for such complex systems as the Earth's climate no equilibirum exists (Loehle 2018).
9. However, when considering larger changes of CO_2 concentration the assumption appears to break down. This needs to be considered when results from climate model experiments with higher CO_2-concentration changes are extrapolated to infer results for lower CO_2-concentration changes and vice versa (e.g., Mitevski et al. 2022)

where F_{2x} and F_{4x} are the radiative forcing at TOA due to doubling and quadrupling of the CO_2 concentration, respectively. The radiative forcing due to a CO_2-concentration doubling is about 4 Wm^{-2} (Charney et al. 1979; Ramanathan et al. 1979; Wilson and Gea-Banacolche 2012) and thus, with each further CO_2 doubling the forcing increases again by 4 Wm^{-2}. (2) A further assumption that is required to be able to estimate ECS from $4\times CO_2$ experiments is that the response of the climate system to a sufficiently small forcing is approximately linear. Given the complexity of the climate system, this should be considered carefully and has been found to be at least somewhat problematic, as higher CO_2 forcings may give rise to processes that remain untriggered at lower forcings, and thus change the equilibrium response (e.g., Mitevski et al. 2021).
However, an advantage of the $4\times CO_2$ experiment over $2\times CO_2$ is the increased signal-to-noise ratio. That is, the CO_2-forcing response emerges more quickly and clearly from the internal variability which potentially masks forced effects if the forcing is low. Moreover, the resulting radiative forcing in $4\times CO_2$ experiments of about 8 Wm^{-2} is similar to that expected at the end of the 21st century in the so-called "business-as-usual" scenarios of the Representative Concentration Pathways (RCP, Vuuren et al. 2011), and more recently in the Shared Socioeconomic Pathways (SSP, Riahi et al. 2017), i.e., RCP8.5 and SSP5-8.5, respectively. It must be noted however, that over the last few years this terminology has been challenged and considerable discussion has ensued about the naming convention (e.g., Field et al. 2021; Hausfather and Peters 2020; Pielke et al. 2022; Pielke and Ritchie 2021a,b; Schwalm et al. 2020; see section 5.2). Indeed, in the AR6 the IPCC acknowledges that, given recent developments, the likelyhood of the RCP8.5 or SSP5-8.5 scenarios is low (IPCC 2021, 238f.), arguably causing the "business-as-usual" terminology to be at least misleading. Nonetheless, in the experimental protocols of CMIP5 and CMIP6 the abrupt4×CO2 experiment is mandatory and intended for the estimation of ECS, while abrupt2×CO2 is not (Eyring et al. 2016; Taylor et al. 2009). Thus, abrupt4×CO2 is available for a large range of fully-coupled climate models and has been widely used to estimate ECS (Andrews et al. 2012; Zelinka et al. 2020).

Having introduced the forcing-feedback-sensitivity framework and the concept of ECS, the next section continues with a brief historical review of estimates of ECS.

1.2 A short history of ECS

Already Arrhenius (1896) estimated the global-mean surface temperature response to a doubling (as well as other factors) of the CO_2 concentration in

the atmosphere (i.e., ECS) based on theoretical calculations including a water vapour feedback (see section 2.2.2), and obtained a value of about 3 to 4 K. Over 80 years later, the seminal Charney et al. (1979) report again estimated ECS using early numerical climate models and arrived at a similar most probable value of about 3 K, but with a range of uncertainty of 1.5 to 4.5 K. It may be noted that rather than being based on a systematic quantification of uncertainty, the Charney et al. (1979) estimate and range are an "educated guess" and are mainly the result of an expert assessment of numerical climate modelling results.

Since then, a large research effort has ensued to narrow down the bounds of ECS, however with limited success. The IPCC's First Assessment Report (FAR) simply adopted the Charney et al. (1979) range, but gave a "best guess" of 2.5 K, however noting that scientists generally were reluctant to give such a value (IPCC 1990, p. 139). By the time of the Second Assessment Report (SAR), the IPCC had found "no strong reasons [...] to change these estimates" (IPCC 1995, p. 34). And again in the Third Assessment Report (TAR) the range of ECS remained unchanged, although with no best estimate explicitly mentioned (IPCC 2001, p. 67). Notably though, in order to estimate the temperature response to different CO_2 concentration changes, the TAR employs simple climate models which are tuned to an ECS of 2.8 K, corresponding to the average over the estimates derived from the GCMs incorporated in the report (IPCC 2001, p. 76).

The first three IPCC assessment reports relied solely on climate models to estimate ECS (IPCC 2021, p. 1007). However, since the Fourth Assessment Report (AR4; IPCC 2007) two more sources have been incorporated: instrumental observational records and paleoclimate evidence (IPCC 2021, p. 1007). The methods to derive ECS from these sources are described in detail in section 1.3. Claiming that the combined assessment of these sources "provides increased confidence" (IPCC 2007, p. 10), the AR4 raised the lower boundary of its ECS estimate, thus lowering its likely ECS range to 2–4.5 K (IPCC 2007, p. 11). The best estimate of 3 K was explicitly provided again (IPCC 2007, p. 11). The difficulty of constraining the bounds of the ECS range was subsequently illustrated when in its Fifth Assessment Report (AR5, IPCC 2013) the IPCC returned to the previous likely range of 1.5–4.5 K because of then new observation-based studies that implied a lower ECS (IPCC 2013, p. 84). The AR5 also refrained from providing a best estimate for the ECS because it found little agreement across the lines of evidence that were assessed (IPCC 2013, p. 85).

In the years since the AR5, increasing efforts to coordinate research on narrowing down the boundaries of ECS and to "achieve more rapid progress" (Stevens et al. 2016, p. 512) have been made, and a change in strategy was proclaimed: Instead of attempting to derive particular values of ECS and building consen-

sus, the focus should shift to ruling out values via a process of hypothesis testing using multiple lines of evidence (Stevens et al. 2016), hereby rooting ECS research more in the fundamental tenets of scientific method (Popper 2002 [1959]). This has culminated in another seminal and monumental review paper incorporating process understanding, historical observations, paleoclimatological evidence, and so-called emergent constraints into a *formal*[10] Bayesian framework to quantify the probability density function of ECS (Sherwood et al. 2020). Notably, GCMs were not used to *directly* generate ECS estimates, reflecting the fact that the range of ECS derived from the latest generation of GCMs has indeed *increased* compared to earlier generations (Zelinka et al. 2020; see also Fig. 1.1).

With their strategy of rejecting unlikely outcomes and their formal Bayesian framework, Sherwood et al. (2020) managed to reduce the *likely* (IPCC terminolgy; Mastrandrea et al. 2010) range of ECS to 2.6–3.9 K. It may be noted that Lewis (2023) using a similar but updated method arrives at a lower and narrower estimate (17–83% range: 1.75–2.7 K), although Sherwood and Forest (2023, p. 3) claim that this resulted from "selective use of evidence". Following Sherwood et al. (2020) as well as the recognition that some CMIP6 members are running "implausibly hot" (Voosen 2021, p. 474; see also Hausfather et al. 2022; Scafetta 2022; Voosen 2019; Zhu et al. 2020), the AR6, in contrast to its predecessors, does not consider direct estimates from climate models in its assessed ECS range (IPCC 2021, pp. 1005, 1007)[11]. However, it may be said that the "process understanding line of evidence builds on and replaces ESM estimates" (IPCC 2021, p. 1008) and, more generally, all the lines of evidence are informed by GCM results (see also section 1.3). Thus, GCMs still feature heavily in the AR6.

Unlike the AR5, the AR6 finds broad agreement among the assessed lines of evidence and thus again gives a best estimate of 3 K and a narrowed *likely* range of 2.5 to 4 K. It may be noted with regards to the political target of 1.5 K that according to the AR6, it is *virtually certain* (> 99%) that ECS is larger than 1.5 K (IPCC 2021, p. 1006). However, ECS should not be equated with the warming by 2100. This is discussed in more detail in chapter 3 (see e.g. also, Roe and Bauman 2013).

Because of its high complexity, the task of narrowing the bounds of climate sensitivity will likely remain a challenge for the foreseeable future. More research on the different sources employed to estimate climate sensitivity will help con-

10. While AR4 and AR5 incorporated instrumental and paleoclimate evidence in addition to GCMs, no *formal* method was used to combine these lines of evidence.
11. Around the topic of "too hot" models and how to treat them there has ensued some discussion, since an "implausibly" high ECS of a given model does not necessarily mean the model should be discounted (Bloch-Johnson et al. 2022; Sanderson and Rugenstein 2022)

strain their uncertainties and this will in turn help to decrease the uncertainty of ECS estimates based on these sources. Better process understanding based on theory, climate models, and observations will allow to synthesise the lines of evidence and support a narrowing of the bounds of ECS. The next section explains in detail how ECS can be estimated from the different sources of evidence.

1.3 Estimating ECS

ECS can be estimated from three sources (e.g., IPCC 2021, p. 992): It can be inferred from (1) instrumental observational records, (2) paleoclimatological evidence, and it can be directly estimated from (3) GCM experiments. Instrumental and paleoclimate data can be combined with GCMs to obtain emergent constraints that may also be used to estimate ECS. In the following these methods are described. Special focus will be given to method (3), as this was the method employed and studied in the present thesis.

1.3.1 Inference from Observational Records

Estimating ECS from observational data is most primarily done based on an energy budget analysis (e.g., Gregory et al. 2002; Otto et al. 2013; Lewis and Curry 2015, 2018, see also IPCC 2021, pp. 995-997). This method employs the forcing-feedback-sensitivity framework as represented in eq. 1.2. Assuming a forcing due to doubling of the CO_2 concentration (F_{2x}), eqs. 1.2 and 1.3 can be combined to eliminate λ and thus yield (e.g., Otto et al. 2013):

$$\mathrm{ECS} = \frac{F_{2x} \times \Delta T_s}{F - N}. \tag{1.5}$$

Note that this implicitly assumes that the feedback parameter λ is constant. Several sources can be queried to obtain estimates for the quantities on the right-hand side of eq. 1.5. They involve instrumental temperature records, the Argo floats (e.g., Roemmich et al. 2019) for the Earth's energy imbalance, as well as GCMs to estimate internal variability (IPCC 2021, p. 996). It may thus be emphasised that GCMs partly inform ECS estimates from observational records.

Generally, estimates of from this method fall on the lower end or even below the likely range as assessed by the IPCC (e.g., Forster 2016; Knutti et al. 2017, IPCC 2021, p. 996). The likely reason for this is the assumption of a constant feedback parameter, thus neglecting a potential change toward a less negative

feedback, which is projected for a future warming (IPCC 2021, p. 996; see section 1.3.3 and chapter 3).

Further methods to infer ECS from observational data include estimation based on the variation of the Earth's TOA radation budget and the response to volcanic eruptions. However, these methods appear to rather infer the short-term response of the climate system and are thus less well suited to directly derive ECS (e.g., Knutti et al. 2017; Sherwood and Forest 2023). For further discussion the reader is referred to the AR6 (IPCC 2021, pp. 998-999), Knutti et al. (2017), and Sherwood et al. (2020).

1.3.2 Inference from Paleoclimate Records

A further source from which estimates of ECS may be derived is paleoclimatological evidence. Here, reconstructions of temperature and radiative forcings from past cold and warm periods are considered. The advantages of the paleoclimate-method over estimates based on the instrumental record (section 1.3.1) are that periods can be chosen where (1) the forcing was large (comparable to CO_2 doubling) and (2) the following response period is sufficiently long that the GSAT stabilises so that a quasi-equilibrium state can be assumed (e.g., IPCC 2021, p. 999, Sherwood et al. 2020, p. 52). Furthermore, the advantage over estimates from GCMs is that paleoclimate records are real-world observation-based data (IPCC 2021, p. 999). However, disadvantages of this method are that paleoclimatological evidence is indirect and proxy-based, which introduces substantial uncertainty. This uncertainty increases the further back in time the analysis extends, as more and more factors need to be considered, such as GHG concentrations, aerosols, vegetation and land surface, topography, land and sea ice, and even continent shape, that may all be different from today's conditions (Sherwood et al. 2020, p. 52). Care must be taken with terminology here: As mentioned in section 1.1, ECS is defined assuming non-changing ice sheets. However, on paleclimatological time scales ice sheets did change and must thus be treated as a forcing when estimating ECS from paleoclimate data (IPCC 2021, p. 999). Another problem with paleoclimatological evidence is that it is mostly highly localised, implying that especially for globally inhomogeneous parameters such as, e.g., sea level or temperature, some amount of climate modelling is necessary to obtain global gridded data (e.g., Sherwood et al. 2020, p. 53). Thus, as with ECS estimates inferred from observations, it may again be emphasised that estimates inferred from paleoclimatological evidence are at least partially informed by GCMs.

Based on the assumption of quasi-equilibrium in a given paleoclimatological period, eq. 1.2 can be reduced to eq. 1.3, and employing the paleoclimate-proxy-based estimates of forcing and temperature a value for ECS can be derived.

This is done for different periods, which, owing to their specific conditions, have different advantages and disadvantages. One of the most widely studied periods is the Last Glacial Maximum (LGM; 23,000-19,000 years ago; e.g., Tierney et al. 2020; Sherwood et al. 2020, pp. 54-57; IPCC 2021, p. 1000). Other periods are the mid-Pliocene Warm Period (mPWP; 3.3-3.0 million years ago; Sherwood et al. 2020, pp. 57-59), or the Paleocene-Eocene Thermal Maximum (PETM; 56 million years ago; Sherwood et al. 2020, pp. 59-61). However, the uncertainties for these warm periods are potentially large and may thus only yield weak constraints on ECS.

ECS estimates based on paleoclimatological evidence appear to fall on the upper end of the range assessed as likely by the AR6 (Knutti et al. 2017, IPCC 2021, pp. 1001-1002), with especially the upper bound being ill-constrained.

Emergent constraints

Both the instrumental record and paleoclimate data can be used to generate so-called "emergent constraints" (ECs) to potentially reduce the range of ECS estimates. This is a relatively new method[12] where in a first step an *observable* characteristic A of the climate system (e.g., the GSAT, the short-wave cloud radiative effect, etc.) is related to ECS across an ensemble of GCMs. The second step is to estimate characteristic A from the observational historical or paleoclimate record to derive a plausible probability density function (PDF) for A. In a third and final step, the GCMs which produce values of A outside the observationally derived PDF for A are discounted, and hereby an observationally constrained estimate of ECS from an ensemble of GCMs is obtained. Thus, observationally constraining ECS from GCMs is essentially a kind of model-weighting (Sherwood et al. 2020, p. 16; Sanderson et al. 2021).
Generally, ECs may be devided into to categories (IPCC 2021, p. 1004; Sanderson et al. 2021): Approaches based (1) on (near-)global indices and (2) on individual process isolation. Studies such as Nijsse et al. (2020) and Tokarska et al. (2020) belong to category 1, and constrain ECS in GCMs based on their fidelity in reproducing the observed late-twentieth century global-mean surface temperature rise. Category 2 includes studies such as Cesana and Del Genio (2021) and Myers et al. (2021), where the short-wave cloud response is constrained in a set of GCMs using satellite observations and so-called "cloud controlling factors". Since short-wave cloud feedback is well-correlated with ECS across GCMs (Caldwell et al. 2018), the constrained short-wave cloud response can be used to constrain ECS.
Several issues have been raised concerning the EC methodology. Generally,

12. According to the review by Brient (2020), the concept was first introduced by Allen and Ingram (2002).

the confidence in a given EC depends on how well the underlying physical mechanism is understood, on the uncertainty of the observational data, and on the robustness of the present-to-future relationship (e.g., Sherwood et al. 2020, p. 16; Brient 2020). A fundamental problem arising from the EC methodolgy is that "bad" GCMs (i.e., outliers with respect to observationally constrained characteristic A) may artificially increase the robustness of the relationship between ECS and A (T. Schneider 2018). However, this may potentially be remedied by more sophisticated statistical methods (Brient 2020). A further general issue with the EC methodolgy is discussed in Sanderson et al. (2021): Confidence in ECs and even a given EC itself may arise as an artifact of common structural assumptions (e.g., similar parametrisations) in the chosen GCM ensemble. This may only be discovered after a new GCM generation becomes available and the ECs from the previous generation are found to not apply anymore (e.g., Caldwell et al. 2018; Schlund et al. 2020). Sanderson et al. (2021, p. 906) conclude that using a single EC to directly constrain projected outcomes can at best be viewed as "a conditional statement that could be proved inaccurate or overconfident by the following generation of models." It should be noted, however, that *all* scientific statements are conditional on the assumptions made to derive them, and they may be proved inaccurate by new findings (Popper 2002 [1959]).

As mentioned above, the EC method is relatively new and several attempts at remedying these issues have been proposed. These include the application of more sophisticated statistical methods (Brient 2020), the combination of multiple ECs (Bretherton and Caldwell 2020), and a transition from a "top-down" to a "bottom-up" approach, focusing more on constraining observable climate characteristics (e.g., cloud radiative flux) instead of high-level properties, such as ECS (Sanderson et al. 2021). Moreover, in his extensive review of the EC literature, Brient (2020) comes to the conclusion that because of disagreements between different ECs and because of their large uncertainties, they have so far not been successful at narrowing the bounds of ECS. Similarly, the AR6 arrives at a likely ECS range of 1.5 to 5 K based on ECs (IPCC 2021, p. 1005), being even larger than the original Charney et al. (1979) range.

As of now, instead of directly constraining ECS, the best application of ECs may be to uncover processes which control intermodel spread and to investigate if process representation in GCMs is adequately diverse, thus potentially informing GCM development (Sanderson et al. 2021).

1.3.3 Direct Estimation From GCM Experiments

ECS can be estimated from abrupt-forcing GCM experiments (section 1.1) in two ways which are described in the following.

Equilibrating Climate Models

The straight-foreward way to estimate ECS in a GCM is to implement a forcing (e.g., a CO_2-doubling) in an experiment in an equilibrium state and then to integrate the model until a new equilibirum is obtained (Paynter et al. 2018; Rugenstein et al. 2020). That is, the model is run until it has warmed sufficiently that the TOA imbalance again approaches zero when averaged over several decades. However, as fully-coupled GCMs include a dynamically active deep ocean with a large heat capacity and thus thermal inertia, this may take up to multiple thousands of simulation years (e.g., Sanderson and Rugenstein 2022). The resulting computational costs still remain prohibitive for establishing organised efforts to run GCMs to equilibirium. However, recently with the Long Run Model Intercomparison Project (LongRunMIP; Rugenstein et al. 2019) progress in this direction has been made. In contrast to e.g. CMIP, though, this is not a coordinated effort with standardised experiments run explicitly for LongRunMIP, but an opportunistic enterprise trying to collect individual experiments that have been conducted without coordination.

A way to circumvent the problem of having to run a fully-coupled GCM for thousands of simulation years to obtain the ECS is to use a slab-ocean model (SOM; e.g., Bitz et al. 2012; Danabasoglu and Gent 2009) instead of a fully-coupled GCM. In a SOM, the dynamical numerical ocean-model component (including the deep ocean) is replaced by a thermally uniform mixed-layer slab with a temporally constant but spatially varying depth and a seasonally and spatially varying lateral heat flux, the so-called Q-flux. Lacking the dynamical ocean component, a SOM obtains a new equilibrium in response to a forcing much faster than a fully-coupled GCM (about 30 to 50 simulation years). Hence, estimating ECS from equilibrating a SOM is computationally much cheaper than from equilibrating a fully-coupled GCM. However, the lack of a dynamical ocean response in a SOM means that potential changes in ocean circulation due to the forcing are neglected. Since such changes may influence the pattern of surface warming which has potentially important consequences for climate feedback (Shell 2012; see chapter 3), this may bias the ECS estimate from a SOM. Aspects of this are investigated in Paper II (Eiselt and Graversen 2023a) and Paper III. However, ECS estimates from SOMs are often shown to be similar to ECS estimated from (close-to) equilibrated fully-coupled GCMs (e.g., Bacmeister et al. 2020; Danabasoglu and Gent 2009; Dunne et al. 2020; Li et al. 2013).

Extrapolation From Shorter Experiments

The second method to estimate ECS from fully-coupled GCM experiments is to run shorter abrupt-forcing experiments on the order of 100 years and then

to "extrapolate" to a new equilibrium. This evades the problem of prohibitive computational costs, but comes with several issues. Essentially, in this method the ECS is estimated by estimating the forcing F and the feedback λ from the model response and then applying eq. 1.3 to obtain ECS. However, it is not obvious how either of these parameters should be exactly defined or how they can be derived from GCM output. This is problematic as the ECS estimate sensitively depends on F and λ (e.g., Rugenstein and Armour 2021).

Hansen et al. (2005) summarised many of the possible definitions of forcing (see their Fig. 2; see also C. J. Smith et al. 2020), the simplest one being the instantaneous flux imbalance at the tropopause after the CO_2 concentration is changed. However, the most widely used method to estimate radiative forcing in climate models is to diagnose the TOA radiative flux imbalance after so-called "adjustments" (Sherwood et al. 2015) on the order of a few years in the stratosphere and the troposphere have been allowed to occur. The forcing thus derived is typically referred to as "effective radiative forcing" (ERF). The point of allowing for these adjustments is to isolate the direct effect of the forcing on climate system characteristics such as temperature profile, humidity, or cloud from the effects mediated by surface temperature changes, i.e., climate feedbacks (Hansen et al. 2005; Sherwood et al. 2015; see chapter 2).
The ERF for a climate model can be derived by running a fixed-SST/sea ice[13] experiment with, e.g., doubled CO_2 concentration (ERF_{2x}). Note that land temperatures are allowed to evolve freely in these experiments which is inconsistent with the adjustments not being mediated by temperature changes. This is a pragmatic choice since prescribing the land temperatures in a climate model is technically challenging, but has been shown to cause non-negligible biases in the ERF estimates (e.g., Andrews et al. 2021). A typical way to circumvent this problem is by employing correctional procedures to adjust the estimates from fixed-SST/sea ice experiments for the land-temperature influence (e.g., Hansen et al. 2005; C. J. Smith et al. 2020).
Having thus obtained ERF_{2x}, the climate feedback parameter λ in abrupt CO_2-doubling experiments in GCMs can be calculated by dividing ERF_{2x} by the GSAT change ΔT_s at, e.g., year 150 (or a multi-year average to account for internal variability). Rugenstein and Armour (2021) refer to this as the "effective feedback parameter" (λ_{eff}; see also Senior and Mitchell 2000). Climate sensitivity can be estimated using ERF_{2x} and λ_{eff} in eq. 1.3, but one must be aware that this assumes *linearity* of the GCM response to the forcing, which is not necessarily given (see below and chapter 3). Thus, the result does not generally correspond to the "true" ECS,[14] and is often referred to as the "effective climate sensitivity" (EffCS; Rugenstein and Armour 2021).

13. In such experiments the sea-surface temperatures (SSTs) and the sea ice are held constant.
14. That is, the ECS that would be obtained after equilibration.

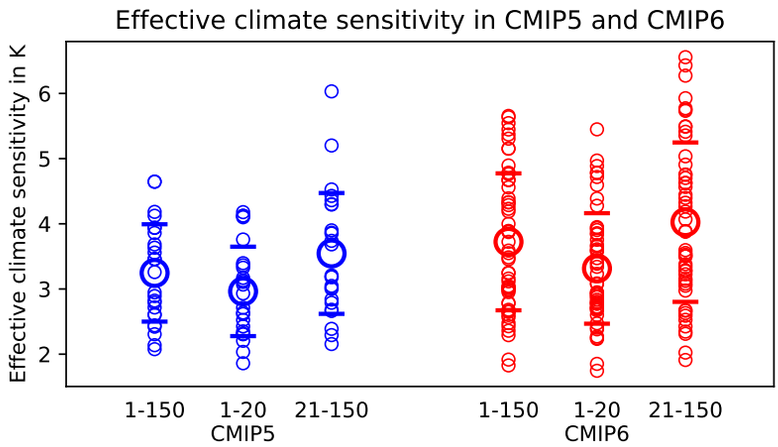

Figure 1.1: Estimates of effective climate sensitivity (EffCS) derived with the Gregory method from the abrupt4xCO2 experiment for (blue) 25 members of CMIP5 and (red) 52 members of CMIP6. To obtain EffCS corresponding to a doubling of the CO_2 concentration, values were divided by two (see text for details). Shown are values for the whole simulation period (years 1-150) and the "early" and "late" periods (years 1-20 and 21-150, respectively). The larger circles correspond to the multi-model mean and the horizontal lines indicate the 1-σ standard deviation across models.

A further method for estimating ECS from a shorter GCM integration is fitting an energy balance model (EBM) to the GCM output (e.g., Murphy 1995; Geoffroy et al. 2013; Geoffroy and Saint-Martin 2020). This involves first solving the EBM equations and then finding ways to fit the parameters of the solution to the GCM output, hereby being a more elaborate effort. However, the method offers the possibility of implementing several different components of the climate system with different heat capacities (e.g., mixed-layer ocean and troposphere), thus representing different time scales of the GCM. Hence, estimates from EBMs may more closely resemble the "true" ECS, since they do not assume linearity of the GCM response.

The most widley used method to estimate climate sensitivity, feedback, and forcing based on abrupt-forcing experiments is known as the Gregory method (Gregory et al. 2004). It makes use of the fact that the simple climate model (eq. 1.2) lends itself well to a linear regression. GCMs have as standard output quantities the TOA radiation and the surface-air temperature (SAT). The Gregory method consists in linearly regressing the global-mean TOA net radiation (N) on the global-mean SAT (GSAT) as shown in Fig. 1.2 for three different GCMs. Accordingly, a scatter plot of N against the GSAT is typically referred to as Gregory plot. The regression yields direct estimates of the forcing (the y-intercept) and the feedback (the slope). Using eq. 1.3, i.e., extrapolating the

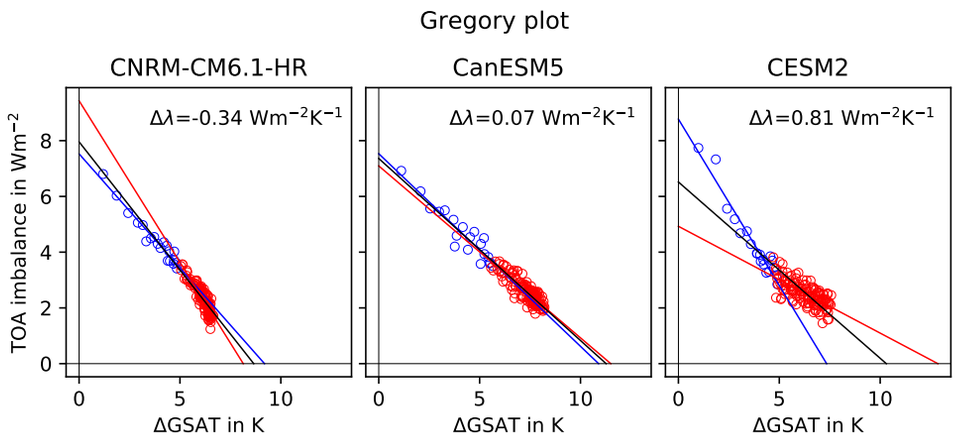

Figure 1.2: Gregory plots for three CMIP6 abrupt4xCO2 experiment members: (left) CNRM-CM6.1-HR with a negative feedback change, (middle) CanESM5 with close to zero feedback change, and (right) CESM2 with a positive feedback change. Years 1-20 of the simulation are shown as blue circles, years 21-150 are shown as red circles. The blue, red, and black lines represent the Gregory regressions for year 1-20, 21-150, and 1-150, respectively. For the parameters derived from the Gregory method see Table 1.1.

relationship to a new equlibirium, yields an estimate of ECS. Again however, this estimate is not the "true" ECS and it is thus also referred to as "effective climate sensitivity" (EffCS; Rugenstein and Armour 2021). Estimates of EffCS for 25 members of CMIP5 and 52 members of CMIP6 obtained from the Gregory method are shown in Fig. 1.1. The forcing derived with the Gregory method is also typically called "effective radiative forcing" (in the following ERF*) since it allows for atmospheric and stratospheric adjustments as described above (Gregory et al. 2004; Hansen et al. 2005; C. J. Smith et al. 2020). Hansen et al. (2005) report that estimates of ERF* based on the first 10 to 30 years of their model experiments yield a good measure of ERF based on other methods (e.g., fixed-SST/sea ice experiments). For abrupt-forcing experiments such as abrupt4xCO2 it has become standard to estimate ERF* based on the first 20 years of the GCM integration (e.g., Andrews et al. 2012; C. J. Smith et al. 2020; Zelinka et al. 2020; IPCC 2013, p. 818). The climate feedback parameter estimated with the Gregory method may be called "differential feedback parameter" ($\tilde{\lambda}$) fullfilling the following relationship (Gregory et al. 2004; Rugenstein and Armour 2021):

$$\tilde{\lambda} = \frac{\partial N}{\partial \Delta T_s}. \tag{1.6}$$

Notably, $\tilde{\lambda}$ is in general not equal to λ_{eff} (Rugenstein and Armour 2021).

EffCS derived from the Gregory method is only a robust estimate of ECS if λ is constant over time. This used to be the general assumption (Boer and Yu 2003b; Gregory et al. 2004), in part because until and including the AR4 (IPCC 2007) most ECS estimates were based on the method of equilibrating slab-ocean models (SOMs), as explained above (see e.g. IPCC 2013, p. 817). The constancy of climate feedback and sensitivity is also implicit in the protocol for the CMIP5 and CMIP6 abrupt4xCO2 experiment, since the requested minimum number of simulation years is 150 (Eyring et al. 2016; Taylor et al. 2009)[15]. Also in more recent studies ECS is estimated with the Gregory method assuming a constant climate feedback (IPCC 2013, p. 817; Zelinka et al. 2020). However, already Murphy (1995) using an early version of the fully-coupled Hadley Centre atmosphere-ocean climate model found that estimating climate feedback and sensitivity with an EBM calibrated to the *earlier* decades in a CO_2-forcing experiment yielded significantly different results than when the EBM was calibrated to the *later* decades. Specifically, climate sensitivity based on years 1-30 of the simulation was 1 K, while based on years 46-75 it was 2.8 K (Murphy 1995). That is, the climate sensitivity increased and the climate feedback became less negative over time. However, in subsequent studies it was found that this is model dependent. While both Senior and Mitchell (2000) and Gregory et al. (2004) in later versions of the Hadley Centre model (HadCM2 and HadCM3, respectively) confirm the increase of climate sensitivity, Watterson (2000) finds little variation of climate sensitivity over time in the Commonwealth Scientific and Industrial Research Organisation (CSIRO) fully-coupled model, and Boer and Yu (2003a) even find a decrease in climate sensitivity in the Centre for Climate Modelling and Analysis (CCCma) fully-coupled model. Figure 1.2 illustrates the three cases by showing the 150-year Gregory plots for models with strengthening feedback (CNRM-CM6.1-HR), roughly constant feedback (CanESM5), and weakening feedback (CESM2). The corresponding values of ERF*, λ, and EffCS are summarised in Table 1.1. In a Gregory plot it is directly apparent that the climate feedback in a given model is not constant over time if the dots do not lie on a straight line. It is typical to divide the 150 years in an early (years 1-20) and a late (years 21-150) period (Andrews et al. 2012) as done in Figs. 1.1 and 1.2 and Table 1.1.

Due to increasing computational resources, in more recent years ECS has been estimated from dozens of fully-coupled GCMs. It has been found that in most models the climate feedback becomes less negative over time, i.e., climate sensitivity increases (e.g., Andrews et al. 2012; Dong et al. 2020; Fig. 1.1). As this may significantly impact the robustness of the standard method of estimating ECS (i.e., the Gregory method applied to 150-year GCM abrupt4xCO2 results), the *time-dependence of climate sensitivity* has now been recognised as one of the most important issues in climate change research, and a large research

15. Rugenstein and Armour (2021) find that 400 years of simulation may be required to estimate ECS to with 5% of its "true" value.

	CNRM-CM6.1-HR	CanESM5	CESM2	units
ERF^*_e	7.53 ± 0.17	7.53 ± 0.26	8.77 ± 0.28	Wm^{-2}
$\tilde{\lambda}_e$	-0.82 ± 0.04	-0.69 ± 0.06	-1.19 ± 0.07	$Wm^{-2}K^{-1}$
$EffCS_e$	9.19	10.9	7.35	K
ERF^*_l	9.43 ± 0.31	7.09 ± 0.23	4.93 ± 0.22	Wm^{-2}
$\tilde{\lambda}_l$	-1.16 ± 0.05	-0.62 ± 0.03	-0.38 ± 0.04	$Wm^{-2}K^{-1}$
$EffCS_l$	8.14	11.5	12.86	K
ERF^*	7.96 ± 0.12	7.36 ± 0.11	6.52 ± 0.18	Wm^{-2}
$\tilde{\lambda}$	-0.92 ± 0.02	-0.65 ± 0.02	-0.63 ± 0.03	$Wm^{-2}K^{-1}$
EffCS	8.66	11.27	10.3	K

Table 1.1: Parameters derived with the Gregory method for the three climate models shown in Fig. 1.2. Subscripts e and l refer to periods early (years 1-20) and late (years 21-150), respectively. No subscript indicates the whole period (years 1-150). The uncertainties correspond to the 1-σ standard deviation.

effort has commenced to understand the underlying reasons. Contributing to this effort is the main aim of this thesis.

The explanation of the the time-dependence of climate sensitivity requires a detailed understanding of the climate system and specifically of the individual physical feedback processes that act to influence the climate system's response to a forcing. Chapter 2 gives an overview of these feedback processes and explains how they impact climate sensitivity. Moreover, it is explained how these feedbacks can be derived from climate model output and estimates are presented based on a suite of members of CMIP5 and CMIP6 and compared to the latest assessment by the IPCC in the AR6 (IPCC 2021, pp. 967-992). Chapter 3 then returns to the time-dependence of climate sensitivity and explains this phenomenon in terms of the pattern effect and physical climate feedbacks based on the current best understanding and recent research.

/2

Climate Feedbacks

The concept of *feedback* is essential in the study of dynamical systems such as the Earth's climate. In the broadest sense, a feedback may be defined as a "process that, when included in the system, makes the forcing a function of the response; in other words, some fraction of the output is fed back into the input" (Roe 2009, p. 96). For the purpose of climate dynamics in this thesis, a feedback here corresponds to the "changes of the net energy budget at the top-of-atmosphere (TOA) in response to a change in the GSAT" (IPCC 2021, p. 967):

$$\lambda_x = \frac{\partial N}{\partial x}\frac{dx}{dT_s}, \tag{2.1}$$

where λ_x is the feedback due to climate variable x. A positive feedback is then a process that increases the GSAT response to a forcing, while a negative feedback inhibits the response. The specific feedbacks considered here are typically called "physical feedbacks" (e.g., IPCC 2021, p. 967) and include the feedbacks due to changes in surface albedo (SA), water-vapour (WV) concentration, temperature lapse rate (LR), and cloud properties. The temperature lapse rate is typically defined as the deviation of the vertical temperature profile from the SAT. The feedback due to the increase (decrease) in outgoing radiation resulting from rising (falling) temperatures is called Planck feedback. It is typically defined by assuming a uniform temperature response (equivalent to the SAT response) across the atmospheric column, and the sum of the LR and the Planck feedbacks is usually called the temperature feedback.

Besides physical feedbacks, the climate system includes biogeochemical and biogeophysical feedbacks associated, e.g., with vegetation or ozone, as well as long-term feedbacks associated with ice sheets (e.g., IPCC 2021, p. 967). These feedbacks are not considered here since they were not included in the traditional definition of ECS and because they act on longer time scales. Thus, these feedbacks are often called "Earth system feedbacks" associated with the "Earth system sensitivity" in contrast to ECS (e.g., IPCC 2021, pp. 967, 2223; section 1.1). It may be noted here, that the AR6 finds that biogeochemical and biogeophysical feedbacks almost cancel each other, implying a small combined impact, however with low confidence (IPCC 2021, pp. 976-977).

According to Roe (2009), the idea of feedbacks in the climate system goes back at least as far as Croll (1864). As mentioned in section 1.2, Arrhenius (1896) in his estimate of ECS takes into account the fact that the surface warming due to a CO_2 forcing would be amplified by a change in the WV concentration *caused* by the surface warming, i.e., a WV feedback. Early climate modelling studies also describe the WV (Manabe and Wetherald 1967) as well as the SA feedback (e.g., Budyko 1969; Sellers 1969; Manabe and Wetherald 1975), the LR feedback (e.g., Cess 1975; Manabe and Wetherald 1975), and cloud feedbacks (e.g., Cess 1975; Manabe and Wetherald 1967).
While some of the first comprehensive reviews of climate feedback mechanisms were conducted by S. H. Schneider and Kellogg (1973) and S. H. Schneider and Dickinson (1974), the first comprehensive *quantification* of all the mentioned physical climate feedbacks appears to have been undertaken by Hansen et al. (1984), using a slab-ocean model. However, as Roe (2009, p. 95) observes, at the time quantification was rarely attempted and the literature was "confusing and, in some places, flat out contradictory", at least when it comes to terminology.
Three methods were typically used to quantify climate feedbacks from climate models: (1) The partial radiative perturbation (PRP) method, where one climate variable is changed at a time from a control state and then the TOA radiative flux change is calculated offline (Wetherald and Manabe 1988; Soden et al. 2008). (2) The cutting feedback loops method (Hall and Manabe 1999; Colman and Soden 2021), where a certain feedback process is suppressed in a climate model experiment, the response of which is then compared to another experiment where the feedback is not suppressed. Examples are the locking of the WV concentration (e.g., Hall and Manabe 1999), the SA or the temperature LR (e.g., Graversen et al. 2014). (3) The SST-perturbation method, where a climate model is run with prescribed SST changes and the resulting TOA radiative flux changes are investigated (Cess et al. 1990).
However, the problem with method (1) is that it is computationally expensive and its implementation is complex, potentially leading to differences in feedback calculations by different users. Method (2) is similarly challenging and may involve decorrelation issues when calculating the radiation (Colman and Soden 2021). Finally, the problem with method (3) is that it only distinguishes

between cloud and non-cloud feedbacks, thus not providing estimates for the other feedbacks such as those due to changes in WV concentration, temperature LR, or SA changes (Soden et al. 2008). Thus, wide-spread quantification of climate feedbacks was hampered by methodological inadequacies.

2.1 The Radiative Kernel Method

A simple and general method for quantifying feedbacks from climate model experiments was introduced by Soden et al. (2008) and is known as the *radiative kernel method*. This method is based on the two following assumptions: (1) the TOA net radiation can be described as a function of independent climate system variables and (2) the radiative flux change due to a small change in these variables is linear. Based on these assumptions, the climate feedback parameter λ in eq. 1.2 can be decomposed into contributions from the individual climate feedbacks mentioned above:

$$N = F + (\lambda_{Pl} + \lambda_{LR} + \lambda_{WV} + \lambda_{SA} + \lambda_{C})\Delta T_s, \tag{2.2}$$

where the subscripts *Pl*, *LR*, *WV*, *SA*, and *C* denote the Planck, lapse-rate, water-vapour, surface-albedo, and cloud feedbacks, respectively.

The so-called radiative kernel for a climate variable can then be derived as follows. In a climate model in an equilibrium state the respective climate variable is changed by some amount and then only the radiation code of the climate model is run. The change in the TOA radiative flux per unit variable change thus calculated is the radiative kernel K for the changed variable. The kernel may then be used to derive the TOA radiative flux change due to the change in the given variable in another climate model experiment (e.g., abrupt4xCO2): by multiplying the kernel on the change of the variable in that experiment.
As an example, the SA radiative kernel may be derived by changing the surface albedo a in a climate model in the pre-industrial control state by $\delta a = 1\%$ and then only running the radiation code. The change in the TOA radiative flux per δa, i.e., the SA radiative kernel K^a, may then be defined as:

$$K^a \equiv \frac{N(a + \delta a, T, w, c) - N(a, T, w, c)}{\delta a} \approx \frac{\partial N}{\partial a}(a, T, w, c) \tag{2.3}$$

where N is the TOA radiative flux as a function of surface albedo a, temperature T, water-vapour concentration w, and a set of cloud properties c. If in a given model experiment (e.g., abrupt4xCO2) the SA changes by 30%, this can be multiplied on K^a to obtain the TOA radiative flux change due to the SA change. The SA feedback parameter λ_{SA} can then be derived by expressing the TOA

radiative flux response to the SA change as:

$$N(\delta a) = \lambda_{SA} \Delta T_s, \tag{2.4}$$

and applying the Gregory method (see section 1.3.3). That is, the TOA flux change due to the SA change is linearly regressed against the GSAT change and the SA feedback parameter corresponds to the slope of this linear regression (e.g., Block and Mauritsen 2013; see Fig. 2.1b). The other kernels and feedbacks can be obtained in the same way, except for the cloud feedback. Cloud effects are strongly non-linear and thus the radiative kernel method as described here is inappropriate. However, within the radiative kernel method framework the cloud feedback can be estimated by adjusting the cloud radiative effect (CRE) by the difference between the all-sky and clear-sky radiative fluxes due to the other feedbacks (Soden et al. 2008). The CRE is defined as the difference between the TOA all-sky (N_{as}) and TOA clear-sky (N_{cs}) radiative flux:

$$\text{CRE} = N_{as} - N_{cs}, \tag{2.5}$$

where fluxes are treated positive downward. The TOA radiative flux change due to clouds $N(\delta c)$ may then be obtained by (Soden et al. 2008):

$$\begin{aligned} N(\delta c) = \text{CRE} + (K^a_{cs} - K^a_{as})\delta a + (K^T_{cs} - K^T_{as})\delta T + \\ (K^w_{cs} - K^w_{as})\delta w + (F^{4x}_{cs} - F^{4x}_{as}), \end{aligned} \tag{2.6}$$

where the subscripts *as* and *cs* denote all-sky and clear-sky, respectively, and F^{4x} represents the forcing due to 4xCO_2.

Even though the radiative kernels are derived from a single model, they are assumed to be applicable to the output of other climate models as well, since the radiation codes of climate models are well-tested and similar across models (e.g., Shell et al. 2008; Soden et al. 2008)[1]. The applicability of the same set of radiative kernels across different climate models and experiments is potentially the biggest advantage of the kernel method compared to the other methods, as this makes feedback quantification from multi-model ensembles such as CMIP computationally much cheaper.
Several radiative kernels based on different climate models (or reanalyses) have been produced (e.g., Soden et al. 2008; Shell et al. 2008; Block and Mauritsen 2013; Huang et al. 2017; C. J. Smith et al. 2018; Pendergrass et al. 2018), and although there is some variation, they tend to yield similar feedbacks (e.g., Soden et al. 2008; Zelinka et al. 2020), confirming their applicability across models.

1. However, e.g., Hahn et al. (2021) find considerable variation of the SA feedback across radiative kernels especially in the polar regions (see sections 2.2.2 and 2.2.3 for more discussion).

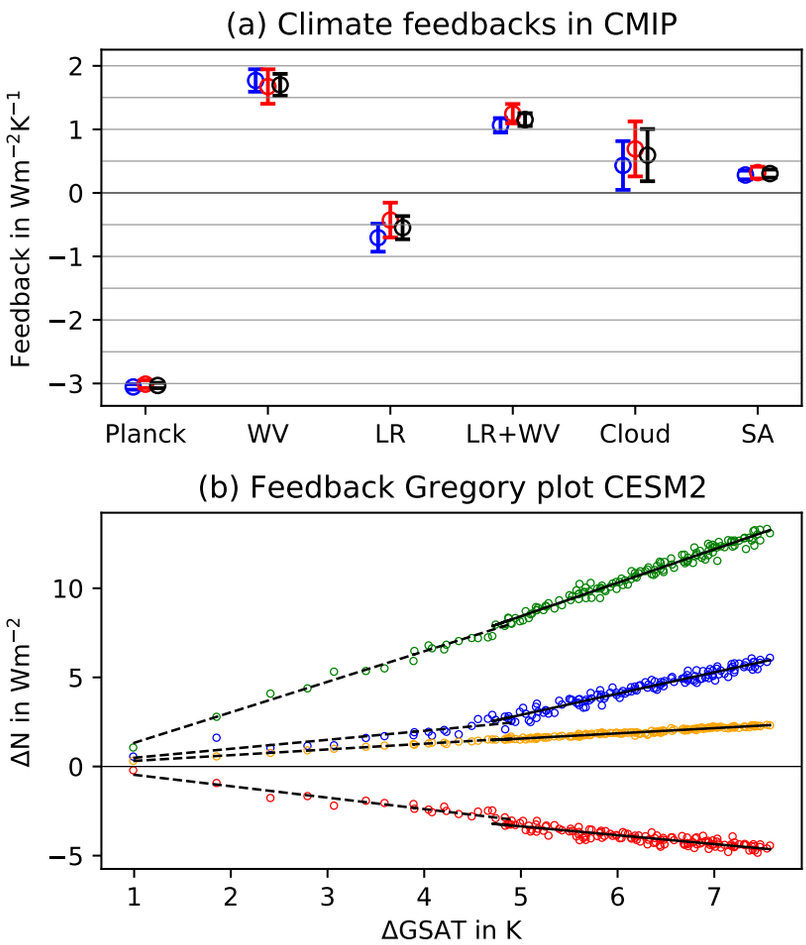

Figure 2.1: (a) Climate feedbacks for 37 members of CMIP5 and CMIP6 derived from the radiative kernels provided by Shell et al. (2008). Depicted are the (blue) early, (red) late, and (black) full period. The bars denote the 1-σ standard deviation from the multi-model mean. (b) Gregory plot of the radiative-kernel-derived TOA radiative fluxes due to the physical feedbacks (except Planck) in CESM2. The black dashed and solid lines correspond to the Gregory regressions for the early and late period, respectively. The values for the feedbacks are listed in Table 2.1.

Lacking corresponding results from the other methods for deriving climate feedbacks—such as the PRP or cutting feedback loops—there is no straightforward procedure for testing the quality of the results of the kernel method for a given GCM. However, as an auxiliary benchmark, Shell et al. (2008) introduced the "clear-sky linearity test" (CSLT). Here, the TOA fluxes due to individual feedbacks as derived from the *clear-sky* radiative kernels are added up to yield the total clear-sky TOA imbalance. This is then compared to the GCM output clear-sky TOA imbalance and if the difference remains below a certain threshold, the set of kernels is accepted as applicable.
Note that directly comparing radiative fluxes, as done by Shell et al. (2008), additionally requires an estimate of the CO_2-forcing, as this impacts the TOA radiative imbalance. As described in section 1.3.3, deriving CO_2-forcing estimates is not straightforward and thus adds a source of potential error to this method. However, this can be avoided by comparing feedbacks instead of radiative fluxes. That is, the sum of the individual feedback parameters derived from the kernel method may be compared to the total feedback parameter derived with the Gregory method applied to the GCM output radiative fluxes. Hence, no estimate of the CO_2-forcing is required. This method is applied, e.g., by Ceppi and Gregory (2017) and in Paper I (Eiselt and Graversen 2022). Note that in Paper I this leads to the rejection of the results of almost 50% of the models that were analysed, thus considerably reducing the data base. Other studies circumvent this problem by considering a residual term. That is, they keep all models regardless of the difference between kernel-derived and model-output fluxes but they explicitly show this difference (i.e., the residual) when showing the feedback results (e.g., Dong et al. 2020; Zelinka et al. 2020).

The advent of the radiative kernel method substantially simplified the quantification of climate feedbacks from GCMs, thus causing an inflation of studies undertaking such quantifications over the last decade (see e.g., the research referenced in Sherwood et al. 2020 and IPCC 2021), including the publications presented in this thesis. Figure 2.1a and Table 2.1 show the climate feedbacks derived with the Shell et al. (2008) kernels in the abrupt4xCO2 experiment averaged over 37 members of CMIP5 and CMIP6. The following sections give an overview of how the individual feedbacks function to influence the climate system response to a forcing.

2.2 Physical Climate Feedbacks

The feedbacks discussed in the following are the Planck feedback (2.2.1), the water-vapour (WV) and the lapse-rate (LR) feedback (2.2.2), the surface-albedo (SA) feedback (2.2.3), and the cloud feedback (2.2.4). Their underlying physical processes are explained. Employing two members of the CMIP6 abrupt4xCO2

	CMIP	CESM2	INM-CM4.8
λ_{Pl}	-3.03 ± 0.04	-3.06 ± 0.0	-3.05 ± 0.01
λ^e_{Pl}	-3.05 ± 0.04	-3.01 ± 0.01	-3.12 ± 0.02
λ^l_{Pl}	-3.01 ± 0.06	-3.06 ± 0.0	-3.03 ± 0.01
$\Delta\lambda_{Pl}$	0.04 ± 0.06	-0.03	0.09
λ_{WV}	1.70 ± 0.17	1.87 ± 0.02	1.53 ± 0.02
λ^e_{WV}	1.77 ± 0.18	1.71 ± 0.05	1.66 ± 0.04
λ^l_{WV}	1.67 ± 0.27	1.88 ± 0.03	1.49 ± 0.04
$\Delta\lambda_{WV}$	-0.1 ± 0.29	0.16	-0.17
λ_{LR}	-0.55 ± 0.18	-0.62 ± 0.02	-0.14 ± 0.03
λ^e_{LR}	-0.71 ± 0.22	-0.64 ± 0.05	-0.49 ± 0.06
λ^l_{LR}	-0.43 ± 0.27	-0.50 ± 0.02	0.15 ± 0.04
$\Delta\lambda_{LR}$	0.28 ± 0.35	0.14	0.63
λ_{LR+WV}	1.15 ± 0.1	1.25 ± 0.02	1.38 ± 0.02
λ^e_{LR+WV}	1.06 ± 0.11	1.07 ± 0.03	1.17 ± 0.05
λ^l_{LR+WV}	1.25 ± 0.15	1.38 ± 0.03	1.64 ± 0.03
$\Delta\lambda_{LR+WV}$	0.18 ± 0.15	0.31	0.47
λ_{SA}	0.30 ± 0.06	0.30 ± 0.0	0.27 ± 0.01
λ^e_{SA}	0.28 ± 0.07	0.32 ± 0.01	0.22 ± 0.01
λ^l_{SA}	0.32 ± 0.09	0.29 ± 0.01	0.32 ± 0.01
$\Delta\lambda_{SA}$	0.04 ± 0.11	-0.03	0.1
λ_C	0.59 ± 0.41	1.03 ± 0.02	-0.07 ± 0.03
λ^e_C	0.43 ± 0.38	0.50 ± 0.06	-0.05 ± 0.05
λ^l_C	0.69 ± 0.43	1.19 ± 0.03	0.08 ± 0.05
$\Delta\lambda_C$	0.26 ± 0.25	0.69	0.13

Table 2.1: Feedback estimates in $Wm^{-2}K^{-1}$ derived from the radiative kernel method using the Shell et al. (2008) kernels. The values in the column "CMIP" correspond to the average and 1-σ standard deviation across 37 members of CMIP5 and CMIP6 (as in Fig. 2.1a). The uncertainties for CESM2 and INM-CM4.8 correspond to the 1-σ standard deviation of the Gregory regression.

experiment (CESM2 and INM-CM-4.8)[2], spatial characteristics of the feedbacks are highlighted and a potential change over time relating to the "pattern effect", as indicated in section 1.3.3, is discussed. The same nomenclature as in section 1.3.3 is adopted, referring to simulation years 1-20 as "early period", simulation years 21-150 as "late period", and to late minus early as "change over time". The full 150-year simulation is referred to as "total".
Finally, values for the individual feedback parameters are given based on 37 members of the CMIP5 and CMIP6 abrupt4xCO2 experiment (see Table 2.1 and Fig. 2.1a). The values are briefly compared to the estimates from the recent comprehensive assessments in Sherwood et al. (2020) and in the AR6 (IPCC 2021, pp. 967-992). Notably, while up until the AR5 (IPCC 2013) only GCM estimates of feedbacks were considered, the AR6, following Sherwood et al. (2020), also considers observational and fine-scale model results (IPCC 2021, p. 967). That is, the estimates of the feedbacks are informed by multiple lines of evidence.

2.2.1 Planck Feedback

The Planck feedback is the feedback that would exist were the atmosphere to warm vertically uniformly with the surface. That is, it is the feedback resulting from assuming ΔT_s in all atmospheric layers. Thus, the Planck feedback can be estimated by substituting the Stefan-Boltzmann law (see eq. 1.1) for N in eq. 2.1 and recognising that in the case of the Planck feedback $x = T_s$:

$$\lambda_{Pl} = \frac{\partial N}{\partial T_s}\frac{dT_s}{dT_s} = \frac{\partial N}{\partial T_s} = \frac{\partial \left[-\epsilon\sigma T_s^4\right]}{\partial T_s} = -4\epsilon\sigma T_s^3. \tag{2.7}$$

Note that the minus sign is required because the Stefan-Boltzmann law gives the *outgoing* radiation, but N is defined here as positive downward. For present-day values this yields a Planck feedback of $\lambda_{Pl} \approx -3.3$ $\mathrm{Wm^{-2}K^{-1}}$ (e.g., IPCC 2021, p. 968; Sherwood et al. 2020, p. 19). The value derived with the radiative kernel method from GCMs is slightly lower (-3.2 $\mathrm{Wm^{-2}K^{-1}}$), which may be due to the assumed cloud distribution in the radiative kernels, as well as water-vapour and surface-warming patterns, influencing the emissivity (e.g., IPCC 2021, p. 968; Sherwood et al. 2020, p. 19). The estimates shown in Fig. 2.1a and Table 2.1 are still lower, which may be due to the fact that they were derived with one specific set of radiative kernels (Shell et al. 2008). Indeed, using another set of radiative kernels (Pendergrass et al. 2018) yields a CMIP average of -3.2 $\mathrm{Wm^{-2}K^{-1}}$. As can be seen in Fig. 2.1a and in Table 2.1, the variation of the Planck feedback across climate models is small compared to the other

2. The model descriptions for CESM2 and INM-CM4.8 are given in Danabasoglu et al. (2020) and Volodin et al. (2018), respectively.

feedbacks. Furthermore, comparing the early, late, and total estimates of the Planck feedback reveals that this feedback changes little over time.

As mentioned above, the Planck feedback may be considered as the feedback that the climate system would have if no other feedbacks were to exist. The other feedbacks thus act to change the Planck feedback, such that the total climate feedback λ differs from λ_{Pl}[3]. They are discussed in the following.

2.2.2 Water-Vapour and Lapse-Rate Feedback

The WV and the LR feedback are often considered together (see e.g. Hansen et al. 1984; Sherwood et al. 2020, p. 20; IPCC 2021, pp. 969-970) because they are both strongly influenced by the development and distribution of water vapour in the atmosphere. However, their relationship varies strongly regionally, also impacting their change over time. Thus, here they are first treated separately, after which their combined characteristics are considered. Note that the focus here is on the traditional WV amount feedback and not the relative humidity feedback introduced by Held and Shell (2012).

Water-Vapour Feedback

The WV feedback was already included in the early climate sensitivity estimate by Arrhenius (1896) and, after some dormancy, its importance was again put into view by Möller (1963). At that time also the idea of the so-called "runaway greenhouse" (see below) was considered, prompted by investigations into the atmosphere of the planet Venus (Held and Soden 2000). However, because of methodological errors, Möller (1963) overestimated the effect of the WV feedback, which was corrected with the introduction of radiative-convective models by Manabe and Wetherald (1967)[4].

The WV feedback results from the fact that due to the Clausius-Clapeyron relation of temperature and WV saturation pressure the atmosphere can contain more water in gaseous form (i.e., water vapour) at higher temperature than at lower temperature. If under a warming the relative humidity of the atmosphere remains roughly constant—which appears to be a sound assump-

3. This framework is equivalent to that applied in Sherwood et al. (2020) and in the AR6 (IPCC 2021), but differs from Roe (2009) where the other climate feedbacks are expressed as adjustment *factors* to the Planck feedback.
4. This paper is regarded as one of the most influential papers in climate science (Colman and Soden 2021). For this and his adjacent climate modelling work, Syukuro Manabe was awarded the shared Nobel Prize in Physics in 2021. (`https://www.nobelprize.org/prizes/physics/2021/summary/`. Visited 27.07.2023.)

tion according to observations and climate model experiments (e.g., Sherwood et al. 2020, p. 20)—this leads to an increase of atmospheric specific humidity, i.e., an increased WV concentration. Since WV is an effective GHG, it acts to inhibit the outgoing long-wave radiation, thus warming the surface and constituting a positive climate feedback. Increased WV concentration also absorbs more short-wave radiation, although the effect on long-wave radiation is much stronger (e.g., Soden et al. 2008). Notably, according to the Clausius-Clapeyron equation the saturation pressure increases slightly less than exponentially with temperature (Held and Soden 2000).

The WV feedback exhibits a vertical structure because changes of the WV concentration in the middle and upper troposphere are more effective at trapping outgoing radiation than in the lower troposphere. This is because the effect of WV in the lower troposphere is masked by WV and clouds aloft (e.g., Held and Soden 2000; Soden et al. 2008).
The horizontal structure of the WV feedback is shown for two members of the CMIP6 abrupt4xCO2 experiment in Fig. 2.2. Although the horizontal structure differs between these two GCMs, two common features are apparent: (1) the WV feedback is larger in the tropics than in the extratropics and (2) the WV feedback is almost exclusively positive. Indeed, as concluded by Held and Soden (2000), there is no empirical or model evidence that the WV feedback is negative. As shown in Fig. 2.1a, in the CMIP model mean the WV is the largest positive physical climate feedback and as noted in Sherwood et al. (2020, p. 20), because of good general agreement between climate model results, observational evidence, and thermodynamic theory, there is high confidence that the WV feedback is positive. The horizontal structure of the WV feedback is at least partly explained by the self-broadening of the WV absorption lines under tropical conditions (Soden et al. 2008). The horizontal structure of the WV feedback may also be related to the vertical distribution of the WV: In the tropics ascending motion elevates WV to higher levels than at higher latitudes, where it remains at lower levels because of more stable atmospheric stratification. As explained above, WV at higher altitude is more effective at trapping outgoing long-wave radiation than at lower altitude. Thus, the WV feedback is more effective in the tropics than in the extratropics.

As can be observed in Figs. 2.1, 2.2, and Table 2.1, the WV feedback may exhibit considerable change over time, both in the global mean and in the feedback pattern. In the CMIP mean considered here, the WV feedback weakens slightly over time, which is at odds with the conclusions in the recent review on WV and LR feedback by Colman and Soden (2021). However, it is clear that the change over time varies considerably across models, again both in the global mean (Table 2.1) and in the feedback pattern (Fig. 2.2), potentially explaining the discrepancy.
There also appears to be some dependence on methodological and time-scale

differences. Block and Mauritsen (2013) find that the WV feedback and its change over time depend on the climate state from which the radiative kernels are generated. They produce two sets of radiative kernels, one based on the pre-industrial control climate, and another based on a $4xCO_2$ climate. They argue that the control kernels may be considered more adequate for the early period (years 1-20) and the $4xCO_2$ kernels for the late period (years 21-150) in an abrupt4xCO2 experiment. Comparing this two-kernel-sets application to the standard method of applying control-state kernels to both early and late periods (as done in this thesis), they find significant differences: For the standard method the WV feedback weakens, corresponding to the results presented here, while for the two-kernel-sets method the WV feedback strengthens. Notably, the opposite is observed for the SA feedback (see section 2.2.3).

The differences of WV feedback and its change between the two models shown in Fig. 2.2 are related to surface-warming patterns (Po-Chedley et al. 2018) and their changes over time, which are themselves related to changes in oceanic and atmospheric circulation and transport, such as the Atlantic meridional overturning circulation (AMOC) and the Hadley circulation, respectively. This is discussed in greater detail in chapter 3 and in Papers I, II, and III. As explained above, cloud masking also impacts the WV feedback (Soden et al. 2008). Comparing Figs. 2.2 and 2.6 indicates that cloud changes may partly be responsible for the changes in the WV feedback pattern.

The WV feedback is the largest individual physical feedback in the current climate. The central estimate of the total-period WV feedback parameter averaged over the CMIP models considered here is 1.70 $Wm^{-2}K^{-1}$, in good agreement with the estimates given in the AR6, referencing both modelling and observational studies (IPCC 2021, p. 969)[5].

As mentioned above, the WV feedback has the theoretical potential to drive the climate system into a runaway greenhouse effect and it is not immediately obvious that this is not the case (Held and Soden 2000). However, according to Wallace and Hobbs (2006, p. 447), in order for a runaway greenhouse to be initiated, the tropical SSTs would first have to be heated from currently about 28 °C to 60 °C.

Lapse-Rate Feedback

The LR feedback is the only consistently globally negative physical feedback (see Fig. 2.1). Although not specifically referred to as lapse-rate feedback it appears to have first been described in Manabe and Wetherald (1975). The first

5. Sherwood et al. (2020) only provide a combined estimate of WV and LR feedback.

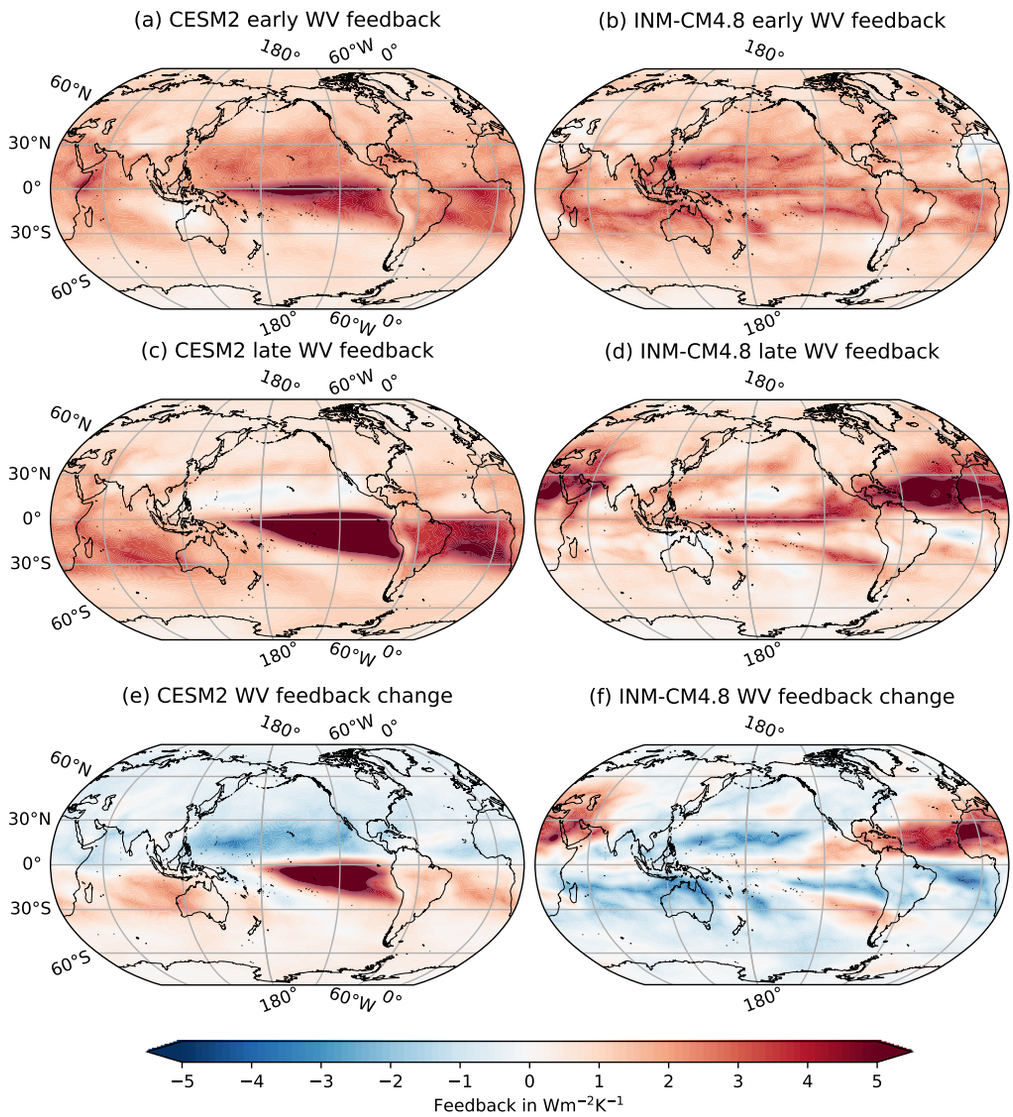

Figure 2.2: Maps of the water-vapour feedback for (a, c, e) CESM2 and (b, d, f) INM-CM4.8. Shown are (a, b) the early (years 1-20) and (c, d) late (years 21-150) feedbacks, as well as (e, f) the feedback change (late minus early). The top-of-atmosphere radiative fluxes were calculated with the radiative kernels provided by Shell et al. (2008) (see text for details).

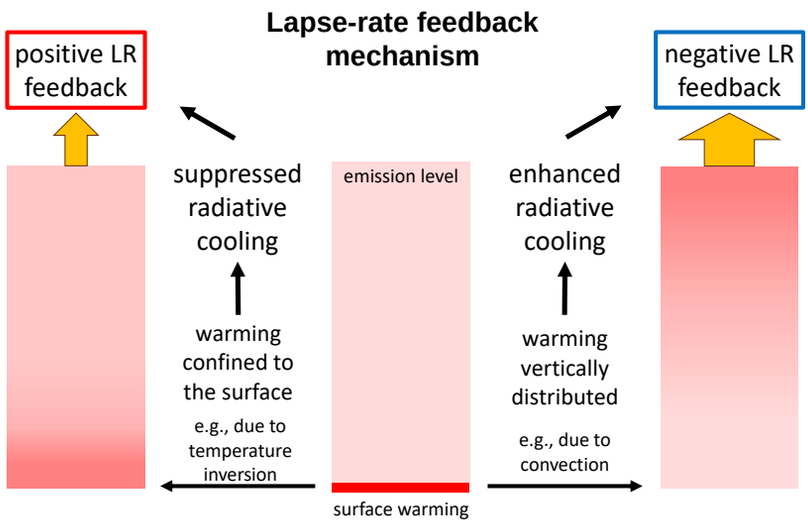

Figure 2.3: Illustration of the lapse-rate (LR) feedback. If a (e.g., CO_2-induced) surface warming is confined to the surface (e.g., due to a temperature inversion), the atmosphere aloft warms less, thus suppressing the radiative cooling and constituting a positive feedback (left). Conversely, if a surface warming is spread throughout the atmospheric column (e.g., due to convection), the atmosphere warms more than the surface and the radiative cooling is enhanced, implying a negative feedback (right).

explicit mention of the "lapse-rate feedback" seems to occur in Cess (1975), although there it is indeed dismissed as insignificant.

The LR feedback orginates if the vertical structure of warming in an atmospheric column is non-uniform (Cess 1975; Manabe and Wetherald 1975). If the surface warms more than the atmosphere aloft, a positive feedback results because the radiation emitted by the whole column is weaker than it would be if it had warmed at the same rate as the surface. If the situation is reversed, a negative feedback results. An illustration of this is shown in Fig. 2.3.

The spatial structure of the LR feedback as seen in Fig. 2.4 was already qualitatively described by Manabe and Wetherald (1975): The feedback is negative in the tropics and becomes positive towards the polar regions. This is because in the tropics a surface warming is efficiently spread vertically due to convective ascending motion, thus warming the free troposphere more than the surface. There the excess energy associated with the warming is more efficiently radiated to space, implying a negative feedback. Conversely, at high latitudes the atmosphere is strongly stably stratified and hence the surface and the free troposphere are coupled less. Thus, a surface warming is confined to lower levels, inhibiting warming aloft, implying a positive feedback. However, aver-

aged globally, the tropical effect dominates the high-latitude effect, resulting in a consistently negative feedback (e.g., Table 2.1, Fig. 2.1; IPCC 2021, p. 969). The central estimate of -0.55 $Wm^{-2}K^{-1}$ for the CMIP member average given in Table 2.1 for the total-period LR feedback parameter is in good agreement with the value estimated by the AR6 based on climate models and observations (IPCC 2021, p. 969)[6]. There is some disagreement about the value between climate models (see Fig. 2.1a and Table 2.1) which will be discussed further below.

It is clear from Figs. 2.1a and 2.4 that the LR feedback may exhibit significant change over time. Indeed, while in some models cloud feedback changes dominate (e.g., CESM2, see Figs. 2.1b and 2.6), for several other models the LR feedback change dominates the total feedback change over time (e.g., Paper I; Armour et al. 2013). In the CMIP mean presented here, the LR feedback change over time is positive, again—as with the WV feedback—being at odds with the conclusions of Colman and Soden (2021). The reasons for these discrepancies were discussed in the section on WV feedback above.

As examples for LR feedback patterns and pattern changes, results from CESM2 and INM-CM4.8 are shown in Fig. 2.4 (values in Table 2.1). Like the WV feedback change, the LR feedback change is associated with spatial pattern changes (Fig. 2.4). Again, these are related to surface-warming changes and oceanic and atmospheric transport (Andrews and Webb 2018; Ceppi and Gregory 2017; Singh et al. 2022; Paper II). However, unlike the WV feedback, the LR feedback exhibits strong changes also in the Arctic. This is related to the melting of sea ice (e.g., Manabe and Wetherald 1975; Graversen et al. 2014; Feldl et al. 2020; Boeke et al. 2021; Jenkins and Dai 2021; Dai and Jenkins 2023; Paper I). Sea ice acts to insulate the atmosphere from the ocean, which in the polar regions is a heat source for the atmosphere. Hence, sea ice inhibits thermal exchange and allows the surface above the sea ice to obtain very low temperatures. The Arctic region is thus strongly stably stratified. Under a CO_2-induced warming, the melting of the sea ice enables the exchange of heat between the atmosphere and the much warmer ocean, thus warming the surface. Due to the stable stratification, however, the warming is confined at the surface, causing a strong positive LR feedback. Comparing panels (e) and (f) in Fig. 2.4 makes clear that the change of the LR feedback in the Arctic varies strongly across models (see also Paper I). This is again related to the sea-ice development. CESM2 generally warms much quicker than INM-CM4.8, thus losing most of its Arctic sea ice already in the early period, causing a larger LR feedback due to the explained mechanism. However, in the later period, since most of the sea ice has melted, only little further melt is possible in CESM2, while more melting can occur in INM-CM4.8, thus reversing the effect on the LR feedback. Notably

6. Again, Sherwood et al. (2020) only provide a combined LR and WV feedback estimate.

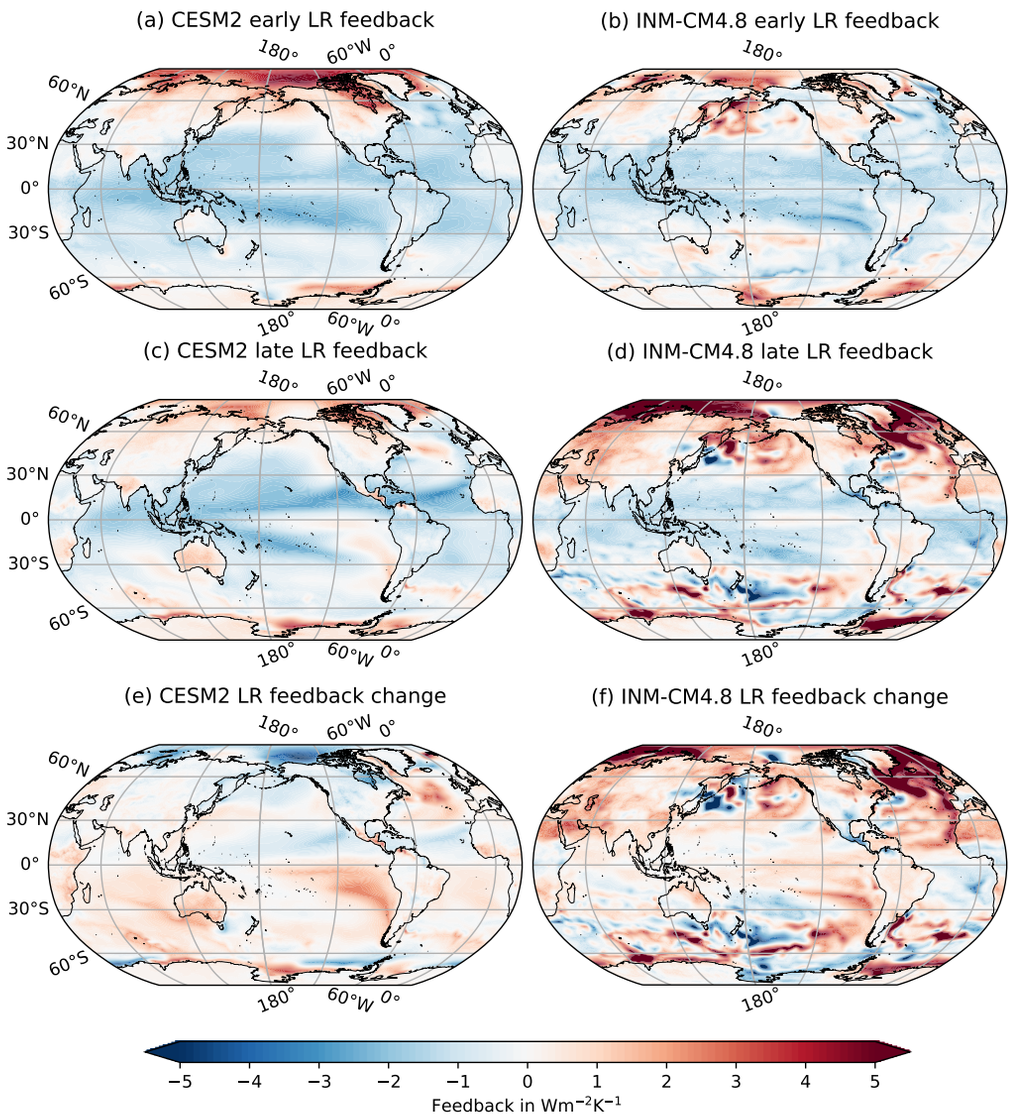

Figure 2.4: As Fig. 2.2 but for the lapse-rate (LR) feedback.

though, this presents a hen-and-egg problem, since the faster melting of sea ice and the resulting feedbacks may be at least one of the *causes* rather than a *result* of the faster warming in CESM2 compared to INM-CM4.8 (see also Paper I).
The pace of the melting of the sea ice may be related to atmospheric and oceanic heat transport. In both CESM2 and INM-CM4.8, the North Atlantic stands out as a region with significant LR feedback change (Fig. 2.4), indicating an impact of the Atlantic meridional overturning circulation (AMOC). The AMOC transports warm water from the tropics into the North Atlantic, thus contributing to locally higher surface temperatures than in other regions at the same latitude. In response to a CO_2 forcing, the AMOC is generally predicted to decline (e.g., Bellomo et al. 2021), impacting the surface warming and thus LR feedback in the North Atlantic. Less northern polar surface warming due to AMOC decline may also hinder sea-ice melt, again influencing the Arctic LR as well as the SA feedback (see also section 2.2.3). This is further discussed in chapter 3 and in Paper II.

The LR feedback has been shown to be an important factor for the so-called *Arctic amplification* (AA), i.e., the accelerated surface warming observed in the Arctic under global warming (Hahn et al. 2021; Pithan and Mauritsen 2014; Stuecker et al. 2018). However, some scientific debate seems to exist regarding the mechanism and the importance of the LR feedback for the AA. This may be connected to methodological as well as terminological issues. While some studies find the LR feedback to be the most important factor for the AA (Pithan and Mauritsen 2014; Stuecker et al. 2018), other studies find its influence to be comparatively weak, at least in isolation (Graversen et al. 2014; Jenkins and Dai 2021). Dai and Jenkins (2023) find that the LR feedback alone is not sufficent to drive significant AA. However, because of the interaction and reciprocal amplification of individual feedbacks (e.g., Graversen et al. 2014; Hansen et al. 1984; Roe 2009), the LR feedback may still be an important contributor to the AA, even if in isolation its impact is weak. Notably, Alexeev (2003) and Alexeev et al. (2005) find that even without the effect of sea-ice melt appreciable AA ensues. Also, regarding studies using radiative kernels to quantify feedbacks it has been found by Hahn et al. (2021) that the relative contributions to the AA from the LR and SA feedbacks depend on the choice of the radiative kernels, possibly contributing to the lack of consensus. Furthermore, confusion may be caused by the specific feedback decompositions and terminology used in the literature. For example, Dai and Jenkins (2023) combine the increased ocean heat release in the winter season due to retreating sea ice with the traditional SA feedback to a “sea-ice feedback”. While self-consistent within studies, the terminological inconsistency across studies may contribute to the debate about the subject.
Pertaining to causal and mechanistic understanding, the importance of atmospheric stability for the AA has been highlighted by several studies (Bintanja

et al. 2011; Pithan and Mauritsen 2014). However, more recent work has found that both the stability and LR feedback changes in the Arctic are controlled by surface warming from the ocean uncovered due to sea-ice melt, and that stability itself is not a strong driver of the LR feedback (Boeke et al. 2021; Dai and Jenkins 2023; Feldl et al. 2020; Jenkins and Dai 2021). According to Dai and Jenkins (2023), the traditional framing of LR feedback and the AA—that a CO_2-induced surface warming is amplified in the Arctic due to stable stratification (Manabe and Wetherald 1975; Hansen et al. 1984)—is at least incomplete and a more accurate framing is: The initial CO_2-induced warming causes some sea-ice melt which uncovers the underlying much warmer ocean. It is the heat release from the uncovered ocean, then, that mainly causes the amplified surface warming, in turn amplified by the stable stratification in the Arctic, confining the ocean-induced warming to the surface (Dai and Jenkins 2023).

Except for the Arctic and some other regions, the general pattern of the LR feedback seems to be mostly inverse to the pattern of the WV feedback (cp. Figs. 2.4 and 2.2), indicating an anticorrelation of these feedbacks. The relation between the LR and WV feedback as well as their combined effect is discussed in the following.

Combined Lapse-Rate and Water-Vapour Feedback

Because of their causal mechanisms, the WV and the LR feedback are closely connected. Indeed, what is referred to as the effect due to water vapour in the early modelling study by Manabe and Wetherald (1967) is in current terminology the combined WV+LR feedback. When the LR feedback was explicitly referred to by Cess (1975), it was believed to not have a global impact, because it would be offset by compensating WV feedback changes due to increased specific humidity under the warming. However, a different and more subtle picture has since emerged. On a global level the WV and LR feedbacks are indeed significantly anticorrelated (Po-Chedley et al. 2018). This may also be seen in the reduced uncertainty of the WV+LR feedback compared to the individual WV and LR feedbacks in Fig. 2.1a. More detailed analysis reveals, that in the tropics, these feedbacks are *not* correlated (Armour et al. 2013). It has been found that this is because while the tropical WV feedback is determined by changes in relative humidity, the tropical LR feedback is unrelated to relative humidity (Po-Chedley et al. 2018). Conversely, in the extratropics both LR and WV feedback appear to be determined by surface-warming patterns, especially in the Southern Hemisphere, and are thus strongly anticorrelated (Colman and Soden 2021; Po-Chedley et al. 2018).

As regards change over time, while the LR and WV feedbacks offset each other to some degree, the total change is robustly positive in the CMIP mean presented here (Fig. 2.1a and Table 2.1). This, unlike for the individual contributions of LR and WV feedback (see the previous sections), concurs with the conclusions of Colman and Soden (2021), and, as they note, seems to apply especially for longer time scales (millennia).

Manabe and Wetherald (1967) estimate that the WV+LR feedback (considered by them only as WV feedback) almost doubles the surface temperature response (i.e., ECS) to a CO_2 forcing, which has since generally been confirmed by more elaborate modelling studies, at least if considered in isolation from other feedbacks (e.g., Hansen et al. 1984; Held and Soden 2000). The central estimate obtained here for the total period from the CMIP model average is 1.15 $Wm^{-2}K^{-1}$ which corresponds to the value given in Sherwood et al. (2020, p. 20), but is smaller than the value of 1.30 $Wm^{-2}K^{-1}$ given by the AR6 (IPCC 2021, p. 970).

For a more extensive treatment of the WV and LR feedbacks, the reader is referred to the seminal review papers by Held and Soden (2000) and Colman and Soden (2021).

2.2.3 Surface-Albedo Feedback

The SA feedback is mostly associated with changes in snow cover and sea ice. Both sea ice and snow exhibit a high albedo, which is generally larger than that of the surface beneath. If the melting of sea ice or snow due to a surface warming uncovers the underlying darker surface (ocean, soil, etc.), the surface albedo declines and more short-wave radiation is absorbed, further warming the surface. Thus, the SA feedback is mostly positive (e.g., Budyko 1969; Sellers 1969; Hall 2004; Winton 2006).

The potentially enormous importance of the SA feedback appears to have been highlighted first in the early modelling studies by Budyko (1969) and Sellers (1969) and, as noted by Lian and Cess (1977), was conventionally thought to be the most important feedback mechanism. With his simple model Budyko (1969) finds that a reduction of the incoming solar radiation by 1.6% would lead to an expansion of the sea ice to a latitude of about 50°, after which the SA feedback becomes large enough to cool the model Earth further. Eventually, this leads to a transition to a so-called “Snowball Earth” (see e.g., Wallace and Hobbs 2006, pp. 50-51; see also Paper III), which may be viewed as an opposite extreme scenario of the above-mention “runaway greenhouse”. However, it was later found that the Budyko (1969) and Sellers (1969) models were too sensitive to the SA feedback because they were neglecting certain factors such

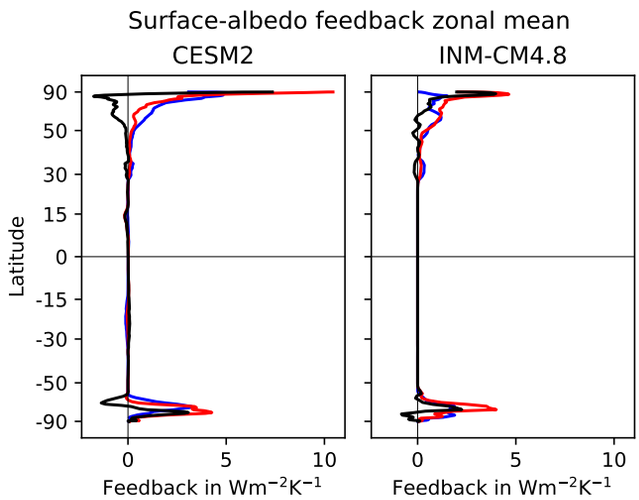

Figure 2.5: Zonal means of surface-albedo (SA) feedback for (left) CESM2 and (right) INM-CM4.8. Early (years 1-20) and late (years 21-150) are depicted as blue and red, respectively. The late-minus-early difference is shown in black.

as cloud shielding, vegetation masking of snow, and the zenith angle variation of albedo, all damping the SA feedback (e.g., Hansen et al. 1984; Lian and Cess 1977).

As Fig. 2.5 illustrates, the largest contributions to the SA feedback come from the Arctic and Antractic sea ice. Figure 2.1a and Table 2.1 indicate that the SA feedback is consistently positive and varies little across models. However, some uncertainty remains which is not borne out here. Generally the largest uncertainty in the SA feedback arises due to cloud masking (IPCC 2021, p. 970). Since the SA feedback is a short-wave radiation feedback, it is strongly influenced by atmospheric short-wave transmissivity which, in turn, is strongly influenced by clouds. However, GCMs have problems simulating clouds in the Arctic, thus possibly biasing the SA feedback (e.g., Sherwood et al. 2020, p. 21). Moreover, vegetation masking of snow is not consistently treated across GCMs, again potentially introducing a bias (Sherwood et al. 2020, p. 21). Further uncertainty of the SA feedback stems from the radiative kernel method: As shown by Hahn et al. (2021), the SA feedback varies considerably depending on which set of radiative kernels is used. This also appears to be related to clouds, as the cloud masking of feedbacks is different across radiative kernels and the net effect of SA and clouds in the Arctic is similar across kernels (Hahn et al. 2021). Thus, methodological difficulty exists pertaining to the partitioning of Arctic feedback between clouds and SA. This has consequences for the contribution of feedbacks to the AA, as discussed in the previous section. Hahn et al. (2021) find that with the radiative kernels that perform best in reproducing radiative

changes due to albedo changes derived from satellite estimates (the kernels by C. J. Smith et al. 2018), a stronger SA feedback than previously thought is estimated.
Another complication with the SA feedback with regards to both the radiative kernel method and GCMs more generally, is feedback state dependence. The SA feedback strongly depends on how much and if sea ice is there to melt: If less sea ice exists to begin with, less sea ice can melt, resulting in a weaker feedback (e.g., Sherwood et al. 2020, p. 21). Additionally, if the radiative kernels used to estimate the SA feedback are *themselves* derived from a climate state with less sea ice, the SA feedback estimate may be weaker (e.g., Block and Mauritsen 2013).

While in the CMIP mean presented here the change of the SA feedback over time is rather small, it may be significant in individual models (Table 2.1), and depend on the set of radiative kernels used (Block and Mauritsen 2013). In CESM2 the SA feedback slightly weakens over time, while it strengths considerably in INM-CM4.8 (Table 2.1; Fig. 2.5). The explanation for this is similar to that for the LR feedback changes in section 2.2.2. In CESM2 the sea ice melts faster than in INM-CM4.8, causing a larger SA feedback in the early period. Conversely, in the late period in CESM2 less sea ice remains to melt, reducing the SA feedback, while in INM-CM4.8 more sea ice is left, indeed *increasing* the SA feedback. The pace of the sea-ice melt may be related to atmospheric and oceanic heat transport, such as the AMOC (see also section 2.2). This is dicussed further in chapter 3 and Paper II.

The SA feedback is widely considered as one of the most important contributors to the AA (e.g., Graversen et al. 2014; Hahn et al. 2021). The issue regarding the strength of individual contributions to the AA was discussed in the section 2.2.2. Here it may only be mentioned that, e.g., Graversen et al. (2014) using one GCM find that the SA feedback explains about 40% of the AA, while the LR feedback explains 15%. Again, however, the topic appears to be terminologically confusing and methodologically challenging (Dai and Jenkins 2023; Hahn et al. 2021; see also the discussion in section 2.2.2 above).

The central estimate of the SA feedback parameter based on the average over the CMIP members considered here is 0.30 $Wm^{-2}K^{-1}$ (Table 2.1) and thus in good agreement with the value given in Sherwood et al. (2020, p. 21). As with the WV+LR feedback however, the estimate provided by the AR6 is again slightly larger (0.35 $Wm^{-2}K^{-1}$, IPCC 2021, p. 971). The reason for the difference may be related to the sources of uncertainty discussed above, especially regarding the radiative kernels. Indeed, using another set of kernels (Pendergrass et al. 2018) yields a central estimate averaged across the here-used CMIP models of 0.48 $Wm^{-2}K^{-1}$, highlighting the kernel dependence of the SA feedback.

2.2.4 Cloud Feedback

Clouds remain the single largest factor of uncertainty in the climate system when it comes to climate feedback and sensitivity (e.g., Caldwell et al. 2016; Zelinka et al. 2020, 2022b; Sherwood et al. 2020, p. 22; see Fig. 2.1a and Table 2.1). The reasons for this are manifold and include the multitude of possible cloud property changes and factors influencing clouds. Furthermore, many important cloud processes occur on small spatial and time scales and thus need to be parameterised in GCMs (e.g., IPCC 2021, p. 972; see also chapter 1). The important influence of clouds as a feedback factor was already recognised by Manabe and Wetherald (1967) and in other early modelling studies (e.g., S. H. Schneider 1972; Cess 1975) and reviews (e.g., S. H. Schneider and Dickinson 1974).

Clouds form when moist air rises and the WV condenses to liquid droplets or ice crystals, which, given sufficient growth, can precipitate as, e.g., rain or snow. A rich taxonomy exists differentiating several cloud types, generally based on cloud height, vertical extent, and phase (e.g., World Meteorological Organization 2017). Following the AR6 (IPCC 2021, p. 971), the general cross-sectional structure of clouds in the atmosphere based on thermodynamic conditions and atmospheric circulation may be described as follows. Close to the equator high SSTs favour the development of deep convective clouds (so-called Cumulonimbus clouds, World Meteorological Organization 2017), with a large vertical extent. These clouds extend high into the troposphere and are topped by anvil and cirrus clouds close to the tropopause. There, because of the temperature inversion, the ascending motion is stopped and diverted poleward, descending in the subtropics, where SSTs are lower. Thus, in the subtropics deep convection is suppressed and the lower tropospheric inversion layer maintained, favouring the formation of shallow (strato-)cumulus clouds. Further poleward in the extratropics, clouds are generally formed along the mid-latitude storm tracks. Notably, liquid cloud droplets do not spontaneously freeze to ice crystals above about -40 °C and thus extratropical clouds are often of mixed phase (IPCC 2021, p. 971).

Clouds weaken both the incoming short-wave and the outgoing long-wave radiation at TOA because of increased reflection and absorption, respectively. Essentially, the short-wave effect of clouds is similar to the SA feedback, while the long-wave effect is similar to the WV feedback. For low clouds the cooling short-wave effect tends to dominate because of their high albedo and cloud-top temperature, while for high clouds the warming long-wave effect is more important, as they are cold and interact little with the incoming short-wave radiation (e.g., Gettleman and Sherwood 2016; Wallace and Hobbs 2006, 447f.).

In response to global temperature changes, clouds may exhibit a multitude of different responses, including changes in cloud altitude, phase, life time, amount, optical depth, and area (e.g., S. H. Schneider 1972, Sherwood et al. 2020, pp. 22-30). These changes may be interconnected: For example, a shift in cloud phase toward more liquid droplets increases cloud optical depth, as liquid droplets tend to be smaller than ice crystals and smaller cloud particles cause a higher optical depth (Stephens 1978).

To constrain estimates of cloud feedback, it is now common to decompose the cloud response into individual cloud regimes based on physical processes and regions (e.g., Gettleman and Sherwood 2016; Zelinka et al. 2022b; Sherwood et al. 2020, pp. 22-30; IPCC 2021, pp. 971-975). The AR6 (IPCC 2021, pp. 971-975), building on the review by Sherwood et al. (2020, pp. 22-30) differentiates seven subcategories of cloud feedbacks based on region (marine, land, tropics, extratropics) and cloud properties (amount, altitude, optical depth; see also Zelinka et al. 2022b). These are here briefly summarised.

Two subcategories of cloud feedback are assessed as robustly positive: The *high-cloud altitude* feedback and the *subtropical marine low cloud* feedback.

The high-cloud altitude feedback arises because under a warming the high clouds are expected to rise due to the "fixed-anvil temperature" (FAT) mechanism (Hartmann and Larson 2002; Yoshimori et al. 2020), leaving their top temperature unchanged. Thus, the long-wave radiation emitted by these clouds does not change under a surface warming, constituting a positive feedback. However, e.g., Seeley et al. (2019) find in their cloud-resolving modelling study that the anvil temperature indeed changes with warming, while, conversely, the tropopause temperature remains fixed (although only if the tropopause is radiatively defined). Seeley et al. (2019) note that this does not imply a FAT. Nonetheless, the AR6 (IPCC 2021, p. 972) considers theory, GCMs, process models, and observational estimates in agreement on the sign of the high-cloud altitude feedback and assesses it to be positive with high confidence.

The subtropical marine low-cloud feedback is almost exclusively a short-wave effect (Sherwood et al. 2020, p. 25) and used to be the largest contributor to the uncertainty of the cloud feedback, especially for GCM feedback estimates (IPCC 2021, pp. 973, 975). However, according to the AR6 (IPCC 2021, p. 973) the uncertainty in this component could be significantly reduced due to the application of a new method called "cloud controlling factors" (e.g., Klein et al. 2017). The mechanism of this feedback is mainly thermodynamic and acts to reduce the low-cloud cover because of a decrease in atmospheric stability resulting from SST increase, thus constituting a positive feedback (IPCC 2021, p. 973; Sherwood et al. 2020, pp. 25-26). The AR6 states that studies employing large eddy simulation (LES) models estimate a subtropical marine low-cloud feedback of up to 0.2 $Wm^{-2}K^{-1}$ global contribution, and adopts this as the best estimate (IPCC 2021, p. 973). Conversely, Radtke et al. (2021), also investigating this feedback in an LES model but across different resolutions (100 m to 5 km), find that the higher positive values only hold for coarser resolution, and

at the highest resolution the feedback is close to zero. They note that recent observational studies (Cesana and Del Genio 2021; Myers et al. 2021) similarly show near-zero subtropical marine low-cloud feedback. However, the reasons for these similar estimates appear to be different: While Radtke et al. (2021) implement a uniform 4 K SST increase in their limited domain LES model and study the impact on low cloud, in the historical data used by Cesana and Del Genio (2021) and Myers et al. (2021) a SST *decline* is observed in the geographical regions where the subtropical marine low-cloud feedback is active (this is dicussed further in chapter 3). That is, the observational cloud response is a response to a local surface *cooling*, while Radtke et al. (2021) study the response to a local surface *warming*. Thus, it is remarkable that Radtke et al. (2021) find a similar subtropical marine low-cloud feedback as Cesana and Del Genio (2021) and Myers et al. (2021). However, Radtke et al. (2021) conclude that it is not necessarily the case that in other LES models the subtropical marine low-cloud feedback converges to zero with increasing resolution, warranting further future study. The AR6 (IPCC 2021, p. 973) assesses this cloud feedback component as robustly positive, although with a large range, including the small values found by Cesana and Del Genio (2021), Myers et al. (2021), and Radtke et al. (2021).

Further cloud feedback components assessed in the AR6 to be positive, however with less confidence, are the *land cloud* feedback, the *mid-latitude cloud amount* feedback, and the *Arctic cloud* feedback.

The land cloud feedback is induced by a decrease of relative humidity over land under global warming. This is because land warms faster than the ocean and the specific humidity increase due to increased evapotranspiration and transport from the ocean is insufficient to compensate the rise of the saturation WV pressure over land following the Clausius-Clapeyron relationship. This leads to a decrease of mainly low-cloud cover over land, inducing a positive feedback (Sherwood et al. 2020, p. 28). However, on a global scale this feedback is estimated to be small, and, even though a robust feature in GCMs, observational evidence is still lacking, leading to low confidence in the estimated values (IPCC 2021, p. 973).

The mid-latitude cloud amount feedback is mainly induced by increasing SSTs in the mid-latitudes, decreasing atmospheric stability and thus low-cloud amount. Observations qualitatively agree, but since the AR6 estimate involves the GCM estimates, the range of this feedback component is large, even encompassing negative values (IPCC 2021, p. 974).

The Arctic cloud feedback stems primarily from low clouds that covary with the sea ice underneath. This feedback strongly varies seasonally, since in the winter the clouds contribute a positive greenhouse effect, and in summer a negative albedo effect. There is general agreement between GCM estimates and observations, although regional variations are large. Thus, this feedback component is assessed to be positive, but with a large range, encompassing also negative values (IPCC 2021, p. 974)

As opposed to the preceding cloud feedback components, the AR6 assesses the *tropical high-cloud amount* feedback (also called *tropical anvil cloud area* feedback; Sherwood et al. 2020, p. 26) and the *extratropical cloud optical depth feedback* to be negative, however with low confidence (IPCC 2021, pp. 972-973). The tropical high-cloud amount feedback arises due to the reduction of the area of the anvil and cirrus clouds accompanying deep convective clouds in the tropics. Since these clouds mainly act to trap outgoing long-wave radiation, their reduction in response to a warming constitutes a negative feedback. This process has been called an infrared "iris effect" (Lindzen et al. 2001) in analogy to the human iris expanding or contracting to regulate incoming light. The iris effect has been the subject of considerable scientific debate (see the review in Lindzen and Choi 2021) and even though the sign of this feedback seems to be robustly negative (e.g., Mauritsen and Stevens 2015), the physical mechanism and the magnitude are poorly understood (Bony et al. 2020; Lindzen and Choi 2021; Mauritsen and Stevens 2015; IPCC 2021, p. 972). Thus, there is little confidence in the estimate of this negative feedback component (IPCC 2021, p. 973). Indeed, according to the AR6, this tropical high cloud amount feedback is now the largest contributor to the uncertainty in the net cloud feedback (IPCC 2021, p. 975).

Finally, the extratropical cloud optical depth feedback is induced by cloud phase changes over the Southern Ocean. Under warming, cloud ice crystals transition to liquid droplets, making the clouds more reflective, thus inducing a negative feedback. A known problem with many CMIP5 models is the creation of too much cloud ice in favour of supercooled liquid droplets, especially in the polar regions (e.g., Cesana et al. 2015). Thus, estimates of the extratropical optical depth feedback from these models are judged to be too large and unreliable. Some models contributing to CMIP6 have an improved representation of the ice–liquid ratio and produce estimates of this feedback more in line with observational estimates (e.g., Bjordal et al. 2020; Kang et al. 2023; Mülmenstädt et al. 2021). The AR6 thus assesses the extratropical cloud optical depth feedback to be weakly negative, however with a large range (IPCC 2021, p. 974).

Summing up the individual contributions and uncertainties, the AR6 obtains a net cloud feedback parameter of 0.42 ± 0.3 $Wm^{-2}K^{-1}$ (IPCC 2021, p. 974). The central estimate from the CMIP models considered here of 0.59 $Wm^{-2}K^{-1}$ is somewhat higher, but falls well within the given range and exhibits similarly large uncertainty (Table 2.1 and Fig. 2.1a). Notably, the two individual models shown in Table 2.1 represent relatively extreme examples, lying outside the AR6 likely range. However, they illustrate the large variance of cloud feedbacks across GCMs. Comparing the two models in terms of regional feedbacks (Fig. 2.6), it is clear that significant cloud feedback pattern differences may exist across GCMs. Considering the late period (years 21-150) general similarities are apparent, such as positive feedbacks in the East Pacific and the North Atlantic. However, in CESM2 the positive effect in the East Pacific is considerably larger than in INM-CM4.8 and is accompanied by positive feedbacks over the South-

ern Ocean which are non-existent in INM-CM4.8. The exceptionally positive feedback in CESM2 is discussed in detail further below.

Generally, the pattern of the cloud feedback is related to the surface-warming pattern induced by the CO_2 forcing and modified by atmospheric and oceanic heat transport. Accordingly, changes in the surface warming pattern over time also cause the cloud feedback to change over time in GCMs (e.g., Andrews and Webb 2018; Ceppi and Gregory 2017; Dong et al. 2022a; Lin et al. 2021; C. Zhou et al. 2017). For many members of the CMIP archives the cloud feedback changes substantially over time, mostly becoming more positive (Fig. 2.1a). Indeed, for some models the strengthening of the positive cloud feedback dominates the weakening of the total climate feedback (e.g., Paper I). CESM2 exemplifies this behaviour (Fig. 2.1b), substantially strengthening its cloud feedback over time in the Southern Ocean and the East Pacific (Fig. 2.6e). Although this effect is rather extreme in CESM2 compared with other GCMs (see the discussion below), it is a general feature in several climate models (Zelinka et al. 2020) and discussed further in chapter 3.

The cloud feedback continues to present a large conundrum when it comes to GCM estimates of climate sensitivity. It appears to be the main reason for the debate on "hot models" (Hausfather et al. 2022; Voosen 2021; see section 1.2), culminating in the decision by the IPCC not to rely on GCMs for estimating climate sensitivity (IPCC 2021, pp. 1005, 1007). Connected with this, cloud feedback is also the reason for the increase in the range of ECS estimates from CMIP5 to CMIP6, despite the increased sophistication of many CMIP6 members (Zelinka et al. 2020).
One of the contributors to this increase in ECS due to cloud feedback is CESM2 (Bacmeister et al. 2020; Bjordal et al. 2020; Gettleman et al. 2019), which is the main tool of study in Papers II and III included in this thesis. This model is an illustrative example that increased model sophistication does not necessarily implicate a better correspondence of the results with observations and that, as shown by Zelinka et al. (2022a) improved skill in representing base-state cloud properties does not necessarily improve cloud feedbacks in GCMs. The stronger positive cloud feedback in CESM2 compared to its predecessor was found to result mostly from low clouds over the Southern Ocean with contributions from the tropics (Bacmeister et al. 2020; Bjordal et al. 2020; Gettleman et al. 2019). Bjordal et al. (2020) showed that this was because of the better representation of the ratio of cloud ice crystals to cloud liquid droplets in CESM2 compared to previous model versions (see also Gettleman et al. 2019), and thus deemed the large positive cloud feedback and the resulting high ECS plausible. However, in a comparison of CESM2 simulations with paleoclimate proxies of the Eocene Climate Optimum and the Last Glacial Maximum (LGM), Zhu et al. (2020, 2021) find that the model runs too hot and too cold, respectively. The excessive LGM cooling is attributed to strong short-wave cloud feedback in the Southern Hemi-

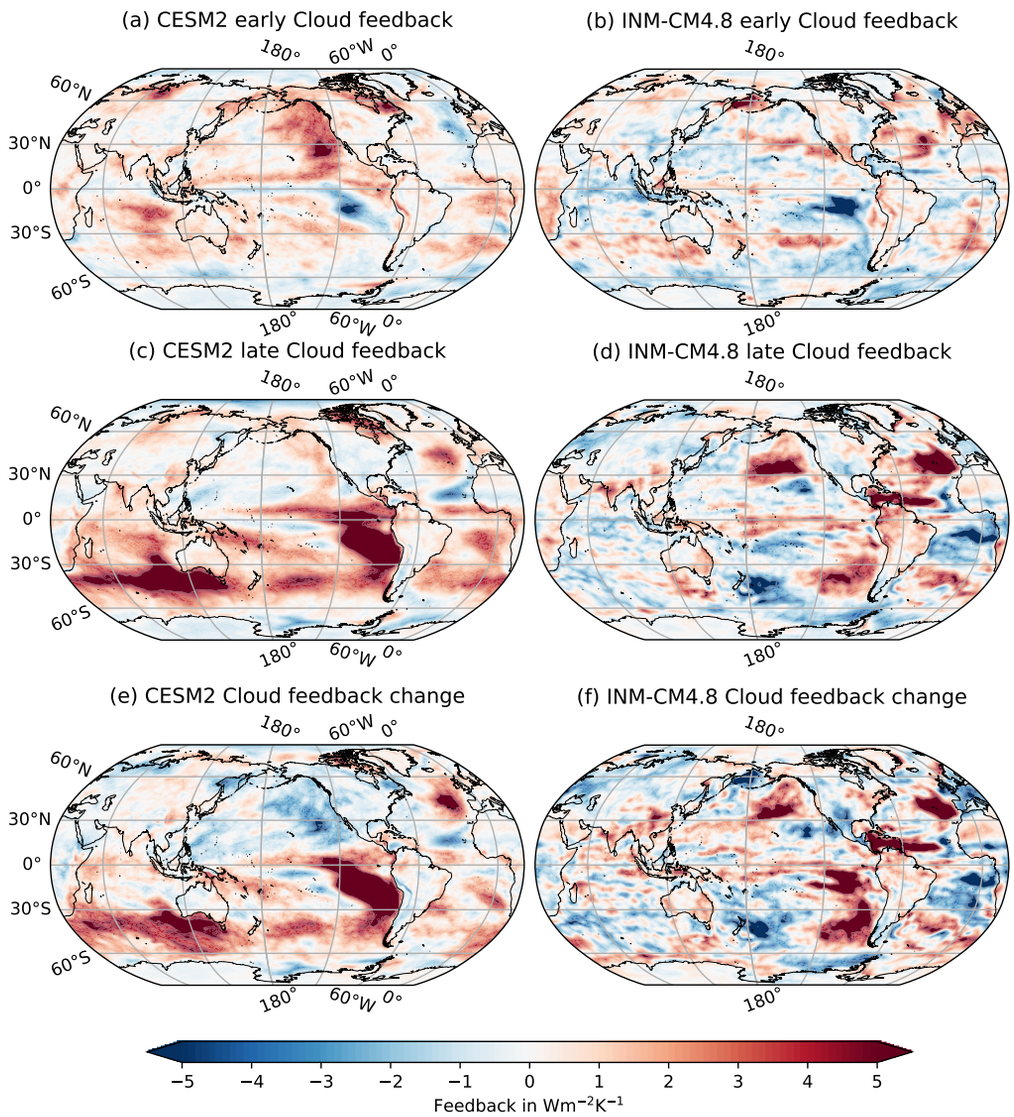

Figure 2.6: As Fig. 2.2 but for the cloud feedback.

sphere subtropics and middle to high latitudes (Zhu et al. 2021). Furthermore, Cesana and Del Genio (2021) and Myers et al. (2023) show that CESM2 (among other CMIP6 members; see also Myers et al. 2021) overestimates the tropical low-cloud feedback[7]. Finally, CESM2 may also underestimate a negative cloud lifetime feedback, as due to the improved ice-liquid ratio clouds tend to be less mixed-phase and should thus diminish less readily from precipitation (Mülmenstädt et al. 2021). Indeed, Shaw et al. (2022), comparing CESM2 simulations with satellite data find what they call a "model error" (Shaw et al. 2022, p. 3) in the CESM2 cloud parametrisation: While a limit is set on the number of ice crystals that can be present in a cloud based on the number of available ice nucleating particles, heterogeneous nucleation processes can still add *mass* to the already existing ice crystals. This may artificially increase their size and cause them to sediment.
Taking this "error" into account and further refining the CESM2 cloud parametrisation (see more discussion in Paper III), Zhu et al. (2022) manage to calibrate CESM2 to an LGM paleoclimate proxy. Estimating ECS with this recalibrated version of CESM2, they obtain a much lower value of 3.9 K as compared to the original 5.3 K, also being more in line with the ECS of 4 K estimated from the previous model version (CESM1). Remarkably though, in a higher-resolution version, CESM2 is found not to reduce its climate sensitivity after implementing the cloud parametrisation correction (McGraw et al. 2023). Similarly, Gettelman et al. (2023) find little change in the higher-resolution version of CESM2 after they implemented their newly developed Parameterization of Unified Microphysics Across Scales (PUMAS), which, among other improvments, corrects the above-discussed error. It must be noted however, that the results of McGraw et al. (2023) and Gettelman et al. (2023) are based only on short integrations of the model with uniformly raised SSTs (+4 K). This discounts the pattern of SST changes, which, especially in CESM2, strongly impacts cloud feedback (Kang et al. 2023; see chapter 3 and Paper III).
In addition to the preceding discussion it must be noted that, while GCMs typically produce too weak cloud feedback in response to surface-warming changes especially in the Southern Ocean (Kim et al. 2022), CESM2, in contrast to its predecessor, manages to reproduce observational cloud feedback estimates when nudged towards observed Southern Ocean SST anomalies (Kang et al. 2023).

The example of CESM2 may be instructive for the use of climate models more generally, since many of the GCMs exhibiting high ECS appear to overestimate observed cloud feedbacks (Cesana and Del Genio 2021; Myers et al. 2021,

7. It must be noted here that this by itself does not necessarily imply problems with CESM2's representation of clouds. As is discussed in this section and in chapter 3, tropical low-cloud feedbacks are found to sensitively depend on surface-warming patterns. Given that CESM2 in the historically forced run does not reproduce observed surface-warming patterns, failure to reproduce low-cloud feedbacks is unsurprising.

2023). This appears to be further confirmed by the recent development of the Energy Exascale Earth System Model (E3SM) from version 1 (Golaz et al. 2019) to version 2 (Golaz et al. 2022): Similar to CESM2 and its Zhu et al. (2022) recalibrated version, this model's ECS decreased from 5.3 K (E3SMv1) to 4.0 K (E3SMv2) because of changes to the cloud parametrisation impacting mainly the marine low-cloud feedback (Golaz et al. 2022).

The above discussion highlights the problem of compensating errors when it comes to climate models, feedbacks, and ECS. That is, the improvement of one process may *negatively* impact the quality of the results because other, compensating errors exist, introducing new biases. While this supports the decision of the IPCC to refrain from using direct GCM estimates of ECS in the AR6, it must be maintained that GCMs are still an important tool for studying the climate system, sensitivity, and feedback. This point is taken up again at the end of the next chapter.

In the preceding review of the physical climate feedbacks and their mechanisms it is apparent that the feedback patterns and their changes over time are related to the pattern of the surface warming. This so-called "pattern effect" (Stevens et al. 2016) has since emerged as the most prominent explanation for the change over time of climate sensitivity and feedback seen in GCM experiments and is the topic of the following chapter.

3

Time-Dependence and the Pattern Effect

The issue of the change over time of climate sensitivity and feedback in climate model experiments has been a topic of increasing research interest for about 30 years (e.g., Murphy 1995, Senior and Mitchell 2000, Boer and Yu 2003a,b, Gregory et al. 2004, Williams et al. 2008, Winton et al. 2010, Andrews et al. 2012, Armour et al. 2013, Andrews et al. 2015, Rugenstein et al. 2016, Ceppi and Gregory 2017, Andrews and Webb 2018, Dong et al. 2019, 2020). The discovery and initial investigation of this time-dependence was based on individual-model studies, generating different explanations, with a general theory remaining elusive (Murphy 1995, Senior and Mitchell 2000, Boer and Yu 2003a,b, Gregory et al. 2004, Williams et al. 2008, Winton et al. 2010, Armour et al. 2013, Rugenstein et al. 2016, Andrews and Webb 2018). A brief review of these explanations is given in the introduction to Paper I.

About two decades after the initial discovery of the time-dependence of climate sensitivity, the idea of the so-called "pattern effect" emerged (Stevens et al. 2016, Andrews et al. 2012, 2015, C. Zhou et al. 2016, Mauritsen 2016) and has since become the accepted explanation (IPCC 2021, p. 989)[1]. The general idea of the pattern effect is that in response to a forcing surface-warming patterns

1. Indeed, in some studies the term "pattern effect" is now implicitly used to mean feedback change and quantified in $\mathrm{Wm^{-2}K^{-1}}$ (e.g., Andrews et al. 2022; Modak and Mauritsen 2023)

evolve over time to favour different regions of varying atmospheric stability. Local surface warming/cooling changes the atmospheric stability, activating different climate feedbacks, notably cloud and lapse-rate feedbacks (C. Zhou et al. 2016; Andrews and Webb 2018; Ceppi and Gregory 2017; Dong et al. 2019, 2020; C. Zhou et al. 2017; see also section 2.2). The shift in surface-temperature patterns thus changes the global "cooling efficiency"—that is, the efficiency with which energy is radiated to space—, hereby causing the time-dependence of global climate feedback and sensitivity.

Different specific geographical regions have been identified as particularly important for the change of climate feedbacks and sensitivity. C. Zhou et al. (2017) and Dong et al. (2019) find the Indo-Pacific warm pool (IPWP) to be the main contributor to the change over time in cloud and total feedback, respectively (see also Andrews and Webb 2018; Dong et al. 2020; Paper I). This is due to the fact that the IPWP is a large region of strong convective activity, and thus readily communicates a surface warming vertically to the free troposphere, where it is easily radiated to space. A further effect is that the warming in the free troposphere is spread around the globe mostly unhindered by topography, increasing the atmospheric stability in other regions. This strengthens both the negative LR feedback as well as the negative low-cloud feedback, since low-cloud cover tends to increase with atmospheric stability (see section 2.2.4; Andrews and Webb 2018; Ceppi and Gregory 2017; Klein et al. 2017; Klein and Hartmann 1993; Mauritsen 2016; Wood and Bretherton 2006; C. Zhou et al. 2016). Thus, if the IPWP region warms more strongly than the rest of the globe, the total climate feedback becomes more negative over time and vice versa. This effect is illustrated in Fig. 3.1.

Lin et al. (2019), Singh et al. (2022), and Papers I and II find that a change over time of the climate feedback is also connected to northern polar warming (see also Paper III), mainly because of the effect of sea-ice melt and the eminent latitudinal redistribution of heat by the Atlantic meridional overturning circulation (AMOC). The northern polar region and especially the Arctic are strongly stably stratified and thus inefficient at cooling. If due to a change in meridional heat transport heat is transferred from the tropics to the polar regions, hereby cooling the former and warming the latter, this weakens the global climate feedback, and vice versa. Singh et al. (2022) and Paper II show that this is mostly due to impacts on the SA and LR feedbacks, with contributions from the cloud feedback.

Most recently, the eminence of the Southern Ocean region for climate feedback has been recognised (Dong et al. 2022a; Kang et al. 2023; Kim et al. 2022; Lin et al. 2021). A warming (cooling) over the Southern Ocean causes local stability decreases (increases) and accompanying decreasing (increasing) cloud cover, constituting a positive feedback (Lin et al. 2021). However, the Southern Ocean surface warming (cooling) is also readily communicated via the climatological mean winds to the stable East Pacific (EP). There, again, the atmospheric

stability decreases (increases), inducing a reduction (intensification) of the low-cloud cover, i.e., a positive feedback. This has been shown to be an important effect, although its magnitude is different for different GCMs and possibly depends sensitively on the cloud parametrisation (Dong et al. 2022a; Kang et al. 2023; Kim et al. 2022; Lin et al. 2021; Zhang et al. 2021; Paper III).

The Pattern Effect in abrupt4xCO2 Most GCMs exhibit a *weakening* of the climate feedback over time in the abrupt4xCO2 experiment (e.g., Andrews et al. 2015; Dong et al. 2020; Paper I). This is explained by the pattern effect as follows. The CO_2-induced surface warming is first rather homogeneous and, importantly, slower in the EP than in the IPWP (e.g., Paper I; Andrews et al. 2015; Dong et al. 2020). This induces a relatively strong negative global feedback according to the mechanism explained above. However, as illustrated in Fig. 3.1b, over time the warming shifts to regions of relatively strong atmospheric stability, particularly the EP, and away from regions of strong convection, particularly the IPWP, thus weakening the feedback (e.g., Andrews et al. 2015; Ceppi and Gregory 2017; Dong et al. 2020; Paper I).

Dong et al. (2020) manage to explain the change over time of the total climate feedback for the members of CMIP5 based on the warming in the IPWP. However, this is less successful for CMIP6, possibly because other regions, such as the Southern Ocean, are more important in some members of CMIP6 (Dong et al. 2020). In Paper I, it is pointed out that after removing several models with particularly strong cloud feedback change, a good correlation is found between relative warming in the IPWP[2] and climate feedback change across models. This appears to be consistent with the findings of Lin et al. (2021), that the Southern Ocean warming in the abrupt4xCO2 experiment also exerts considerable influence on the total and in particular cloud feedback change. That is, models with strong cloud feedback (change) may be impacted more by changes in surface warming over the Southern Ocean, thus confounding the explanation of the feedback change due to IPWP warming. A case in point is CESM2, which is found to exhibit an exceedingly strong cloud response over the Southern Ocean and in the EP to Southern Ocean SST changes (Fig. 2.6a,c,e; Paper III). In section 2.2.4 the potential exaggeration of the cloud feedback in CESM2 due to a problem in the cloud parametrisation was discussed.

A further contribution to the change over time in climate feedback in the abrupt4xCO2 experiment results from changes in meridional energy transport, particularly due to a change in the AMOC (Lin et al. 2019; Singh et al. 2022; Paper II). A general feature in GCMs is an initial decline of the AMOC in response to the quadrupling of the CO_2 concentration (e.g., Bellomo et al. 2021),

2. That is, relative to global-mean warming.

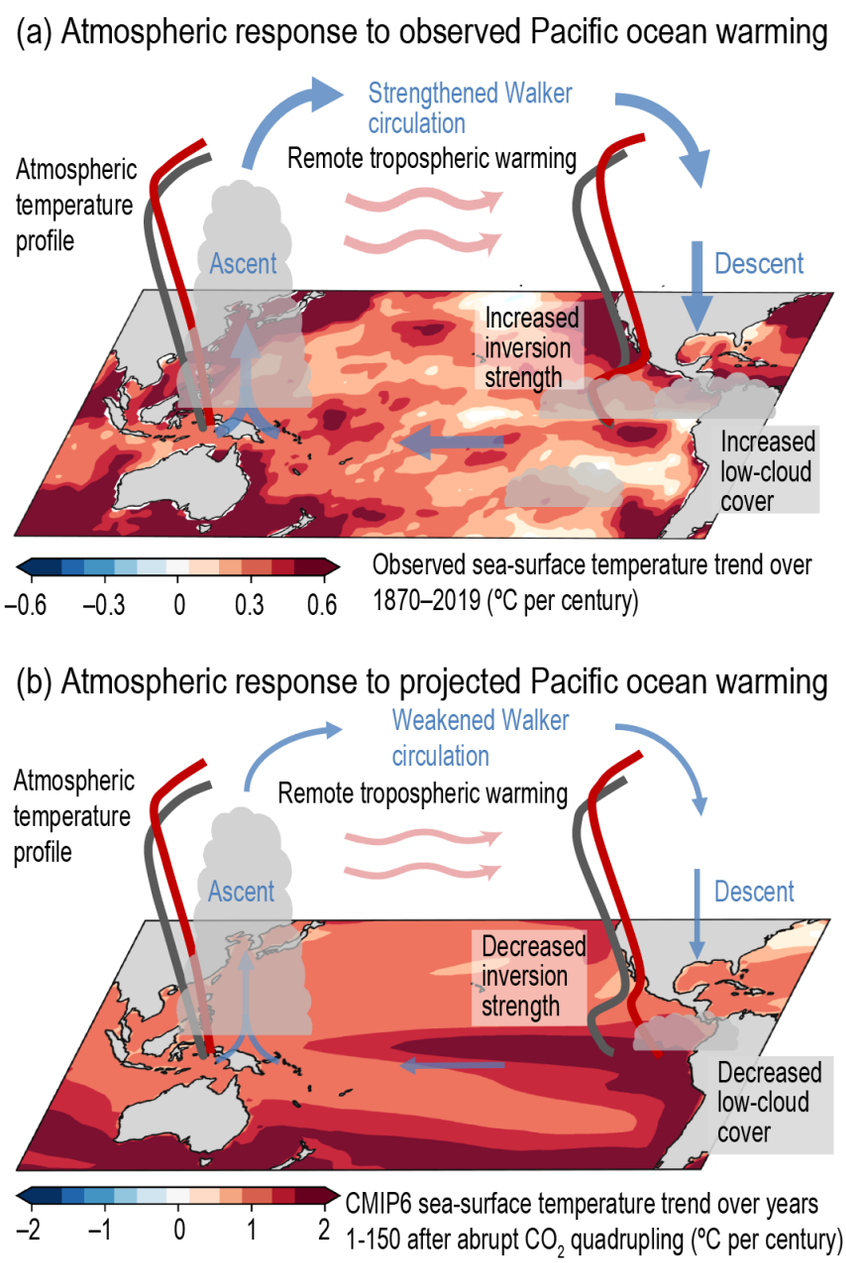

Figure 3.1: Illustration of (a) the historical pattern effect and (b) the pattern effect as simulated by global climate models in the abrupt4xCO2 forcing experiment. (Figure 7.14 in IPCC 2021, p. 990).

implying a reduced heat transport from the tropics into the northern polar region. However, after some time, some GCMs partially recover the AMOC, thus again increasing the poleward heat transport and hence northern high-latitude surface warming (e.g., Lin et al. 2019; Bellomo et al. 2021; Paper II). According to the mechanism explained above, this induces a positive feedback change. Both CESM2 and, more intensely, INM-CM4.8 exhibit an AMOC recovery (not shown), which may be the reason for the increase over time of their positive LR and cloud feedbacks in the North Atlantic (Figs. 2.4 and 2.6), contributing to the weakening of their global climate feedback (Table 2.1).

The Historical Pattern Effect

The pattern effect has also been used to explain the variation of climate feedback over the historical period (C. Zhou et al. 2016, 2017; Gregory and Andrews 2016; Mauritsen 2016; Andrews et al. 2018; Dong et al. 2019; Kang et al. 2023). It is well established that the surface temperature of the Earth in the *global mean* has increased over the last century, and especially in the last four decades (e.g., IPCC 2021, 59f.). However, *regional* SAT and SST trends vary and a weak but significant cooling has been observered since 1980 in the Southern Ocean and the East Pacific (e.g., Zhang et al. 2021). Notably, this is not captured by historical GCM simulations (e.g., Wills et al. 2022), nor does it occur in the idealised abrupt4xCO_2 experiment (e.g., Dong et al. 2021). There is currently no consensus on an explanation for this observed cooling and multiple hypotheses have been put forward. These include, as listed by Kang et al. (2023): natural internal variability associated with deep ocean convection, stratospheric ozone depletion, Antarctic ice-sheet meltwater, and a driving mechanism due to a La Niña-like cooling in the EP. Recent work shows that the cooling in the Southern Ocean may drive the cooling in the EP region (Kang et al. 2023; Kim et al. 2022; Paper III) and that feedback loops may exist by which the cooling in these two regions reinforces itself (Dong et al. 2022a; Kang et al. 2023; Paper III).

The Southern Ocean and EP cooling together with the relatively large warming in the IPWP explain the variation of the net climate feedback parameter over the historical period according to the mechanisms described above (Andrews et al. 2018, 2022; Dong et al. 2019; Kang et al. 2023). This is illustrated in Fig. 3.1a: The relatively large warming in the IPWP is efficiently spread vertically through convective ascent. In the free troposphere, the warming is easily transported over the EP, where it increases atmospheric stability. This is further enhanced by the observed surface cooling in the EP. These effects, together with the observed Southern Ocean cooling act to cause anomalously negative cloud and LR feedbacks, especially over the EP and the Southern Ocean, strengthening the negative global climate feedback.

The Pattern Effect – Historical, abrupt4xCO2, and the Future

As discussed in chapter 1, estimates of ECS inferred from the historical record are typically lower than estimates based on GCMs. The pattern effect helps to resolve this discrepancy (C. Zhou et al. 2016, 2017; Gregory and Andrews 2016; Mauritsen 2016; Andrews et al. 2018, Dong et al. 2021): While over the historical period the Southern Ocean and the EP have cooled and the IPWP has relatively strongly warmed, strengthening the climate feedback, in the abrupt4xCO2 simulation the inverse situation obtains, thus exhibiting an increasingly weakened climate feedback over time.

It follows from the above discussion that the comparatively small estimates of ECS inferred from observations (see sections 1.3.1 and 1.3.3) depend on the surface-warming pattern and specifically on the Southern Ocean and EP cooling. It has been suggested that the Southern Ocean and the EP will eventually warm (e.g., Wills et al. 2022, and references therein), which would imply a greater (eventual) global warming (i.e., higher ECS) than observational ECS estimates suggest (Andrews et al. 2018; C. Zhou et al. 2021). However, as the causes for the observed SST pattern changes are yet unknown (see Lee et al. 2022 for a recent review), several future scenarios are plausible (Andrews et al. 2022; Dong et al. 2021). It may be that the Southern Ocean and EP cooling trends result from internal variability and thus they could reverse in the near future, weakening the climate feedback and accelerating global warming (e.g., Watanabe et al. 2021). Conversely, the regional cooling response could be consistently induced by the CO_2 forcing, meaning that it might continue under continued CO_2 forcing (e.g., Seager et al. 2019; Lee et al. 2022). This would imply that GCMs overestimate the weakening of the climate feedback and simulate too high ECS.

It must be noted that there is considerable uncertainty regarding the magnitude of the historical pattern effect, as different SST datasets exhibit somewhat different regional warming and cooling trends (e.g., Lewis and Mauritsen 2021). However, this may also depend on the specific GCM used to investigate the impact of the surface-temperature pattern on climate feedback (compare Lewis and Mauritsen 2021 with Andrews et al. 2022). Accounting for the dataset variability, Modak and Mauritsen (2023) recently obtained an ECS estimate of 3.2 K similar to the Andrews et al. (2018) estimate, however with a larger 5th-95th percentile range of 1.8-11.0 K compared to 1.5-8.1 K.

Beyond ECS, there is also the question of the *timing* of a possible reversal of the Southern Ocean and EP cooling trends. That is, if the Southern Ocean and EP cooling are, e.g., mainly caused by Antarctic ice-sheet meltwater input, they may continue well into the 21st century, reducing the warming over the coming decades (e.g. Dong et al. 2022b). This would mean that even though

ECS estimates of GCMs may be correct because the cooling trends *eventually* reverse, this would happen only far into the future, implying that the *transient* global warming approaches the ECS only slowly. Consequently, ECS would be less instructive for assessing 21st century climate change than some studies indicate (e.g., Grose et al. 2018).

The preceding discussion chronicles the progress made regarding the explanation of the change of global climate feedback and sensitivity over time. The pattern effect and its impact on climate feedback have been intensely studied and are now increasingly understood. However, the discussion also highlights several important holes in the understanding of the pattern effect. This mainly concerns the question of how the specific observed surface-warming pattern arises and how it will develop in the future. Given the potentially large implications of this for estimates of ECS and future warming in general, continued research on this topic is imperative.

A Comment on the Role of GCMs

While in chapters 1 and 2 the potential problems with GCM estimates of ECS and climate feedbacks where discussed, the current chapter makes it clear that GCMs were instrumental for the discovery, investigation, and understanding of the pattern effect. GCMs helped to realise and understand (1) that climate feedbacks may change over time, (2) that the pattern evolution of the surface warming influences the global-mean warming itself, and (3) the meaning of this for historical and potential future warming. Thus, while it appears a reasonable decision of the IPCC and climate science in general to move away from the sole reliance on GCMs for estimating ECS, they still play a vital role in the study of the climate system.

4

Summary of Papers

The main scientific contribution of this thesis consists in the three included papers (chapters 6, 7, and 8). The following sections give a summary of the papers, setting them into the wider context presented in the first three chapters of the thesis as well as providing a connecting framework.

4.1 Paper I

Eiselt, K.-U., and R. G. Graversen (2022): Change in climate sensitivity and its dependence on the lapse-rate feedback in $4\times CO_2$ climate model experiments. *J. Climate*, **35**, 2919-2932, https://doi.org/10.1175/JCLI-D-21-0623.s1.

The general aim of this study is to investigate the reasons of the time-dependence of climate feedback in the abrupt4xCO2 experiment as conducted by the fully-coupled global climate models (GCMs) participating in phases 5 and 6 of the Coupled Model Intercomparison Project (CMIP). The study was conducted when the concept of the pattern effect (chapter 3) as an explanation for the change of climate feedback over time had not gained the prominence it now has. However, in the paper the "stability hypothesis" is invoked, based on the explanation in Ceppi and Gregory (2017): Patterns of surface warming change over time to favour more stable over less stable regions, thus weakening the cooling efficiency, implying a weaker stabilising feedback. Hence, the stability hypothesis corresponds to the pattern effect.

The global total feedback for each CMIP member is estimated using the Gregory method (section 1.3.3) and the individual physical feedbacks are derived with the radiative kernel method (section 2.1). This is done separately for the early (years 1-20) and the late (years 21-150) period of the simulations. To investigate the change over time, the result for the early period is subtracted from that for the late period. Generally, it is observed that the total feedback change over time in most members is either due to a change in the lapse-rate (LR) feedback and/or the cloud feedback. Removing the GCMs with particularly large cloud feedback change causes the correlation of the total with the LR feedback change over time to increase significantly. This is interpreted to generally confirm the "stability hypothesis" (i.e., the pattern effect), as the LR feedback is strongly related to atmospheric stability. Notably, clouds are also strongly connected to atmospheric stability (section 2.2.4 and chapter 3). However, since clouds in GCMs are strongly dependent on parametrisations, which may be different across models, the connection between atmospheric stability and cloud feedback across GCMs is deemed less straightforward (section 2.2.4) and the focus is laid on the LR feedback. Thus, the above-mentioned models with a strong cloud feedback are removed from the further analysis

To investigate differences across models, GCMs are grouped according to the magnitude of their LR feedback change ($\Delta\lambda_{LR}$). Averages over the two groups G1 with weak ($\Delta\lambda_{LR} < 0.1$ $\mathrm{Wm^{-2}K^{-1}}$) and G2 with strong ($\Delta\lambda_{LR} > 0.5$ $\mathrm{Wm^{-2}K^{-1}}$) LR feedback change are calculated and compared. In a regional analysis it is found that with respect to relative warming change, atmospheric stability change, and individual feedback change, G1 and G2 are most different in the Arctic. Further analysis reveals that this is connected to sea-ice cover changes: Sea-ice changes are the main cause of the surface-albedo (SA) feedback (section 2.2.3) and because of its insulating effect sea ice strongly inhibits atmosphere-ocean coupling, impacting the LR feedback as well (section 2.2.2). As G1 loses most of its sea ice in the early period, it exhibits strong positive early-period SA and LR feedbacks, especially in the Arctic. However, in the late period little sea ice is left to melt further, hereby reducing both the SA and LR feedbacks, implying a negative feedback change. The reverse situation obtains in G2, where in the late period more sea ice remains and continues to melt, inducing continued positive SA and LR feedbacks. Consistent with the faster sea-ice melt in G1 than in G2 it is found that G1 warms more in the global mean than G2. This presents a hen-and-egg problem: Does the faster global-mean warming cause the stronger sea-ice melt or does the faster sea-ice melt contribute to induce a faster global-mean warming? Theories for both conjectures are considered and briefly reviewed but no conclusion is reached as causal direction is difficult to analyse in fully-coupled GCM experiments.

The final part of the analysis considers correlations across all GCMs of the change over time of regional warming relative to the global-mean warming

with the change over time of total climate feedback. The investigated regions include the Arctic, as well as regions that were highlighted in earlier literature (Andrews and Webb 2018; Dong et al. 2019; see also chapter 3): the East Pacific (EP), the West Pacific (WP), and the Indo-Pacific warm pool (IPWP). While no correlation is found for the EP, substantial *negative* correlations are found for the WP and especially the IPWP, while a robust *positive* correlation is found for the Arctic. Thus, the results of Dong et al. (2019, 2020) highlighting the influence of the IPWP on the total climate feedback change are confirmed (see chapter 3), but the potential influence of the Arctic is again suggested. However, as mentioned above, causality cannot be established in this analysis.

4.2 Paper II

Eiselt, K.-U., and R. G. Graversen (2023a): On the control of Northern Hemispheric feedbacks by AMOC: Evidence from CMIP and slab-ocean modeling. *J. Climate*, **36**, 6777-6795 , https://doi.org/10.1175/JCLI-D-22-0884.1.

An essential shortcoming of Paper I was the inability to establish causality because of the sole reliance on already published data from fully-coupled GCMs. Thus, the general aim of Paper II is to investigate causality by conducting proprietary model experiments with a simpler model.

The study begins with a further investigation of the GCM groups G1 and G2 defined based on the change over time of the LR feedback in Paper I. It is evident from Fig. 3 in Paper I that G1 and G2 are different in the Arctic (the focus of Paper I), but also that they strongly differ in the North Atlantic. An important factor influencing the climate in the North Atlantic is the Atlantic meridional overturning circulation (AMOC). This is an ocean circulation in the Atlantic driven by the formation of dense waters in the Arctic, and it transports warm waters from the tropics northward. A potential change of the AMOC induced by a CO_2 forcing may significantly impact local surface warming and feedbacks, potentially influencing the Arctic sea-ice melt as well. Multiple studies have established this connection (e.g., Jackson et al. 2015; Bellomo et al. 2021) and even a connection of the AMOC to the total global climate feedback change (Lin et al. 2019). The underlying hypothesis for the impact of a change of the AMOC on the global total feedback is consistent with the pattern effect (chapter 3): As the polar regions tend to be considerably more stably stratified than the tropics, a CO_2-forcing induced change in AMOC leads to the redistribution of surface warming between more and less stable regions and thus impacts the total global climate feedback (Lin et al. 2019).

The analysis of the model groups G1 and G2 with respect to the AMOC reveals significant differences between the model groups in the AMOC's pre-industrial control state as well as in its response to the abrupt quadrupling of CO_2. G2 exhibits a much stronger pre-industrial AMOC than G1 and it also changes the AMOC more in response to the CO_2 forcing. Another important finding is that after the initial rapid decline of the AMOC seen in both G1 and G2, only G2 exhibits a subsequent *recovery* of the AMOC. This is potentially significant as it is consistent with the stronger weakening of the climate feedback in G2 than in G1: While initially the AMOC declines more in G2 than in G1, inhibiting the distribution of heat from the less stable tropics to the more stable North Atlantic, the later recovery has the reverse effect, thus weakening the global total feedback over time. Further analysis is conducted, establishing that the local North Atlantic changes of surface temperature, stability, and surface energy balance are consistent with the expected influence of the AMOC change. Consequently, it is hypothesised that the difference of the change of the AMOC between G1 and G2 may contribute to the difference in change over time of their global total feedbacks.

To establish causality, in this study the Community Earth System Model version 2 (CESM2) in the slab-ocean model (SOM) configuration is employed to run proprietary experiments. The difference between a SOM and a fully-coupled GCM is that in the SOM the ocean is not dynamic, meaning that the ocean heat transport is prescribed and constant in time (except for a seasonal cycle). The ocean heat transport is typically referred to as Q-flux. Since a change of the AMOC is effectively a change in ocean heat transport it can be "mimicked" in a SOM by changing the Q-flux. Causality can thus be investigated: Atmospheric changes in a SOM experiment with mimicked AMOC change are solely due to the AMOC change. However, in the real world AMOC changes may feed back on themselves (e.g., Todd et al. 2020) which cannot be investigated in a SOM, implying potential differences between the effects of the mimicked AMOC change and "real" AMOC change in a fully-coupled model.

To investigate if the difference in climate feedback change between the GCM groups G1 and G2 may be explained by their difference in AMOC change, the following procedure was pursued: (1) Calculation of the difference of the change of the AMOC between G1 and G2 after the fast-paced early decline; (2) estimation of the energy flux change implied by this difference in AMOC change via a simple order-of-magnitude approach; (3) running of two SOM experiments, one with quadrupled CO_2 concentration (called no-dQ) and one with both quadrupled CO_2 concentration *and* implemented Q-flux change, corresponding to the AMOC change (called dQ); (4) comparison of the *differences* between dQ and no-dQ with the *differences* between G2 and G1 in terms of surface warming and radiative fluxes due to individual physical feedbacks. It must be noted that an exact quantitative reproduction of the difference between G2

and G1 cannot be expected, partly owing to the structural differences between SOMs and fully-coupled GCMs (e.g., different thermal inertias), as well as due to the fact that a single model is employed trying to investigate the differences between multi-model groups.

The general finding from this procedure is that the differences between G2 and G1 in surface warming and radiative fluxes can be qualitatively reproduced with SOM experiments in the Northern Hemisphere, and especially in the Northern extratropics. Significant impacts of the Q-flux change on the Arctic sea-ice extent are also found, qualitatively consistent with the differences between G2 and G1: Like G1, no-dQ loses most of its Arctic sea ice, while dQ, like G2, retains a considerable amount through the whole simulation. Related to the Arctic sea-ice extent, especially the differences of the Northern extratropical LR and SA as well as short-wave cloud radiative fluxes between G2 and G1 are qualitatively well reproduced in the SOM experiments. However, in the tropics as well as the Southern Hemisphere the SOM experiments perform less well. This may be due to the specific implementation of the "mimicking" of the AMOC change in the SOM experiment. Another reason may be that the SOM does not allow for a dynamic ocean response, potentially influencing warming and radiative-flux changes across the globe. Finally, there may be other differences between climate models (parametrisations, resolution, etc.) that may influence the response to a CO_2 forcing, but which are not considered here.

In conclusion, it may be said that the differences in surface warming and radiative fluxes induced by a change in AMOC in the SOM are consistent with the differences between the GCM groups G2 and G1. Thus, the different AMOC response between these two groups may partly drive the difference in feedback change over time. Another way of putting this is to say that the AMOC change influences the pattern effect. It is found here that this is mainly due to the influence of the AMOC on direct surface warming, as well as the LR, SA, and cloud feedbacks, likely related to impacts on the sea-ice extent. This is consistent with other recent studies (Lin et al. 2019; Singh et al. 2022). It must be noted again that this study investigates only one causal direction: The influence of the AMOC on other climate variables. A potential inverse relationship such as, e.g., sea-ice melt influencing the AMOC (e.g., W. Liu et al. 2019), is not investigated.

As a final point it is remarked that, consistent with earlier findings (e.g., Gregory et al. 2005; He et al. 2017), the pre-industrial AMOC strength correlates well with the pace and strength of the AMOC decline in response to the CO_2 forcing. In this context, the concept of "capacity to change" (Kajtar et al. 2021) is invoked, i.e., a given stronger process in the equilibrium state (stronger AMOC) may have the capacity to change more in response to a forcing, thus impacting the

forced response more. On the basis of this, a plea is made for more research into both the base-state[1] as well as the drivers of AMOC, as its potentially significant impact on the climate is increasingly confirmed.

4.3 Paper III

Eiselt, K.-U., and R. G. Graversen (2023b): On the impact of net-zero forcing Q-flux changes. *Submitted to Climate Dynamics*.

The purpose of this study is to apply the methodology of Paper II to more aspects of the pattern effect, motivated by historical surface-warming/cooling patterns (Zhang et al. 2021) as well as by studies showing disparate global impacts of warming/cooling in different geographical regions and teleconnections between regions (Lin et al. 2021; Dong et al. 2022a; Kim et al. 2022; Kang et al. 2023). Furthermore, the recognition of a strong change in the cloud feedback between the model used in Paper II (CESM2) and its predecessor (CESM1; see also the discussion in section 2.2.4) prompted a model comparison of the effect of different surface-warming/cooling patterns between the two model versions. To more specifically study the impact of the cloud parametrisation, additionally a recalibrated version of CESM2 (here called CESM2-Z22; Zhu et al. 2022) kindly provided freely online (Zhu 2021) is employed, which differs structurally from CESM2 only in terms of cloud parametrisation.

Three main experiments are run for this study: (1) The AMOC experiment from Paper I, but at pre-industrial CO_2 levels (dQ-AMOC), (2) an experiment with negative and positive Q-flux change in the East and the West Pacific, respectively (dQ-EP-WP), and (3) effectively the dQ-AMOC experiment "flipped" at the equator with the negative Q-flux region situated in the South Atlantic (dQ-SA-TA). The latter two experiments may be thought of as qualitatively mimicking the historically observed SST decline in the EP and the Southern Ocean. However, while dQ-AMOC was based on an order-of-magnitude estimation of the AMOC decline in response to $4xCO_2$ in fully-coupled climate model experiments (see the appendix in Paper II), no such estimation is conducted for dQ-EP-WP and dQ-SA-TA. Instead, to establish comparability between experiments, the same regionally integrated negative and positive Q-flux changes are applied as in dQ-AMOC. Notably, this leads to much stronger cooling than is historically observed.

1. Consider the discussion around the historical development of the AMOC in Caesar et al. (2021), Kilbourne et al. (2022), and Caesar et al. (2022)

A general finding, consistent with earlier studies (F. Liu et al. 2020), is that the negative Q-flux change is more effective at cooling than the positive Q-flux change is at warming. This is because especially in high latitudes cooling induced by the Q-flux change is amplified by the growth of sea ice and accompanying surface-albedo, lapse-rate, and cloud feedbacks.

A further finding is that, consistent with the notion of the pattern effect, different patterns of warming/cooling indeed have different effects on the global-mean temperature, and this is due to the different feedbacks activated by these patterns. dQ-AMOC exhibits a relatively small impact on the global-mean temperature compared to the other two experiments. In contrast, dQ-SA-TA has an immense impact, even causing a transition to a Snowball Earth state (see section 2.2.3). This is likely related to the differential geographic characteristics of the northern and southern polar regions: The Arctic sea ice is confined by continents and thus hindered in its expansion, while the Antarctic sea ice can freely grow across the Southern Ocean, facilitated by climatological circumpolar winds. In in the Southern Hemisphere mid-latitudes these winds are deflected at the mountainous western coasts of the continents (Africa, South America) into the tropical Atlantic and the tropical EP. Thus, after sufficient expansion of the Antarctic sea ice, these regions are efficiently cooled. Especially in the EP the SST decline induces an increase in low-cloud cover, reflecting more sunlight (see section 2.2.4), inducing further surface cooling (Dong et al. 2022a; Kim et al. 2022; Kang et al. 2023). This SST-cloud feedback loop accelerates the Antarctic sea-ice growth into the EP, eventually leading to the Earth being almost fully covered by sea ice, i.e., a Snowball Earth. Notably, also in dQ-EP-WP a Snowball Earth transition is encountered, although this occurs more slowly than in dQ-SA-TA. The reason for this is that the Q-flux in dQ-EP-WP does not immediately lead to sea-ice growth. However, a strong cooling cloud effect in the East and central Pacific is induced, giving rise to an initially slow but accelerating sea-ice growth, resulting in the eventual transition to a Snowball Earth.

In additional experiments with quadrupled CO_2 concentration the non-linearity of the combined response to Q-flux and CO_2 forcing is investigated. While dQ-AMOC is relatively linear, strong non-linearities are found in dQ-EP-WP and dQ-SA-TA, primarily related to the Snowball-Earth transition. However, a weaker, but considerable non-linear response is also found in a dQ-SA-TA experiment where the sea ice is deactivated. This is done by changing the freezing temperature of salt-water in the model code to a very low value. The non-linearity found in this experiment is due to the cloud feedback, especially over the Southern Ocean and the EP.

To at least tentatively investigate the dependence of the pattern effect on the climate model, the three main experiments (dQ-AMOC, dQ-EP-WP, dQ-SA-TA)

are repeated with CESM1 and CESM2-Z22. While the impacts of the same Q-flux changes are generally similar across models, CESM2 reacts more quickly to the Q-flux changes and exhibits especially larger cloud radiative flux changes than the other two models (again see the discussion in section 2.2.4). Since the largest structural difference between these models is the cloud parametrisation this indicates that differences in the pattern effect across models are related to their differences in cloud parametrisation.

Concluding, the article advises caution against the over-interpretation of the results, especially pertaining to the Snowball-Earth transition: In an experiment with a similar energy redistribution to dQ-SA-TA in the fully-coupled version of CESM2 (i.e., including a dynamical ocean model) the global impact is small.

5

Conclusion and Outlook

In this final chapter a summary and conclusions of the scientific work conducted for this thesis are offered. Problems especially concerning global climate models and their usage pertaining to ECS are discussed. Concluding, the scope is widened beyond ECS, and an attempt is made to give recommendations for future research.

5.1 Summary and Conclusions

The general purpose of this work is to contribute to the understanding of the climate system and its change due to an external forcing. Particularly, the focus is on the *change of climate feedback/sensitivity over time* in climate model simulations. In other words: On the non-linearity of the climate system response to a forcing. It has long been clear that the climate does not respond linearly to a forcing (Murphy 1995; Senior and Mitchell 2000; Gregory et al. 2004), but only over the last ten years has there emerged a robust theory explaining the why and how—that is, the *pattern effect* (e.g., Andrews et al. 2012; C. Zhou et al. 2016; Stevens et al. 2016). The principle research tool for establishing this theory were numerical global climate models (GCMs). Consistently, GCMs in the fully-coupled as well as in the slab-ocean configuration are the primary research tool in this thesis. The main aim of the thesis and the papers presented herein is to study the pattern effect, investigate its impacts specifically on physical

climate feedbacks, and contribute to the elucidation of its consequences for the climate system.

In general terms the findings of Papers I, II, and III can be summarised as follows. Paper I finds that in fully-coupled GCMs the changes over time in the cloud and the lapse-rate feedback dominate the time-dependence of the total feedback and sensitivity. Focusing on those models for which the lapse-rate feedback change is most important, the Indo-Pacfic warm-pool (IPWP) as well as the Arctic are identified as important regions potentially influencing the climate feedback change. This confirms both the eminence of the IPWP for climate feedback change found in Dong et al. (2019, 2020) and also, consistent with Lin et al. (2019), indicates the importance of a change in the Atlantic meridional overturning circulation (AMOC) for climate feedback change, although this was not discussed in Paper I. Paper II addresses this second point, confirming the importance of the AMOC for climate feedback change in fully-coupled models via correlation analysis as well as in a slab-ocean model (SOM) by mimicking the AMOC change found in the fully-coupled models. Feedback changes induced by the mimicked AMOC change are found to be qualitatively similar to differences in feedback changes between fully-coupled models, providing evidence for the importance of the AMOC for climate feedback and sensitivity. Finally, Paper III builds on Paper II and investigates the impact of different spatial redistributions of heat across the globe, again in a SOM. It is found that in contrast to the AMOC-type energy redistribution (i.e., in the tropical and North Atlantic), energy redistributions in the Southern Ocean and the tropical Pacific strongly affect the cloud feedback. A feedback loop between the Southern and East Pacific is found, intensifying the cooling initially induced by the energy redistribution, consistent with Dong et al. (2022a), Kang et al. (2023), and Kim et al. (2022). The magnitude of this effect in the SOM is strongly dependent on the presence of sea ice, potentially even taking the model into a Snowball-Earth state. Furthermore, the impact of the same spatial energy redistribution is found to be model-dependent, specifically on the cloud parametrisation.

A general conclusion from the theory of the pattern effect is that the *global* climate sensitivity depends on *regional* climate changes. Regarding simulation of climate change in response to a forcing this means that inadequacy of GCMs in terms of simulating regional changes may imply inadequacy in simulating the global response, including climate sensitivity. This is further discussed in section 5.2.
Combining the findings from Paper I—that is, when models with strong cloud feedback changes are excluded the relative IPWP warming correlates well with total feedback change over time—with the findings of Dong et al. (2020)—that is, while the feedback can be reasonably reconstructed from IPWP warming in CMIP5, this is not the case in CMIP6, likely due to cloud feedback—may indicate the dependence of the pattern effect on cloud parametrisation. This is confirmed

in Paper III by comparing three different versions of the Community Earth System Model (CESM), which primarily differ in their cloud parametrisation.
Based on the previous two conclusions, it is consistent that while most models *weaken* their feedback over time, there is still considerable diversity, with some models exhibiting little change and a few models even negative change (see section 1.3.3 and Paper I). Especially as it pertains to the lapse-rate feedback, this may be related to differences of AMOC development across models (Paper II).
A final, different type of conclusion, mainly from Paper III, is that caution is advised when using SOMs in experiments with Q-flux changes directly affecting the sea ice, especially in GCMs with more sensitive—potentially more realistic (Kang et al. 2023)—cloud parametrisations: The response of a SOM may be unrealistically large because of feedback loops that are impeded in fully-coupled models by ocean circulation changes (Paper III).

5.2 Future research: Problems and Recommendations

The theory of the pattern effect appears qualitatively fairly robust. However, as is typically the case with solutions to scientific problems (Deutsch 2011), also the pattern effect gives rise to new problems/questions which need further investigation. Based on the conclusions given in section 5.1, the following two questions may be raised:

1. Are the effects of the same pattern changes similar across GCMs or do they depend sensitively on parametrisations (see Paper III; Kang et al. 2023)[1]?

2. How do specific surface-warming/cooling patterns arise (e.g., Bellomo et al. 2021; Lee et al. 2022; Paper II)?

It was made clear in chapter 3 that the answers to these questions are highly important for robustly gauging future development of the climate system. The following offers some comments and recommendations for future research.

Parametrisations, especially of cloud processes, are one of the largest fundamental problems with GCMs (e.g., Stevens and Bony 2013). In Kang et al. (2023) and Paper III it was shown that the cloud parametrisation can have an

1. The upcoming Green's Function Model Intercomparison Project (GFMIP; Bloch-Johnson et al. 2023) may give important insights on this question.

important impact on how effective the pattern effect is at changing the global climate feedback. This highlights the importance of robust parametrisations in GCMs. However, a solution appears elusive, since increased complexity of GCMs and more sophisticated parametrisations rather than reducing model biases have "multiplied the ways in which these biases introduce uncertainties in climate simulations" (Stevens and Bony 2013, p. 1054). In a similar vein, the spread of ECS has *increased* from the previous to the current, more sophisticated, generation of GCMs, the cloud parametrisation being the main cause (Zelinka et al. 2020). As discussed in section 2.2.4, subsequent investigation has revealed issues with the more sophisticated cloud parametrisation of some of the newer GCMs from CMIP6 (Golaz et al. 2022; Mülmenstädt et al. 2021; Shaw et al. 2022; Zhu et al. 2020, 2021, 2022).

Palmer and Stevens (2019, p. 24392) even contend that the current GCMs are "not fit for purpose" (see also Parker 2009)—the purpose being regional climate projections and the assessment of large-scale changes from small-scale processes. According to Palmer and Stevens (2019), the methods applied to GCMs to alleviate their inadequacies—that is, empirically correcting or sampling models (e.g., via emergent constraints, see section 1.3) and incrementally improving model resolution and parametrisations—do not address the fundamental problem. The solution they suggest, called the *unified approach* by Katzav and Parker (2015), appears conceptually simple but suffers from practical difficulty: Echoing a call to action by Shukla et al. (2010), they suggest revolutionary change of GCM resolution from $O(100)$ km to $O(1)$ km. The fundamental improvement with this change is that at resolutions of $O(1)$ km essential processes (deep convection, ocean mesoscale eddies, etc.) that contribute to the model uncertainties and biases described above can be directly simulated and no longer need to be parameterised. However, achieving such high model resolution in the near future requires large international efforts reminiscent of those commissioned for particle accelerators or nuclear fusion (Shukla et al. 2010). These include not only the increase in computing power but also potential recoding of the GCMs to account for a more parallelised computer architecture, the development of new parametrisations and data assimilation procedures adequate for the $O(1)$ km scale, as well as new ways of handling the increased data output of the high-resolution models (Shukla et al. 2010). Katzav and Parker (2015) caution that even after the high-resolution models are implemented, it may take up to decades until meaningful statistics on these models' performance are collected and evaluated. They further note that apprehensions have been raised that efforts such as those championed by Palmer and Stevens (2019) and Shukla et al. (2010) intended to increase prediction skill *later* divert resources from efforts intended to decrease vulnerability *now*. An additional problem is that the resolution increases envisaged by Palmer and Stevens (2019) and Shukla et al. (2010) may not yield the advertised increased precision (see also Emanuel 2020). For example, Radtke et al. (2021) find that changing the resolution of a large eddy simulation (LES) model

from 5 km to 100 m considerably changes the subtropical marine low-cloud feedback (see section 2.2.4) from being robustly positive to near-zero. This feedback component is one of largest contributors to the uncertainty in ECS estimates in GCMs, and hence these changes appear significant. As a final problem, the limited knowledge of the current climate state should be mentioned. As an example, the Atlantic meridional overturning circulation (AMOC) has been shown to potentially significantly affect regional and global climate (e.g., Jackson et al. 2015; Bellomo et al. 2021; Paper II). However, the historical state and development of the AMOC is ill-constrained and debated (see Caesar et al. 2021 and the resulting discussion: Kilbourne et al. 2022, Caesar et al. 2022). This puts potentially incommensurable limitations on the verifiability of the high-resolution GCMs and on their capability to "sharpen" (Palmer and Stevens 2019, p. 24391) our view of our future climate.
More than a decade after the call to action by Shukla et al. (2010) considerable progress has been made and $O(1)$ km resolution models have been successfully run, but still only short simulations appear to be "within reach" (Hohenegger et al. 2023, p. 781; Mauritsen et al. 2022). For the time being, the usefulness of these high-resolution models may thus concentrate on assessing existing lower-resolution models (Sherwood and Forest 2023, Mauritsen et al. 2022).

Given these problems with the unified approach of revolutionising model resolution, other strategies of climate model improvement have been put forward. These include the *hierarchical approach* (Held 2005, 2014; Katzav and Parker 2015) and the *pluralist approach* (Katzav and Parker 2015). The idea of the hierarchical approach is to start by generating simpler, physics-based models and then build increasingly complex models on top of these, such that traceable model hierarchies are created (Held 2005; for a similar sentiment see Emanuel 2020)[2]. This may increase process understanding and ground climate models more robustly in basic physics. Katzav and Parker (2015) raise several issues with this strategy. For example, the extent to which such physically grounded model hierarchies can be built is likely limited, because of the inevitability of parametrisations in comprehensive models. Furthermore, the insights gained from simple models do not necessarily apply to more complex models due to potential non-linear feedbacks. Finally, Katzav and Parker (2015, p. 480) contend, that the hierarchical approach would require a "cultural change in climate modeling" from the traditional way of incrementally improving high-end models to systematic analysis of simpler models. Notably, a "cultural change" is required also in the unified and the pluralist approach. The pluralist approach calls for a substantial increase in model diversity to be able to more robustly quantify uncertainty. This may partly be done by data-driven models,

2. Particularly regarding ECS, Stevens and Kluft (2023) appears an impressive and promising effort in this direction.

which could give insights into "drivers and sensitivities of emergent climatic phenomena" (Katzav and Parker 2015, p. 482). Again however, problems with this approach include the difficulty of knowing how the model diversification should be accomplished, and how the "space of model structures" should be sampled (Katzav and Parker 2015, p. 482).

The here reviewed three strategies for model improvement have the potential of substantially contributing to the answers of the questions posed at the beginning of this section: Increased resolution and less reliance on parametrisations (especially for clouds) would increase the confidence in the radiative effect of the pattern effect (question 1). Furthermore, especially higher ocean resolution has been shown to lead to better representation of surface-warming patterns (question 2; e.g., Bilgen and Kirtman 2020; Roberts et al. 2020; Winton et al. 2014). Given the inevitability of parametrisations even at higher resolutions and the accompanying uncertainties, hierarchical models may provide a clearer and more precise formulation of the pattern effect as a theory, increasing confidence in its origins (question 2) and its consequences (question 1). More model diversity may help to establish to what extent observed patterns are the result of internal variability or historically forced (question 2), and, more generally, discover potentially previously unconsidered processes and connections.

The above review of problems with GCMs and the prominent suggested future strategies indicates that the current strategy of model development and application may remain dominant for some time. Given this situation, recommendations are made further below for more incremental GCM improvement and usage. However, before turning to specific recommendations, a further recently highlighted problem with GCMs, especially pertaining to estimating ECS, warrants discussion. One of the basic assumptions in the forcing-feedback-sensitivity framework for estimating ECS in GCM forcing experiments is that the system is originally in an equilibrium state (section 1.1). Indeed, this is not generally the case in GCMs, meaning that many GCMs do not exhibit equilibrium conditions in the pre-industrial control experiment (piControl) from which the forcing experiments (e.g., abrupt4xCO_2) are branched (e.g., Irving et al. 2021). These GCMs exhibit a net imbalance at TOA in the control state, implying a energy leaks. This is problematic, because one cannot necessarily expect a forced GCM to return to the same state of imbalance as before the forcing. Indeed, exploiting the millenial time-scale LongRunMIP (Rugenstein et al. 2019) data, Sanderson and Rugenstein (2022) find that not even all GCMs that do exhibit zero net imbalance in the piControl experiment return to the equilibrium state. This may be related to "model tuning" (e.g., Hourdin et al. 2017; Mauritsen et al. 2012; Mauritsen and Roecker 2020). In order to faithfully represent the climate system, GCMs typically undergo a tuning process in which uncertain or even unobservable parameters are changed and optimised until certain observable climate variables are simulated sufficiently close to

their observed values. Such climate variables are, e.g., the global mean surface temperature (Mauritsen et al. 2012), but climate sensitivity has also been used (Mauritsen and Roecker 2020). According to Hourdin et al. (2017), parameters in the cloud parametrisation are most often adjusted in the tuning process to optimise the target variable, followed by snow and sea-ice albedo, ocean mixing, and orographic drag. Sanderson and Rugenstein (2022) distinguish two types of models: (1) structurally balanced models, in which all energy leaks were systematically eliminated, and (2) tuned balanced models, in which parameters were tuned such that the model is in balance in the piControl experiment. The latter kind of models may by inappropriate for climate forcing experiments to estimate ECS. This should be taken into account when considering ECS estimates from GCMs and only such GCMs should be used for which confidence in a structurally balanced equilibrium state is established (Sanderson and Rugenstein 2022).

Tentative Recommendations and Suggestions

The recommendations offered here consider two different types of research:

1. The estimation of ECS, i.e., the *equilibrium* response to a forcing and

2. the projection of climate change over the next 50 to 100 years, i.e., the *transient* response to a forcing.

Notably, these two issues are related in important ways. That is, ECS is often used to make inferences about the transient response, e.g., in climate emulators to estimate remaining carbon budgets (e.g., IPCC 2021, pp. 994-995; Rypdal et al. 2021) or in Integrated Assessment Models (IAMs) to investigate socio-economic climate impacts (e.g., National Academies of Sciences, Engineering, and Medicine 2016). Thus, even though the usefulness of ECS for estimating the near-term climate response has been challenged (Knutti et al. 2017; Sanderson 2020; Sanderson and Rugenstein 2022; Sherwood and Forest 2023; Tokarska et al. 2020), under current practice it remains highly relevant and thus its estimation should be improved. The following gives recommendations for future research on ECS based on the currently available type of GCM.
As found by Sanderson and Rugenstein (2022), several GCMs are not *structurally* balanced and thus potentially do not return to an equilibrium state after a forcing. Thus, it may be recommended that for estimating ECS only such models be used that are structurally balanced. More fundamentally, in the development process it should be attempted to structurally balance models and to use as little tuning as possible. Further recommendations concern the time-dependence of climate sensitivity. As discussed in section 1.3.3, this hampers a robust estimation of the "true" ECS from shorter ($\mathcal{O}(100)$ years)

GCM simulations. Rugenstein and Armour (2021) find that 400 simulation years of abrupt4xCO2 may be enough to estimate the "true" ECS in a GCM to within 5% accuracy with the Gregory method (see section 1.3.3). Following Rugenstein and Armour (2021) it is thus recommended to expand the CMIP abrupt4xCO2 protocol from 150 years (Eyring et al. 2016) to at least 400 years. Another recommendation for CMIP arises from the recognition of two points: (1) The high-emissions "business-as-usual" scenarios RCP8.5 and SSP5-8.5 appear increasingly unlikely (Hausfather et al. 2022; Hausfather and Peters 2020; Pielke et al. 2022; Pielke and Ritchie 2021a) and (2) the climate system response does not change *monotonically* with CO_2-forcing (Bloch-Johnson et al. 2021; Mitevski et al. 2021). Thus, additionally to the current abrupt4xCO2 experiment, the CMIP protocol should include the abrupt2xCO2 experiment intended for the estimation of ECS. The radiative forcing due to $2xCO_2$ is about 4 Wm^{-2} and thus more in line with the currently more plausible RCP4.5 or SSP2-4.5 scenarios (Hausfather and Peters 2020; IPCC 2021, p. 239; according to Pielke et al. 2022 the SSP 3.4 scenarios are most plausible). It must be noted that in abrupt2xCO2 more years may be necessary to robustly determine "true" ECS than in abrupt4xCO2 (Dai et al. 2020), potentially implying an upward adjustment of the 400-year recommendation given above.

These recommendations do not address the problem of the failure of GCMs to correctly simulate the historical surface warming patterns. Notably, further, potentially related inadequacies of historical simulations exist. These include too much global warming especially during the recent 2000-2012 hiatus period (see e.g., Modak and Mauritsen 2021 for a brief review) and excessive tropical upper tropospheric warming (McKitrick and Christy 2018, 2020), potentially being "smoking guns" for climate modelling (American Physical Society 2014, p. 487). The latter point is especially interesting in the context of the relative lack of warming of the IPWP in GCMs compared to observations. As has been discussed in chapter 3, due to strong convective activity, the IPWP tends to control upper tropospheric warming. Thus, the relative lack of warming of the IPWP in historical GCM simulations compared to observations seems somewhat inconsistent with the excessive upper tropospheric warming. A further issue with GCMs, likely related to their too strong historical global warming simulation (e.g., Scafetta 2022), is the inadequacy of the representation of longer-term (decadal to multi-decadal) ocean cycles such as the Pacific Decadal Oscillation (PDO) or the Atlantic Multidecadal Oscillation (AMO) (e.g., Coburn and Pryor 2021). The AMO has been shown to impact global-mean temperatures by between 0.1 and 0.4 K (Tung and J. Zhou 2013, Chylek et al. 2016). These oscillations may have consequences for ECS: Correcting the global-mean temperature for the AMO index, e.g., Tung and J. Zhou (2013) found that only about half of the temperature increase since 1950 was due to anthropogenic influences. Based on this, Loehle (2014) estimated an ECS of about 2 K, a comparatively low value, however in line with other ECS estimates derived from observations

(e.g., Otto et al. 2013, Lewis and Curry 2018; see section 1.3.1). Notably, this was subsequently found to be dataset dependent (Cawley et al. 2015; Loehle 2015). A final, somewhat different issue with current GCMs to be raised here is that they essentially assume flat geometry in the atmosphere. This has been found to cause an underestimation of the solar radiation intercepted by the sun, leading to global and regional forcing differences (Prather and Hsu 2019), which in light of the pattern effect may be non-negligible for the climate state and climate sensitivity.

On the basis of these problems further recommendations are as follows. In future model development higher value should be concentrated on the correct representation of longer-term variability related to ocean oscillations. This pertains especially to model tuning: If a GCM is tuned to the observed temperature record, which may be substantially influenced by oscillations like the AMO, but such oscillations are inadequately represented in the GCM, this may lead to "overtuning" to anthropogenic forcing, potentially biasing the GCM's climate sensitivity[3]. Here the hierarchical approach suggested by Held (2005, 2014) discussed above may help to ground GCMs more on a physical basis and better represent longer-term internal variability related to ocean oscillations.

Moving beyond ECS, the usage of GCMs to project the near-term future (2050-2100) is here addressed, as this appears the most immediate problem climate science should give reliable guidance on. It was mentioned above that ECS estimates are often used to investigate the near-term future in climate emulators or IAMs. An important question regarding these methods is if they are indeed able to give a robust range of potential real-world outcomes of climate trajectories. Using GCMs under future emissions scenarios may offer such a range of trajectories, but also here caution is advised. For example, internal variability, especially from the ocean oscillations mentioned above, needs to be considered. As was indicated, these oscillations are somewhat under-represented in GCMs, but, more importantly, they are not necessarily in phase across GCMs. This means that only considering multi-model means, while isolating the forced response, would tend to average out the impact of these oscillations (Curry and Webster 2011), hereby reducing the range of considered real-world trajectories. Simulations by GCMs may thus be supplemented by semi-empirical scenarios, attempting to cover the full range of potential outcomes (see e.g., Curry 2023, pp. 97-107). Notably, Curry (2023) mostly finds a cooling influence form natural variability, as well as potential volcanic eruptions, and solar forcing changes. However, the pattern effect, especially in terms of the East-to-West Pacific surface-warming gradient (see chapter 3) is not considered, which may negate this influence if the gradient reverses in the near future, causing an

3. Hence, while the *exact prediction* of internal varialbility may be viewed as "icing on the cake" (Held 2014, p. 1206), this is not true of the *sufficient representation* of internal variaiblity in GCMs.

acceleration of the warming (e.g., Dong et al. 2021; C. Zhou et al. 2021). Finally, it has been found that a strong reduction of human emissions may lead to a short-term warming effect (due to decreased cooling from aerosols; Dvorak et al. 2022)[4]. Again, for possible real-world trajectories of global climate change this needs to be considered.

Concluding, a shift in thinking about climate sensitivity research, especially pertaining to ECS, is here suggested. Instead of viewing research on ECS as directly applicable to policy decisions, it may be more appropriate to think of it as "basic research" (see e.g., Emanuel 2020, p. 3), offering fundamental insights into the climate system. This is not to diminish the role of this endeavour, since as Emanuel (2020, p. 3) notes: "[L]ong-term progress depends [...] on the rare leaps forward made possible by basic research." One might consider the pattern effect as one these "leaps forward" in the understanding of the climate system.

As regards the more applied, policy-relevant side of climate science, the focus should be on real-world possibilities and outcomes, with a robust treatment of risk and uncertainty (e.g., Curry and Webster 2013; Aven 2020; Curry 2023). One path for applied climate science is the emissions-scenario approach, which has been dominant since its adoption by the IPCC's First Assessment Report (IPCC 1990; Pielke and Ritchie 2021a,b). However, while the approach itself seems sound, there appears to be a lock-in on certain specific scenarios which may be problematic if indeed these scenarios diverge from plausible future pathways. It is thus to be recommended that higher priority be placed on *plausibility* in the construction of future scenarios and to not use a potentially misleading "business-as-usual" terminology, concentrating climate science on scenarios with little footing in real-world developments[5] (Hausfather and Peters 2020; Pielke et al. 2022; Pielke and Ritchie 2021a,b; Ritchie and Dowlatabadi 2017). This is not to say that research on these scenarios should be halted, but since scientific resources in terms of computation and researcher time are limited, they should not be the main focus, as is current practice (Pielke and Ritchie 2021b)[6]. Recently, the idea of using "storylines" to assess and communicate uncertainties, risks, and potential impacts has gained increasing prominence (T. G. Shepherd et al. 2018). Note that this is the same concept championed by Stevens et al. (2016) for the shift in strategy in ECS research (see section 1.2). As it pertains to climate impacts, the concept of storylines is built on the recognition of the

4. However, given China's recent abandoning of the Paris Agreement (C. Shepherd et al. 2023), it appears unlikely that the cessation of anthropogenic emissions is achieved in the near future.
5. Consider especially the apparently inappropriate treatment of coal in the "business-as-usual" scenarios (Ritchie and Dowlatabadi 2017).
6. While Pielke and Ritchie (2021b) have been strongly criticised for their contentions (Field et al. 2021), their main points pertaining to the discussion here (see the text) appear sound.

asymmetry between semantic and experiential/episodic memory or knowledge. That is, experiencing, e.g., an extreme weather event is much more effective at conveying the risk such an event poses than just knowing (semantically) about such an event. A consequence is that adaptation policies are often only enacted *after* a certain severe event occurred with devastating impact (see the examples give in T. G. Shepherd et al. 2018). As climate change may induce shifts in weather patterns and thus previously not experienced weather events (at least locally), the construction and "playing through" of specific storylines may incentivise appropriate adaptation measures to build resilience against extreme events. Given the often inadequate preparedness for extreme weather even under the current climate, applying the storyline approach appears beneficial also in the absence of local weather pattern changes. Combining the storyline and the emissions-scenario approach may give people the tools to adapt to climate change locally and help in decision-making for implementing the most effective mitigation measures on a more global level.

As opposed to a “tame” problem where general agreement exists on both the problem and the solution, climate change has been called a “wicked” problem (Rittel and Webber 1973; Curry and Webster 2013; Head 2019), involving “a conflict of visions” (Sowell 2007 [1987]) between concerned parties regarding the status of the problem and its solution(s). Indeed, *solutions* to such problems may not exist and the best that can be achieved is a “prudent trade-off” (Sowell 2007 [1987], p. 25). This confronts humanity with “deep uncertainty” (e.g., Simpson et al. 2016) which, even as climate and other sciences progress, will likely fundamentally remain, and thus needs to be carefully treated when it comes to decision-making (e.g., Curry 2023, pp. 153-156, 181-191). Even though, rather than applied research directly pertaining to decision-making, the current thesis presents more “basic” research (Emanuel 2020), it is hoped that a “meaningful” (T. G. Shepherd and Lloyd 2021) contribution has been made to the understanding of the climate system.

Closing Remarks: A Note on Epistemology

It appears to be a tendency in modern scientific research to just “do the science”, albeit without reflection on the kinds of claims that can be made on the basis of the employed methods and tools. That is, there is little reflection on *epistemology*. However, given that especially climate science concerns potentially disruptive climate change on the one hand, and potentially equally disruptive policy changes on the other hand, an inquiry into the epistemological status of climate science may be fruitful.
In some circles, the epistemology of climate models (one of the most important tools of climate science) seems to be vigorously discussed, particularly in terms of their *falsifiability* (e.g., Curry and Webster 2013; Loehle 2018; Parker 2009;

American Physical Society 2014; Happer and Lindzen 2023). Falsifiability—that is, the capability of a hypothesis to be rejected on the basis of empirical evidence—is an important epistemological concept for science (e.g., Popper 2002 [1959]). Disagreement appears to exist on the question of if climate models can be falsified (compare e.g., American Physical Society 2014, p. 179 and Happer and Lindzen 2023 with Curry and Webster 2013; Loehle 2018), although this may be related to a lack of clarity in the usage of the term "falsifiability". A more nuanced view is presented by Parker (2009) who argues that climate models can be tested for "adequacy-for-purpose" but not for truth or empirical adequacy. This is because it is clear *a priori* that climate models cannot represent reality or be considered empirically adequate, due to the significant simplifications necessary in their construction (see also Loehle 2018).

One of the main purposes of climate models is the prediction of the development of climate variables (e.g., temperature) under future changes (e.g., of the CO_2 concentration) to the climate system. However, the testing of the adequacy of a climate model for this specific purpose is non-trivial (Parker 2009). One plausible test may be to investigate if the climate model can adequately simulate the climate variable in question over the historical period for which observational records exist. According to Parker (2009, p. 242), such a test may provide the "least controversial" case of model "disconfirmation": When the climate model consistently cannot simulate the climate variable in question to within a specified margin of error compared to observational data. Thus, this may be the extent to which climate models can be "falsified".

However, caution is advised in too readily rejecting a given climate model on this basis. Especially for emergent, higher-level quantities[7] and during non-stationary/transient states, the *constancy principle of causality*—that is, same cause, same effect (e.g., Hoppe 1983, p. 11)[8]—may not be applicable. For example, while under pre-industrial (i.e., stationary) conditions multi-decadal changes of the global-mean surface temperature appear to be driven by ocean oscillations, under increasing CO_2 forcing (i.e., in a transient state) this may shift and the CO_2 forcing may become the main driver. If a climate model is adequate for simulating temperature changes in response to CO_2 forcing but inadequate for simulating temperature changes caused by internal ocean oscillations, this implies that the model, while inadequate over most of the historical period, may become adequate-for-purpose under future emissions. However, these considerations yield little epistemological gain, since it is not clear (1) when and even if this shift to adequacy might occur, and (2) if the model indeed is adequate for simulating the temperature response to CO_2 forcing. The temperature response of the complex climate system to the CO_2

7. Because of the myriad processes included in a climate model even the temperature may be viewed as such an emergent, higher-level quantity.
8. As Hoppe (1983, p. 12) notes: The applicability of the constancy principle of causation is a requirement for strict falsification.

forcing is strongly dependent on feedbacks (see chapter 2) and, at least in a transient state, on the patterns of the temperature change (chapter 3), implying considerable uncertainty. A further problem is the tuning of climate models, which, due to its *ad hoc* character, introduces some uncertainty about model fidelity (Parker 2009)[9]. Tuning has the potential to make models too sensitive or too insensitive to CO_2 forcing, e.g., if internal variability is underestimated (e.g., Curry and Webster 2011) or negative historical forcings (e.g. due to anthropogenic aerosols) are overestimated (e.g., Stevens 2017; Robson et al. 2022; Curry and Webster 2011).
Thus, while climate models should not be outright rejected when they are supposedly "falsified", as argued by some (e.g., Happer and Lindzen 2023), caution is advised against the overinterpretation of modelling results, especially when it comes to attributing cause and effect, and a claim made on the basis of climate models should be qualified by epistemic humility, "made complete with a statement of its own limitations" (Thompson and L. A. Smith 2019, p. 10).

Moving beyond the epistemology of climate models to climate science and its interface with social science, politics, and public policy, according to Hulme (2011, p. 256) an "epistemological slippage" has occurred, leading to the resurrection of climate determinism in the guise of climate reductionism: "[T]he knowledge claims of climate modelers are transferred [...] to the putative knowledge claims of the social, economic, and political analysts." However, "[t]hese models and calculations allow for little human agency, little recognition of evolving, adapting, and innovating societies, and little endeavor to consider the changing values, and practices of humanity." This is an expression of the inapplicability of the constancy principle of causality particularly in the social sciences (Hoppe 1983). Studies of climate change impacts should thus be exceptionally cautious and always frame their conclusions in terms of their assumptions, lest they be catastrophised or downplayed.

The recent increasing focus on the storyline approach, both in the physical climate sciences (Stevens et al. 2016) and at the climate science–policy interface (T. G. Shepherd et al. 2018), appears an attempt at shifting climate science towards a more robust standard of *rejecting* unlikely outcomes instead of the building of *consensus*[10] in certain estimated values of climate system metrics

9. According to Curry and Webster (2013, p. 1670), tuning "provides a means for the model to avoid being falsified." However, as noted at the beginning of this section, the meaning of "falsified" is somewhat unclear here.
10. Potentially even more so than GCMs, the epistemological basis of the consensus approach warrants inquiry. However, this is beyond the scope of this thesis, as the focus here is specifically on climate models. For critical examinations of the consensus approach in climate science see, e.g., Curry and Webster (2011, 2013), Oppenheimer et al. (2007), and van der Slujis et al. (2010).

(e.g., ECS), which consequently come to dominate public policy (e.g., Roe and Bauman 2013; Hulme 2011). This warrants a certain degree of hope for a more sophisticated and nuanced approach, particulary with regards to climate modelling, leading to more robust scientific claims and a better treatment of uncertainty in a climate science trying to earn “the public’s trust” (Warren 2022).

Bibliography

Alexeev, V. A. (2003). "Sensitivity to CO_2 doubling of an atmospheric GCM coupled to an oceanic mixed layer: a linear analysis." In: *Climate Dyn.* 20, pp. 775–787. DOI: 10.1007/s00382-003-0312-x.

Alexeev, V. A., P. L. Langen, and J. R. Bates (2005). "Polar amplification of surface warming on an aquaplanet in "ghost forcing" experiments without sea ice feedbacks." In: *Climate Dyn.* 24, pp. 655–666. DOI: 10.1007/s00382-005-0018-3.

Allen, M. R. and W. J. Ingram (2002). "Contraints on future changes in climate and the hydrological cycle." In: *Nature* 419, pp. 224–231. DOI: 10.1038/nature01092.

Andrews, T., J. M. Gregory, M. J. Webb, and K. E. Taylor (2012). "Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models." In: *Geophys. Res. Lett.* 39, L09712. DOI: 10.1029/2012GL051607.

Andrews, T., J. M. Gregory, and M. J. Webb (2015). "The dependence of radiative forcing and feedback on evolving patterns of surface temperature change in climate models." In: *J. Climate* 28, pp. 1630–1648. DOI: 10.1175/JCLI-D-14-00545.1.

Andrews, T. et al. (2018). "Accounting for changing temperature patterns increases historical estimates of climate sensitivity." In: *Geophys. Res. Lett.* 45, pp. 8490–8499. DOI: 10.1029/2018GL078887.

Andrews, T. et al. (2021). "Effective radiative forcing in a GCM with fixed surface temperatures." In: *J. Geophys. Res. Atmos.* 39, e2020JD033880. DOI: 10.1029/2020JD033880.

Andrews, T. et al. (2022). "On the effect of historical SST patterns on radiative feedback." In: *J. Geophys. Res. Atmos.* 127, e2022JD036675. DOI: 10.1029/2022JD036675.

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

American Physical Society (2014). *Climate Change Statement Review Workshop. Transcript of Porceedings*. URL: https://www.aps.org/policy/statements/upload/climate-seminar-transcript.pdf.

Armour, K., C. M. Blitz, and G. H. Roe (2013). "Time-varying climate sensitivity from regional feedbacks." In: *J. Climate* 26, pp. 4518–4534. DOI: 10.1175/JCLI-D-12-00544.1.

Arrhenius, S. (1896). "On the influence of carbonic acid in the air upon the temperature of the ground." In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 5, pp. 237–276. DOI: 10.1080/14786449608620846.

Aven, T. (2020). "Climate change risk - what is it and how should it be expressed." In: *Journal of Risk Research* 23, pp. 1387–1404. DOI: 10.1080/13669877.2019.1687578.

Bacmeister, J. T. et al. (2020). "CO_2 Increase Experiments Using the CESM: Relationship to Climate Sensitivity and Comparison of CESM1 to CESM2." In: *Journal of Advances in Modeling Earth Systems* 12, e2020MS002120. DOI: 10.1029/2020MS002120.

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." In: *Nat. Commun.* 12, 3659. DOI: 10.1038/s41467-021-24015-w.

Bilgen, S. I. and B. P. Kirtman (2020). "Impact of ocean model resolution on understanding the delayed warming of the Southern Ocean." In: *Environ. Res. Lett.* 15, 114012. DOI: 10.1088/1748-9326/abbc3e.

Bintanja, R., R. G. Graversen, and W. Hazeleger (2011). "Arctic winter warming amplified by the thermal inversion and consequent low infrared cooling to space." In: *Nat. Geosci.* 4, pp. 758–761. DOI: 10.1038/NGEO1285.

Bitz, C. M. et al. (2012). "Climate sensitivity of the Community Climate System Model, Version 4." In: *J. Climate* 25, pp. 3053–3070. DOI: 10.1175/JCLI-D-11-00290.1.

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

Bloch-Johnson, J. et al. (2021). "Climate sensitivity increases under higher CO_2 levels due to feedback temperature dependence." In: *Geophys. Res. Lett.* 48, e2020GL089074. DOI: 10.1029/2020GL089074.

Bloch-Johnson, J., M. Rugenstein, J. Gregory, B. B. Cael, and T. Andrews (2022). "Climate impact assessment should not discount 'hot' models." In: *Nature* 608, p. 667. DOI: 10.1038/d41586-022-02241-6.

Bloch-Johnson, J. et al. (Mar. 2023). "The Green's Function Model Intercomparison Project (GFMIP) protocol." In: *Authorea*. DOI: 10.22541/essoar.167839939.92474288/v2.

Block, K. and T. Mauritsen (2013). "Forcing and feedback in the MPI-ESM-LR coupled model under abruptly quadrupled CO_2." In: *J. Adv. in Model. Earth Syst.* 5, pp. 676–691. DOI: 10.1002/jame.20041.

Boeke, R. C., P. C. Taylor, and S. A. Sejas (2021). "On the Nature of the Arctic's Positive Lapse-Rate Feedback." In: *Geophys. Res. Lett.* 48, e2020GL091109. DOI: 10.1029/2020GL091109.

Boer, G. J. and B. Yu (2003a). "Climate sensitivity and climate state." In: *Climate Dyn.* 21, 167–176. DOI: 10.1007/s00382-003-0323-7.

– (2003b). "Dynamical aspects of climate sensitivity." In: *Geophys. Res. Lett.* 30, 1135. DOI: 10.1029/2002GL016549.

Bony, S. et al. (2020). "Observed modulation of the tropical radiation budget by deep convective organization and lower-tropospheric stability." In: *AGU Advances* 1, e2019AV000155. DOI: 10.1029/2019AV000155.

Bretherton, C. S. and P. M. Caldwell (2020). "Combining emergent constraints for climate sensitivity." In: *J. Climate* 33, pp. 7413–7430. DOI: 10.1175/JCLI-D-19-0911.1.

Brient, F. (2020). "Reducing uncertainties in climate projections with emergent constraints: Concepts, examples and prospects." In: *Adv. Atmos. Sci.* 37, pp. 1–15. DOI: 10.1007/s00376-019-9140-8.

Budyko, M. I. (1969). "The effect of solar radiation variations on the climate of the Earth." In: *Tellus* 5, pp. 611–619. DOI: 10.1111/j.2153-3490.1969.tb00466.x.

Caesar, L., G. D. McCarthy, D. J. R. Thornally, N. Cahill, and S. Rahmstorf (2021). "Current Atlantic meridional overturning circulation weakest in last millenium." In: *Nat. Geosci.* 14, pp. 118–120. DOI: 10.1038/s41561-021-00699-z.

– (2022). "Reply to: Atlantic circulation change still uncertain." In: *Nat. Geosci.* 15, pp. 168–170. DOI: 10.1038/s41561-022-00897-3.

Caldwell, P. M., M. D. Zelinka, K. E. Taylor, and K. Marvel (2016). "Quantifying the Sources of Intermodel Spread in Equilibrium Climate Sensitivity." In: *J. Climate* 29, pp. 513–524. DOI: 10.1175/JCLI-D-15-0352.1.

Caldwell, P. M., M. D. Zelinka, and S. A. Klein (2018). "Evaluating emergent constraintes on equilibrium climate sensitivity." In: *J. Climate* 31, pp. 3921–3942. DOI: 10.1175/JCLI-D-17-0631.1.

Cawley, G. C., K. Cowtan, R. G. Way, P. Jacobs, and A. Jokimäki (2015). "On a minimal model for estimating climate sensitivity." In: *Ecological Modelling* 297, pp. 20–25. DOI: 10.1016/j.ecolmodel.2014.10.018.

Ceppi, P. and J. M. Gregory (2017). "Relationship of tropospheric stability to climate sensitvity and Earth's observed radiation budget." In: *Proc. Natl. Acad. Sci. (USA)* 114, pp. 13126–13131. DOI: 10.1073/pnas.1714308114.

Cesana, G. and A. D. Del Genio (2021). "Observational constraint on cloud feedbacks suggests moderate climate sensitivity." In: *Nat. Clim. Change* 11, pp. 213–218. DOI: 10.1038/s41558-020-00970-y.

Cesana, G., D. E. Waliser, X. Jiang, and J.-L. F. Li (2015). "Multimodel evaluation of cloud phase transition using satellite and reanalysis data." In: *J. Geophys. Res. Atmos.* 120, pp. 7871–7892. DOI: 10.1002/2014JD022932.

Cess, R. D. (1975). "Global climate change: an investigation of atmospheric feedback mechanisms." In: *Tellus* 3, pp. 193–198. DOI: 10.1111/j.2153-3490.1975.tb01672.x.

Cess, R. D. et al. (1990). "Intercomparison and interpretation of climate feedback processes in 19 atmospheric general circulation models." In: *J. Geophys. Res.* 95, pp. 16601–16615. DOI: 10.1029/JD095iD10p16601.

Charney, J. G. et al. (1979). *Carbon Dioxide and Climate: a Scientific Assessment*. Report of an Ad Hoc Study Group on Carbon Dioxide and Climate. National Academy of Sciences.

Chylek, P., J. D. Klett, M. K. Dubey, and N. Hengartner (2016). "The role of Atlantic Multi-decadal Oscillation in the global mean temperature variability." In: *Climate Dyn.* 47, pp. 3271–3279. DOI: 10.1007/s00382-016-3025-7.

Coburn, J. and S. C. Pryor (2021). "Differential credibility of climate modes in CMIP6." In: *J. Climate* 34, pp. 8145–8164. DOI: 10.1175/JCLI-D-21-0359.1.

Colman, R. and B. J. Soden (2021). "Water vapor and lapse rate feedbacks in the climate system." In: *Rev. Mod. Phys.* 93, 045002. DOI: 10.1103/RevModPhys.93.045002.

Connolly, R. et al. (2021). "How much has the Sun influenced Northern Hemisphere temperature trends? An ongoing debate." In: *Res. Astron. Astrophys.* 21, 131. DOI: 10.1088/1674-4527/21/6/131.

Croll, J. (1864). "On the physical cause of the change of climate during geological epochs." In: *Philos. Mag.* 28, pp. 121–137. DOI: 10.1080/14786446408643733.

Curry, J. A. (2023). *Climate uncertainty and risk. Rethinking our response*. New York, NY, USA: Anthem Press. 290 pp.

Curry, J. A. and P. J. Webster (2011). "Climate science and the Uncertainty Monster." In: *Bull. Amer. Meteor. Soc.* 92. DOI: 10.1175/2011BAMS3139.1.

– (2013). "Climate change: No consensus on consensus." In: *CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources* 8, 001. DOI: 10.1079/PAVSNNR20138001.

Dai, A., D. Huang, B. E. Rose, J. Zhu, and X. Tian (2020). "Improved methods for estimating equilibrium climate sensitivity from transient warming simulations." In: *Climate Dyn.* 54, pp. 4515–4543. DOI: 10.1007/s00382-020-05242-1.

Dai, A. and M. T. Jenkins (2023). "Relationships among Arctic warming, sea-ice loss, stability, lapse rate feedback, and Arctic amplification." In: *Climate Dyn.* DOI: 10.1007/s00382-023-06848-x.

Danabasoglu, G. et al. (2020). "The Community Earth System Model Version 2 (CESM2)." In: *Journal of Advances in Modeling Earth Systems* 12, e2019MS001916. DOI: 10.1029/2019MS001916.

Danabasoglu, G. and P. R. Gent (2009). "Equilibrium climate sensitivity: Is it accurate to use a slab ocean model?" In: *J. Climate* 22, pp. 2494–2499. DOI: 10.1175/2008JCLI2596.1.

Deutsch, David (2011). *The Beginning of Infinity. Explanations that Transform the World.* UK: Allen Lane. 496 pp.

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." In: *J. Climate* 32, pp. 5471–5491. DOI: 10.1175/JCLI-D-18-0843.1.

Dong, Y. et al. (2020). "Intermodel spread in the pattern effect and its contribution of climate sensitivity in CMIP5 and CMIP6 models." In: *J. Climate* 33, pp. 7755–7775. DOI: 10.1175/JCLI-D-19-1011.1.

Dong, Y. et al. (2021). "Biased estimates of equilibrium climate sensitivity and transient climate response derived from historical CMIP6 simulations." In: *Geophys. Res. Lett.* 48, e2021GL095778. DOI: 10.1029/2021GL095778.

Dong, Y., K. C. Armour, D. S. Battisti, and E. Blanchard-Wrigglesworth (2022a). "Two-Way Teleconnections between the Southern Ocean and the Tropical Pacific via a Dynamic Feedback." In: *J. Climate* 35, pp. 2667–2682. DOI: 10.1175/JCLI-D-22-0080.1.

Dong, Y., A. G. Pauling, S. Sadai, and K. C. Armour (2022b). "Antarctic ice-sheet meltwater reduces transient warming and climate sensitivity through the sea-surface temperature pattern effect." In: *Geophys. Res. Lett.* 49, e2022GL101249. DOI: 10.1029/2022GL101249.

Dunne, J. P. et al. (2020). "Comparison of equilibrium climate sensitivity estimates from slab ocean, 150-year, and long simulations." In: *Geophys. Res. Lett.* 47, e2020GL088852. DOI: 10.1029/2020GL088852.

Dvorak, M. T. et al. (2022). "Estimating the timing of geophysical commitment to 1.5 and 2.0 °C of global warming." In: *Nat. Clim. Change* 12, pp. 547–552. DOI: 10.1038/s41558-022-01372-y.

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." In: *J. Climate* 35, pp. 2919–2932. DOI: 10.1175/JCLI-D-21-0623.1.

– (2023a). "On the control of Northern Hemispheric feedbacks by AMOC: Evidence from CMIP and slab-ocean modeling." In: *J. Climate* 36, pp. 6777–6795. DOI: 10.1175/JCLI-D-22-0884.1.

– (2023b). *On the impact of net-zero forcing Q-flux change*. Submitted to Climate Dynamics.

Emanuel, K. (2020). "The relevance of theory for contemporary research in atmospheres, oceans, and climate." In: *AGU Advances* 1, e2019AV000129. DOI: 10.1029/2019AV000129.

Eyring, V. et al. (2016). "Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization." In: *Geosci. Model Dev.* 9, pp. 1937–1958. DOI: 10.5194/gmd-9-1937-2016.

Feldl, N., S. Po-Chedley, H. K. A. Singh, S. Hay, and P. J. Kushner (2020). "Sea ice and atmospheric circulation shape the high-latitude lapse rate feedback." In: *npj Climate and Atmospheric Science* 3, 41. DOI: 10.1038/s41612-020-00146-7.

Field, C., M. McNutt, K. Marvel, G. A. Schmidt, and P. H. Jacobs (2021). "Evaluating Climate Scenarios." In: *Issues Sci. Technol.* 38, pp. 5–6, 8–9.

Forster, P. M. (2016). "Inference of climate sensitivity from analysis of Earth's energy budget." In: *Ann. Rev. Earth Planet. Sci.* 44, pp. 85–106. DOI: 10.1146/annurev-earth-060614-105156.

Geoffroy, O. et al. (2013). "Transient climate response in a two-layer energy-balance model. Part I: Analytical solution and parameter calibration using CMIP5 AOGCM experiments." In: *J. Climate* 26, pp. 1841–1857. DOI: 10.1175/JCLI-D-12-00195.1.

Geoffroy, O. and D. Saint-Martin (2020). "Equilibrium- and transient-state dependencies of climate sensitivity: Are they important for climate projections?" In: *J. Climate* 33, pp. 1863–1879. DOI: 10.1175/JCLI-D-19-0248.1.

Gettleman, A. et al. (2019). "High climate sensitivity in the Community Earth System Model Version 2 (CESM2)." In: *Geophys. Res. Lett.* 46, pp. 8329–8337. DOI: 10.1029/2019GL083978.

Gettelman, A. et al. (2023). "Inportance of ice nucleation and precipitation on climate with the Parameterization of Unified Microphysics Across Scales version 1 (PUMASv1)." In: *Geoscientific Model Development* 16, pp. 1735–1754. DOI: 10.5194/gmd-16-1735-2023.

Gettleman, A. and S. C. Sherwood (2016). "Processes responsible for cloud feedback." In: *Curr. Clim. Change Rep.* 2, pp. 179–189. DOI: 10.1007/s40641-016-0052-8.

Golaz, J.-C. et al. (2019). "The DOE E3SM Coupled Model Version 1: Overview and evaluation at standard resolution." In: *Journal of Advances in Modeling Earth Systems* 11, pp. 2089–2129. DOI: 10.1029/2018MS001603.

Golaz, J.-C. et al. (2022). "The DOE E3SM Model Version 2: Overview of the Physical Model and Initial Model Evaluation." In: *Journal of Advances in Modeling Earth Systems* 14, e2022MS003156. DOI: 10.1029/2022MS003156.

Graversen, R. G., P. L. Langen, and T. Mauritsen (2014). "Polar Amplification in CCSM4: Contributions from the Lapse Rate and Surface Albedo Feedbacks." In: *J. Climate* 27, pp. 4433–4450. DOI: 10.1175/JCLI-D-13-00551.1.

Gregory, J. M. and T. Andrews (2016). "Variation in climate sensitivity and feedback parameters during the historical period." In: *Geophys. Res. Lett.* 43, pp. 3911–3920. DOI: 10.1002/2016GL068406.

Gregory, J. M., R. J. Stouffer, S. C. B. Raper, P. A. Stott, and N. A. Rayner (2002). "An observationally based estimate of the climate sensitivity." In: *J. Climate* 15, pp. 3117–3121. DOI: 10.1175/1520-0442(2002)015<3117:AOBEOT>2.0.CO;2.

Gregory, J. M. et al. (2004). "A new method for diagnosing radiative forcing and climate sensitivity." In: *Geophys. Res. Lett.* 31, L03205. DOI: 10.1029/2003GL018747.

Gregory, J. M. et al. (2005). "A model intercomparison of changes in the Atlantic thermohaline circulation in response to increasing atmospheric CO_2 concentration." In: *Geophys. Res. Lett.* 32, L12703. DOI: 10.1029/2005GL023209.

Grose, M. R., J. Gregory, R. Colman, and T. Andrews (2018). "What climate sensitivity index is most useful for projections?" In: *Geophys. Res. Lett.* 45, pp. 1559–1566. DOI: 10.1002/2017GL075742.

Hahn, L. C., K. C. Armour, M. D. Zelinka, C. M. Bitz, and A. Donohoe (2021). "Contributions to Polar Amplification in CMIP5 and CMIP6 Models." In: *Front. Earth Sci.* 9, 710036. DOI: 10.3389/feart.2021.710036.

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

Hall, A. and S. Manabe (1999). "The role of water vapor feedback in unperturbed climate variability and global warming." In: *J. Climate* 12, pp. 2327–2346. DOI: 10.1175/1520-0442(1999)012<2327:TROWVF>2.0.CO;2.

Hansen, J. et al. (1984). *Climate sensitivity: Analysis of feedback mechanisms*. Climate Processes and Climate Sensitivity, Geophys. Monogr., Vol. 29. Amer. Geophys. Union, pp. 130–162. DOI: 10.1029/GM029p0130.

Hansen, J. et al. (2005). "Efficacy of climate forcings." In: *J. Geophys. Res.* 110, D18104. DOI: 10.1029/2005JD005776.

Happer, W. and R. Lindzen (2023). *Re: Proposed Fossil Fuel Power Plant Rule: "New Source Performance Standards for Greenhouse Gas Emissions From New, Modified, and Reconstructed Fossil Fuel-Fired Electric Generating Units; Emission Guidelines for Greenhouse Gas Emissions From Existing Fossil Fuel-Fired Electric Generating Units; and Repeal of the Affordable Clean Energy Rule" (the "Proposed Rule")*. URL: https://co2coalition.org/wp-content/uploads/2023/07/Happer-Lindzen-EPA-Power-Plants-2023-07-19.pdf.

Hartmann, D. L. and K. Larson (2002). "An important constraint on tropical cloud - climate feedback." In: *Geophys. Res. Lett.* 29, 1971. DOI: 10.1029/2002GL015835.

Hausfather, Z., K. Marvel, G. A. Schmidt, J. W. Nielsen-Gammon, and M. Zelinka (2022). "Climate simulations: recognize the 'hot model' problem." In: *Nature* 605, pp. 26–29. DOI: 10.1038/d41586-022-01192-2.

Hausfather, Z. and G. P. Peters (2020). "RCP8.5 is a problematic scenario for near-term emissions." In: *Proc. Natl. Acad. Sci. (USA)* 117, pp. 27791–27792. DOI: 10.1073/pnas.2017124117.

He, J., M. Winton, G. Vecchi, L. Jia, and M. Rugenstein (2017). "Transient Climate Sensitivity Depends on Base Climate Ocean Circulation." In: *J. Climate* 30, pp. 1493–1504. DOI: 10.1175/JCLI-D-16-0581.1.

Head, B. W. (2019). "Forty years of wicked problems literature: forging closer links to policy studies." In: *Policy and Society* 32. DOI: 10.1080/14494035.2018.1488797.

Held, I. M. (2005). "The gap between simulation and understanding in climate modeling." In: *Bull. Amer. Meteor. Soc.* 86. DOI: 10.1175/BAMS-86-11-1609.

– (2014). "Simplicity and complexity." In: *Science* 343, pp. 1206–1207. DOI: 10.1126/science.1248447.

Held, I. M. and K. M. Shell (2012). "Using relative humidity as a state variable in climate feedback analysis." In: *J. Climate* 25, pp. 2578–2582. DOI: 10.1175/JCLI-D-11-00721.1.

Held, I. M. and B. J. Soden (2000). "Water vapor feedback and global warming." In: *Annu. Rev. Energy Environ.* 25, pp. 441–475. DOI: 10.1146/annurev.energy.25.1.441.

Hohenegger, C. et al. (2023). "ICON-Sapphire: simulating the components of the Earth system and their interactions at kilometer and subkilometer scales." In: *Geoscientific Model Development* 16, pp. 779–811. DOI: 10.5194/gmd-16-779-2023.

Hoppe, H.-H. (1983). *Kritik der kausalwissenschaftlichen Sozialforschung. Untersuchungen zur Grundlegung von Soziologie und Ökonomie*. Studien zur Sozialwissenschaft. Opladen: Westdeutscher Verlag. 108 pp.

Hourdin, F. et al. (2017). "The art and science of climate model tuning." In: *Bull. Amer. Meteor. Soc.* 98. DOI: 10.1175/BAMS-D-15-00135.1.

Huang, Y., Y. Xia, and X. Tan (2017). "On the pattern of CO_2 radiative forcing and poleward energy transport." In: *J. Geophys. Res. Atmos.* 122, pp. 10578–10593. DOI: 10.1002/2017JD027221.

Hulme, M. (2011). "Reducing the future to climate: A story of climate determinism and reductionism." In: *Osiris* 26, pp. 245–266. DOI: 10.1086/661274.

IPCC (1990). *Climate Change: The IPCC Scientific Assessment: Report prepared for IPCC by Working Group I*. Ed. by J. T. Houghton, G. J. Jenkins, and J. J. Ephraums. https://www.ipcc.ch/site/assets/uploads/2018/03/ipcc_far_wg_I_full_report.pdf. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. 414 pp.

– (1995). *Climate Change 1995: The Science of Climate Change. Contribution of WGI to the Second Assessment Report of the Intergovernmental Panel on Climate Change*. Ed. by J. T. Houghton et al. https://www.ipcc.ch/site/assets/uploads/2018/02/ipcc_sar_wg_I_full_report.pdf. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. 588 pp.

– (2001). *Climate Change 1995: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*. Ed. by J. T. Houghton et al. https://www.ipcc.ch/site/assets/uploads/2018/03/WGI_TAR_full_report.pdf. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. 893 pp.

– (2007). *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Ed. by S. Solomon et al. https://www.ipcc.ch/site/assets/uploads/2018/05/ar4_wg1_full_report-1.pdf. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. 996 pp.

– (2013). *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Ed. by T. F. Stocker et al. https://www.ipcc.ch/

site/assets/uploads/2018/02/WG1AR5_all_final.pdf. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. 1535 pp.

– (2018). *Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty*. Vol. In Press. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. DOI: https://doi.org/10.1017/9781009157940.

– (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, United Kingdom and New York, NY, USA: Cambridge University Press. DOI: 10.1017/9781009157896.

Irving, D., W. Hobbs, J. Church, and J. Zika (2021). "A mass and energy conservation analysis of drift in the CMIP6 ensemble." In: *J. Climate* 34, pp. 3157–3170. DOI: 10.1175/JCLI-D-20-0281.1.

Jackson, L. C. et al. (2015). "Global and European climate impacts of a slowdown of the AMOC in a high resolution GCM." In: *Climate Dyn.* 45, pp. 3299–3316. DOI: 10.1007/s00382-015-2540-2.

Jenkins, M. T. and A. Dai (2021). "The impact of sea-ice loss on Arctic climate feedbacks and their role for Arctic amplification." In: *Geophys. Res. Lett.* 48, e2021GL094599. DOI: 10.1029/2021GL094599.

Kajtar, J. B. et al. (2021). "CMIP5 intermodel relationships in the baseline Southern Ocean climate system and with future projections." In: *Earth's Future* 9, e2020EF001873. DOI: 10.1029/2020EF001873.

Kang, S. M. et al. (2023). "Subtropical clouds key to Southern Ocean teleconnections to the tropical Pacific." In: *Proc. Natl. Acad. Sci. (USA)* 120, e2300881120. DOI: 10.1073/pnas.2300881120.

Katzav, J. and W. S. Parker (2015). "The future of climate modeling." In: *Climatic Change* 132, pp. 475–487. DOI: 10.1007/s10584-015-1435-x.

Kilbourne, K. H. et al. (2022). "Atlantic circulation change still uncertain." In: *Nat. Geosci.* 15, pp. 165–167. DOI: 10.1038/s41561-022-00896-4.

Kim, H., S. M. Kang, J. E. Kay, and S.-P. Xie (2022). "Subtropical clouds key to Southern Ocean teleconnections to the tropical Pacific." In: *Proc. Natl. Acad. Sci. (USA)* 119, e2200514119. DOI: 10.1073/pnas.2200514119.

Klein, S. A., A. Hall, J. R. Norris, and R. Pinus (2017). "Low-cloud feedbacks from cloud-controlling factors: A review." In: *Surv. Geophys.* 38, pp. 1307–1329. DOI: 10.1007/s10712-017-9433-3.

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

Knutti, R., J. Rogelj, J. Sedláček, and E. M. Fischer (2016). "A scientific critique of the two-degree climate change target." In: *Nat. Geosci.* 9, pp. 13–19. DOI: 10.1038/NGEO2595.

Knutti, R., M. A. A. Rugenstein, and G. C. Hegerl (2017). "Beyond equilibrium climate sensitivity." In: *Nat. Geosci.* 10, pp. 727–744. DOI: 10.1038/NGEO3017.

Lee, S., M. L'Heureux, A. T. Wittenberg, R. Seager, and P. A. O'Gorman abd N C. Johnson (2022). "On the future zonal contrasts of equatorial Pacific climate: Perspectives from observations, simulations, and theories." In: *npj Climate and Atmospheric Science* 5, 82. DOI: 10.1038/s41612-022-00301-2.

Lewis, N. (2023). "Objectively combining climate sensitivity evidence." In: *Climate Dyn.* 60, pp. 3139–3165. DOI: 10.1007/s00382-022-06468-x.

Lewis, N. and J. A. Curry (2015). "The implications for climate sensitivity of AR5 forcing and heat uptake estimates." In: *Climate Dyn.* 45, pp. 1009–1023. DOI: 10.1007/s00382-014-2342-y.

– (2018). "The impact of recent ocean heat uptake data on estimates of climate sensitivity." In: *J. Climate* 31, pp. 6051–6071. DOI: 10.1175/JCLI-D-17-0667.1.

Lewis, N. and T. Mauritsen (2021). "Negligible unforced historical pattern effect on climate feedback strength found in HadISST-based AMIP simulations." In: *J. Climate* 34, pp. 39–55. DOI: 10.1175/JCLI-D-19-0941.1.

Li, C., J.-S. von Storch, and J. Marotzke (2013). "Deep-ocean heat uptake and equilibrium climate sensitivity." In: *Climate Dyn.* 40, pp. 1071–1086. DOI: 10.1007/s00382-012-1350-z.

Lian, M. S. and R. D. Cess (1977). "Energy balance climate models: A a reappraisal of ice-albedo feedback." In: *J. Atmos. Sci.* 34, pp. 1058–1062. DOI: 10.1175/1520-0469(1977)034<1058:EBCMAR>2.0.CO;2.

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." In: *Geophys. Res. Lett.* 46, pp. 12331–12339. DOI: 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." In: *Geophys. Res. Lett.* 48, e2021GL093302. DOI: 10.1029/2021GL093302.

Lindzen, R. and Y.-S. Choi (2021). "The iris effect: A review." In: *Asia-Pac. J. Atmos. Sci.* 58, pp. 159–168. DOI: 10.1007/s13143-021-00238-1.

Lindzen, R., M.-D. Chou, and A. Y. Hou (2001). "Does the Earth have an adaptive infrared iris?" In: *Bull. Amer. Meteor. Soc.* 82. DOI: 10.1175/1520-0477(2001)082<0417:DTEHAA>2.3.CO;2.

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

Liu, F. et al. (2020). "Sensitivity of Surface Temperature to Oceanic Forcing via q-Flux Green's Function Experiments. Part III: Asymmetric response to warming and cooling." In: *J. Climate* 33, pp. 1283–1297. DOI: 10.1175/JCLI-D-19-0131.1.

Loehle, C. (2014). "A minimal model for estimating climate sensitivity." In: *Ecological Modelling* 276, pp. 80–84. DOI: 10.1016/j.ecolmodel.2014.01.006.

– (2015). "Global temperature trend adjusted for unforced variability." In: *Universal Journal of Geoscience* 3, pp. 183–187. DOI: 10.13189/ujg.2015.030601.

– (2018). "The epistemological status of general circulation models." In: *Climate Dyn.* 50, pp. 1719–1731. DOI: 10.1007/s00382-017-3717-7.

Lüning, S. and F. Vahrenholt (2017). "Paleoclimatological context and reference level of the 2 °C and 1.5 °C Paris Agreement long-term temperature limits." In: *Fontiers in Earth Science* 5, 104. DOI: 10.3389/feart.2017.00104.

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

– (1975). "The effects of doubling of the CO_2 concentration on the climate of a general circulation model." In: *J. Atmos. Sci.* 32, pp. 3–15. DOI: 10.1175/1520-0469(1975)032<0003:TEODTC>2.0.CO;2.

Mastrandrea, M. D. et al. (2010). *Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties*. IPCC. URL: https://www.ipcc.ch/publication/ipcc-cross-working-group-meeting-on-consistent-treatment-of-uncertainties/.

Matthes, K. et al. (2017). "Solar forcing for CMIP6 (v3.2)." In: *Geoscientific Model Development* 10, pp. 2247–2302. DOI: 10.5194/gmd-10-2247-2017.

Mauritsen, T. (2016). "Clouds cooled the Earth." In: *Nat. Geosci.* 9, pp. 865–867. DOI: 10.1038/ngeo2838.

Mauritsen, T. et al. (2012). "The tuning of a global model." In: *Journal of Advances in Modeling Earth Systems* 4, M00A01. DOI: 10.1029/2012MS000154.

Mauritsen, T. et al. (2022). "Early development and tuning of a global coupled cloud resolving model, and its fast response to increasing CO_2." In: *Tellus A: Dynamic Meteorology and Oceanography* 74, pp. 346–363. DOI: 10.16993/tellusa.54.

Mauritsen, T. and E. Roecker (2020). "Tuning the MPI-ESM1.2 global climate model to improve the match with instrumental record warming by lowering its climate sensitivity." In: *Journal of Advances in Modeling Earth Systems* 12, e2019MS002037. DOI: 10.1029/2019MS002037.

Mauritsen, T. and B. Stevens (2015). "Missing iris effect as apossible cause of muted hydrological change and high climate sensitivity in models." In: *Nat. Geosci.* 8, pp. 346–351. DOI: 10.1038/NGEO2414.

McGraw, Z. et al. (Feb. 2023). "On the links between ice nucleation, cloud phase, and climate sensitivity in CESM2." In: *Authorea*. DOI: 10.22541/essoar.167214452.25853014/v2.

McKitrick, R. and K. Christy (2018). "A test of the tropical 200- to 300-hPa warming rate in climate models." In: *Earth and Space Science* 5, pp. 529–536. DOI: 10.1029/2018EA000401.

McKitrick, R. and K. Christy (2020). "Pervasive warming bias in CMIP6 tropospheric layers." In: *Earth and Space Science* 7, e2020EA001281. DOI: 10.1029/2020EA001281.

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." In: *Geophys. Res. Lett.* 48, e2020GL090861. DOI: 10.1029/2020GL090861.

Mitevski, I., L. M. Polvani, and C. Orbe (2022). "Asymmetric warming/cooling response to CO_2 increase/decrease mainly due to non-logarithmic forcing, not feedbacks." In: *Geophys. Res. Lett.* 49, e2021GL097133. DOI: 10.1029/2021GL097133.

Modak, A. and T. Mauritsen (2021). "The 2000-2012 global warming hiatus more likely with a low climate sensitivity." In: *Geophys. Res. Lett.* 48, e2020GL091779. DOI: 10.1029/2020GL091779.

– (2023). "Better-constrained climate sensitivity when accounting for dataset dependency on pattern effect estimates." In: *Atmos. Chem. and Phys.* 23, pp. 7535–7549. DOI: 10.5194/acp-23-7535-2023.

Möller, F. (1963). "On the influence of changes in the CO2 concentration in air on the radiation balance of the Earth's surface and on the climate." In: *J. Geophys. Res.* 68, pp. 3877–3886. DOI: 10.1029/JZ068i013p03877.

Mülmenstädt, J. et al. (2021). "An underestimated negative cloud feedback from cloud lifetime changes." In: *Nat. Clim. Change* 11, pp. 508–513. DOI: 10.1038/s41558-021-01038-1.

Murphy, J. M. (1995). "Transient response of the Hadley Centre coupled ocean-atmosphere model to increasing carbon dioxide. Part III: Analysis of global-mean response using simple models." In: *J. Climate* 8, pp. 496–514. DOI: 10.1175/1520-0442(1995)008<0496:TROTHC>2.0.CO;2.

Myers, T. A. et al. (2021). "Observational constraints on low cloud feedback reduce uncertainty of climate sensitivity." In: *Nat. Clim. Change* 11, pp. 501–507. DOI: 10.1038/s41558-021-01039-0.

Myers, T. A., M. D. Zelinka, and S. A. Klein (2023). "Observational constraints on the cloud feedback pattern effect." In: *J. Climate*, Accepted Manuscript. DOI: 10.1175/JCLI-D-22-0862.1.

National Academies of Sciences, Engineering, and Medicine (2016). *Assessment of Approaches to Updating the Social Cost of Carbon: Phase 1 Report on a Near-Term Update. Committee on Assessing Approaches to Updating the Social Cost of Carbon, Board on Environmental Change and Society*. Washington, DC: The National Academies Press. DOI: 10.17226/21898.

Nijsse, F. J. M. M., P. M. Cox, and M. S. Williamson (2020). "Emergent constraints on transient climate response (TCR) and equilibrium climate sensitivity (ECS) from historical warming in CMIP5 and CMIP6 models." In: *Earth Syst. Dynam.* 11, pp. 737–750. DOI: 10.5194/esd-11-737-2020.

Oppenheimer, M., B. C. O'Neikk, M. Webster, and S. Agrawala (2007). "The limits of consensus." In: *Science* 317, pp. 1505–1506. DOI: 10.1126/science.1144831.

Otto, A. et al. (2013). "Energy budget constraints on climate response." In: *Nat. Geosci.* 6, pp. 415–416. DOI: 10.1038/ngeo1836.

Palmer, T. and B. Stevens (2019). "The scientific challenge of understanding and estimating climate change." In: *Proc. Natl. Acad. Sci. (USA)* 116, pp. 24390–24395. DOI: 10.1073/pnas.1906691116.

Parker, W. S. (2009). "Confirmation and adequacy-for-purpose in climate modelling." In: *Proceedings of the Aristotelian Society Supplementary* 83, pp. 233–249. DOI: 10.1111/j.1467-8349.2009.00180.x.

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." In: *J. Geophys. Res. Atmos.* 123, pp. 1921–1941. DOI: 10.1002/2017JD027885.

Pendergrass, A. M., A. Conley, and F. M. Vitt (2018). "Surface and top-of-atmosphere radiative feedback kernels for CESM-CAM5." In: *Earth Syst. Sci. Data* 10, pp. 317–324. DOI: 10.5194/essd-10-317-2018.

Pielke, R. Jr., M. G. Burgess, and J. Ritchie (2022). "Plausible 2005-2100 emissions scenarious project between 2 °C and 3 °C of warming by 2100." In: *Environ. Res. Lett.* 17, 024027. DOI: 10.1088/1748-9326/ac4ebf.

Pielke, R. Jr. and J. Ritchie (2021a). "Distorting the view of our climate future: The misuse and abuse of climate pathways and scenarios." In: *Energy Research & Social Science* 72, 101890. DOI: 10.1016/j.erss.2020.101890.

– (2021b). "How climate scenarios lost touch with reality. A failure of self-correction in science has compomised climate science's ability to provide plausible views of our collective future." In: *Issues Sci. Technol.* 37, pp. 74–84.

Pithan, F. and T. Mauritsen (2014). "Arctic amplification dominated by temperature feedbacks in contemporary climate models." In: *Nat. Geosci.* 7, pp. 181–184. DOI: 10.1038/ngeo2071.

Po-Chedley, S. et al. (2018). "Sources of intermodel spread in the lapse rate and water vapor feedbacks." In: *J. Climate* 31, pp. 3187–3206. DOI: 10.1175/JCLI-D-17-0674.1.

Popper, Karl (2002 [1959]). *The Logic of Scientific Discovery*. 11 New Fetter Lane, London: Routlege. 544 pp.

Prather, M. J. and J. C. Hsu (2019). "A round Earth for climate models." In: *Proc. Natl. Acad. Sci. (USA)* 116, pp. 19330–19335. DOI: 10.1073/pnas.1908198116.

Radtke, J., T. Mauritsen, and C. Hohenegger (2021). "Shallow cumulus cloud feedback in large eddy simulations - bridging the gap to storm resolving models." In: *Atmos. Chem. and Phys.* 21, pp. 3275–3288. DOI: 10.5194/acp-21-3275-2021.

Ramanathan, V., M. S. Lian, and R. D. Cess (1979). "Increased atmospheric CO_2: Zonal and seasonal estimates of the effect on the radiation energy balance and surface temperature." In: *J. Geophys. Res.* 84, pp. 4949–4958. DOI: 10.1029/JC084iC08p04949.

Riahi, K. et al. (2017). "The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview." In:

Global Environmental Change 42, pp. 153–168. DOI: 10.1016/j.gloenvcha.2016.05.009.

Ritchie, J. and H. Dowlatabadi (2017). "Why do climate change scenarios return to coal?" In: *Energy* 140, pp. 1276–1291. DOI: 10.1016/j.energy.2017.08.083.

Rittel, H. W. J. and M. M. Webber (1973). "Dilemmas in a General Theory of Planning." In: *Policy Sciences* 4. DOI: 10.1007/BF01405730.

Roberts, M. J. et al. (2020). "Sensitivity of the Atlantic Meridional Overturning Circulation to Model Resolution in CMIP6 HighResMIP Simulations and Implications for Future Changes." In: *Journal of Advances in Modeling Earth Systems* 12, e2019MS002014. DOI: 10.1029/2019MS002014.

Robson, J. et al. (2022). "The role of anthropogenic aerosol forcing in the 1850-1985 strengthening of the AMOC in CMIP6 historical simulations." In: *J. Climate* 35, pp. 843–863. DOI: 10.1175/JCLI-D-22-0124.1.

Roe, G. H. (2009). "Feedbacks, timescales, and seeing red." In: *Ann. Rev. Earth Planet. Sci.* 37, pp. 93–115. DOI: 10.1146/annurev.earth.061008.134734.

Roe, G. H. and Y. Bauman (2013). "Climate sensitivity: should the climate tail wag the policy dog?" In: *Climate Change* 117, pp. 647–662. DOI: 10.1007/s10584-012-0582-6.

Roemmich, D. et al. (2019). "On the future of Argo: A global, full-depth, multi-disciplinary array." In: *Frontiers in Marine Science* 6, 439. DOI: 10.3389/fmars.2019.00439.

Romps, D. M., J. T . Seeley, and J. P. Edman (2022). "Why the forcing fom carbon dioxide scales as the logarithm of its concentration." In: *J. Climate* 35, pp. 4027–4047. DOI: 10.1175/JCLI-D-21-0275.1.

Rugenstein, M. and K. Armour (2021). "Three flavors of radiative feedbacks and their implications for estimating climate sensitivity." In: *Geophys. Res. Lett.* 48, e2021GL092983. DOI: 10.1029/2021GL092983.

Rugenstein, M., K. Caldeira, and R. Knutti (2016). "Dependence of global radiative feedbacks on evolving patterns of surface heat fluxes." In: *Geophys. Res. Lett.* 43, pp. 9877–9885. DOI: 10.1002/2016GL070907.

Rugenstein, M. et al. (2019). "LongRunMIP - motivation and design for a large collection of millennial-length GCM simulations." In: *Bull. Amer. Meteor. Soc.* 100. DOI: 10.1175/BAMS-D-19-0068.1.

Rugenstein, M. et al. (2020). "Equilibrium climate sensitivity estimated by equilibrating climate models." In: *Geophys. Res. Lett.* 47, e2019GL083898. DOI: 10.1029/2019GL083898.

Rypdal, M. et al. (2021). "Estimating remaining carbon budgets using temperature responses informed by CMIP6." In: *Front. Clim.* 3, 686058. DOI: 10.3389/fclim.2021.686058.

Sanderson, B. M. (2020). "Relating climate sensitivity to indices to projection uncertainty." In: *Earth Syst. Dynam.* 11, pp. 721–735. DOI: 10.5194/esd-11-721-2020.

Sanderson, B. M. et al. (2021). "The potential for structural errors in emergent constraints." In: *Earth Syst. Dynam.* 12, pp. 899–918. DOI: 10.5194/esd-12-899-2021.

Sanderson, B. M. and M. Rugenstein (2022). "Potential bias in effective climate sensitivity from state-dependent energetic imbalance." In: *Earth Syst. Dynam.* 13, pp. 1715–1736. DOI: 10.5194/esd-13-1715-2022.

Scafetta, N. (2022). "Advanced testing of low, medium, and high ECS CMIP6 GCM simulations versus ERA5-T2m." In: *Geophys. Res. Lett.* 48, e2022GL097716. DOI: 10.1029/2022GL097716.

Schlund, M., A. Lauer, P. Gentine, S. C. Sherwood, and V. Eyring (2020). "Emergent constraints on equilibrium climate sensitivity in CMIP5: Do they hold for CMIP6?" In: *Earth Syst. Dynam.* 11, pp. 1233–1258. DOI: 10.5194/esd-11-1233-2020.

Schneider, S. H. (1972). "Cloudiness as global climatic feedback mechanism: The effects on the radiation balance and surface temperature of variations in cloudiness." In: *J. Atmos. Sci.* 29, pp. 1413–1422. DOI: 10.1175/1520-0469(1972)029<1413:CAAGCF>2.0.CO;2.

Schneider, T. (2018). *Statistical Inference with Emergent Constraints*. URL: https://climate-dynamics.org/statistical-inference-with-emergent-constraints/.

Schneider, S. H. and R. E. Dickinson (1974). "Climate modeling." In: *Reviews in Geophysics and Space Physics* 12, pp. 447–493. DOI: 10.1029/RG012i003p00447.

Schneider, S. H. and W. W. Kellogg (1973). "The chemical basis for climate change." In: *Chemistry of the Lower Atmosphere*. Ed. by S. I. Rasool. New York: Plenum, pp. 203–249.

Schwalm, C. R., S. Glendon, and P. B. Duffy (2020). "Reply to Husfather and Peters: RCP8.5 is neither problematic nor misleading." In: *Proc. Natl. Acad. Sci. (USA)* 117, pp. 27793–27794. DOI: 10.1073/pnas.2018008117.

Seager, R. et al. (2019). "Strengthening tropical Pacific zonal sea surface temperature gradient consistent with rising greenhouse gases." In: *Nat. Clim. Change* 9, pp. 517–525. DOI: doi.org/10.1038/s41558-019-0505-x.

Seeley, J. T., N. Jeevanjee, and D. M. Romps (2019). "FAT or FiTT: Are anvil clouds or the tropopause temperature invariant." In: *Geophys. Res. Lett.* 46, pp. 1842–1850. DOI: 10.1029/2018GL080096.

Sellers, W. D. (1969). "A global climatic model based on the energy balance of the Earth-atmosphere system." In: *J. Appl. Meteor.* 8, pp. 392–400. DOI: 10.1175/1520-0450(1969)008<0392:AGCMBO>2.0.CO;2.

Senior, C. A. and J. F. B. Mitchell (2000). "The time-dependence of climate sensitivity." In: *Geophys. Res. Lett.* 27, pp. 2685–2688. DOI: 10.1029/2000GL011373.

Shaw, J., Z. McGraw, O. Bruno, T. Storelvmo, and S. Hofer (2022). "Using satellite observations to evaluate model microphysical representation of Arctic mixed-phase clouds." In: *Geophys. Res. Lett.* 49, e2021GL096191. DOI: 10.1029/2021GL096191.

Shell, K. M. (2012). "Consistent differences in climate feedbacks between atmosphere–ocean GCMs and atmospheric GCMs with slab-ocean models." In: *J. Climate* 26, pp. 4264–4281. DOI: 10.1175/JCLI-D-12-00519.1.

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." In: *J. Climate* 21, pp. 2269–2282. DOI: 10.1175/2007JCLI2044.1.

Shepherd, T. G. et al. (2018). "Storylines: an alternative approach to representing uncertainty in physical aspects of climate change." In: *Climatic Change* 151, pp. 555–571. DOI: 10.1007/s10584-018-2317-9.

Shepherd, C., E. Rauhala, and C. Mooney (July 2023). "As the world sizzels, China says it woll deal with climate its own way." In: *The Washington Post*. Visited 04.09.2023. URL: https://www.washingtonpost.com/climate-environment/2023/07/19/climate-change-heat-wave-china/.

Shepherd, T. G. and E. A. Lloyd (2021). "Meaningful climate science." In: *Climatic Change* 169, 17. DOI: 10.1007/s10584-021-03246-2.

Sherwood, S. C. et al. (2015). "Adjustments in the forcing-feedback framework for understanding climate change." In: *Bull. Amer. Meteor. Soc.* 96. DOI: 10.1175/BAMS-D-13-00167.1.

Sherwood, S. C. et al. (2020). "An assessment of Earth's climate sensitivity using multiple lines of evidence." In: *Rev. Geophys.* 58, e2019RG000678. DOI: 10.1029/2019RG000678.

Sherwood, S. C. and C. E. Forest (2023). "Opinion: Can uncertainty in climate sensitivity be narrowed further?" In: *EGUsphere* 2023, pp. 1–12. DOI: 10.5194/egusphere-2023-1647.

Shukla, J. et al. (2010). "Toward a new generation of worl climate research and computing facilities." In: *Bull. Amer. Meteor. Soc.* 91. DOI: 10.1175/2010BAMS2900.1.

Simpson, M. et al. (2016). "Decision analysis for management of natural hazards." In: *Ann. Rev. Environ. Resour.* 41, pp. 489–516. DOI: 10.1146/annurev-environ-110615-090011.

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

Smith, C. J. et al. (2018). "Understanding rapid adjustments to diverse forcing agents." In: *Geophys. Res. Lett.* 45, pp. 12023–12031. DOI: 10.1029/2018GL079826.

Smith, C. J. et al. (2020). "Effective radiative forcing and adjustments in CMIP6 models." In: *Atmos. Chem. and Phys.* 20, pp. 9591–9618. DOI: 10.5194/acp-20-9591-2020.

Soden, B. J. et al. (2008). "Quantifying climate feedbacks using radiative kernels." In: *J. Climate* 21, pp. 3504–3520. DOI: 10.1175/2007JCLI2110.1.

Sowell, T. (2007 [1987]). *A Conflict of Visions. Ideological Origins of Political Struggles*. New York, NY, USA: Basic Books. 330 pp.

Stefani, F. (2021). "Solar and anthropogenic influences on climate: Regression analysis and tentative predictions." In: *Climate* 9, 163. DOI: 10.3390/cli9110163.

Stephens, G. L. (1978). "Radiation profiles in extended water clouds. II: Parameterization schemes." In: *J. Atmos. Sci.* 35, pp. 2123–2132. DOI: 10.1175/1520-0469(1978)035<2123:RPIEWC>2.0.CO;2.

Stevens, B. (2017). "Clouds unfazed by haze." In: *Nature* 546, pp. 483–484. DOI: 10.1038/546483a.

Stevens, B. and S. Bony (2013). "What are climate models missing?" In: *Science* 30, pp. 1053–1054. DOI: 10.1126/science.1237554.

Stevens, B., S. C. Sherwood, S. Bony, and M. J. Webb (2016). "Prospects for narrowing bounds on Earth's climate sensitivity." In: *Earth's Future* 4, pp. 512–522. DOI: 10.1002/2016EF000376.

Stevens, B. and L. Kluft (2023). "A colorful look at climate sensitivity." In: *EGUsphere* 2023, pp. 1–24. DOI: 10.5194/egusphere-2022-1460.

Stuecker, M. F. et al. (2018). "Polar amplification dominated by local forcing and feedbacks." In: *Nat. Clim. Change* 8, pp. 1076–1081. DOI: 10.1038/s41558-018-0339-y.

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

Thompson, E. L. and L. A. Smith (2019). "Escape from model-land." In: *Economic: The Open-Access, Open-Assessment E-Journal* 13, (2019-40), pp. 1–15. DOI: 10.5018/economics-ejournal.ja.2019-40.

Tierney, J. E. et al. (2020). "Glacial cooling and climate sensitivity revisited." In: *Nature* 584, pp. 569–585. DOI: 10.1038/s41586-020-2617-x.

Todd, A. et al. (2020). "Ocean-only FAFMIP: Understanding regional patterns of ocean heat content and dynamic sea level change." In: *Journal of Advances in Modeling Earth Systems* 12, e2019MS002027. DOI: 10.1029/2019MS002027.

Tokarska, K. B. et al. (2020). "Past warming trend constrains future warming in CMIP6 models." In: *Sci. Adv* 6, eaaz9549. DOI: 10.1126/sciadv.aaz9549.

Tung, K.K. and J. Zhou (2013). "Using data to attribute episodes of warming and cooling in intrumental records." In: *Proc. Natl. Acad. Sci. (USA)* 110, pp. 2058–2063. DOI: 10.1073/pnas.1212471110.

United Nations (1998). *Kyoto Protocol to the United Nations framework convention on climate change*. URL: https://unfccc.int/resource/docs/convkp/kpeng.pdf.

– (2015). *Paris Agreement*. URL: https://unfccc.int/sites/default/files/english_paris_agreement.pdf.

van der Slujis, J. P., R. van Est, and M. Riphagen (2010). "Beyond consensus: Reflections from a democratic perspective on the interaction between climate policies and science." In: *Current Opinion and Environmental Sustainability* 2, pp. 409–415. DOI: 10.1016/j.cosust.2010.10.003.

Volodin, E. et al. (2018). "Simulation of the modern climate using the INM-CM48 climate model." In: *Russian Journal of Numerical Analysis and Mathematical Modelling* 33, pp. 367–374. DOI: 10.1515/rnam-2018-0032.

Voosen, P. (2019). "New climate model forecast warming surge." In: *Science* 364, p. 6437. DOI: 10.1126/science.364.6437.222.

– (2021). "Climate panel confronts implausibly hot models." In: *Science* 373, p. 6554. DOI: 10.1126/science.373.6554.474.

Vuuren, D. P. van et al. (2011). "The Representative Concentration Pathways: An overview." In: *Climatic Change* 109, pp. 5–31. DOI: 10.1007/s10584-011-0148-z.

Wallace, J. M. and P. V. Hobbs (2006). *Atmospheric Science: An Introductory Survey*. San Diego, CA, USA: Academic Press. 483 pp.

Warren, W. S. (2022). "Earning the public's trust." In: *Sci. Adv* 8, eabo6347. DOI: 10.1126/sciadv.abo6347.

Watanabe, M., J.-L. Dufresne, Y. Kosaka, T. Mauritsen, and H. Tatebe (2021). "Enhanced warming constrained by past trends in equatorial Pacific sea surface temperature gradient." In: *Nat. Clim. Change* 11, pp. 33–39. DOI: 10.1038/s41558-020-00933-3.

Watterson, I. G. (2000). "Interpretation of simulated global warming using a simple model." In: *J. Climate* 13, pp. 202–215. DOI: 10.1175/1520-0442(2000)013<0202:IOSGWU>2.0.CO;2.

Wetherald, R. T. and S. Manabe (1988). "Cloud feedback processes in a general circulation model." In: *J. Atmos. Sci.* 45, pp. 1397–1415. DOI: 10.1175/1520-0469(1988)045<1397:CFPIAG>2.0.CO;2.

Williams, K. D., W. J. Ingram, and J. M. Gregory (2008). "Time variation of effective climate sensitivity in GCMs." In: *J. Climate* 21, pp. 5076–5089. DOI: 10.1175/2008JCLI2371.1.

Wills, R. C. J., Y. Dong, C. Proistosecu, K. C. Armour, and D. S. Battisti (2022). "Systematic climate model biases in the large-scale patterns of recent sea-surface temperature and sea-level pressure change." In: *Geophys. Res. Lett.* 49, e2022GL100011. DOI: 10.1029/2022GL100011.

Wilson, D. J. and J. Gea-Banacolche (2012). "Simple model to estimate the contribution of atmospheric CO_2 to the Earth's greenhouse effect." In: *Amer. J. Phys.* 80, pp. 306–315. DOI: 10.1119/1.3681188.

Winton, M. (2006). "Surface albedo feedback estimates from the AR4 climate models." In: *J. Climate* 19, pp. 359–365. DOI: 10.1175/JCLI3624.1.

Winton, M., K. Takahashi, and I. M. Held (2010). "Importance of Ocean Heat Uptake Efficacy to Transient Climate Change." In: *J. Climate* 23, pp. 2333–2344. DOI: 10.1175/2009JCLI3139.1.

Winton, M. et al. (2014). "Has coarse ocean resolution biased simulations of transient climate sensitivity?" In: *Geophys. Res. Lett.* 41, pp. 8522–8529. DOI: 10.1002/2014GL061523.

World Meteorological Organization (2017). *International Cloud Atlas. Manual on the observation of clouds and other meteors*. Tech. rep. WMO-No. 407.

World Meteorological Organization. URL: https://cloudatlas.wmo.int/en/home.html.

Wood, R. and C. S. Bretherton (2006). "On the Relationship between Stratiform Low Cloud Cover and Lower-Tropospheric Stability." In: *J. Climate* 19, pp. 6425–6432. DOI: 10.1175/JCLI3988.1.

Yoshimori, M., F. H. Lambert, M. J. Webb, and T. Andrews (2020). "Fixed anvil temperature feedback: Positive, zero, or negativ?" In: *J. Climate* 33, pp. 2719–2739. DOI: 10.1175/JCLI-D-19-0108.1.

Zelinka, M. D. et al. (2020). "Causes of higher climate sensitivity in CMIP6 models." In: *Geophys. Res. Lett.* 47, e2019GL085782. DOI: 10.1029/2019GL085782.

Zelinka, M. D., S. A. Klein, Yi Qin, and T. A. Meyers (2022a). "Evaluating climate models' cloud feedbacks against expert judgement." In: *J. Geophys. Res. Atmos.* 127, e2021JD35198. DOI: 10.1029/2021JD035198.

Zelinka, M. D., I. Tan, L. Oreopoulos, and G. Tselioudis (2022b). "Detailing cloud property feedbacks with a regime-based decomposition." In: *Climate Dyn.* 60, pp. 2983–3003. DOI: 10.1007/s00382-022-06488-7.

Zhang, X., C. Deser, and L. Sun (2021). "Is There a Tropical Response to Recent Observed Southern Ocean Cooling?" In: *Geophys. Res. Lett.* 48, e2020GL091235. DOI: 10.1029/2020GL091235.

Zhou, C., M. D. Zelinka, and S. A. Klein (2016). "Impact of decadal cloud variations on the Earth's energy budget." In: *Nat. Geosci.* 9, pp. 871–874. DOI: 10.1038/NGEO2828.

– (2017). "Analyzing the dependence of global cloud feedback on the spatial pattern of sea surface temperature change with a Green's function approach." In: *J. Adv. in Model. Earth Syst.* 9, pp. 2174–2189. DOI: 10.1002/2017MS001096.

Zhou, C., M. D. Zelinka, A. E. Dessler, and M. Wang (2021). "Greater committed warming after accounting for the pattern effect." In: *Nat. Clim. Change* 9, pp. 132–136. DOI: 10.1038/s41558-020-00955-x.

Zhu, J. (2021). *Simulation data of the DECK, CMIP6 historical, and PMIP4 LGM simulations using the Paleoclimate-Calibrated CESM2. Version 1.0.* DOI: 10.5065/bdr7-wt42.

Zhu, J., C. J. Poulsen, and B. L. Otto-Bliesner (2020). "High climate sensitivity in CMIP6 model not supported by paleoclimate." In: *Nat. Clim. Change* 10, pp. 378–379. DOI: 10.1038/s41558-020-0764-6.

Zhu, J. et al. (2021). "Assessment of equilibrium climate sensitivity of the Community Earth System Model Version 2 through simulation of the Last Glacial Maximum." In: *Geophys. Res. Lett.* 14, e2020GL091220. DOI: 10.1029/2020GL091220.

Zhu, J. et al. (2022). "LGM paleoclimate constraints inform cloud parameterizations and equilibrium climate sensitivity in CESM2." In: *Journal of Advances in Modeling Earth Systems* 14, e2021MS002776. DOI: 10.1029/2021MS002776.

6

Paper I

Eiselt, K.-U., and R. G. Graversen (2022): Change in climate sensitivity and its dependence on the lapse-rate feedback in $4\times CO_2$ climate model experiments. *J. Climate*, **35**, 2919-2932, https://doi.org/10.1175/JCLI-D-21-0623.s1.

Change in Climate Sensitivity and Its Dependence on the Lapse-Rate Feedback in 4 × CO_2 Climate Model Experiments

KAI-UWE EISELT[a] AND RUNE GRAND GRAVERSEN[a,b]

[a] *Institute for Physics and Technology, University of Tromsø, Tromsø, Norway*
[b] *Norwegian Meteorological Institute, Tromsø, Norway*

(Manuscript received 13 August 2021, in final form 18 January 2022)

ABSTRACT: Robust estimates of climate sensitivity are important for decision-making on mitigation of climate change. However, climate sensitivity and its governing processes are still subject to large uncertainty. Recently it has been established that climate sensitivity changes over time in numerical climate model experiments with abrupt quadrupling of the CO_2 concentration. Here we conduct an analysis of such experiments from a range of climate models from phases 5 and 6 of the Coupled Model Intercomparison Project (CMIP). Climate feedbacks associated with clouds, lapse rate, Planck radiation, surface albedo, and water vapor and their changes over time are diagnosed based on a radiative kernel method. We find two clearly distinct model groups, one with weak and one with strong lapse-rate feedback change. The Arctic is the region showing the largest differences between these two model groups, with respect to both warming change and individual feedback changes. We retrace this change to the development over time of the Arctic sea ice, which impacts both the surface-albedo and lapse-rate feedbacks. Generally, models that warm quickly, both globally and in the Arctic, also quickly lose their Arctic sea ice and change their total global-mean climate feedback only little, and vice versa. However, it remains unclear if the Arctic changes are a cause or rather a by-product of the total global-mean feedback change. Finally, we find support for the results of previous studies finding that the relative warming in the tropical Indo-Pacific region may control the change of total climate feedback over time.

KEYWORDS: Climate change; Climate sensitivity; Feedback; Radiative fluxes; Regional effects; Climate models; Coupled models; Model comparison

1. Introduction

The response of Earth's climate system to a forcing due to an increase in atmospheric greenhouse gas concentration has been of interest for more than 100 years (Arrhenius 1896). Equilibrium climate sensitivity (ECS), which is the equilibrium global-mean temperature increase due to a doubling of the CO_2 concentration in the atmosphere, is a traditional measure for the climate system's response to such a forcing (Charney et al. 1979; Sherwood et al. 2020). It is often estimated based on numerical climate model experiments. A straightforward model-based method for deriving ECS is to start from a model state in radiative equilibrium, then implementing a forcing corresponding to a doubling of the CO_2 concentration in the atmosphere, and finally running the model until radiative equilibrium is restored. However, in complex Earth system models (ESMs), which realistically represent the climate system, more than 1000 simulation years are necessary for the model to reach a new equilibrium because the deep ocean has a large heat capacity and thus warms only slowly. Hence, large computational resources are required in order to equilibrate a perturbed ESM (e.g., Paynter et al. 2018; Rugenstein et al. 2020). Gregory et al. (2004) introduced an alternative but indirect approach (known as the Gregory method) to estimate ECS from a much smaller amount of simulation years in instantaneous CO_2 forcing experiments, hence without the need of running the model to a new equilibrium. The global-mean top-of-atmosphere (TOA) radiative imbalance is expressed in terms of global-mean surface-air temperature (SAT) change as follows:

$$N = F + \alpha \Delta T_s, \tag{1}$$

where N is the TOA radiative imbalance, F is the forcing (e.g., due to an increase in CO_2), α is the climate feedback parameter, and ΔT_s is the SAT change. The climate feedback parameter α represents the total feedback of the climate system and is usually estimated by regressing global-mean TOA imbalance on global-mean SAT change, and climate sensitivity is estimated by extrapolating this relationship to zero imbalance [i.e., where N in Eq. (1) is zero]. Sherwood et al. (2020) call the climate sensitivity thus estimated the *effective climate sensitivity* (S), to separate it from ECS derived from equilibrating an ESM. The climate system's response and hence its climate sensitivity is controlled by its total feedback, which is negative provided that the climate system is stable.

As mentioned by Gregory et al. (2004), and later confirmed for several other ESMs (e.g., Andrews et al. 2015), the estimates for S and α depend on the interval of simulation years included in the ΔT_s–N regression. In most simulations, S is smaller when based on the years directly following the forcing than when based on years later in the simulation, meaning

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JCLI-D-21-0623.s1.

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

DOI: 10.1175/JCLI-D-21-0623.1

that S increases over time. Correspondingly, the total feedback estimate becomes less negative over time.

The time dependence of climate sensitivity and feedback has been extensively researched but is still not well understood. Understanding the mechanism leading to a weakening of total feedback and hence an increase in S will help to gauge the likelihood of the possible occurrence of this shift in ongoing and future climate change. These model-based results are of great importance since they call into question the reliability of climate sensitivity estimates based on historical data, which, by necessity, cannot take into account future feedback changes. The estimates based on observations and proxy data typically fall on the lower end of expected climate sensitivity (Sherwood et al. 2020, and references therein).

Senior and Mitchell (2000) report a change over time of climate sensitivity in their CO_2-doubling experiment with an early coupled climate model, and argue that this is due to a change of cloud feedback over time that is associated with stability changes resulting from the delayed warming of the Southern Hemisphere compared to the Northern Hemisphere. Williams et al. (2008) take an alternative approach and introduce the concept of *effective forcing*, which is the greenhouse gas forcing adjusted for processes that occur on time scales that are short compared to those of climate stabilization. They argue that after the initial adjustment process, the ΔT_s–N relation is linear on centennial time scales and hence that the change over time of climate sensitivity is an artifact of not accounting for the relatively fast forcing adjustments. However, Winton et al. (2010) challenge this notion, since the adjustments occur on time scales on the order of decades where the "oceanic adjustment" is an important factor as well. They propose a time-varying *ocean heat uptake efficacy*, considering ocean heat uptake as a forcing that accounts for the change in the ΔT_s–N relationship over time. Armour et al. (2013) show that the evolution of global total feedback can be explained by a change over time in spatial weighting of time-invariant local feedbacks due to evolving patterns of surface warming. However, Ceppi and Gregory (2017) find that this cannot explain the global total feedback changes in a suite of ESMs that they analyze.

The perhaps most recent hypothesis for explaining the time dependence of climate feedback and S as presented by Ceppi and Gregory (2017) is the following: Upon CO_2 quadrupling, sea surface temperature (SST) warming patterns change over time to increasingly favor stably stratified regions where surface warming remains trapped near the surface and is not readily communicated vertically and horizontally, thus inducing weaker global cooling efficiency over time. In the following we refer to this as the stability hypothesis. The southeastern tropical Pacific (EP) is a region of descent (stably stratified) and the tropical western Pacific (WP) a region of ascent (close to neutrally stratified). Across members of phase 5 of the Coupled Model Intercomparison Project (CMIP5) a *delayed* warming in the EP and an *early* warming in the WP in response to an instantaneous forcing are observed (e.g., Andrews et al. 2015). A physical mechanism is suggested by Zhou et al. (2016) and further developed by Andrews and Webb (2018) that explains how this change in warming pattern causes a weakening of total feedback over time: The earlier warming in the WP readily affects higher levels in the atmosphere whereby energy is efficiently radiated out to space. Over time the warming shifts more to the EP and because this region is stably stratified, the heat remains trapped close to the surface, implying a less negative lapse-rate feedback and a reduction of the cooling efficiency of the Earth system. Moreover, the surface warming in the stably stratified regions decreases the stability there, reducing the low cloud cover (see also Wood and Bretherton 2006). Since low clouds mainly act to reflect incoming solar radiation back to space, a reduction in low cloud cover enhances the absorption of solar radiation, implying a positive cloud feedback. This means that as the warming shifts from WP to EP over time, the total feedback becomes less negative, due to a change in both lapse-rate and cloud feedback. Andrews and Webb (2018) confirm this mechanism for one atmosphere-only model (HadGEM2-A; Martin et al. 2011) by performing experiments with prescribed surface warmings in the dedicated regions. Ceppi and Gregory (2017) show that changes over time in climate feedbacks (especially cloud and lapse rate) are consistent with this mechanism and the corresponding stability reduction over a range of CMIP5 members. Andrews and Webb (2018) construct the tropical Pacific warming pattern index (TPI) where the mean warming change in the WP is subtracted from that in the EP. If the mechanism proposed by Zhou et al. (2016) and Andrews and Webb (2018) is a main driver of feedback change over time across models we would expect a positive correlation of the TPI change with the change of total feedback across models. However, Andrews and Webb (2018) do not find such a correlation, but suggest that this lack of linkage might be because the index is not refined well enough, or that other regions are drivers of the total feedback change. Zhou et al. (2017) introduced a Green's function approach to investigate the influence of regional warming on global cloud feedback evolution and Dong et al. (2019) expand this to global total feedback. They find that the influence of the EP as well as of northern and southern polar regions on total feedback is small and argue that this is because these regions are stably stratified and local changes are confined there. However, their analysis points to the Indo-Pacific warm pool (IPWP) region, including the WP and large parts of the Indian Ocean, as having the largest impact on the change over time of total feedback. In a follow-up study, Dong et al. (2020) succeed to reconstruct the change of total feedback from the IPWP surface warming reasonably well for members of CMIP5, but not of CMIP6. Dong et al. (2020) suggest that the failure of reconstruction for CMIP6 results from strong EP and Southern Hemisphere midlatitude cloud feedbacks.

Like Dong et al. (2020), we perform a feedback analysis of instantaneous CO_2 forcing experiments in a suite of ESMs from CMIP5 and CMIP6. We adopt their terminology in assigning the years 1–20 to the *early* period and the years 21–150 to the *late* period (see also Ceppi and Gregory 2017). "Change over time" in the following refers to the change from early to late period. Unlike in previous studies, which

either use individual models (e.g., Andrews and Webb 2018; Dong et al. 2019) or compare CMIP5 and CMIP6 (e.g., Dong et al. 2020), we group models according to their feedback changes in order to find prominent similarities and differences that potentially can explain the differences in total feedback change. Given our model grouping, we examine the influence of surface warming, stability, and feedback changes in various specific regions on total feedback change and we investigate correlations across models.

The remainder of this study is structured as follows. Section 2 gives a short overview over the model simulation data we consider and section 3 describes the radiative kernel method used to estimate individual climate feedbacks. In section 4 we present the results and offer some discussion. Section 5 concludes with a summary.

2. Models and experiments

We use data from the CMIP5 and CMIP6 abrupt4xCO2 experiments where the CO_2 concentration is abruptly increased by a factor of 4 in comparison to a preindustrial equilibrium condition. The minimum length of this experiment in CMIP5 and CMIP6 is 150 years although there exist longer simulations for some models (Taylor et al. 2009; Eyring et al. 2016). Hence, we only analyze the first 150 years for each available model run. All abrupt4xCO2 experiments have a corresponding preindustrial control experiment (piControl) that is in radiative equilibrium. For each of the quantities, following Zelinka et al. (2020), we take a 21-yr running mean over this control run and subtract from the abrupt4xCO2 run to derive the changes due to the CO_2 forcing and to remove a possible model drift. All models used in the present study are listed and referenced in Tables S1 and S2 in the online supplemental material.

3. Methods

Radiative kernel method

The radiative kernel method (Soden et al. 2008) can be used to derive the magnitude of individual radiative feedbacks activated due to a temperature change associated with a forcing of the climate system. It is based on the assumptions that 1) the TOA net radiation (N) can be described as a function of independent climate system variables (i.e., these variables do not interact or interact only little) and that 2) the radiative flux change due to a small change in one of these variables is linear. As an example, the surface albedo (SA) kernel can hence be defined as

$$N(a+\delta a,T,w,c) - N(a,T,w,c) = N(\delta a) = \frac{\partial N}{\partial a}(a,T,w,c)\delta a \equiv K^a \delta a, \tag{2}$$

where a is the SA, T the temperature, w the water vapor (WV) mixing ratio, c a set of cloud properties, and δa a small SA perturbation; K^a represents the SA kernel.

Radiative kernels for a specific climate system variable are derived from a climate model or a reanalysis. In a given climate state with TOA imbalance $N(a, T, w, c)$, the variable in question is perturbed by a specific amount (e.g., 1% for SA per grid cell) and then only the radiation code is executed. This yields the TOA imbalance of the given climate state with perturbed SA, $N(a + \delta a, T, w, c)$, where δa is 1%. The radiative kernel, K^a, can then be calculated from Eq. (2). Since radiative transfer schemes in climate models are well tested and fairly similar across models, a set of kernels based on a given model is expected to be applicable across models (e.g., Shell et al. 2008). This is supported by the fact that feedbacks calculated with radiative kernels derived from different climate models yield mostly similar results (e.g., Soden et al. 2008), although results for SA feedback can vary considerably across kernels (Donohoe et al. 2020; Hahn et al. 2021). Radiative kernels can be generated for surface and air temperature, WV mixing ration, and SA. These are all either direct outputs of the here-used CMIP model simulations or can be calculated from them.

Since the effects of clouds on radiation are strongly nonlinear, the radiative kernels described here are not appropriate for cloud feedback [for radiative kernels for cloud feedback using cloud-top pressure and optical thickness that are not available in the simulations used here, see Zelinka et al. (2012, 2016)]. However, the radiative flux change due to cloud changes, $N(\delta c)$, can still be estimated by adjusting the change in cloud radiative effect (CRE; calculated as TOA all-sky minus TOA clear-sky radiation, for TOA radiation being positive downward) by the cloud masking of the other radiative feedbacks. This procedure is described in Soden et al. (2008) and can be performed by adding the difference of kernel-derived clear-sky minus all-sky radiative flux changes to the CRE as calculated from model output:

$$N(\delta c) = \mathrm{CRE} + (K^a_{\mathrm{cs}} - K^a_{\mathrm{as}})\delta a + \left(K^T_{\mathrm{cs}} - K^T_{\mathrm{as}}\right)\delta T + (K^w_{\mathrm{cs}} - K^w_{\mathrm{as}})\delta w + \left(F^{4x}_{\mathrm{cs}} - F^{4x}_{\mathrm{as}}\right). \tag{3}$$

The subscripts "as" and "cs" denote all-sky and clear-sky, respectively, and F^{4x} represents the forcing due to a quadrupling of the CO_2 concentration. For a derivation of Eq. (3) see appendix A.

The above-given assumptions of the kernel method allow a decomposition of the total feedback parameter α in Eq. (1) into individual feedback parameters related to physical processes and we can rewrite the equation as

$$N = F + (\alpha_a + \alpha_T + \alpha_w + \alpha_c)\Delta T_s, \tag{4}$$

where α_a, α_T, α_w, and α_c are the feedback parameters due to a change in SA, surface and air temperature, WV mixing ratio, and cloud properties, respectively. This decomposition together with the kernel-derived radiative responses for the individual variables yields the possibility of performing the Gregory method for each individual feedback and the feedback parameters themselves can be derived via a linear regression, similar to the total feedback parameter (e.g., Block and Mauritsen 2013). As an example we can write for the radiative effect of the SA feedback:

$$N(\delta a) = \alpha_a \Delta T_s, \tag{5}$$

where $N(\delta a)$ is the TOA radiative flux change due to the change in SA, which is derived with the radiative kernels according to Eq. (2). Note that the forcing terms in Eqs. (3) and (4) are constant in abrupt forcing experiments. Thus, they do not influence the feedback which is calculated as the linear regression slope. The temperature feedback (α_T) may be split into the Planck feedback and the lapse-rate (LR) feedback. The Planck feedback can be derived by assuming a uniform temperature change equal to the surface temperature change throughout the atmospheric column. It can hence be calculated by using the surface (skin) temperature change on all atmospheric levels and multiplying it on the air temperature kernel. Subtracting the Planck feedback from the air temperature feedback gives the LR feedback. This feedback is strongly linked to tropospheric stability because it originates from the stratification of warming in the atmosphere. In a stably stratified region a surface warming is confined at lower levels, which implies that, excluding other influences, the warming at higher levels in that region is lower (i.e., the stability decreases). This results in a positive LR feedback, since the lapse rate reduces the cooling efficiency. Hence, a reduction in tropospheric stability is typically connected with an increase in LR feedback. On the other hand, if there is an upper-level warming, while the surface warming is comparatively weaker, meaning an increase in stability, a negative LR feedback ensues since the cooling to space is hereby stronger than it would have been under uniform warming. An increase in stability therefore tends to be associated with a decrease in LR feedback.

If the assumptions of the kernel method are appropriate, the sum of the individual feedback parameters should be approximately equal to the total feedback as derived from model output radiative fluxes using the Gregory method. To test if these assumptions are sufficiently fulfilled in a given model simulation, we employ the clear-sky linearity test (CSLT; Shell et al. 2008). This test consists of comparing the sum of the kernel-derived clear-sky feedbacks (i.e., $\alpha_a^{cs} + \alpha_T^{cs} + \alpha_w^{cs}$) with the total clear-sky feedback derived from the Gregory method using model output clear-sky radiative fluxes. In appendix B we show that in the kernel decomposition used here it does not matter if all-sky or clear-sky feedbacks are compared. Following Ceppi and Gregory (2017), we exclude models whose kernel-derived clear-sky feedback sum deviates more than 15% from the total clear-sky feedback in *both* the early and late period calculated independently.

We calculated the feedbacks for 66 members of CMIP5 and CMIP6 listed in Tables S1 and S2 using six different sets of radiative kernels (Soden et al. 2008; Shell et al. 2008; Block and Mauritsen 2013; Huang et al. 2017; Pendergrass et al. 2018; Smith et al. 2018). A maximum of 37 members pass the CSLT for the Shell et al. (2008) kernels and we focus our analysis on these results. There is, however, only a small variation in feedbacks derived from different kernels. It is found that three outliers (MIROC-ES2L, GISS-E2-R, and GISS-E2.2-G) lie on opposite sides of the distribution of models, especially for the WV and LR feedbacks. We exclude these models since they may artificially enhance correlations across models of different aspects that are investigated in the present study. Where in the following we mention "all models" we refer to the remaining 34 models. The whole analysis was repeated with a CSLT error threshold of 20% for which 50 models pass using the Shell et al. (2008) kernels. Generally, the results are little different from the original analysis and the conclusions do not change. The same is true if the complete analysis is conducted using the Pendergrass et al. (2018) kernels for which 36 (CSLT threshold 15%) and 44 models (CSLT threshold 20%) pass the CSLT. Note that SA feedback and its change over time calculated with the Shell et al. (2008) kernels exhibit the smallest standard deviation across models of all tested kernels while for the Pendergrass et al. (2018) kernels they exhibit the largest standard deviation. Hence, we conclude that the above-mentioned dependence of SA feedback on the chosen set of radiative kernels does not significantly impact our results.

We apply a simple mask for the stratosphere by excluding values on pressure levels exceeding a lower threshold that varies with the cosine of the latitude between 300 hPa at the poles and 100 hPa at the equator. Thus, our feedbacks do not include stratospheric adjustments.

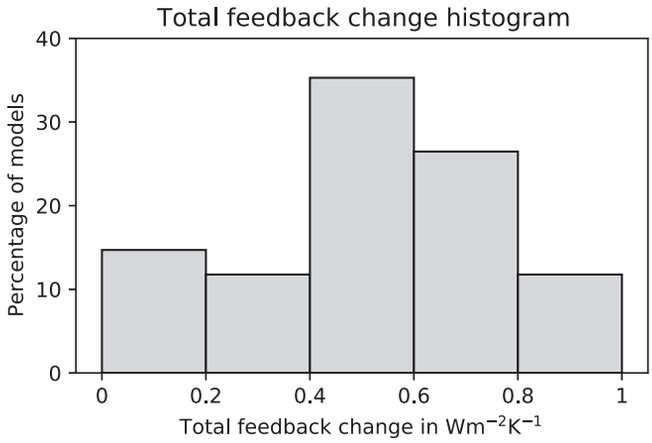

FIG. 1. Distribution of total feedback change from early (years 1–20) to late period (year 21–150) across models ("all models"; refer to section 3 for details). For the distribution of all 66 models with available data see Fig. S1 in the online supplemental material.

4. Results

We now first present the results of the kernel feedback analysis and then proceed to compare groups of models distinguished by their feedback changes. Informed by this analysis we investigate the impact of regional warming and stability changes on total feedback change.

a. *Feedback analysis*

Figure 1 shows the distribution of total feedback change from early to late period derived from the Gregory method across all models. Most models concentrate

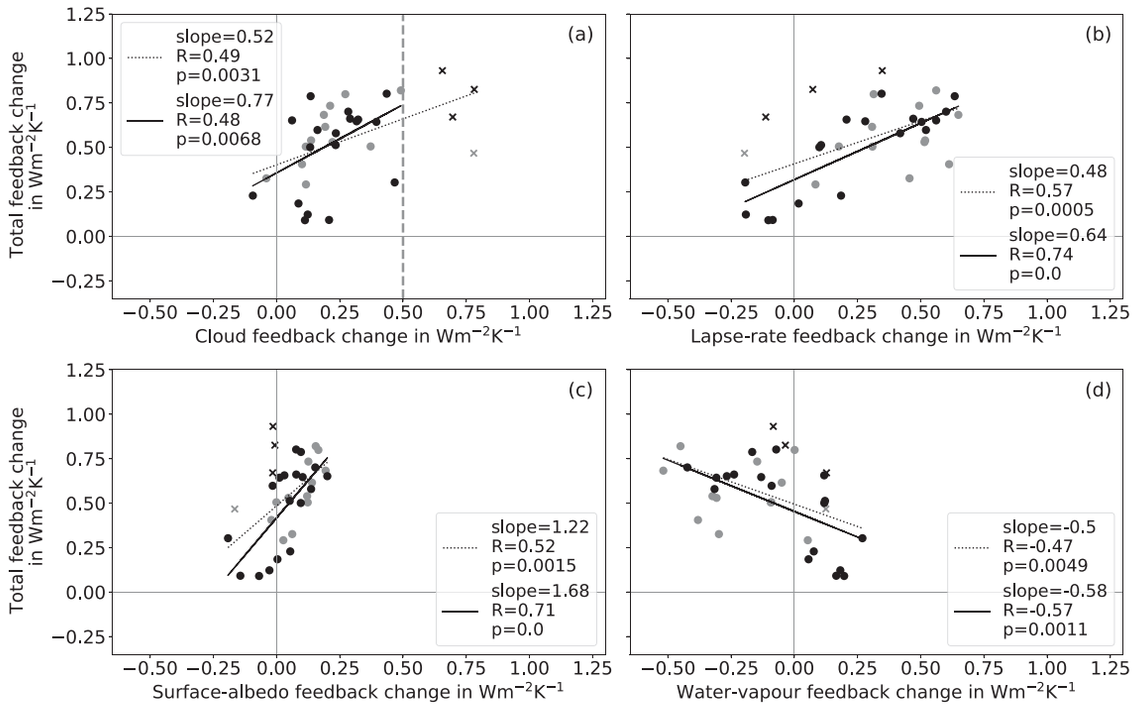

FIG. 2. Total vs individual kernel-derived feedback changes from early (years 1–20) to late period (years 21–150). The dotted black line represents the linear regression for all models and the solid black line represents the regression if the strong cloud feedback change models are excluded (see text for discussion). These models are indicated by × markers. In (a) the dashed gray vertical line shows the threshold value (0.5 W m^{-2} K^{-1}) for "strong cloud feedback change". Members of CMIP5 are depicted in gray and members of CMIP6 in black. The R and p values are the correlation coefficient and significance based on a two-sided p value calculated from a Wald test with a t distribution, respectively.

around 0.5 W m^{-2} K^{-1} change but there is considerable spread with some models showing little feedback change while others show considerable change of close to 1 W m^{-2} K^{-1}. Notably, none of the investigated models exhibits a negative change. The few models that do in fact exhibit negative change fail the CSLT or are considered outliers (see section 3 as well as Tables S1 and S2 and Fig. S1).

We now proceed with the kernel feedback analysis, and decompose the total feedback into individual components to investigate the cause for the total feedback change. Figure 2 shows the intermodel linkage between individual feedback changes and the total (i.e., kernel sum) feedback change. For some models with a large feedback change, the change of the cloud feedback is particularly strong ($\Delta\alpha_c > 0.5$ W m^{-2} K^{-1}; Fig. 2a and Table S3). From comparison with the other feedbacks it is clear that for these models, the change in cloud feedback is the dominating mechanism. These models are shown with × markers in Fig. 2 and correlations are shown both for all models (black dotted line) and for models excluding the strong cloud feedback change models (black solid line). In the latter set of models, a large total feedback change mostly results from a large LR feedback change (Fig. 2b), although the cloud feedback change seems important as well. For this set of models there appears to be a strong and significant correlation between the LR feedback change and the total feedback change. The SA (Fig. 2c) and Planck (not shown) feedback changes are generally smaller but their correlation with the total feedback change is strong provided that the models with large cloud feedback change are excluded. The WV feedback change is anticorrelated with the total feedback change (Fig. 2d), and this anticorrelation is again strong if the strong cloud feedback change models are excluded. We conclude that the set of models with strong cloud feedback change responds somewhat differently to a forcing compared to the remainder of models in that the cloud feedback change is exceptionally large and explains most of the total feedback change while the other feedback changes are less important for these models. Most models in this group (this also holds in the case of a CSLT threshold of 20%) have particularly large positive cloud feedback change over the Southern Ocean (not shown). This region has been noted for positive cloud feedback change due to cloud solid–liquid phase change due to recent updates of cloud-physics parameterization (i.e., feedback changes not directly related to stability changes; Bjordal et al. 2020). Hence, we disregard the models with strong cloud feedback change from our analysis. For the other models there appears to be a strong dependence of the total feedback change on the change of the LR feedback indicating that the change of this feedback explains a large part of the total feedback change. In the following section we thus investigate this group more extensively.

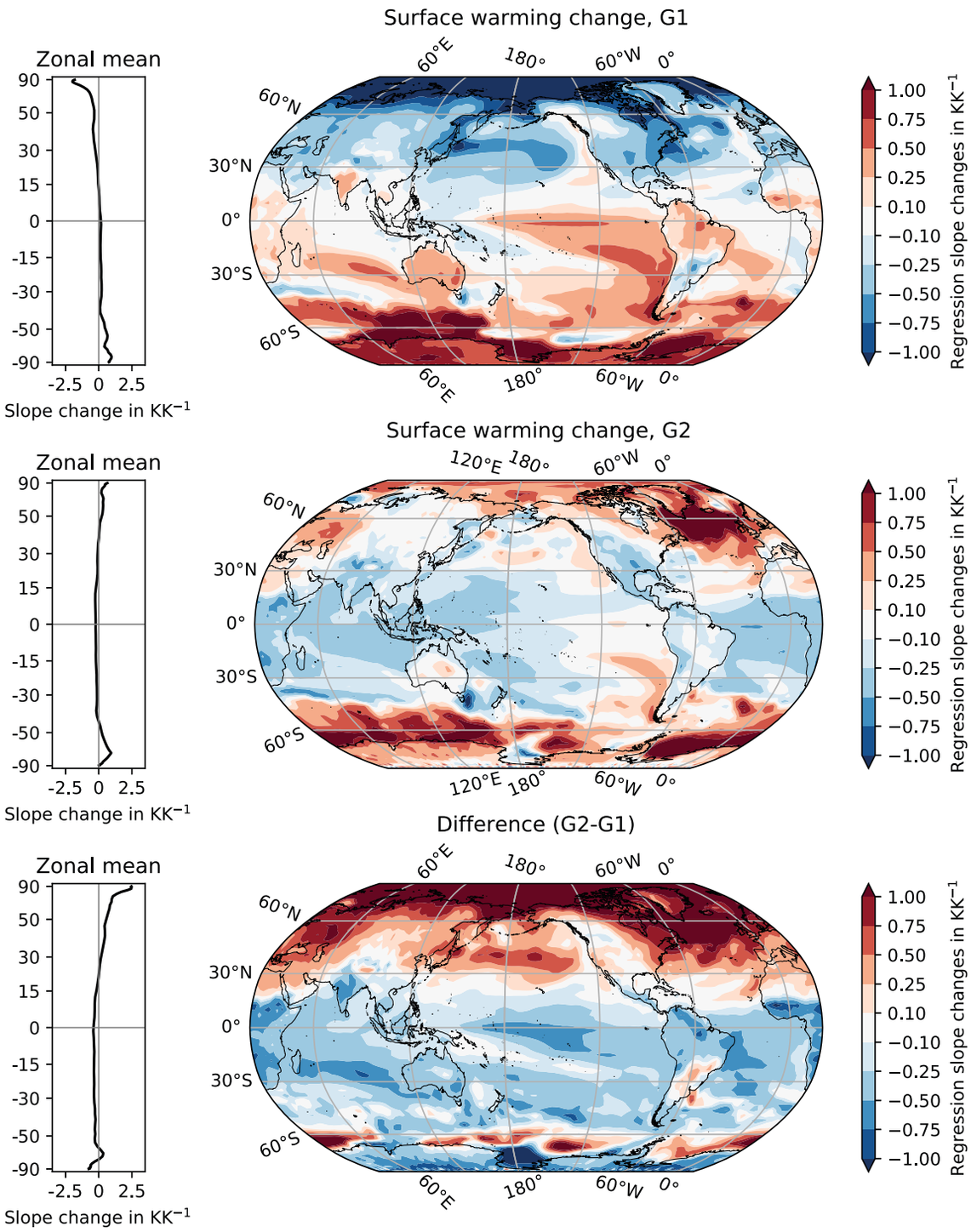

FIG. 3. (right) Maps and (left) zonal means of surface-warming change relative to the global-mean change time averaged over models with (top) weak (G1) and (middle) strong (G2) lapse-rate feedback change, and (bottom) their difference (G2 − G1). The surface-warming change is calculated by first regressing local annual-mean surface temperature on global annual-mean surface temperature for the periods of years 1–20 (early) and 21–150 (late), and then subtracting the results of the early from the late period (cf. Ceppi and Gregory 2017). Note that the *y* axis in the zonal-mean plot is scaled by the cosine of the latitude to obtain an equal-area perspective. See Figs. S2 and S3 for the maps for the early and late period, respectively.

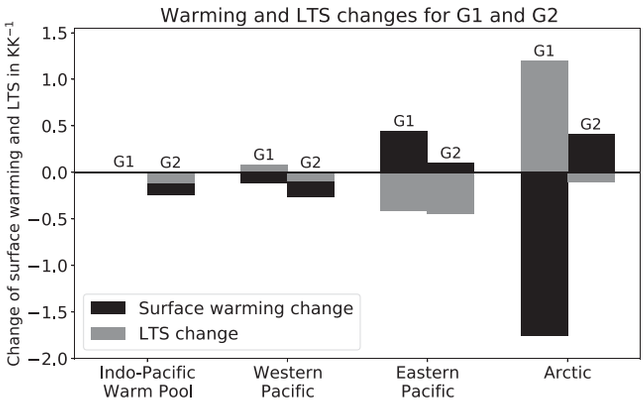

FIG. 4. Indo-Pacific warm pool, western Pacific, eastern Pacific, and Arctic surface-warming change (black) as well as lower-tropospheric stability (LTS; see text for details) change (gray) averaged over models with weak (G1) and strong (G2) lapse-rate feedback change (refer to the text for details). As in Fig. 3, the change of surface warming and LTS is calculated similarly by first regressing local surface temperature and LTS on global-mean surface temperature for both the early (years 1–20) and late period (years 21–150) and then subtracting the results of the early from the late period (cf. Ceppi and Gregory 2017). Note that while the Arctic, eastern Pacific, and western Pacific regions are of similar size, the Indo-Pacific warm pool region is much larger (see text for discussion and Table S6).

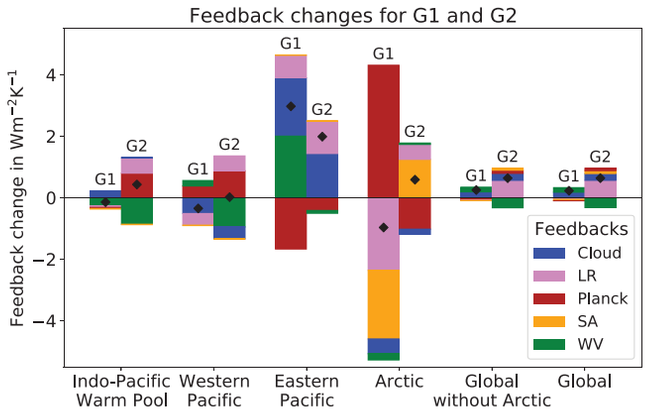

FIG. 5. Indo-Pacific warm pool, western Pacific, eastern Pacific, Arctic, global mean without Arctic, and global-mean feedback changes from the early (years 1–20) to the late period (years 21–150) decomposed into parts indicated by the color coding averaged over models with weak (G1) and strong (G2) lapse-rate feedback change (refer to the text for details). Feedbacks are added on top of each other for each sign and ordered according to their individual magnitudes. The black diamonds represent the total local feedback change.

b. *Comparing two model groups*

Excluding the models with strong cloud feedback change, we now more closely examine the remainder for which a strong coupling between the LR feedback change and the total feedback change is found. To investigate general differences between models, we divide this group into two subgroups, models with a weak LR feedback change (G1; $\Delta\alpha_{LR} > 0.1$ W m^{-2} K^{-1}; see Table S4) and models with a strong LR feedback change (G2; $\Delta\alpha_{LR} > 0.5$ Wm^{-2} K^{-1}; see Table S5). We conduct a sensitivity analysis by varying the thresholds by + and −0.05 W m^{-2} K^{-1}, respectively, and find that the overall conclusions do not change. Notably, since for these models the LR feedback change dominates the total feedback change, G1 tends to include models with weak total feedback change whereas G2 models with strong total feedback change. The slight positive feedback change in G1 is due to small positive changes in cloud and WV feedback that are counteracted by even smaller negative changes in Planck, LR, and SA feedback.

Similar to Ceppi and Gregory (2017) and Andrews and Webb (2018), we focus first on surface temperature. Figure 3 shows the change from early to late period of local surface temperatures regressed on global-mean surface temperature averaged over G1 (top panel) and G2 (middle panel), as well as the difference between the two groups (G2 minus G1, bottom panel). It is clear that both groups exhibit positive warming change relative to the global-mean change in the eastern subtropical Pacific as well as over the Southern Ocean and Antarctica. Furthermore, G1 exhibits stronger positive changes relative to the global average across almost the whole Southern Hemisphere, while G2 shows mostly negative changes between 30°S and 30°N. Accordingly, the difference between G1 and G2 is mostly negative south of 30°N with the exception of some parts north of Antarctica. The IPWP is noticeably different between the two groups, with little change in G1 but negative change in G2. However, the largest difference between the groups occurs in the Arctic. Here G1 exhibits negative change while the G2 change is positive. This means that the models showing little LR feedback change tend to quickly warm the Arctic in response to the abrupt greenhouse gas forcing while the Arctic warming relative to global mean during the later period is weaker. In contrast, there is a delay in relative Arctic warming in the models having large LR feedback change and this relative warming increases over time.

We now proceed with an investigation of individual regions with respect to their warming and stability change. As a stability metric we use the lower tropospheric stability (LTS), which is taken as the difference between surface and 700-hPa potential temperature (Klein and Hartmann 1993). First, we examine the Arctic since this region exhibits the largest difference between the two model groups. We also consider the regions pointed out in Andrews and Webb (2018) and Dong et al. (2019), namely the EP and the WP as well as the IPWP (see section 1). The warming and LTS changes averaged over the four regions are shown in Fig. 4 for both model groups. Note that the Arctic, the EP, and the WP are of roughly similar size, while the IPWP is considerably larger (~20 times as large; see Table S6).

In the Arctic (which we define as 75°–90°N) the largest surface-warming and LTS changes as compared to the other regions are found. Moreover, also here the biggest difference between the two model groups is apparent. G1 exhibits a strong negative warming change over time, while in G2 the change is positive but much weaker. The changes in LTS are opposite those of the surface warming. In accordance with accelerated early relative Arctic surface warming, the Arctic

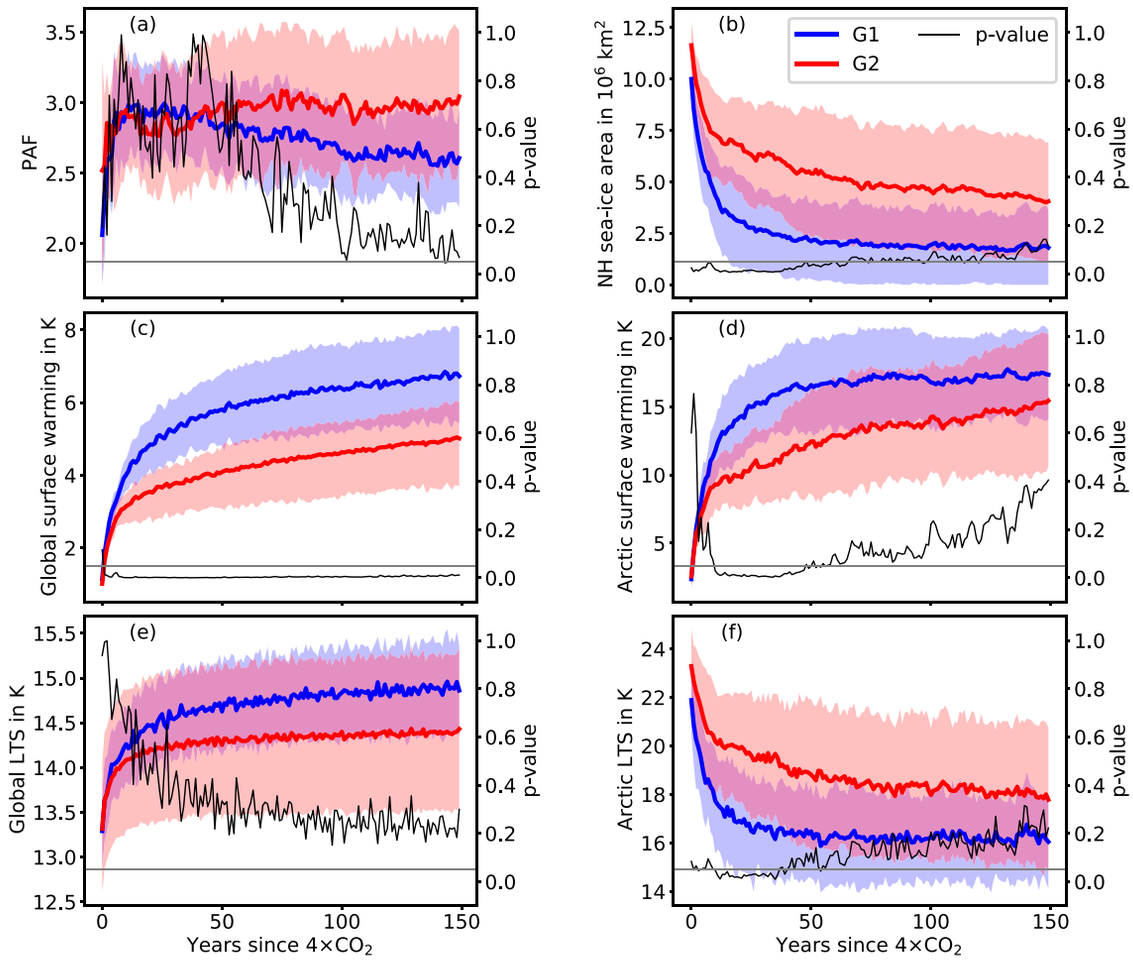

FIG. 6. (a) Polar amplification factor (PAF), (b) Northern Hemisphere (NH) sea ice area, (c) global mean, and (d) Arctic surface temperature anomaly, and (e) global mean and (f) Arctic lower-tropospheric stability (LTS; see text) averaged over models with weak (G1; blue) and strong (G2; red) lapse-rate feedback change (refer to the text for details). The lines indicate the multimodel means and the shading denotes the ±1-sigma spread. The PAF is calculated as the Arctic (75°–90°N) surface temperature anomaly divided by that of the global mean. The Northern Hemisphere sea ice area is calculated by multiplying the sea ice fraction in a grid cell by the grid cell area and then integrating over the NH. The black 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 of 0.05. No sea ice data were available for one member of G2 (BCC-CSM2-MR) so this model is excluded in (b).

LTS in G1 decreases strongly in the early period but as the relative Arctic surface warming subsequently slows down (cf. Fig. 6a), the decrease of LTS is much weaker in the late period, leading to the positive change of LTS seen in Fig. 4. In G2 the relative Arctic surface warming does not slow down (cf. also Fig. 6a) and the Arctic LTS decreases similarly in both early and late period, leading to the small negative change seen in Fig. 4.

The EP exhibits a positive surface warming change in G1, which is much larger than in G2. However, the magnitude of the change of G1 in the EP is generally much smaller than in the Arctic. Regarding the LTS change in the EP, G1 and G2 show smaller difference as compared to the surface warming. Note that the change is slightly more negative in G2. Hence, even though the surface in the EP warms less strongly in G2, the EP stability in G2 decreases more than in G1. Since total global-mean feedback changes more in G2 than in G1, this may indicate that the delayed EP surface warming, although a robust feature across ESMs in response to an abrupt greenhouse gas concentration increase, does not significantly influence total feedback change. This supports the findings of Dong et al. (2019) that it is the IPWP surface warming rather than the tropical east–west SST gradient that is important for total global feedback change.

As expected for regions of ascent, the LTS changes only little in the WP and the IPWP in both groups. In G1, both the WP and IPWP exhibit little or no surface warming change, while there is a negative change in both regions in G2. Since G1 consists mostly of models with weak total feedback change and G2 mostly of models with strong total feedback change this lends support to the findings of Dong et al. (2019, 2020) that the IPWP surface warming largely controls the total feedback change. Since the IPWP is a region of ascent, a surface warming there is quickly spread vertically

and in upper levels also horizontally through the atmosphere and hence increases the global cooling efficiency. In contrast, energy associated with a surface warming in other, generally more stable regions is not radiated to space as efficiently. The surface warming change in the IPWP thus determines the total global feedback change to a significant degree. Since G2 experiences a reduction in IPWP surface warming relative to the global average while in G1 it is almost constant, this implicates a reduction of the magnitude of the negative total feedback in G2 while the G1 total feedback should only change little, which is what we see.

We further investigate the changes of the kernel-derived feedbacks in the respective regions. Figure 5 shows the individual feedback changes averaged over the regions IPWP, WP, EP, and Arctic as well as the globe without the Arctic, and the entire globe, for G1 and G2.

For both groups the most important feedback changes in the Arctic are the Planck, SA, and LR feedbacks. However, the groups differ in the magnitude of the individual changes and, more importantly, the sign of each individual feedback change is different between groups. As expected from the surface warming change (Fig. 4), the Planck feedback change is positive in G1 and negative in G2. In contrast, the LR and SA feedback changes are both negative in G1 while they are both positive in G2. The sum of the LR and SA feedback change is stronger than the Planck feedback change in both groups so that for G1 the Arctic feedback change is in total negative, while for G2 it is positive.

The LR and SA feedbacks in the Arctic are both strongly connected to the melting of sea ice and they are known to interact considerably with each other (e.g., Graversen et al. 2014). As Earth warms, the sea ice melts and exposes the darker ocean surface, reducing the albedo and hence causing a positive feedback. Furthermore, the sea ice acts as an insulating layer impairing heat exchange between the atmosphere and the much warmer ocean so that above the sea ice low atmospheric temperatures far below the melting point of sea ice may prevail especially during the dark seasons. The melting of sea ice therefore strongly enhances the heat exchange between atmosphere and ocean, leading to a strong increase in Arctic surface temperatures. Because the Arctic is stably stratified, this warming is trapped at the surface and a positive LR feedback is induced. Hence, the melting of sea ice controls the evolution of both the LR and SA feedbacks. Thus, a plausible explanation for the large difference in Arctic feedback changes in G1 and G2 is that as a result of the stronger early-period Arctic warming in G1 (Fig. 6d), the sea ice quickly melts (Fig. 6b), causing strong positive LR and SA feedbacks. In the later period most of the ice has already melted, resulting in a weakening of both the LR and SA feedback in the Arctic. Accordingly, these feedbacks are much less positive in the late period, inducing the strong negative change of these feedbacks in G1 seen in Fig. 5. Furthermore, the strong early warming leads to a considerable reduction in atmospheric stability, causing an increase of the local cooling efficiency from the early to the late period, which hampers further warming (Fig. 6f). In G2, on the other hand, the Arctic warming is weaker in the early period, inducing much less sea ice melt and weaker positive LR and SA feedbacks. Also, the Arctic remains more stable in G2 than in G1. As the Arctic in the late period continues to warm more strongly in G2 than in G1 (Figs. 6a,d), more sea ice melts (Fig. 6b), hereby strengthening the LR and SA feedbacks and causing the positive change as shown in Fig. 5. As the Arctic surface warming is weaker in G2 than in G1 (Fig. 6d), the stratification remains more stable and the local cooling efficiency weaker (Fig. 6f), enhancing the surface warming also in the late period.

In the EP the two model groups exhibit differences especially in WV and Planck feedback change but in both groups the total feedback change is strongly positive. Notably, the difference in total feedback change in the EP between G1 and G2 is the opposite of the global total feedback change. The feedback changes in the WP and the IPWP are similar in G2, while in G1 the individual feedback changes have opposite sign between the two regions. However, the total feedback change for both regions in G1 and G2 is small compared to the EP and the Arctic. The differences in total feedback between G1 and G2 in the regions IPWP, WP, and Arctic are similar to the difference in global-mean total feedback change between G1 and G2 and hence appear important for the difference between the two model groups. Conversely, the difference in the EP is opposite that of the global mean, indicating that this region is less important for the general between-group difference.

By construction, in the global mean the positive LR feedback change dominates for G2 with some smaller positive changes in cloud and SA feedback. The WV feedback change is negative and somewhat compensates the other feedbacks but there still remains a considerable positive total feedback change. In contrast, the total G1 feedback change is only slightly positive. The SA and LR feedback changes are small but negative and compensated for by the positive cloud and WV feedback changes. In spite of the large changes in the Arctic, excluding this region from the global mean has little effect on the feedback decomposition, owing to the fact that the Arctic (>75°N) covers less than 2% of Earth's surface area. Thus, in this simple global-mean perspective, the Arctic changes, though large, have only little influence. However, previous research has pointed to remote influences of Arctic changes, especially on the development of the Hadley circulation (Feldl and Bordoni 2016; Feldl et al. 2017). Generally, positive high-latitude feedbacks reduce the meridional temperature gradient, hereby weakening the meridional heat transport, which at low latitudes is partly accomplished by the thermally direct cell constituting the Hadley circulation. Stronger positive Arctic SA and LR feedbacks are thus associated with a stronger weakening of the Hadley circulation. We find evidence for this coupling since, on the one hand, in G1 the meridional overturning streamfunction in the Northern Hemisphere tropics weakens more strongly in the early period but then remains constant, and on the other hand, in G2 this streamfunction weakens less in the early period but continues to weaken further in the late period (Fig. S4).

Other research highlights the influence of low-latitude changes on the Arctic. For example, Yoo et al. (2012) show that convective heating in the IPWP enhances poleward

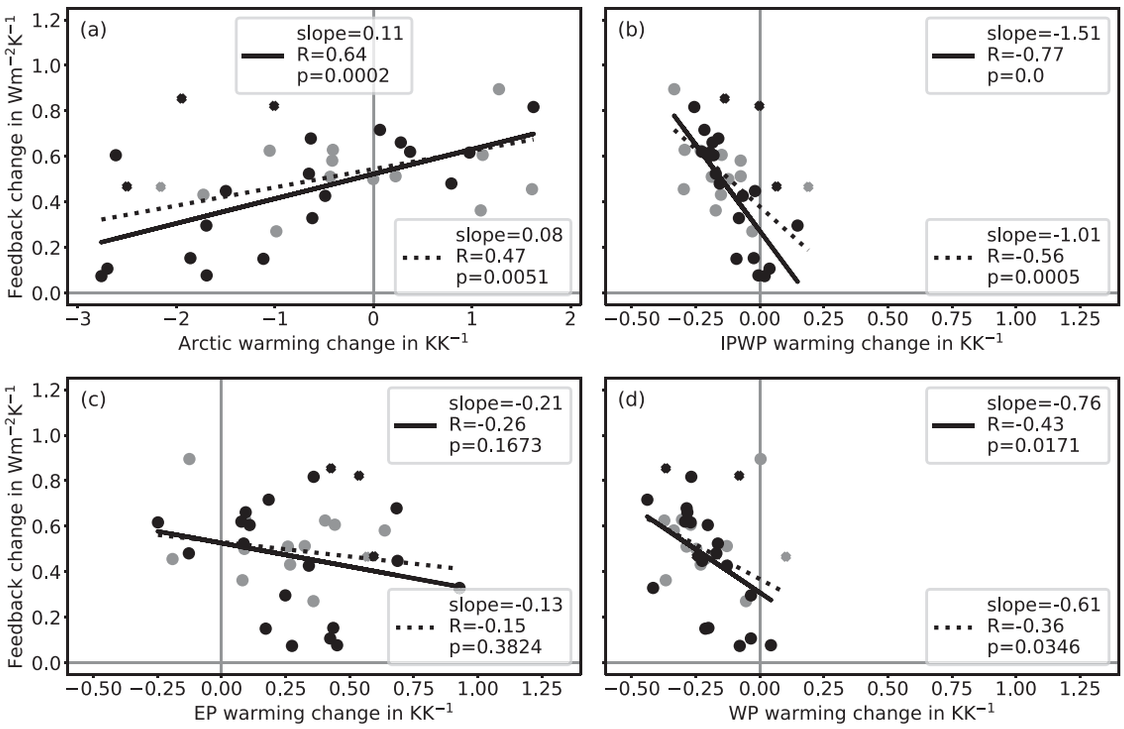

FIG. 7. As in Fig. 2, but for total feedback change derived with the Gregory method vs surface-warming change in (a) the Arctic, (b) the Indo-Pacific warm pool (IPWP), (c) the eastern Pacific (EP), and (d) the western Pacific (WP). Note that the *x*-axis scale in (a) differs from that in (b)–(d).

propagating Rossby waves leading to increased Arctic surface air temperatures. Lee (2014) presents a theory for this, according to which poleward heat transport intensifies and increases Arctic temperatures given that a greenhouse gas forcing leads to enhanced localized convection in the tropics.

In summary, previous literature highlights both the influence of the Arctic on lower latitudes and vice versa. The effects of the coupling of these regions on our results have not been disentangled in the present study but are left for future investigations.

We now move away from the two-model-group frame and consider correlations across all models.

c. *Regional warming and feedbacks across models*

Similar to the approach by Andrews and Webb (2018) using the TPI, we investigate correlations of the change over time of the surface warming of the four regions described above with the change over time of total feedback across models (Fig. 7). If the surface warming in a specific region has a major influence on the global cooling efficiency, as the stability hypothesis states, and as shown for the IPWP in Dong et al. (2019), this should be supported by a correlation of the change of the surface warming in that region relative to global-mean surface warming with the change of total feedback over time. As Fig. 7 (black solid lines) shows, the

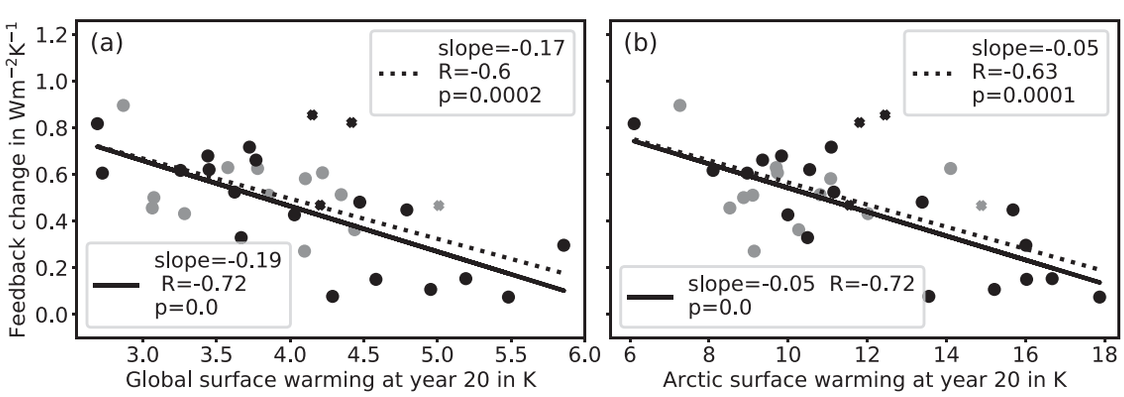

FIG. 8. As in Fig. 2, but for total feedback change derived with the Gregory method vs (a) global- and (b) Arctic-mean surface warming at year 20 (averaged over years 18–22) of the abrupt4xCO2 simulation. Note the different *x*-axis scales.

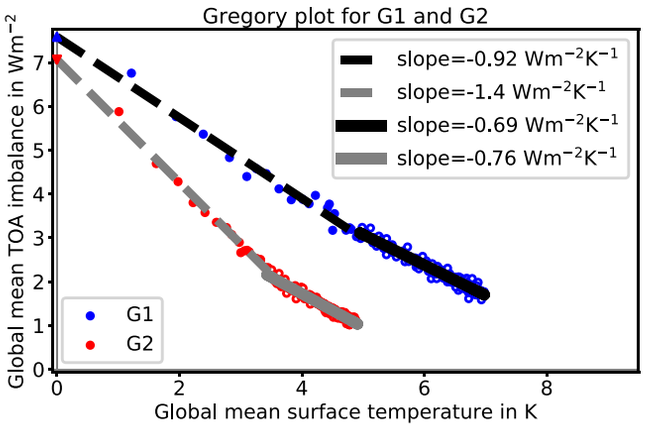

FIG. 9. Gregory plot for models with weak (G1; blue) and strong (G2; red) lapse-rate feedback change. The filled circles represent the early period (years 1–20) and the unfilled circles the late period (years 21–150). The dashed and solid lines show the early- and late-period linear regressions, respectively. The regression slope indicates the total feedback (see text for details).

strongest correlation is found for the IPWP and there is a slightly weaker correlation for the Arctic. There is a moderate correlation for the WP but close to no correlation for the EP, supporting the suggestion (see section 4b) that the latter region has little influence on the global-mean total feedback change. Note that the correlations weaken considerably if the models with a strong cloud feedback change are included (Fig. 7, black dotted lines). However, if the cloud feedback is excluded from the kernel-derived feedback sum, the correlations for the Arctic, WP, and IPWP, but not the EP, are strong regardless of these models being included or not (Fig. S5). This provides support for the stability hypothesis in general and the findings of Dong et al. (2019) and Dong et al. (2020) in particular (i.e., the IPWP surface-warming evolution largely controls total climate feedback change). Here CMIP5 and CMIP6 are not investigated separately as is done by Dong et al. (2020). They point out that while they can reconstruct the total feedback evolution over time from the IPWP surface warming evolution for CMIP5, this fails for CMIP6. With the caveat that we use a somewhat different set of models here, our results support the suggestion in Dong et al. (2020) that the failure to reconstruct total feedback evolution from IPWP warming in CMIP6 is because of a particularly large cloud feedback change in a number of members of CMIP6. If these members are excluded, the total feedback evolution may still be reasonably well reconstructed from the IPWP surface warming. Considering that due to a likely missing negative cloud lifetime feedback component (Mülmenstädt et al. 2021) some members of CMIP6 may overestimate the cloud feedback change, adjusting the cloud physics parameterization to include this feedback component might make a reconstruction of the total feedback evolution from IPWP surface warming as suggested by Dong et al. (2019, 2020) successful for CMIP6 as well.

As pointed out above, the surface warming change in the Arctic is significantly positively correlated with the change in total feedback. This means that the more a model increases its Arctic warming relative to the global mean over time, the more it changes its total feedback over time. Since the Arctic is a stably stratified region, this fits the stability hypothesis, since the surface warming over time shifts to more stable regions, decreasing Earth's cooling efficiency and hence weakening the total feedback. However, as pointed out in section 4b, it is difficult to establish causality between Arctic and global changes and the large differences in Arctic warming could be a by-product of the following: Models that warm strongly in the global mean in the early period (such as the members of G1) generally exhibit an even stronger early warming in the Arctic, while in the later period the global warming has a larger pace than the Arctic warming (Figs. 6c,d). Notably, these tend to be the models that have a weak total feedback change (Fig. 8). On the other hand, models with strong total feedback change tend to warm less quickly, and while they exhibit similar Arctic amplification their *absolute* Arctic warming is weaker (see Figs. 6a,c,d for the G1–G2 comparison). As explained in section 4b, the Arctic warming is strongly connected to sea ice changes and the resulting SA and LR feedbacks, and due to this double feedback loop, small differences in surface warming are quickly enhanced locally. Hence, the difference in early warming across models (the stronger the early warming, the weaker the total feedback change; Fig. 8) explains the relationship of local Arctic warming and feedback evolution with total feedback change across models. However, as discussed in section 4b, Arctic changes have been found to have remote influences (Feldl and Bordoni 2016; Feldl et al. 2017) and hence may have important indirect influences on feedback change. Further research is needed to establish the relevance of the large changes in the Arctic for the total global feedback change.

5. Discussion and conclusions

This study investigates why different numerical climate models change their total feedback differently over time using members of the CMIP5 and CMIP6 archives. We perform a radiative kernel analysis to decompose the total feedback into individual parts associated with different feedback processes and group the models according to similarities in individual feedbacks.

We investigate the differences between a group of models with weak (G1) and strong (G2) lapse-rate feedback change. It is revealed that the Arctic is the region with the largest difference between these groups and a region with large warming, stability, and individual feedback changes over time. These changes as well as their differences between the groups are strongly linked to Arctic sea ice changes. It is found that members of G1 warm much more quickly and exhibit faster Arctic sea ice melt, triggering stronger positive early-period Arctic surface-albedo and lapse-rate feedbacks than those of G2. Since the Arctic sea ice after a few decades has mostly vanished in the G1 models, the surface-albedo and lapse-rate feedbacks are much weaker in the late period for those models. Conversely, in the members of G2, more Arctic sea ice remains and the surface-albedo and lapse-rate feedbacks

become more positive in the late period. We furthermore find evidence for influence of Arctic changes on lower-latitude circulation, in accordance with previous studies (Feldl and Bordoni 2016; Feldl et al. 2017). However, it is difficult to determine if the Arctic changes and their low-latitude influence have causal relevance for the total feedback change.

As the members of G1 warm up much faster than the members of G2, they also experience much more rapid changes. Hence, the G1 members quickly reach a new warmer climate state with almost constant total feedback on the time scales considered here. In contrast, the members of G2, due to their slower warming, experience changes over climate states on longer time scales that become apparent in the present study. Evidence for this is provided by the fact that whereas the early-period feedback in G2 is much more negative than in G1 and the early-period warming of G2 is smaller than in G1, the late-period feedbacks of both groups are similar (Fig. 9).

Additionally, we generally find support for the results of previous studies (Dong et al. 2019, 2020) that the relative surface warming in the tropical Indo-Pacific is well correlated with total feedback change across models, which fits the previous finding that the surface warming in this region may control the total feedback change via the mechanism explained by the stability hypothesis.

Future research should focus on disentangling the cause-and-effect relationships. Questions of interest include the following: Why do some models warm much more quickly than others? Why do models warming slower have a large feedback change over time? How much does the sensitivity to warming of Arctic sea ice vary across models? How strong is the influence of Arctic changes on the Hadley circulation and how do changes in the Hadley circulation affect global feedbacks?

The change of climate feedback in numerical climate models as investigated here is of relevance, since it affects estimates of Earth's sensitivity to a forcing (e.g., due to anthropogenic greenhouse gas concentration changes in the atmosphere). Understanding the reasons behind climate feedback change will make it possible to compare the climate model results with real-world changes under ongoing climate change and hence improve the robustness of climate sensitivity estimates.

Acknowledgments. We thank Hege-Beate Fredriksen and Martin Rypdal for useful discussions and suggestions. 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 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 three anonymous reviewers and the editor, Isaac Held, for their help in improving the work.

Data availability statement. The original 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 feedbacks using the radiative kernels is described in section 3. Scripts and data generated during this research are stored at Nird storage facilities provided by UNINETT Sigma2, and are available on request from the corresponding author.

APPENDIX A

Derivation of Adjusted Cloud Radiative Effect

Here we present a brief derivation of Eq. (3). We start by splitting up the clear-sky and all-sky top-of-atmosphere (TOA) imbalances (N_{cs} and N_{as}, respectively) into contributions of individual climate state variables:

$$N_{cs} = N_{cs}(\delta a) + N_{cs}(\delta T) + N_{cs}(\delta w) + F_{cs}^{4x}, \tag{A1}$$

$$N_{as} = N_{as}(\delta a) + N_{as}(\delta T) + N_{as}(\delta w) + N(\delta c) + F_{as}^{4x}. \tag{A2}$$

Note that to fully represent the TOA imbalance, we need to include the forcing due to the quadrupling of the CO_2 concentration in both cases (i.e., F_{cs}^{4x} and F_{as}^{4x}). Since, as described in section 3, the cloud radiative effect (CRE) is defined as $N_{as} - N_{cs}$ we now subtract Eq. (A1) from Eq. (A2) to derive this quantity:

$$\begin{aligned}\mathrm{CRE} = N_{as} - N_{cs} &= N_{as}(\delta a) + N_{as}(\delta T) + N_{as}(\delta w) + N(\delta c)\\ &\quad + F_{as}^{4x} - N_{cs}(\delta a) - N_{cs}(\delta T) - N_{cs}(\delta w) - F_{cs}^{4x}.\end{aligned} \tag{A3}$$

Solving for $N(\delta c)$, substituting CRE, and rearranging yields

$$\begin{aligned}N(\delta c) = \mathrm{CRE} &+ N_{cs}(\delta a) - N_{as}(\delta a) + N_{cs}(\delta T) - N_{as}(\delta T)\\ &+ N_{cs}(\delta w) - N_{as}(\delta w) + F_{cs}^{4x} - F_{as}^{4x}.\end{aligned} \tag{A4}$$

Using Eq. (2) for the surface-albedo flux and its equivalents for the other climate state variables with both clear-sky and all-sky kernels, Eq. (3) is obtained.

APPENDIX B

Why Clear-Sky and Not All-Sky Linearity Test?

We here briefly show that in the kernel decomposition used in the present study the difference between total kernel-derived feedback and total feedback derived from model output radiative fluxes is the same for both clear-sky and all-sky conditions. We start from Eq. (A2), but now indicate the quantities involving radiative kernels by the superscript K:

$$N_{as}^{K} = N_{as}^{K}(\delta a) + N_{as}^{K}(\delta T) + N_{as}^{K}(\delta w) + N^{K}(\delta c) + F_{as}^{4x}. \tag{B1}$$

Note that $N^K(\delta c)$ is not solely based on radiative kernels since it involves CRE, which is calculated from model

output radiative fluxes. Substituting Eq. (2) for the surface-albedo flux and its equivalents for temperature and water vapor as well as Eq. (3) for the cloud flux, this expression reduces to

$$N_{\mathrm{as}}^{K} = \mathrm{CRE} + K_{\mathrm{cs}}^{a}\delta a + K_{\mathrm{cs}}^{w}\delta w + K_{\mathrm{cs}}^{T}\delta T + F_{\mathrm{cs}}^{4x}. \tag{B2}$$

On the right-hand side we can now substitute N_{cs}^{K} as well as the definition of CRE (i.e., $N_{\mathrm{as}}^{M} - N_{\mathrm{cs}}^{M}$) and obtain

$$N_{\mathrm{as}}^{K} - N_{\mathrm{as}}^{M} = N_{\mathrm{cs}}^{K} - N_{\mathrm{cs}}^{M}, \tag{B3}$$

where the superscript M denotes a model output radiative flux quantity. The CSLT is chosen here since this test is the direct comparison between the sum of the kernel-derived feedbacks and the total feedback derived from model output radiative fluxes with the Gregory method.

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.

——, J. M. Gregory, and M. J. Webb, 2015: The dependence of radiative forcing and feedback on evolving patterns of surface temperature change in climate models. *J. Climate*, **28**, 1630–1648, https://doi.org/10.1175/JCLI-D-14-00545.1.

Armour, K. C., C. M. Blitz, and G. H. Roe, 2013: Time-varying climate sensitivity from regional feedbacks. *J. Climate*, **26**, 4518–4534, https://doi.org/10.1175/JCLI-D-12-00544.1.

Arrhenius, S., 1896: On the influence of carbonic acid in the air upon the temperature of the ground. *London Edinburgh Dublin Philos. Mag. J. Sci.*, **41**, 237–276, https://doi.org/10.1080/14786449608620846.

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.

Block, K., and T. Mauritsen, 2013: Forcing and feedback in the MPI-ESM-LR coupled model under abruptly quadrupled CO_2. *J. Adv. Model. Earth Syst.*, **5**, 676–691, https://doi.org/10.1002/jame.20041.

Ceppi, P., and J. M. Gregory, 2017: Relationship of tropospheric stability to climate sensitivity 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, 20 pp.

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.

——, 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.

Donohoe, A., E. Blanchard-Wrigglesworth, A. Schweiger, and P. J. Rasch, 2020: The effect of atmospheric transmissivity on model and observational estimates of the sea ice albedo feedback. *J. Climate*, **33**, 5743–5765, https://doi.org/10.1175/JCLI-D-19-0674.1.

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.

Feldl, N., and S. Bordoni, 2016: Characterizing the Hadley circulation response through regional climate feedbacks. *J. Climate*, **29**, 613–622, https://doi.org/10.1175/JCLI-D-15-0424.1.

——, ——, and T. M. Merlis, 2017: Coupled high-latitude climate feedbacks and their impact on atmospheric heat transport. *J. Climate*, **30**, 189–201, https://doi.org/10.1175/JCLI-D-16-0324.1.

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.

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.

Hahn, L. C., K. C. Armour, M. D. Zelinka, C. M. Bitz, and A. Donohoe, 2021: Contributions to polar amplification in CMIP5 and CMIP6 models. *Front. Earth Sci.*, **9**, 710036, https://doi.org/10.3389/feart.2021.710036.

Huang, Y., Y. Xia, and X. Tan, 2017: On the pattern of CO_2 radiative forcing and poleward energy transport. *J. Geophys. Res. Atmos.*, **122**, 10 578–10 593, https://doi.org/10.1002/2017JD027221.

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.

Lee, S., 2014: A theory for polar amplification from a general circulation perspective. *Asia-Pac. J. Atmos. Sci.*, **50**, 31–43, https://doi.org/10.1007/s13143-014-0024-7.

Martin, G. M., and Coauthors, 2011: The HadGEM2 family of Met Office Unified Model climate configurations. *Geosci. Model Dev.*, **4**, 723–757, https://doi.org/10.5194/gmd-4-723-2011.

Mülmenstädt, J., and Coauthors, 2021: An underestimated negative cloud feedback from cloud life-time changes. *Nat. Climate 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.

Pendergrass, A. M., A. Conley, and F. M. Vitt, 2018: Surface and top-of-atmosphere radiative feedback kernels for CESM-CAM5. *Earth Syst. Sci. Data*, **10**, 317–324, https://doi.org/10.5194/essd-10-317-2018.

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.

Senior, C. A., and J. F. B. Mitchell, 2000: The time-dependence of climate sensitivity. *Geophys. Res. Lett.*, **27**, 2685–2688, https://doi.org/10.1029/2000GL011373.

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**, e2019RG000, https://doi.org/10.1029/2019RG000678.

Smith, C. J., and Coauthors, 2018: Understanding rapid adjustments to diverse forcing agents. *Geophys. Res. Lett.*, **45**, 12 023–12 031, https://doi.org/10.1029/2018GL079826.

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.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2009: A summary of the CMIP5 experiment design. Tech. Rep. 33 pp., https://pcmdi.llnl.gov/mips/cmip5/docs/Taylor_CMIP5_design.pdf?id=98.

Williams, K. D., W. J. Ingram, and J. M. Gregory, 2008: Time variation of effective climate sensitivity in GCMs. *J. Climate*, **21**, 5076–5090, https://doi.org/10.1175/2008JCLI2371.1.

Winton, M., K. Takahashi, and I. M. Held, 2010: Importance of ocean heat uptake efficacy to transient climate change. *J. Climate*, **23**, 2333–2344, https://doi.org/10.1175/2009JCLI3139.1.

Wood, R., and C. S. Bretherton, 2006: On the relationship between stratiform low cloud cover and lower-tropospheric stability. *J. Climate*, **19**, 6425–6432, https://doi.org/10.1175/JCLI3988.1.

Yoo, C., S. Lee, and S. B. Feldstein, 2012: Arctic response to an MJO-like tropical heating in an idealized GCM. *J. Atmos. Sci.*, **69**, 2379–2393, https://doi.org/10.1175/JAS-D-11-0261.1.

Zelinka, M. D., S. A. Klein, and D. L. Hartmann, 2012: Computing and partitioning cloud feedbacks using cloud property histograms. Part I: Cloud radiative kernels. *J. Climate*, **25**, 3715–3735, https://doi.org/10.1175/JCLI-D-11-00248.1.

——, 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.

——, 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.

Zhou, C., M. D. Zelinka, and S. A. Klein, 2016: Impact of decadal cloud variations on the Earth's energy budget. *Nat. Geosci.*, **9**, 871–874, https://doi.org/10.1038/ngeo2828.

——, ——, and ——, 2017: Analyzing the dependence of global cloud feedback on the spatial pattern of sea surface temperature change with a Green's function approach. *J. Adv. Model. Earth Syst.*, **9**, 2174–2189, https://doi.org/10.1002/2017MS001096.

7

Paper II

Eiselt, K.-U., and R. G. Graversen (2023a): On the control of Northern Hemispheric feedbacks by AMOC: Evidence from CMIP and slab-ocean modeling. *J. Climate*, **36**, 6777-6795 , https://doi.org/10.1175/JCLI-D-22-0884.1.

On the Control of Northern 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*

(Manuscript received 1 December 2022, in final form 20 June 2023, accepted 21 June 2023)

ABSTRACT: The climate sensitivity of Earth and the radiative climate feedback both change over time as a result of 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 from the abrupt4xCO2 experiment of phases 5 and 6 of the Coupled Model Intercomparison Project (CMIP) are distinguished: one group showing weak feedback change and the other showing 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 in which the AMOC change is mimicked by changing the ocean heat transport. It is found that in the Northern Hemisphere extratropics 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.

KEYWORDS: Ocean circulation; Climate; Climate change; Climate sensitivity; Climate models

1. Introduction

Equilibrium climate sensitivity (ECS), that is, the magnitude of the warming of 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-the-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 vapor (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 decrease the surface albedo of Earth and thus cause a positive feedback, raising the climate sensitivity (e.g., Hall 2004; Winton 2006; Graversen and Wang 2009). The WV feedback is also positive since at higher temperature water molecules evaporate more readily, increasing atmospheric water vapor, 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 from different warming at different altitudes in the atmosphere. If an atmospheric column warms more aloft than at the surface below, this increases Earth's cooling efficiency and thus constitutes a negative feedback. In contrast, a positive feedback results if the lower atmospheric layers warm more than those aloft. Because of 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). Last, 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 intermodel 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

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JCLI-D-22-0884.s1.

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

DOI: 10.1175/JCLI-D-22-0884.1

does not lead to the cloud feedback being in better agreement with expert judgement.

Despite 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 (preindustrial) 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 nonlinear 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 favor 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-Pacific warm pool (IPWP) as one region that appears particularly important for the pattern effect. The IPWP is characterized 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 phases 5 and 6 of the Coupled Model Intercomparison Project (CMIP) 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 extrapolar 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 with 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 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 Ocean. Members of CMIP5 and CMIP6 show a large spread in both the preindustrial 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), whereas 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 analyze 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 composed 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 the appendix) while also abruptly raising the CO_2 concentration by a factor of 4 (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 the result with 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. In addition, 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 in which AMOC recovery is mimicked and compared with those without AMOC recovery. Last, 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 parameterization) 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 discuss 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 are 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-yr 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

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 (hereinafter called CESM2-SOM). In this setup, the Community Atmosphere Model, version 6 (CAM6; Danabasoglu et al. 2020), on an ~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-yr fully coupled CESM2 preindustrial-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-yr simulation was continued at our server from a 300-yr preindustrial control simulation conducted at the National Center for Atmospheric Research (NCAR) servers that 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 freshwater 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 increasing the Q-flux in the tropical Atlantic and decreasing 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 farther 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 for 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 (the appendix). 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 nonlinearly with the Q-flux change (see Fig. S1 in the online supplemental material). This may be due to nonlinear aspects of the feedback changes triggered by the Q-flux change. For instance, the Q-flux changes nonlinearly impact the Arctic sea ice melt (see section 4 and Fig. S1). The nonlinear nature of the response to the Q-flux change complicates a quantitative analysis and we thus emphasize 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, that is, 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 (hereinafter 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 preindustrial-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

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

where a is the SA, δa is the SA perturbation (typically 1%), T is the temperature, w is the WV mixing ratio, c is 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 nonlinear 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 streamfunction (based on the variables named *msftmyz* or *msftyyz* in CMIP5 and *msftmz* or *msftyz* in

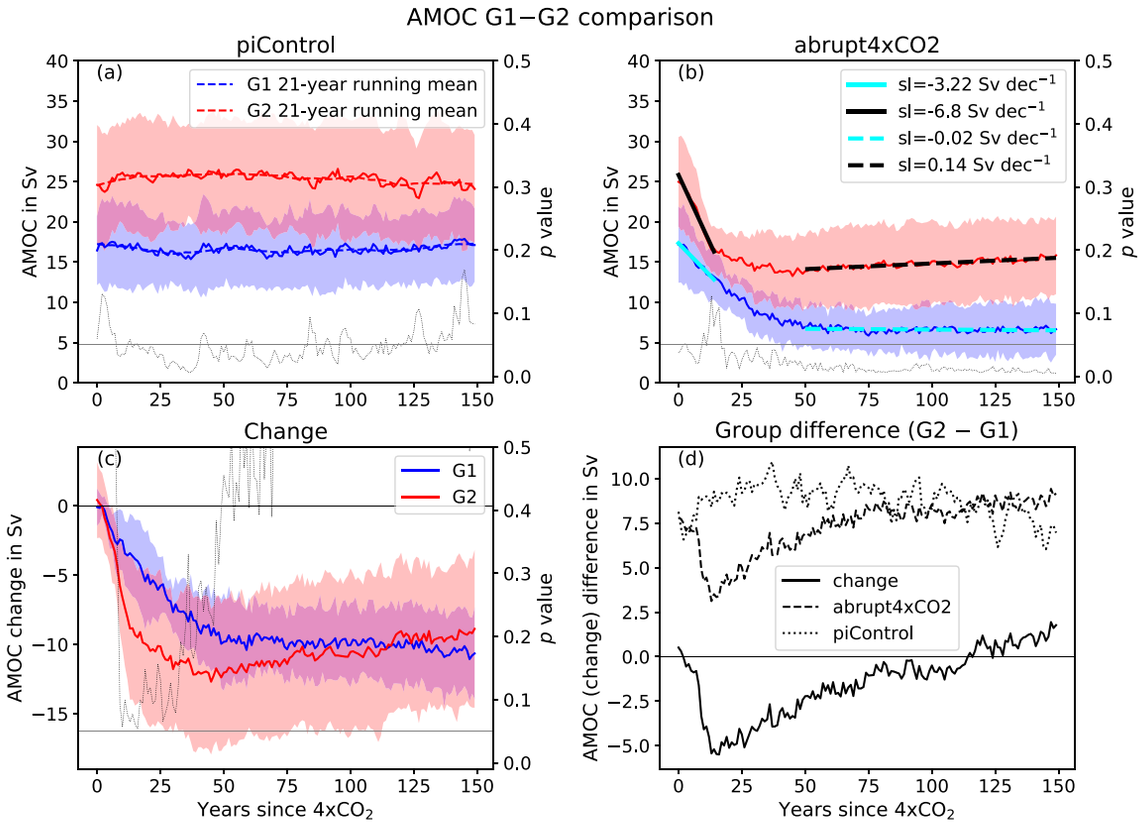

FIG. 1. Atlantic meridional overturning circulation (AMOC) index for G1 and G2 in (a) piControl (both annual and 21-yr running mean; the period corresponds to the abrupt4xCO2 simulation) and (b) abrupt4xCO2, and (c) the difference between abrupt4xCO2 and a 21-yr running mean of the piControl simulation. Also shown is (d) the G2-minus-G1 difference. In (a)–(c), the blue and red solid lines show the G1 and G2 means, respectively; the shading indicates the ±1-sigma (std dev) spread of models; 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. Linear trends for the early (years 1–15; solid) and late (years 51–150; dashed) periods are included in (b). For details of the AMOC index calculation, see section 3b.

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-gridcell 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 longwave and shortwave 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 hypothesize that the difference between the two model groups when it comes to the response in SEB, temperature, and atmospheric stability in the midlatitude 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 midlatitude North Atlantic SEB*

The piControl state AMOC is considerably stronger in G2 than in G1 [on average by 8.77 Sv (1 Sv ≡ 10^6 m^3 s^{-1}); 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

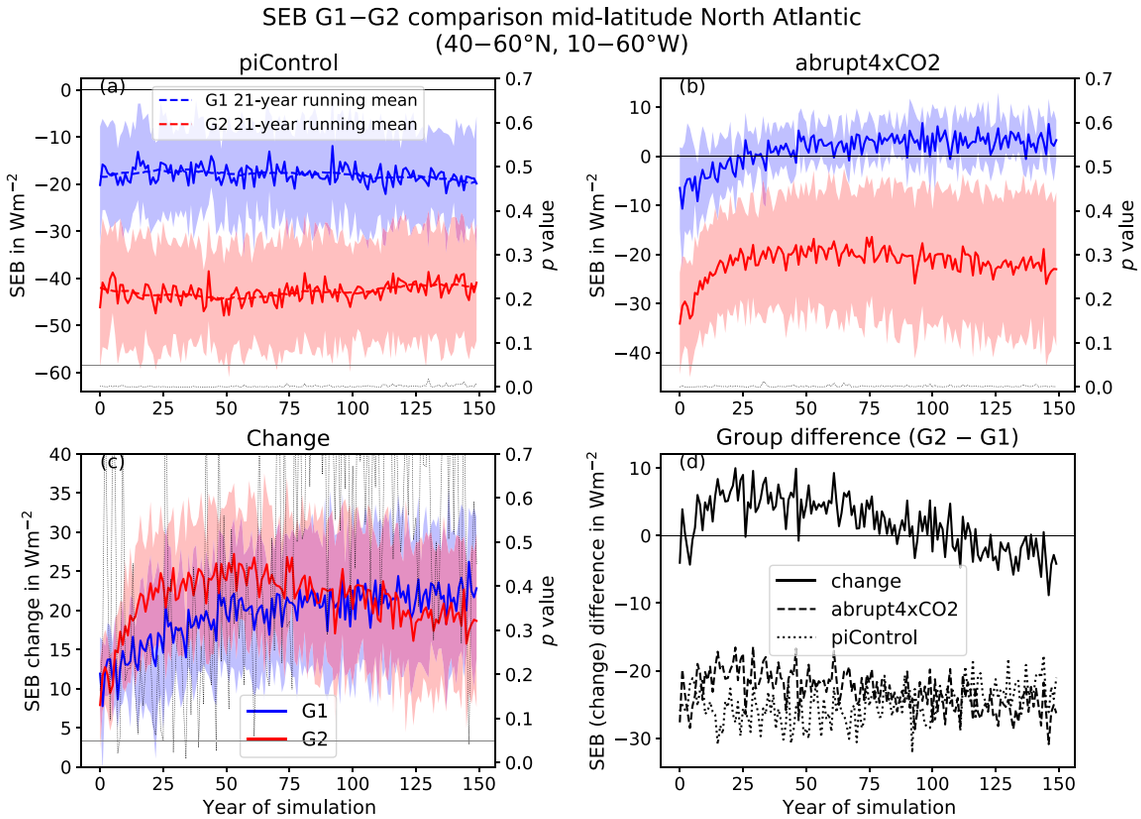

FIG. 2. As in Fig. 1, but for surface energy balance (SEB; units are positive down) averaged over the midlatitude North Atlantic (40°–60°N, 10°–60°W). Note that the members of G1 and G2 for which AMOC was not available were *not* excluded here; however, excluding these members does not qualitatively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

G1 (see Figs. 1b–d as well as Figs. 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 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 (Figs. 1c,d). A similar development obtains for the OHR change (Figs. 2c,d). 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 Tables S1 and S2, respectively, in the online supplemental material.

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 (cf. 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 midlatitude 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 in the online supplemental material). 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 nonlocal factors and is thus less sensitive to the AMOC-induced surface warming. Hence, it is also in agreement with the AMOC impact

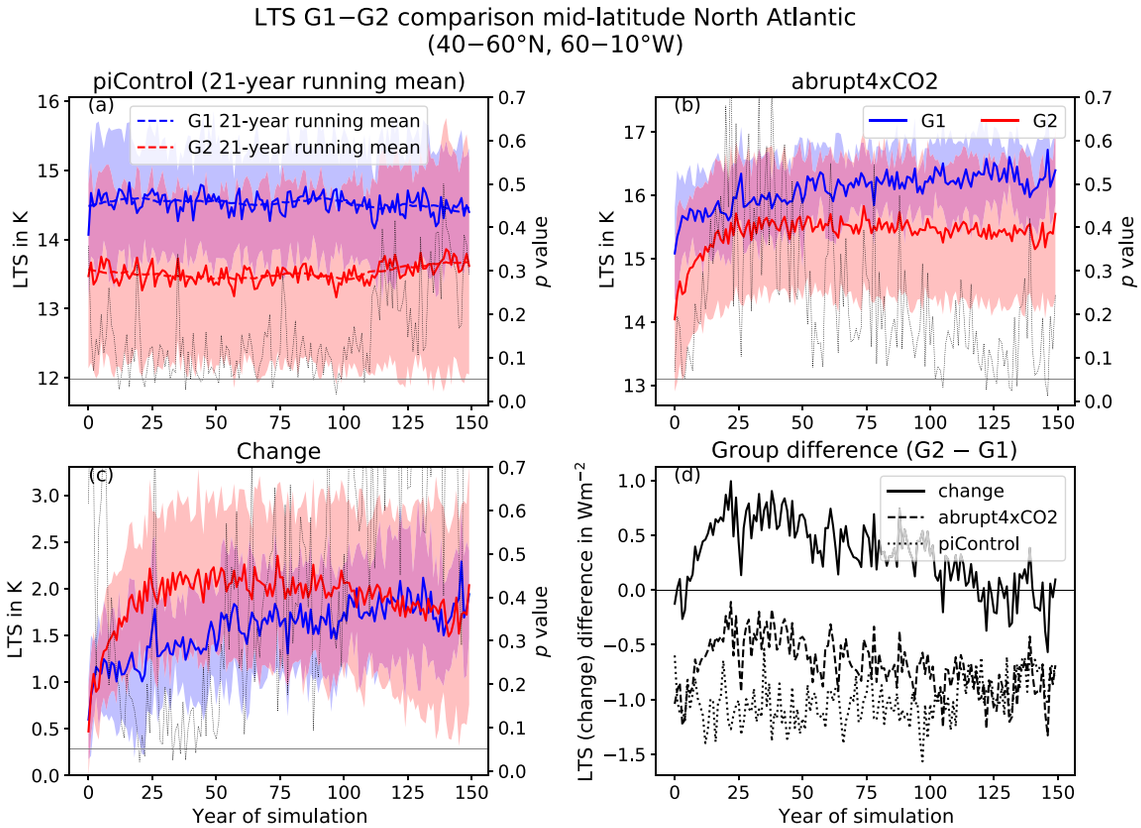

FIG. 3. As in Fig. 1, but for lower-tropospheric stability (LTS) averaged over the midlatitude North Atlantic (40°–60°N, 10°–60°W). The LTS was calculated as the difference between the 700-hPa and surface potential temperature (see Fig. S3 in the online supplemental material for similar plots of the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

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 (Figs. S3b,c in the online supplemental material). However, in G2 the SAT plateaus after a few years, whereas 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 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 (Figs. 3e,f). Thus, the 700-hPa temperature is more similar in the two model groups than is the SAT, although G1 again exhibits more warming. Consistent with the discussion above, this indicates that the 700-hPa temperature is partly driven by nonlocal factors and to a lesser extent by the AMOC.

The change in LTS (Figs. 3b,c) depends on the difference in changes in temperature at the two levels. Because of 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 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), relative 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 integrated* 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

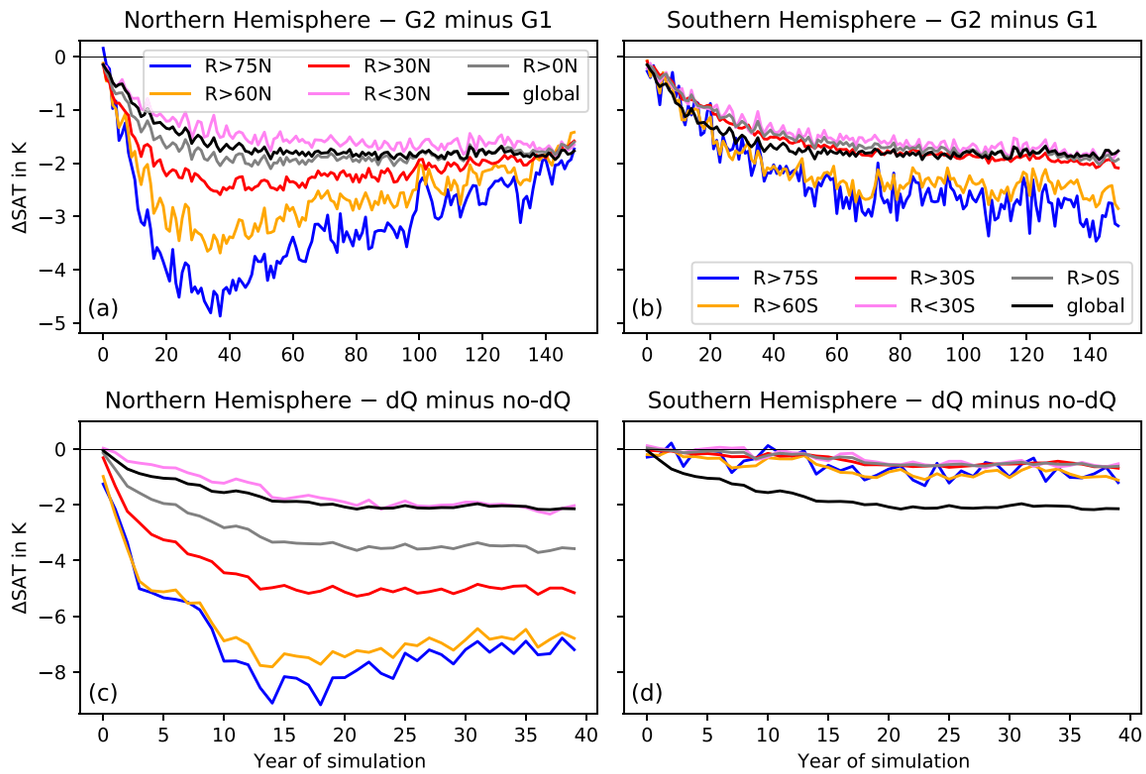

FIG. 4. Difference in surface air temperature (a),(b) between G2 and G1 and (c),(d) between the CESM2-SOM experiments *dQ* and no-*dQ* for (left) the NH and (right) the SH. Shown are zonal averages over latitude regions $R > 75°$, $R > 60°$, $R > 30°$, $R < 30°$ (i.e., 0°–30°), and $R > 0°$ for both hemispheres. In the legend, the degree sign is replaced by N in the case of the NH and by S in the case of the SH. Figure S4 in the online supplemental material shows the *p* values of the differences as determined from a two-sided Welch's *t* test.

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 with CESM2-SOM experiments that 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 hereinafter the degree sign is replaced by N in the case of the Northern Hemisphere (NH) and by S in the case of the Southern Hemisphere (SH).

1) SURFACE AIR TEMPERATURE

Figures 4a and 4b show 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 $R > 0$N and $R > 0$S are similar, the differences are larger farther poleward, with the most extreme differences in the Arctic ($R > 75$N). 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 $R > 75$N and $R > 60$N), but after about year 40, they start decreasing and continue to do so throughout the remainder of the 150-yr 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. Last, 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

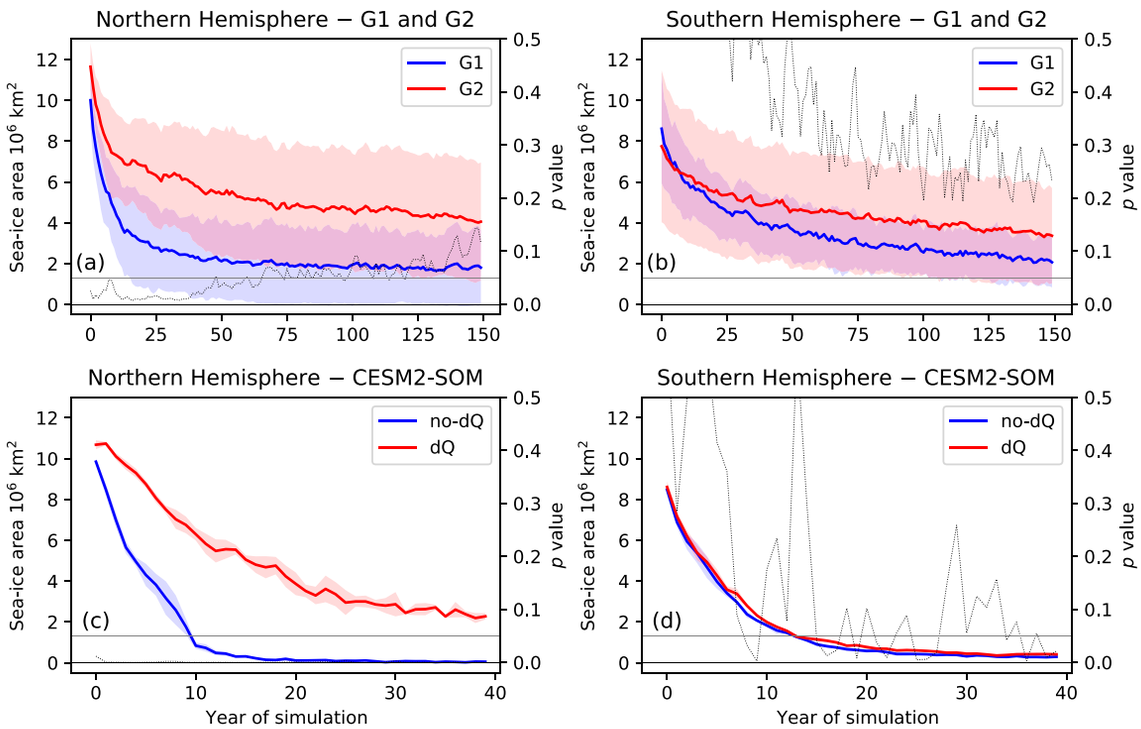

FIG. 5. Sea ice area development, integrated over (left) the NH and (right) the SH, in (a),(b) G1 and G2 and in (c),(d) no-dQ and dQ. The shading denotes the ±1-sigma spread across group members in (a) and (b) and across ensemble members in (c) and (d). See the text for details. No sea ice data were available for one member of G2 (BCC-CSM2-MR), and therefore 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 (c) the p value is almost invisible because it is close to zero.

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 $R > 75$N. Conversely in G2, more sea ice remains, which continues to melt in concert with the steady rise of SAT in $R > 75$N. This causes the difference in SAT between G2 and G1 to decline in the later years of the CMIP simulations (especially in $R > 75$N and $R > 60$N). 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 $R > 75$S and $R > 60$S between G2 and G1 are generally smaller (except in the last simulation years) than in $R > 75$N and $R > 60$N, respectively, and they remain almost constant after about year 50 in the SH.

Figures 4c and 4d present 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 farther north that one samples the results, the larger is the magnitude of the difference between the dQ and the no-dQ experiment, and the differences in $R > 0$N are distinctly larger than in $R > 0$S. Notably, however, while the difference between G2 and G1 in $R > 75$N and $R > 60$N 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 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 $R > 30$N, $R > 60$N, and $R > 75$N. Indeed, in additional experiments with CESM2-SOM where the energy transport recovers after the initial 1-yr decline over the course of years 2–30 of the simulation, the magnitude of the difference between SATs in $R > 30$N and $R > 60$N to the no-dQ experiment also decreases after the initial increase (see Fig. S5 in the online supplemental material). 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-yr 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, in the online supplemental material). However, after about year 25, the Arctic sea ice area

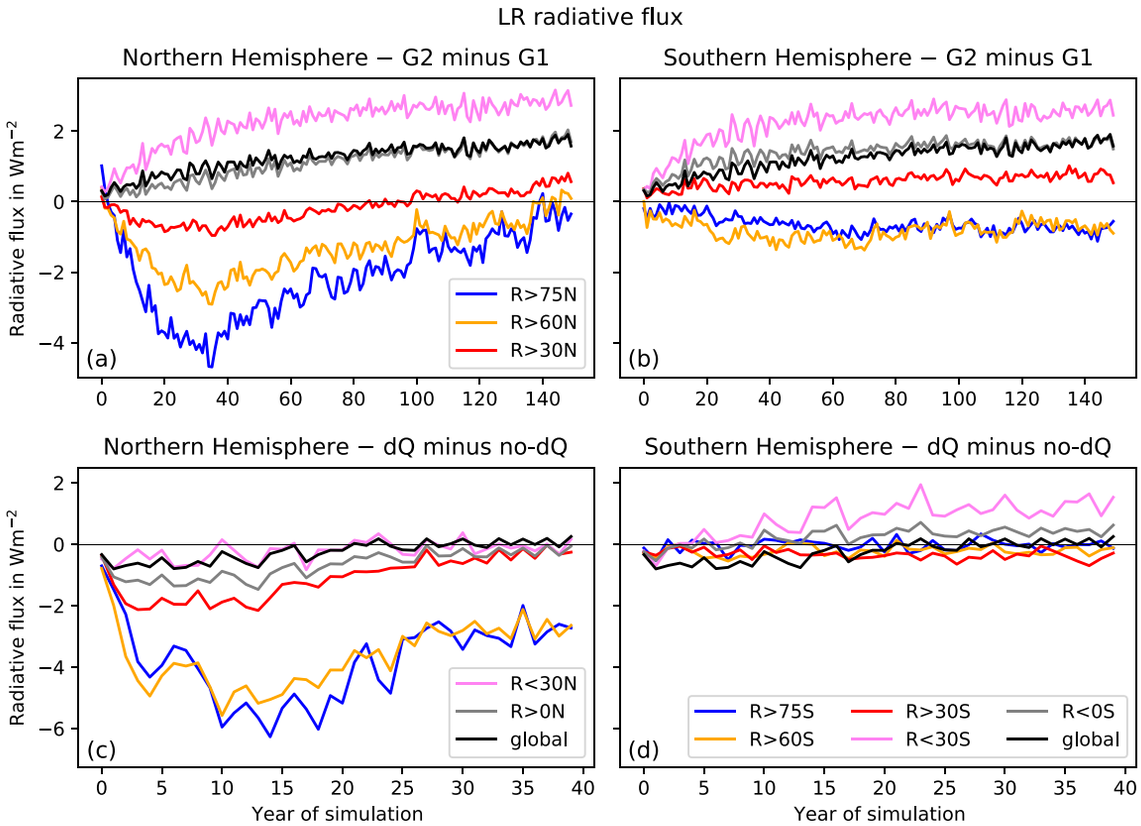

FIG. 6. As in Fig. 4, but for lapse-rate-feedback-induced top-of-the-atmosphere radiative-flux differences.

remains almost constant, consistent with the constant SAT difference between dQ and no-dQ in $R > 30$N, $R > 60$N, and $R > 75$N. Thus, a consistent conjecture for the differences in temperature development in the Arctic, and more generally in the NH extratropics, is that a stronger decline in AMOC in G2 moderates the sea ice loss relative 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 in the online supplemental material).

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 (Figs. 4c,d). 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 extratropical SATs ($R > 30$S, $R > 60$S, and $R > 75$S) are larger than the differences between the CESM2-SOM experiments (cf. Figs. 4b,d). We note that the development of Antarctic sea ice appears unaffected by AMOC recovery (Fig. S6b in the online supplemental material).

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 extratropics are qualitatively reproducible by CESM2-SOM experiments with mimicked AMOC change while differences in the tropics and in the SH require other explanations.

The G2-minus-G1 differences in LR and SA feedback-induced TOA radiative fluxes (Figs. 6a,b and 7a,b, respectively) in $R > 75$N and $R > 60$N, as well as $R > 75$S and $R > 60$S 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 $R > 75$N and $R > 60$N, the differences in both SA and LR fluxes exhibit an initial increase in magnitude, but after about year 40 they decrease. In $R > 75$S and $R > 60$S, the change of the difference is initially slower but continues to slowly increase in magnitude over the whole CMIP simulation period. In the tropics ($R < 30$N and $R < 30$S), the differences in SA feedback-induced TOA radiative flux are

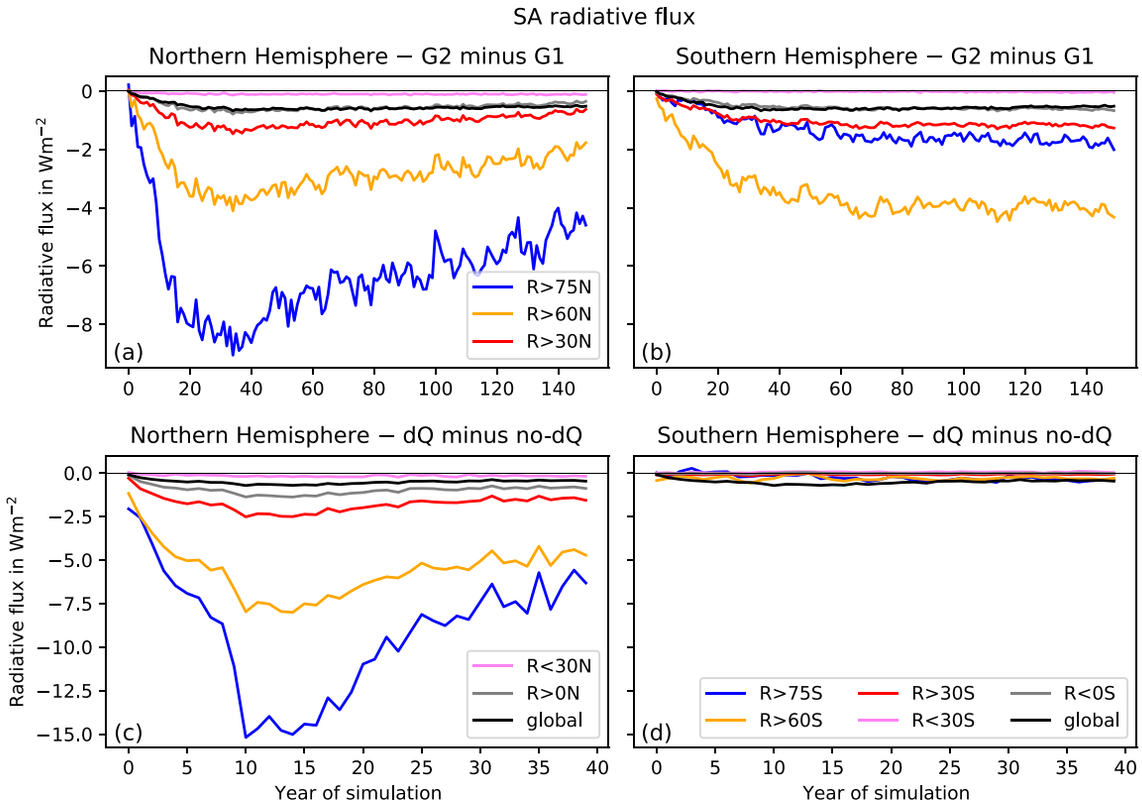

FIG. 7. As in Fig. 4, but for surface-albedo-feedback-induced top-of-the-atmosphere radiative-flux differences.

negligible. However, the differences in LR feedback-induced TOA radiative flux are positive in both $R < 30$N and $R < 30$S and exhibit similar development in these regions. Furthermore, in $R > 30$N the difference in LR TOA radiative flux is initially negative and then becomes increasingly positive while it is always positive in $R > 30$S 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. 6c,d and 7c,d, respectively) are qualitatively similar to the G2-minus-G1 differences in LR and SA TOA fluxes in the NH extratropics (i.e., $R > 30$N, $R > 60$N, $R > 75$N). 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 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 extratropics appears qualitatively more similar to the G2-minus-G1 difference since both show about full recovery (Fig. S9 in the online supplemental material). However, for the SA radiative flux, the recovery after the initial increase is much stronger than for the CMIP model difference (Fig. S10 in the online supplemental material), which is consistent with the development of Arctic sea ice in the experiment explained above (Fig. S6a in the online supplemental material). 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 (Figs. 5b,d), it is consistent that the difference in Antarctic SA flux is not reproduced either. The differences in LR flux in the tropics ($R < 30$N and $R < 30$S) are discussed in more detail in section 4c(3).

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 (Figs. 8a,b). However, similar to the LR and SA fluxes, the differences in the NH regions increase initially faster in magnitude and then remain either constant ($R < 30$N) or start slowly decreasing ($R > 30$N, $R > 60$N, $R > 75$N), while in the SH regions the differences increase

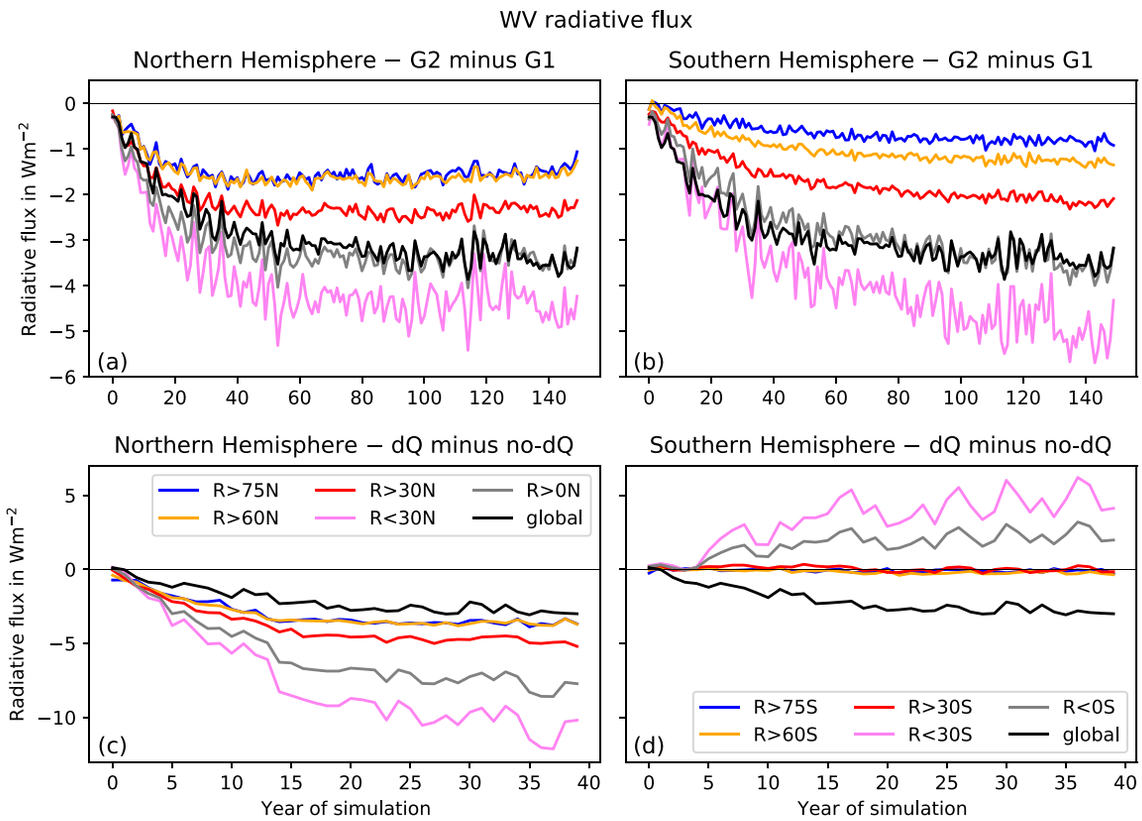

FIG. 8. As in Fig. 4, but for water-vapor-feedback-induced top-of-the-atmosphere radiative-flux differences.

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 extratropical regions. This is also true for the *dQ* minus no-*dQ* differences (Figs. 8c,d). In general, 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 $R > 30$S, $R > 60$S, and $R > 75$S there is little or no difference in WV TOA radiative flux between *dQ* and no-*dQ*, but in $R < 30$S 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 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 4c(3).

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 cloud area, cloud droplet size, cloud phase, cloud height, etc.) whereby these changes are more difficult to interpret.

The longwave cloud (CLW) radiative-flux differences between G2 and G1 are generally smaller than those of the other radiative fluxes (Figs. 9a,b). The differences between the CESM2-SOM experiments in the extratropics 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 between the CESM2-SOM in the tropics are similar to the WV flux differences. The shortwave 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 $R > 75$S 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 ($R < 30$N), 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 4c(3).

In general, it may be concluded that for the NH extratropics, 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 response between G2 and G1. However, differences in the tropics are not captured by the SOM experiments. In the following, the *dQ* and no-*dQ* experiments are further analyzed 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 ($R < 30$N) and the southern tropics ($R < 30$S) are consistent with a concomitant change in the Hadley cell and the difference in

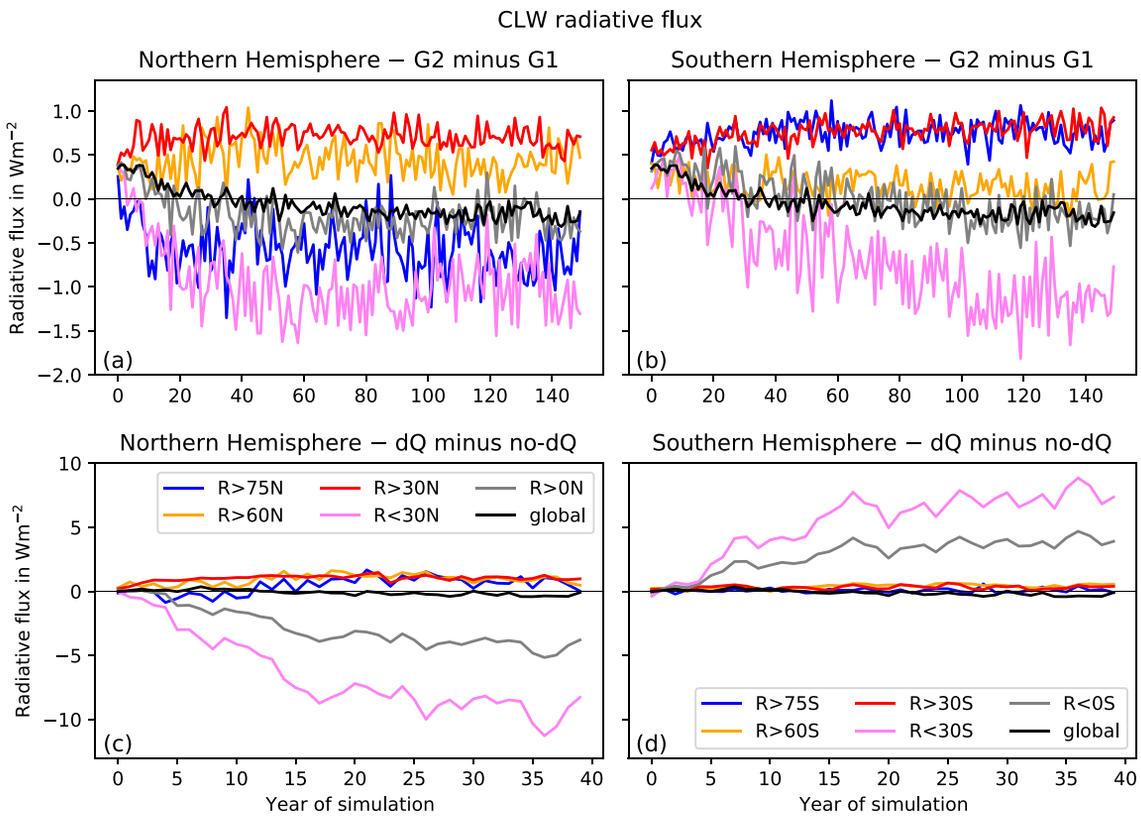

FIG. 9. As in Fig. 4, but for longwave-cloud-feedback-induced top-of-the-atmosphere radiative-flux differences.

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).

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 (Fig. 11a; Figs. S11a and S12 in the online supplemental material). 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 (Fig. 11b; Figs. S11b and S12). As a consequence, there is more ascending motion in the $R < 30$S and more descending motion in $R < 30$N in dQ relative to no-dQ (Fig. 11c; see Figs. S13 and S14 in the online supplemental material 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 in the online supplemental material) leads to a lower specific humidity in $R < 30$N 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 $R < 30$S, although to a lesser degree (see Fig. S16 in the online supplemental material and Figs. 8c,d). Furthermore, the increase in descent in $R < 30$N and ascent in $R < 30$S implies that the cloud content is reduced in $R < 30$N but increased in $R < 30$S (Fig. S17 in the online supplemental material) so that in the former region more longwave radiation can escape to space and more shortwave 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 $R < 30$N, but positive in $R < 30$S (Figs. 9c,d), while the opposite obtains for the CSW flux difference (Figs. 10c,d). Last, in comparing the difference in temperature lapse rates between dQ and no-dQ in the tropics, it is found that the difference in $R < 30$N is positive at lower and negative at higher levels, while the difference in the $R < 30$S is always negative (Fig. S18 in the online supplemental material). This is consistent with the LR feedback induced flux difference being close to zero in the $R < 30$N and positive in $R < 30$S (Figs. 6c,d). However, the governing processes for the difference in lapse rate are less clear. We speculate that due to the larger cloud amount in the $R < 30$S 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 longwave 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

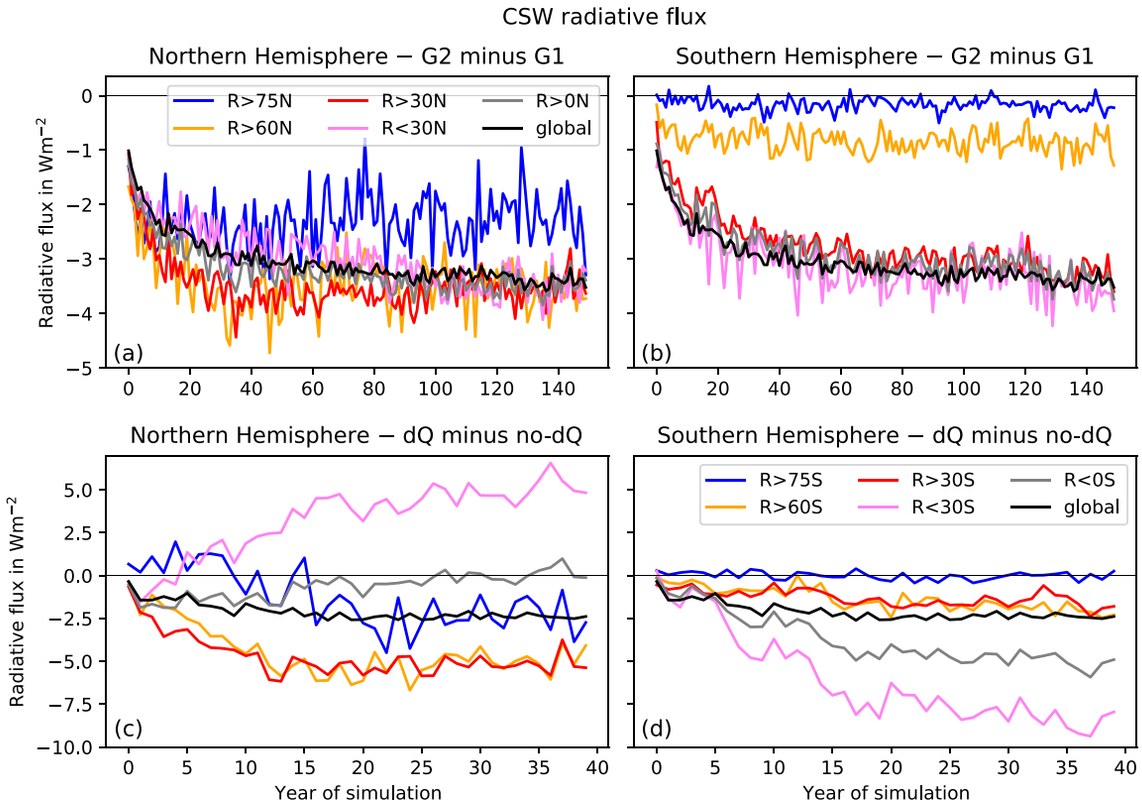

FIG. 10. As in Fig. 4, but for shortwave-cloud-feedback-induced top-of-the-atmosphere radiative-flux differences.

$R < 30$S. 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 with regard to 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 in the online supplemental material), although the latter is weaker. In comparing the change of Hadley circulation separately for G1 and G2 (Figs. S19a,b) as well as dQ and no-dQ (Figs. 11a,b) 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

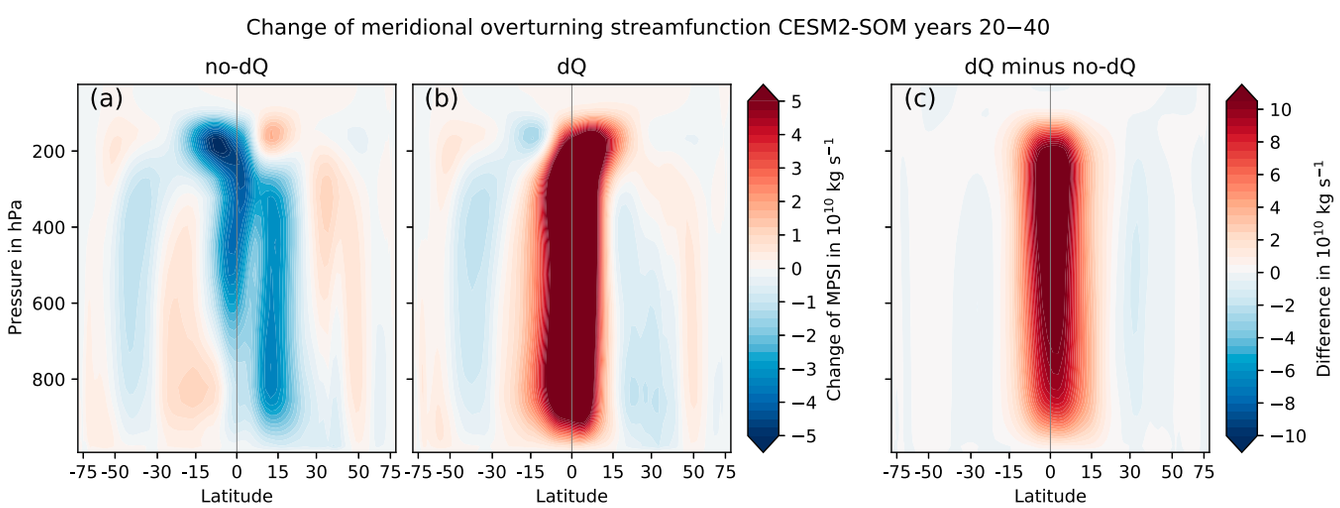

FIG. 11. Change of the atmospheric meridional overturning streamfunction for (a) no-dQ and (b) dQ, and (c) their difference. The streamfunctions are averaged over years 20–40 of the simulations.

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 in the online supplemental material). 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 in the online supplemental material). 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-yr 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. Discussion and conclusions

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-, water-vapor-, and cloud-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 in the online supplemental material). Notably though, the differences in surface temperature and 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 setup is the difference between their preindustrial states. As pointed out in section 4a and shown in Fig. 1a, G2 exhibits a significantly stronger preindustrial 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 preindustrial 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 preindustrial 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 relative to G1 shows only a slightly and statistically non-significantly smaller sea ice cover in the Labrador Sea (Fig. S23 in the online supplemental material). Recently, Lin et al. (2023) found that the base-state stratification in the Labrador Sea influences the AMOC decline. A stronger AMOC causes

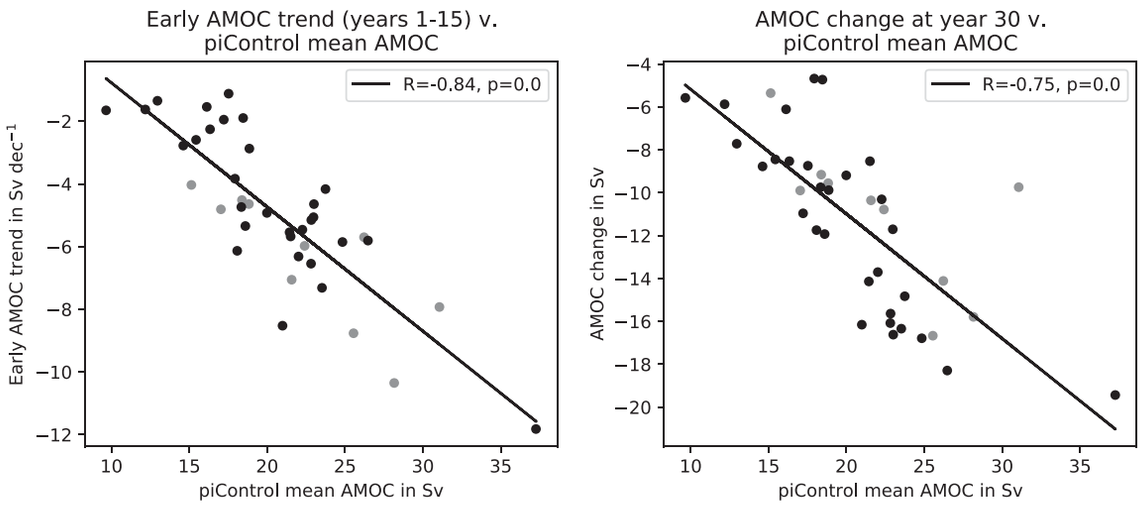

FIG. 12. Preindustrial control mean AMOC vs (left) AMOC trend over years 1 to 15 and (right) change of AMOC averaged over years 28–32. Members of CMIP5 and CMIP6 are depicted in gray and black, respectively.

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 farther 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 Seas, DWF can move farther 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 panel) in the online supplemental material] but the lapse rate feedback change exhibits considerable correlation with the late AMOC trend [$R = 0.5$, $p = 0.001$; Fig. S24 (right panel); 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 preindustrial AMOC with late AMOC trend are weak ($R = -0.1$ and 0.08, respectively; see Fig. S25 in the online supplemental material).

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 thank Patrick Stoll for helpful discussions and comments. We acknowledge the World Climate Research Programme (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 are 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 (https://esgf-node.llnl.gov/search/cmip6/) archives. 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 of the information needed to replicate the simulations; we used model version 2.1.3, freely available online (https://www.cesm.ucar.edu/models/cesm2). Pre- and postprocessing code for the CMIP and CESM2-SOM data are available at zenodo (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 online (https://doi.org/10.5281/zenodo.7937804).

APPENDIX

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 Table A1; also see Table S3 in the online supplemental material) 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

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

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

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

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

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

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 the potential temperature (with the ocean surface as a reference level). Assuming that the streamfunction vanishes at 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 the 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 in the online supplemental material 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 W m^{-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 2b) is about 25 W m^{-2}.

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. *London Edinburgh Dublin Philos. Mag. J. Sci.*, **5**, 237–276, https://doi.org/10.1080/14786449608620846.

Bellomo, K., M. Angeloni, S. Corti, and J. 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*, Vol. 10, Academic Press, 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*, **29**, 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 sensitivity 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. Rep. of an ad hoc study group on carbon dioxide and climate, Woods Hole, MA, 18 pp., https://geosci.uchicago.edu/~archer/warming_papers/charney.1979.report.pdf.

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

——, 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). *J. Adv. Model. Earth Syst.*, **12**, e2019MS001916, https://doi.org/10.1029/2019MS001916.

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.

——, 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.

Eiselt, K.-U., and R. G. Graversen, 2022: Change in climate sensitivity and its dependence on lapse-rate feedback in 4 × 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., 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.

——, 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.

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.

——, and Coauthors, 2005: A model intercomparison of changes in the Atlantic thermohaline circulation in response to increasing atmospheric CO_2 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, 116 pp.

IPCC, 2021: *Climate Change 2021: The Physical Science Basis*. Cambridge University Press, 2409 pp., https://report.ipcc.ch/ar6/wg1/IPCC_AR6_WGI_FullReport.pdf.

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.

——, 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. *J. Adv. Model. Earth Syst.*, **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., M. S. Wigmosta, H. Wu, M. Huang, Y. Ke, A. M. Coleman, and L. R. Leung, 2013: A physically based runoff routing model for land surface and Earth system models. *J. Hydrometeor.*, **14**, 808–828, 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.

——, ——, 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.

——, B. E. J. Rose, and Y.-T. Hwang, 2023: Mean state AMOC affects AMOC weakening through subsurface warming in the Labrador Sea. *J. Climate*, **36**, 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 atmosphere 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.

——, and ——, 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. Climate 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. *J. Adv. Model. Earth Syst.*, **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.

——, 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., and Coauthors, 2010: The Parallel Ocean Program (POP) reference manual: Ocean component of the Community Climate System Model (CCSM) and Community Earth System Model (CESM). LANL Tech. Rep. LAUR-10-01853, 141 pp., https://opensky.ucar.edu/islandora/object/manuscripts%3A825/datastream/PDF/download/The_Parallel_Ocean_Program__POP__reference_manual__Ocean_component_of_the_Community_Climate_System_Model__CCSM_.citation.

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.

——, ——, 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. J., S. C. Sherwood, S. Bony, and M. J. Webb, 2016: Prospects for narrowing bounds on Earth's equilibrium 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., 33 pp., 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. *J. Adv. Model. Earth Syst.*, **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.

——, 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. G., A. R. 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.

——, ——, and ——, 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.

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.

——, 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.

——, 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.

8

Paper III

Eiselt, K.-U., and R. G. Graversen (2023b): On the impact of net-zero forcing Q-flux changes. *Submitted to Climate Dynamics*.

On the Impact of Net-Zero Forcing Q-flux Change

Kai-Uwe Eiselt[1*] and Rune Grand Graversen[1,2]

[1]Department of Physics and Technology, University of Tromsø, Tromsø, Norway.
[2]Norwegian Meteorological Institute, Tromsø, Norway.

*Corresponding author(s). E-mail(s): kai-uwe.eiselt@uit.no;
Contributing authors: rune.graversen@uit.no;

Abstract

Numerical climate model simulations suggest that global warming is enhanced or hampered by the *spatial pattern* of the warming itself. This phenomenon is known as the "pattern effect" and has in recent years become the most promising explanation for the change over time of climate sensitivity in climate models. Under historical global warming, different patterns of surface-temperature change have emerged, notably a yet unexplained cooling in the Southern Ocean and the East Pacific. Historical climate model simulations notoriously fail to reproduce this cooling, which may contribute to the deviation of the simulated global-mean warming from the observed record. Here we qualitatively investigate the potential impact of historical and other surface-temperature pattern changes by changing the ocean heat transport convergence (Q-flux) in a slab-ocean model. The Q-flux changes are always implemented such that in the global mean they impose no net forcing. Consistent with earlier studies we find that the impact of a negative Q-flux change in the Southern Ocean has a stronger effect than in other regions because of a feedback loop between sea-surface temperatures (SSTs) and clouds in the Southern Ocean and the stably stratified regions in the tropics. The SST-cloud feedback loop facilitates the expansion of the Antarctic sea ice, indeed taking the model into a Snowball-Earth state. The intensity of this effect is found to be model dependent, especially due to differences in the cloud parametrisation. In experiments with deactivated sea ice the impact of the negative Q-flux change is much weaker.

Keywords: Climate sensitivity, Feedback, Pattern effect, Climate modeling, Ocean heat transport

1 Introduction

Global warming may be enhanced or hampered due to a change in warming patterns between regions. This process is known as the "pattern effect" [1] and was discovered in numerical global climate model (GCM) simulations [2, 3]. It has in recent years garnered prominence in climate change research. The mechanism of the pattern effect works as follows: Surface warming patterns evolve over time to favour or disfavour regions of different atmospheric stability, giving rise to feedback processes that enhance or hamper cooling. Thus, the change in local warming patterns induces the global "cooling efficiency" to be reduced or increased over time [e.g., 2–10]. In most GCMs a CO_2-induced surface warming shifts over time to regions which are characterised by relatively strong atmospheric stability, inhibiting cooling due to lapse-rate and cloud feedbacks [e.g., 3, 4].

The efficiency with which the Earth is cooling in response to a forcing-induced warming is called the climate feedback. Employing the forcing

F (e.g., due to an increase in the CO_2 concentration) and the climate feedback parameter λ, the climate system is often represented by the simple linear relation:

$$\Delta N = F + \lambda \Delta T_s, \tag{1}$$

where ΔN is the change in the top-of-the-atmosphere (TOA) radiative balance, and ΔT_s is the change of the surface-air temperature (SAT). This framework is typically used to derive the so-called equilibrium climate sensitivity (ECS) from global climate models (GCMs), which is possibly the most widely used single-number metric to describe the response of the climate system to a forcing. The ECS is usually defined as the global-mean SAT (GSAT) response to a doubling of the CO_2 concentration in the atmosphere and may be derived from Eq. 1 by assuming a new equilibrium condition ($\Delta N = 0$):

$$\mathrm{ECS} = -\frac{F_{2x}}{\lambda}, \tag{2}$$

where F_{2x} corresponds to the forcing due to the doubled CO_2 concentration.

In the framework defined by Eq. 1, the changing global cooling efficiency over time implies a changing climate feedback parameter and thus, according to Eq. 2, a changing climate sensitivity over time. This time-dependence of climate sensitivity was discovered long a ago [11–14], but due to the lack of coordinated GCM experiments different conjectures were put forward and no general theory was formulated [see e.g., 15, for a brief literature overview]. Over the last ten years, from investigation of the coordinated experiments of the Coupled Model Intercomparison Project (CMIP) phases 5 [16] and 6 [17] as well as targeted individual model studies and comparisons with observations, the pattern effect finally emerged as a reasonable explanation for a important part of the time dependence of climate sensitivity [2–5, 7, 8, 10]

Several specific geographical regions have been highlighted in the literature as either exhibiting abnormal historical warming changes, or as having a significant impact on climate feedback. These regions include the Indo-Pacific Warm Pool (IPWP), which has warmed comparatively strongly over the historical period and has been found to significantly influence climate feedback [4, 6–8, 10]. This is because the IPWP exhibits strong convective activity, meaning that a surface warming is quickly distributed to higher atmospheric layers. Here the warming pattern is readily spread horizontally to over large areas and the associated internal energy is efficiently radiated into space. Thus, a relatively strong warming in the IPWP implies a relatively strong negative climate feedback [10].

Another region is the Eastern Pacific (EP), which in spite of global warming has exhibited a cooling trend over the last forty years [e.g., 18]. Some studies also indicate this latter region's possible importance for climate feedback changes, especially in more recent climate models, mainly due to its effect on short-wave cloud feedback from low-cloud-cover changes [4, 5, 9]. Clouds have both warming and cooling impacts on the climate: Warming because they impede long-wave radiation from escaping to space and cooling because they reflect incoming solar radiation. Since low clouds are relatively warm, their impact on the long-wave radiation and thus their warming effect is small. Conversely, low clouds have a strong cooling effect because they are optically dense, meaning they reflect a large fraction of the solar radiation back to space [e.g., 19]. Tropical marine low cloud cover has been found to be strongly controlled by sea-surface temperature (SST) and atmospheric stability [e.g., 20]. Hence, the surface warming in the EP may be important in driving the cloud feedback. The short-wave cloud feedback tends to dominate the uncertainty in the total feedback across GCMs, also in the latest generation [e.g., 21], implying the EP as potentially important when it comes to climate feedback change. Notably, it is unclear why despite global warming the EP has cooled. Competing hypotheses exist such as that the cooling is a result of internal variability but it has also been proposed that the cooling is a forced response to CO_2-concentrations changes [22, 23].

An additional region is the Southern Ocean, which, in parts, has exhibited a similar cooling trend as the EP over the same period and has been shown to have a potentially large impact on climate feedback and sensitivity through local and teleconnection effects. That is, similar to the EP, SSTs and stability influence the local cloud

feedback, but it has also been found that a warming/cooling in the Southern Ocean affects the EP, influencing the cloud feedback there through SST and stability changes, as described above [24–27]. As for the EP, it is unclear why the Southern Ocean has historically cooled. Suggested explanations include internal variability [28] and freshwater input from Antarctic ice-sheet melt [29]. Notably, due to the teleconnection and feedbacks described in [25], [26], and [27], the EP cooling is potentially driven by the Southern Ocean cooling.

Finally, the North Atlantic is the focus of many studies as it is strongly influenced by the Atlantic meridional overturning circulation (AMOC), impacting the local surface warming [30–33]. Indeed, a lack of warming in the North Atlantic (the so-called North Atlantic Warming Hole, NAWH) is found in many climate models in CO_2-forcing experiments and is thought to be connected to a decline of the AMOC [e.g., 32]. Several studies indicate the influence of a decline of the AMOC on climate sensitivity, mainly due to impacts on sea ice, surface-albedo, and lapse-rate feedback [30, 31, 33].

To test the impact, of the warming/cooling patterns we perform slab-ocean model (SOM) experiments where energy is redistributed horizontally across the Earth's surface by changing the ocean heat-flux convergence (the so-called Q-flux). The experiments are motivated by the observed changes in surface-warming patterns described above as well as by projected changes under CO_2-forcing. We repeat the AMOC-mimicking experiment described by [33], where positive and negative Q-flux changes are implemented in the tropical and the North Atlantic, respectively. This experiment is based on an order-of-magnitude estimation of the Q-flux change induced by differences in AMOC change between fully-coupled abrupt CO_2-quadrupling experiments. However, in contrast to [33], the CO_2 remains at preindustrial levels (284.7 ppm), to isolate the effect of the Q-flux change. Motivated by the observed historical cooling in the EP and the relative warming in the IPWP, we perform an experiment with negative and positive Q-flux change in these regions, respectively. Similarly, in order to investigate the historical cooling in the Southern Ocean, experiments are run with negative Q-flux change in parts of and the whole Southern Ocean and positive Q-flux change in the parts of or the whole tropics. In contrast to the AMOC-mimicking experiment no quantitative analysis is performed to relate the Q-flux changes to results from fully-coupled model experiments for these regions. However, in order to maintain comparability, the Q-flux changes integrated over the individual regions are equivalent to those in the AMOC-mimicking experiment. We note that the cooling induced by the negative Q-flux changes in the EP and the Southern Ocean is much stronger than is observed historically. To investigate the impact of the Q-flux change magnitude we run additional experiments with Q-flux changes in the Southern Ocean that are similar to historical changes of Southern Ocean heat uptake [e.g. 34]. Importantly, in order to not to introduce a global net forcing from the Q-flux changes, the integrated net Q-flux change in all experiments is zero.

2 Models and experiments

2.1 CESM and the slab-ocean model

The Community Earth System Model version 2.1.3 (CESM2; [35]) in the slab-ocean model (SOM) configuration is applied. A SOM is similar to a fully-coupled GCM in that it couples several dynamical model components of the climate system. However, the difference is that in a SOM the dynamical ocean component is replaced by an isothermal mixed-layer ocean with a spatially varying mixed-layer depth (MLD) and a seasonally and spatially varying lateral ocean heat transport, called Q-flux. The Q-flux is derived from a fully-coupled GCM control simulation under equilibrium conditions from the MLD, the SSTs, and the ocean heat uptake of the dynamical ocean mode component [36].

In a SOM, the SST tendency $\frac{\partial \mathrm{SST}}{\partial t}$ is determined as follows [36]:

$$\frac{\partial \mathrm{SST}}{\partial t} = \frac{F_{net} + Q_{flx}}{\rho_0 c_p h}, \tag{3}$$

where F_{net} comprises the fluxes from the atmosphere to the ocean and/or sea ice, Q_{flx} is the ocean heat transport convergence (hereafter Q-flux), ρ_0 is the density of seawater, c_p the ocean heat capacity, and h is the MLD. Note that F_{net} is determined dynamically. Note that F_{net} is determined dynamically, while the Q-flux is prescribed.

Hence, SST patterns may be changed in a SOM by changing the Q-flux.

To test the sensitivity of our results to model version, we additionally employ an earlier version of the same model [CESM1; 37], as well as a version of CESM2 with a changed cloud parametrisation [CESM2-Z22; 38]. The cloud parametrisation of CESM2-Z22 was explicitly calibrated to exhibit greater skill in simulating the Last Glacial Maximum (LGM) climate than CESM2. Notably, also CESM1 simulates the LGM climate more realistically than does CESM2 [39]. The differences in cloud parametrisations have been implicated as the main reason for the differences in ECS between the three model versions [CESM2: 5-6 K; CESM1: 3-4 K, CESM2-Z22: ∼4 K; 38, 40–42]. For detailed model descriptions as well as a discussion of the differences between models see Text S1 in the online supplemental material.

2.2 Experiment design

Building on [33], we conduct several experiments with a Q-flux change in a particular region, always compensated by an opposite-sign Q-flux change in another region to ensure net-zero global forcing. Hence, the imposed climate response is solely due to regional energy redistribution and not to a global forcing. The CO_2 concentration is left at the pre-industrial level, in order to isolate the effect of the pattern changes. Our three main experiments are listed in the following. See Table 1 for a summary of all experiments conducted for this study.

dQ-AMOC Here, negative and positive Q-flux changes are implemented in the North Atlantic (50-80°N, 285-25°W) and the tropical Atlantic (15°S-15°N, 285-25°W), respectively (Fig. S1 in the online supplemental material). This experiment is equivalent to the "dQ" experiment in [33], although without 4xCO_2. The purpose of the experiment is to mimic the decline of the AMOC, as simulated by several CMIP5 and CMIP6 members in response to 4xCO_2 (see the appendix in [33] for the derivation of the Q-flux change induced by the AMOC change).

dQ-EP-WP A negative Q-flux change is implemented in the EP (0-30°S, 255-280°W) and the compensating positive Q-flux change in the IPWP (15°S-15°N, 60-180°W; Fig. S2). The experiment may be thought of as mimicking the historical cooling of the EP and the comparatively fast warming of the IPWP, although the SST changes induced by the Q-flux change are much larger than those historically observed.

dQ-SA-TA Several experiments have been conducted with negative Q-flux change in parts of or the whole Southern Ocean and a compensating positive Q-flux change in parts of or the whole tropics. These experiments may be thought of as mimicking the historical cooling and/or increased ocean heat uptake of the Southern Ocean [e.g., 34]. Here, we focus on the example of an experiment that is similar to dQ-AMOC, only "flipped" at the equator (dQ-SA-TA) with the region in the South Atlantic being 50-70°S and 305-30°W, and in the tropical Atlantic 15°S-15°N and 285-25°W (Fig. S3). As in dQ-EP-WP, the local cooling induced by the Q-flux change is much larger than the historical trends in the Southern Ocean. Further experiments with smaller and potentially more realistic Q-flux changes are conducted (Table 1).

Further experiments To test the linearity of the response to the pattern and CO_2-concentration changes, the three main experiments described above are rerun with quadrupled CO_2-concentration. In addition, a simulation is performed where only CO_2 is quadrupled and no Q-flux change is implemented. To test the potential impact of sea ice on the linearity of the response, the dQ-SA-TA experiment is rerun again with deactivated sea ice.

Many studies have the weakness of investigating the impact of Q-flux changes in only one model [e.g., 24, 26, 33, 43–46]. Our focus here is also on one model (CESM2), but to allow for at least some model diversity we additionally perform the main experiments with a previous version of the same model (CESM1). In addition, to investigate the potential impact of the cloud parametrisation on the pattern effect, we perform the main experiments also with CESM2-Z22 which is equivalent to CESM2, except for adjustments to the cloud microphysics parametrisation scheme (see section 2.1 and Text S1 in the online supplemental material).

Table 1 Summary of Q-flux change experiments. The columns with "Snowball" only concern CESM2.

name	negative δQ region	positive δQ region	δQ in Wm^{-2}	Snowball 1xCO_2	Snowball 4xCO_2
dQ-AMOC	50-80°N; 285-25°W	15°S-15°N; 285-25°W	-51, 26	no	no
dQ-EP-WP	0-30°S; 255-280°W	15°S-15°N; 60-180°W	-50, 11	yes	no
dQ-SA-TA	50-70°S; 305-30°W	15°S-15°N; 285-25°W	-49, 26	yes	yes
dQ-SEP-TA	50-70°S; 200-280°W	15°S-15°N; 285-25°W	-51, 26	yes	yes
dQ-SEP-EP	50-70°S; 200-280°W	0-30°S; 250-280°W	-51, 49	yes	no
dQ-SO-T-3Wm^{-2}	50-70°S	15°S-15°N	-3, 1.2	yes	(N/A)
dQ-SO-T-2Wm^{-2}	50-70°S	15°S-15°N	-2, 0.8	no	(N/A)

3 Methods

3.1 Radiative kernel method

The radiative kernel method [47, 48] can be used to decompose the climate model response into individual physical feedbacks and is based on two main assumptions: (1) The radiative response ΔN of the climate system at TOA can be expressed as a function of independent climate variables:

$$\Delta N = F + (\lambda_{Pl} + \lambda_{LR} + \lambda_{SA} + \lambda_{WV} + \lambda_C)\Delta T_s, \quad (4)$$

where λ_{Pl}, λ_{LR}, λ_{SA}, λ_{WV}, and λ_C are the feedback parameters due to changes in Planck radiation, temperature lapse rate (LR), surface albedo (SA), water-vapour (WV) mixing ratio, and cloud properties, respectively. (2) The dependence of the TOA radiative flux on a change in one of the climate variables is linear. Radiative kernels are derived by changing one of these climate variables in the radiation code of a climate model. Since radiation codes are well tested and similar across models [47], one set of radiative kernels can be used to derived the radiative fluxes from different climate models and experiments. Under the assumptions introduced above the kernel can then be multiplied on the change of the variable in question in an experiment with another climate model, hereby deriving the TOA radiative flux change due to the change in this variable. The feedback parameter λ_i for a climate variable i may be obtained by regressing the kernel-derived TOA radiative flux change due to the change in i on the GSAT change [14, 49]. A more detailed description of the kernel method is presented in Text S2 in the online supplemental material.

Here we use the set of radiative kernels provided by [48]. For details on the choice of radiative kernels see [15].

4 Results

In the following we discuss the results of the Q-flux change experiments described in section 2.2. We mainly focus on CESM2, but in section 4.3 results are compared across CESM1, CESM2, and CESM2-Z22.

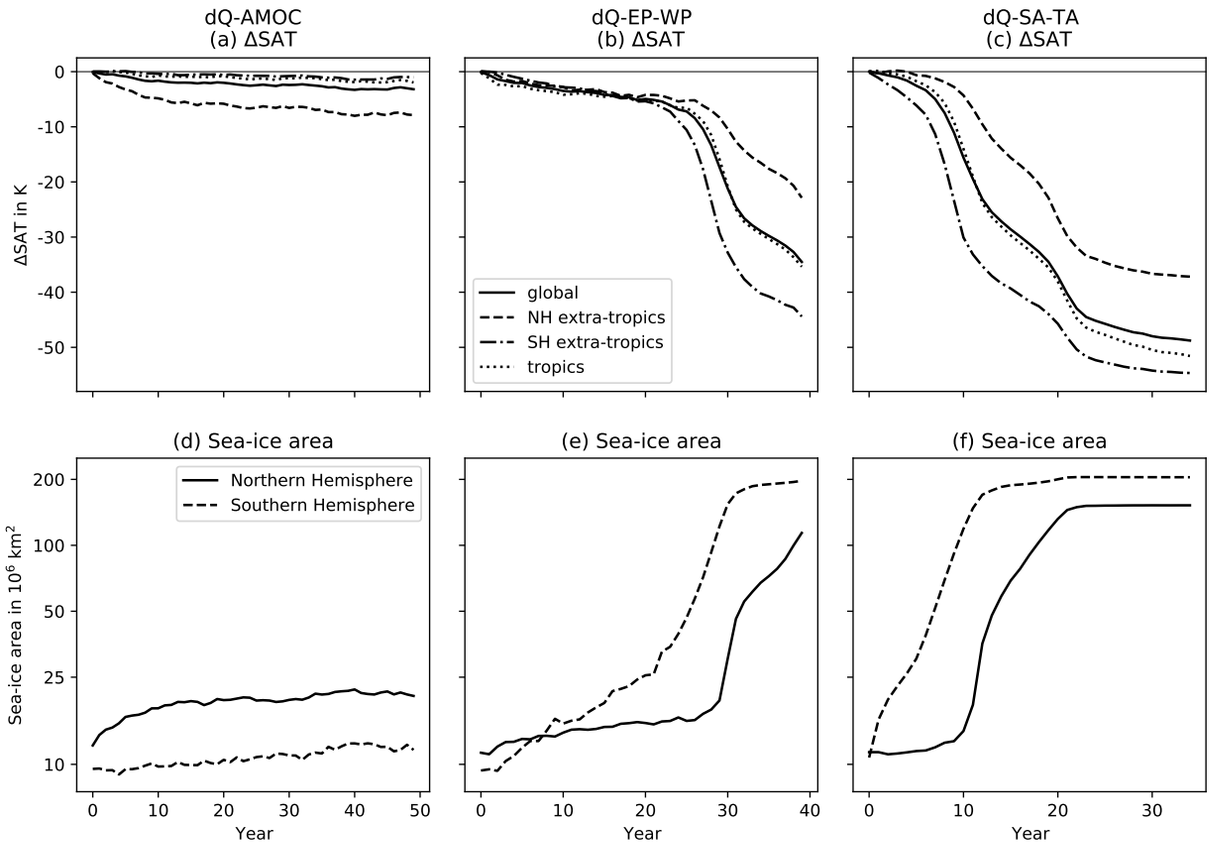

Fig. 1 Surface-air temperature response (ΔSAT) and sea-ice area for (a,d) dQ-AMOC, (b,e) dQ-EP-WP, and (c,f) dQ-SA-TA. For the ΔSAT are shown (solid) global means, (dotted) tropical means (30°N-30°S), and extra-tropical means over the (dashed) Northern Hemisphere (>30°N) and (dash-dotted) Southern Hemisphere (>30°S). Northern Hemisphere and Southern Hemisphere areas are depicted as solid and dashed lines, respectively in panels (d-f). The sea-ice area is shown on a logarithmic scale.

4.1 Response to different Q-flux pattern changes

In this section we compare the response of CESM2 to the three different net-zero-forcing Q-flux change experiments dQ-AMOC, dQ-EP-WP, and dQ-SA-TA. The comparison is conducted mainly in terms of the SAT change and individual feedbacks.

dQ-AMOC The mimicked change of the AMOC leads to a decline of the GSAT of about 3 K after 50 years of simulation (Fig. 1a). The effect is strongest in the NH extra-tropics (-8 K), while there is much less of an impact in the tropics (-2 K) and the SH extra-tropics (-1 K). Figure 2a shows the individual climate feedbacks and their evolution over time. Initially, the LR and cloud feedback are largest, followed by the SA and WV feedbacks. However, because the impact of the Q-flux change in this experiment is comparatively weak on the global level, especially the cloud feedback exhibits strong uncertainty in the first few years. In a zonal-mean decomposition (Figs. S4 and S5 in the online supplemental material), it is revealed that the LR and SA effects concentrate in the NH polar region (>50°N), while the effect of the cloud feedback is spread over most of the NH extra-tropics, with some additional effect in the deep tropics, due to partly compensating long-wave and short-wave cloud effects. The positive LR and SA feedbacks appear to be related to the Arctic sea ice, which quickly expands over the first decade of the simulation (Fig. 1d). The increased sea-ice area leads to more reflection of sunlight, causing a negative SA radiative flux, inducing a

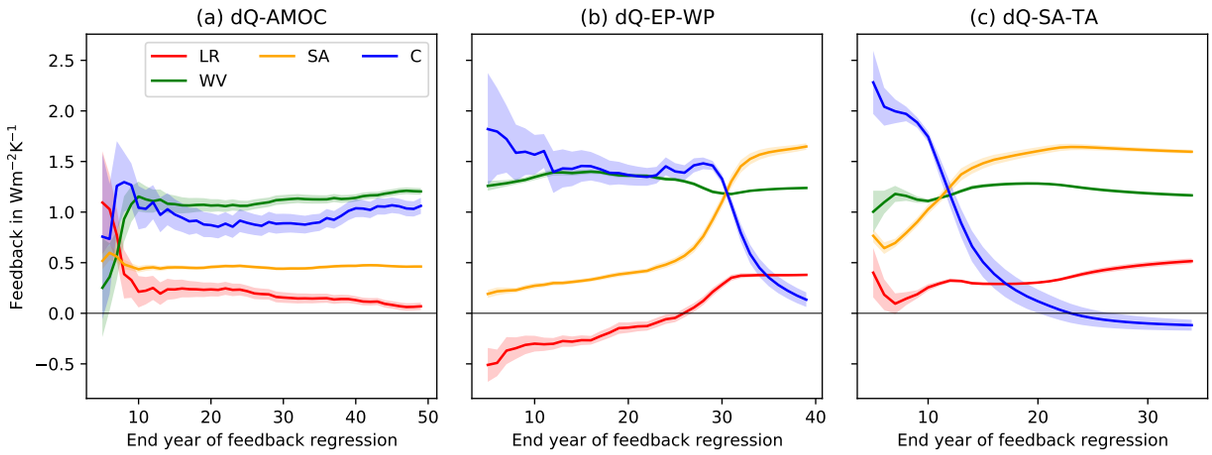

Fig. 2 Feedback time series in (a) dQ-AMOC, (b) dQ-EP-WP, and (c) dQ-SA-TA as simulated by CESM2. Shown are the (red) lapse-rate, (green) water-vapour (WV), (orange) surface-albedo (SA), and (blue) cloud (C) feedbacks. The feedbacks are calculated by regressing the radiative kernel derived top-of-the-atmosphere (TOA) radiative fluxes on the global-mean surface-air temperature [14, 49]. The starting year of the regression is always year 1 while the end year is continuously incremented from year 5 to the final year. The radiative kernels provided by [48] were used. Note the different x-axis scales.

positive (i.e., cooling) feedback. Moreover, the sea ice insulates the atmosphere from the ocean. In the polar regions the ocean acts as a heat source for the atmosphere. Thus, an increase in sea-ice cover induces a cooling influence at the surface, increasing atmospheric stability, hereby leading to a positive LR feedback. A positive LR feedback also results without the generation of new sea ice because of the relative surface cooling from the negative North Atlantic Q-flux change. With the saturation of the growth of the Arctic sea ice (Fig. 1d), the LR and SA feedbacks decline (Fig. 2a).

dQ-EP-WP The response of the GSAT in dQ-EP-WP is much stronger than in dQ-AMOC. The GSAT decreases by about 7 K at year 25 of the simulation, after which the decline rapidly accelerates, taking the model into a Snowball Earth state (Fig. 1b,e). The main impact of the Q-flux change is on the cloud and WV feedbacks, with a minor contribution from the SA feedback and a negative contribution from the LR feedback (Fig. 2b). Note that dQ-EP-WP is different from dQ-AMOC and dQ-SA-TA (see below) since it exhibits a negative LR feedback while the other two experiments exhibit a positive LR feedback (Fig. 2). This is associated with the spatial structure of the LR feedback, being generally negative at low and positive at high latitudes, due to convection at low and strong atmospheric stability at high latitudes [50]. In dQ-EP-WP the impact of the Q-flux change is a strong surface cooling of the tropics, affecting the convective regions and cooling the higher atmospheric layers. This leads to a reduction of the outgoing long-wave radiation, hereby inducing a globally negative (i.e., warming) LR feedback. However, due to the persistent general cooling the sea ice grows continually (Fig. 1e), amplifying the surface cooling at the high latitudes, which, due to the stability in these regions is not readily vertically distributed. Hence, the surface cools more strongly than the upper atmospheric layers, eventually giving rise to a globally positive LR feedback. In contrast to dQ-EP-WP, in the other experiments the sea-ice effect is active immediately from the start of the simulation, because the Q-flux change impact is concentrated at the high latitudes. Thus, the LR feedback is positive throughout the simulation in these experiments. The differences between the experiments highlight the importance of the latitudinal aspects of the pattern effect in influencing global climate feedback, especially regarding the LR feedback [see also 31, 33, 46].

The strongly positive cloud feedback is concentrated in the tropics (Figs. S4 and S5), and particularly in the East Pacific (EP; Fig. S6). The EP is a relatively stably stratified region and the Q-flux-change-induced surface cooling in dQ-EP-WP further increases the stability. The SSTs and

the estimated inversion strength (EIS; a measure for atmospheric stability; [51]) are strong cloud controlling factors, especially influencing tropical low cloud [20]. Note again that clouds generally have two climate feedback effects: A cooling effect due to increased reflection of incoming solar radiation and a warming effect by blocking thermal radiation from exiting to space. Low clouds mainly cause the former effect as they are optically thick and thus exhibit a high albedo, while they are also relatively warm, implying a weak impact on the outgoing thermal radiation [e.g., 19]. The effect on short-wave radiation is especially strong for liquid clouds, since liquid cloud droplets are smaller and more densely packed than cloud ice crystals, causing a higher cloud optical depth and thus albedo [52]. Hence, the surface cooling and the accompanying increase in EIS in the EP induce increased local low-cloud cover, reflecting more of the incoming sunlight (Figs. S6 and S7), hereby amplifying the local surface cooling. The intense surface cooling in the EP is advected westward across the central tropical Pacific (see Fig. S8b for the climatological winds), affecting EIS, low-cloud cover, and short-wave cloud radiative flux (CRF_{SW}) to the west of the Q-flux anomaly.

As compared to the EP, the positive Q-flux change implemented in the IPWP does not as strongly impact the SSTs there. Indeed, because of the strong convective characteristics of this region, heat is quickly spread vertically. In the free troposphere, it is readily distributed horizontally, in fact potentially contributing to the increasing EIS in the EP [e.g., 5–8, 10]. Energy associated with the heating in the free troposphere is also radiated to space more readily than is energy at the surface, hereby contributing to the global cooling.

The continued cooling in the tropics also impacts other regions (Fig. 1b), and in the Southern Ocean the cooling induces an expansion of the Antarctic sea ice (Fig. 1e). Through the SA feedback, this contributes further to the global cooling (Fig. 2b), eventually leading to an acceleration of the sea-ice expansion, driving the transition to the Snowball-Earth state (Fig. 1e). This transition proceeds similarly in the dQ-SA-TA experiment and is described in more detail below.

dQ-SA-TA The strongest response to the Q-flux change is found in the dQ-SA-TA experiment. The initial decline of the GSAT is similar to the dQ-EP-WP experiment, with the strongest feedback again being due to clouds (Fig. 2c). However, the pattern of the cloud feedback is different from that in the dQ-EP-WP experiment, by being spread over the SH subtropics and middle latitudes (Figs. S4 and S5). In addition a much stronger initial SA feedback due to the Antarctic sea-ice growth is encountered (Figs. 1f and 2c), which is induced by the negative Q-flux change in the South Atlantic. As the Antarctic sea ice expands, the SA feedback increases, while the cloud feedback declines, as clouds dissolve over the expanding ice (Figs. 2c and 3).

Considering regional radiative fluxes, initially there is a weak but extensive negative CRF_{SW} over the Southern Indian Ocean as well as a stronger negative CRF_{SW} over the Southern subtropical Atlantic (Fig. 3a). This appears to be due to the strong surface cooling in the South Atlantic induced by the Q-flux change and amplified by local sea-ice growth, which is quickly advected via climatological winds along the African and Australian coasts over the subtropical Atlantic and Indian Ocean (Fig. S8b). We note that the Southeast Pacific is initially (i.e., in the first about 5 years) not as strongly impacted as the other mentioned regions and exhibits relatively less cooling and only a small negative CRF_{SW}. This may be associated with the Antarctic sea ice exhibiting its smallest northward extent in the Southeast Pacific until about year 7 (Fig. 3c). From year 6 and onward, both the tropical Atlantic and Indian Ocean show increasingly negative CRF_{SW}, and from year 8 a similar effect is seen in the EP. This is consistent with the northern boundary of the Antarctic sea ice being situated at a similar latitude (circa 45°S) across the whole Southern Ocean at year 8 (Fig. 3d). The process in the EP is similar to that described for the tropical Atlantic and Indian Ocean: The cooling of the South Atlantic due to the negative Q-flux change and local feedbacks (SA, LR, cloud) is advected around Antarctica by climatological winds (Fig. S8b). Successively, this leads to an expansion of the sea ice also in the Southeast Pacific, which, again facilitated by local feedbacks, accelerates the cooling there. Through climatological winds along the coast of South America (Fig. S8b), this cooling impacts the EP region, locally inducing lower SSTs, causing an increase of EP stability.

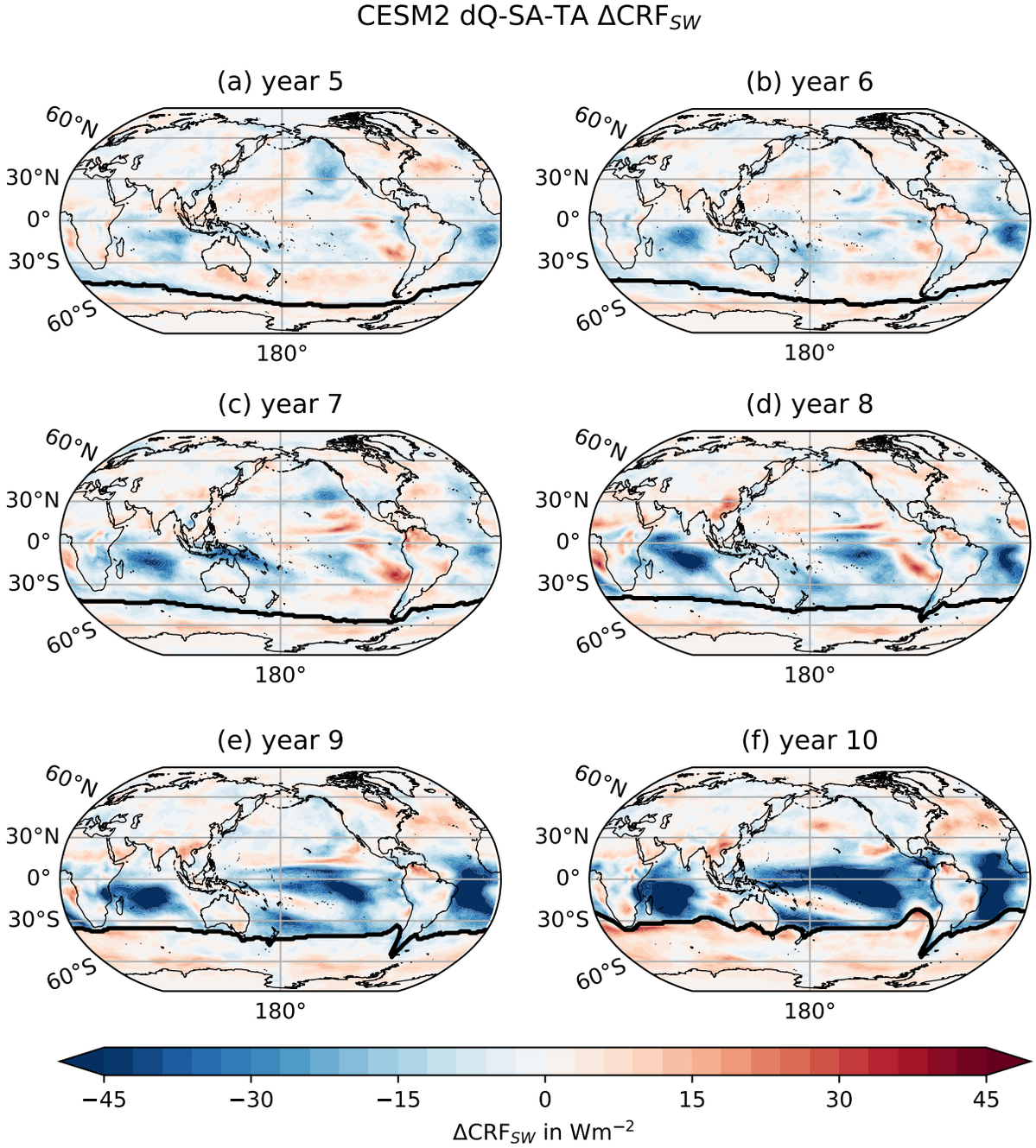

Fig. 3 Maps of annual means of the short-wave cloud radiative flux change ($\Delta\mathrm{CRF}_{SW}$) for years 5 to 10 of the dQ-SA-TA experiment as simulated by CESM2. The black line indicates the annual mean Antarctic sea-ice extent.

As mentioned above, tropical marine low clouds in stable regions are sensitive to SST and stability changes [e.g., 20, 51, 53]. The decrease of the SST and the increase of the stability lead to more low-cloud cover, reflecting the incoming solar radiation. A further SST cooling initiates a positive SST-cloud feedback loop: lower SSTs, more low cloud, less insolation, lower SSTs, more

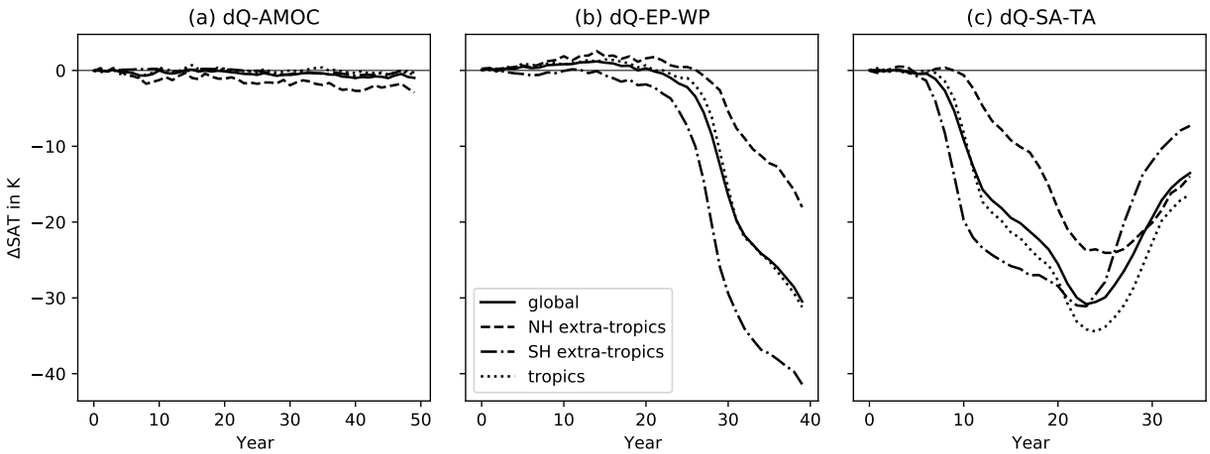

Fig. 4 Differences in surface-air temperature change (ΔSAT) between dQ+4x and dQ&4x (dQ+4x minus dQ&4x) for (a) dQ-AMOC, (b) dQ-EP-WP, and (c) dQ-SA-TA. Shown are (solid) global means, (dotted) tropical means (30°N-30°S), and extra-tropical means over the (dashed) Northern Hemisphere (>30°N) and (dash-dotted) Southern Hemisphere (>30°S). Note the different x-axis scales.

low cloud, etc. [25–27]. Over years 7 to 11 the tropical CRF_{SW} becomes increasingly negative, leading to the accelerated decline of the GSAT, and continuously causing the Antarctic sea ice to advance more equatorward (Fig. 3c-f). Finally, the cooled tropical regions together with the winds advecting along the west coasts of the continents facilitate the quick expansion of the Antarctic sea ice, leading to almost full coverage of the Southern Hemisphere by year 12 of the simulation (not shown).

As mentioned above, several dQ-SO-T-type experiments were conducted (see Table 1). The response of these experiments follows a similar trajectory as described for dQ-SA-TA. That is, implementing a negative Q-flux change in the Southeast Pacific with the positive Q-flux change being either in the tropical Atlantic (dQ-SEP-TA) or the tropical East Pacific (dQ-SEP-EP) causes an almost equivalent transition to a Snowball Earth as dQ-SA-TA (not shown). In another experiment where the negative Q-flux change is spread over the whole Southern Ocean and the positive Q-flux change over the whole tropics (dQ-SO-T-3Wm^{-2}), the response is linear and weak for over 30 years, after which an abrupt shift occurs, taking the model into a Snowball Earth. This shift appears to be driven by the SST-cloud feedback loop, especially in the East Pacific, as explained above (Figs. S9 and S10). Notably, in the latter experiment the Q-flux change in the individual regions is only about 1/4 of the Q-flux change in dQ-SA-TA. We report that in a further experiment (dQ-SO-T-2Wm^{-2}) with -2 Wm^{-2} Q-flux change over the whole Southern Ocean and a compensating positive change over the whole tropics, no Snowball Earth transition is instigated. Remarkably, it is generally only when the cooling SST-cloud feedback loop in the EP is initiated that the model rapidly intensifies the negative CRF_{SW} and accelerates the sea-ice expansion, facilitating the transition to a Snowball Earth state (see especially dQ-SO-T-3Wm^{-2}; Figs. S9 and S10).

4.2 (Non-)Linearity of responses to Q-flux change and CO_2-forcing

Three different experiments are employed to test the linearity of the response to the Q-flux change and the CO_2 forcing (which induces the Q-flux change at least for dQ-AMOC). These include the Q-flux change experiments (dQ), the CO_2 experiment (4xCO_2), and the combined Q-flux-change and CO_2-forcing experiment (dQ&4x). To investigate linearity, we add up the responses of the dQ and 4xCO_2 experiments (hereafter referred to as dQ+4x) and compare with the dQ&4x experiment. If the responses to dQ and 4xCO_2 add linearly, dQ&4x and dQ+4x should yield similar results. dQ+4x and dQ&4x are compared for

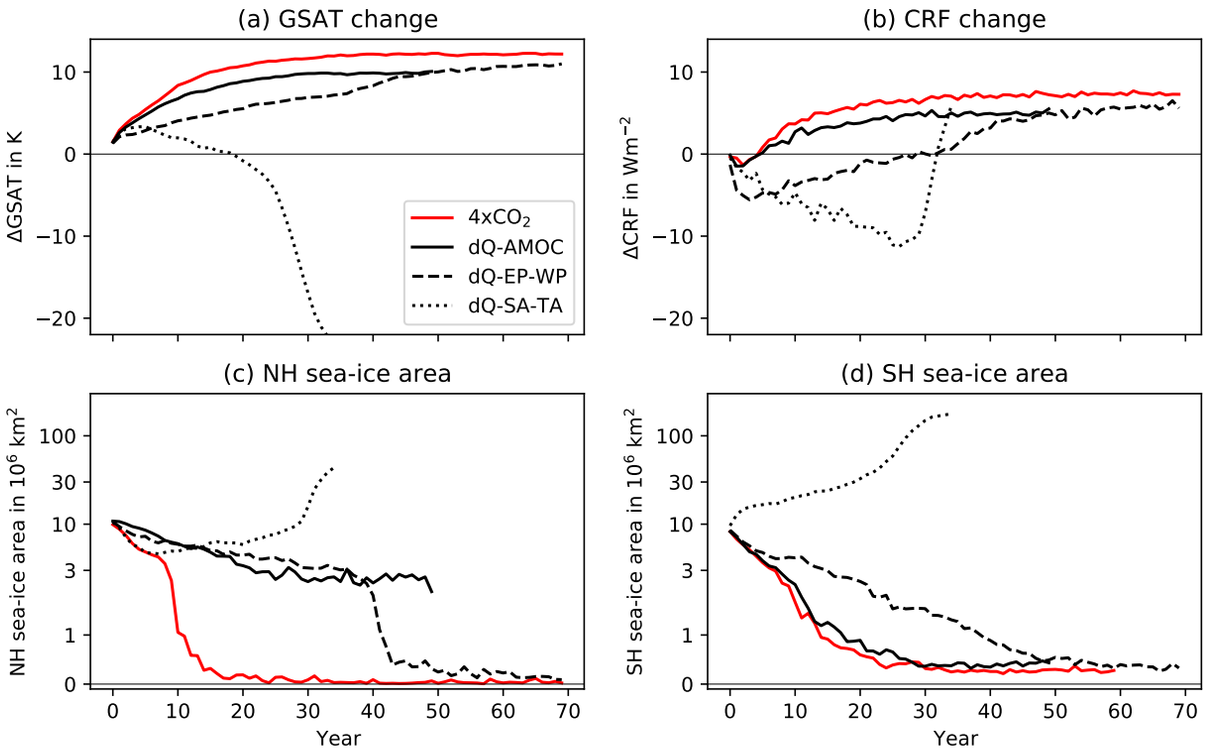

Fig. 5 Global means of (a) surface-air temperature change (ΔGSAT) and (b) cloud radiative flux (ΔCRF), as well as (c) Northern Hemisphere sea-ice area, and (d) Southern Hemisphere sea-ice area for the CESM2 experiments (solid red) 4xCO_2, (solid black) dQ-AMOC, (dashed) dQ-EP-WP, and (dotted) dQ-SA-TA. All dQ experiments were run with quadrupled CO_2-concentration. The sea-ice area is shown on a logarithmic scale.

our three main experimental set-ups dQ-AMOC, dQ-EP-WP, and dQ-SA-TA.

As shown in Fig. 4, the GSAT response in all three set-ups is initially linear, i.e., dQ+4x and dQ&4x exhibit only small differences of $|\Delta| < 1$ K. In dQ-AMOC this difference is largest in the Northern Hemisphere ($\sim$ -2.5 K), growing slightly throughout the 50-year simulation. This small non-linearity in the GSAT response appears to be driven by differences in sea-ice development that drive differences in SA and LR fluxes (Fig. S11a): In both the 4xCO_2 and the dQ&4x experiments the Arctic sea-ice cover quickly declines, remaining constant after about 15 and 25 years, respectively (Fig. 5c). Conversely, in the dQ-AMOC experiment at 1xCO_2, the Arctic sea ice quickly grows (Fig. 1d). The growth decelerates over time, but it continues at least until year 40. In dQ+4x the sea-ice growth in dQ-AMOC compensates the loss in 4xCO_2 more than in the combined dQ&4x experiment. Thus, in dQ&4x more sea ice is lost than in dQ+4x, engendering the above-mentioned SA and LR feedbacks, driving the deviation from a linear response to dQ-AMOC and 4xCO_2. A smaller but over time growing contribution to the non-linearity comes from the Antarctic sea ice: dQ&4x and 4xCO_2 exhibit almost equivalent Antarctic sea-ice decline, i.e., dQ-AMOC has little effect on Antarctic sea ice under concomitant 4xCO_2 (Fig. 5d). However, in dQ-AMOC with 1xCO_2, after some time the Antarctic sea ice starts slowly growing, triggering LR and SA feedbacks similar to the Arctic sea ice, thus contributing to the GSAT difference between dQ+4x and dQ&4x.

In contrast to dQ-AMOC, dQ-EP-WP and dQ-SA-TA quickly exhibit strongly non-linear responses (Fig. 4b,c). These are related to the Antarctic sea-ice expansion in dQ-EP-WP and dQ-SA-TA under 1xCO_2, which is driven by the cloud-SST feedback loop in the Southern Ocean and the tropics as explained in section 4.1. While

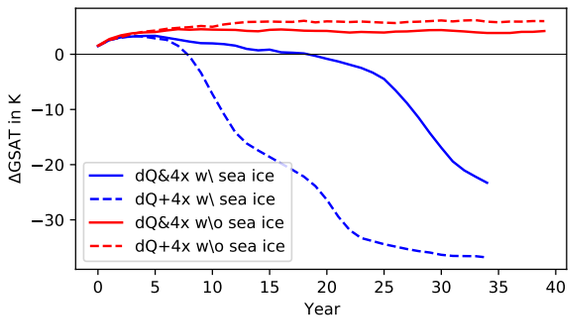

Fig. 6 Global-mean surface-air temperature change (ΔGSAT) in the dQ-SA-TA experiment (blue) with sea ice and (red) without sea ice for (solid) dQ&4x and (dashed) dQ+4x.

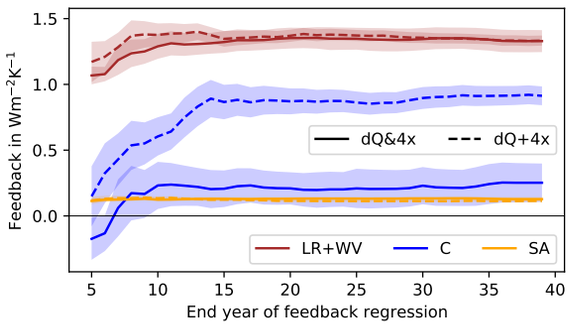

Fig. 7 As Fig. 2 but for dQ-SA-TA under (solid) dQ&4x and (dashed) dQ+4x. Note that here the lapse-rate and water-vapour feedbacks are summed (they are generally anti-correlated; [54]) to increase the clarity of the figure.

in the dQ-SA-TA set-up eventually the Antarctic sea ice expands and a Snowball Earth state is obtained also under 4xCO$_2$ forcing (Fig. 5), this is delayed compared to the 1xCO$_2$ simulation: In dQ&4x the rapid sea-ice expansion starts only after year 25, while under 1xCO$_2$ this occurs already around year 8 (Fig. 1c,f). Thus, in the dQ-SA-TA set-up in terms of GSAT the dQ+4x minus dQ&4x difference quickly becomes strongly negative, expressing the non-linearity of the response. However, after the Antarctic sea-ice expansion accelerates also in dQ&4x (Fig. 5d) the difference between dQ+4x and dQ&4x declines and starts approaching zero. Thus, while the *transient* response to dQ-SA-TA and 4xCO$_2$ is strongly non-linear, the *equilibrium* response may be linear, although the simulations conducted here are not long enough to confirm this.

dQ-EP-WP is different from dQ-SA-TA because under 4xCO$_2$ it does not enter a Snowball Earth state (Fig. 5). Hence, the difference in terms of GSAT between dQ+4x and dQ&4x becomes strongly negative (similar to dQ-SA-TA) when the Antarctic sea ice starts its accelerated expansion in dQ-EP-WP under 1xCO$_2$, but, unlike in dQ-SA-TA, the difference never recovers. Hence, in dQ-EP-WP *both* the transient and the equilibrium response to dQ and 4xCO$_2$ may be described as non-linear.

The difference of the response to dQ-EP-WP and dQ-SA-TA under 1xCO$_2$ and 4xCO$_2$ highlights the importance of the location of the Q-flux change and of the impact of the sea ice. Similar to dQ-AMOC, it is the effect of the high-latitude Q-flux change in dQ-SA-TA which leads to the maintenance of the sea-ice cover despite the quadrupled CO$_2$-concentration. Conversely, even though the negative Q-flux changes in dQ-EP-WP and dQ-SA-TA are of similar magnitude, in dQ-EP-WP they do not impact the sea ice *directly*. Under 4xCO$_2$ the impact of the CO$_2$-forcing on the sea ice, facilitated particularly by strongly positive extra-tropical CRF$_{SW}$ (Fig. S12), is thus stronger than the Q-flux impact, causing sea-ice retreat and hence additional warming SA and cloud radiative fluxes (Fig. S13b), prohibiting the transition to a Snowball Earth. In contrast, in dQ-SA-TA the Q-flux change directly impacts the Antarctic sea ice more than the CO$_2$-forcing, causing its local expansion and the accompanying cooling SA, LR, and cloud radiative fluxes (Fig. S13c). As explained above, the effect eventually spreads to the tropics, similar to the 1xCO$_2$ dQ-SA-TA experiment, causing the expansion of the Antarctic sea ice and the transition to the Snowball-Earth state.

To further investigate the effect of the sea ice, we rerun the dQ-SA-TA experiments (under 1x and 4xCO$_2$) and the 4xCO$_2$ experiment, but with no sea ice present. In this set-up, dQ+4x shows 2 K more warming than dQ&4x, thus exhibiting a much more linear response than the experiments with sea ice (Fig. 6). The difference is primarily caused by the cloud feedback (Fig. 7), especially over the Southern Ocean (Fig. S14) and the EP (not shown). The CO$_2$-forcing and the Q-flux change have competing impacts on the cloud radiative fluxes, especially over the Southern Ocean: The CO$_2$-forcing induces a positive

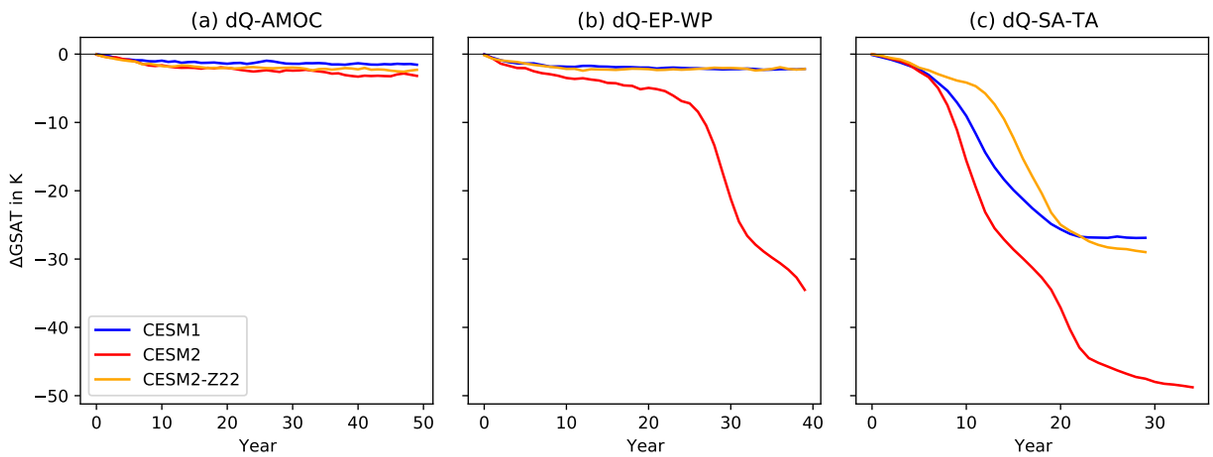

Fig. 8 Comparison of (blue) CESM1, (red) CESM2, and (orange) CESM2-Z22 in terms of global-mean surface-air temperature change (ΔGSAT) for cases (a) dQ-AMOC, (b) dQ-EP-WP, and (c) dQ-SA-TA. Note the different x-axis scales.

CRF, while the Q-flux change induces a negative CRF. A self-reinforcing SST-cloud feedback loop is initiated, warming in 4xCO_2 and cooling in dQ, as described in section 4.1. Without sea ice, a stronger feedback loop is induced in 4xCO_2 than in dQ, leading to a positive total CRF in dQ+4x. In dQ&4x the feedback loop is muted because of the compensating impact of the Q-flux change and the CO_2-forcing, implying a smaller CRF. Thus, the sum of the cloud effects in the separate dQ and 4xCO_2 experiments is larger than the effect in the combined experiment, giving rise to the difference in cloud feedback seen in Fig. 7.

4.3 Comparing different model versions

Here our aim is to investigate the dependence of the impact of Q-flux changes on model version. To this end, we rerun our three main experiments (see section 2.2) with CESM1, i.e., the previous version of CESM2, as well as with CESM2-Z22, which is structurally equivalent to CESM2 except for the cloud parametrisation [38].

The GSAT responses to dQ-AMOC and dQ-SA-TA are qualitatively similar in all three models, i.e., a moderate cooling and a transition to a Snowball Earth state, respectively (Fig. 8). Conversely, there is a large difference in dQ-EP-WP in CESM2 compared with CESM1 and CESM2-Z22. That is, CESM2 cools much quicker than the other two models, and transitions to a Snowball Earth state, whereas CESM1 and CESM2-Z22 only exhibit a moderate cooling, similar to their response in dQ-AMOC (see Fig. S15 for the sea-ice area).

Considering the radiative fluxes derived from the kernel method, the differences in GSAT response between the models appear to be explained by differences in the short-wave cloud response. CESM2 generally exhibits the strongest negative CRF_{SW} of the three models, especially in dQ-EP-WP and dQ-SA-TA (Fig. S16). Notably, the WV radiative flux is also strongly negative but the intensification of the CRF_{SW} precedes that of the WV flux, and we thus suggest that the WV flux may be viewed as a response to the temperature change which is mainly initiated by the CRF_{SW}.

The above-discussed differences in model response to the Q-flux changes may be related to several differences in model characteristics between the three model versions. First, it should be noted that the base-state Q-flux and the mixed-layer depth (MLD) in CESM1 are different from CESM2 and CESM2-Z22 (Fig. S17 and S18, respectively). Regions of large base-state Q-flux may be less impacted by the same Q-flux change as regions with small base-state Q-flux, since the Q-flux change would be *relatively* smaller. However, generally the annual-mean Q-flux in CESM1 and CESM2 appears similar in the regions of Q-flux change considered in this study (Fig. S17c).

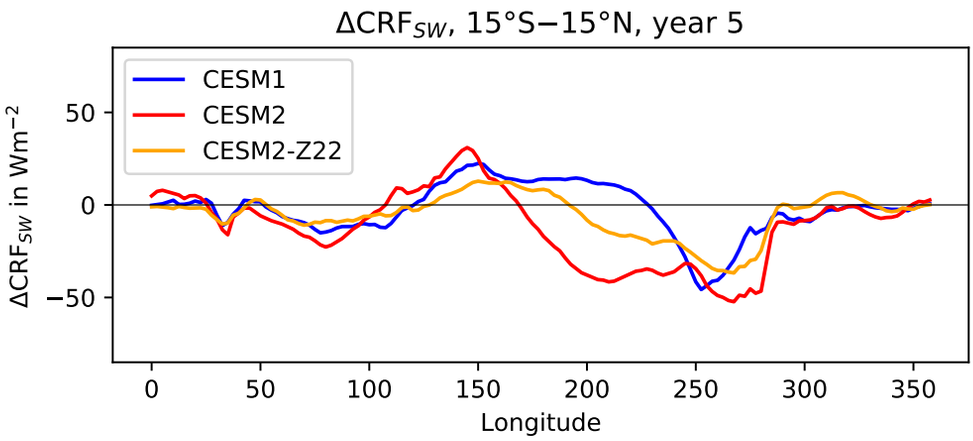

Fig. 9 Change of the short-wave cloud radiative flux (ΔCRF_{SW}) in the inner tropics (15°S-15°N) for (blue) CESM1, (red) CESM2, and (orange) CESM2-Z22 in year 5 of the respective dQ-EP-WP simulation.

The influence of the MLD on the impact of the Q-flux on SSTs is described by Eq. 3: The larger the MLD, the smaller the impact of a given Q-flux change, meaning that the MLD effectively acts as a kind of thermal inertia. Regarding dQ-AMOC, the area-weighted MLD in the North Atlantic region of negative Q-flux change is somewhat lower in CESM2 than in CESM1 (178 m compared to 201 m), potentially facilitating the faster cooling and regional sea-ice expansion in CESM2. Conversely, pertaining to dQ-SA-TA, in the South Atlantic as well as in the Southern Ocean in general the MLD is larger in CESM2 than in CESM1, which should decelerate the cooling and sea-ice response in CESM2. However, as seen in Figs. 8, respectively, CESM2 cools quicker and expands the Antarctic sea ice faster than CESM1. This indicates that other factors such as a stronger cloud feedback (Fig. S19) triggered by the Q-flux-induced surface cooling facilitate the sea-ice expansion in CESM2. Note that the base-state Q-flux and MLD are equivalent in CESM2 and CESM2-Z22.

CESM1 and CESM2 are also different in terms of base-state climatological wind patterns: The subtropical easterlies are slightly stronger in CESM2 than in CESM1 (Fig. S8c). This may facilitate the larger westward expansion of the cooling and the accompanying negative CRF_{SW} in CESM2 (Fig. 9 and cf. Figs. S6 and S20).

Differences between the models also exist in terms of base-state sea-ice area (Fig. S21). Notably, the base-state Antarctic sea-ice area in CESM2-Z22 is about 30% smaller than in CESM1 and CESM2. This may explain the slower response of CESM2-Z22 compared to the other models in dQ-SA-TA: More sea ice needs to build up before the SST-cloud feedback loop described in section 4.1, which accelerates the sea-ice expansion, is triggered. Thus, it takes longer for CESM2-Z22 to enter the Snowball Earth state than for CESM1 and CESM2. However, we confirm a weaker cloud effect in CESM2-Z22 than in CESM2 also in the absence of sea ice (Fig. S22).

Finally, the differences in cloud parametrisation cause differences in base-state cloud distribution (Fig. S23) as well as different cloud responses between the three models versions [38]. In the base state, CESM1 has the least cloud liquid water path (LWP). CESM2 exhibits much more LWP, with the CESM2-Z22 values being between the two other models. The difference between the models are concentrated mostly in the extra-tropics (>30°) over the oceans. The difference in cloud ice water path (IWP) between CESM2 and CESM2-Z22 are small, but both exhibit much less IWP than CESM1. While the LWP and IWP distributions are more realistic in CESM2 than in CESM1 [e.g., 40, 41], according to [38], they are even more realistic in CESM2-Z22 than in CESM2. In

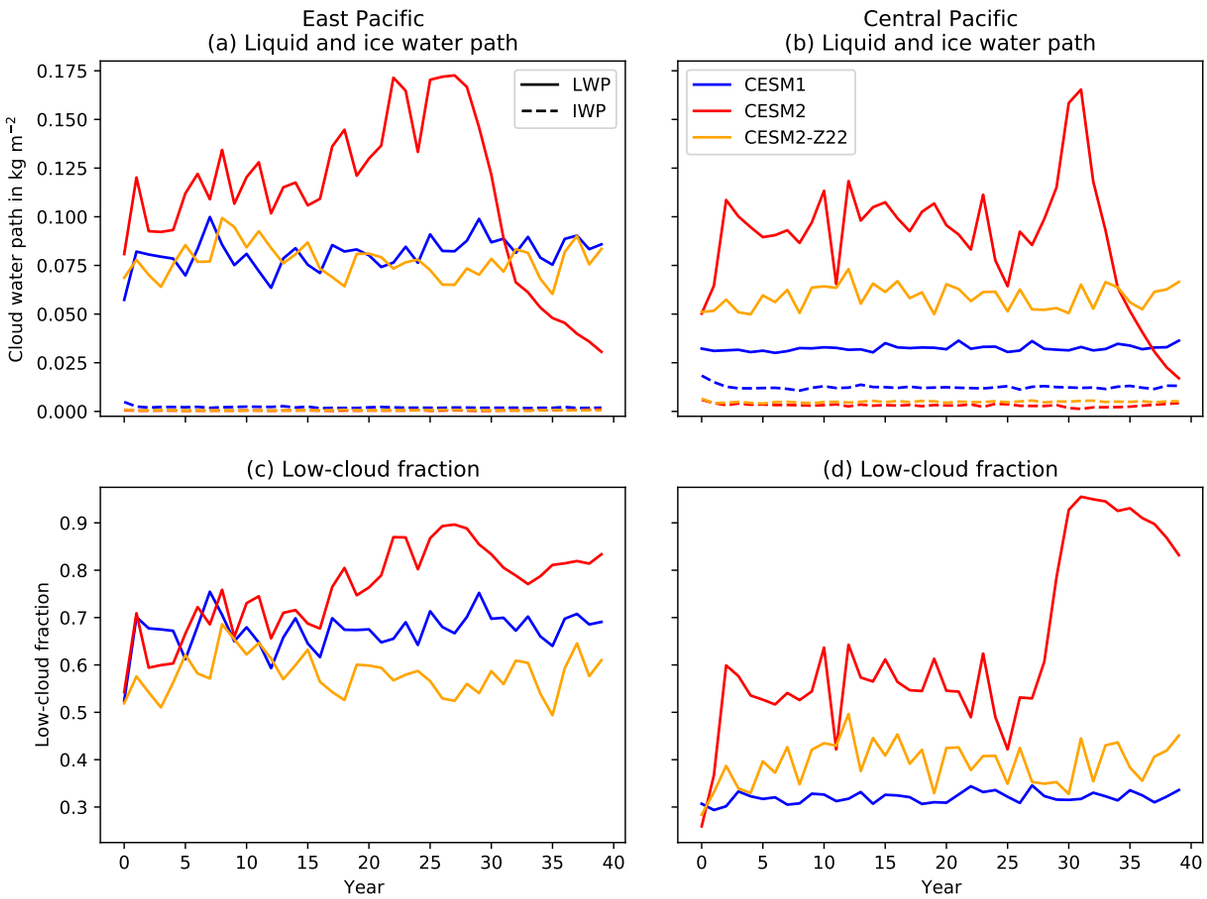

Fig. 10 Annual means of (a) East Pacific and (b) central Pacific cloud (solid) liquid and (dashed) ice water path (LWP and IWP, respectively), as well as (c) East Pacific and (d) central Pacific low-cloud fraction for (blue) CESM1, (red) CESM2, and (orange) CESM2-Z22.

response to dQ-EP-WP in the EP, the strongest LWP response is exhibited by CESM2, with a continuous increase throughout the simulation, while in CESM1 and CESM2-Z22 the LWP remains almost constant (Fig. 10a). The low-cloud cover response in the EP is similar to the LWP response (Fig. 10c). In the central Pacific both the LWP and the low-cloud cover almost instantly more than double in CESM2, while there is only a slight and slow increase in CESM2-Z22 and little response in CESM1 (Fig. 10b,d). The effect of low liquid clouds on CRF_{SW} was explained in section 4.1 and the changes discussed here are consistent with Fig. 9. As mentioned above, the differences between CESM1 and CESM2 may be impacted by different MLD and base-state Q-fluxes, explaining part of the difference of the cloud response. However, there is no difference in these quantities between CESM2 and CESM2-Z22, indicating that the different cloud response is mostly due to the difference in cloud parametrisation. The main changes in the cloud microphysics parametrisation undertaken by [38] for CESM2-Z22 are the removal of the cloud ice number concentration delimiter and a substepping of the microphysics scheme, i.e., a shorter time-step. The change of the ice delimiter should mainly impact the ice clouds. However, in dQ-EP-WP in the EP and the central Pacific there is very little difference between CESM2 and CESM2-Z22 in terms of IWP (Fig. 10a,b), suggesting that this change to the cloud microphysics is not the cause for the difference between the models. Indeed, [38] find that the substepping of the microphysics scheme weakens the cloud feedback in their simulations of

the LGM climate. However, the reasons for this are unclear. [38] mention that substepping has been found to affect rain evaporation and self-collection processes but no definitive conclusions are given. While we cannot give any definitive conclusions either, we report that in the East and central Pacific the precipitation in dQ-EP-WP declines more strongly in CESM2 than in CESM2-Z22 (Fig. S24), indicating that less precipitation leads to more LWP observed in CESM2 than in CESM2-Z22, inducing the stronger CRF_{SW}.

5 Conclusion

In this study we investigate the impact on the climate of a redistribution of energy across the Earth's surface. The investigation is motivated by the historically observed changes in surface-warming/cooling patterns and the idea of the pattern effect, i.e., that the patterns of surface warming affect the global-mean warming itself. Experiments are performed with a slab-ocean model with changed ocean heat transport convergence (Q-flux) to qualitatively mimic surface-warming changes. The impact on the global-mean surface temperature as well as individual radiative feedbacks in the climate system is investigated.

Even though the positive and negative Q-flux changes cancel perfectly by construction, all experiments exhibit a global cooling. This "asymmetry" in the response to positive and negative Q-flux changes is consistent with the findings of [45]. They attribute the asymmetry mostly to the effect of sea ice. However, here we also find an asymmetry, albeit much weaker, in the absence of sea ice, which is mainly due to the cloud feedback (Fig. S25). [45] use CESM1 while our simulations with deactivated sea ice are performed with CESM2. Since these models differ significantly in the cloud response to Q-flux changes (section 4.3) this may explain the different results.

We find another asymmetry between cooling in the Northern and Southern Hemispheres, similar to the results of [24]: A cooling in the North Atlantic, such as may be induced by a decline in the Atlantic meridional overturning circulation (AMOC; e.g., [33]), gives rise to a much weaker global cooling than a similar local cooling in the South Atlantic. This is related to a geographically less constrained sea-ice development in the Southern Ocean than in the Arctic, as well as a strong feedback loop between the Southern Ocean and the stably stratified regions of the tropics (especially the East Pacific): Cooling in the Southern Ocean is advected via climatological winds into the stably stratified regions of the tropics, decreasing local sea-surface temperatures and increasing stability, causing enhanced cloud feedback due to increased low-cloud formation. More low clouds lead to more reflection of sunlight, inducing further local cooling, which in turn induces more low-cloud formation [25–27].

Since in addition to the global-mean warming a CO_2-forcing may induce changes in surface-warming *patterns* [22, 23, 33], the linearity of the combined response to the Q-flux changes and CO_2-forcing is tested. We find that strong non-linearities are introduced by the sea-ice expansion and the accompanying surface-albedo feedback. In an experiment with deactivated sea ice, non-linearities are much smaller but non-negligible. In the absence of sea ice, the remaining non-linearities are introduced mainly by the cloud feedback, especially over the Southern Ocean and the East Pacific.

The dependence of the effects of the Q-flux changes on the specific employed climate model are investigated by comparing different versions of the same model, differing mainly in terms of cloud parametrisation. A significant dependence on these differences of the response to different Q-flux changes is found. This implies that the pattern effect in different climate models depends in particular on the cloud parametrisation, potentially associated with clouds being the largest contributor to differences between climate models in terms of the response to CO_2-forcing [21].

Finally, we caution against over-interpretation of our results. In an additional experiment, where a surface cooling of 10 Wm^{-2} is implemented over the whole Southern Ocean with a compensating surface warming in the tropics in the fully-coupled version of CESM2 by changing long-wave surface fluxes, the global-mean surface-air temperature decreases by only $\sim$1 K after 50 years of simulation (Fig. 11). However, in such experiments with a full dynamical ocean, the ocean can react and set up a compensating circulation. Hence, it is thought-provoking, that a transition to a Snowball Earth is induced by comparatively small redistributions of surface energy in slab-ocean models,

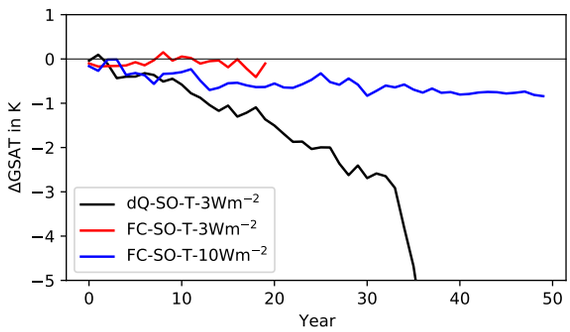

Fig. 11 Global-mean surface-air temperature change (ΔGSAT) in (black) dQ-SO-T-3Wm^{-2}, (red) FC-SO-T-3Wm^{-2}, and (blue) FC-SO-T-10Wm^{-2}. "FC" indicates the fully-coupled version of CESM2.

where by construction a compensating dynamical ocean circulation change cannot occur.

Supplementary information. This article has an accompanying supplementary file.

Statements and Declarations

Competing interests

The authors have no relevant financial or non-financial interests to disclose.

Funding

The data 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 CESM simulations were performed on the FRAM supercomputer at the University of Troms (UiT) provided by UNINETT Sigma2, under the project NN9348k. 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.

Availability of data and materials

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 CESM simulations (see Code availability); we used model versions 1.2.2 and 2.1.3, freely available at https://www2.cesm.ucar.edu/models/cesm1.2/ and https://www.cesm.ucar.edu/models/cesm2, respectively.

Code availability

The code adjustments for CESM2-Z22 are available at https://doi.org/10.5065/bdr7-wt42. Pre- and postprocessing code for the CESM data are available at zenodo (https://doi.org/10.5281/zenodo.7950682), as are the scripts for changing the CESM code to run the experiments presented here (https://doi.org/10.5281/zenodo.8334523).

Authors' contributions

Kai-Uwe Eiselt and Rune G. Graversen conceived and designed the study. Material preparation, data collection and analysis were performed by Kai-Uwe Eiselt. The first draft of the manuscript was written by Kai-Uwe Eiselt with comments and revisions from Rune G. Graversen.

Ethics approval

Not applicable.

Consent to participate

Not applicable.

Consent for publication

The authors consent to the publication of the manuscript.

References

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

[2] Andrews, T., Gregory, J.M., Webb, M.J., Taylor, K.E.: Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models. Geophys. Res. Lett. **39** (2012) https://doi.org/10.1029/2012GL051607

[3] Andrews, T., Gregory, J.M., Webb, M.J.: The dependence of radiative forcing and feedback on evolving patterns of surface temperature change in climate models. J. Climate **28**, 1630–1648 (2015) https://doi.org/10.1175/JCLI-D-14-00545.1

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

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

[6] Mauritsen, T.: Clouds cooled the earth. Nat. Geosci. **9**, 865–867 (2016) https://doi.org/10.1038/ngeo2838

[7] Zhou, C., Zelinka, M.D., Klein, S.A.: Impact of decadal cloud variations on the earth's energy budget. Nat. Geosci. **9**, 871–874 (2016) https://doi.org/10.1038/NGEO2828

[8] Zhou, C., Zelinka, M.D., Klein, S.A.: Analyzing the dependence of global cloud feedback on the spatial pattern of sea surface temperature change with a Greens function approach. J. Adv. in Model. Earth Syst. **9**, 2174–2189 (2017) https://doi.org/10.1002/2017MS001096

[9] Ceppi, P., Gregory, J.M.: Relationship of tropospheric stability to climate sensitvity and Earth's observed radiation budget. Proc. Natl. Acad. Sci. (USA) **114**, 13126–13131 (2017) https://doi.org/10.1073/pnas.1714308114

[10] Dong, Y., Proistosescu, C., Armour, K.C., Battisti, D.S.: 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 (2019) https://doi.org/10.1175/JCLI-D-18-0843.1

[11] Murphy, J.M.: Transient response of the Hadley Centre coupled ocean-atmosphere model to increasing carbon dioxide. Part III: Analysis of global-mean response using simple models. J. Climate **8**, 496–514 (1995) https://doi.org/10.1175/1520-0442(1995)008⟨0496:TROTHC⟩2.0.CO;2

[12] Senior, C.A., Mitchell, J.F.B.: The time-dependence of climate sensitivity. Geophys. Res. Lett. **27**, 2685–2688 (2000) https://doi.org/10.1029/2000GL011373

[13] Boer, G.J., Yu, B.: Climate sensitivity and climate state. Climate Dyn. **21** (2003) https://doi.org/10.1007/s00382-003-0323-7

[14] Gregory, J.M., Ingram, W.J., Palmer, M.A., Jones, G.S., Stott, P.A., Thorpe, R.B., Lowe, J.A., Johns, T.C., Williams, K.D.: A new method for diagnosing radiative forcing and climate sensitivity. Geophys. Res. Lett. **31** (2004) https://doi.org/10.1029/2003GL018747

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

[16] Taylor, K.E., Stouffer, R.J., Meehl, G.A.: A summary of the CMIP5 experiment design. Technical report (December 2009). https://pcmdi.llnl.gov/mips/cmip5/docs/Taylor_CMIP5_design.pdf?id=98

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

[18] Zhang, X., Deser, C., Sun, L.: Is there a tropical response to recent observed Southern Ocean cooling? Geophys. Res. Lett. **48** (2021) https://doi.org/10.1029/2020GL090861

[19] Wallace, J.M., Hobbs, P.V.: Atmospheric Science: An Introductory Survey. Academic Press, 525 B Street, Suite 1900, San Diego, CA (2006)

[20] Klein, S.A., Hall, A., Norris, J.R., Pinus, R.: Low-cloud feedbacks from cloud-controlling factors: A review. Surv. Geophys. **38**, 1307–1329 (2017) https://doi.org/10.1007/s10712-017-9433-3

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

[22] Lee, S., L'Heureux, M., Wittenberg, A.T., Seager, R., N C. Johnson, P.A.O.: On the future zonal contrasts of equatorial pacific climate: Perspectives from observations, simulations, and theories. npj Climate and Atmospheric Science **5** (2022) https://doi.org/10.1038/s41612-022-00301-2

[23] Seager, R., Cane, M., Henderson, N., Lee, D.-E., Abernathy, R., Zhang, H.: Strengthening tropical Pacific zonal sea surface temperature gradient consistent with rising greenhouse gases. Nat. Clim. Change **9**, 517–525 (2019) https://doi.org/10.1038/s41558-019-0505-x

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

[25] Kim, H., Kang, S.M., Kay, J.E., Xie, S.-P.: Subtropical clouds key to Southern Ocean teleconnections to the tropical Pacific. Proc. Natl. Acad. Sci. (USA) **119** (2022) https://doi.org/10.1073/pnas.2200514119

[26] Dong, Y., Armour, K.C., Battisti, D.S., Blanchard-Wrigglesworth, E.: Two-way teleconnections between the Southern Ocean and the tropical Pacific via a dynamic feedback. J. Climate **35**, 2667–2682 (2022) https://doi.org/10.1175/JCLI-D-22-0080.1

[27] Kang, S.M., Yu, Y., Deser, C., Zhang, X., Kang, I.-S., Lee, S.-S., Rodgers, K.B., Ceppi, P.: Subtropical clouds key to Southern Ocean teleconnections to the tropical Pacific. Proc. Natl. Acad. Sci. (USA) **120** (2023) https://doi.org/10.1073/pnas.2300881120

[28] Zhang, L., Delworth, T.L., Cooke, W., Yang, X.: Natural variability of Southern Ocean convection as a driver of observed climate trends. Nat. Climate Change **9**, 59–65 (2019) https://doi.org/10.1038/s41558-018-0350-3

[29] Dong, Y., Pauling, A.G., Sadai, S., Armour, K.C.: Antarctic ice-sheet meltwater reduces transient warming and climate sensitivity through the sea-surface temperature pattern effect. Geophys. Res. Lett. **49** (2022) https://doi.org/10.1029/2022GL101249

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

[31] Lin, Y.-J., Hwang, Y.-T., Ceppi, P., Gregory, J.M.: Uncertainty in the evolution of climate feedback traced to strength of the Atlantic meridional overturning circulation. Geophys. Res. Lett. **46**, 12331–12339 (2019) https://doi.org/10.1029/2019GL083084

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

[33] Eiselt, K.-U., Graversen, R.G.: On the control of North-Hemispheric feedbacks by AMOC, evidence from CMIP and slab-ocean modeling. J. Climate, (2023) https://doi.org/10.1175/JCLI-D-22-0884.1

[34] Huguenin, M.F., Holmes, R.M., England, M.H.: Drivers and distribution of global ocean heat uptake over the last half century. Nat. Commun. **13** (2022) https://doi.org/10.1038/s41467-022-32540-5

[35] Danabasoglu, G., Lamarque, J.., Bacmeister, J., Bailey, D.A., DuVivier, A.K., Edwards, J., Emmons, L.K., Fasullo, J., Garcia, R., Gettelman, A., Hannay, C., Holland, M.M., Large, W.G., Lauritzen, P.H., Lawrence, D.M., Lenaerts, J.T.M., Lindsay, K., Lipscomb, W.H., Mills, M.J., Neale, R., Oleson, K.W., OttoBliesner, B., Phillips, A.S., Sacks, W., Tilmes, S., Kampenhout, L., Vertenstein, M., Bertini, A., Dennis, J., Deser, C., Fischer, C., FoxKemper, B., Kay, J.E., Kinnison, D., Kushner, P.J., Larson, V.E., Long, M.C., Mickelson, S., Moore, J.K., Nienhouse, E., Polvani, L., Strand, P.J.R.W.G.: The Community Earth System Model Version 2 (CESM2). Journal of Advances in Modeling Earth Systems **12** (2020) https://doi.org/10.1029/2019MS001916

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

[37] Hurrell, J.W., Holland, M.M., Gent, P.R., Ghan, S., Kay, J.E., Kushner, P.J., Lamarque, J.-F., Large, W.G., Lawrence, D., Lindsay, K., Lipscomb, W.H., Long, M.C., Mahowald, N., Marsh, D.R., Neale, R.B., Rasch, P., Vavrus, S., Vertenstein, M., Bader, D., Collins, W.D., Hack, J.J., Kiehl, J., Marshall, S.: The Community Earth System Model: A framework for collaborative research. Bull. Amer. Meteor. Soc. **94**, 1339–1360 (2013) https://doi.org/10.1175/BAMS-D-12-00121.1

[38] Zhu, J., Otto-Bliesner, B.L., Brady, E.C., Gettelman, A., Bacmeister, J.T., Neale, R.B., Poulsen, C.J., Shaw, J.K., McGraw, Z.S., Kay, J.E.: LGM paleoclimate constraints inform cloud parameterizations and equilibrium climate sensitivity in CESM2. Journal of Advances in Modeling Earth Systems **14** (2022) https://doi.org/10.1029/2021MS002776

[39] Zhu, J., Otto-Bliesner, B.L., Brady, E.C., Poulsen, C.J., Tierney, J.E., Lofverstrom, M., DiNezio, P.: Assessment of equilibrium climate sensitivity of the community earth system model version 2 through simulation of the last glacial maximum. Geophys. Res. Lett. **14** (2021) https://doi.org/10.1029/2020GL091220

[40] Gettelman, A., Hannay, C., Bacmeister, J.T., Neale, R.B., Pendergrass, A.G., Danabasoglu, G., Lamarque, J.-F., Fasullo, J.T., Bailey, D.A., Lawrence, D.M., Mills, M.J.: High climate sensitivity in the Community Earth System Model Version 2 (CESM2). Geophys. Res. Lett. **46**, 8329–8337 (2019) https://doi.org/10.1029/2019GL083978

[41] Bacmeister, J.T., Hannay, C., Medeiros, B., Gettelman, A., Neale, R., Fredriksen, H.B., Lipscomb, W.H., Simpson, I., Bailey, D.A., Holland, M., Lindsay, K., Otto-Bliesner, B.: CO_2 increase experiments using the CESM: Relationship to climate sensitivity and comparison of CESM1 to CESM2. Journal of Advances in Modeling Earth Systems **12** (2020) https://doi.org/10.1029/2020MS002120

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

[43] Liu, F., Lu, J., Garuba, O., Leung, L.R., Luo, Y., Wan, X.: Sensitivity of surface temperature to oceanic forcing via q-flux Greens function experiments. Part I: Linear response function. J. Climate **31**, 3625–3641 (2018) https://doi.org/10.1175/JCLI-D-17-0462.1

[44] Liu, F., Lu, J., Garuba, O., Leung, L.R., Harrop, B.E., Luo, Y.: Sensitivity of surface temperature to oceanic forcing via q-flux Greens function experiments. Part II: Feedback decomposition and Polar Amplification. J. Climate **31**, 6745–6761 (2018) https://doi.org/10.1175/JCLI-D-18-0042.1

[45] Liu, F., Lu, J., Garuba, O., Leung, L.R., Harrop, B.E., Luo, Y.: Sensitivity of surface temperature to oceanic forcing via q-flux

Greens function experiments. Part III: Asymmetric response to warming and cooling. J. Climate **33**, 1283–1297 (2020) https://doi.org/10.1175/JCLI-D-19-0131.1

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

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

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

[49] Block, K., Mauritsen, T.: Forcing and feedback in the MPI-ESM-LR coupled model under abruptly quadrupled CO_2. J. Adv. in Model. Earth Syst. **5**, 676–691 (2013) https://doi.org/10.1002/jame.20041

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

[51] Wood, R., Bretherton, C.S.: On the relationship between stratiform low cloud cover and lower-tropospheric stability. J. Climate **19**, 6425–6432 (2006) https://doi.org/10.1175/JCLI3988.1

[52] Stephens, G.L.: Radiation profiles in extended water clouds. II: Parameterization schemes. Journal of the Atmospheric Sciences **35**, 2123–2132 (1978) https://doi.org/10.1175/1520-0469(1978)035⟨2123:RPIEWC⟩2.0.CO;2

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

[54] Po-Chedley, S., Armour, K.C., Bitz, C.M., Zelinka, M.D., Santer, B.D., Fu, Q.: Sources of intermodel spread in the lapse rate and water vapor feedbacks. J. Climate **31**, 3187–3206 (2018) https://doi.org/10.1175/JCLI-D-17-0674.1