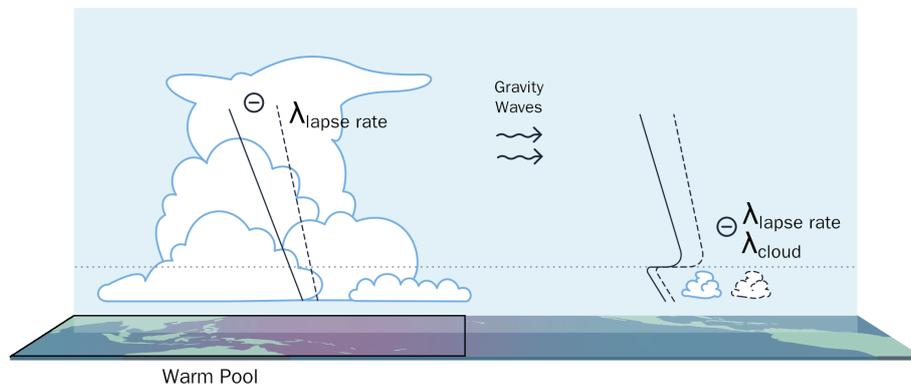# Asymmetries between the climate responses to CO_2 and stratospheric aerosol forcing

Dissertation
with the aim of achieving a doctoral degree
at the Faculty of Mathematics, Informatics and Natural Sciences
Department of Earth Sciences
at Universität Hamburg

submitted by
Moritz Günther

Hamburg, 2024

Department of Earth Sciences

Date of Oral Defense:
Reviewers:

Members of the examination commission:

Chair of the Subject Doctoral Committee
Earth System Sciences:

                                            Prof. Dr. Hermann Held

Dean of Faculty MIN:

                                            Prof. Dr.-Ing. Norbert Ritter

ABSTRACT

Understanding Earth's temperature response to radiative perturbations is pivotal in climate science, as emphasized in the World Climate Research Program's Grand Challenge on Clouds, Circulation and Climate Sensitivity. Surprisingly, a radiative perturbation from stratospheric aerosol produces a smaller temperature response than a radiative perturbation of the same magnitude from CO_2. Resolving this apparent paradox is the core motivation behind this thesis, with the ambition to understand the atmospheric processes that cause this climate response asymmetry.

The climate response asymmetry can be quantified by the feedback parameter, which is more negative (i.e., more stabilizing) for stratospheric aerosol than CO_2 forcing. The pattern effect framework provides a modern interpretation of variations of the feedback parameter *in time*. I apply this framework to explain differences of the feedback parameter *among forcing agents*, in this particular case between CO_2 and stratospheric aerosol forcing. I employ an earth system modelling approach, testing hypotheses with simulations with the general circulation model MPI-ESM 1.2, and demonstrating the identified key mechanisms in the CMIP6 multi model ensemble.

I identify two essential ingredients to explain the climate response asymmetry between CO_2 and stratospheric aerosol forcing: temperature change in the warm pool (tropical Indian and Western Pacific Ocean, 30°S - 30°N, 50°E - 160°W), and the Brewer-Dobson circulation. I show that the warm pool is the dominant region causing feedback differences between CO_2 and stratospheric aerosol forcing. In an ensemble of 120 decadal simulations forced with CO_2 and stratospheric aerosol forcing, a simple measure for warm pool temperature change explains 50 % of the variance of the feedback parameter. I prove that the warm pool cools particularly strongly from stratospheric aerosol forcing, and that the stratospheric circulation is a key ingredient to explain this. Stratospheric aerosol heats the stratosphere, leading to an acceleration of the Brewer-Dobson circulation and hence to an increased poleward energy transport. This provides additional cooling to the tropical surface that is not reflected in the forcing pattern at the top of the atmosphere, highlighting the role of surface forcing for understanding the evolution of temperature patterns.

My results engender conclusions about climate feedback beyond the response to stratospheric aerosol forcing. To understand temperature pattern formation, the surface forcing must be considered, which is in contrast to the established method of measuring forcing at the top of the atmosphere. Although climate feedback mainly refers to processes at the surface and in the troposphere, I show that the stratospheric circulation influences surface forcing patterns and thus temperature patterns and feedback.

ZUSAMMENFASSUNG

Die Temperaturantwort der Erde auf Störungen der Strahlungsbilanz (sog. *Forcing*) zu verstehen, ist in der Klimawissenschaft von entscheidender Bedeutung, wie in der Grand Challenge zu Wolken, Zirkulation und Klimasensitivität des World Climate Research Program hervorgehoben wird. Überraschenderweise erzeugt ein Forcing durch stratosphärisches Aerosol eine geringere Temperaturantwort als ein ebenso großes Forcing durch CO_2. Die Kernmotivation dieser Arbeit ist es, dieses scheinbare Paradox aufzulösen, verbunden mit dem Anspruch, die dieser Asymmetrie zugrundeliegenden atmosphärischen Prozesse zu verstehen.

Die Asymmetrie der Klimaantwort kann durch den Feedback-Parameter quantifiziert werden, welcher für Forcing durch stratosphärisches Aerosol negativer (also stabilisierender) ist als für CO_2. Der Erklärungsansatz des Mustereffekts (*Pattern Effect*) bietet eine moderne Interpretation von Variationen des Feedback-Parameters *mit der Zeit*. Ich wende diesen Ansatz an, um Unterschiede des Feedback-Parameters *zwischen Verursachern von Forcings* zu erklären, in diesem speziellen Fall zwischen Forcing durch CO_2 und Forcing durch stratosphärisches Aerosol. Ich wähle einen Erdsystemmodellierungsansatz, teste Hypothesen mit Hilfe von Simulationen mit dem globalen Klimamodell MPI-ESM 1.2 und zeige die identifizierten Schlüsselmechanismen im Multimodellensemble CMIP6 auf.

Ich identifiziere zwei essentielle Bestandteile zur Erklärung der Asymmetrie der Klimaantwort zwischen Forcing durch CO_2 und stratosphärisches Aerosol: Temperaturänderungen im Warm-Pool (tropischer Indischer und Westpazifischer Ozean, 30°S - 30°N, 50°O - 160°W) und die Brewer-Dobson-Zirkluation. Ich zeige, dass der Warm-Pool die dominierende Region ist, welche die Unterschiede des Feedback-Parameters zwischen CO_2 und stratosphärischem Aerosol verursacht. In einem Ensemble mit 120 dekadischen Simulationen mit Forcing durch CO_2 und stratosphärisches Aerosol erklärt ein einfaches Maß für Temperaturänderung im Warm-Pool 50 % der Varianz des Feedback-Parameters. Ich beweise, dass sich der Warm-Pool durch Forcing durch stratosphärisches Aerosol besonders stark abkühlt und dass die stratosphärische Zirkulation ein Schlüsselbestandteil der Erklärung dafür ist. Stratosphärisches Aerosol heizt die Stratosphäre, was zu einer Beschleunigung der Brewer-Dobson-Zirkulation und somit zu einem erhöhten Energietransport zu den Polen führt. Dieser verursacht zusätzliche Kühlung an der tropischen Erdoberfläche, welches sich nicht im Forcing-Muster am Oberrand der Atmosphäre widerspiegelt, was die Bedeutung des Forcings an der Oberfläche für das Verständnis der Entwicklung von Temperaturmustern hervorhebt.

Meine Ergebnisse erlauben Schlussfolgerungen, welche über die Antwort auf Forcing durch stratosphärisches Aerosol hinausgehen. Um die Erzeugung von Temperaturmustern zu verstehen, muss das Forcing an der Oberfläche betrachtet werden, was im Widerspruch zur etablierten Methode steht, nach der Forcing am Oberrand der Atmosphäre bestimmt wird. Obwohl Feedbackprozesse hauptsächlich an der Oberfläche und in der Troposphäre stattfinden, zeige ich, dass die stratosphärische Zirkulation Forcing-Muster an der Oberfläche und demzufolge auch Temperaturmuster und Feedback beeinflusst.

PUBLICATIONS RELATED TO THIS DISSERTATION

The following first-author publications are part of this dissertation and included in the appendix:

M. Günther, H. Schmidt, C. Timmreck, and M. Toohey (2022). "Climate Feedback to Stratospheric Aerosol Forcing: The Key Role of the Pattern Effect." *Journal of Climate* 35.24, pp. 4303–4317. DOI: `10.1175/JCLI-D-22-0306.1`

M. Günther, H. Schmidt, C. Timmreck, and M. Toohey (2024). "Why Does Stratospheric Aerosol Forcing Strongly Cool the Warm Pool?" *under review with minor revisions in Atmospheric Chemistry and Physics*. DOI: `10.5194/egusphere-2024-429`

Another single-author publication is currently in preparation and will be submitted after the submission of this dissertation:

M. Günther (2024). "Separating Forced Response from Internal Variability with Independent Component Analysis." *in prep.*

Two further publications to which I contributed as a PhD student are currently in preparation:

R. Hegde, M. Günther, C. A. Kroll, and H. Schmidt (2024). "Surface Temperature Dependence of Stratospheric Sulphate Aerosol Forcing and Feedback." *in prep.*

D. J. Burt, D. Putrasahan, M. Günther, and T. Ilyina (2024). "Identifying Episodic Regional Sea Surface Temperature Extremes in the Dominant-Mode of Latest Permian Internal Climate Variability in the Max-Planck Institute Earth System Model." *in prep. for Climate of the Past*

ACKNOWLEDGMENTS

I want to thank thank Hauke for his exceptional supervision. I learned a lot from Hauke about the climate, but also about science in general. I am thankful for his calm and patient advice, for giving me space to develop my thoughts freely into whichever direction seemed most promising to me, for his ability to guide my thoughts into clear questions and explanations, and for not letting me get away with imprecise statements. Claudia is an inspiring co-supervisor who taught me much about volcanoes and the stratosphere, and provided me with genuinely motivating support. To Matt I am thankful for always taking the time to think things through 10 % more thoroughly than I did, which led to very valuable discussions. Also thanks for the invitation to Canada, where I spent two fun and instructive months. Furthermore, I have learned a lot from Bjorn's way of thinking, his persistence in aiming to understand the big picture, and his thought-provoking questions. Another important pillar of support comes from the supervisors of my bachelor's and master's thesis, Jan and Ronny, who bolstered my earliest steps in science.

The people who teach me science

The second group of people I owe thanks are those who make science possible - often by working in the background. I want to specifically mention Antje and Angela. I don't want to forget the people who maintain wikipedia pages, python libraries, clean my office, and repair broken stuff in the building - this work is somewhat invisible, but nonetheless essential. I also want to thank all people who carry climate-related demands to the streets and parliaments. Without their work, a large part of climate science would only be for the libraries.

The people doing the invisible work behind science

Third, I reach out to thank the people who made all non-science-related aspects of my life fun, and without whom my work on this dissertation would have been a glum endeavor. This includes my friends from Hamburg, among whom I want to single out Luca, Tomer, Enora, Abisha, Helene, Martin, Hernan, Clara, Janina, Paul, Clarissa, George, Diego, Marius, Freddy, both Meikes, and also David, to whom I am extra grateful for his very detailed and on-point comments on an earlier version of this thesis. I am thankful to have my friends from Leipzig, who have been and will always be invited at my place: Nico, Johannes, Simba, Pauline, Flora, Steff, Alex, and Arjuna. I got further support from my Shapira house friends Devon, Yoshi, and Danielle. However, the person I am most thankful to is Yuval, for his loving support, for moving from another continent to Hamburg, for patiently finding (and if necessary faking) interest in my nerd stuff, and for sharing large parts of life with me.

The people outside science who make it fun

CONTENTS

Part I

UNIFYING ESSAY

INTRODUCTION

Perturbations to Earth's climate from stratospheric aerosol have always influenced human life and will continue to do so in the future. Aerosol can be added to the stratosphere by large volcanic eruptions. In the future, artificial cooling by deliberate injection of aerosol into the stratosphere may be seen as an adequate way to moderate climate change from greenhouse gases. In this thesis, I investigate the longer-term climate impacts of stratospheric aerosol forcing. They arise from two processes: the reflection of solar shortwave radiation which leads to a cooling of the Earth's surface; and the absorption of solar and terrestrial longwave radiation, which causes a local heating of the stratosphere and a greenhouse effect. While increasing CO_2 concentrations heats the planet by trapping longwave radiation, stratospheric aerosol cools the Earth by reflecting sun light (partly offset by the longwave effect).

The primary impetus motivating this thesis is the finding that the Earth seems to be surprisingly stable to radiative perturbations from stratospheric aerosol. A normalized measure for Earth's sensitivity to radiative forcing is the feedback parameter, which scales inversely with the surface temperature change following a radiative perturbation. Numerous modeling studies find an up to 40 % more negative feedback parameter for stratospheric aerosol forcing than for CO_2 forcing (e. g., Boer et al., 2006; Gregory et al., 2016; Marvel et al., 2016), indicating that the temperature response to stratospheric aerosol forcing is substantially weaker than the temperature response to a CO_2 forcing of equal magnitude. This is surprising because one might expect that the temperature change required for Earth to balance a radiative perturbation only depends on the perturbation's magnitude.

This seeming paradox points out a lack in our knowledge of climate feedback processes in general, and hence in our understanding of climate changes. It raises the question of how a unit loss or gain of power can lead to different global mean temperature responses, depending on the way that power is added or drained from the climate system. Apparently, there are unknown asymmetries between the climate responses to CO_2 and stratospheric aerosol forcing. It is crucial to resolve this issue for two reasons: first, it will help us to better understand the climate response to stratospheric aerosol forcing; second, if we really want to understand radiative forcing and feedback in general, we must be able to explain why different forcings cause different feedbacks. These thoughts lead to the guiding research question of the present thesis:

Why does stratospheric aerosol forcing cause stronger feedback than CO_2 forcing?

BACKGROUND

First, I take a step back and introduce useful concepts for the treatment of the research question, situating it within the broader context of ongoing research.

For a long time, climate scientists have asked how much Earth's temperature changes in response to forcings such as altered CO_2 concentrations or additional stratospheric aerosol. The answer to this question is fundamental for characterizing the climate system, predicting future climate change, and assessing climate risks. The question reaches back to the 19th century (Arrhenius, 1896), and is still under intense debate today (Sherwood et al., 2020). In the following, I

- conceptualize a scheme to think about responses to radiative perturbations,

- explain why this is an important topic to explore,

- show the evolution and gaps of our knowledge of forcing and response, and

- point out how resolving asymmetries between the climate responses to CO_2 and stratospheric aerosol forcing can advance our understanding of forcing and response.

2.1 THE ENERGY BALANCE FRAMEWORK

A well-established and useful way to think about climate changes builds on Earth's energy balance, which consists of incoming shortwave (SW) and outgoing longwave (LW) radiation. In a steady state, the global-mean net flux N at the top of the atmosphere (TOA) is zero: absorbed SW radiation must equal outgoing LW radiation (Kiehl and Trenberth, 1997). This balance can be disturbed by changes to radiatively active components of the Earth, e. g. greenhouse gas concentrations or stratospheric aerosol. Such a "radiative forcing", or simply "forcing", leads to a gain or loss of energy. The Earth heats up as it gains energy, increasing the emitted LW radiation from the surface according to the Stefan-Boltzmann law (Boltzmann, 1884; Stefan, 1879). A new steady state is reached when the surface temperature has changed enough to modify the outgoing LW radiation so that the radiation budget is closed again.

Forcing F is defined as the global mean radiative flux perturbation at the TOA at zero temperature change, and T is the global mean change in 2-meter-air temperature. They are linked via the linearized equation (Gregory et al., 2002)
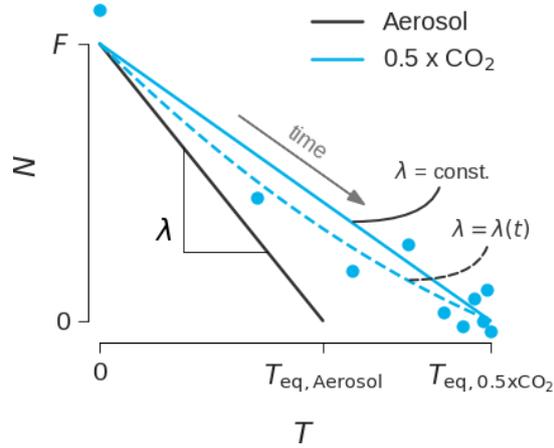
Figure 2.1: Idealized $N(T)$ plot (Gregory plot) for constant forcing of magnitude F, once forced by stratospheric aerosol and once by halving CO_2 concentrations. N and T are deviations from the initial steady state. All values are negative, but have been multiplied by -1 to facilitate comparison with the usual Gregory plots from warming scenarios. The blue points show annual-mean $N(T)$ for an abrupt 0.5 x CO_2 scenario. The solid blue line illustrates an estimate of λ under the assumption of constant feedback. Most climate models show a time-dependent $\lambda(t)$ which qualitatively follows the dashed blue line. The black line describes $N(T)$ from volcanic aerosol forcing, only showing the constant λ simplification. This figure's "data" (the dots) is made up for illustration purposes. Note that an actual equilibration would take 1000s of years.

$$\overbrace{}^{\text{radiative perturbation (independent of temperature)}}$$
$$N = \underset{\text{temperature-mediated response}}{F + \lambda T} \tag{2.1}$$

The change in TOA radiation N (in $W\,m^{-2}$) induced by surface temperature change T (in K) is the feedback parameter $\lambda = \partial_T N$ (in $W\,m^{-2}\,K^{-1}$), which is a measure for the climate's stability. The more negative λ, the less temperature change is needed to balance a given radiative forcing. In a world without atmosphere, λ would be equal to $-\partial_T \sigma T^4 = -4\sigma T^3$, where σ is the Stefan-Boltzmann constant. However, radiative feedbacks modify λ to be less negative than the pure Planck response. Radiative feedbacks arise from temperature-dependent components of the climate system which in turn affect radiation, namely clouds, the lapse rate, (absolute or relative) humidity, and surface albedo.

An advantage of Eq. 2.1 is the clean separation of the TOA flux anomaly into processes that depend on temperature change (λT) and those that do not depend on temperature change (F). The components of Eq. 2.1 can be conveniently illustrated in a scatter plot showing $N(T)$. As an example, I show this for the case of cooling by an abrupt halving of the CO_2 concentration in Fig. 2.1 (dark blue line). After the

forcing is introduced, the global-mean surface temperature changes and affects N in a way to oppose F, until a new steady state is reached ($N = 0$).

If Eq. 2.1 holds, the global-mean equilibrium temperature change for a given forcing F can be determined by $T_{eq} = F/\lambda$. Consequently, knowledge of λ would enable us to predict temperature impacts of any climate perturbation, given that the forcing is known. The task to predict Earth's temperature change in response to any forcing therefore essentially boils down to determining λ.

However, the simple linear picture of Eq. 2.1 often does not well describe model output (Murphy, 1995; Senior and Mitchell, 2000) and observations (Gregory and Andrews, 2016), which has prompted many propositions to modify the energy balance approach.

At first, the definition of forcing has been improved to better predict surface temperature changes. The original *instantaneous forcing* is defined as the mean TOA flux difference between a forced and an unforced state, but with all other climate variables (especially humidity, temperatures, clouds, and sea ice) fixed. This does not account for atmospheric adjustments that happen on time scales of months to weeks independent of surface temperatures, such as the stratospheric heating in response to aerosol. Instantaneous forcing is not a good predictor of temperature change, because the rapid adjustments change the TOA flux without affecting temperatures, effectively breaking the separation of Eq. 2.1 into temperature-dependent and -independent processes. Numerous suggestions to overcome this issue have been made, including measuring the forcing at the tropopause and allowing for adjustments of the stratosphere (*adjusted forcing,* Forster et al., 1997; IPCC, 2001). The research community settled for *effective forcing* as the standard forcing definition (Forster et al., 2016; IPCC, 2022). Effective forcing is the TOA flux change after introducing the forcing agent but at fixed sea surface temperatures (SST) and sea ice (Hansen et al., 2005), allowing for adjustments of the whole atmosphere. Thus, rapid adjustments are accounted for in the forcing term and separated from the surface temperature-mediated feedback processes. However, this comes at the expense of increased model-dependence, because diagnosing effective forcing involves the whole atmospheric component and not just the radiation model. In this work, I exclusively use the *effective forcing* definition, and use this term interchangeably with *forcing*. The forcing is approximately equal to the extrapolation of $N(T)$ for $T \rightarrow 0$ in the $N(T)$ plot (Fig. 2.1).

However, even the effective forcing definition cannot satisfy the requirement that the temperature response be proportional to the forcing. In the early 2000s, it was shown that the feedback parameter varies on the order of 10s of percent depending on the forcing agent (Hansen et al., 2005; Joshi and Shine, 2003). Since then, observational and modeling studies have repeatedly confirmed the dependence of

Source	$\lambda_{CO_2}/\lambda_{Aerosol}$
Hansen et al., 2005	1.00
	0.88
Boer et al., 2006	0.81*
Merlis et al., 2014	0.84* - 0.94*
	0.69* - 0.92*
Gregory et al., 2016	0.69 ± 0.09
Marvel et al., 2016	0.73
	5-95 % CI: [-0.61, 2.06]
Modak et al., 2016	<1
Ceppi and Gregory, 2019	0.45*
	0.71*
Gregory et al., 2020	<1
Zhao et al., 2021	0.66* (low lat.) - 1.16* (high lat.)
Salvi et al., 2023	1.19*
Zhou et al., 2023	≈ 0.75*

Table 2.1: Efficacy of stratospheric aerosol forcing (ratio of CO_2-feedback to aerosol-feedback) in previous studies.

the feedback parameter on forcing agents (e.g. Marvel et al., 2016; Modak et al., 2016; Zhou et al., 2023).

Stratospheric aerosol forcing has consistently been shown to cause stronger feedback than CO_2 forcing. I document previously reported ratios of feedback from CO_2 and stratospheric aerosol forcing, also termed efficacy (Hansen et al., 2005), in Table 2.1. These studies are based on a variety of climate models and aerosol forcing profiles. Almost all of them point towards stronger feedback to forcing from stratospheric aerosol than from CO_2. This implies that aerosol forcing would lead to less temperature change than CO_2 forcing of the same magnitude (Fig. 2.1, black line). It is a clear indication of a climate response asymmetry and provides the empirical basis of the thesis' main research question.

Furthermore, feedback has been shown to vary with time under increasing CO_2 concentrations (Murphy, 1995), and even under constant CO_2 forcing (Senior and Mitchell, 2000). Radiative feedback has also varied throughout the historical period (Gregory et al., 2020; Rugenstein et al., 2016a). In almost all climate models, λ weakens over time, i. e., becomes less negative (Geoffroy et al., 2013), similar to the dashed blue line in Fig. 2.1.

The non-constancy of λ has been identified at least since 1995 (Murphy, 1995). Yet, for two more decades the assumption of constant feedback persisted in the research community, which long assumed

that the climate response λT only depends on global mean surface temperature (Gregory, 2004, Eq. 2.1). In the following section, I introduce the modern view on this: the pattern effect framework, which recognizes the importance of temperature patterns for radiative feedback. A field of two-dimensional temperature change not only has a global mean, but also a spatial pattern, and the word "temperature pattern" refers to the latter. Since the slow time scales are governed by the ocean, temperature patterns are often equated with SST patterns.

2.2 THE PATTERN EFFECT

The importance of evolving SST patterns for explaining the time-varying λ was recognized in the 2010s and led to the development of the "pattern effect" framework, coined as such in 2016 (Stevens et al., 2016). Like two sides of the same coin, two major explanatory approaches have developed (Rugenstein et al., 2023b), which emphasize different climate processes, but are largely equivalent (Haugstad et al., 2017).

The first approach is based on the recognition that the efficacy of atmospheric cooling by the ocean depends on the location of the heat uptake (Rose and Rayborn, 2016). Evolving patterns of ocean heat uptake lead to changing efficacies, and therefore changing feedbacks (Geoffroy et al., 2013; Held et al., 2010; Lin et al., 2021; Rugenstein et al., 2016a; Winton et al., 2010).

The second approach explains radiative feedback in terms of surface temperature patterns (Andrews et al., 2015; Dong et al., 2019; Gregory et al., 2020; Zhou et al., 2017). In a review and opinion paper, Rugenstein et al. (2023b) summarize the main pattern effect idea: "Much like applying a force uniformly over someone's entire body will elicit a very different reaction than tickling the soles of that person's feet, a degree of global warming spread out evenly will cause a different radiative response than if that same warming were concentrated in a climate sweet spot (a location where surface warming produces efficient radiative damping)" (i. e., negative feedback).

One such climate sweet spot is the equatorial Indian and Western Pacific Ocean (30°S - 30°N, 50°E - 160°W; Dong et al., 2019). Due to the prevailing high SST, it is a major source region of deep convection, and has been termed the Warm Pool (WP) or Earth's heat engine (De Deckker, 2016). The location of the WP and its exceptionally stabilizing feedback processes are shown in Fig. 2.2. Any temperature change at the WP surface is efficiently communicated by convection to the free troposphere, which approximately follows a moist adiabatic temperature profile (solid line in Fig. 2.2). As a consequence of the moist adiabatic vertical profile, any temperature change at the surface is amplified aloft (dashed line in Fig. 2.2; Karoly et al., 1994). Since the top-heavy temperature change is favorable for radiating energy to
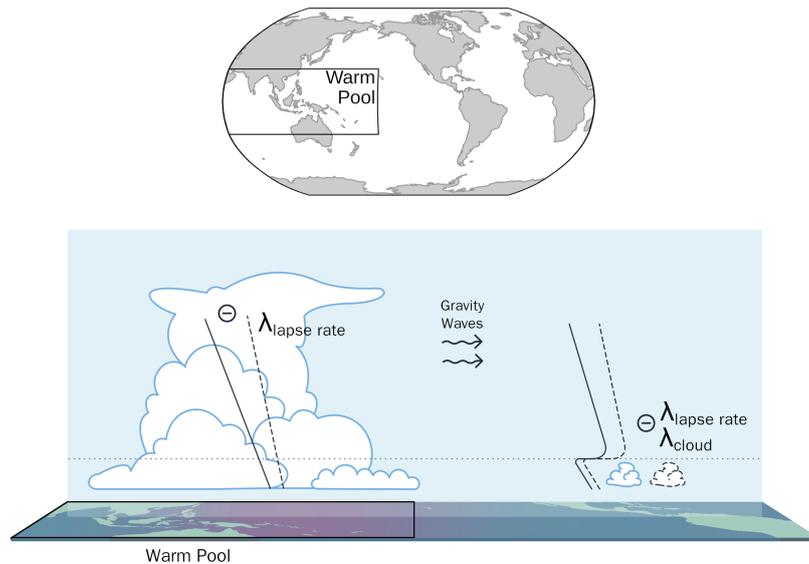
Figure 2.2: Top: Location of the Warm Pool (WP). Bottom: The pattern effect of the tropical Pacific Ocean. Solid lines represent the lapse rates of the climate base state, dashed lines a hypothetical state that follows from surface warming of the warm pool, which is located in the left half of the figure. Minus signs represent the negative lapse rate and cloud feedbacks.

space, this effect dampens the temperature response and constitutes a negative lapse rate feedback (Held and Soden, 2000). The tropical free troposphere cannot sustain large horizontal temperature gradients, and hence the same lapse rate is communicated via gravity waves to the tropical subsidence regions (Bretherton and Smolarkiewicz, 1989), e. g., the Eastern Pacific (right side of Fig. 2.2). A warming at the Western Pacific surface therefore leads to higher free tropospheric temperatures above the relatively cool ocean of the tropical subsidence regions. This enhanced inversion provides favorable conditions for the formation of stratus and stratocumulus clouds (Bretherton, 2015), in particular in the central and Eastern Pacific subsidence regions of the Walker circulation. Since the low clouds' main effect is the reflection of SW radiation, this constitutes a negative cloud feedback (Gettelman and Sherwood, 2016).

The local and remote lapse-rate and cloud feedbacks that originate from the Western Pacific are so strong that they exert a dominant control on Earth's global-mean feedback (Andrews et al., 2015; Dong et al., 2019). The more temperature change is concentrated in the WP, the more this region's stabilizing feedbacks will affect the global mean response. Note that the sign of the feedback is independent of the sign of the forcing, so that the WP feedback processes are stabilizing under both negative and positive forcing. They cool the Earth when it is heated, and they heat it when it is cooled.

The findings I summarized in the previous paragraphs bear important implications, which also serve as motivations for this work. First, they present a problem for the energy balance framework, because they contradict the assumption that λ is a planetary constant, which uniquely describes the Earth system's stability to radiative perturbations. Second, simple zero- or one-dimensional models cannot accurately describe transient climate change if they do not allow for varying feedback parameters. Third, the feedback from observed or paleo-climate change poorly predicts future climate change unless changing SST patterns are taken into account (Alessi and Rugenstein, 2023; Rugenstein et al., 2016a; Sherwood et al., 2020). Regarding the topic of this thesis, this also applies to estimating climate sensitivity from the observed climate response following volcanic eruptions (Andrews et al., 2022; Boer et al., 2006; Pauling et al., 2023).

The pattern effect framework has long been applied to the variation of the feedback parameter with time, which has been subject to extensive research (e.g. Andrews et al., 2018; Andrews and Webb, 2018; Armour et al., 2013; Held et al., 2010; Stevens et al., 2016; Winton et al., 2010). In contrast, the dependence of feedback on forcing agents due to specific SST patterns has received less attention, and an understanding of this phenomenon is only beginning to evolve. Ceppi and Gregory (2019) point out that different forcing agents cause different SST patterns with different effects on large-scale tropospheric stability, which affects lapse rate and cloud feedbacks. Gregory et al. (2020) also suggest a relation between temperature patterns and feedback to volcanic eruptions. Recently, the dependence of the feedback parameter on the forcing agent has been linked to the pattern effect using Green's functions (Zhou et al., 2023, published after the paper of appendix A). I address the lack of pattern effect studies related to feedback differences between forcing agents in chapter 3, where I identify differences in WP temperature change as primary causes for the feedback differences between CO_2 and stratospheric aerosol forcing.

2.3 STRATOSPHERIC AEROSOL VS. CO_2 FORCING

The energy balance framework and the pattern effect framework have mostly been developed around CO_2 forcings. They provide useful concepts for the treatment of the research question, in particular for linking the radiative feedback to surface temperature patterns. However, in order to find out what distinguishes the climate response to stratospheric aerosol forcing from that to CO_2 forcing on a deeper mechanistic level, it is useful to understand how exactly these forcings come about and how they differ. Therefore, I briefly introduce the main mechanisms of CO_2 and stratospheric aerosol forcing and point out

differences that could lead to the asymmetries between their climate responses.

2.3.1 *Stratospheric aerosol forcing*

Stratospheric aerosol forcing arises when sulfate aerosol is added to the background aerosol layer of the stratosphere. This can happen due to the chemical reactions to volcanically injected sulfur, or by deliberate injection of sulfur compounds[1] as a strategy to offset climate impacts from greenhouse gas warming. The radiative forcing arises because the aerosol alters the radiative balance in two ways: first, aerosol scatters incoming SW radiation, which increases the Earth's albedo. Since the troposphere only weakly absorbs SW radiation (Wild et al., 2015), the main consequence is decreased absorption of SW radiation at the surface, which constitutes a negative forcing. Second, the aerosol absorbs LW radiation, which heats the aerosol layer and causes a positive forcing due to the greenhouse effect. The SW outweighs the LW effect, hence stratospheric aerosol tends to cool the Earth. Volcanic aerosol has an e-folding lifetime in the stratosphere of about one year (Robock, 2000), and so does the resulting radiative forcing.

The last volcanic eruption that caused substantial global-mean cooling happened in June 1991, when Mt. Pinatubo on the Philippines erupted and injected an estimated 5 - 10 Tg of sulfur into the stratosphere (Timmreck, 2018) (compare this to $0.1\,\mathrm{Tg}\,\mathrm{yr}^{-1}$ in volcanically quiescent periods, Rasch et al., 2008). In the following year, Earth cooled by approximately 0.4 K (Thompson et al., 2009). A volcanic eruption's global climate impacts on time scales longer than a few weeks depend on the sulfur emission, not on the erupted magma mass or the amount of ash that is emitted (Timmreck, 2018).

2.3.2 CO_2 *forcing*

In contrast to stratospheric aerosol forcing, CO_2 forcing affects only LW radiation. CO_2 absorbs terrestrial LW radiation from the relatively warm surface and emits it from the typically colder atmosphere. The difference in emission temperatures of surface and atmosphere is responsible for the greenhouse effect. CO_2 forcing is usually associated with positive forcing because CO_2 concentrations on Earth are rising. However, decreasing CO_2 concentrations would lead to a negative forcing, similar to stratospheric aerosol forcing. Less CO_2 in the stratosphere also reduces the emissivity of the stratosphere, leading to spread out stratospheric heating (Manabe and Wetherald, 1975). CO_2

---

[1] Other aerosol types, such as alumina, calcite, or diamond, are also under consideration for climate engineering (Sukhodolov et al., 2024). This thesis exclusively treats sulfate aerosol, which is responsible for the cooling effect of volcanic eruptions, and the most commonly researched material for climate engineering.

forcing persists for up to hundreds of thousands of years (Archer and Brovkin, 2008; Inman, 2008).

At this point, the research question is posed, and the mechanisms by which CO_2 and stratospheric aerosol forcing affect the climate are clear. The necessary tools to resolve the climate response asymmetry between CO_2 and stratospheric aerosol forcing have been introduced. In the following two chapters, I summarize the results from the papers in the appendix. I establish the link between strong temperature change in the WP and strong feedback to stratospheric aerosol forcing, building upon the pattern effect framework introduced in this section (chapter 3). I then connect the WP temperatures to the stratospheric circulation changes that arise from stratospheric aerosol forcing (chapter 4).

STRONG FEEDBACK TO STRATOSPHERIC AEROSOL FORCING: THE KEY ROLE OF THE PATTERN EFFECT

The overarching goal of the thesis is to explain why stratospheric aerosol forcing causes stronger feedback than CO_2 forcing. Since the pattern effect framework has proven useful to understand variations of the feedback parameter under CO_2 forcing, I take it as a starting point for investigating the feedback differences between CO_2 and stratospheric aerosol forcing. I ask:

Can the different feedback parameters to CO_2 and stratospheric aerosol forcing be explained by the pattern effect?
Yes. In this chapter, I show how temperatures in the WP modulate the feedback parameter, and that this is the cause for the strong feedback to stratospheric aerosol forcing. For details, see the paper in appendix A.

3.1 MODEL AND SIMULATIONS

I do this by performing simulations with three different forcings in the coupled general circulation model MPI-ESM 1.2 (Mauritsen et al., 2019). The CO_2 simulations are performed under an abrupt, constant halving or doubling of CO_2 concentrations, termed *0.5 × CO₂* and *2 × CO₂*, as an analog to CMIP's *abrupt-4 × CO₂* experiments (Eyring et al., 2016). For simulating aerosol forcing, I use the "Easy Volcanic Aerosol" forcing generator (EVA, Toohey et al., 2016) to derive idealized aerosol optical properties, that describe the time-mean aerosol field in the three years after a tropical volcanic eruption. I prescribe these aerosol optical properties abruptly as a constant forcing to the *Aero* simulation. The time-constant forcing of *Aero* resembles forcing from solar radiation management rather than from a volcanic eruption. It allows an apples-to-apples comparison to the time-constant CO_2 forcings, and to identify differences between the forcing agents regardless of their time scales. For each forcing (*0.5 × CO₂*, *2 × CO₂*, *Aero*), I conduct one simulation spanning 1000 years and an ensemble of 40 simulations with independent initial conditions, each lasting 10 years. I also perform one simulation each with fixed SST over 30 years to diagnose the effective forcing.
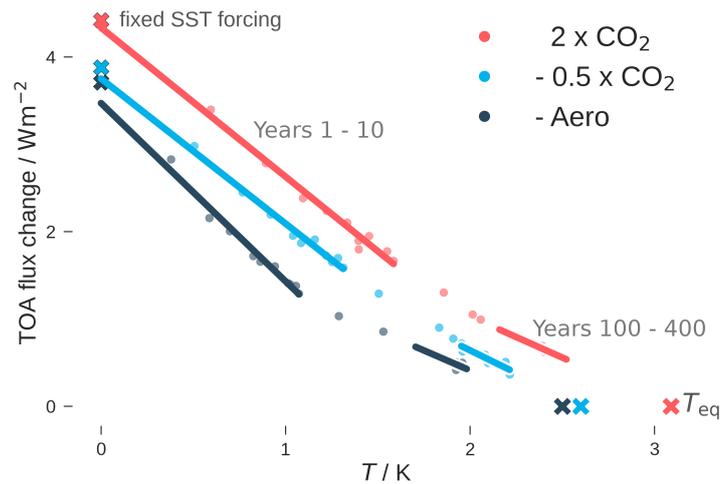
Figure 3.1: N(T) (Gregory) plot. All values from *0.5 × CO₂* and *Aero* exper-
iments are multiplied by -1 in order to show them in the same
quadrant as the $2 \times CO_2$ results. Linear regressions are shown for
the early period (years 1–10) and the late period (years 100–400).
Crosses mark the fixed SST effective forcings and the equilibrium
temperatures, which were extrapolated from a linear regression
of N(T) over years 100–1000. The first 10 points of each simulation
type are the ensemble averages from the 40-member ensembles
of the first decade. Afterwards, due to the lack of an ensemble,
each point shows 40-year averages in order to reduce the impact
of internal variability. The slope of each regression line is the
feedback parameter.

3.2   FEEDBACK IN MPI-ESM

The $N(T)$ plot is shown in Fig. 3.1. All forcings produce relatively
strong feedback in the first decade, indicated by the relatively steep
slopes in the left half of Fig. 3.1. During the first decade, stratospheric
aerosol forcing produces 23 % stronger feedback than the CO_2 forcings,
consistent with previous studies' findings (Table 2.1). On the centennial
time scale, the feedback weakens, and the feedback parameters of the
three simulations differ only within uncertainty. Fig. 3.1 shows that
in MPI-ESM feedback differences between aerosol and CO_2 forcing
cannot be purely explained by the different duration of short-term
volcanic aerosol and constant CO_2 forcing, as all forcings are constant
in my experiments. However, the figure does not provide any insights
into the underlying mechanisms.

3.3   THE KEY ROLE OF THE PATTERN EFFECT

I demonstrate in Fig. 3.2 that the pattern effect is responsible for
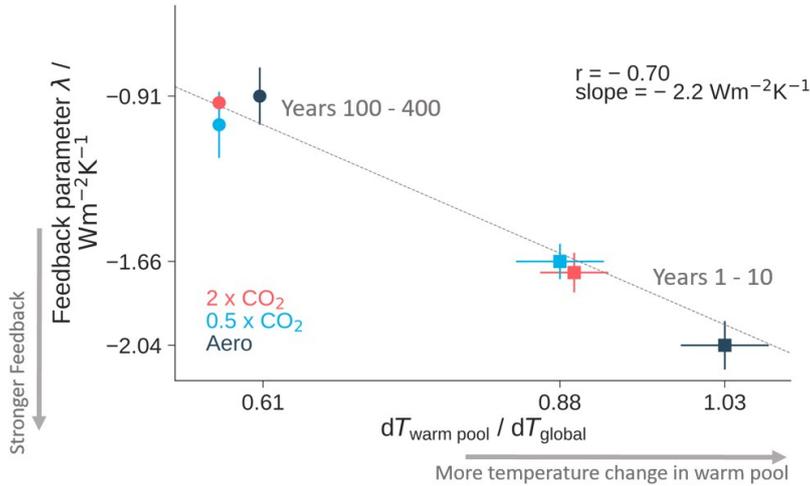the feedback differences between CO_2 and aerosol forcing. With a

Figure 3.2: Scatter plot of feedback parameter vs ratio of WP mean to global mean temperature change. Feedback parameters and associated standard errors are calculated by: regression over 10 years, followed by ensemble averaging, for years 1-10; regression over 40-year-averages for years 100-400, in order to eliminate natural variability. The gray dashed line is a linear regression through the feedback parameter values from all 3 x 40 = 120 ensemble members from the first decade.

correlation of r = –0.7, strong decadal feedback is associated with strong temperature change in the WP, which is the central result of the paper of appendix A. This finding directly relates to the feedback-strengthening processes illustrated in Fig. 2.2 and the accompanying text. In the aerosol simulations, an anomalously strongly cooling WP can effectively unfold its strong negative feedback processes over the tropical oceans. In MPI-ESM these feedback differences manifest mostly in the lapse rate feedback, partly mitigated by the opposing water vapor feedback. In contrast, the feedback dependence on time in CO_2 simulations has mostly been attributed to changing cloud feedback (Andrews et al., 2015; Dong et al., 2020).

On the centennial time scale, the temperature changes are less pronounced in the warm pool compared to the global mean, which explains the weakening feedback (Fig. 3.2). This can be attributed to delayed temperature change in the high latitudes (Armour et al., 2016; Rugenstein et al., 2016a; Senior and Mitchell, 2000), which diminishes the WP's role in setting the global mean feedback.

The WP is not the region with the largest temperature pattern differences between *Aero* and the *CO₂* forcings. Much greater differences can be found, e. g., over the Southern Ocean and the Northern Pacific (not shown here, for details see Fig. A.4). However, the local and remote radiative effects originating from unit changes in WP SST by far outweigh the radiative impact of SST in other regions (Dong et al., 2019). Therefore, even small WP differences can govern the global-

		CO$_2$	Aerosol
My Simulations		0.88 ± 0.02	1.03 ± 0.02
EVA-ENS	40 TgS		1.19 ± 0.01
	20 TgS		1.32 ± 0.01
	10 TgS		1.49 ± 0.04
	5 TgS		1.70 ± 0.58
	2.5 TgS		2.50 ± 1.67
CMIP6		0.97 ± 0.04	1.10 ± 0.03 (Krakatau)
			1.07 ± 0.05 (Pinatubo)

Table 3.1: Ratio of WP temperature change to global mean temperature change, obtained by regression. In my simulations, the regression is performed over the first ten years. In EVA-ENS and the CMIP6 $4 \times$ CO$_2$ simulations, the regression is performed over the first three years. In the CMIP6 historical simulations, regressions are performed over years 1883 - 85 (Krakatau) and 1991 - 93 (Pinatubo).

mean radiative feedback. I find that 10 % stronger temperature change in the WP relative to the global mean leads to about $0.2\,\mathrm{W\,m^{-2}\,K^{-1}}$ (about 10 %) more negative feedback. The ratio of WP to global mean warming explains about half the variance of the feedback parameter across all decadal simulations in MPI-ESM. This value is surprisingly high, given that this ratio is a very simple metric for the temperature pattern. While the WP boundaries are based on previous findings (e. g., Andrews and Webb, 2018; Dong et al., 2019), they were not tuned to maximize the correlation with the feedback parameter.

3.4   STRONG WARM POOL COOLING IN LESS IDEALIZED SIMULATIONS

Finally, I show that the key results hold not only in my strongly idealized simulations, but also in simulations with more realistic volcanic forcing conditions. I test this with the MPI "idealized volcanic forcing ensemble" (EVA-ENS Azoulay et al., 2021), which consists of simulations with volcanic aerosol forcing obtained from EVA, the same tool I used to create the idealized forcing files for my simulations. However, the EVA-ENS forcing depends on time in a way that is realistic for volcanic eruptions, and retains the seasonally enhanced transport to the winter hemisphere. There are 100 ensemble members each for injection masses between 2.5 and 40 Tg sulfur (TgS). In comparison, the Mt. Pinatubo eruption from 1991 injected an estimated 5 - 10 TgS (Timmreck, 2018). Table 3.1 shows that the EVA-ENS simulations share my simulations' behavior of strong WP temperature change compared

to the global mean, indicating that this effect persists under more realistic volcanic forcing conditions.

Not only EVA-ENS, but also the CMIP6 multi model ensemble of historical simulations corroborates my findings. In the multi-model mean, the WP temperature change normalized by the global mean is about 10 percentage points stronger in volcanically impacted years of the historical simulations, than under constant $4 \times CO_2$ forcing (Table 3.1).

Coming back to the research question, I have shown the existence of an asymmetry between the climate response to CO_2 and stratospheric aerosol forcing, using model simulations from other author teams and myself. Stratospheric aerosol forcing causes stronger feedback than CO_2 forcing, although this may only be true for the first decade of the forcing. The strong feedback arises due to pattern effects, because stratospheric aerosol forcing produces pronounced cooling of the WP.

4

WHY DOES STRATOSPHERIC AEROSOL FORCING STRONGLY COOL THE WARM POOL?

Having established that WP temperature changes are the preeminent cause for the feedback differences, another question immediately arises:

Why does stratospheric aerosol forcing strongly cool the warm pool?
In this chapter, I show that a surface perspective on the forcing must be taken to understand the temperature pattern that results from stratospheric aerosol forcing. The acceleration of the Brewer-Dobson circulation (BDC) plays a distinctive role in transporting energy from the tropical surface to the extratropics, which is a necessary ingredient to explain the enhanced WP cooling. The results that I present here are a summary of the paper in appendix B.

4.1 HYPOTHESES

I introduce hypotheses that could explain the enhanced warm pool cooling, and develop methods to test them. The most obvious hypothesis is that differences in the temperature pattern could arise from differences in the effective forcing pattern, which is typically measured at the TOA (hence called "TOA forcing" hereafter). The idea behind the TOA forcing hypothesis can be understood from Fig. 4.1 (a) and (b), which show the pattern of effective forcing at the TOA. The TOA effective forcing over the WP is 40 % stronger than the global mean in *Aero*, while this difference is only 15 % in *0.5 × CO$_2$*. In *Aero*, the tropics are subjected to stronger forcing than the global mean, due to the higher aerosol concentration and stronger insolation. It seems intuitive that the local forcing maximum in the tropics, particularly in the WP, could lead to a cooling maximum in the WP.

Alternatively, it is possible that the spectral differences of the forcings cause the pattern differences. The fact that aerosol forcing primarily affects SW radiation, while CO$_2$ interacts with LW radiation, has been speculated to be a relevant distinction (Bony et al., 2006).

Other hypotheses are centered around the stratospheric heating that arises in consequence to the aerosol's LW absorption. A warmer cold point allows more water vapor to enter the stratosphere (Joshi and Shine, 2003; Kroll et al., 2021), potentially impacting the temperature response (Lee et al., 2023). Furthermore, the local heating of the aerosol layer alters the stratospheric meridional temperature gradient, potentially affecting the polar vortex' position and strength (Azoulay
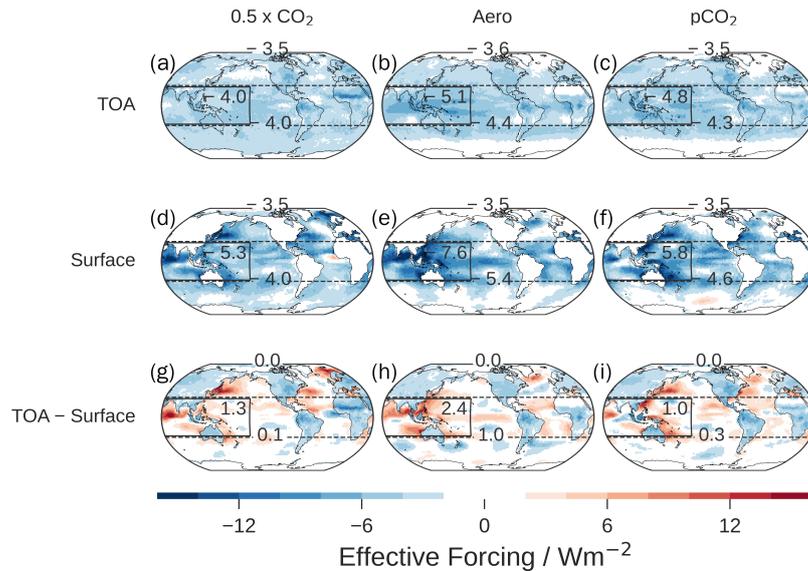
Figure 4.1: Effective forcing diagnosed at the TOA (top row), surface (middle row), and their difference (bottom row). WP and tropics are shown with solid and dashed lines, and the field average over these regions is shown in the WP box and in the dashed line, respectively. The global mean is shown at the top of each panel. Note that the surface forcing values over land are not masked, but actually (close to) zero.

et al., 2021; Bittner et al., 2016; Graf et al., 2007; Toohey et al., 2014). Model studies suggest an acceleration of the BDC as a consequence of stratospheric aerosol forcing (e. g., Aquila et al., 2013; Garcia et al., 2011; Garfinkel et al., 2017; Pitari and Mancini, 2002), but observational studies yield mixed results (e. g., Diallo et al., 2012; Graf et al., 2007; SPARC, 2022; Schnadt Poberaj et al., 2011). Within the wave-driven BDC, air enters the stratosphere in the tropics, where the forced up-welling leads to an adiabatic cooling of the environment. After moving polewards, the air descends in the extratropical stratosphere where it causes adiabatic heating (Holton et al., 1995). While there has been research on the impact of stratospheric aerosol on the stratospheric circulation (Diallo et al., 2017; Garfinkel et al., 2017; SPARC, 2022), it is unclear how this affects radiative feedback and temperature patterns.

I expect to find the origin of the temperature pattern differences in one or more mechanisms from the pool of possibilities I outlined. In this chapter I test these hypotheses with coupled and fixed-SST simulations with the MPI-ESM.

4.2   TESTING THE TOA FORCING HYPOTHESIS

First, I reject the hypothesis that the WP-enhanced TOA forcing pattern from stratospheric aerosol is the sole cause of the WP-enhanced temperature change. If the TOA forcing pattern hypothesis were true,
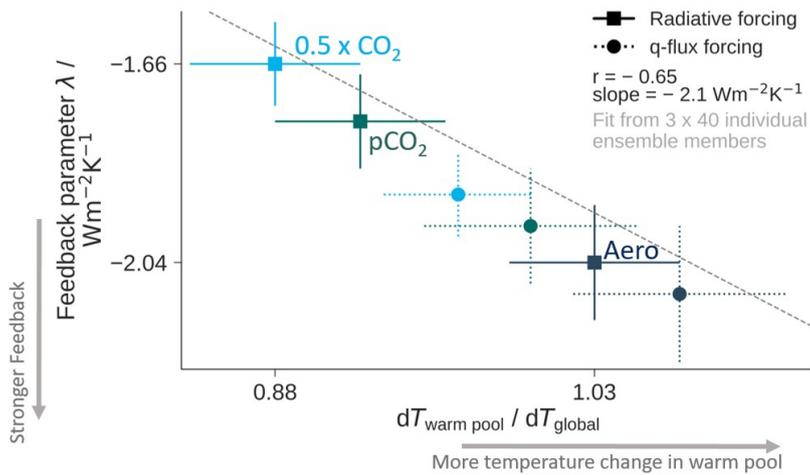
Figure 4.2: Scatter plot of feedback parameter vs ratio of WP mean to global mean temperature change. Squares represent results from the radiatively forced simulations (described in chapter 3. The circles are from the q-flux-forced simulations, which I discuss in section 4.6 Feedback parameters and associated standard errors are calculated by regression over 10 years, followed by ensemble averaging. The gray dashed line is a linear regression through the feedback parameter values from all 3 x 40 = 120 ensemble members from the first decade.

a CO_2-forced simulation with a WP-enhanced TOA forcing pattern would also produce stronger temperature changes in the WP. I develop such a "patterned CO_2" simulation (pCO_2[1]) where CO_2 concentrations are not a single number, but a field, which depends on longitude and latitude. I iteratively adjust the CO_2 concentrations until they match the target forcing field from the stratospheric aerosol. Since the aerosol forcing varies seasonally due to changing insolation, the CO_2 field depends not only on longitude and latitude, but also on the month. The pCO_2 simulation has generally reduced CO_2 concentrations (since the mean forcing is negative), which vary considerably. The resulting TOA effective forcing field of pCO_2 is similar to that of *Aero*, but not quite as WP-enhanced (Fig. 4.1), which I address later. Although unrealistic, making CO_2 not well mixed provides a direct test of the TOA forcing pattern hypothesis.

Fig. 4.2 shows the first decade of the *0.5 × CO₂*, *pCO₂* and *Aero* simulations in the space of WP temperature change and feedback which I developed in chapter 3 (compare also Fig. 3.2). Although the TOA forcing pattern of pCO_2 is almost as WP-enhanced as that of *Aero*, its temperature pattern and consequently the feedback rather resemble *0.5 × CO₂*. The pCO_2 simulation's output follows the relationship between the temperature change in the WP and the feedback

---

1 In this thesis, pCO_2 always refers to the patterned CO_2 simulations, and never to the otherwise common partial pressure of CO_2

parameter established in Fig. 3.2, corroborating the results of chapter 3.

A statistical comparison of the ratios of WP-mean temperature change to global-mean temperature change underpins this result quantitatively. I perform Student's t-tests to test the null hypothesis that these ratios (1 value per ensemble member, i. e. 40 for each forcing type) are drawn from the same distribution. The null hypothesis is rejected with $p = 10^{-6}$ when comparing *Aero* and pCO_2, but is not rejected when comparing *0.5 × CO$_2$* and pCO_2 ($p = 0.2$). The significant difference between pCO_2 and *Aero* persists even when correcting for the fact that the forcing field of pCO_2 is not quite as WP-enhanced as that of *Aero*, albeit at a higher p-value ($p = 0.02$, for details see appendix A).

This analysis shows that despite their similar TOA forcing patterns, pCO_2 and *Aero* produce considerably different temperature patterns (Fig. 4.2), in conflict with the TOA forcing hypothesis. This indicates the existence of another aerosol-specific process that contributes to the temperature pattern differences.

4.3   THE ROLE OF SURFACE FORCING

The hypothesis was grounded in the idea that stronger forcing in the WP could lead to stronger temperature change in the WP. Forcing is traditionally measured at the TOA (Forster et al., 2016). However, the surface temperatures respond to fluxes at the surface, not fluxes at the TOA. I present results that show that the surface forcing pattern can better explain the discrepancy between the temperature patterns. Effective forcing is traditionally defined at the TOA, but can be diagnosed at the surface as well. While there are no turbulent fluxes (sensible heat, latent heat) at the TOA, they must be taken into account for the surface forcing.

The surface effective forcing of *Aero* is twice as strong in the WP as in the global mean, and thus even more WP-enhanced than the TOA effective forcing. In comparison, the surface effective forcing of *0.5 × CO$_2$* and pCO_2 is only slightly stronger in the WP than in the global mean. Before showing that this discrepancy indeed explains *Aero*'s amplified WP temperature change, I explain the processes that cause the differences between surface forcing and TOA forcing.

The difference between TOA forcing (F_{TOA}, first row of Fig. 4.1) and surface forcing ($F_{surface}$, second row of Fig. 4.1) is the forcing on the atmosphere (F_{atm}, third row of Fig. 4.1). F_{atm} has a pattern because the atmosphere transports energy and thus redistributes forcing. However, it is zero in the global mean, because the atmosphere cannot act as a substantial energy source or sink on long time scales, due to its small heat capacity. A similar argument can explain why surface forcing is zero over land, which has a small heat capacity, too (see second row

of Fig. 4.1). A key ingredient for the following arguments is the fact that the difference between TOA and surface forcing is balanced by atmospheric energy transport, i.e. the horizontal heat flux divergence Q.

$$F_{\text{TOA}} - F_{\text{surface}} = F_{\text{atm}} \tag{4.1}$$
$$F_{\text{atm}} + Q = 0 \tag{4.2}$$

Both sides of Eq. 4.1 globally average to zero. In *Aero*, the atmospheric energy transport out of the WP is slightly more than $1\,\text{W}\,\text{m}^{-2}$ stronger than in the CO_2 simulations (compare WP averages in the lower row of Fig. 4.1). Hence, the surface forcing in *Aero* is especially strong in the WP, because the atmosphere effectively transports energy out of the WP to other regions, particularly to the extratropics. I separate the energy transport anomaly into a meridional and a tropical-zonal component. The meridional component represents the energy transport from the tropics to the extratropics, while the tropical-zonal component represents the energy transport from the WP to the tropical non-WP regions.

The energy transport distinctions between *Aero* and the CO_2 simulations arise mainly from the meridional component, which amounts to $1\,\text{W}\,\text{m}^{-2}$ in *Aero*, but only $0.1\,\text{W}\,\text{m}^{-2}$ and $0.3\,\text{W}\,\text{m}^{-2}$ in *0.5 × CO2* and *pCO2*, respectively. Large-scale meridional energy transport is typically associated with the Hadley circulation and eddies. Surprisingly, they do not play major roles in effectuating the anomalous energy transport that arises from stratospheric aerosol forcing. Instead, the BDC turns out to be a key player that greatly affects the surface forcing, despite acting in the stratosphere.

4.4 MERIDIONAL ENERGY TRANSPORT VIA AN ACCELERATED BREWER-DOBSON CIRCULATION

I argue that the meridional energy transport in *Aero* mainly arises from an acceleration of the BDC, which comes to pass as an adjustment irrespective of surface temperature changes. Within the climatological BDC, air rises in the tropical stratosphere, then moves poleward, and descends in the extratropical stratosphere. In the upwelling branch, the expanding air causes adiabatic cooling K (in $\text{K}\,\text{s}^{-1}$) of the environment. The cooling is proportional to the residual mean vertical velocity $\bar{w}^*$ (i.e., the strength of the BDC), and the difference between the background lapse rate $\frac{\partial T}{\partial z}$ and the dry adiabatic lapse rate $-\frac{g}{c_{\text{p}}}$ (Birner and Charlesworth, 2017):

$$K = -\bar{w}^* \left(\frac{\partial T}{\partial z} + \frac{g}{c_p} \right) \tag{4.3}$$

$$\tag{4.4}$$

	$0.5 \times CO_2$	Aero	pCO_2
ΔQ_{adi} / Wm^{-2}	0.0	-0.9	-0.1
ΔQ_{adi} due to $\Delta \bar{w}^*$ / Wm^{-2}	0.2	-0.7	0.1
ΔQ_{adi} due to $\Delta \frac{\partial T}{\partial z}$ / Wm^{-2}	-0.2	-0.2	-0.2
$-$ Meridional energy transport / Wm^{-2} (from Fig. 4.1 lower row)	-0.1	-1.0	-0.3

Table 4.1: Anomalous adiabatic cooling in the stratosphere at fixed SST, calculated according to Eq. 4.5, and averaged over the tropics (30°N to 30°S). The second and third rows show the contributions from changes in the upwelling speed and the lapse rate.

where g is the gravitational acceleration and c_p the specific heat capacity of dry air. The residual mean vertical velocity $\bar{w}^*$ is obtained from a transformed Eulerian mean analysis (e.g. Butchart, 2014). The analysis here is restricted to the tropics, where the influence of eddies is small, and therefore $\bar{w}^*$ is similar to the vertical velocity w. The adiabatic cooling rate can be converted to a heat flux convergence due to adiabatic cooling Q_{adi} (in W m^{-2}). This is accomplished by integrating the adiabatic cooling over the stratosphere and using the hydrostatic approximation $\rho dz = -dp/g$:

$$Q_{\text{adi}} = -\frac{c_p}{g} \int_{100 \text{ hPa}}^{1 \text{ hPa}} K(p) \, dp \qquad (4.5)$$

Using a first-order Taylor expansion, the changes in K can be further decomposed into contributions from changes in $\bar{w}^*$ and $\frac{\partial T}{\partial z}$.

The tropically averaged anomalous adiabatic cooling within the BDC's rising branch amounts to 0.9 W m^{-2} in *Aero* (Table 4.1). Globally, the BDC is not an energy sink. The same energy is released as adiabatic heating in the sinking branch of the BDC in the extratropics. The described process therefore constitutes an energy transport from the tropics to the extratropics. In *Aero*, the energy transport via the BDC explains the meridional atmospheric energy transport of 1 W m^{-2} seen in Fig. 4.1 almost entirely. Partitioning the anomalous adiabatic cooling into contributions from changes in lapse rate and vertical velocity shows that the effect mainly arises from an acceleration of the BDC (Table 4.1). The energy loss from adiabatic cooling in the tropical stratosphere is communicated to the troposphere via radiation. If there was no adiabatic cooling, the tropical stratosphere would heat up even more and cause additional downward LW radiation which would heat the tropical troposphere, and hence the surface. In contrast to *Aero*, in the CO$_2$ simulations there is only a small adiabatic cooling in the tropics, and hence little meridional energy transport.

4.5 COMPARISON TO OTHER MODELS AND OBSERVATIONS

Is this mechanism specific to MPI-ESM or shared among climate models? The additional adiabatic cooling from the BDC also appears in post-eruption years in historical simulations of the CMIP6 multi model ensemble. Since I restrict the CMIP6 analysis to the tropics where the eddy contribution is small, I simplify it by using w instead of $\bar{w}^*$, which allows me to work with monthly-mean instead of daily or 6-hourly output. In MPI-ESM, this causes an underestimation of $\approx$ 10 %, and I expect similar errors in CMIP. Compared to the piControl simulation, adiabatic cooling of the tropical stratosphere in the years following the Krakatau eruption and Pinatubo eruption is stronger than in the control simulation by $(0.12 \pm 0.02)\,\mathrm{W\,m}^{-2}$ and $(0.13 \pm 0.03)\,\mathrm{W\,m}^{-2}$, respectively (multi-model mean ±standard error). The additional adiabatic cooling of the post-eruption years amounts to $0.16\,\mathrm{W\,m}^{-2}$ and $0.25\,\mathrm{W\,m}^{-2}$ when restricting the analysis to models with at least 10 realizations in order to minimize the influence of internal variability.

It follows that the acceleration of the BDC, arising as an adjustment to stratospheric aerosol forcing, causes a substantial energy export from the tropics to the extratropics. This leads to a pronounced negative surface forcing in the tropics. The findings that I obtain with the MPI-ESM simulations qualitatively agree with the CMIP6 multi model mean and modeling studies, which have consistently shown an acceleration of the BDC following volcanic eruptions (e. g., Aquila et al., 2013; Garcia et al., 2011; Garfinkel et al., 2017; Pitari and Mancini, 2002).

Observational and reanalysis studies show mixed results. Some studies find enhanced wave activity after volcanic eruptions (Graf et al., 2007; Schnadt Poberaj et al., 2011), which can cause an acceleration of the BDC. Others do not find BDC anomalies (Diallo et al., 2012; SPARC, 2022; Seviour et al., 2012). However, reanalysis products may not be suitable to study the BDC because they do not assimilate stratospheric aerosol, leaving them ignorant to heating rate anomalies (Abalos et al., 2015). Furthermore, observations might miss the height of the maximum vertical velocity change (Toohey et al., 2014).

4.6 STRONGER SURFACE FORCING IN THE WP CAUSES STRONGER LOCAL COOLING

In order to link the results on the forcing pattern to the temperature pattern and therefore the feedback (see chapter 3), it remains to be shown that the enhanced forcing at the WP surface causes enhanced temperature change at the WP surface. This may seem obvious. However, a localized surface flux can lead to much stronger remote than local surface temperature responses (e. g., Lin et al., 2021; Liu

et al., 2018a,b, 2022). I establish the link between locally enhanced forcing and temperature change in the WP using a separate set of simulations which are forced by surface fluxes instead of radiative perturbations. The surface forcings diagnosed from the *0.5 × CO₂*, *Aero*, and *pCO₂* simulations (Fig. 4.1 middle row) are prescribed as a heat source / sink (q-flux forcing, or "ghost forcing" Hansen et al., 1997), but no changes are made to the atmospheric aerosol or CO_2 concentrations. If stronger surface forcing in the WP is the reason for stronger temperature changes in the WP, then this effect should appear in these q-flux-forced simulations as well. Indeed, Fig. 4.2 shows that simulations with stronger WP surface forcing also cause stronger WP temperature change (*Aero > pCO₂ > 0.5 × CO₂*). However, the q-flux-forced simulations generally produce stronger temperature change in the WP than their radiatively forced counterparts. Although knowledge of the surface forcing is not enough to correctly simulate the temperature patterns, the q-flux-forced simulations support the hypothesis that strong surface forcing in the WP leads to strong temperature change in the WP.

CLOSING THOUGHTS

5.1 SUMMARY

In the beginning of chapter 2 I motivated the thesis with the question how much the Earth changes its temperatures in reaction to radiative forcing. I delineated the current understanding of climate feedback and pointed out an apparently paradoxical asymmetry between the climate response to CO_2 and stratospheric aerosol forcing. Resolving this peculiarity not only holds significance for gaining deeper insights into the climate response to stratospheric aerosol forcing, but can also advance our understanding of climate feedback and the pattern effect in broader terms. Consequently, with the objective of adding to the discourse surrounding forcing, feedback, and the pattern effect, I set out to answer the overarching research question

Why does stratospheric aerosol forcing cause stronger feedback than CO_2 forcing?

I formulated mechanisms and research hypotheses which I tested with the general circulation model MPI-ESM 1.2. In the beginning of this section, I summarize my contribution to answering the main research question, guided by the two sub-research questions that I formulated in this thesis. I then discuss the broader impact on comprehending climate feedback and the pattern effect, and finally present research avenues that I deem promising for closing the remaining gaps.

Can the different feedback parameters to CO_2 and stratospheric aerosol forcing be explained by the pattern effect?

Yes. I show that 50 % of the variance in feedback parameter is explained by a simple index for the temperature pattern, the ratio of warm pool to global mean temperature change, in MPI-ESM. In agreement with most modelling studies, I find stronger feedback to stratospheric aerosol forcing than to CO_2 forcing. In MPI-ESM, feedback to stratospheric aerosol is time-dependent and weakens considerably on a decadal to centennial time scale, so that the feedback differences between stratospheric aerosol and CO_2 forcing are only pronounced within one decade. The warm pool plays a superior role in setting Earth's global mean feedback. Warm pool temperatures react more strongly to stratospheric aerosol than CO_2 forcing, which allows the strong stabilizing feedbacks that originate from the warm pool to

unfold efficiently.

Why does stratospheric aerosol forcing strongly cool the warm pool?

The strong warm pool cooling due to stratospheric aerosol is partially explained by the pronounced negative forcing at the TOA, resulting from high aerosol load, high insolation, low background albedo, and cloud adjustments. However, this explanation, albeit intuitive, does not suffice. The pattern of surface temperature change is more closely linked to the pattern of forcing at the surface than at the TOA. Stratospheric aerosol causes particularly strong surface forcing in the warm pool, and consequentially a strong local surface cooling. The Brewer-Dobson circulation plays a major role in this process. In climate models, the Brewer-Dobson circulation accelerates due to the differential stratospheric heating, and transports additional energy from the tropical surface to the extratropics via the stratosphere.

The identified key processes, such as the strong warm pool temperature change and enhanced adiabatic cooling from stratospheric aerosol forcing, are also present in the CMIP6 multi model ensemble. However, observational evidence is weak (see also chapter 4). The only climatically relevant eruptions since 1850 are those of Krakatau (1883), Katmai (1912), Agung (1963), El Chichon (1982), and Pinatubo (1991) (Timmreck, 2018), the latter being the only one in the satellite era. Temperature patterns are naturally so variable, e.g., due to ENSO, that a few volcanic eruptions do not suffice to make a clear statement. In any case, if we believe that stratospheric aerosol forcing causes an acceleration of the Brewer-Dobson circulation, then the results from this thesis show that the consequent adiabatic cooling anomaly causes pronounced negative forcing at the tropical surface. However, if the acceleration of the Brewer-Dobson circulation turns out not to be a real-world phenomenon, the consequences for the surface forcing pattern and even the strong feedback to stratospheric aerosol forcing itself may have to be reevaluated. It should also be kept in mind that the aerosol in my model is not transported, and described by a highly idealized profile without seasonal dependence. However, I don't expect this to impact the results qualitatively. Richter et al. (2017) find only small differences in stratospheric upwelling and temperatures when comparing aerosol-forced simulations in a single model with and without interactive ozone chemistry.

5.2 CONCLUSIONS

Moving forward, what is the broader significance of this thesis' findings? The first motivation I stated was to shed light on the climate response to stratospheric aerosol forcing, in particular to resolve the

climate response asymmetry between CO_2 and stratospheric aerosol forcing. My results spotlight the warm pool's prominent role for feedback differences between CO_2 and stratospheric aerosol forcing, which are only pronounced on decadal time scales. This implies that solar geoengineering by stratospheric aerosol injection might get more effective over time, as the high latitude temperature change strengthens at the expense of changes over the warm pool. However, since the explanations are based on temperature change patterns, it is unclear how they carry over to a world of stagnating temperatures. Stratospheric heating has already been recognized as a climatically relevant adjustment to stratospheric aerosol forcing with impacts on, e. g., the tropical overturning circulation (Ferraro et al., 2014), midlatitude circulation, and the hydrological cycle (Simpson et al., 2019). My results add to this by pointing out how the ensuing acceleration of the Brewer-Dobson circulation impacts the troposphere and surface climate. Additionally, this implies that solar dimming is a deficient substitute for modelling stratospheric aerosol forcing, which has been argued before, but on different grounds (Ferraro et al., 2014; Simpson et al., 2019; Visioni et al., 2021).

I had also suggested that investigating the asymmetry between the climate response to CO_2 and stratospheric aerosol forcing may provide insights into climate feedback in general. Has this goal been achieved? I show that feedback to stratospheric aerosol forcing is different from feedback to CO_2 forcing and that this feedback difference changes over time. Consequently, even if there were enough observations of volcanic eruptions to eliminate the influence of internal variability, they could not be used to constrain equilibrium climate sensitivity (in agreement with, e. g., Gregory et al., 2020; Merlis et al., 2014; Pauling et al., 2023; Stowasser et al., 2006; Wigley et al., 2005). Furthermore, my findings highlight the role of surface forcing for the formation of temperature change patterns, in agreement with Kang and Xie (2014). By contrasting the surface forcing pattern to the TOA forcing pattern, I challenge previous studies suggesting that the TOA forcing pattern might explain feedback to aerosol forcing (Salvi et al., 2022, 2023), solar forcing (Kaur et al., 2023; Modak et al., 2016), and temperature change patterns in general (Liu et al., 2022). The established method of diagnosing effective forcing at the TOA is useful for global means, but not sufficient for understanding the formation of surface temperature patterns. Differences between the effective forcing at the TOA and at the surface arise from circulation adjustments. The anomalous circulations produce atmospheric heat transport changes, which can redistribute forcing between different regions of the Earth. Future studies investigating the relationship between forcing patterns and temperature change patterns should take this finding into consideration. The fact that the Brewer-Dobson circulation can affect temperatures at the surface reveals an important aspect of stratosphere-troposphere-surface

coupling, highlighting the importance of the stratosphere for surface processes. It is interesting to note that knowledge of the surface forcing is not sufficient to reconstruct the exact temperature change pattern (Fig. 4.2), although this was shown to be the case in an aquaplanet slab ocean simulation (Haugstad et al., 2017). This may be related to the lack of surface forcing over land, or the need to prescribe also other air-sea fluxes such as momentum or freshwater flux to the dynamical ocean. CESM2 simulations with historical forcing point towards a critical role of wind stress forcing (McMonigal et al., 2023).

In chapter 2 and throughout this thesis I have shown attempts to explain the non-constancy of the feedback parameter, and to reconcile Eq. 2.1 ($N = F + \lambda T$) with models and reality. Taking a step back, one can ask whether the energy balance framework is at all a useful way of thinking about the climate response to forcing, given all the weaknesses and inconsistencies. I do think it is. Linearization is often our best chance to get a grasp of non-linear processes. Furthermore, the energy balance framework has been very successful, especially when taking into account how simple it is. The challenge lies in maximizing improvements to the skill with minimal modifications, such as separating the Earth into Warm Pool and non-Warm Pool regions. The quest to reconcile reality with $N = F + \lambda T$ has led to so many discoveries in Earth system science (see chapter 2), and keeps doing so, that it seems unreasonable to abandon this tool. Nevertheless, it is essential to bear in mind its limitations and take it for no more than it is - a simple yet relatively good approximation.

5.3   LOOKING AHEAD

Which challenges lie ahead? Drawing upon the energy balance framework, enormous progress has been made in understanding the links between temperature patterns and TOA radiative flux over the last two decades. Notable advances on understanding how SST patterns affect feedbacks come from approaches with SST Green's functions (Dong et al., 2020, 2019; Williams et al., 2023; Zhou et al., 2017). I am optimistic about the forthcoming Green's Function Model Intercomparison Project (Bloch-Johnson et al., 2024) to further our understanding how feedback, circulation, and the hydrological cycle depend on SST patterns.

While the link between SST and feedback is understood increasingly well, deficiencies in simulated SST patterns pose a significant barrier for progress. Coupled climate models still systematically fail to reproduce the observed temperature pattern (Rugenstein et al., 2023a; Seager et al., 2019; Wills et al., 2022). This limits the credibility of the models' projected future SST patterns, adding substantial uncertainty to estimates of equilibrium climate sensitivity (Alessi and Rugenstein, 2023; Andrews et al., 2018; Rugenstein et al., 2020; Sherwood et al.,

2020), and hampering adaptation and mitigation efforts (Lee et al., 2022). There is an urgent need to improve climate models to produce temperature patterns consistent with observations, but also to gain process understanding of the mechanisms governing temperature pattern formation in the tropical Pacific.

Considerable progress with respect to understanding temperature pattern formation has been achieved with q-flux Green's functions (Lin et al., 2021; Liu et al., 2018a,b, 2022) and pacemaker experiments (Kang et al., 2023; Zhang et al., 2019). This strand of research generally draws attention to the importance of high latitude ocean heat uptake, emphasizing the role of the Southern Ocean. However, a comprehensive framework for temperature pattern formation is still missing, but highly desirable in order to understand past and future climate change. The research community should aim to develop a physical understanding about what future temperature patterns we expect, and how exactly they would affect climate feedback. This includes resolving the discrepancy of simulated versus observed historical SST patterns (Seager et al., 2019; Wills et al., 2022). The advent of global storm resolving models that explicitly simulate convection and ocean eddies may help improve the representation of processes that are critical for pattern formation. Correctly simulating deep convection is important for processes relevant for temperature pattern formation (Sohn et al., 2016), but currently deficient across models due to the limitations of convective parameterizations (Holloway et al., 2012). Preliminary results from higher-resolution (although not convection-resolving) historical simulations already show an encouraging improvement of the simulated Southern Ocean and Eastern tropical Pacific SST pattern, which is attributable to the resolution increase (DiNezio, 2023).

If (when) we succeed to establish one comprehensive framework for temperature pattern formation and pattern effects, we can claim to understand the Earth system a whole lot better than before, and will be able to make more accurate and physically founded predictions about near- and long-term climate change. Such a framework will have to include atmosphere (Dong et al., 2019), ocean (Rugenstein et al., 2016b), air-sea interactions (Hu et al., 2022), and potentially land processes (Bloch-Johnson et al., 2024). This thesis hopefully contributes a small piece to this huge puzzle: by highlighting the warm pool's role for the pattern effect of stratospheric aerosol forcing; by identifying circulation adjustments that cause differences between TOA and surface forcing, stressing the importance of the latter; and by revealing a pathway that couples TOA forcing, the atmosphere, and surface forcing.

Part II

APPENDIX

STRONG FEEDBACK TO STRATOSPHERIC AEROSOL FORCING: THE KEY ROLE OF THE PATTERN EFFECT

This work was published as:

STRONG FEEDBACK TO STRATOSPHERIC AEROSOL FORCING: THE KEY ROLE OF THE PATTERN EFFECT

Moritz Günther[1], Hauke Schmidt[1], Claudia Timmreck[1], Matthew Toohey[2]

[1]*Max Planck Institute for Meteorology, Hamburg, Germany*
[2]*Institute of Space and Atmospheric Studies, University of Saskatchewan, Saskatoon, Canada*

ABSTRACT

Volcanic aerosol forcing has previously been found to cause a weak global mean temperature response, compared to CO_2 radiative forcing of equal magnitude: its efficacy is supposedly low, but for reasons which are not fully understood. In order to investigate this, we perform idealized, time-invariant stratospheric sulfate aerosol forcing simulations with the MPI-ESM-1.2 and compare them to $0.5 \times CO_2$ and $2 \times CO_2$ runs. While the early decades of the aerosol forcing simulations are characterized by strong negative feedback (i.e. low efficacy), the feedback weakens on the decadal to centennial time scale. Although this effect is qualitatively also found in CO_2-warming simulations, it is more pronounced for stratospheric aerosol forcing. The strong early and weak late cooling feedbacks compensate, leading to an equilibrium efficacy of approximately one in all simulations. The $0.5 \times CO_2$ cooling simulations also exhibit strong feedback changes over time, albeit less than in the idealized aerosol forcing simulations. This suggests that the underlying cause for the feedback change is not exclusively specific to aerosol forcing. One critical region for the feedback differences between simulations with negative and positive radiative forcing is the tropical Indo-Pacific warm pool region ($30°S$ - $30°N$, $50°E$ - $160°W$). In the first decades of cooling, the temperature change in this region is stronger than the global average, while it is stronger outside of it for $2 \times CO_2$ warming. In cooling scenarios, this leads to an enhanced activation of the warm pool region's strongly negative lapse rate feedback.

SIGNIFICANCE STATEMENT

Large volcanic eruptions can enhance the scattering aerosol layer in the stratosphere, which leads to a global cooling for a few years.

Surprisingly, the earth has been found to cool less from radiative flux perturbations from stratospheric aerosol forcing, compared to how much it warms due to increases in CO_2 concentration. We find that specific surface temperature change patterns after volcanic eruptions cause this effect. The temperature change in the tropical Indian and West Pacific Ocean determines how much global temperature change is needed to regain radiative equilibrium. Our findings contribute to understanding the climate response to volcanic eruptions, and are relevant for understanding the mechanisms of climate change due to changes in CO_2 concentration.

A.1 INTRODUCTION

Following large volcanic eruptions, volcanic aerosols are added to the natural stratospheric aerosol layer and persist for a few years. The aerosol layer globally increases the albedo in the shortwave (SW) and absorption in the longwave (LW) spectrum. This constitutes a net negative radiative forcing, which cools the earth. We study volcanic aerosol forcing (VAF) in the broader sense of stratospheric sulfate aerosol forcing (SSAF), independent of the volcano-specific characteristic forcing evolution in time.

VAF has been found to produce smaller temperature change per unit radiative forcing than CO_2 forcing (Boer et al., 2006; Ceppi and Gregory, 2019; Gregory and Andrews, 2016; Gregory et al., 2020, 2016; Hansen et al., 2005; Marvel et al., 2016; Modak et al., 2016; Zhao et al., 2021). This is equivalent to the statement that the feedback parameter is more negative for VAF than for CO_2 forcing. In model simulations with a slab ocean, and with fixed sea surface temperatures (SST) and sea ice, the strongly negative feedback parameter has been shown to be connected to differences in atmospheric stability, which ultimately arise from different SST patterns (Ceppi and Gregory, 2017, 2019). We aim to compare the climate feedbacks to SSAF and positive and negative CO_2 radiative forcing on annual to centennial time scales in a coupled atmosphere-ocean model, explain the differences between them, and identify the key regions that cause the distinct climate responses.

First, we introduce the necessary building blocks for our explanation: Efficacy, feedback parameter, and the pattern effect (Section A.2). This is followed by a literature review, showing that the strong feedback to VAF is a broad consensus from earlier studies (Section A.3).

In order to tackle the question why the feedback parameter appears to be stronger for VAF, we perform three types of simulation with the coupled climate model Max Planck Institute Earth System Model 1.2 (MPI-ESM 1.2, Mauritsen et al. (2019)): Abrupt $0.5 \times CO_2$, abrupt $2 \times CO_2$ and an idealized SSAF simulation (Section A.4). With these simulations, we can not only shed light on the differences between

the responses to stratospheric aerosol and CO_2 forcing, but also more generally elaborate disparities between the responses to negative and positive radiative forcing.

We show the different time-dependencies of the feedback parameter for simulations with negative and positive radiative forcing in Section A.5 and establish a relationship to the pattern effect. Thereafter we discuss the origin of the modified temperature change pattern. Finally, we shift the focus back from SSAF to VAF in order to show that our results also hold for more realistic simulations of volcanic eruptions. For this purpose, we compare our simulations to a large ensemble of volcanically forced simulations (MPI EVA ensemble, Azoulay et al. (2021)) and simulations from the Coupled Model Intercomparison Project phase 6 (CMIP6, Eyring et al. (2016), Section A.6).

A.2 CONCEPTS

Efficacy and Feedback

We describe the increased feedback parameter of VAF following Hansen et al. (2005): VAF has been claimed to have lower than unity efficacy, where efficacy is the ratio of global mean near-surface air temperature change (T) per unit radiative forcing (F) of VAF and $2 \times CO_2$ forcing:

$$\text{Efficacy} = \frac{T/F}{T_{2\times CO_2}/F_{2\times CO_2}} \tag{A.1}$$

A forcing agent produces less temperature change than CO_2 forcing of equal magnitude if the efficacy is lower than one.

The concept of efficacy is closely related to the feedback parameter $\lambda = dN/dT$, where N is the top of the atmosphere (TOA) flux anomaly. Forcing, feedback and surface temperature are linked by the linearized equation $N = F + \lambda T$. A more negative feedback parameter λ indicates a more stable climate system, since for a strongly negative feedback less surface temperature change is required to offset a given radiative forcing. Therefore, the efficacy of a forcing in equilibrium can be computed as the ratio of its inverse feedback parameter to the inverse feedback parameter of a $2 \times CO_2$ forcing (see Eq. (A.1), Zhao et al. (2021)):

$$\text{Efficacy} = \frac{\lambda_{CO_2}}{\lambda} \tag{A.2}$$

Rugenstein and Armour (2021) discuss three closely related, but distinct definitions of the feedback parameter, of which only two are relevant to this work: the equilibrium and the differential feedback parameter. For an illustration, see their Table 1.

The equilibrium feedback parameter is defined as the ratio of forcing and equilibrium temperature change and therefore compares two equilibrated states. The differential feedback parameter is obtained by a regression of $N(T)$ over a certain number of years. Its value depends on the chosen period for the regression and can differ more from the equilibrium feedback parameter the more $N(T)$ deviates from a linear relationship.

While Hansen et al. (2005) define efficacy with respect to the equilibrium efficacy, we use the term in a broader sense as a ratio of feedback parameters in general. Efficacy can depend sensitively on the employed feedback parameter definition. We use the words *equilibrium efficacy* and *differential efficacy* to refer to efficacies calculated with Eq. (A.2), using the definitions of the equilibrium feedback parameter and the differential feedback parameter, respectively.

The pattern effect

Variations of the feedback parameter have not only been investigated with respect to the forcing agent, such as CO_2, VAF/SSAF, etc., but also with respect to changes over time in CO_2 step forcing experiments (Senior and Mitchell, 2000). For most climate models, a weakening of feedback over time has been reported (Andrews et al., 2015, 2012; Armour, 2017). N has been found to not only depend on the global average T, but also on its pattern, which gives rise to the term "pattern effect" (Stevens et al., 2016). The pattern of SST anomalies is subject to changes over time (Andrews et al., 2018; Armour et al., 2013; Ceppi and Gregory, 2019; Dong et al., 2019; Zhou et al., 2016), therefore the feedback parameter changes with time, too. Indeed, the feedback parameter spread in CMIP5 and CMIP6 models can be explained by SST pattern differences (Dong et al., 2020). This has been linked to changes in cloud feedback and lapse rate feedback (Andrews et al., 2018, 2015; Andrews and Webb, 2018; Ceppi and Gregory, 2019; Stevens et al., 2016; Zhang et al., 2010), which depend on the spatial distribution of SST.

A useful framework to understand the pattern effect is provided by Armour et al. (2013), who suggest that some regions intrinsically exhibit stronger feedback than others, independent of time or state. Only the temperature change pattern varies in time and activates the local feedbacks differently. The global feedback is the temperature change-weighted average over the local feedbacks. Therefore, the global feedback gets weaker, as the temperature change gets relatively stronger in regions with weak local feedback (Rugenstein et al., 2016a). Zhou et al. (2017) and Dong et al. (2019) expand on this view by using a Green's Function approach, which accounts for the fact that local SST perturbations can modify feedbacks also in remote regions (Andrews

and Webb, 2018; Ceppi and Gregory, 2017; Lin et al., 2019). In this work, we will refer mostly to this framework.

There is evidence that the strong feedback to VAF could be related to the specific temperature change pattern after volcanic eruptions. Gregory and Andrews (2016) propose that the variations of the differential feedback parameter during the historical period could be related to specific SST patterns following volcanic eruptions, among other reasons. The efficacy of VAF has been connected to changes in tropospheric stability, which arise from different SST patterns and lead to changes in lapse rate and cloud feedback (Ceppi and Gregory, 2017, 2019). Similarly, the efficacy of anthropogenic aerosol forcing has been interpreted with respect to changes in stability which arise from specific SST patterns (Salvi et al., 2022).

More generally, the feedback parameter was found to depend on the latitude of the radiative forcing: Forcing in higher latitudes produces weaker feedbacks (Ceppi and Gregory, 2019; Forster et al., 2000; Hansen et al., 1997; Rose et al., 2014; Rose and Rayborn, 2016; Rugenstein et al., 2016a), possibly because of the resulting high-latitude temperature change pattern (Haugstad et al., 2017; Salvi et al., 2022). In a multi-model analysis Po-Chedley et al. (2018) derive a theoretical basis for a relationship between the meridional temperature change pattern and the global feedback parameter, which was already found by Soden and Held (2006). In the model of Po-Chedley et al. (2018) more tropical temperature change leads to stronger global mean lapse rate and water vapor feedback. More specifically, Dong et al. (2019) show that it is the warm pool region of the tropical Indian Ocean and the tropical Western Pacific Ocean that dominates the negative feedback. There, the strong convective coupling yields a strongly negative lapse rate feedback due to the moist adiabatic lapse rate (Andrews and Webb, 2018; Bintanja et al., 2012; Manabe and Wetherald, 1975; Pithan and Mauritsen, 2014), partially compensated by the positive water vapor feedback. Together with strong negative cloud feedbacks, absent positive surface albedo feedback, and strong negative Planck feedback due to warm background conditions, this region exhibits the most negative local feedback, and controls the global feedback strength. Changes in feedback parameter over time in abrupt CO_2 forcing simulations can be attributed to changes in the relative warming of the Indo-Pacific warm pool region to the global warming (Dong et al., 2020).

A.3   EARLIER STUDIES INDICATE THAT THE EFFICACY OF VAF IS LOWER THAN UNITY

There are several modelling studies which report the efficacy or the feedback parameter of VAF. All of them point towards a low efficacy of VAF. An overview is given in Table A.1.

Source	Efficacy	Method
Hansen et al. (2005)	1.00	Differential
	0.88	Equilibrium
Marvel et al. (2016)	0.73	Differential
	5-95 % CI: [-0.61, 2.06]	
Ceppi and Gregory (2019)	0.45*	Differential
	0.71*	Differential
Boer et al. (2006)	0.81*	Differential
Gregory et al. (2016)	0.69 ± 0.09	Differential
Merlis et al. (2014)	0.84* - 0.94*	Equilibrium
	0.69* - 0.92*	Differential
Zhao et al. (2021)	0.66* (low lat.) - 1.16* (high lat.)	Equilibrium
Modak et al. (2016)	<1	Differential
Gregory et al. (2020)	<1	Differential

Table A.1: Efficacy of VAF in previous studies.
* Values marked with an asterisk are not directly reported in the source, but calculated by us based on the reported feedback parameters. In the case of Boer et al. (2006), the 4 x CO_2 feedback parameter was taken from Andrews et al. (2015). For definitions of differential and equilibrium feedback parameter, see Section A.2 or Rugenstein and Armour (2021).

No two author teams use exactly the same method. The results stem from different models, and while some use time-constant forcings (Ceppi and Gregory, 2019; Gregory et al., 2016; Hansen et al., 2005; Zhao et al., 2021), others use transient forcings (Boer et al., 2006; Marvel et al., 2016). The differential feedback parameters are regressed over a different numbers of years, between two years (Ceppi and Gregory, 2019) and the entire historical period (Gregory et al., 2020; Marvel et al., 2016). In two cases, the feedback parameter of CO_2 forcing is taken from a 4 x CO_2 simulation instead of the $2 \times CO_2$ simulation (Boer et al., 2006; Gregory et al., 2016).

Additional studies suggest or show a lower-than-unity efficacy without reporting a specific value (Gregory et al., 2020; Modak et al., 2016). We conclude from this literature review that VAF likely exhibits a low efficacy.

A.4 METHODS

Model and experiments

We employ the coupled climate model MPI-ESM 1.2 in its low resolution version. It consists of the atmosphere component ECHAM6 (192

	$0.5 \times CO_2$	$2 \times CO_2$	Idealized SSAF
Years 1 - 10, coupled	18	18	18
Years 1 - 1000, coupled	1	1	1
Years 1 - 30, fixed SST	1	1	1

Table A.2: Type and number of performed simulations.

$\times$ 96 grid points horizontally, 47 vertical levels, Stevens et al. (2013)), coupled to the ocean component MPIOM (256 $\times$ 220 grid points horizontally, 40 vertical levels, Jungclaus et al. (2013)). Furthermore, land processes and ocean biogeochemistry are simulated in JSBACH (Reick et al., 2021) and HAMOCC (Ilyina et al., 2013), respectively. Since no interactive atmospheric chemistry processes are included, aerosols and trace gases are prescribed with monthly fields.

Based on a control simulation with pre-industrial conditions (piControl), three sets of simulations are performed (Table A.2). For each forcing type ($0.5 \times CO_2$, $2 \times CO_2$, Idealized SSAF) we prescribe a step-like radiative forcing and perform 18 simulations of the first 10 years, one simulation of 1000 years, and one simulation of 30 years with fixed SST.

The aerosol optical properties for the idealized SSAF are calculated with the EVA forcing generator (Toohey et al., 2016). The monthly and zonal mean fields of aerosol extinction, single scattering albedo and the asymmetry factor are pre-computed offline for a sulfur injection of 20 Tg, once for a January and once for a July eruption. We then shift the July eruption by 6 months in order to phase match it with the January eruption, and compute the average of both. This eliminates the seasonal transport asymmetry, but retains a realistic annual-average sulfate transport towards the poles. We then average over the first 3 post-eruption years and prescribe this profile to MPI-ESM. It is time-constant by construction, but representative of the time-averaged forcing structure after a volcanic eruption. The sulfur mass is chosen so that the time-averaged global-mean effective forcing is approximately the same as in the $0.5 \times CO_2$ experiment, in order to preclude forcing strength-dependent effects on the feedback (Rohrschneider et al., 2019).

The simulations with fixed SST are only used to calculate the effective radiative forcing. The 18-member ensembles of the first decade are intended to reduce the uncertainty in regressions over the early phase of the temperature change, where there are only few data points.

Calculation of effective forcing, feedback parameter and temperature change pattern

We calculate the effective forcing as the average TOA flux imbalance of the simulations with fixed climatological SST (Forster and Taylor, 2006). Although the SST are fixed, there is non-negligible temperature change over land. Therefore, we correct the effective forcing by subtracting λT in each year from the TOA flux imbalance, where λ is the differential feedback parameter of the first decade from the coupled runs (see Section A.4).

The equilibrium feedback parameters are calculated as described in Section A.2. Since the simulations still deviate from equilibrium by about $0.3\,\mathrm{W\,m^{-2}\,K^{-1}}$ after 1000 years, we extrapolate $N(T)$ to zero based on a linear fit over the years 100 - 1000. The intercept with the T-axis yields the equilibrium surface temperature change. This choice is roughly equivalent to Rugenstein and Armour (2021)'s method M5, which they apply to estimate the equilibrium temperature change in almost equilibrated models. Their method of extrapolating the years 100 to 400 from $N(T)$ to $N = 0$ underestimates the equilibrium T in our simulations, likely due to the pronounced curvature of $N(T)$ in our cooling simulations.

Additionally, we obtain differential feedback parameters by regressing over the first ten years (early period) or years 100 - 400 (late period). The latter choice is based on the recommendations by Rugenstein and Armour (2021). The results are not sensitive to the exact choice of boundary years for the periods. Derived quantities for the first ten years, e.g. feedback parameter and temperature change pattern, are calculated from the ensemble average, as opposed to first calculating them individually for each ensemble member and averaging thereafter following the recommendations of Gregory et al. (2020). The results are the same to within a few percent when the order of regression and averaging is reversed.

Analogous to the differential feedback parameter, differential temperature change patterns are calculated by regressing the local against the global average change in T. Based on the findings of Dong et al. (2020, 2019) we define the warm pool index (WPI) as the regression slope of the temperature change in the tropical warm pool region (30°S - 30°N, 50°E - 160°W) against the global surface temperature change, similar to the ratio of tropical to global temperature change of Soden and Held (2006).

All regressions are performed over annual-mean values. The method to compute errors of the feedback parameters depends on the feedback parameter definition. The uncertainty of the equilibrium feedback parameter is equal to the uncertainty of F/T_{eq}, propagated from the standard deviations of F and T_{eq}. The differential feedback parameters
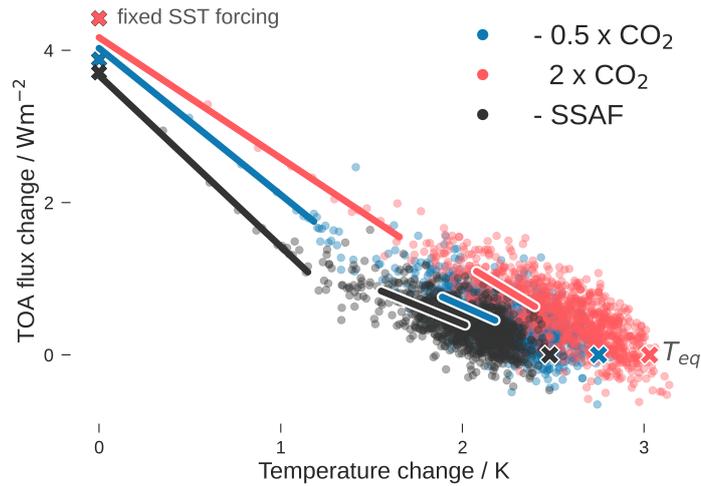
Figure A.1: $N(T)$ (Gregory) plots. All values from $0.5 \times CO_2$ and SSAF experiments are multiplied by -1 in order to show them in the same quadrant as the $2 \times CO_2$ results. Linear regressions are shown for the early period (Years 1 - 10) and the late period (years 100 - 400). Crosses mark the fixed SST effective forcings and the equilibrium temperatures, which were extrapolated from a linear regression of $N(T)$ over years 100-1000. The first ten points of each simulation type are the ensemble averages from the 18-member ensembles of the first decade. The slope of each regression line is the differential feedback parameter.

are computed from the slope of $N(T)$, therefore their uncertainties are described by the least square regression slope's standard deviation.

A.5 RESULTS

Cooling is characterized by strong early feedback

As pointed out in Section A.3, other authors have found lower-than-unity efficacies for VAF, but not under a uniform procedure. In order to emphasize the importance of the methodological approach, we show $N(T)$ plots (Fig. A.1), the global equilibrium and differential feedback parameters (Fig. A.2 (a)), and the corresponding efficacies from our simulations (Fig. A.2 (b)).

The effective radiative forcing magnitude for CO_2 halving is approximately 15 % lower than for CO_2 doubling, in line with the previously reported 10 % from Chalmers et al. (2022). All quantities we show are normalized by effective forcing or temperature change, so that this does not compromise our findings.

The equilibrium feedback parameter is very similar in all three simulations. In equilibrium, SSAF produces only slightly less temperature
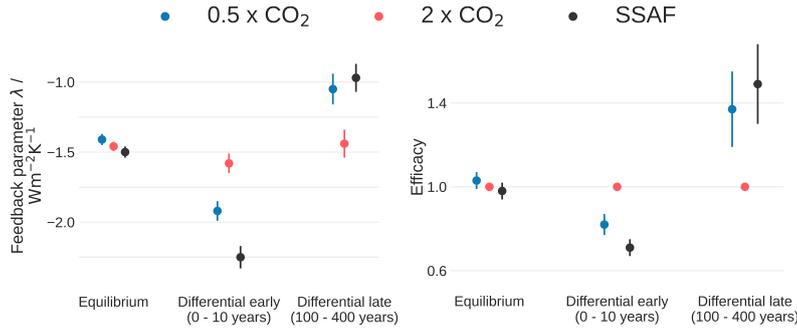
Figure A.2: Feedback parameter (a) and efficacy (b), calculated from the equilibrium method and the differential method for the early and late period. The efficacy of $2 \times CO_2$ simulations is 1 by definition.

change per unit forcing than a halving or doubling of CO_2. Differences between the three forcing agents are pronounced in the early differential feedback parameter, which is obtained from an ensemble average of the regression slopes of $N(T)$ over the first 10 years of each ensemble member. This method is closest to what most authors did to calculate the efficacy of VAF (see Section A.3). We find a more negative feedback parameter in both cooling cases for the early period. By contrast, the late regression feedback parameter (years 100 - 400) is less negative for the cooling cases than for the $2 \times CO_2$ simulation.

These findings suggest the following picture: Compared to the $2 \times CO_2$ warming case, the cooling simulations ($0.5 \times CO_2$ and SSAF) exhibit stronger feedback in the first decades. Therefore the efficacy is low in transient simulations of VAF, as previously reported (Section A.3). Later on, the feedbacks weaken and become even weaker than in the warming case. Since most authors focus on one specific time period for regressing the differential feedback parameter, this change in feedback strength has not been reported before for cooling simulations. All in all, the early stronger and the late weaker feedback almost compensate, so that the equilibrium feedback parameters of SSAF, CO_2-induced cooling, and warming are almost equal, i.e. the equilibrium efficacy is approximately one. Efficacy differences exist in transient states, but not in equilibrium.

Note that also in the $2 \times CO_2$ simulation the feedback parameter decreases over time, albeit less than in the other two simulations. The weakening of the feedback of $2 \times CO_2$ simulations is a well-known feature of the majority of CMIP5 and CMIP6 models (Andrews et al., 2012; Armour, 2017; Armour et al., 2013; Ceppi and Gregory, 2017; Senior and Mitchell, 2000; Winton et al., 2010) and explained by pattern effects (see Section A.2).

The feedback parameter change over time is more pronounced in our cooling simulations than in the warming simulations, and largest for SSAF. Two questions arise immediately:
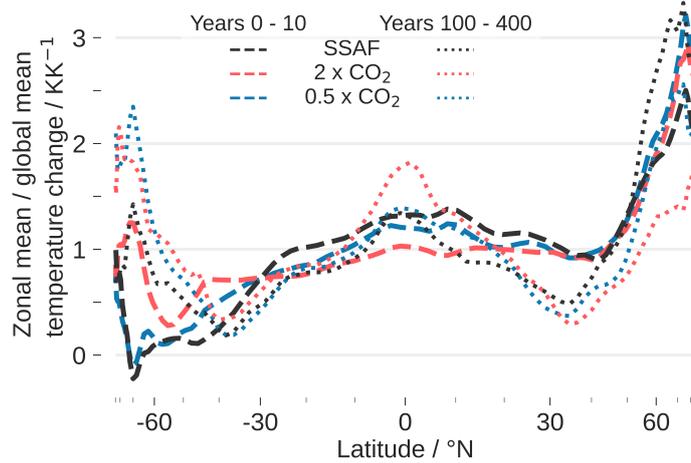
Figure A.3: Zonal mean temperature change pattern for the early (years 0 - 10) and the late (years 100 - 400) period. Values greater than one indicate stronger temperature change than the global average. A value of zero indicates no temperature change.

- Why do the cooling simulations produce initially stronger global mean feedback than the warming simulation?

- Which processes change over time, that lead to the enhanced weakening of the feedback parameter in the cooling simulations?

In the following we will argue that these questions can be answered on grounds of different temperature change patterns and their temporal evolution.

The temperature change pattern explains feedback differences

The temperature change patterns are shown as zonal averages in Fig. A.3, and spatially resolved in Fig. A.4. All patterns show some common characteristics in the early period: Amplified temperature change in the Arctic, intermediate temperature change in the tropics, and small temperature change in the southern mid and high latitudes, except close to the south pole. The qualitative temperature change pattern with reduced Southern Ocean and Antarctic temperature response after volcanic eruptions was found before by Yang et al. (2019) and confirmed in a multi-model large ensemble comparison (Pauling et al., 2021). In the late period, the pattern becomes more "El-Niño-like", consistent with the long-term temperature change pattern in other models (Cai and Whetton, 2001; Held et al., 2010).

Despite the generally similar shape of the patterns, there are important differences: While the zonal mean temperature change is smaller or equal to the global average at almost all latitudes in the tropics in the first decade of the $2 \times CO_2$ simulation, the opposite is true in both
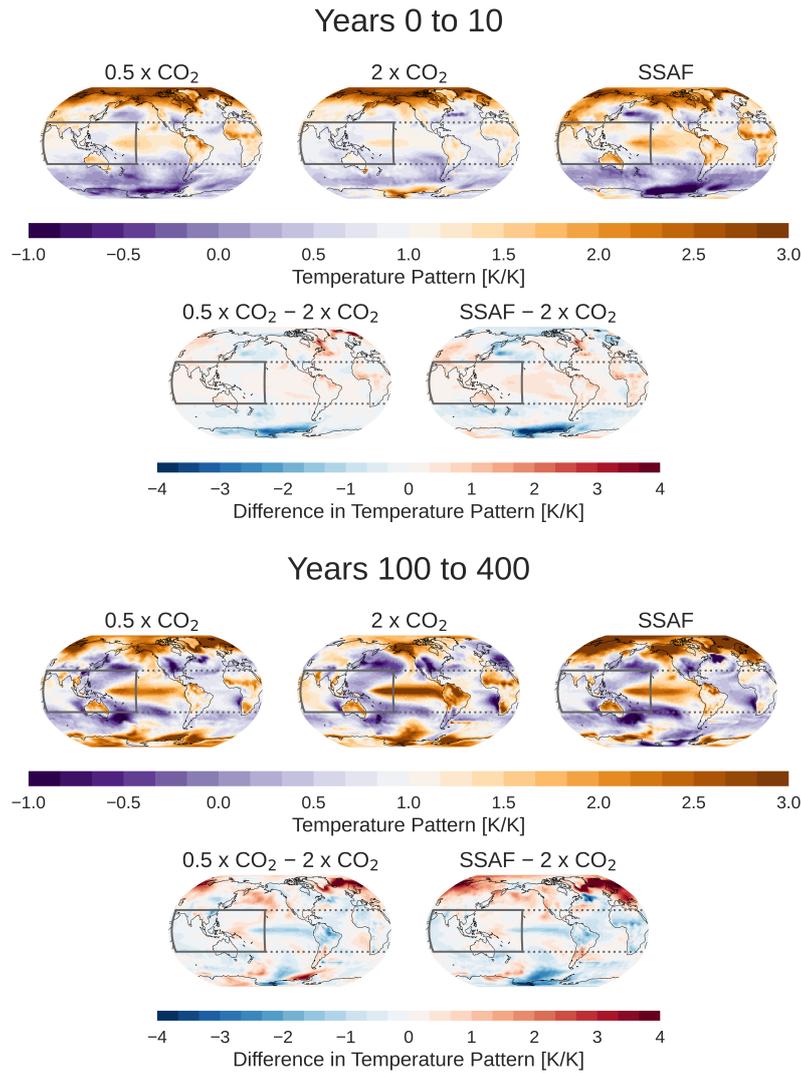
Figure A.4: Spatially resolved temperature change pattern for a) the early period (years 0 - 10), and b) the late period (years 100 - 400). The lower panels of each subfigure show the difference to the $2 \times CO_2$ simulation. Values greater than one indicate stronger temperature change than the global average. A value of zero indicates no temperature change. The solid and dotted lines indicate the boundaries of the warm pool region and the tropics, respectively.

cooling cases. This picture is reversed over time: In the late period, the deep tropical temperature change is enhanced compared to the global mean for the $2 \times CO_2$ simulation (see Fig. A.3, difference between dashed and dotted line between $20°S$ and $20°N$). The opposite is true for the cooling simulations, which have a lower temperature change in the late period than in the early period in most tropical latitudes, compared to the global average. During the late in contrast to the early period, the southern high latitudes show much larger temperature

changes in both cooling experiments, but especially in the $0.5 \times CO_2$ case.

We propose that the global feedback is strongly negative in the cooling simulations in the early period, because the temperature change is concentrated in the tropical warm pool region (30°S - 30°N, 50°E - 160°W). In this region the combined lapse rate and water vapor feedback is stronger than in the global average (Dong et al., 2020, 2019; Po-Chedley et al., 2018), and the surface albedo feedback is zero. Using the radiative kernels from Block and Mauritsen (2013), we perform a kernel decomposition of the total feedback into its components (Soden et al., 2008). This yields that lapse-rate, water vapor and surface albedo feedback together account for 85 % of the feedback differences between $2 \times CO_2$ and SSAF radiative forcing in the early period (not shown). The same feedback processes cause 100 % of the feedback differences between $2 \times CO_2$ and $0.5 \times CO_2$, where additional small differences in cloud and Planck feedback compensate. In contrast to the multi-model comparison of Dong et al. (2019), where the cloud feedback plays an important role for feedback parameter differences, it is approximately zero in all our simulations during the early period.

The most striking differences in the temperature patterns are not located in the warm pool region, but rather at higher latitudes. However, temperature changes in the warm pool have much larger global radiative effects than changes in any other region, such that even moderate differences in the warm pool region can dominate the global response (Dong et al. (2019), their Fig. 5). For this reason we focus on the warm pool region, although the temperature pattern differences there are only moderate.

We show the relationship between the temperature change pattern and the feedback parameter in Fig. A.5, where the WPI is a simple measure for the temperature change pattern. A WPI greater than one means that the tropical warm pool region warms or cools more than the global average. The feedback parameter is more negative in simulations with enhanced temperature change in the tropical warm pool region. The scatter points of SSAF and $2 \times CO_2$ simulations are well separated in the early period. The negative correlation between WPI and feedback parameter exists not only for the three simulation types together, but also for each of the three ensembles individually.

As time passes, the temperature change pattern in the SSAF and $0.5 \times CO_2$ simulations is shifted away from the warm pool region, and more generally from the tropics towards the extratropics. This enhances the less negative lapse-rate feedback of the non-warm pool regions and the positive surface albedo feedback of the high latitudes. For this reason, the overall feedback is initially stronger when cooling, and becomes weaker later on.

However, the scatter points from the late period (squares in Fig. A.5) do not fall on the regression line from the early period. Apparently, the
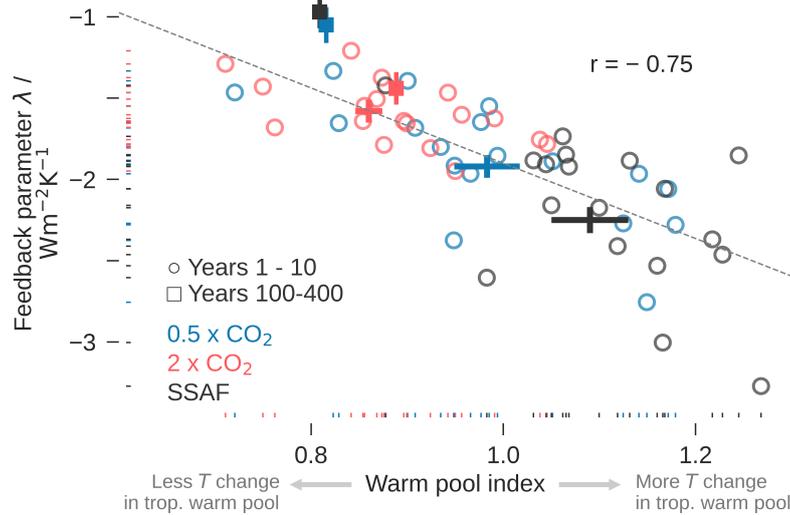
Figure A.5: Scatter plot of feedback parameter vs WPI (ratio of temperature change in the tropical warm pool region to global mean temperature change). Each circle represents one ensemble member from the early period; crosses show ensemble mean $\pm$ standard error; squares and error bars mark values from the late period (years 100 - 400) $\pm$ standard error. The gray dashed line is a linear regression through all $3 \times 18 = 54$ ensemble members from the first decade. Marginal distributions are plotted on the axes.

simple WPI cannot explain completely the shift from strong to weak feedback between the decadal and centennial time scale. This shift has been found to be related to the temperature change pattern over ascent and descent regions within the tropics (Dong et al., 2020, 2019), and to the delayed southern ocean temperature change (Senior and Mitchell, 2000). Additionally, the relationship between local SST and global radiation changes might not be invariant to climate change, which could alter the relationship between WPI and feedback parameter in warmer or cooler climates.

Although the WPI cannot explain the change in feedback from the decadal to the centennial time scale entirely, it does explain the efficacy differences between simulations within each period. While we point out the importance of the tropical warm pool region for efficacy differences between $0.5 \times CO_2$ forcing, $2 \times CO_2$ forcing, and SSAF, we acknowledge that temperature change in other regions also influences the feedback. Interestingly, the cloud feedback of all our simulations is approximately the same in the early period (0.0 to 0.1 +/- $0.1 \, \mathrm{W \, m^{-2} \, K^{-1}}$), although cloud feedbacks are the most important contribution to the differences between early and late feedback of CMIP5 and CMIP6 $4 \times CO_2$ simulations (Dong et al., 2020), including MPI-ESM (Block and Mauritsen, 2013). Either MPI-ESM is an outlier in this regard, or the processes that lead to differences between early
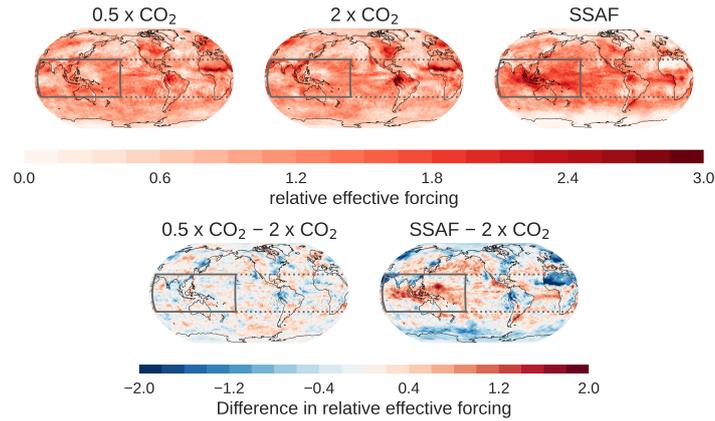
Figure A.6: Relative effective forcing (effective forcing normalized by its global average). The lower panels show the difference to the $2 \times CO_2$ simulation. The solid and dotted lines indicate the boundaries of the warm pool region and the tropics, respectively.

and late feedback are not exactly the same processes that distinguish SSAF, $0.5 \times CO_2$, and $2 \times CO_2$ forcing in the early period.

Repeating the analysis including the whole tropics instead of only the warm pool region, (i.e. using the "ratio of tropical to global warming" from Po-Chedley et al. (2018) and Soden and Held (2006)) yields similar results, albeit with weaker correlations (s. supporting information, Fig. S1). The conclusions are the same for both variants. While there is a theoretical foundation for the link between the ratio of tropical to global warming and the feedback parameter (Po-Chedley et al., 2018), this relation might actually originate from the dominance of the tropical warm pool region. Of course, WPI and the ratio of tropical to global warming are correlated.

SST pattern changes can lead to changes in large-scale tropospheric stability (Ceppi and Gregory, 2019). They can be quantified by the differential stability change dS/dT, where S is the area-averaged estimated inversion strength (EIS, Wood and Bretherton (2006)) between 50°S - 50°N. Increased stability leads to more low clouds in the tropical subsidence regions and a more negative lapse rate feedback (Ceppi and Gregory, 2019). Exchanging the WPI for dS/dT in Fig. A.5 yields a qualitatively similar result (s. supporting information, Fig. S2). The SST pattern measure WPI and the stability measure dS/dT are well correlated (r = 0.81, s. supporting information, Fig. S3), which corroborates the relationship between changes in SST patterns, stability, and feedback (Ceppi and Gregory, 2019; Salvi et al., 2022).

Sign, pattern of radiative forcing, and stratospheric heating cause temperature change pattern differences

While we have shown that the differences in temperature change patterns explain the differences between the feedback parameters within the same period, it is not clear what causes the former. We propose that three factors combine to elicit different WPI values.

One logical explanation would be that the WPI differences are caused by different radiative forcing patterns, which we show in Fig. A.6. The high aerosol concentration and the strong insolation in the tropics cause the SSAF to be stronger in the tropics than in the extratropics. Specifically the warm pool region experiences stronger relative effective forcing from aerosols than from either doubling or halving the CO_2 concentration. Since this pattern only weakly appears in the instantaneous forcing, the warm pool-enhanced effective forcing pattern must largely originate from atmospheric adjustments (not shown). It could lead to more temperature change in the tropical warm pool region. However, the two simulations with altered CO_2 concentrations have almost the same forcing pattern (see Fig. A.6), yet their temperature change patterns are very different (see Fig. A.3, A.4), and so are the feedback parameters (see Fig. A.2, A.5). Hence, the forcing pattern alone is not sufficient.

Another conceivable explanation is that the temperature change pattern differences arise from the distinction between negative and positive radiative forcing. Ocean heat transport is projected to decrease with warming (Li et al., 2013; Previdi et al., 2021), which leads to more negative radiative feedbacks (Singh et al., 2022), consistent with our interpretation of the decreased activation of extratropical, more positive feedbacks. If the poleward ocean heat transport strengthens in cooler conditions, that would explain why the temperature change shifts from the tropical warm pool region to the extratropics on the centennial time scale. The reduced ocean heat transport in warmer conditions could be the reason that the temperature change pattern, as measured by the WPI, remains relatively constant in the $2 \times CO_2$ simulation. Furthermore, warming stabilises the ocean and therefore suppresses vertical mixing, whereas cooling leads to destabilisation. This might cause differences in mixing time scales, surface fluxes, and consequentially in SST patterns. The enhanced mixing from cooling would enhance the relative importance of ocean heat uptake, compared to radiative feedbacks. Ocean heat uptake has been interpreted as a forcing with weak feedback (Rose et al., 2014; Winton et al., 2010). Therefore a cooling/warming asymmetry of the relative role of ocean heat uptake could affect the global mean feedback. We present no evidence corroborating or rejecting this hypothesis, but suggest it as a possible reason for a cooling vs. warming feedback asymmetry. Still, if cooling vs. warming were the only important distinction, the

$0.5 \times CO_2$ and SSAF ensemble means would not be separated in Fig. A.5.

Thirdly, the altitude of the aerosol layer has been shown to influence the efficacy through stratospheric heating and water vapor feedback (Zhao et al., 2021). Therefore the heating by absorption of LW radiation influences the feedback parameter, but our results show that this process is mediated by the temperature change pattern. This would reconcile the findings of Zhao et al. (2021) with those of Gregory and Andrews (2016), Shindell et al. (2015) and Haugstad et al. (2017), who show that the surface temperature change pattern and not the existence/non-existence of aerosol determines the feedback strength.

Our results do not allow to conclude what exactly causes the distinct temperature change patterns, but it is likely a combination of the sign (possibly also the magnitude) of the radiative forcing, its pattern, and the stratospheric heating from the aerosols. None of these factors is sufficient to explain the differences on its own.

a.6   eva ensemble and cmip6 simulations confirm increased warm pool temperature change
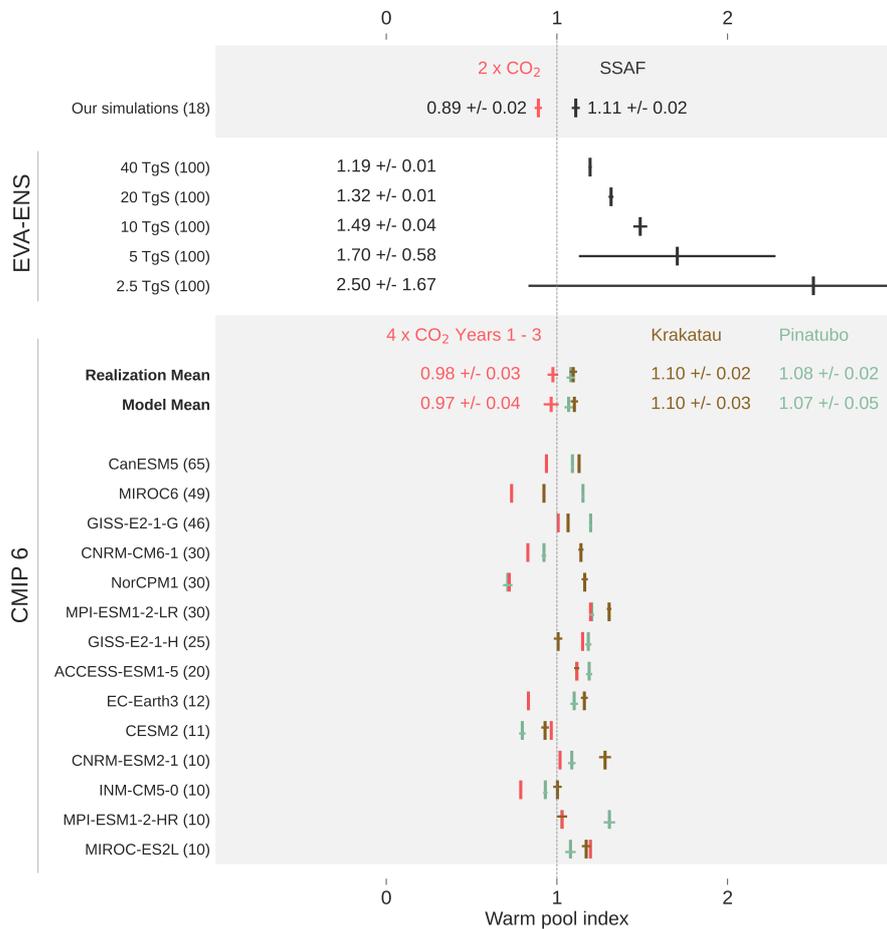
EVA ensemble

Figure A.7: Comparison of the WPI in our 18-member, 10-years ensemble of idealized SSAF simulations, simulations from EVA-ENS, and volcanic periods of the CMIP6 historical simulations. Numbers in parentheses indicate the number of ensemble members, models with less than 10 members are not shown. Error bars represent standard errors and are slightly shifted up and down in order to avoid indistinguishable overlaps. For $4 \times CO_2$ there are generally only one or a few realizations per model. The errorbar of the 2.5 TgS EVA-ENS simulations exceeds the figure limits.

Our SSAF simulations are highly idealized: They are based on time-independent radiative forcing and do not reflect the seasonally varying transport of stratospheric aerosol. In order to verify if the enhanced temperature change in the tropical warm pool region also holds for time-dependent VAF, we compare them to the MPI "Idealized Volcanic Forcing Ensemble" (EVA-ENS). Azoulay et al. (2021) created 100-member ensembles of idealized equatorial volcanic eruption with sulfur injections of 2.5 to 40 Tg. For comparison, it has been estimated that the Mt. Pinatubo eruption injected approximately 5 - 10 TgS (Timmreck et al., 2018) into the stratosphere. Their aerosol optical depth distributions are, as in our case, obtained from the EVA forcing generator. The simulations of EVA-ENS assume an eruption in June 1991, and are branched off of 100 different realizations of the MPI Grand Ensemble of historical simulations (Maher et al., 2019). We determine the temperature change pattern in years 1991 - 1993 and compute the WPI. The results are shown in Fig. A.7. For low sulfur injections of 5 Tg or less, the standard error is very large and no reliable statement can be made on the spatial structure of the temperature change. For injections of 10 TgS or more, the temperature change is robustly concentrated in the tropical warm pool region. After the idealized volcanic eruptions with injections greater than 5 TgS the WPI is approximately between 1.2 and 1.5. This is higher than the values from the idealized SSAF simulation (1.1), which could be related to the even shorter time scale of 3 years for the simulations of EVA-ENS, as compared to the time scale of 10 years of the idealized SSAF simulations. For injections of less or equal to 5 TgS, the variability becomes too large to make a reliable statement.

Volcanic eruptions in CMIP6 historical simulations

For a multi-model comparison we examine the CMIP6 "historical" simulations and warming signals from the "abrupt 4 x CO_2" simulations. We use the years 1883-1885 and 1991-1993 from the CMIP6 historical simulations for VAF, because in these periods the radiative forcings from the eruptions of Krakatau in May - August 1883 and Mt. Pinatubo in June 1991 dominate the total effective forcing from all sources. The signal-to-noise ratio is still poor and the average simulated annual mean global mean cooling amounts to only 0.3 K for Krakatau, and 0.2 K for Mt. Pinatubo. This is on the order of internal variability, so that individual realizations may exhibit only small negative or even positive temperature anomalies. We compute the WPI for all ensemble members from the participating models with at least 10 historical realizations. After removing 3 outliers with a WPI $\gg 6$, we compare these WPI values to those from the first three years of the abrupt 4 x CO_2 simulations. The results for each model and for the whole ensemble are shown in Fig. A.7. Figures showing results for

each individual simulation from all models (including those with less than 10 realizations) are provided in the supporting information (Fig. S4, S5).

There are multiple sensible possibilities of averaging the results: e.g. giving each realization the same weight (realization average), or giving each model the same weight (model average). Both methods yield qualitatively the same result: The temperature change pattern is more concentrated in the tropical warm pool region in the periods of historical VAF than in the 4 x CO_2 simulations.

However, the WPI spread is large for the Pinatubo and Krakatau periods of the historical simulations, both within models and between models. Most models agree that the WPI is larger in periods of VAF, especially those with more realizations, but not all of them do. Many realizations can be found with higher WPI in the abrupt 4 $\times CO_2$ simulations than in the historical simulations during periods of VAF. If the variability of the CMIP6 simulations is representative of the real world, observed temperature change patterns and therefore also observed feedback parameters and efficacies for any individual eruption might substantially deviate from the expected value. The high WPI of VAF is robust in the multi-model ensemble mean, and in some but not all single-model ensemble means. Even in our own ensemble of highly idealized forcing scenarios some realizations exhibit a WPI, which is substantially different from the ensemble mean, causing an overlap of the WPI values of SSAF and 2 $\times CO_2$ simulations (but no overlap of the estimated uncertainties, see Fig.A.5).

The comparison to the CMIP historical simulations and the EVA ensemble shows that the increased WPI of SSAF in the early period is also found in simulations of VAF.

0.5×CO_2 simulations from CFMIP

In order to verify our results from the 0.5 $\times$ CO_2 simulations, we analyze a set of 5 models which performed 0.5 $\times$ CO_2 and 2 $\times$ CO_2 simulations in the framework of the Cloud Feedback Model Intercomparison Project Phase 3 (CFMIP3, Webb et al. (2017). Since the CFMIP simulations span only 150 years, the feedback parameter can only be computed for the early period. Three out of four models that provide the necessary output to compute the feedback parameter show an at least moderately stronger feedback in the 0.5 $\times$ CO_2 simulations than in the 2 $\times$ CO_2 simulations. However, there is no clear relationship between WPI and feedback parameter. While this seems to indicate that our results are not supported by the CFMIP3 ensemble, we stress that these are only five models with only one realization per model (two 2 $\times$ CO_2 simulations for GISS-E2-1-G) of relatively weak forcings. In our 18-member ensembles, the ranges of feedback parameters are large and there is an overlap between the ensembles of cooling

	$0.5 \times CO_2$		$2 \times CO_2$	
	WPI	λ / $W\,m^{-2}\,K^{-1}$	WPI	λ / $W\,m^{-2}\,K^{-1}$
MRI-ESM2-0	0.87	—	0.65	—
CNRM-CM6-1	0.75	-1.10	0.92	-1.03
IPSL-CM6A-LR	0.88	-1.16	0.82	-1.40
GISS-E2-1-G	0.79	-1.66	0.92 | 0.93	-1.04 | -1.60
MIROC6	0.84	-1.89	0.94	-1.61

Table A.3: Early WPI and feedback parameter for the simulations from CFMIP. MRI-ESM2-0 did not provide the necessary output to compute the feedback parameter. GISS-E2-1-G provides two realizations of the $2 \times CO_2$ experiment

and warming. A larger ensemble from CFMIP would be necessary to conclude this more definitely. If indeed MPI-ESM is an outlier in that regard, then the conclusions on the importance of the sign (or magnitude) of the radiative forcing pattern might be wrong. Mitevski et al. (2022) found no feedback differences between $0.5 \times CO_2$ forcing and $2 \times CO_2$, although this might be explained by the fact that they examine the differential feedback parameter of years 1 - 150, which is exactly between our early period (years 1-10, characterized by strong feedback) and our late period (years 100-400, characterized by weak feedback). On the other hand, our results match those from Chalmers et al. (2022), who also report stronger feedbacks in $0.5 \times CO_2$ than in $2 \times CO_2$ simulations, although they find an elevated importance of cloud feedbacks in addition to lapse rate and surface albedo feedbacks.

A.7 DISCUSSION AND SUMMARY

There is a broad consensus in the literature that VAF is characterized as producing stronger feedback than $2 \times CO_2$ forcing, i.e. low efficacy. Using idealized simulations of $0.5 \times CO_2$ forcing, $2 \times CO_2$ forcing, and step-like SSAF, we tackle the question why the feedback parameter appears to be stronger for VAF.

The finding that the feedback to SSAF is stronger than that of $2 \times CO_2$ forcing holds only for differential feedback parameters in the early decade, while on the centennial time scale the feedback weakens considerably. This weakening is more pronounced in the $0.5 \times CO_2$ and SSAF simulations than in the $2 \times CO_2$ simulation. In equilibrium, there are almost no efficacy differences between SSAF, CO_2-warming and CO_2-cooling, due to a compensation of early strong and late weak cooling feedbacks.

This pronounced change in feedback over time might have implications for climate engineering by solar radiation management. Due to the low early efficacy a comparably large negative radiative forcing

would be necessary to balance the surface temperature change from CO_2 on the decadal time scale. On the centennial time scale the efficacy of SSAF is larger than unity. Therefore a particular CO_2 radiative forcing could be balanced by a weaker (in magnitude) negative radiative forcing from aerosols in the long term. However, it is not clear how our results might change in the presence of strong CO_2 radiative forcing. Since part of the pronounced changes in the efficacy of aerosol forcing seems to be related to the fact that the earth is cooling, this would not be observed in a world of approximately constant or slightly rising mean temperature.

The responses to cooling and warming differ and there are substantive feedback changes over time. Hence it is not possible, or at least not straightforward, to estimate the equilibrium climate sensitivity (ECS) to CO_2 forcing from the observed response to volcanic eruptions. This has been argued before, but partly on different grounds (Boer et al., 2006; Gregory et al., 2020; Kummer and Dessler, 2014; Merlis et al., 2014; Stowasser et al., 2006; Wigley et al., 2005). Feedback differences were often interpreted to originate from the SW/LW nature of the radiative forcing (Bony et al., 2006; Joshi et al., 2003) or from differences in ocean heat uptake (Boer et al., 2006; Stowasser et al., 2006), but less from its sign.

Changes of the feedback parameter in time are related to varying temperature change patterns. The temperature change pattern causes the differences in feedback strength between warming and cooling simulations, which explains the early low efficacy of SSAF. Compared to a doubling of the CO_2 concentration, a halving of the CO_2 concentration, and even more so a cooling with SSAF, leads to stronger temperature changes in the tropical warm pool region in the first decade, relative to the global mean. The slightly enhanced temperature change in the warm pool region substantially increases near-global stability S and strengthens the global feedback parameter. The differences between cooling and warming simulations in the first decade mainly originate from lapse rate, water vapor, and surface albedo feedback. In the cooling simulations, the temperature change progresses to the high southern latitudes on the decadal to centennial time scale, which leads to an activation of the less negative / more positive high-latitude feedbacks. Therefore, the global feedback weakens more in the cooling simulations than in the $2 \times CO_2$ simulation. In our simulations, this effect is common to CO_2-induced and SSAF-induced cooling, but more pronounced in the SSAF case. This highlights that the radiative forcing's pattern, its sign (possibly also the magnitude), and the stratospheric heating from the aerosols (Zhao et al., 2021) likely contribute to setting the temperature change pattern. The atmospheric and oceanic circulations blur differences in the forcing patterns when translating them to a temperature change pattern.

A comparison with effects of the Mt. Pinatubo and Krakatau eruptions in the CMIP6 historical simulations shows that the enhanced temperature change in the tropical warm pool region is an average feature of climate models after volcanic eruptions, although the spread is large. Moreover, we find this characteristic in the EVA ensemble, a large ensemble of simulations of idealized volcanic eruptions in MPI-ESM 1.2-LR.

Our simulations and the eruptions of Krakatau and Pinatubo in the CMIP6 historical simulations are characterized by tropically concentrated SSAF. Can the results be transferred to solar forcing and extratropical SSAF?

Other studies have found the efficacy of solar forcings to be lower than unity (Hansen et al., 2005; Modak et al., 2016; Schmidt et al., 2012), specifically when cooling. We speculate that this might be explained by the fact that solar forcing predominantly affects the tropics, and therefore leads to a tropically enhanced temperature change pattern, similar to SSAF. It is furthermore possible that the forcing pattern helps explain non-unity efficacies of other forcing agents (Ceppi and Gregory, 2019; Hansen et al., 2005). For *anthropogenic* aerosol forcing, the link between an extratropically concentrated radiative forcing pattern and *high* efficacy has already been demonstrated Salvi et al., 2022.

Zhao et al. (2021) found that extratropically concentrated volcanic aerosols lead to a less negative equilibrium feedback parameter than in the case of mostly tropical aerosol load. Therefore, the finding of low efficacy from SSAF might not hold for eruptions at higher latitudes. Extratropical eruptions likely elicit a temperature change pattern, which is shifted towards the extratropics, and therefore cause weaker feedback. In that case, the efficacy could increase and be on the order of or even larger than one. While differences in LW feedbacks are also important in the simulations of Zhao et al. (2021), the efficacy is mostly explained by differences in clear-sky and cloudy-sky SW feedbacks, suggesting an elevated importance of SW-cloud and surface albedo feedbacks.

The exact reason why the temperature patterns of cooling and warming differ remains open. In our simulations, they involve different degrees of polar amplification. The mechanisms that drive polar amplification are, e.g., lapse rate feedback, ice albedo feedback (Pithan and Mauritsen, 2014), and a changing balance of moist vs. dry static poleward energy transport (Alexeev et al., 2005; Armour et al., 2019; Hahn et al., 2021). An explanation for the differences in WPI between the radiative forcings and its changes in time could be based on these processes, but is beyond the scope of this work. The physical origin of the differences between the transient temperature change patterns of warming and cooling should be the focus of further research.

A.8 ACKNOWLEDGMENTS

DATA AND CODE AVAILABILITY

All primary model output and scripts used in the analysis are available online (`https://www.wdc-climate.de/ui/entry?acronym=DKRZ_LTA_066_ds00004`)

B

WHY DOES STRATOSPHERIC AEROSOL FORCING STRONGLY COOL THE WARM POOL?

This work is currently under review and available as a preprint:

WHY DOES STRATOSPHERIC AEROSOL FORCING STRONGLY COOL THE WARM POOL?

Moritz Günther[1], Hauke Schmidt[1], Claudia Timmreck[1], Matthew Toohey[2]

[1]*Max Planck Institute for Meteorology, Hamburg, Germany*
[2]*Institute of Space and Atmospheric Studies, University of Saskatchewan, Saskatoon, Canada*

ABSTRACT

Previous research has shown that stratospheric aerosols cause only a small temperature change per unit forcing because they produce stronger cooling in the tropical Indian and Western Pacific Ocean than in the global mean. The enhanced temperature change in this so-called "warm pool" region activates strongly negative local and remote feedbacks, which dampen the global mean temperature response. This paper addresses the question why stratospheric aerosol forcing affects warm pool temperatures more strongly than CO_2 forcing, using idealized MPI-ESM simulations. We show that the aerosol's enhanced effective forcing at the top of the atmosphere (TOA) over the warm pool contributes to the warm pool-intensified temperature change, but is not sufficient to explain the effect. Instead, the pattern of surface effective forcing, which is substantially different from the effective forcing at the TOA, is more closely linked to the temperature pattern. Independent of surface temperature changes the aerosol heats the tropical stratosphere, accelerating the Brewer-Dobson circulation. The intensified Brewer-Dobson circulation exports additional energy from the tropics to the extratropics, which leads to a particularly strong negative forcing at the tropical surface. These results show how forced circulation changes can affect the climate response by altering the surface forcing pattern. Furthermore, they indicate that the established approach of diagnosing effective forcing at the TOA is useful for global means, but a surface perspective on the forcing must be adopted to understand the evolution of temperature patterns.

B.1   INTRODUCTION

Stratospheric sulfate aerosol forcing can arise naturally from volcanic eruptions, or artificially from deliberate injection of sulfur into the stratosphere. The aerosol increases reflection of shortwave (SW) radiation, which constitutes a negative forcing and cools the Earth. Sulfate
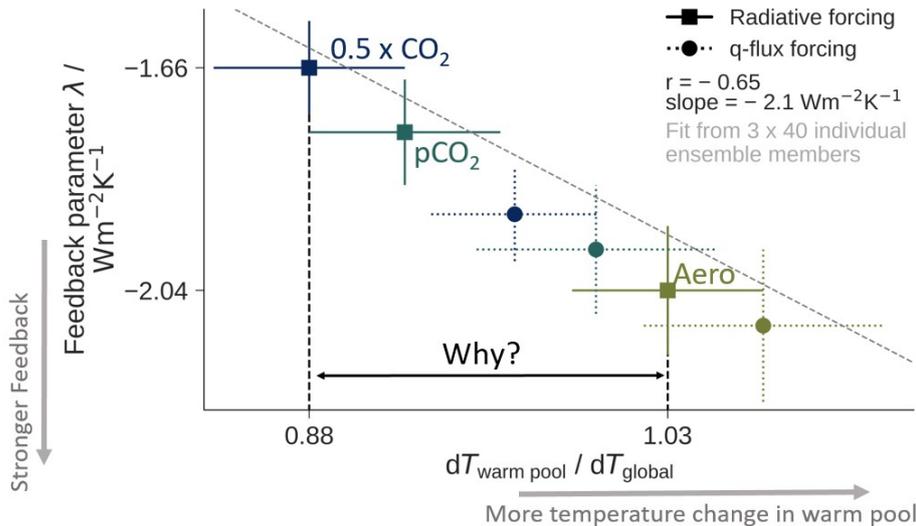
Figure B.1: Relationship between the feedback parameter and the relative WP temperature change. The quantity on the x-axis is a measure for how strongly the WP cools relatively to the global mean. The symbols indicate the mean, and the lines indicate the standard errors. Squares represent results from the radiatively forced simulations. Circles represent results from the q-flux forced simulations which are discussed in section B.3.5. Stratospheric aerosol forcing (Aero) causes stronger WP temperature change and hence stronger feedback than $0.5 \times CO_2$ forcing. The gray dashed fit line is calculated from a regression through a total of 120 radiatively forced simulations, 40 for each forcing agent (see section B.2).

aerosol also absorbs near-infrared and terrestrial longwave (LW) radiation, causing a smaller positive forcing and radiative heating in the stratosphere. Radiative forcing from stratospheric aerosol produces stronger feedback and hence a smaller temperature change per unit forcing than radiative forcing from CO_2 (e.g. Gregory et al., 2016; Günther et al., 2022; Hansen et al., 2005; Zhao et al., 2021). The pronounced negative feedback to volcanic eruptions contributes to variations in Earth's radiative feedback parameter over the historical period, where high volcanic activity coincides with strong global-mean feedback (Gregory and Andrews, 2016; Gregory et al., 2020; Salvi et al., 2023).

Modelling studies have shown that the strong feedback to stratospheric aerosol forcing arises from enhanced changes in warm pool (WP) temperatures relative to the global mean (Günther et al., 2022). The WP comprises the equatorial Indian and Western Pacific Ocean (30°S - 30°N, 50°E - 160°W) and is the main region of deep convection due to its high sea surface temperature (SST). The amplified temperature change in the WP increases the tropical to mid-latitude inversion strength and activates strong negative lapse rate and cloud feedbacks (Ceppi and Gregory, 2019). The cloud and lapse rate feedback processes that originate from the WP temperature change are powerful

enough to impact the global-mean radiative feedback, which explains how the pronounced WP cooling from stratospheric aerosol can cause substantially more negative feedback than CO_2 forcing.

However, it has remained unclear why aerosol forcing impacts WP temperatures more strongly than CO_2 forcing, which stands as the principal incentive for pursuing this study (see Fig. B.1). We explore hypotheses that could explain the causes of the different temperature patterns.

The most obvious hypothesis is that the different top of the atmosphere (TOA) effective forcing patterns cause the temperature pattern differences. Model studies have focused on the impact of aerosol from large tropical eruptions, which lead to aerosol optical depths that are largest in the low latitudes. Since both aerosol optical depth and incoming solar radiation peak in the tropics, they could combine to produce intensified low-latitude radiative forcing. In comparison, CO_2 forcing is relatively spatially uniform. The tropically enhanced forcing pattern from aerosols has been proposed to be the reason for the pronounced temperature changes in the tropics, in particular in the WP (Günther et al., 2022; Salvi et al., 2023).

Alternatively, the different temperature patterns of aerosol and CO_2 forcing could originate from other distinctive features of the forcing agents. It has been argued that spectral differences could play a role, since aerosol forcing predominantly affects SW radiation, while CO_2 exclusively affects LW radiation (Bony et al., 2006; Joshi and Shine, 2003). Günther et al. (2022) also speculated about a fundamental difference in feedback strength to positive vs. negative forcing, however, an extension of the ensemble analysed in their study made this hypothesis less plausible.

Another essential discrepancy between aerosol and CO_2 forcing is the heating of the stratosphere and upper troposphere due to the aerosols' absorption of radiation. The diabatic heating leads to a cold point warming, which allows more water vapor to enter the stratosphere (Joshi and Shine, 2003; Kroll et al., 2021), with potential impacts on the temperature response (Lee et al., 2023). The heating can furthermore alter the energy balance and the meridional temperature gradient in the upper troposphere and stratosphere. This has consequences for the strength and position of the polar vortex (e.g. Azoulay et al., 2021; Bittner et al., 2016; Graf et al., 2007; Toohey et al., 2014), and can lead to an acceleration of the Brewer-Dobson circulation (BDC), although different studies yield conflicting results (Garfinkel et al., 2017). Within the wave-driven BDC, air moves upward in the tropical stratosphere. In the tropics, forced upwelling leads to an adiabatic cooling of the environment that depends on the vertical velocity and the temperature gradient (Birner and Charlesworth, 2017). The air then moves polewards and descends in the extratropical stratosphere where it causes adiabatic heating (Holton et al., 1995). Changes to

the BDC due to stratospheric aerosol forcing have been the subject of previous research (e.g. Diallo et al., 2017; Garfinkel et al., 2017; Richter et al., 2017; SPARC, 2022), but the consequences for radiative feedback and temperature patterns have not been explored yet.

Motivated by the temperature pattern's importance for radiative feedbacks, we investigate which of the distinctions between CO_2 and aerosol forcing cause the differences in the temperature change patterns, particularly with respect to the WP. Using coupled climate model simulations, we present arguments that the pattern of TOA effective forcing is only weakly related to the pattern of surface temperatures and the radiative feedback. Instead, the surface forcing is more relevant for explaining the temperature pattern. We show that the contrast between adiabatic cooling in the tropical stratosphere and heating in the extratropical stratosphere from an accelerated BDC causes additional negative forcing at the tropical surface, which contributes to enhanced cooling of the tropics.

B.2 SIMULATIONS AND METHODS

B.2.1 *Model*

We perform simulations with the climate model MPI-ESM 1.2 in the low-resolution setup (Mauritsen et al., 2019). The atmosphere component ECHAM6 (Stevens et al., 2013) is resolved with 1.875° x 1.875° at 47 levels. It is coupled to the ocean component MPIOM (Jungclaus et al., 2013), which runs on a bipolar grid with a resolution of 1.5° near the equator. MPI-ESM also includes modules for land processes and ocean biogeochemistry (Ilyina et al., 2013; Reick et al., 2021). Since no interactive atmospheric chemistry processes are included, aerosols and trace gases are prescribed with monthly climatological fields that represent unforced pre-industrial conditions.

B.2.2 *Simulations*

We perform simulations with three forcings: An abrupt halving of the CO_2 concentration ("$0.5 \times CO_2$"), an abrupt increase of the stratospheric aerosol concentration ("Aero"), and a patterned CO_2 simulation with spatially and seasonally varying CO_2 concentrations ("pCO_2").

The Aero simulations are designed to represent the time-mean forcing induced by a strongly idealized tropical volcanic eruption, or by deliberate stratospheric aerosol injection. We derive monthly and zonal mean fields of aerosol optical properties from the EVA forcing generator (Toohey et al., 2016) for one January and one July eruption, both with an injection mass of 20 Tg sulfur. The July eruption is then shifted by 6 months, and the average of both phase-matched erup-
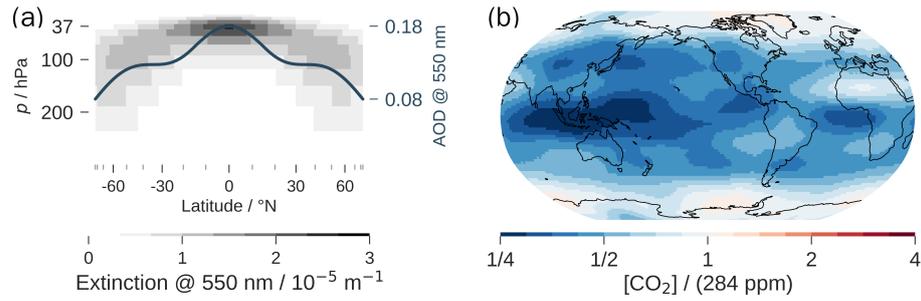
Figure B.2: Forcing input for the simulations. (a) Aerosol radiative properties that serve as input for the Aero simulation: Extinction as function of latitude and pressure, and aerosol optical depth (AOD, vertically integrated extinction) as function of latitude, computed with EVA (Toohey et al., 2016). (b) Annual-mean field of CO_2 concentrations (in units of the pre-industrial CO_2 concentration) that serves as input for the pCO$_2$ simulation. In the actual simulation, monthly varying fields are used (appendix, Figs. B.7 and B.8). The fields were computed with the algorithm described in appendix B.5.1.

tions is computed, in order to remove seasonal transport asymmetries while preserving a realistic poleward mass transport. We prescribe the average of the first three post-eruption years as time-invariant forcing to MPI-ESM. Constructing the aerosol forcing to be step-like in time allows for a consistent comparison to the 0.5 x CO_2 forcing. The aerosol is only coupled to the radiation, not transported by the model, does not evolve in time, nor does it interact directly with clouds or ozone. While these restrictions certainly limit realism, they allow us to isolate the effects of stratospheric aerosol in an idealized, interpretable framework. The most important radiative properties of the aerosol input are shown in Fig. B.2 (a).

In pCO$_2$, CO_2 concentrations at each grid box and month are chosen such that they give rise to an effective TOA forcing field which is approximately equal to the TOA radiative forcing of Aero in space and time. The rationale for this experiment's design is as follows: If the WP-enhanced TOA forcing pattern of stratospheric aerosol is responsible for the WP-enhanced temperature pattern, then the same effect should appear in a CO_2-forced simulation with WP-enhanced forcing pattern. The iterative process that was used to determine the CO_2 concentrations is described in appendix B.5.1. The annual-mean input field of spatially varying CO_2 concentrations is shown in Fig. B.2 (b). The CO_2 concentrations are lowest over the WP, and slightly higher than the pre-industrial value over the poles. The resulting field of effective forcing shares these broad features and is shown in Fig. B.3 (a).

To test the hypothesis that the effective forcing pattern from stratospheric aerosol causes the enhanced WP temperature change, we perform the three sets of simulations summarized in Table B.1.

B.2.2.1 *Coupled simulations with radiative forcing*

From a 1000-year control simulation with pre-industrial conditions (piControl), we branch one simulation for each forcing ($0.5 \times CO_2$, Aero, pCO_2) every 25 years, leading to a total of 3 x 40 ensemble members, each run for 10 years.

B.2.2.2 *Simulations with fixed SST and sea ice*

As an analog to the coupled simulations, for each forcing we perform one 100-years simulation with SST and sea ice concentrations fixed to climatological control values. By subtracting the mean climate state in these perturbed simulations from the model's mean control climate state (piClim-control), we can diagnose effective forcing at the TOA, at the surface, and adjustments (Forster et al., 2016; Sherwood et al., 2015). Results from the fixed SST simulations are averaged over all 100 simulated years except the first to allow for rapid adjustments. Forster et al., 2016 recommend 30 years to reliably diagnose the global mean effective forcing. We find that 100 years are necessary to determine also the spatial pattern of the effective forcing, especially at the surface, where interannual variability is strong.

   In addition to the fixed SST simulation with the Aero forcing, we perform a simulation with non-absorbing aerosol forcing and only with fixed SST and sea ice, in order to isolate the effects that arise from the stratospheric heating, in particular the acceleration of the BDC (section B.3.3). For this simulation we take the forcing from Aero, but set the single scattering albedo (ratio of scattering to total extinction) to one everywhere. The total extinction is then multiplied by (1 - initial single scattering albedo) in order to avoid increases in the reflectivity. For slightly different approaches to isolate the stratospheric heating effects, see Simpson et al. (2019) and Wunderlin et al. (2024). The focus of this study will be on the absorbing aerosol forcing (Aero).

B.2.2.3 *Coupled simulations with q-flux forcing*

Forcing the climate system not radiatively but with a "ghost forcing" (Hansen et al., 1997) at the surface allows for an examination of the way the surface forcing pattern affects the temperature pattern, without any perturbations to the atmosphere's radiative properties. We derive the surface effective forcing from the fixed SST simulations as the difference between all surface fluxes (radiative and turbulent) of the perturbed simulations and piClim-control. We then prescribe these flux anomalies as an additional heat source / sink ("q-flux") to the ocean and compute an ensemble of 40 simulations for each forcing agent, where each simulation lasts for 10 years. Note that the atmosphere is still fully coupled to the dynamical ocean.

Table B.1: Overview over the MPI-ESM simulations

	$0.5 \times CO_2$	Aero	pCO$_2$	Aero (non-absorbing)
Coupled: radiative forcing	40 x 10 years	40 x 10 years	40 x 10 years	—
Fixed SST: radiative forcing	100 years	100 years	100 years	100 years
Coupled: q-flux forcing	40 x 10 years	40 x 10 years	40 x 10 years	—

B.2.3 *CMIP6 output*

We complement the dedicated MPI-ESM simulations with output of the piControl and historical simulations from phase 6 of the Coupled Model Intercomparison Project (CMIP6; Eyring et al., 2016) to test the results on other models. We include all 23 models that provide the necessary output to compute adiabatic cooling in the stratosphere according to Eq. B.2 (see section B.2.4.4). For models with multiple realizations the ensemble-mean is calculated after applying Eq. B.2, so that each model is weighted equally. A list of all models and the number of ensemble members is found in Table B.3 in the appendix.

B.2.4 *Defining forcing, feedback, WP-enhancement, and adiabatic cooling*

B.2.4.1 *Effective forcing*

Effective forcing is defined as the time-mean flux change in a perturbed simulation compared to an unperturbed control simulation, both with the same prescribed SST and sea ice (Forster et al., 2016). It is traditionally measured at the TOA, where it consists of SW and LW flux changes. We also diagnose effective forcing at the surface, where additionally the sensible and latent heat fluxes must be taken into account.

B.2.4.2 *Feedback parameter*

We employ the definition of the "differential feedback parameter" following Rugenstein and Armour (2021) as $\lambda = \frac{\partial N}{\partial T}$ with global-mean TOA flux N and global-mean near-surface air temperature T, obtained by regression over ten years. The differential feedback parameter characterizes the transient response to the forcing on a time scale of ten years and bears only very limited implications for the long-term or equilibrium response.

B.2.4.3 *Warm pool Index (WPI)*

Given the elevated role of the WP, spatial patterns can be meaningfully measured with a simple WP index (WPI), which indicates how

strongly a quantity is concentrated in the WP. For patterns of effective forcing F, we define it as $\text{WPI}_F = F_{\text{WP}}/F_{\text{global}}$. Temperatures vary with time, so for patterns of temperature change T we define $\text{WPI}_T = \mathrm{d}T_{\text{WP}}/\mathrm{d}T_{\text{global}}$, obtained by regression over 10 years. Values greater than one indicate greater forcing or temperature change in the WP than in the global mean.

B.2.4.4 *Adiabatic cooling in the stratosphere*

Upwelling in the tropical stratosphere causes adiabatic cooling of rate K (in Ks^{-1}), which is proportional to the residual mean vertical velocity $\bar{w}^*$ and the deviation of the temperature profile $\frac{\partial T}{\partial z}$ from a dry adiabat $-\frac{g}{c_p}$ (Birner and Charlesworth, 2017):

$$K = -\bar{w}^* \left(\frac{\partial T}{\partial z} + \frac{g}{c_p} \right) \tag{B.1}$$

with the specific heat capacity of dry air c_p and gravitational acceleration g. The residual mean vertical velocity $\bar{w}^*$ is obtained from a transformed Eulerian mean analysis (e.g. Butchart, 2014). $\bar{w}^*$ combines the mass flux contributions from the mean velocity w and the eddies (Butchart, 2014), and therefore better represents the mass flux than w alone. Using the hydrostatic approximation $\rho dz = -dp/g$, we calculate the integrated adiabatic cooling of the stratosphere as power flux density Q_{adi} (in Wm^{-2}) according to

$$\begin{aligned} Q_{\text{adi}} &= c_p \int_{\text{strat.}} K(z)\rho(z) \, dz \\ &= -\frac{c_p}{g} \int_{100 \text{ hPa}}^{1 \text{ hPa}} K(p) \, dp \end{aligned} \tag{B.2}$$

By vertically integrating over the stratosphere we effectively treat it as one layer that causes adiabatic cooling. Changing the lower limit to 70 hPa changes the numbers by up to 25 %, but not in a way that would affect the conclusions.

The changes in K can be decomposed linearly into contributions from changes in $\bar{w}^*$ and $\frac{\partial T}{\partial z}$:

$$\Delta_{\bar{w}^*} K = \left. \frac{\partial K}{\partial \bar{w}^*} \right|_0 \Delta \bar{w}^* = -\left(\left. \frac{\partial T}{\partial z} \right|_0 + \frac{g}{c_p} \right) \Delta \bar{w}^* \tag{B.3}$$

$$\Delta_{\frac{\partial T}{\partial z}} K = \left. \frac{\partial K}{\partial \frac{\partial T}{\partial z}} \right|_0 \Delta \frac{\partial T}{\partial z} = -\bar{w}^*|_0 \Delta \frac{\partial T}{\partial z} \tag{B.4}$$

The notation $|_0$ indicates values in the unperturbed state, i.e. from the piClim-control or piControl simulation. Plugging these cooling rates into Eq. B.2 yields the adiabatic cooling changes ΔQ_{adi} due to changes in $\bar{w}^*$ and $\frac{\partial T}{\partial z}$.
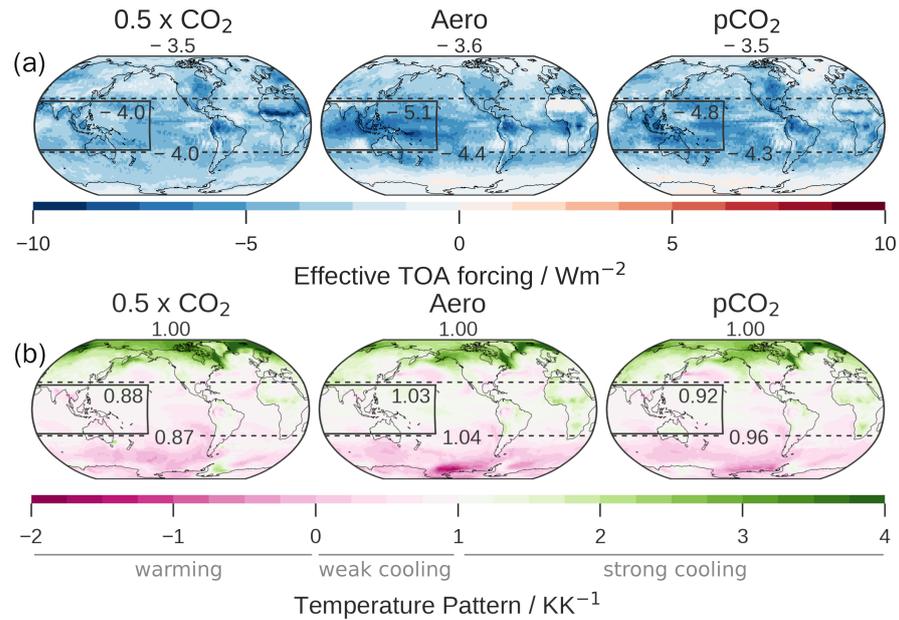
Figure B.3: (a) Annual-mean TOA effective forcing obtained from the difference between the forced simulations with fixed SST and the control simulation with fixed SST. (b) Ensemble-mean 2-meter temperature change patterns (ratio of local to global mean temperature change) of the radiatively forced, coupled simulations. Values greater than one (green) indicate stronger than global average cooling, lower than one (pink) weaker than average cooling. Values lower than zero indicate warming. WP and tropics are shown with solid and dashed lines, and the field average over these regions is shown in the WP box and in the dashed line, respectively. The global mean is shown at the top of each panel. Standard errors are negligible compared to the shown precision in (a). In (b), standard errors of the means over the WP and the tropics are $\approx 0.04 - 0.05\ \mathrm{KK}^{-1}$.

B.3 RESULTS

B.3.1 *Effects of the TOA forcing pattern on the temperature change pattern*

First, we explore the hypothesis that the WP-enhanced forcing pattern from volcanic aerosol causes the WP-enhanced temperature response. The TOA effective forcing fields of the radiatively forced simulations are shown in Fig. B.3 (a). All fields average globally to approximately $-3.5\ \mathrm{Wm}^{-2}$, but exhibit different patterns. Compared to the relatively uniform $0.5 \times CO_2$ TOA effective forcing pattern, Aero exhibits a pronounced forcing pattern. The TOA forcing of Aero is $1.5\ \mathrm{Wm}^{-2}$ more negative in the WP than in the global mean, mainly due to three effects (appendix Fig. B.10): first, stronger instantaneous forcing in the tropics than extratropics due to higher aerosol concentration and insolation ($-1\ \mathrm{Wm}^{-2}$); second, weaker LW effect over the WP than the whole tropics ($-0.2\ \mathrm{Wm}^{-2}$) because the LW effect is weaker over high

clouds than over low clouds or the surface; third, more negative SW cloud adjustments over the WP than the whole tropics (-0.3 Wm^{-2}). The last point may be model-dependent and differs from Marshall et al. (2020), who find strong positive SW cloud adjustments in UK-ESM.

By design, the pCO_2 TOA forcing field shares the Aero forcing field's main features, although it is slightly less enhanced over the WP, which we will address later. The TOA effective forcing fields of Aero and pCO_2 agree well not only in the annual mean, but also in each month (appendix, Figs. B.7 and B.8).

The temperature change patterns of all coupled 10-year simulations are broadly similar (Fig. B.3 (b)). Temperature change is amplified in the Arctic, moderate in low latitudes including the WP, and suppressed over the Southern Ocean. The most relevant region for the feedback is the WP, where small differences have substantial impacts for global feedback, with changes of roughly -0.2 $\mathrm{Wm}^{-2}\mathrm{K}^{-1}$ per 0.1 WPI_T points in MPI-ESM (Fig. B.1). The definition $\mathrm{WPI}_T = \mathrm{d}T_{\mathrm{WP}}/\mathrm{d}T_{\mathrm{global}}$ is equivalent to the WP average of the temperature pattern shown in the black box in Fig. B.3 (b). Despite the fact that the temperature pattern differences among the simulations over the WP are relatively small compared to those in other regions, these small changes dominate the global mean radiative feedback parameter (Dong et al., 2019; Günther et al., 2022).

Although Aero and pCO_2 are almost equally strongly forced in the WP, the WPI of the temperature pattern of pCO_2 (0.92) is smaller than that of Aero (1.03). We perform Student's t-tests on the distributions of WPI_T from the 40 ensemble members of each forcing agent under the null hypothesis that they are drawn from distributions with the same average. While the WPI_T values of $0.5 \times CO_2$ and pCO_2 are not significantly different (p = 0.2), the WPI_T values of Aero and pCO_2 are distinct (p = 10^{-6}). Although the TOA forcing patterns of pCO_2 and Aero are similar, their temperature change patterns are not, when measured by the WPI_T (Fig. B.1). This contradicts the hypothesis that the WP-enhanced TOA forcing pattern of Aero causes the WP-enhanced temperature change pattern. If that were the case, the WP should cool equally strongly in Aero and pCO_2.

This result is limited by the fact that the forcing pattern of pCO_2 is not quite as WP-enhanced as in Aero: The WPI_F (WP forcing divided by global mean forcing) is only 1.36 in pCO_2, but 1.44 in Aero. However, even when applying a correction factor of 1.44/1.36 to the WPI_T values of pCO_2, the WPI_T values remain significantly different from Aero, albeit with higher p-value (p = 0.02).

In summary, the pCO_2 simulation with a TOA forcing pattern almost as WP-heavy as Aero, does not produce a temperature change pattern as WP-heavy as Aero. Instead, its temperature change pattern is rather similar to the $0.5 \times CO_2$ simulation (see also Fig. B.1). Hence,
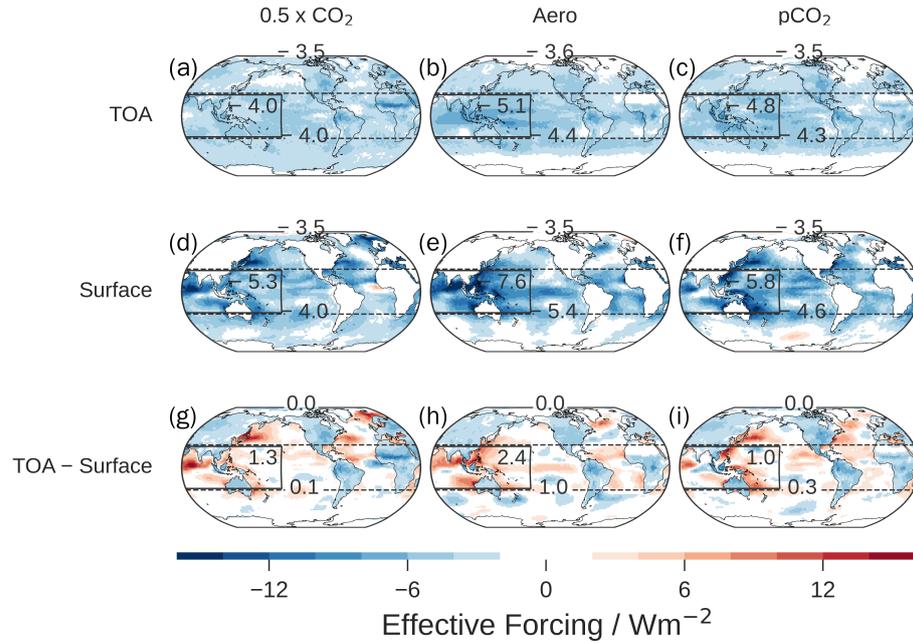
Figure B.4: Effective forcing of the radiatively forced simulations at TOA (row 1), surface (row 2), and their difference (row 3). The difference between TOA and surface effective forcing is the effective forcing on the atmosphere, and can be interpreted as a redistribution of negative forcing from blue to red regions, when going from the TOA to the surface. WP and tropics are shown with solid and dashed lines, and the field average over these regions is shown in the WP box and in the dashed line, respectively. The global mean is shown at the top of each panel. The standard errors are generally negligible compared to the shown precision, except: WP mean of surface and TOA–surface: ≈ 0.2 Wm^{-2}, and tropical mean of surface and TOA–surface: ≈ 0.1 Wm^{-2}.

we arrive at the conclusion that another process which is specific to aerosol forcing must cause the temperature pattern differences.

B.3.2 *Surface forcing pattern*

The rationale of the TOA-forcing hypothesis was that stronger forcing in the WP could lead to stronger temperature change in the WP. However, the ocean does not directly respond to the forcing at the TOA, but the forcing at the surface, which might therefore be more relevant. In the following section, we examine the surface effective forcing and how it differs from the TOA effective forcing.

There are multiple constraints to surface forcing: land and atmosphere have small heat capacities and therefore cannot act as energy reservoirs on time scales of the 100 years fixed SST simulation. If there were substantial fluxes into the atmosphere or the land, they would heat up or cool down until the fluxes become zero. The time scale of these adjustments is fast due to the small heat capacities of land

and atmosphere. Only the ocean (due to its fixed SST) and the TOA can support non-zero fluxes in the steady state of the fixed SST simulations. These considerations imply that the surface effective forcing over land must be zero. The fact that the global-mean flux into the atmosphere is zero implies that the global-mean effective forcings at the TOA and the surface must be equal (see also Eq. B.5).

Fig. B.4 shows the annual mean effective radiative forcing at the TOA, the surface, and their difference, diagnosed from the fixed SST simulations. The surface forcing exhibits a richer spatial structure than the TOA forcing. Furthermore, the WP-intensification of the forcing in Aero is much more pronounced at the surface than at the TOA: At the WP surface, it is more than twice as strong as in the global mean. In comparison, the surface forcing in pCO_2 is only slightly more WP-enhanced than in $0.5 \times CO_2$.

The difference between TOA and surface forcing is the effective forcing that acts on the atmosphere, which is zero in the global mean, but has a pattern. It must be locally balanced by changes in the horizontal heat flux divergence Q, since the atmosphere has no relevant sinks or sources of energy on long time scales.

$$F_{\text{TOA}} - F_{\text{surface}} = F_{\text{atm}} \tag{B.5}$$
$$F_{\text{atm}} + Q = 0 \tag{B.6}$$

F_{TOA} and F_{surface} denote the effective forcing *at* the TOA or *at* the surface, respectively. F_{atm} denotes the effective forcing *on* the atmosphere. All effective forcing fields are a function of longitude and latitude. Both sides of Eq. B.5 globally average to zero.

The forcing on the atmosphere equals the change in horizontal atmospheric heat flux divergence at fixed SST. Taking a perspective from the TOA looking down to the surface, the atmosphere shifts negative forcing from grid points with negative values towards grid points with positive values. Negative forcing is redistributed from columns over land to columns over ocean in all simulations (Fig. B.4, bottom row). This effect arises somewhat artificially from the fact that effective forcing is diagnosed at fixed SST concentrations and sea ice concentrations, but not fixed land temperatures. In Aero there is an additional convergence of negative forcing at the WP surface (2.4 ± 0.2 Wm^{-2} compared to 1.0 ± 0.2 Wm^{-2} in pCO_2 and 1.3 ± 0.2 Wm^{-2} in $0.5 \times CO_2$).

We argue that the surface forcing is the critical factor that distinguishes aerosol from CO_2 forcing. The differences between TOA and surface forcing are imposed by heat transport changes that arise from anomalous circulations. They come about as adjustments which are specific to the forcing agent. We hypothesize that the WP surface in Aero is so strongly forced because the anomalous atmospheric circulation leads to an anomalous energy transport out of the WP, or,

equivalently, moves positive forcing away from the WP surface. In the following sections, we aim to (1) explain how the differences in atmospheric forcing divergence arise from the circulation changes, and (2) determine if these differences cause the distinctions among the temperature change patterns.

B.3.3 *Explaining the atmospheric forcing divergence*

What explains the different structures of the forcing on the atmosphere (bottom row of Fig. B.4)? Most of the spatial structure arises from variations in the latent heat flux at the surface (not shown). Therefore, it is mostly the latent heat flux that reacts to energetic constraints from the atmosphere, consistent with previous studies (Fajber et al., 2023; Fajber and Kushner, 2021). However, this does not explain why the atmosphere redistributes energy, and which circulations accomplish this energy transport.

Eq. B.6 states that the forcing on the atmosphere is balanced by heat transport. We focus on the anomalous energy export out of the WP, and on the question why this anomaly is stronger for Aero. To this end we partition the energy export into a meridional component, i.e. the energy transport from the tropics to the extratropics, and a tropical-zonal component, i.e. the transport from the WP to tropical non-WP regions. In the "TOA - Surface" row of Fig. B.4, the meridional energy transport is equal to the average over the tropics (1.0 ± 0.1 Wm^{-2} in Aero, compared to 0.1 ± 0.1 Wm^{-2} in $0.5 \times CO_2$), and the zonal energy transport is measured by the difference between the WP mean and the tropical mean (1.4 ± 0.2 Wm^{-2} in Aero compared to 1.2 ± 0.3 Wm^{-2} in $0.5 \times CO_2$).

Obvious explanations for these transports could be changes in gross moist stability, the Hadley or Walker circulation, or eddy-energy flux. Indeed, the tropical tropospheric zonal overturning circulation has been shown to *weaken* as a consequence of stratospheric heating (Ferraro et al., 2014; Simpson et al., 2019). However, a weaker zonal overturning circulation cannot be the reason for increased energy flux out of the WP, unless it is overcompensated by increases in gross moist stability. We argue that the circulation changes associated to the atmospheric energy budget do not project onto the Hadley and Walker circulations. Instead, the meridional transport is accomplished via the BDC, and the tropical-zonal transport arises from an anomalous ocean-land circulation. This only applies to the direct circulation adjustments, and we do not make any statement about temperature-dependent changes to the Hadley or the Walker circulation from stratospheric aerosol forcing. The meridional and tropical-zonal component will be treated separately in the following.

Table B.2: Anomalous adiabatic cooling in the stratosphere calculated according to Eq. B.2 and averaged over the tropics (30°N to 30°S). Negative values represent cooling. Standard errors are only shown where they are at least on the order of the precision shown, i.e. $O(0.1 \text{ Wm}^{-2})$. The lower rows show the contributions from changes in the upwelling speed and the lapse rate, respectively.

	$0.5 \times CO_2$	Aero	pCO$_2$	Aero (non-absorbing)
ΔQ_{adi} / Wm^{-2}	0.0	-0.9	-0.1 ± 0.1	0.0
ΔQ_{adi} due to $\Delta \bar{w}^*$ / Wm^{-2}	0.2	-0.7	0.1 ± 0.1	0.0
ΔQ_{adi} due to $\Delta \partial T / \partial z$ / Wm^{-2}	-0.2	-0.2	-0.2	0.0

B.3.3.1 *Meridional energy transport: adiabatic cooling from the Brewer-Dobson circulation*

We propose that the meridional transport of energy from the tropics to the extratropics arises from increased adiabatic cooling in the tropics via the BDC. Stratospheric aerosol causes a meridional heating gradient in the stratosphere, which affects the wave propagation and therefore the wave driving of the BDC (Garcia and Randel, 2008). An increase in adiabatic cooling could arise from an acceleration of the vertical velocity (i.e. the BDC) or an increase of the vertical temperature gradient (see Eq. B.1). The changes in adiabatic cooling in the fixed SST simulations compared to piClim-control, calculated according to Eq. B.2, are shown in Table B.2. In the tropics, the BDC causes an additional adiabatic cooling of 0.9 Wm^{-2} in Aero, 0.0 Wm^{-2} in $0.5 \times CO_2$, and 0.1 Wm^{-2} in pCO$_2$. In Aero, most of this is driven by changes in the upwelling speed, and only a small part is due to changes in the stratospheric lapse rate. While the upwelling speed influences the adiabatic cooling proportionally according to Eq. B.2, changes in the temperature profile influence adiabatic cooling only in so far as they change the difference between the actual lapse rate and the dry adiabatic lapse rate. Since this difference is already quite large in the unperturbed stratosphere (10 K/km dry adiabatic lapse rate vs. -2 K/km stratospheric lapse rate), moderate changes to the stratospheric lapse rate will only have a small effect on the adiabatic cooling. However, eventually all changes to adiabatic cooling are driven by changes to the global stratospheric temperature distribution, since it is the differential heating between equator and pole that affects wave propagation and therefore the speed of the BDC (Garcia and Randel, 2008). A measure for the BDC strength is the tropically averaged residual mean vertical velocity $\bar{w}^*$, which increases in Aero by 10±2 % at 70 hPa and 24±2 % at 30 hPa. In the CO$_2$-forced simulations, these changes are on the order of the uncertainty.

Adiabatic cooling is not a sink of energy in the global energy budget. The energy is released during the sinking motion in the extratropics,

and therefore constitutes an energy transport from the tropics to the extratropics (Richter et al., 2017). The mechanism of energy export due to the BDC explains the meridional energy transport of slightly less than 1 Wm^{-2} from tropics to extratropics seen in Fig. B.4 (h). The forcing anomalies are communicated from the stratosphere to the tropical tropopause layer and the free troposphere via radiation, and then passed on via convection to the surface.

We turn to the additional simulation with fixed SST and a prescribed aerosol forcing similar to Aero, where the aerosol is modified to only scatter, but not absorb radiation. This precludes the aerosol from heating the stratosphere, which should in turn prevent the BDC mechanism. Consistent with our expectations, we find no anomalous energy transport from the tropics to the extratropics (Table B.2).

Note that an increase in adiabatic cooling does not imply that the tropical stratosphere becomes colder - it just warms less than it would if there was no acceleration of the BDC. The stratosphere heats in Aero and the CO_2-forced simulations, leading to a small additional adiabatic cooling from the increased lapse rate. However, only in Aero the BDC accelerates considerably, which provides the bulk of the adiabatic cooling effect.

B.3.3.2 *Tropical-zonal energy transport: Ocean-to-land circulation*

All simulations exhibit a zonal energy transport from the WP to the tropical non-WP regions. This tropical-zonal energy transport is not a big contributor to the differences between the Aero and $0.5 \times CO_2$ forcing patterns and therefore not a focus of our study. Nevertheless, we briefly lay out the reasons for this anomalous circulation. The explanations are somewhat rooted in the way forcing is diagnosed in models, and only partially apply to the real world.

Since land temperatures vary freely in the fixed SST simulations, the land cools down rapidly, which leads to an enhanced energy flux from the atmosphere to the land. For the atmospheric energy budget to be closed, this energy must be replenished from the ocean, which can draw from an infinite energy reservoir due to the fixed SST ocean surface. Consequentially, an anomalous energy transport arises from ocean towards land (see appendix Fig. B.9 and accompanying text). This is accomplished by land-to-ocean winds at the surface, and ocean-to-land winds aloft, which is also the direction of the energy flow. Since deep convection is impeded over non-WP regions, the anomalous circulation predominantly transports energy from WP ocean regions to the tropical land regions. For this reason, there is an additional energy export from the WP to the non-WP regions in all fixed SST simulations.

Previous studies have noted the importance of land-ocean temperature contrast and a resulting monsoon-like circulation in the fast response to abrupt forcing (Heede et al., 2020; Modak et al., 2016).

While this circulation arises artificially in our fixed-SST simulations from the fact that land temperatures are not fixed, it would appear similarly in reality because land reacts much faster than ocean. This effect is important on a time scale of several months (Modak et al., 2016), which is comparable to the time scale of volcanic aerosol forcing. Our simulations show that this circulation causes an energy export out of the WP, and might therefore contribute to a strengthening of the feedback to volcanic eruptions, simply due to the time scale they act on. For long-term forcing such as anthropogenic greenhouse gases and solar radiation management, this effect would be less important. This raises the question how the different time scales of volcanic aerosol and CO_2 forcing affect the feedback.

Further explanations and a figure showing the anomalous circulation can be found in appendix B.5.2.

B.3.3.3 *Sum of meridional and zonal terms*

In total, the atmosphere exports $1.3 \, \mathrm{Wm^{-2}}$ out of the WP in the $0.5 \times CO_2$ simulation, but $2.4 \, \mathrm{Wm^{-2}}$ in Aero, which leads to enhanced negative forcing at the WP surface in Aero (Fig. B.4). In Aero, the atmosphere absorbs $0.9 \, \mathrm{Wm^{-2}}$ of positive forcing in the tropics and exports it to the extratropics via the BDC (Table B.2). The energy transport via the BDC explains most of the meridional energy transport, and the difference between the surface forcing in the Aero and $0.5 \times CO_2$ experiments. The cooling of the land surface causes an additional zonal transport of energy from the WP ocean to tropical land regions via the free troposphere in all simulations. While the anomalous meridional transport via the BDC does not show substantial year-to-year variation, the zonal energy transport varies substantially interannually. Even using 100 years of fixed SST simulations, its standard error (standard deviation in time divided by square root of sample size) is on the order of 10 to 15 %, which is an order of magnitude higher than the standard error of the meridional transport. This implies the existence of substantial interannual variability, which originates from the atmosphere alone. Furthermore, the tropical-zonal energy transports of Aero and $0.5 \times CO_2$ differ only within one standard error, and arise at least partly from the specifics of fixed SST simulations, where temperatures are only prescribed at the ocean surface, but not at the land surface. We therefore emphasize the BDC changes as the more important result, and caution with the interpretation of the zonal energy transport. In the following, we examine the forcing patterns and the meridional energy transport mechanism in CMIP6 models.
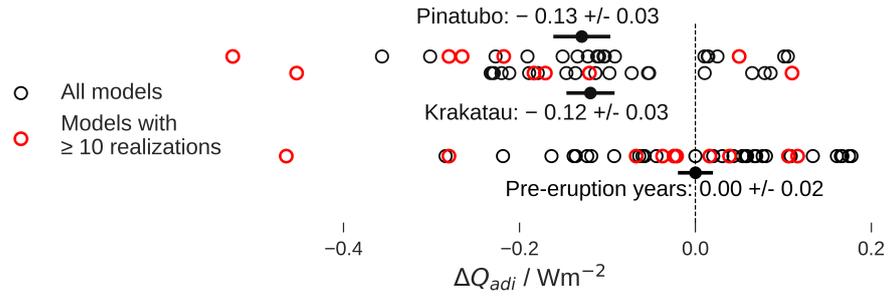
Figure B.5: Tropical mean changes in stratospheric adiabatic cooling in the CMIP6 historical simulations in the post-eruption years of Krakatau (1884) and Pinatubo (1992) and the pre-eruption years (1882 and 1990), compared to the piControl simulations. The value from each model is shown by a circle. The green circles are from the five models with at least ten realizations. Together, these models account for 80 % of all realizations. All values are calculated using Eq. B.2. The error-bars show the multi-model mean and standard error. Negative values indicate increased adiabatic cooling.

B.3.4 *Testing the energy export mechanism with CMIP6 models*

Salvi et al., 2023 provide an analysis of CMIP6 forcing patterns at the TOA and the surface (see their Fig. 10). For aerosol forcing from the Krakatau and Pinatubo eruptions, they find enhanced forcing in the tropics at the TOA and at the surface. However, while the WP-enhancement of aerosol forcing at the surface is clearly visible, it it not substantially stronger than in the case of greenhouse gas-forcing. This calls into question the existence of the BDC mechanism in CMIP6 models.

We explicitly test if the energy export due to the BDC increases after volcanic eruptions. 23 models provide the necessary output to compute the adiabatic cooling in the historical coupled simulations. We compute the changes to the adiabatic cooling in the tropics according to Eq. B.2 in the first post-eruption year of the Krakatau and Pinatubo eruption, respectively (1884 and 1992). Since we restrict the computations to the tropics where the eddy contribution is small, we simplify the analysis by using w instead of $\bar{w}^*$ for the CMIP6 analysis. In MPI-ESM, this leads to an underestimate of the additional adiabatic cooling from aerosol forcing of approximately 10 %.

Most CMIP6 models show a moderate increase in adiabatic cooling, with a multi-model average of 0.12 / 0.13 Wm^{-2} for the Krakatau / Pinatubo eruption, respectively (Fig. B.5) The multi-model mean +/- standard error does not overlap with zero, indicating the presence of a significant effect. The 20 to 80 % intervals are: [-0.22, -0.01] for Krakatau, [-0.25, 0.01] for Pinatubo, [-0.11, 0.05] for the pre-eruption years. The adiabatic cooling in the pre-eruption years (1882 and 1990) is indistinguishable from the control simulation, which implies that

the additional adiabatic cooling in the post-eruption years is indeed caused by the eruption and not some other historical forcing agent.

The effect is smaller than what we find in the fixed SST simulations of MPI-ESM: Assuming a peak global mean effective forcing from Krakatau and Pinatubo of approximately 1.8 Wm^{-2} (Salvi et al., 2023), the BDC cools the tropics by about 7 % of the global mean forcing in CMIP6, compared to about 25 % in our simulations. This difference is diminished when only taking into account the models with at least 10 realizations, where the influence of internal variability is reduced. In these models, the transport from the BDC is -0.16 [-0.24, -0.07] Wm^{-2} for Krakatau and -0.25 [-0.33, -0.16] Wm^{-2} for Pinatubo, corresponding to roughly 9 % and 14 % of the global mean forcing, respectively. Apart from model-differences, the remaining disparity between our results and the CMIP6 results could arise for three systematic reasons: First, using w instead of $\bar{w}^*$ leads to a small underestimation. Second, the aerosol is short-lived and not all of the post-eruption year is equally strongly affected by the presence of the aerosol. Third, the CMIP6 estimate is likely biased low, because the historical coupled simulations cool in the post-eruption year, and there is a positive correlation between temperatures and the strength of the BDC (Garfinkel et al., 2017). The last point does not compromise our finding that the BDC reshapes the surface forcing pattern, since the forcing is defined at zero surface temperature change.

In agreement with our CMIP6 analysis, model studies consistently show an acceleration of the BDC after volcanic eruptions (Aquila et al., 2013; Garcia et al., 2011; Garfinkel et al., 2017; Muthers et al., 2016; Pitari and Mancini, 2002; Pitari and Rizi, 1993; Toohey et al., 2014). In contrast, studies using reanalysis or observations provide mixed results. Some studies find enhanced wave activity in at least one hemisphere (Graf et al., 2007; Schnadt Poberaj et al., 2011), others do not find stratospheric circulation changes after volcanic eruptions (Diallo et al., 2012; SPARC, 2022; Seviour et al., 2012). Since reanalysis products do not assimilate aerosol data and are hence ignorant to the ensuing heating rate anomalies in the stratosphere, they might not be a suitable tool to study the links between stratospheric heating and the BDC (Abalos et al., 2015). The absence of the upwelling effect in observational records might also be related to the choice of the metric: Toohey et al. (2014) argue that the upwelling change might be most pronounced in the middle and upper stratosphere, while observational studies focus on upwelling in the lower stratosphere.

b.3.5 *Does the surface forcing pattern cause the surface temperature pattern?*

If the previously demonstrated differences in the surface forcing cause the temperature pattern differences, then these differences should
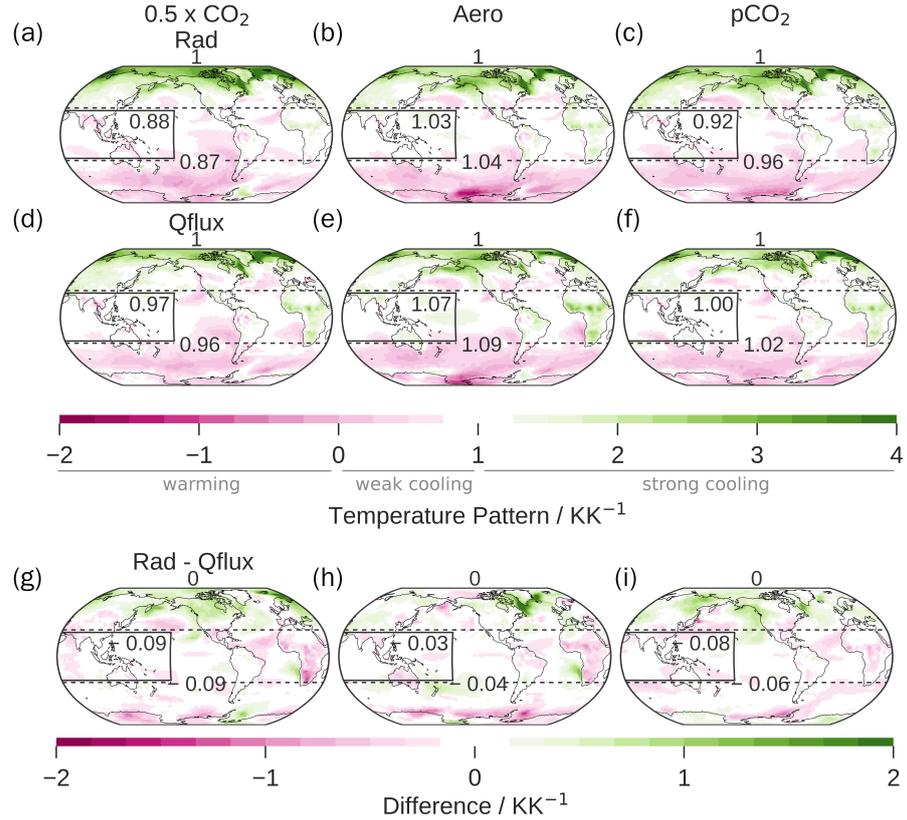
Figure B.6: Ensemble-mean 2-meter temperature change patterns (ratio of local to global mean temperature change) of the radiatively forced simulations (a - c), q-flux-forced simulations (d - f), and their difference (g - i). In panels (a) to (f), Values greater than one (green) indicate stronger than global average cooling, lower than one (pink) weaker than average cooling. Values lower than zero indicate warming. In (a) - (f), the standard errors of the means over the WP and the tropics are $\approx 0.04 - 0.05$ KK^{-1}. For the differences (g - i), the standard errors are ≈ 0.07 KK^{-1}.

appear in surface-forced simulations without any changes to aerosol or CO_2 concentrations. While it seems intuitive, it is not straightforward that stronger surface forcing in the WP causes stronger temperature change in the WP. Results from q-flux Green's functions have shown that the temperature response to a localized surface flux is typically non-local (Lin et al., 2021; Liu et al., 2018a,b, 2022). We use the q-flux simulations to establish the link between local forcing and local temperature change in the WP. They are forced by a surface heat sink / source, each of them with a global mean of -3.5 Wm^{-2}, but with the patterns diagnosed from the radiative forcing simulations (Fig. B.4 d-f).

Results from the q-flux simulations in the space of temperature pattern and feedback are shown in Fig. B.1. The simulations with more strongly concentrated surface forcing in the WP also cause stronger temperature change in the WP (Aero > pCO$_2$ > 0.5 $\times$ CO$_2$). These

results support our hypothesis that stronger surface forcing in the WP leads to stronger surface temperature change in the WP. However, the temperature patterns differ between the q-flux forced and the corresponding radiatively forced simulations (Fig. B.6), in particular with respect to temperature changes in the WP and the tropics. All q-flux forced simulations cool more strongly in the WP and in the tropics than their radiatively forced counterparts. Generally, the pattern differences between radiatively and q-flux forced simulations of the same forcing agent are in the same order of magnitude as the differences among the forcing agents. Possible reasons for these deviations are: (i) the lack of changes to the atmospheric CO_2 concentrations / aerosol load, which affects the vertical structure of the atmosphere and the general circulation; (ii) the lack of forcing over land; (iii) the fact that we only include heat fluxes, but no momentum or freshwater fluxes to force the ocean surface. Aquaplanet simulations show little difference between radiatively forced and q-flux forced simulations (Haugstad et al., 2017), rendering hypothesis (i) - the absence of CO_2/ aerosol in the atmosphere - as a cause for the differences unlikely. On the other hand, both CO_2 forced simulations show a more positive PDO-like temperature change pattern in the radiatively forced, compared to the q-flux-forced simulations, which could be an indication that this part of the temperature pattern is due to the direct effect of CO_2. In CESM2 simulations with historical forcing, the absence of wind stress forcing causes statistically significant changes to the SST pattern (McMonigal et al., 2023) towards a more WP-enhanced temperature pattern, consistent with the bias in our simulations and with hypothesis (iii), the momentum forcing-hypothesis.

The exact causes are beyond the scope of this study and warrant further research. Despite the shortcomings, we interpret the results from the q-flux forced simulations as a support for our hypothesis that strong surface forcing in the WP leads to strong surface cooling in the WP.

B.4 DISCUSSION AND CONCLUSIONS

In this study, we identify a mechanism that redistributes energy from the tropics to the extratropics via an accelerated BDC due to stratospheric heating. Using MPI-ESM simulations of idealized CO_2 and aerosol forcing we explain the pronounced WP cooling from stratospheric aerosol forcing with the strongly negative forcing it causes at the WP surface. This finding enhances our understanding of the formation of the WP-enhanced temperature change pattern in response to stratospheric aerosol forcing, which has previously been shown to cause strongly negative feedback (Günther et al., 2022; Salvi et al., 2023; Zhou et al., 2023).

The effective forcing from stratospheric aerosol is more negative in the WP than in the global mean already at the TOA. Furthermore, there is a substantial export of energy from the tropics to the extratropics via the stratosphere, effectively removing additional energy from the tropical surface. The stratospheric energy export emanates mainly from an acceleration of the BDC, which leads to increased adiabatic cooling in the tropical stratosphere and adiabatic heating in the extratropical stratosphere. Changes in the BDC ultimately arise from the differential heating between the tropical and the extratropical stratosphere, which affects wave activity and therefore the strength of the stratospheric pump (Graf et al., 2007; Holton et al., 1995; Schnadt Poberaj et al., 2011).

The time-constant forcing we use to model stratospheric aerosol forcing is reminiscent of strategies to cool the Earth with solar radiation management by deliberate injection of reflective aerosol into the stratosphere. Depending on the location and absorptivity of the used aerosol the BDC will accelerate and lead to stronger cooling of the tropics than the extratropics. Tropical overcooling is a notorious problem of solar radiation management, unless more sophisticated injection strategies are used (Kravitz et al., 2019; Laakso et al., 2017). The importance of the BDC for the climate response corroborates the finding from previous studies that solar dimming is an imperfect substitute for simulating aerosol forcing (Ferraro et al., 2014; Simpson et al., 2019; Visioni et al., 2021).

In order to highlight differences between aerosol and CO_2 forcing independent of the forcing pattern, we created a patterned CO_2 simulation, which approximately reproduces the TOA effective forcing pattern of stratospheric aerosol. This was achieved by varying the CO_2 concentration in space and time. The aerosol and the patterned CO_2 simulation have more negative TOA radiative forcing in the WP and the tropics than in the global mean. Despite their similar TOA effective forcing patterns, they exhibit substantial temperature pattern differences. We therefore argue that the increased energy export out of the tropics due to the acceleration of the BDC is essential for the emergence of the tropically enhanced temperature change pattern and strong feedback to stratospheric aerosol forcing. This shows that the TOA forcing perspective is not sufficient to explain the temperature patterns. The TOA perspective has previously been used as an explanation for the feedback to aerosol forcing (Salvi et al., 2022, 2023), to solar forcing (Kaur et al., 2023; Modak et al., 2016), and for explaining temperature change patterns in general (Liu et al., 2022). We show that the temperature pattern is more closely linked to the pattern of forcing at the surface than at the TOA. The TOA forcing perspective is established in climate science, which is appropriate as long as the focus is on global means, but the patterns of TOA and surface forcing can be substantially different. Differences between them

arise from changes in the atmospheric heat transport, which lead to a considerable redistribution of forcing between different regions of the Earth. This should be kept in mind when addressing the relationship between forcing patterns and temperature change patterns in future studies.

The comparison of radiatively forced simulations with simulations that were forced with an equivalent heat flux forcing at the surface, reveals the existence of a link between surface forcing pattern and surface temperature response. Stronger surface forcing in the WP produces stronger temperature change in the WP. Still, knowledge of the surface effective forcing in our simulations is not enough to reproduce the exact temperature response.

For the interpretation of our results it should be kept in mind that the aerosol in our model has a highly idealized profile with no seasonal dependence, and is not transported. The ozone profile is fixed, although ozone is affected by the presence of aerosol and has been shown to affect the BDC (Garfinkel et al., 2017; Pitari and Rizi, 1993; Schnadt Poberaj et al., 2011). Comparing aerosol-forced simulations in a single model with and without interactive ozone chemistry, Richter et al. (2017) find slightly higher upper stratospheric upwelling in the simulation without atmospheric chemistry, and almost no tropical temperature differences.

Furthermore, shifting the aerosol profile in altitude or latitude would likely modify the effect on the BDC, so that our results may be dependent on the specific aerosol profile we chose. Aerosol that is injected at greater altitude has been found to cause less negative feedback (Lee et al., 2023; Zhao et al., 2021). Lower injections allow more water vapor to enter the stratosphere, because they more strongly affect cold point temperatures. Lee et al. (2023) argue that this leads to a negative water vapor feedback. Since the increased stratospheric water vapor from cold point heating appears on a time scale of months (Kroll et al., 2021) and independent of surface temperature, we suggest that this does not constitute a feedback, but rather an adjustment. According to our results, the altitude dependence of the feedback could be related to the altitude dependence of the effect of stratospheric heating on the BDC. In addition to the dependence of feedback on altitude, we also expect a dependence on the meridional profile. Extratropical eruptions would not only cause a less WP-enhanced TOA forcing pattern. They also affect the BDC differently (Richter et al., 2017) and might therefore lack the WP-enhancement of the surface forcing. Our results are therefore not necessarily applicable to aerosol forcing with pronounced hemispheric asymmetries.

In recent years, much progress has been made to understand how patterns of SST affect radiative fluxes. Especially SST Green's functions provide a detailed picture about the importance of tropical convective regions for radiative feedbacks. However, less is known about how

these SST patterns come about. Simulations with q-flux Green's functions and slab ocean and pacemaker experiments indicate that heat fluxes over the Southern Ocean play an elevated role for SST pattern formation (Hu et al., 2022; Hwang et al., 2017; Kang et al., 2023; Lin et al., 2021; Liu et al., 2018a,b, 2022). Dynamic ocean and atmosphere processes make the temperature pattern time-dependent, even for constant forcings (Heede et al., 2020). Yet, the mechanistic picture of the connection between forcing pattern and SST pattern is still incomplete. While many pieces are missing on the way to complete this picture, we contribute to filling this gap by identifying relevant processes that cause differences between TOA and surface forcing, emphasizing the relevance of the latter, and by pointing out the atmospheric pathway from TOA forcing to surface forcing to surface temperature pattern specifically for stratospheric aerosol forcing.

B.5   APPENDIX

B.5.1   *Finding a field of CO_2 concentrations that matches the effective TOA forcing of Aerosol*

B.5.1.1   *Approach*

The goal is to find a field of CO_2 concentrations whose effective forcing is equal to the effective forcing field from the stratospheric aerosol ("Aero"), which is known. The idea of the algorithm is to start with an initial guess, compute its effective forcing, and then iteratively increase the CO_2 concentration wherever the effective forcing is too negative, and decrease it wherever the effective forcing is too positive, taking into account the logarithmic dependence of forcing on CO_2 concentration.

B.5.1.2   *Algorithm*

Let x be the CO_2 concentration in units of the pre-industrial CO_2 concentration (284 ppm), as function of longitude, latitude, and time. The goal is to find a target field x_t, whose effective TOA forcing F matches a given target F_t. The indices t, i will be used in the following to indicate initial and target fields.

Instantaneous CO_2 radiative forcing approximately follows the relationship

$$F \approx c \cdot \log_2(x) \tag{B.7}$$

Eq. B.7 holds approximately in the global mean with $c \approx 3.7 \text{ Wm}^{-2}$ (Myhre et al., 1998). However, the instantaneous radiative forcing at each location is determined by the local difference between the temperatures at the surface and the tropopause (Jeevanjee et al., 2021),

resulting in a non-uniform TOA instantaneous forcing pattern, even for a uniform change in CO_2 concentration. Atmospheric adjustments and noise cause further departures of the effective forcing from the instantaneous forcing, both in its pattern and global mean.

From Eq. B.7 it follows that

$$F_t / F_i \approx \log_2(x_t) / \log_2(x_i) \tag{B.8}$$

$$x_t \approx 2^{F_t / F_i \cdot \log_2(x_i)} \tag{B.9}$$

$$x_t = x_i^{F_t / F_i} + \epsilon \tag{B.10}$$

(see also Xia and Huang, 2017). In the last step, the approximation symbol is replaced by inclusion of an error term ϵ. The effective forcing field F of any forcing can be computed as the difference of TOA fluxes between a perturbed and an unperturbed simulation with fixed SST and sea ice. Therefore, for any x_i, F_i can be determined. Using Eq. B.10, one can compute a field x_t that will have the desired forcing field F_t, up to an error term ϵ. Using x_t as the new x_i, ϵ can be minimized by repeatedly applying Eq. B.10 to each horizontal grid point. The target F_t remains the same in all iterations.

It is not a priori clear that this algorithm converges. In fact, a few modifications must be made due to errors from adjustments and noise. We apply these modifications in every step.

1. In Eq. B.10, $F_t / F_i \rightarrow \pm\infty$ for $F_i \rightarrow 0$. To address this problem we set F_i in the calculation to ± 0.5 Wm^{-2} wherever its absolute value is smaller than 0.5 Wm^{-2}.

2. In our specific case F_t is generally negative, but positive in some places, especially near the poles. In case $F_i < 0 < F_t$, problems arise when $x_i > 1$. From Eq. B.7 this is not generally expected ($x_i > 1$ is associated with $F_i > 0$), but can happen due to adjustments or noise. This violation of the assumptions leads to local divergence of the algorithm. One way to fix this is to set $x_t = x_i + c$ wherever $F_i < 0 < F_t \wedge x_i > 1$, instead of applying Eq. B.10. We arbitrarily choose $c = 0.5$. The idea is to force the algorithm to increase the CO_2 concentration when it is too low in cases where it would normally decrease it. Similarly, we set $x_t = x_i - c$ wherever $F_i > 0 > F_t \wedge x_i < 1$ to account for the opposite case.

3. After applying Eq. B.10, we apply a moving average filter to $\log_2(x)$ with window length of 36° longitude and 19° latitude in order to smooth the spatial variations.

4. After the moving average filter, we restrict CO_2 concentrations to $1/4 < x_t < 4$ in the interest of avoiding too extreme variations.

With the modifications in place, the algorithm converged according to our subjective judgment after three iterations. The resulting field of CO_2 concentrations is shown in Fig. 2 (b) of the main manuscript, the effective TOA forcing field is shown in Fig. 3 (a) of the main manuscript. Both clearly share large-scale features, e.g. the most negative forcing and the most strongly reduced CO_2 concentration over the WP, but they differ on smaller scales due to adjustments and noise. Monthly-mean fields of effective TOA forcing and CO_2 concentrations of pCO$_2$ in comparison to Aero are shown in Figs. B.7 and B.8.
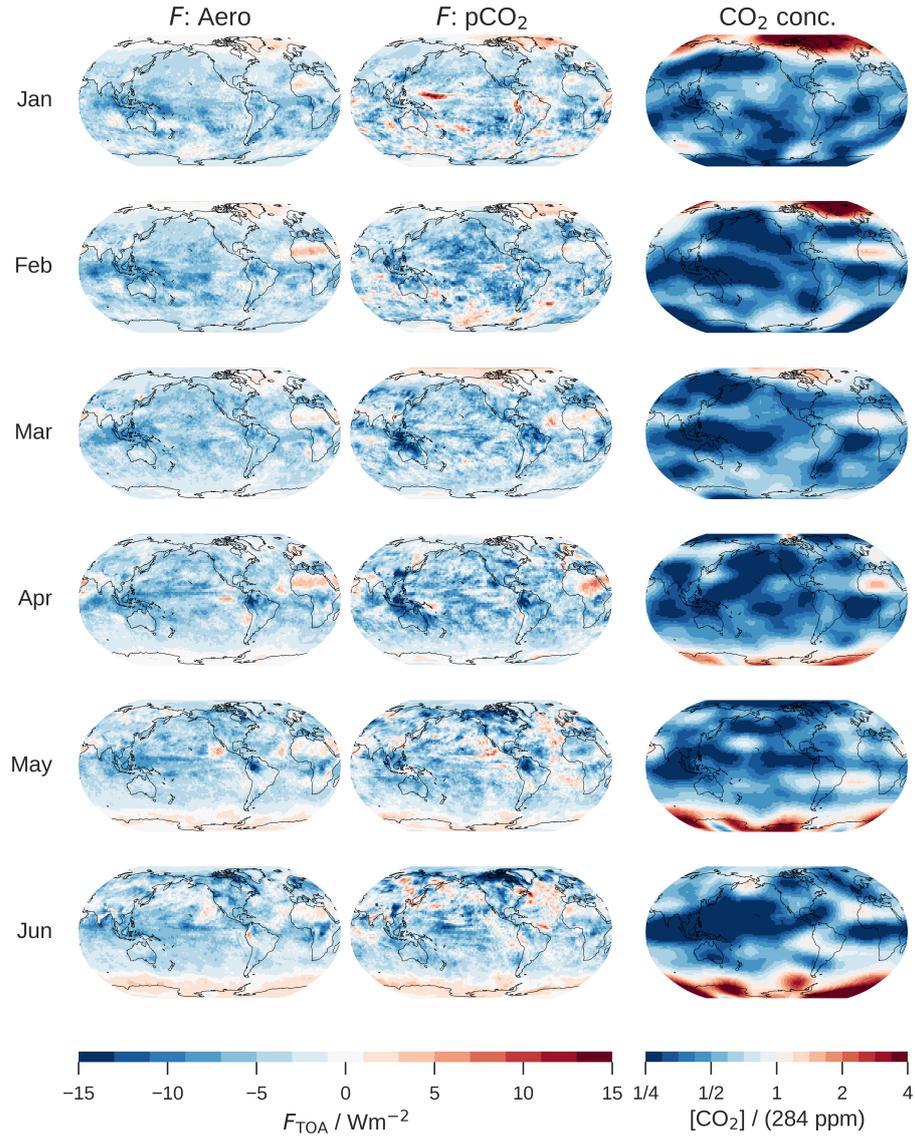
Figure B.7: Monthly comparison of the effective TOA forcings of Aero and pCO$_2$: January - June. The right column shows the field of CO_2 concentrations that results in the effective forcing of the middle column. Note the seasonal dependence of the forcing, with positive forcing over the pole of the winter hemisphere. The seasonal dependence arises despite the time-invariant aerosol profile from the seasonally varying insolation.
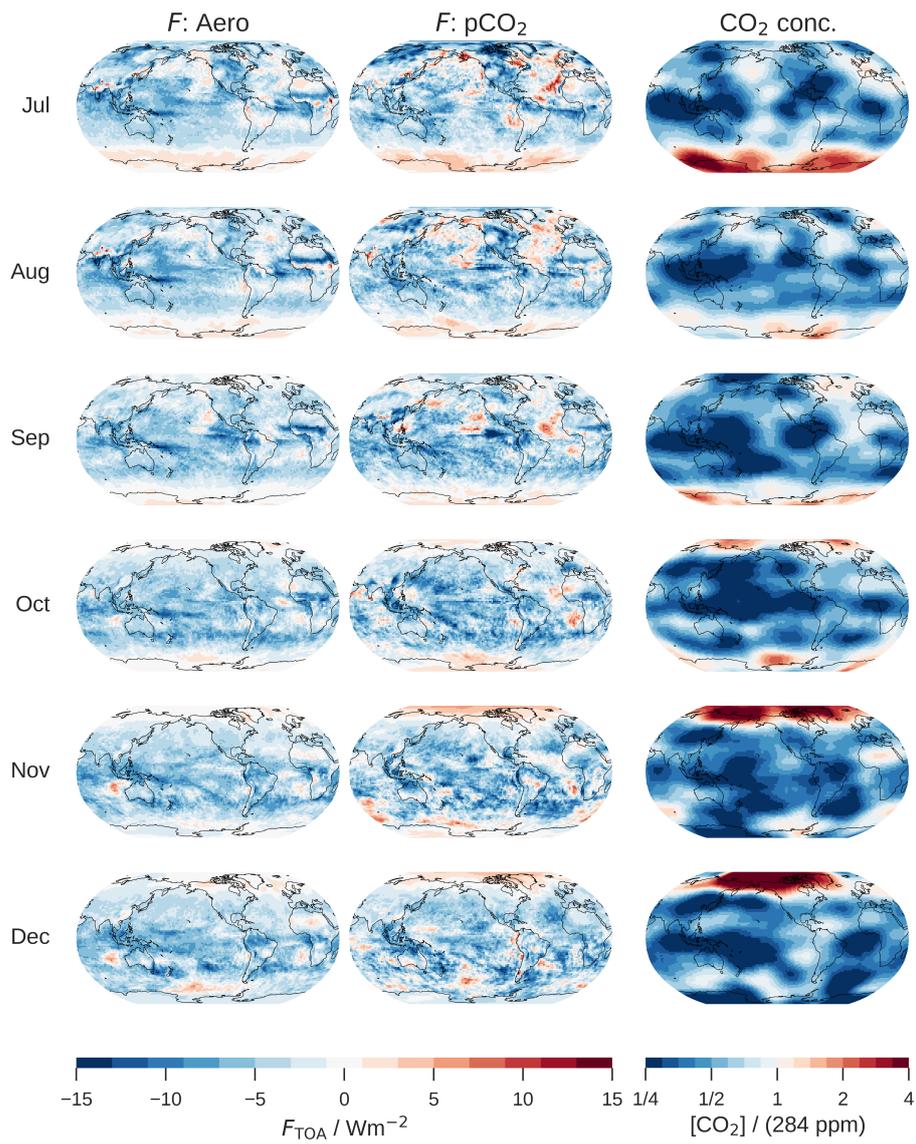
Figure B.8: Same as Fig. B.7, but for July - December

B.5.2   *Tropical-zonal circulation from ocean to land*

Although not the main focus of this study, we provide a short explanation why there is a tropical-zonal energy transport from the WP to tropical non-WP regions in all fixed SST simulations (see section 3.3 of the main manuscript).

Fig. B.9 shows histograms of variables that indicate an anomalous ocean-land circulation in the fixed SST simulations. Over land, the atmosphere loses energy by radiative and turbulent fluxes almost everywhere, because the land can cool down in the fixed SST simulation, while the ocean can not. This energy loss is compensated by adiabatic heating due to more pronounced downward motion over land.

The energy that is transported to the atmosphere over land is supplied from the atmosphere over ocean, which gains more energy from turbulent and radiative fluxes. The air rises more strongly over ocean, associated with more precipitation and hence convective heating. The circulation must necessarily be closed by movement of air from ocean to land aloft, and from land to ocean near the surface.

The top and middle row show that the WP ocean regions provide proportionally more energy than the tropical non-WP ocean regions, indicated by more positive forcing on the atmosphere and a stronger increase in precipitation. Since the WP is a major region of deep convection, it effectively couples the surface to the free troposphere, which enables energy transport from the surface to the free troposphere and subsequently to the land regions. The prevailing inversion over, e.g., the Eastern Pacific impedes this energy transport. Therefore the majority of the energy transport happens from WP ocean regions to land regions.

This picture is qualitatively similar for the Aero, 0.5 x CO_2 and pCO_2 simulations, because it does not directly depend on the presence of the forcing agent. Instead, it emerges as a consequence of the air-sea contrast that arises from fixing SST, but not land temperatures. A real-world analog might be the fast response to forcing, where land temperatures react more quickly than SST, such as after volcanic eruptions.
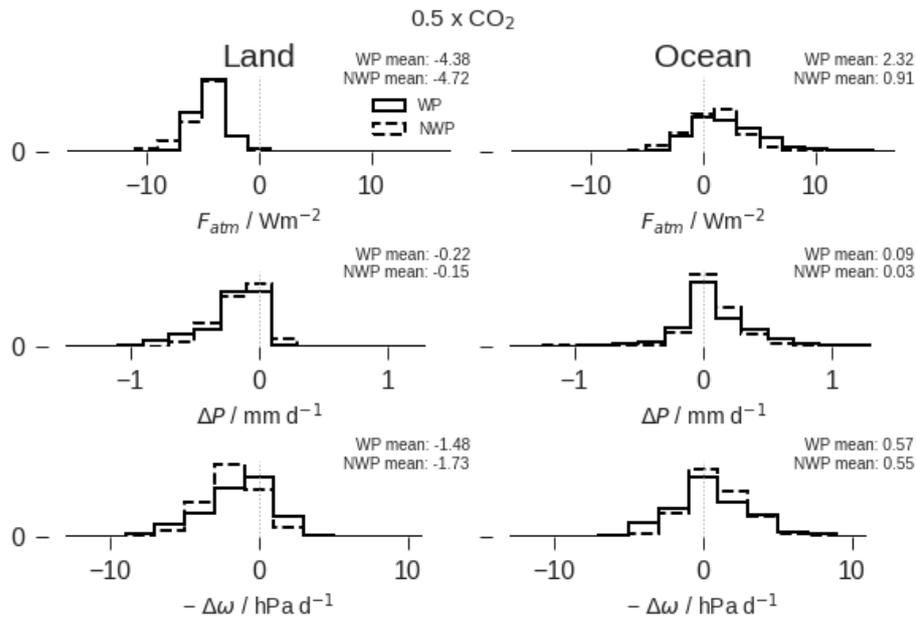
Figure B.9: Distribution of changes over the tropics, separated into land / ocean and WP / tropical non-WP regions, indicating an ocean-land circulation. Values are taken from the 0.5 x CO_2 simulation (Aero and pCO_2 qualitatively similar). All histograms show changes of the fixed SST simulation compared to piClim-control. Top: Effective forcing on the atmosphere (= TOA - surface forcing). Middle: Precipitation. Bottom: Negative pressure velocity at 500 hPa (positive values indicate rising motion). NWP = tropical non-WP regions.

B.6   COMPONENTS OF THE TOA FORCING OF AERO
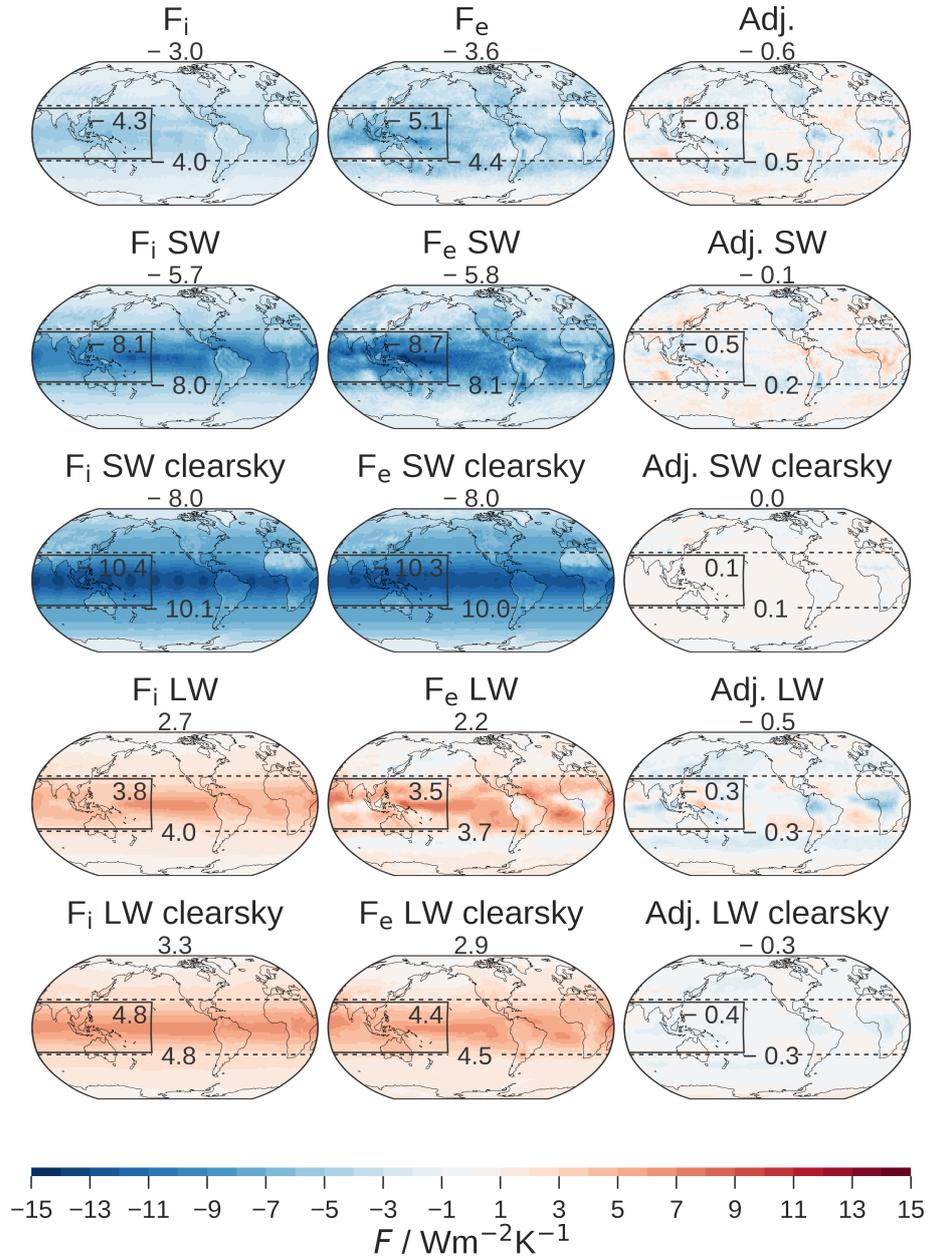
Figure B.10: For Aero, we show instantaneous forcing (first column), effective forcing (second column), and the difference between the two (= adjustments, third column). We further split this into SW and LW contributions, and separate out the clear-sky.

B.6.1 *CMIP6 models*

	Ensemble members
ACCESS-CM2	3
BCC-CSM2-MR	3
BCC-ESM1	3
CESM2-FV2	3
CESM2-WACCM-FV2	3
CESM2-WACCM	3
CIESM	3
CMCC-CM2-SR5	1
CNRM-CM6-1-HR	1
CanESM5	65
FGOALS-f3-L	3
FGOALS-g3	6
GFDL-CM4	1
GFDL-ESM4	3
GISS-E2-1-G	47
GISS-E2-1-H	25
INM-CM4-8	1
INM-CM5-0	10
MIROC6	50
NESM3	5
NorESM2-MM	3
SAM0-UNICON	1
TaiESM1	1

Table B.3: CMIP6 models with *hist* simulations included in the CMIP analysis

AUTHOR CONTRIBUTIONS AND COMPETING INTERESTS

MG and HS conceived this study, building on ideas formulated by HS, CT, and MT. MG performed the simulations, the analysis, and wrote the initial paper draft. All authors discussed the results and revised the paper together.

The authors declare that they have no conflict of interest. Matthew Toohey is a member of the editorial board of Atmospheric Chemistry and Physics.

DATA AND CODE AVAILABILITY

The code and data used for this study will be made available upon publication.

I used ChatGPT (OpenAI, 2021) to generate feedback and suggestions for linguistic and style improvements. ChatGPT was not involved in the scientific research presented in this thesis. I did not copy output from ChatGPT, but integrated some of its suggestions into my own writing.

I would like to give credit to the developers of `CDO` (Schulzweida, 2023), `python` (Python Software Foundation, 2020), and the python-packages `xarray` (Hoyer and Hamman, 2017), `numpy` (Harris et al., 2020), `matplotlib` (Hunter, 2007), and `cartopy` (Met Office, 2010–2015) which are the main analysis tools I used.

The figure on the title page (Fig. 2.2) was created with the help of Yvonne Schrader, MPI-M, based on my idea.

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede and Ivo Pletikosić. `classicthesis` is available for both LaTeX and LyX:

```
https://bitbucket.org/amiede/classicthesis/
```

BIBLIOGRAPHY

Abalos, M., B. Legras, F. Ploeger, and W. J. Randel (2015). "Evaluating the Advective Brewer-Dobson Circulation in Three Reanalyses for the Period 1979–2012." *Journal of Geophysical Research: Atmospheres* 120.15, pp. 7534–7554. DOI: 10.1002/2015JD023182.

Alessi, M. J. and M. A. A. Rugenstein (2023). "Surface Temperature Pattern Scenarios Suggest Higher Warming Rates Than Current Projections." *Geophysical Research Letters* 50.23, e2023GL105795. DOI: 10.1029/2023GL105795.

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." *Climate Dynamics* 24.7, pp. 655–666. DOI: 10.1007/s00382-005-0018-3.

Andrews, T., J. M. Gregory, D. Paynter, L. G. Silvers, C. Zhou, T. Mauritsen, M. J. Webb, K. C. Armour, P. M. Forster, and H. Titchner (2018). "Accounting for Changing Temperature Patterns Increases Historical Estimates of Climate Sensitivity." *Geophysical Research Letters* 45.16, pp. 8490–8499. DOI: 10.1029/2018GL078887.

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." *Journal of Climate* 28.4, pp. 1630–1648. DOI: 10.1175/JCLI-D-14-00545.1.

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." *Geophysical Research Letters* 39.9. DOI: 10.1029/2012GL051607.

Andrews, T. and M. J. Webb (2018). "The Dependence of Global Cloud and Lapse Rate Feedbacks on the Spatial Structure of Tropical Pacific Warming." *Journal of Climate* 31.2, pp. 641–654. DOI: 10.1175/JCLI-D-17-0087.1.

Andrews, T. et al. (2022). "On the Effect of Historical SST Patterns on Radiative Feedback." *Journal of Geophysical Research: Atmospheres* 127.18, e2022JD036675. DOI: 10.1029/2022JD036675.

Aquila, V., L. D. Oman, R. Stolarski, A. R. Douglass, and P. A. Newman (2013). "The Response of Ozone and Nitrogen Dioxide to the Eruption of Mt. Pinatubo at Southern and Northern Midlatitudes." *Journal of the Atmospheric Sciences* 70.3, pp. 894–900. DOI: 10.1175/JAS-D-12-0143.1.

Archer, D. and V. Brovkin (2008). "The Millennial Atmospheric Lifetime of Anthropogenic CO2." *Climatic Change* 90.3, pp. 283–297. DOI: 10.1007/s10584-008-9413-1.

Armour, K. C. (2017). "Energy Budget Constraints on Climate Sensitivity in Light of Inconstant Climate Feedbacks." *Nature Climate Change* 7.5, pp. 331–335. DOI: 10.1038/nclimate3278.

Armour, K. C., C. M. Bitz, and G. H. Roe (2013). "Time-Varying Climate Sensitivity from Regional Feedbacks." *Journal of Climate* 26.13, pp. 4518–4534. DOI: 10.1175/JCLI-D-12-00544.1.

Armour, K. C., J. Marshall, J. R. Scott, A. Donohoe, and E. R. Newsom (2016). "Southern Ocean Warming Delayed by Circumpolar Upwelling and Equatorward Transport." *Nature Geoscience* 9.7 (7), pp. 549–554. DOI: 10.1038/ngeo2731.

Armour, K. C., N. Siler, A. Donohoe, and G. H. Roe (2019). "Meridional Atmospheric Heat Transport Constrained by Energetics and Mediated by Large-Scale Diffusion." *Journal of Climate* 32.12, pp. 3655–3680. DOI: 10.1175/JCLI-D-18-0563.1.

Arrhenius, S. (1896). "On the Influence of Carbonic Acid in the Air upon the Temperature of the Ground." *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 41.251, pp. 237–276. DOI: 10.1080/14786449608620846.

Azoulay, A., H. Schmidt, and C. Timmreck (2021). "The Arctic Polar Vortex Response to Volcanic Forcing of Different Strengths." *Journal of Geophysical Research: Atmospheres* 126.11, e2020JD034450. DOI: 10.1029/2020JD034450.

Bintanja, R., E. C. van der Linden, and W. Hazeleger (2012). "Boundary Layer Stability and Arctic Climate Change: A Feedback Study Using EC-Earth." *Climate Dynamics* 39.11, pp. 2659–2673. DOI: 10.1007/s00382-011-1272-1.

Birner, T. and E. J. Charlesworth (2017). "On the Relative Importance of Radiative and Dynamical Heating for Tropical Tropopause Temperatures." *Journal of Geophysical Research: Atmospheres* 122.13, pp. 6782–6797. DOI: 10.1002/2016JD026445.

Bittner, M., H. Schmidt, C. Timmreck, and F. Sienz (2016). "Using a Large Ensemble of Simulations to Assess the Northern Hemisphere Stratospheric Dynamical Response to Tropical Volcanic Eruptions and Its Uncertainty." *Geophysical Research Letters* 43.17, pp. 9324–9332. DOI: 10.1002/2016GL070587.

Bloch-Johnson, J. et al. (2024). "The Green's Function Model Intercomparison Project (GFMIP) Protocol." *Journal of Advances in Modeling Earth Systems* 16.2, e2023MS003700. DOI: 10.1029/2023MS003700.

Block, K. and T. Mauritsen (2013). "Forcing and Feedback in the MPI-ESM-LR Coupled Model under Abruptly Quadrupled CO2." *Journal of Advances in Modeling Earth Systems* 5.4, pp. 676–691. DOI: 10.1002/jame.20041.

Boer, G. J., M Stowasser, and K Hamilton (2006). "Inferring Climate Sensitivity from Volcanic Events." *Climate Dynamics*, p. 22.

Boltzmann, L. (1884). "Ableitung Des Stefan'schen Gesetzes, Betreffend Die Abhängigkeit Der Wärmestrahlung von Der Temperatur

Aus Der Electromagnetischen Lichttheorie." *Annalen der Physik* 258.6, pp. 291–294. DOI: `10.1002/andp.18842580616`.

Bony, S. et al. (2006). "How Well Do We Understand and Evaluate Climate Change Feedback Processes?" *Journal of Climate* 19.15, pp. 3445–3482. DOI: `10.1175/JCLI3819.1`.

Bretherton, C. S. (2015). "Insights into Low-Latitude Cloud Feedbacks from High-Resolution Models." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 373.2054, p. 20140415. DOI: `10.1098/rsta.2014.0415`.

Bretherton, C. S. and P. K. Smolarkiewicz (1989). "Gravity Waves, Compensating Subsidence and Detrainment around Cumulus Clouds." *Journal of the Atmospheric Sciences* 46.6, pp. 740–759. DOI: `10.1175/1520-0469(1989)046<0740:GWCSAD>2.0.CO;2`.

Butchart, N. (2014). "The Brewer-Dobson Circulation." *Reviews of Geophysics* 52.2, pp. 157–184. DOI: `10.1002/2013RG000448`.

Cai, W. and P. H. Whetton (2001). "A Time-Varying Greenhouse Warming Pattern and the Tropical–Extratropical Circulation Linkage in the Pacific Ocean." *Journal of Climate* 14.16, pp. 3337–3355. DOI: `10.1175/1520-0442(2001)014<3337:ATVGWP>2.0.CO;2`.

Ceppi, P. and J. M. Gregory (2017). "Relationship of Tropospheric Stability to Climate Sensitivity and Earth's Observed Radiation Budget." *Proceedings of the National Academy of Sciences* 114.50, pp. 13126–13131. DOI: `10.1073/pnas.1714308114`.

– (2019). "A Refined Model for the Earth's Global Energy Balance." *Climate Dynamics* 53.7-8, pp. 4781–4797. DOI: `10.1007/s00382-019-04825-x`.

Chalmers, J., J. E. Kay, E. A. Middlemas, E. A. Maroon, and P. DiNezio (2022). "Does Disabling Cloud Radiative Feedbacks Change Spatial Patterns of Surface Greenhouse Warming and Cooling?" *Journal of Climate* 35.6, pp. 1787–1807. DOI: `10.1175/JCLI-D-21-0391.1`.

De Deckker, P. (2016). "The Indo-Pacific Warm Pool: Critical to World Oceanography and World Climate." *Geoscience Letters* 3.1, p. 20. DOI: `10.1186/s40562-016-0054-3`.

DiNezio, P. (2023). "Anthropogenic Forcing of the Tropical Pacific Cooling Trend." ECS and Cloud Feedback Virtual Symposium 26 (https://www.youtube.com/watch?v=Eug7r-3rDqM).

Diallo, M., B. Legras, and A. Chédin (2012). "Age of Stratospheric Air in the ERA-Interim." *Atmospheric Chemistry and Physics* 12.24, pp. 12133–12154. DOI: `10.5194/acp-12-12133-2012`.

Diallo, M. et al. (2017). "Significant Contributions of Volcanic Aerosols to Decadal Changes in the Stratospheric Circulation." *Geophysical Research Letters* 44.20, pp. 10,780–10,791. DOI: `10.1002/2017GL074662`.

Dong, Y., K. C. Armour, M. D. Zelinka, C. Proistosescu, D. S. Battisti, C. Zhou, and T. Andrews (2020). "Intermodel Spread in the Pattern Effect and Its Contribution to Climate Sensitivity in CMIP5 and

CMIP6 Models." *Journal of Climate* 33.18, pp. 7755–7775. DOI: 10.1175/JCLI-D-19-1011.1.

Dong, Y., C. Proistosescu, K. C. Armour, and D. S. Battisti (2019). "Attributing Historical and Future Evolution of Radiative Feedbacks to Regional Warming Patterns Using a Green's Function Approach: The Preeminence of the Western Pacific." *Journal of Climate* 32.17, pp. 5471–5491. DOI: 10.1175/JCLI-D-18-0843.1.

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." *Geoscientific Model Development* 9.5, pp. 1937–1958. DOI: 10.5194/gmd-9-1937-2016.

Fajber, R., A. Donohoe, S. Ragen, K. C. Armour, and P. J. Kushner (2023). "Atmospheric Heat Transport Is Governed by Meridional Gradients in Surface Evaporation in Modern-Day Earth-like Climates." *Proceedings of the National Academy of Sciences* 120.25, e2217202120. DOI: 10.1073/pnas.2217202120.

Fajber, R. and P. J. Kushner (2021). "Using "Heat Tagging" to Understand the Remote Influence of Atmospheric Diabatic Heating through Long-Range Transport." *Journal of the Atmospheric Sciences* 78.7, pp. 2161–2176. DOI: 10.1175/JAS-D-20-0290.1.

Ferraro, A. J., E. J. Highwood, and A. J. Charlton-Perez (2014). "Weakened Tropical Circulation and Reduced Precipitation in Response to Geoengineering." *Environmental Research Letters* 9.1, p. 014001. DOI: 10.1088/1748-9326/9/1/014001.

Forster, P. M. F., R. S. Freckleton, and K. P. Shine (1997). "On Aspects of the Concept of Radiative Forcing." *Climate Dynamics* 13.7, pp. 547–560. DOI: 10.1007/s003820050182.

Forster, P. M., M. Blackburn, R. Glover, and K. P. Shine (2000). "An Examination of Climate Sensitivity for Idealised Climate Change Experiments in an Intermediate General Circulation Model." *Climate Dynamics* 16.10, pp. 833–849. DOI: 10.1007/s003820000083.

Forster, P. M., T. Richardson, A. C. Maycock, C. J. Smith, B. H. Samset, G. Myhre, T. Andrews, R. Pincus, and M. Schulz (2016). "Recommendations for Diagnosing Effective Radiative Forcing from Climate Models for CMIP6." *Journal of Geophysical Research: Atmospheres* 121.20, pp. 12,460–12,475. DOI: 10.1002/2016JD025320.

Forster, P. M. F. and K. E. Taylor (2006). "Climate Forcings and Climate Sensitivities Diagnosed from Coupled Climate Model Integrations." *Journal of Climate* 19.23, pp. 6181–6194. DOI: 10.1175/JCLI3974.1.

Garcia, R. R. and W. J. Randel (2008). "Acceleration of the Brewer–Dobson Circulation Due to Increases in Greenhouse Gases." *Journal of the Atmospheric Sciences* 65.8, pp. 2731–2739. DOI: 10.1175/2008JAS2712.1.

Garcia, R. R., W. J. Randel, and D. E. Kinnison (2011). "On the Determination of Age of Air Trends from Atmospheric Trace Species."

Journal of the Atmospheric Sciences 68.1, pp. 139–154. DOI: `10.1175/2010JAS3527.1`.

Garfinkel, C. I., V. Aquila, D. W. Waugh, and L. D. Oman (2017). "Time-Varying Changes in the Simulated Structure of the Brewer–Dobson Circulation." *Atmospheric Chemistry and Physics* 17.2, pp. 1313–1327. DOI: `10.5194/acp-17-1313-2017`.

Geoffroy, O., D. Saint-Martin, G. Bellon, A. Voldoire, D. J. L. Olivié, and S. Tytéca (2013). "Transient Climate Response in a Two-Layer Energy-Balance Model. Part II: Representation of the Efficacy of Deep-Ocean Heat Uptake and Validation for CMIP5 AOGCMs." *Journal of Climate* 26.6, pp. 1859–1876. DOI: `10.1175/JCLI-D-12-00196.1`.

Gettelman, A. and S. C. Sherwood (2016). "Processes Responsible for Cloud Feedback." *Current Climate Change Reports* 2.4, pp. 179–189. DOI: `10.1007/s40641-016-0052-8`.

Graf, H.-F., Q. Li, and M. A. Giorgetta (2007). "Volcanic Effects on Climate: Revisiting the Mechanisms." *Atmospheric Chemistry and Physics* 7.17, pp. 4503–4511. DOI: `10.5194/acp-7-4503-2007`.

Gregory, J. M. (2004). "A New Method for Diagnosing Radiative Forcing and Climate Sensitivity." *Geophysical Research Letters* 31.3, p. L03205. DOI: `10.1029/2003GL018747`.

Gregory, J. M. and T. Andrews (2016). "Variation in Climate Sensitivity and Feedback Parameters during the Historical Period." *Geophysical Research Letters* 43.8, pp. 3911–3920. DOI: `10.1002/2016GL068406`.

Gregory, J. M., T. Andrews, P. Ceppi, T. Mauritsen, and M. J. Webb (2020). "How Accurately Can the Climate Sensitivity to $$\textbackslash Hbox {\vphantom}CO\vphantom{}_{2}$$ Be Estimated from Historical Climate Change?" *Climate Dynamics* 54.1-2, pp. 129–157. DOI: `10.1007/s00382-019-04991-y`.

Gregory, J. M., T. Andrews, P. Good, T. Mauritsen, and P. M. Forster (2016). "Small Global-Mean Cooling Due to Volcanic Radiative Forcing." *Climate Dynamics* 47.12, pp. 3979–3991. DOI: `10.1007/s00382-016-3055-1`.

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." *Journal of Climate* 15.22, pp. 3117–3121. DOI: `10.1175/1520-0442(2002)015<3117:AOBEOT>2.0.CO;2`.

Günther, M., H. Schmidt, C. Timmreck, and M. Toohey (2022). "Climate Feedback to Stratospheric Aerosol Forcing: The Key Role of the Pattern Effect." *Journal of Climate* 35.24, pp. 4303–4317. DOI: `10.1175/JCLI-D-22-0306.1`.

– (2024). "Why Does Stratospheric Aerosol Forcing Strongly Cool the Warm Pool?" *under review with minor revisions in Atmospheric Chemistry and Physics*. DOI: `10.5194/egusphere-2024-429`.

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." *Frontiers in Earth Science* 9, p. 725. DOI: `10.3389/feart.2021.710036`.

Hansen, J., M. Sato, and R. Ruedy (1997). "Radiative Forcing and Climate Response." *Journal of Geophysical Research: Atmospheres* 102.D6, pp. 6831–6864. DOI: `10.1029/96JD03436`.

Hansen, J. et al. (2005). "Efficacy of Climate Forcings." *Journal of Geophysical Research* 110.D18, p. D18104. DOI: `10.1029/2005JD005776`.

Harris, C. R. et al. (2020). "Array Programming with NumPy." *Nature* 585.7825, pp. 357–362. DOI: `10.1038/s41586-020-2649-2`.

Haugstad, A. D., K. C. Armour, D. S. Battisti, and B. E. J. Rose (2017). "Relative Roles of Surface Temperature and Climate Forcing Patterns in the Inconstancy of Radiative Feedbacks." *Geophysical Research Letters* 44.14, pp. 7455–7463. DOI: `10.1002/2017GL074372`.

Heede, U. K., A. V. Fedorov, and N. J. Burls (2020). "Time Scales and Mechanisms for the Tropical Pacific Response to Global Warming: A Tug of War between the Ocean Thermostat and Weaker Walker." *Journal of Climate* 33.14, pp. 6101–6118. DOI: `10.1175/JCLI-D-19-0690.1`.

Held, I. M. and B. J. Soden (2000). "Water Vapor Feedback and Global Warming." *Annual Review of Energy and the Environment* 25.1, pp. 441–475. DOI: `10.1146/annurev.energy.25.1.441`.

Held, I. M., M. Winton, K. Takahashi, T. Delworth, F. Zeng, and G. K. Vallis (2010). "Probing the Fast and Slow Components of Global Warming by Returning Abruptly to Preindustrial Forcing." *Journal of Climate* 23.9, pp. 2418–2427. DOI: `10.1175/2009JCLI3466.1`.

Holloway, C. E., S. J. Woolnough, and G. M. S. Lister (2012). "Precipitation Distributions for Explicit versus Parametrized Convection in a Large-Domain High-Resolution Tropical Case Study." *Quarterly Journal of the Royal Meteorological Society* 138.668, pp. 1692–1708. DOI: `10.1002/qj.1903`.

Holton, J. R., P. H. Haynes, M. E. McIntyre, A. R. Douglass, R. B. Rood, and L. Pfister (1995). "Stratosphere-Troposphere Exchange." *Reviews of Geophysics* 33.4, pp. 403–439. DOI: `10.1029/95RG02097`.

Hoyer, S. and J. Hamman (2017). "Xarray: N-D Labeled Arrays and Datasets in Python." 5.1 (1), p. 10. DOI: `10.5334/jors.148`.

Hu, S., S.-P. Xie, and S. M. Kang (2022). "Global Warming Pattern Formation: The Role of Ocean Heat Uptake." *Journal of Climate* 35.6, pp. 1885–1899. DOI: `10.1175/JCLI-D-21-0317.1`.

Hunter, J. D. (2007). "Matplotlib: A 2D Graphics Environment." *Computing in Science & Engineering* 9.3, pp. 90–95. DOI: `10.1109/MCSE.2007.55`.

Hwang, Y.-T., S.-P. Xie, C. Deser, and S. M. Kang (2017). "Connecting Tropical Climate Change with Southern Ocean Heat Uptake." *Geophysical Research Letters* 44.18, pp. 9449–9457. DOI: `10.1002/2017GL074972`.

IPCC (2001). *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*. In collab. with R. N. Cooper, J. T. Houghton, J. J. McCarthy, and B. Metz. Vol. 81. DOI: `10.2307/20033020`.

– (2022). *Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. 1st ed. Cambridge University Press. DOI: `10.1017/9781009157896`.

Ilyina, T., K. D. Six, J. Segschneider, E. Maier-Reimer, H. Li, and I. Núñez-Riboni (2013). "Global Ocean Biogeochemistry Model HAMOCC: Model Architecture and Performance as Component of the MPI-Earth System Model in Different CMIP5 Experimental Realizations." *Journal of Advances in Modeling Earth Systems* 5.2, pp. 287–315. DOI: `10.1029/2012MS000178`.

Inman, M. (2008). "Carbon Is Forever." *Nature Climate Change* 1.812, pp. 156–158. DOI: `10.1038/climate.2008.122`.

Jeevanjee, N., J. T. Seeley, D. Paynter, and S. Fueglistaler (2021). "An Analytical Model for Spatially Varying Clear-Sky CO2 Forcing." *Journal of Climate* 34.23, pp. 9463–9480. DOI: `10.1175/JCLI-D-19-0756.1`.

Joshi, M., K. Shine, M. Ponater, N. Stuber, R. Sausen, and L. Li (2003). "A Comparison of Climate Response to Different Radiative Forcings in Three General Circulation Models: Towards an Improved Metric of Climate Change." *Climate Dynamics* 20.7-8, pp. 843–854. DOI: `10.1007/s00382-003-0305-9`.

Joshi, M. M. and K. P. Shine (2003). "A GCM Study of Volcanic Eruptions as a Cause of Increased Stratospheric Water Vapor." *Journal of Climate* 16.21, pp. 3525–3534. DOI: `10.1175/1520-0442(2003)016<3525:AGSOVE>2.0.CO;2`.

Jungclaus, J. H., N. Fischer, H. Haak, K. Lohmann, J. Marotzke, D. Matei, U. Mikolajewicz, D. Notz, and J. S. von Storch (2013). "Characteristics of the Ocean Simulations in the Max Planck Institute Ocean Model (MPIOM) the Ocean Component of the MPI-Earth System Model." *Journal of Advances in Modeling Earth Systems* 5.2, pp. 422–446. DOI: `10.1002/jame.20023`.

Kang, S. M., P. Ceppi, Y. Yu, and I.-S. Kang (2023). "Recent Global Climate Feedback Controlled by Southern Ocean Cooling." *Nature Geoscience* 16.9 (9), pp. 775–780. DOI: `10.1038/s41561-023-01256-6`.

Kang, S. M. and S.-P. Xie (2014). "Dependence of Climate Response on Meridional Structure of External Thermal Forcing." *Journal of Climate* 27.14, pp. 5593–5600. DOI: `10.1175/JCLI-D-13-00622.1`.

Karoly, D., J. Cohen, G. Meehl, J. Mitchell, A. Oort, R. Stouffer, and R. Wetherald (1994). "An Example of Fingerprint Detection of Greenhouse Climate Change." *Climate Dynamics* 10.1, pp. 97–105. DOI: `10.1007/BF00210339`.

Kaur, H., G. Bala, and A. K. Seshadri (2023). "Why Is Climate Sensitivity for Solar Forcing Smaller than for an Equivalent CO2 Forcing?" *Journal of Climate* 36.3, pp. 775–789. DOI: `10.1175/JCLI-D-21-0980.1`.

Kiehl, J. T. and K. E. Trenberth (1997). "Earth's Annual Global Mean Energy Budget." *Bulletin of the American Meteorological Society* 78.2, pp. 197–208. DOI: `10.1175/1520-0477(1997)078<0197:EAGMEB>2.0.CO;2`.

Kravitz, B. et al. (2019). "Comparing Surface and Stratospheric Impacts of Geoengineering With Different SO2 Injection Strategies." *Journal of Geophysical Research: Atmospheres* 124.14, pp. 7900–7918. DOI: `10.1029/2019JD030329`.

Kroll, C. A., S. Dacie, A. Azoulay, H. Schmidt, and C. Timmreck (2021). "The Impact of Volcanic Eruptions of Different Magnitude on Stratospheric Water Vapor in the Tropics." *Atmospheric Chemistry and Physics* 21.8, pp. 6565–6591. DOI: `10.5194/acp-21-6565-2021`.

Kummer, J. R. and A. E. Dessler (2014). "The Impact of Forcing Efficacy on the Equilibrium Climate Sensitivity." *Geophysical Research Letters* 41.10, pp. 3565–3568. DOI: `10.1002/2014GL060046`.

Laakso, A., H. Korhonen, S. Romakkaniemi, and H. Kokkola (2017). "Radiative and Climate Effects of Stratospheric Sulfur Geoengineering Using Seasonally Varying Injection Areas." *Atmospheric Chemistry and Physics* 17.11, pp. 6957–6974. DOI: `10.5194/acp-17-6957-2017`.

Lee, S., M. L'Heureux, A. T. Wittenberg, R. Seager, P. A. O'Gorman, and N. C. Johnson (2022). "On the Future Zonal Contrasts of Equatorial Pacific Climate: Perspectives from Observations, Simulations, and Theories." *npj Climate and Atmospheric Science* 5.1 (1), pp. 1–15. DOI: `10.1038/s41612-022-00301-2`.

Lee, W. R., D. Visioni, E. M. Bednarz, D. G. MacMartin, B. Kravitz, and S. Tilmes (2023). "Quantifying the Efficiency of Stratospheric Aerosol Geoengineering at Different Altitudes." *Geophysical Research Letters* 50.14, e2023GL104417. DOI: `10.1029/2023GL104417`.

Li, C., J.-S. von Storch, and J. Marotzke (2013). "Deep-Ocean Heat Uptake and Equilibrium Climate Response." *Climate Dynamics* 40.5, pp. 1071–1086. DOI: `10.1007/s00382-012-1350-z`.

Lin, Y.-J., Y.-T. Hwang, P. Ceppi, and J. M. Gregory (2019). "Uncertainty in the Evolution of Climate Feedback Traced to the Strength of the Atlantic Meridional Overturning Circulation." *Geophysical Research Letters* 46.21, 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." *Geophysical Research Letters* 48.9, e2021GL093302. DOI: `10.1029/2021GL093302`.

Liu, F., J. Lu, O. A. Garuba, Y. Huang, L. R. Leung, B. E. Harrop, and Y. Luo (2018a). "Sensitivity of Surface Temperature to Oceanic

Forcing via Q-Flux Green's Function Experiments. Part II: Feedback Decomposition and Polar Amplification." *Journal of Climate* 31.17, pp. 6745–6761. DOI: `10.1175/JCLI-D-18-0042.1`.

Liu, F., J. Lu, O. Garuba, L. R. Leung, Y. Luo, and X. Wan (2018b). "Sensitivity of Surface Temperature to Oceanic Forcing via Q-Flux Green's Function Experiments. Part I: Linear Response Function." *Journal of Climate* 31.9, pp. 3625–3641. DOI: `10.1175/JCLI-D-17-0462.1`.

Liu, F., J. Lu, and L. R. Leung (2022). "Neutral Mode Dominates the Forced Global and Regional Surface Temperature Response in the Past and Future." *Geophysical Research Letters* 49.15, e2022GL098788. DOI: `10.1029/2022GL098788`.

Maher, N. et al. (2019). "The Max Planck Institute Grand Ensemble: Enabling the Exploration of Climate System Variability." *Journal of Advances in Modeling Earth Systems* 11.7, pp. 2050–2069. DOI: `10.1029/2019MS001639`.

Manabe, S. and R. T. Wetherald (1975). "The Effects of Doubling the CO2 Concentration on the Climate of a General Circulation Model." *Journal of the Atmospheric Sciences* 32.1, pp. 3–15. DOI: `10.1175/1520-0469(1975)032<0003:TEODTC>2.0.CO;2`.

Marshall, L. R., C. J. Smith, P. M. Forster, T. J. Aubry, T. Andrews, and A. Schmidt (2020). "Large Variations in Volcanic Aerosol Forcing Efficiency Due to Eruption Source Parameters and Rapid Adjustments." *Geophysical Research Letters* 47.19. DOI: `10.1029/2020GL090241`.

Marvel, K., G. A. Schmidt, R. L. Miller, and L. S. Nazarenko (2016). "Implications for Climate Sensitivity from the Response to Individual Forcings." *Nature Climate Change* 6.4 (4), pp. 386–389. DOI: `10.1038/nclimate2888`.

Mauritsen, T. et al. (2019). "Developments in the MPI-M Earth System Model Version 1.2 (MPI-ESM1.2) and Its Response to Increasing CO2." *Journal of Advances in Modeling Earth Systems* 11.4, pp. 998–1038. DOI: `10.1029/2018MS001400`.

McMonigal, K., S. Larson, S. Hu, and R. Kramer (2023). "Historical Changes in Wind-Driven Ocean Circulation Can Accelerate Global Warming." *Geophysical Research Letters* 50.4, e2023GL102846. DOI: `10.1029/2023GL102846`.

Merlis, T. M., I. M. Held, G. L. Stenchikov, F. Zeng, and L. W. Horowitz (2014). "Constraining Transient Climate Sensitivity Using Coupled Climate Model Simulations of Volcanic Eruptions." *Journal of Climate* 27.20, pp. 7781–7795. DOI: `10.1175/JCLI-D-14-00214.1`.

Met Office (2010–2015). *Cartopy: A Cartographic Python Library with a Matplotlib Interface*.

Mitevski, I., L. M. Polvani, and C. Orbe (2022). "Asymmetric Warming/Cooling Response to CO2 Increase/Decrease Mainly Due To Non-Logarithmic Forcing, Not Feedbacks." *Geophysical Research Letters* 49.5, e2021GL097133. DOI: `10.1029/2021GL097133`.

Modak, A., G. Bala, L. Cao, and K. Caldeira (2016). "Why Must a Solar Forcing Be Larger than a CO2 Forcing to Cause the Same Global Mean Surface Temperature Change?" *Environmental Research Letters* 11.4, p. 044013. DOI: 10.1088/1748-9326/11/4/044013.

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." *Journal of Climate* 8.3, pp. 496–514. DOI: 10.1175/1520-0442(1995)008<0496:TROTHC>2.0.CO;2.

Muthers, S., A. Kuchar, A. Stenke, J. Schmitt, J. G. Anet, C. C. Raible, and T. F. Stocker (2016). "Stratospheric Age of Air Variations between 1600 and 2100." *Geophysical Research Letters* 43.10, pp. 5409–5418. DOI: 10.1002/2016GL068734.

Myhre, G., E. J. Highwood, K. P. Shine, and F. Stordal (1998). "New Estimates of Radiative Forcing Due to Well Mixed Greenhouse Gases." *Geophysical Research Letters* 25.14, pp. 2715–2718. DOI: 10.1029/98GL01908.

OpenAI (2021). *ChatGPT: A Large-Scale Generative Model for Open-Domain Chat*. Version 3.5.

Pauling, A. G., C. M. Bitz, and K. C. Armour (2023). "The Climate Response to the Mt. Pinatubo Eruption Does Not Constrain Climate Sensitivity." *Geophysical Research Letters* 50.7, e2023GL102946. DOI: 10.1029/2023GL102946.

Pauling, A. G., M. Bushuk, and C. M. Bitz (2021). "Robust Inter-Hemispheric Asymmetry in the Response to Symmetric Volcanic Forcing in Model Large Ensembles." *Geophysical Research Letters* 48.9, e2021GL092558. DOI: 10.1029/2021GL092558.

Pitari, G. and E. Mancini (2002). "Short-Term Climatic Impact of the 1991 Volcanic Eruption of Mt. Pinatubo and Effects on Atmospheric Tracers." *Natural Hazards and Earth System Sciences* 2.1/2, pp. 91–108. DOI: 10.5194/nhess-2-91-2002.

Pitari, G. and V. Rizi (1993). "An Estimate of the Chemical and Radiative Perturbation of Stratospheric Ozone Following the Eruption of Mt. Pinatubo." *Journal of the Atmospheric Sciences* 50.19, pp. 3260–3276. DOI: 10.1175/1520-0469(1993)050<3260:AEOTCA>2.0.CO;2.

Pithan, F. and T. Mauritsen (2014). "Arctic Amplification Dominated by Temperature Feedbacks in Contemporary Climate Models." *Nature Geoscience* 7.3 (3), pp. 181–184. DOI: 10.1038/ngeo2071.

Po-Chedley, S., K. C. Armour, C. M. Bitz, M. D. Zelinka, B. D. Santer, and Q. Fu (2018). "Sources of Intermodel Spread in the Lapse Rate and Water Vapor Feedbacks." *Journal of Climate* 31.8, pp. 3187–3206. DOI: 10.1175/JCLI-D-17-0674.1.

Previdi, M., K. L. Smith, and L. M. Polvani (2021). "Arctic Amplification of Climate Change: A Review of Underlying Mechanisms." *Environmental Research Letters* 16.9, p. 093003. DOI: 10.1088/1748-9326/ac1c29.

Python Software Foundation (2020). *Python*. Version 3.9.

Rasch, P. J., S. Tilmes, R. P. Turco, A. Robock, L. Oman, C.-C. J. Chen, G. L. Stenchikov, and R. R. Garcia (2008). "An Overview of Geoengineering of Climate Using Stratospheric Sulphate Aerosols." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 366.1882, pp. 4007–4037. DOI: `10.1098/rsta.2008.0131`.

Reick, C. H., V. Gayler, D. Goll, S. Hagemann, M. Heidkamp, J. E. M. S. Nabel, T. Raddatz, E. Roeckner, R. Schnur, and S. Wilkenskjeld (2021). "JSBACH 3 - The Land Component of the MPI Earth System Model: Documentation of Version 3.2." DOI: `10.17617/2.3279802`.

Richter, J. H., S. Tilmes, M. J. Mills, J. J. Tribbia, B. Kravitz, D. G. MacMartin, F. Vitt, and J.-F. Lamarque (2017). "Stratospheric Dynamical Response and Ozone Feedbacks in the Presence of SO2 Injections." *Journal of Geophysical Research: Atmospheres* 122.23, pp. 12,557–12,573. DOI: `10.1002/2017JD026912`.

Robock, A. (2000). "Volcanic Eruptions and Climate." *Reviews of Geophysics* 38.2, pp. 191–219. DOI: `10.1029/1998RG000054`.

Rohrschneider, T., B. Stevens, and T. Mauritsen (2019). "On Simple Representations of the Climate Response to External Radiative Forcing." *Climate Dynamics* 53.5-6, pp. 3131–3145. DOI: `10.1007/s00382-019-04686-4`.

Rose, B. E. J., K. C. Armour, D. S. Battisti, N. Feldl, and D. D. B. Koll (2014). "The Dependence of Transient Climate Sensitivity and Radiative Feedbacks on the Spatial Pattern of Ocean Heat Uptake." *Geophysical Research Letters* 41.3, pp. 1071–1078. DOI: `10.1002/2013GL058955`.

Rose, B. E. J. and L. Rayborn (2016). "The Effects of Ocean Heat Uptake on Transient Climate Sensitivity." *Current Climate Change Reports* 2.4, pp. 190–201. DOI: `10.1007/s40641-016-0048-4`.

Rugenstein, M. A. A. and K. C. Armour (2021). "Three Flavors of Radiative Feedbacks and Their Implications for Estimating Equilibrium Climate Sensitivity." *Geophysical Research Letters* 48.15, e2021GL092983. DOI: `10.1029/2021GL092983`.

Rugenstein, M. A. A., K. Caldeira, and R. Knutti (2016a). "Dependence of Global Radiative Feedbacks on Evolving Patterns of Surface Heat Fluxes." *Geophysical Research Letters* 43.18, pp. 9877–9885. DOI: `10.1002/2016GL070907`.

Rugenstein, M. A. A., J. Sedláček, and R. Knutti (2016b). "Nonlinearities in Patterns of Long-Term Ocean Warming." *Geophysical Research Letters* 43.7, pp. 3380–3388. DOI: `10.1002/2016GL068041`.

Rugenstein, M., S. Dhame, D. Olonscheck, R. J. Wills, M. Watanabe, and R. Seager (2023a). "Connecting the SST Pattern Problem and the Hot Model Problem." *Geophysical Research Letters* 50.22, e2023GL105488. DOI: `10.1029/2023GL105488`.

Rugenstein, M., M. Zelinka, K. B. Karnauskas, P. Ceppi, and T. Andrews (2023b). "Patterns of Surface Warming Matter for Climate Sensitivity." *Eos* 104. DOI: 10.1029/2023EO230411.

Rugenstein, M. et al. (2020). "Equilibrium Climate Sensitivity Estimated by Equilibrating Climate Models." *Geophysical Research Letters* 47.4, e2019GL083898. DOI: 10.1029/2019GL083898.

SPARC (2022). *SPARC Reanalysis Intercomparison Project (S-RIP) Final Report.* SPARC Report 10. WCRP-6/2021.

Salvi, P., P. Ceppi, and J. M. Gregory (2022). "Interpreting Differences in Radiative Feedbacks From Aerosols Versus Greenhouse Gases." *Geophysical Research Letters* 49.8, e2022GL097766. DOI: 10.1029/2022GL097766.

Salvi, P., J. M. Gregory, and P. Ceppi (2023). "Time-Evolving Radiative Feedbacks in the Historical Period." *Journal of Geophysical Research: Atmospheres* 128.20, e2023JD038984. DOI: 10.1029/2023JD038984.

Schmidt, H. et al. (2012). "Solar Irradiance Reduction to Counteract Radiative Forcing from a Quadrupling of CO\textsubscript2: Climate Responses Simulated by Four Earth System Models." *Earth System Dynamics* 3.1, pp. 63–78. DOI: 10.5194/esd-3-63-2012.

Schnadt Poberaj, C., J. Staehelin, and D. Brunner (2011). "Missing Stratospheric Ozone Decrease at Southern Hemisphere Middle Latitudes after Mt. Pinatubo: A Dynamical Perspective." *Journal of the Atmospheric Sciences* 68.9, pp. 1922–1945. DOI: 10.1175/JAS-D-10-05004.1.

Schulzweida, U. (2023). "CDO User Guide." DOI: 10.5281/zenodo.10020800.

Seager, R., M. Cane, N. Henderson, D.-E. Lee, R. Abernathey, and H. Zhang (2019). "Strengthening Tropical Pacific Zonal Sea Surface Temperature Gradient Consistent with Rising Greenhouse Gases." *Nature Climate Change* 9.7 (7), pp. 517–522. DOI: 10.1038/s41558-019-0505-x.

Senior, C. A. and J. F. B. Mitchell (2000). "The Time-Dependence of Climate Sensitivity." *Geophysical Research Letters* 27.17, pp. 2685–2688. DOI: 10.1029/2000GL011373.

Seviour, W. J. M., N. Butchart, and S. C. Hardiman (2012). "The Brewer–Dobson Circulation Inferred from ERA-Interim." *Quarterly Journal of the Royal Meteorological Society* 138.665, pp. 878–888. DOI: 10.1002/qj.966.

Sherwood, S. C. et al. (2020). "An Assessment of Earth's Climate Sensitivity Using Multiple Lines of Evidence." *Reviews of Geophysics* 58.4, e2019RG000678. DOI: 10.1029/2019RG000678.

Sherwood, S. C., S. Bony, O. Boucher, C. Bretherton, P. M. Forster, J. M. Gregory, and B. Stevens (2015). "Adjustments in the Forcing-Feedback Framework for Understanding Climate Change." *Bulletin of the American Meteorological Society* 96.2, pp. 217–228. DOI: 10.1175/BAMS-D-13-00167.1.

Shindell, D. T., G. Faluvegi, L. Rotstayn, and G. Milly (2015). "Spatial Patterns of Radiative Forcing and Surface Temperature Response." *Journal of Geophysical Research: Atmospheres* 120.11, pp. 5385–5403. DOI: 10.1002/2014JD022752.

Simpson, I. R., S. Tilmes, J. H. Richter, B. Kravitz, D. G. MacMartin, M. J. Mills, J. T. Fasullo, and A. G. Pendergrass (2019). "The Regional Hydroclimate Response to Stratospheric Sulfate Geoengineering and the Role of Stratospheric Heating." *Journal of Geophysical Research: Atmospheres* 124.23, pp. 12587–12616. DOI: 10.1029/2019JD031093.

Singh, H., N. Feldl, J. E. Kay, and A. L. Morrison (2022). "Climate Sensitivity Is Sensitive to Changes in Ocean Heat Transport." *Journal of Climate* -1 (aop), pp. 1–60. DOI: 10.1175/JCLI-D-21-0674.1.

Soden, B. J. and I. M. Held (2006). "An Assessment of Climate Feedbacks in Coupled Ocean–Atmosphere Models." *Journal of Climate* 19.14, pp. 3354–3360. DOI: 10.1175/JCLI3799.1.

Soden, B. J., I. M. Held, R. Colman, K. M. Shell, J. T. Kiehl, and C. A. Shields (2008). "Quantifying Climate Feedbacks Using Radiative Kernels." *Journal of Climate* 21.14, pp. 3504–3520. DOI: 10.1175/2007JCLI2110.1.

Sohn, B.-J., S. Lee, E.-S. Chung, and H.-J. Song (2016). "The Role of the Dry Static Stability for the Recent Change in the Pacific Walker Circulation." *Journal of Climate* 29.8, pp. 2765–2779. DOI: 10.1175/JCLI-D-15-0374.1.

Stefan, J. (1879). "Über Die Beziehung Zwischen Der Warmestrahlung Und Der Temperatur." *Sitzungsberichte der Mathematisch-Naturwissenschaftlichen Classe der Kaiserlichen Akademie der Wissenschaften* LXXIX, pp. 391–428.

Stevens, B., S. C. Sherwood, S. Bony, and M. J. Webb (2016). "Prospects for Narrowing Bounds on Earth's Equilibrium Climate Sensitivity." *Earth's Future* 4.11, pp. 512–522. DOI: 10.1002/2016EF000376.

Stevens, B. et al. (2013). "Atmospheric Component of the MPI-M Earth System Model: ECHAM6." *Journal of Advances in Modeling Earth Systems* 5.2, pp. 146–172. DOI: 10.1002/jame.20015.

Stowasser, M., K. Hamilton, and G. J. Boer (2006). "Local and Global Climate Feedbacks in Models with Differing Climate Sensitivities." *Journal of Climate* 19.2, pp. 193–209. DOI: 10.1175/JCLI3613.1.

Sukhodolov, T., S. Vattioni, F. Stefanetti, I. Schuring, J. Sedlacek, and G. Chiodo (2024). "Solid Particle SAI with a Fully Coupled Atmosphere-Ocean-Aerosol-Chemistry-Climate Model SOCOLv4.0." EGU24. Copernicus Meetings. DOI: 10.5194/egusphere-egu24-18905.

Thompson, D. W. J., J. M. Wallace, P. D. Jones, and J. J. Kennedy (2009). "Identifying Signatures of Natural Climate Variability in Time Series of Global-Mean Surface Temperature: Methodology and Insights." *Journal of Climate* 22.22, pp. 6120–6141. DOI: 10.1175/2009JCLI3089.1.

Timmreck, C. (2018). "Climatic Effects of Large Volcanic Eruptions." Universität Hamburg Hamburg. DOI: 10.17617/2.2566000.

Timmreck, C. et al. (2018). "The Interactive Stratospheric Aerosol Model Intercomparison Project (ISA-MIP): Motivation and Experimental Design." *Geoscientific Model Development* 11.7, pp. 2581–2608. DOI: 10.5194/gmd-11-2581-2018.

Toohey, M., K. Krüger, M. Bittner, C. Timmreck, and H. Schmidt (2014). "The Impact of Volcanic Aerosol on the Northern Hemisphere Stratospheric Polar Vortex: Mechanisms and Sensitivity to Forcing Structure." *Atmospheric Chemistry and Physics* 14.23, pp. 13063–13079. DOI: 10.5194/acp-14-13063-2014.

Toohey, M., B. Stevens, H. Schmidt, and C. Timmreck (2016). "Easy Volcanic Aerosol (EVA v1.0): An Idealized Forcing Generator for Climate Simulations." *Geoscientific Model Development* 9.11, pp. 4049–4070. DOI: 10.5194/gmd-9-4049-2016.

Visioni, D., D. G. MacMartin, and B. Kravitz (2021). "Is Turning Down the Sun a Good Proxy for Stratospheric Sulfate Geoengineering?" *Journal of Geophysical Research: Atmospheres* 126.5, e2020JD033952. DOI: 10.1029/2020JD033952.

Webb, M. J. et al. (2017). "The Cloud Feedback Model Intercomparison Project (CFMIP) Contribution to CMIP6." *Geoscientific Model Development* 10.1, pp. 359–384. DOI: 10.5194/gmd-10-359-2017.

Wigley, T. M. L., C. M. Ammann, B. D. Santer, and S. C. B. Raper (2005). "Effect of Climate Sensitivity on the Response to Volcanic Forcing." *Journal of Geophysical Research: Atmospheres* 110.D9. DOI: 10.1029/2004JD005557.

Wild, M., D. Folini, M. Z. Hakuba, C. Schär, S. I. Seneviratne, S. Kato, D. Rutan, C. Ammann, E. F. Wood, and G. König-Langlo (2015). "The Energy Balance over Land and Oceans: An Assessment Based on Direct Observations and CMIP5 Climate Models." *Climate Dynamics* 44.11, pp. 3393–3429. DOI: 10.1007/s00382-014-2430-z.

Williams, A. I. L., N. Jeevanjee, and J. Bloch-Johnson (2023). "Circus Tents, Convective Thresholds, and the Non-Linear Climate Response to Tropical SSTs." *Geophysical Research Letters* 50.6, e2022GL101499. DOI: 10.1029/2022GL101499.

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." *Geophysical Research Letters* 49.17, e2022GL100011. DOI: 10.1029/2022GL100011.

Winton, M., K. Takahashi, and I. M. Held (2010). "Importance of Ocean Heat Uptake Efficacy to Transient Climate Change." *Journal of Climate* 23.9, pp. 2333–2344. DOI: 10.1175/2009JCLI3139.1.

Wood, R. and C. S. Bretherton (2006). "On the Relationship between Stratiform Low Cloud Cover and Lower-Tropospheric Stability." *Journal of Climate* 19.24, pp. 6425–6432. DOI: 10.1175/JCLI3988.1.

Wunderlin, E., G. Chiodo, T. Sukhodolov, S. Vattioni, D. Visioni, and S. Tilmes (2024). "Side Effects of Sulfur-Based Geoengineering Due To Absorptivity of Sulfate Aerosols." *Geophysical Research Letters* 51.4, e2023GL107285. DOI: 10.1029/2023GL107285.

Xia, Y. and Y. Huang (2017). "Differential Radiative Heating Drives Tropical Atmospheric Circulation Weakening." *Geophysical Research Letters* 44.20, pp. 10,592–10,600. DOI: 10.1002/2017GL075678.

Yang, W., G. A. Vecchi, S. Fueglistaler, L. W. Horowitz, D. J. Luet, G. Muñoz, D. Paynter, and S. Underwood (2019). "Climate Impacts From Large Volcanic Eruptions in a High-Resolution Climate Model: The Importance of Forcing Structure." *Geophysical Research Letters* 46.13, pp. 7690–7699. DOI: 10.1029/2019GL082367.

Zhang, L., W. Han, K. B. Karnauskas, G. A. Meehl, A. Hu, N. Rosenbloom, and T. Shinoda (2019). "Indian Ocean Warming Trend Reduces Pacific Warming Response to Anthropogenic Greenhouse Gases: An Interbasin Thermostat Mechanism." *Geophysical Research Letters* 46.19, pp. 10882–10890. DOI: 10.1029/2019GL084088.

Zhang, R., S. M. Kang, and I. M. Held (2010). "Sensitivity of Climate Change Induced by the Weakening of the Atlantic Meridional Overturning Circulation to Cloud Feedback." *Journal of Climate* 23.2, pp. 378–389. DOI: 10.1175/2009JCLI3118.1.

Zhao, M., L. Cao, G. Bala, and L. Duan (2021). "Climate Response to Latitudinal and Altitudinal Distribution of Stratospheric Sulfate Aerosols." *Journal of Geophysical Research: Atmospheres* 126.24, e2021JD035379. DOI: 10.1029/2021JD035379.

Zhou, C., M. Wang, M. D. Zelinka, Y. Liu, Y. Dong, and K. C. Armour (2023). "Explaining Forcing Efficacy With Pattern Effect and State Dependence." *Geophysical Research Letters* 50.3, e2022GL101700. DOI: 10.1029/2022GL101700.

Zhou, C., M. D. Zelinka, and S. A. Klein (2016). "Impact of Decadal Cloud Variations on the Earth's Energy Budget." *Nature Geoscience* 9.12, 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." *Journal of Advances in Modeling Earth Systems* 9.5, pp. 2174–2189. DOI: 10.1002/2017MS001096.

DECLARATION ON OATH - EIDESSTATTLICHE VERSICHERUNG

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

I hereby declare upon oath that I have written the present dissertation independently and have not used further resources and aids than those stated.

Hamburg, April 2024

Moritz Günther