THE UNIVERSITY OF CHICAGO

CAUSES AND IMPLICATIONS OF TIME-VARYING CLIMATE SENSITIVITY

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF THE GEOPHYSICAL SCIENCES

BY
JONAH BLOCH-JOHNSON

CHICAGO, ILLINOIS
DECEMBER 2018

To my parents, Jeff and Ellen.

Spaziergang

Schon ist mein Blick am Hügel, dem besonnten,
dem Wege, den ich kaum begann, voran.
So fasst uns das, was wir nicht fassen konnten,
voller Erscheinung, aus der Ferne an –

und wandelt uns, auch wenn wirs nicht erreichen,
in jenes, das wir, kaum es ahnend, sind;
ein Zeichen weht, erwidernd unserm Zeichen ...
Wir aber spüren nur den Gegenwind.

The Walk

My eyes already touch the sunny hill,
going far ahead of the road I have begun.
So we are grasped by what we cannot grasp;
it has its inner light, even from a distance –

and changes us, even if we do not reach it,
into something else, which hardly sensing it, we already are;
a gesture waves us on, answering our own wave...
but what we feel is the wind in our faces.

-Rainer Maria Rilke (t. Robert Bly)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

A PhD thesis is a long, plodding, wind-buffeted walk up the sunny hill. I could not have had the endurance to do this alone, and I want to talk a moment to thank the people who kept me going.

My parents, Ellen Bloch and Jeffrey Johnson, have always supported and encouraged me, even as my path has curved and looped. I have been inspired by their hardworking examples, their scholarship, and their kind hearts.

I have been fortunate to have spent my education around a succession of great math teachers, from Bob Reveri in my youth, to my high school teacher Charles Worrall, who provided admonishing encouragement, to my undergraduate teachers, including Jim Walsh at Oberlin, whose lively pedagogy made ideas pop, Peter J. Thomas, also at Oberlin, who first introduced me to the beauty of dynamical systems, and my senior thesis advisor Dave Bayer at Columbia, who made math seem as creative a pursuit as any art.

Ray Pierrehumbert served as my first contact in the world of climate science, and he knew that David Archer's accessible and compelling introduction textbook would have me hooked. Ray was my thesis advisor for the first years of my PhD, and was a fantastic teacher, breaking down complex problems into simple pieces, and communicating it all with an infectious sense of wonder. Ray made me feel that the abstract and even persnickety intellectual style of the pure mathematics world that I was trained in could be useful in as messy and applied a field as climate science. This thesis began with my attempts to understand his skepticism of the linear view of global warming.

Once I arrived in Chicago, I found myself relying again, and again, and again, on my friend and colleague Daniel Koll. Whether working with me late nights in Hinds, collaborating on various climate data hackathons, debating with me about the nature of blackbody radiation, or guiding me through the strange world of academia, Daniel was an everpresent support throughout this journey. I really do not know how I could have done this without his help.

After Ray left for Oxford, I began working with Dorian Abbot. Dorian always encouraged

me to dream big, work efficiently, and to be true to myself, and it's hard to imagine a better set of commandments for research science. He also taught not to be scared of simple questions and thought experiments, ensuring that any knowledge of a system rested on an understanding of its first-order behavior. Dorian has exhibited superhuman stoicism and patience as the advisor of a very peripatetic student, sticking with me as I wandered off to work on Rossbypalooza and LongRunMIP, persevering with me through years spent working towards one simple offline radiative calculation, willing to meet even when I had nothing to show, ready to celebrate when I made a breakthrough or reached a vista.

Liz Moyer was an invaluable mentor and committee member, helping me think through the broader consequences of my work, sharing her wisdom on how to communicate science effectively, and connecting me to academics in disciplines across the university. Malte Jansen, another committee member, helped me find excitement in my research while asking searching questions that opened up new angles. Noboru Nakamura came to me with the opportunity to organize a summer school around climate and statistics, Rossbypalooza, which introduced me to two great colleagues and friends, Cristi Proistosescu, who got me thinking about nonlocal feedbacks, and B. B. Cael, who has been a sympathetic ear as I've developed my thesis. Tiffany Shaw helped me question some fundamental assumptions in my research. Outside of my department, professors Thorsten Mauristen, Kyle Armour, and Tim Cronin stood out as scientists willing to give me a chance, to talk to me about my research and to share insights and ideas from their own.

During the past few years, Maria Rugenstein and I have worked with scientists at many institutions to create the LongRunMIP archive. Maria has brought hard work, good humor, and her own deep well of patience to what has been a rewarding collaboration conducted across many dozens of Skype calls and email chains. She has also been a supportive friend. Data from LongRunMIP served as the basis for Chapters 2 and 5. I thank Maria for helping to plant the seeds for this research, and for being willing to talk with me about time-varying climate feedbacks for what feels like centennial to millennial timescales.

The extremes of the doctoral degree can warp one's perception of the world, one's sense of proportion, one's mental metabolism. I have been blessed to have people in my life who have pulled me back, again and again, towards balance, perspective, and happiness.

Elissa Miller-Kay, a few measures ahead and half way around the world, was always up for a much-needed late-night/midday dispensation of calming sagacity, whooping tropical bird included. Jon Levin listened patiently through many, many rants, helping me improvise my way onwards, even when the task was "unrewarding," and kept me thinking about the bigger picture. Anna Leuchtenberger made me feel like my work had value, and provided poetic inspiration, including the Rilke poem that serves as the text's epigraph. Sarah Litvin gave solidarity as we labored on our dissertations in tandem. Phyllis Ma was a steadfast friend and an ambassador from a different, more artistic world and worldview. Nadia Chana noodled, and smoodled, and gave the requisite number of rambunctious hugs. Lauren Beitler helped me think through my ideas carefully and introduced me to some of my favorite places in Chicago. Eric Potash made a fantastic hiking buddy, whether in a San Francisco parking lot or out along the Point, asking good questions all the way.

Asya Rahlin inspired me with her compassion, determination, and idealism as a scientist, and taught me the importance of self-care. M Wu helped me to keep my head in the clouds and my spirits there too. Navah X. Farahat showed me that scientists can have rich lives outside their work. Shining Li provided joyful encouragement with the right mix of existential "WHAT the –". Marcelle Pierson helped me to be in the here and now and to think twice about using water to cool down a steering wheel. Gwen Smuda gave me the gift of Jones, and some gallows humor for a practitioner of "catastrophe math." Beth Peachey at the keys kept me going as my musical partner-in-crime, whether in Sheepdog with Rachael Weasley, the Karaoke All-Stars with Ted Gordon and John Walsh, or Pangolin Stack with Miguel Lorica. All of my housemates at Haymarket Co-op kept me fed and hopeful.

To the many more who I could name, know that I am thankful for you and the help you have given me. I thank everyone who has joined me for part of this walk.

ABSTRACT

Climate studies often assume that a key metric of global warming, the equilibrium climate sensitivity ($\Delta T_{2x,eq}$, the long-term warming from doubling atmospheric CO_2), is constant, and that it has the same value as the instantaneous climate sensitivity ($\Delta T_{2x,inst}$, an estimate of $\Delta T_{2x,eq}$ made using time series of surface temperature and net top-of-atmosphere radiative flux). Recent studies have shown that both assumptions should be reconsidered: in computer simulations, $\Delta T_{2x,eq}$ can depend on the background state under anthropogenic levels of forcing, and $\Delta T_{2x,inst}$ often differs from $\Delta T_{2x,eq}$. These discrepancies suggest that past estimates of climate sensitivity and of future warming may be incorrect. In this thesis, we explore the causes of these two discrepancies and develop new models that account for them, allowing for more accurate forecasting of future warming.

We explore the first issue by extending a simple energy balance model to account for the fact that the feedback processes that determine the climate sensitivity can change strength in a warmer world, causing $\Delta T_{2x,eq}$ to vary. We use a measure of this feedback temperature dependence, a, to show that positive values of a predict the large increases in $\Delta T_{2x,eq}$ under successive doublings of CO_2 seen in some general circulation models (GCMs). Using this simple model, we show that the range of values of a seen in GCMs implies that observational probabilistic forecasts of climate sensitivity underestimate the risk of high warming. The degree of this underestimate depends on how sensitive the planet is initially. We perform offline calculations on the ECHAM6.1 model to demonstrate that changes in equilibrium climate sensitivity are partly driven by feedback temperature dependence through increases in the water vapor feedback. Perturbing the convective parameters in ECHAM6.1 demonstrates that a small uncertainty in present day climate sensitivity can translate into large uncertainties in sensitivity at higher forcings when a is positive.

We explore the second discrepancy by developing an energy balance model that accounts for the spatially nonuniform nature of the Earth's radiative feedbacks (the primary cause in the difference between $\Delta T_{2x,eq}$ and $\Delta T_{2x,inst}$). We demonstrate a method for estimating

these spatial radiative feedbacks from interannual variability by using multiple regression of top-of-atmosphere fluxes against local and non-local surface temperature. Our method can separate the global feedback into local and nonlocal components, and we show that most models have strong negative nonlocal feedbacks associated with warming in regions of tropical convection. Since warming is initially more weighted towards the tropics, these negative feedbacks make initial values of $\Delta T_{2x,inst}$ lower than $\Delta T_{2x,eq}$.

These two issues are compounding, in that if we are underestimating the $\Delta T_{2x,eq}$ associated with the present climate due to spatially varying feedbacks, this makes it much more likely that we are underestimating the value $\Delta T_{2x,eq}$ will have in the warmer future due to feedback temperature dependence. While the chapters of this thesis use results from computer simulations of climate to assess the power of these effects, they also point towards ways that observations and physical reasoning can be used to measure their strength. Regardless of the method, the present work makes clear that we must account for these two causes of time-varying climate sensitivity to properly forecast future warming.

CHAPTER 1

INTRODUCTION

Humans and much of the other life with which we are connected have a certain envelope of background conditions in which we can survive. This thin range of temperatures, humidities, pressures, and so forth are abundant on the Earth's surface, but are generally uncommon in the universe or even on other planetary surfaces. Unsurprisingly, various people, whether out of curiosity, thankfulness, or concern, have attempted to explain the fortunate situation we find ourselves in.

Starting in the early 19^{th} century, some of these people developed a plausible theory for why the Earth's surface maintains its habitable conditions (Fourier, 1827). With time, it has been found that their theory provides explanations not only for why the Earth's climate is the way it is (Manabe and Wetherald, 1967), but also why the climates of other planets and moons in our solar system are the way they are (e.g., Sagan, 1962). The theory, which can be described as the "energy balance" theory of the climate, also makes predictions about how the climate of planets would change as a result of being pushed in various ways, such as changing the composition of its atmosphere (Tyndall, 1861) or the orientation of its orbit with respect to its star (Milankovitch, 1930). Some of these people, after considering this model, forecasted that anthropogenic CO_2 emissions could provide a push that would cause the surface of the Earth to warm (Arrhenius, 1896; Callendar, 1938). It now appears that this century-old forecast was accurate (Hegerl et al., 2007).

Making this forecast precise has proven to be a challenge. Some initial studies found a range of possible outcomes (Charney et al., 1979), and many decades of research have, for the most part, merely confirmed that outcomes throughout this range are plausible (Knutti et al., 2017). While many of these forecasts have been made using computer simulations of the climate (e.g. Manabe and Wetherald, 1967), there has been some hope that since global warming has begun in earnest, its signal will have risen enough above the noise to allow us to directly estimate the sensitivity of the climate to CO_2 increases from observations (Gregory

et al., 2002). Recent studies along these lines have in fact narrowed the range of possible values, and strikingly have suggested that the range of likely forecasts is lower than that suggested by computer simulations (Otto et al., 2013; Lewis and Curry, 2015). This finding would have the twin benefits of providing progress in our attempt to narrow our forecasts, as well as more breathing room for humanity to respond to climate change.

However, subsequent study has found that the lower observational forecasts are due in part to an oversimplification in the application of the energy balance theory (Armour, 2017; Proistosescu and Huybers, 2017; Goodwin, 2018). While the potential danger of these simplifying assumptions was acknowledged from the beginning (Gregory et al., 2002, 2004), they have often been ignored. It is only in recent years that people have begun the work of extending this energy balance model to account for the true complexity of the climate system (Armour et al., 2012).

The goal of my thesis research has been to develop some of these extensions to the energy balance model of climate change. While I have tested their utility using computer simulations, I have developed them in the hope that they can help our observational forecasts of future warming. Chapter 2 lays out some of the ideas and terminology I will use in the rest of the thesis. Chapters 3 and 4 explore the implications of one of these extensions, feedback temperature dependence, which can have a large effect on the probability of the unlikely but worst-case scenarios in which the sensitivity of the planet is very large. Chapter 5 introduces a method by which spatially nonuniform feedbacks can be predicted from the behavior of the unforced climate. The success of this method suggests that it may be possible to directly address this oversimplification in the observational estimates of climate sensitivity described above. Chapter 6 concludes and notes how the findings of the previous chapters suggest the critical importance of both of these extensions working in concert.

For those who are familiar with climate science and the energy balance model of the climate, you can skip the next section, which is here for the benefit of the lay reader to explain the basic principles at work in this theory.

1.1 Energy balance determines the climate

One of the most fundamental scientific principles that appears to govern our world is that there is a quantity, called energy, that appears to be conserved, impossible to create or destroy on its own. Energy can take different forms, and energy can move between these forms, but when all the energy of a system is counted, it remains unchanged, unless some has been taken out or put into the system in some way. A common form of energy is the energy associated with movement.

It has become apparent that all objects are composed of a multitude of individual molecules, and that these molecules are constantly jiggling around, and therefore moving. The temperature of an object, as we experience it, appears to correspond with the amount of jiggling "movement energy" that the average molecule in that object has. When an object gets hotter, it has more jiggling energy in them. Objects can gain or lose energy in various ways. If they gain energy, the amount of jiggling and therefore temperature will tend to increase, while if they lose energy, the amount of jiggling will decrease and they will cool.

It turns out that all objects are constantly losing energy in the form of electromagnetic radiation. Electromagnetic radiation, which will be simply called radiation in this thesis, can come in various flavors; visible, ultraviolet, and infrared light are common examples. In fact, "light" can also be used as a synonym for electromagnetic radiation, with the understanding that not all forms of this light can be seen with human eyes. All objects give off this radiation, which flies away as photons. As a result, all objects are constantly cooling off. Further, the hotter they are, the faster they lose energy and cool. Our planet would therefore be in a sorry state if it weren't for a compensating source of energy, which comes almost entirely from our Sun (there is a relatively minuscule amount that comes from the planet's interior). The Sun's photons that strike the planet are mostly absorbed by it.

Combining these principles tells us how a planet gets its temperature. If a planet is absorbing more sunlight than it is giving up, it warms, and as it warms it gives off more of its own radiation, narrowing the imbalance. As long as the planet is gaining energy, it

will keep warming, until the energy being given off and taken in are in balance. The same principle works in the opposite direction; if the planet is giving off more energy than it is taking in, it will cool until it is balance once more. The Earth has its climate because for this temperature these two flows of energy are in balance; this is the energy balance theory of climate.

Carbon dioxide inhibits the amount of radiation leaving the planet without directly affecting the amount absorbed by the planet much, and so the surface temperature needed to create the same amount of outgoing radiation is larger than it otherwise would be. As a result, our planet is warmer than it otherwise would be. While the planet was roughly in equilibrium in the recent past, industrial activity, which increases the atmospheric CO_2 concentration, has had the effect of creating an imbalance in the amount of energy coming in and out, so that there is a net gain of energy. This radiative imbalance forces the system to change, and so we call it a radiative forcing.

We seek to understand how the climate will change in the future due to this and other radiative forcings. While there are many messy uncertainties involved in forecasting the forcing itself, we can consider the impact of arbitrary forcing scenarios by first studying a more idealized scenario: starting with preindustrial conditions, and then abruptly increasing the carbon dioxide. As a common point of reference, it has become standard to study the total global mean surface warming caused by doubling the atmospheric CO_2 concentration. This quantity is called the climate sensitivity, and as its names suggests, it is considered a metric that can tell us more broadly how sensitive the climate is to CO_2 forcing, considering that many other climatic properties are considered to scale with temperature.

Radiative forcing and climate sensitivity are tied together by the climate feedback, which is a measure of the ability of warming to undo the radiative perturbation caused by the forcing in the first place. In other words, it is the dependence of the flows of energy on the surface temperature itself. We mentioned above that a warmer world gives off more energy, but there are many other feedbacks that can play role, such as the absorption of additional

sunlight due to melting ice in a warmer world.

Many studies and methods in the past have assumed that the climate sensitivity is a fixed quantity, but a growing body of recent work has suggested that it can change with time, primarily because the climate feedback can change with time. My thesis is concerned with why this occurs. It turns out that there are different causes of time-varying feedbacks, chief among them feedback temperature dependence and the pattern effect. These causes have very different consequences for our forecasts of future CO_2 warming.

CHAPTER 2

TERMINOLOGY: EQUILIBRIUM VS. TRANSIENT NONLINEARITY

The chapters in this thesis come back again and again to the time varying nature of climate sensitivity and climate feedbacks. As a result, they share a lot of terminology. I am including this chapter to lay out our definitions of the various terms associated with these concepts. While some of these definitions may seem abstract, they end up allowing us to disentangle physical phenomena with significantly different implications.

Let Ω be a mathematical object containing all climatic quantities of an atmosphere and ocean. Ω can either be an instantaneous snapshot or an average taken over a period. Let us say that Ω_{inst}, an instantaneous snapshot, is a function of an initial condition Ω_0, the amount of time that has elapsed since this initial condition t, and the amount of CO_2 in the air. We hold all other forcings fixed, including insolation, aerosols, and other greenhouse gases besides water vapor, and assume throughout this thesis that land surface conditions and ice sheets are fixed as well. In other words, we can write $\Omega_{inst} = \Omega_{inst}(\Omega_0, t, d_{CO_2})$, where

$$d_{CO_2}(pCO_2) \equiv log_2(pCO_2/pCO_{2,pi}) \tag{2.1}$$

is the number of atmospheric CO_2 doublings relative to the preindustrial CO_2 concentration $pCO_{2,pi} \approx 280ppm$. Let us define the equilibrium climate associated with a given pair of initial conditions and CO_2 level d_{CO_2} as

$$\Omega_{eq}(\Omega_0, d_{CO_2}) \equiv \lim_{t\to\infty} \Omega_{inst}(\Omega_0, t, d_{CO_2}) \tag{2.2}$$

and let us define the climate that the Earth had in the years before the Industrial Revolution as Ω_{pi}, the preindustrial.

For each Ω, we can take the globally averaged surface temperature $T(\Omega)$. We can then

define the *equilibrium climate sensitivity* at a given CO_2 concentration, d_{CO_2}, to be

$$\Delta T_{2x,eq} \equiv T(\Omega_{eq}(\Omega_{pi}, d_{CO_2} + 1)) - T(\Omega_{eq}(\Omega_{pi}, d_{CO_2})), \tag{2.3}$$

where the equilibrium warming from doubling CO_2 relative to preindustrial is simply

$$\Delta T_{2x,eq} \equiv T(\Omega_{eq}(\Omega_{pi}, 2)) - T(\Omega_{pi}). \tag{2.4}$$

For each Ω, we can also take the globally averaged net top-of-atmosphere radiative flux $R(\Omega)$ (defined downwards positive, so that it is equivalent to the absorbed sunlight less the radiated earthlight per unit area, which is the rate of net energy gain per unit area). Note that by our assumption of equilibrium above, $R(\Omega_{eq}(\Omega_0, d_{CO_2})) = 0$ for all Ω_0 and d_{CO_2}.

We can generally define the radiative forcing $F(\Omega, d_{CO_2})$ associated with setting the CO_2 concentration to d_{CO_2} as follows:

$$F(\Omega, d_{CO_2}) \equiv R(\Omega_{inst}(\Omega, 0, d_{CO_2})) - R(\Omega) \tag{2.5}$$

so that when Ω is a climate in equilibrium, the last term drops away, e.g. the forcing associated with doubling CO_2 relative to preindustrial is simply $F_{2x} \equiv F(\Omega_{pi}, 2) = R(\Omega_{pi}, 0, 2)$. Let us assume that increasing the CO_2 concentration always leads to a positive radiative forcing F, which should be true for all climate states that we will be considering in this work.

Suppose we had a climate that is in preindustrial equilibrium, Ω_{pi}. Suppose we abruptly increase d_{CO_2}, so that we have a positive F, and because $R(\Omega_{pi}) = 0$, a positive $R(\Omega)$. The planet will gain energy, and therefore start to warm. Let us assume that changes in R are primarily caused either by quantities that typically change with d_{CO_2} (the direct effects of CO_2 changes on outgoing radiation, in some cases changes to the stratospheric lapse rate) or quantities that typically change with the surface temperature T (all other climatic variables). (Note that there may be some changes in cloudiness that might scale more with d_{CO_2} than

T, but here we associate them with T). This motivates us to define the climate feedback $\lambda \equiv \partial R/\partial T$.

The climate feedback is the sum of many individual components, each representing a different mechanism by which changes in T affect R. Typically the most substantial effect is that a warmer planet directly emits more infrared radiation due to the Planck effect, which contributes a large negative term to λ. Increasing T also increases humidity, which counteracts this increased radiation, contributing a smaller positive term to λ. Changes to vertical profiles of temperature, ice cover, and cloudiness all contribute terms as well. Since the Planck effect is typically the largest in magnitude, λ is typically negative. This ensures that as T increases, R decreases, so that eventually R is zero once again, and the planet is in equilibrium. Therefore, the strength of λ determines how much warming is necessary to equilibrate to a given F.

Generally, λ (and R) depends on Ω – that is, it can change with d_{CO_2}, T, or even different Ωs which share the same T and d_{CO_2}. The oversimplification referred to above is to ignore this and assume that the forcings associated with anthropogenic warming are small enough that λ is fixed. This assumption lets us do many things; for example, we can easily solve for the warming ΔT associated with a given forcing applied to the preindustrial:

$$\begin{aligned} 0 &= R(T(\Omega_{eq}(\Omega_{pi}, d_{CO_2})), d_{CO_2}) && (2.6)\\ &= R(T(\Omega_{pi}) + \Delta T, d_{CO_2}) && (2.7)\\ &\approx R(\Omega_{pi}, 0, d_{CO_2}) + \lambda \Delta T && (2.8)\\ &= F(\Omega_{pi}, d_{CO_2}) + \lambda \Delta T && (2.9) \end{aligned}$$

which we can rearrange as

$$\Delta T = -F(\Omega_{pi}, d_{CO_2})/\lambda \tag{2.10}$$

In particular, $\Delta T_{2x} = -F_{2x}/\lambda$.

This assumption also allows us to apply estimates of λ made in one context to its value

in other contexts. For example, let us say we have a time series $\Omega_{inst}(\Omega_0, t, d_{CO_2})$, where d_{CO_2} is held fixed. Let us define the instantaneous climate feedback

$$\lambda_{inst} \equiv \partial T(\Omega_{inst}(\Omega_0, t, d_{CO_2}))/\partial R(\Omega_{inst}(\Omega_0, t, d_{CO_2})) \tag{2.11}$$

where this partial derivative can be estimated either by taking a regression of R_{inst} against T_{inst} or by taking finite differences of R_{inst} and T_{inst} at times t_1 and t_2. Let us define the instantaneous climate sensitivity $\Delta T_{2x,inst}$ as $-F/\lambda_{inst}$. This has been used as an estimate of the equilibrium climate sensitivity (Gregory et al., 2002; Murphy et al., 2009).

In recent years, it has been acknowledged that λ changes under anthropogenic levels of forcing (Andrews et al., 2012; Armour et al., 2012; Rose et al., 2014; Bloch-Johnson et al., 2015). These changes can occur for different reasons, with different implications. To explain this, we need to take a few steps that might seem a bit abstract, but actually allow us to make a very critical distinction between different types of time-varying climate sensitivity.

First, we will argue that it is unlikely that there can be two different Ω_{eq} associated with the same T (that is, $T(\Omega_{eq,1}) = T(\Omega_{eq,2})$ implies $\Omega_{eq,1} = \Omega_{eq,2}$). Note that this is different than saying whether there can be two different Ω_{eq} associated with the same d_{CO_2}; there clearly can be, as this allows for hysteresis of the sort seen in Budyko-Sellars models (Budyko, 1969; Sellers, 1969), in which the equilibrium climate associated with a CO_2 level can depend on if one starts in an ice-free or Snowball Earth state.

Next, we assume that if $d_1 = d_2$ and $T_1(\Omega_{eq,1}) = T_2(\Omega_{eq,2})$, then $\Omega_{eq,1} = \Omega_{eq,2}$. Although it is possible that this kind of bistability can exist – for example, perhaps there can be different circulation patterns that create different pole-to-equator temperature gradients but the same T for a given d_{CO_2} – this seems like an exotic possibility that has not been found in the literature to our knowledge. If we accept this assumption, we have then only to show that if $\Omega_{eq,1}$ and $\Omega_{eq,2}$ have different values of d_{CO_2}, then $T(\Omega_{eq,1})$ and $T(\Omega_{eq,2})$ differ to prove that there only be one Ω_{eq} associated with the a given T. We show this using proof

by contradiction.

We stated above that we assume any increase in CO_2 causes a positive radiative forcing, and that any planetary energy gain leads to some degree of surface warming, no matter how small. Let us assume that forcing is additive; that is,

$$F(\Omega_0, d_2) = F(\Omega_0, d_1) + F(\Omega_{inst}(\Omega_0, t, d_1), d_2) \tag{2.12}$$

and that in spite of any hysteresis, it is always possible to get from one equilibrium climate state to another through a series of forcings (which need not all be of the same sign; for example, in the Snowball Earth example above, one could move between climate states with the same d_{CO_2} but different T by forcing in one direction and then back in the other). Suppose that we have $\Omega_{eq,1}(\Omega_{0,1}, d_1)$ and $\Omega_{eq,2}(\Omega_{0,2}, d_2)$ such that $d_1 \neq d_2$ but $T(\Omega_{eq,1}) = T(\Omega_{eq,2})$. Assume, without loss of generality, that $d_1 < d_2$. Apply a series of forcings to $\Omega_{eq,1}$ so that it ends up as $\Omega_{eq,2}$. By the additive property above, the sum of all of these forcings will be $F(\Omega_{eq,1}(\Omega_{0,1}, d_1), d_2)$. Since $d_2 > d_1$, this will have to be positive. However, the total change in temperature is 0. Indeed, if we were to then reduce the CO_2 level to be slightly less than d_2 but still above d_1, this would cause a slight cooling. We would then still have a positive forcing with respect to $\Omega_{eq,1}$, but a net cooling. This contradicts our assumption that any positive forcing causes at least some warming.

Our assumptions therefore imply that for each T there is a unique Ω_{eq}, and we can define a function $\Omega_{eq}(T)$. We can also define $R_{eq}(T, d_{CO_2}) = R(\Omega(\Omega_{eq}(T), 0, d_{CO_2})) = F(\Omega_{eq}(T), d_{CO_2})$. Most importantly, we can define the equilibrium climate feedback $\lambda_{eq}(T) = \partial R_{eq}(T, d_{eq}(T))/\partial T$. We can now find the equilibrium warming response to a given forcing applied to the preindustrial by solving for ΔT in the following:

$$F(\Omega_{pi}, d_{CO_2}) + \int_{T(\Omega_{pi})}^{T(\Omega_{pi})+\Delta T} \lambda_{eq}(T) dT = 0 \tag{2.13}$$

which reduces to $\Delta T = -F/\lambda_{eq}$ when λ_{eq} is constant with temperature.

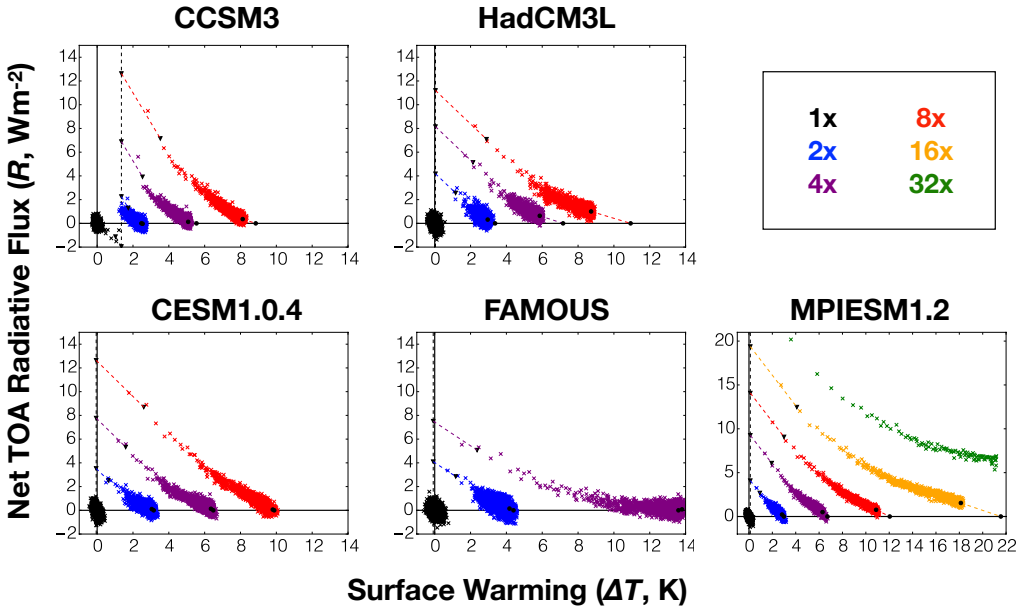

Figure 2.1: Abrupt CO_2 simulations of five different coupled general circulation models. For each simulation, annual averages of R_{inst} vs. T_{inst} are shown. The local tangent of each curve increases over time and as the models warm, so that λ_{inst} increases everywhere. However, for some models, equilibrium warming appears to be proportional to the forcing no matter the forcing level, implying that λ_{eq} is constant. This illustrates the distinction between transient nonlinearity (changing λ_{inst}) and equilibrium nonlinearity (changing λ_{eq}). Note that models with just transient nonlinearity exhibit an abrupt shift towards a higher feedback, while models with equilibrium nonlinearity appear to have a more gradual increase. Runs are drawn from the LongRunMIP archive; for more information, see Chapter 5.

This allows us to demonstrate the critical and subtle point: λ_{inst} (and therefore $\Delta T_{2x,inst}$) can be inconstant while λ_{eq} (and therefore $\Delta T_{2x,eq}$) is constant. An example can be seen in the CCSM3 and CESM1.0.4 panels of Figure 2.1. Each collection of "x"'s of the same color represents years of a run of an abruptly forced simulation. For each run, we can estimate λ_{inst} by taking the local tangent of R vs. T. λ_{inst} appears to increase somewhat abruptly around halfway through each run's warming. As a result, $\Delta T_{2x,inst}$ abruptly increases at this point. However, each run appears to have the same ratio of forcing to equilibrium warming, so that each subsequent doubling of CO_2 results in the same amount of warming. This implies that $\Delta T_{2x,eq}$ is the same across all of these temperatures, so that λ_{eq} most be constant as well.

We can directly estimate $\lambda_{eq,pi}$ by assuming that $R_{eq}(T, d_{CO_2}) = R_{eq}(T, 1) + F(\Omega_{pi}, d_{CO_2})$, which is equivalent to assuming that λ_{eq} is independent of CO_2 level, an assumption we explore in Chapter 4. We can then use the estimates of $F(\Omega_{pi}, d_{CO_2})$ and $\Delta T(\Omega_{eq}(\Omega_{pi}, d_{CO_2}))$ in Figure 2.1 to constrain $R_{eq}(T, d_{CO_2})$ for each model. We plot curves of $R_{eq}(T, d'_{CO_2})$ for each model in Figure 2.2, where d'_{CO_2} is the highest doubling in a run for that model. The slopes of these curves are $\lambda_{eq}(T)$ for each model respectively, allowing us to compare them to slopes of the instantaneous values of R and T showing the evolution of $\lambda_{inst}(T_{inst})$. For all models but FAMOUS, there is some degree of deviation between the slopes of the two lines.

The reason for this discrepancy is as follows: the derivation of λ_{eq} assumes that there is a specific Ω_{eq} associated with each T and d_{CO_2}, which in turn implies a specific R_{eq}. No such constraint holds in an instantaneous run; different Ω_{inst}s associated with the same T and d_{CO_2} can occur, with different spatial and seasonal patterns of surface temperature and other climatic variables. Since these can have different values of R, it stands to reason their R can change with T in different ways. Thus there can be a "pattern effect" that creates a changing $\Delta T_{2x,inst}$ while we have a constant $\Delta T_{2x,eq}$, where the relative amounts of warming in different regions differs from the equilibrium response, setting off feedback processes with different strengths. (Note that the equilibrium response can have a changing pattern too, and if this pattern involved a shift between regions of different feedback strengths, this would cause the equilibrium feedback to change. These equilibrium pattern changes are folded into feedback temperature dependence however, as they are part of $\Omega_{eq}(T)$.)

On the other hand, changing λ_{eq} does tend to imply a changing λ_{inst}. Both FAMOUS and MPIESM1.2 have λ_{eq}s that get clearly more positive with warming. Plots of instantaneous R_{inst} and T_{inst} under high forcing seem to exhibit this behavior. It stands to reason that if the feedbacks associated with the "equilibrium climates" change in a warmer world, that some of those same changes will be felt even in the presence of a pattern effect.

When $\Delta T_{2x,eq}$ changes with forcing, we call this *equilibrium nonlinearity*. When λ_{eq}

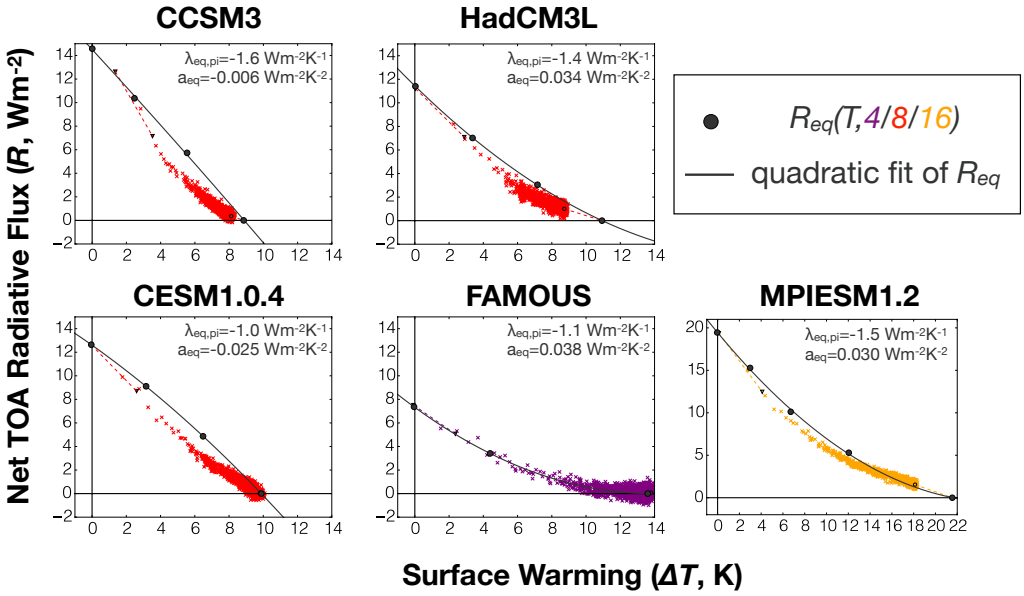

Figure 2.2: The same as Figure 2.1, except now only the runs with the largest CO_2 increase d_{CO_2} for each model are shown (excepting MPIESM1.2, where we use second largest, as the largest appears to undergo a runaway; see Chapter 3 for more details) and are compared with estimates of $R_{eq}(T, d_{CO_2})$. Once more, there is a distinction between models with changing λ_{eq} (the slope of R_{eq}) vs. those with changing λ_{inst}. The same model can have different values of R for the same value of T because the equilibrium pattern of warming can differ from the spatial pattern seen instantaneously. Values of $\lambda_{eq,pi}$ and a_{eq} are estimates of the preindustrial feedback and feedback temperature dependence as defined in Chapter 3.

changes with temperature, we say it has *feedback temperature dependence*, which is one of several potential causes of equilibrium nonlinearity. When a climate has a $\Delta T_{2x,inst}$ that changes with time, we say that it has *transient nonlinearity*. When it has transient nonlinearity in excess of what would be expected from its equilibrium nonlinearity, this is evidence of a pattern effect. In general, equilibrium nonlinearity implies transient nonlinearity but transient nonlinearity does not imply equilibrium nonlinearity. As we saw in Figure 2.2, some models have feedback temperature dependence while others don't, while most models appear to exhibit transient nonlinearity (Andrews et al., 2014).

With those definitions in place, we can now lay out the rest of the thesis. In Chapter 3, we discuss why feedback temperature dependence plays an outsize role in determining the risk of high equilibrium climate sensitivity. In Chapter 4, we use offline calculations to

determine why feedback temperature dependence exists in ECHAM6.1 and use a perturbed physics ensemble to illustrate more concretely why it has a larger effect on equilibrium climate sensitivity than other sources of equilibrium nonlinearity. In Chapter 5, we turn to the pattern effect, and show that in spite of the many λ_{inst} associated with the same T, the spatial pattern of feedbacks, Λ, may be fixed, and independent of the particular spatial pattern of warming associated with a given $\Omega_{inst}(\Omega_{pi}, t, d_{CO_2})$. We demonstrate how Λ may be derived directly from records of interannual variability, given sufficient years of data, by using multiple regression. We then summarize our conclusions in Chapter 6.

CHAPTER 3

FEEDBACK TEMPERATURE DEPENDENCE DETERMINES THE RISK OF HIGH WARMING

3.1 Introduction

As discussed in the proceeding Chapters, studies of equilibrium climate sensitivity often assume that the long-term warming caused by increasing atmospheric CO_2 to some new fixed level is proportional to the radiative forcing associated with that CO_2 increase (e.g., Andrews et al., 2012; Otto et al., 2013). This allows the general response to be characterized by the equilibrium climate sensitivity ($\Delta T_{2\times}$), the warming caused by doubling CO_2 from preindustrial levels. This assumption can be summarized as supposing the climate has no equilibrium nonlinearity, emphasizing that an equilibrium linear world can still have transient nonlinearity (see Figures 2.1 and 2.2).

This linear assumption is always made with the caveat that, for large enough forcings, feedbacks do change strength, often as a result of changes in temperature (e.g., Hansen et al., 1984), making the equilibrium warming response nonlinear with forcing. An extreme case is given by the runaway greenhouse (e.g., Nakajima et al., 1992), in which as the world warms, the water vapor feedback strengthens, eventually causing the climate to jump to a warmer state. A warmer world also has a weaker surface albedo feedback, due to the disappearance of snow and ice (Wetherald and Manabe, 1975).

Studies have explored the importance of equilibrium nonlinearity by running general circulation models (GCMs) to equilibrium under different levels of CO_2 forcing Manabe and Bryan (e.g., 1985); Colman et al. (e.g., 1997); Caballero and Huber (e.g., 2013). While many models do not show much difference in sensitivity under different amounts of CO_2 forcing, some do. For example, the ECHAM6 model warms about as much from its third doubling of CO_2 as from its first and second doubling combined, with each doubling having roughly the same forcing (Meraner et al., 2013). This invites the question of under what conditions

the linear approximation breaks down.

We assess the limits of equilibrium linearity by exploring the behavior of a zero-dimensional energy balance model with a term representing feedback temperature dependence, following Roe and Baker (2007) and Zaliapin and Ghil (2010). We estimate a reasonable range of values for this dependence by diagnosing it from GCM experiments and by using physical arguments. We find that a positive feedback temperature dependence can cause Earth's sensitivity to increase substantially from its linear approximation, with some parameter choices leading to a jump to a warmer state under only one or two doublings of CO_2. In particular, the linear approximation breaks down for the long tail of high sensitivity cases, which play a large role in assessments of economic risk (Weitzman, 2011). We show that observational estimates of climate sensitivity likely underestimate the risk of high warming by neglecting equilibrium nonlinearity, and that understanding this nonlinearity is essential for reducing our uncertainty of these high risk scenarios.

3.2 Model setup and methods

We use a zero-dimensional energy balance model of equilibrium climate sensitivity (Figure 3.1). Let T, d_{CO_2}, and R be defined as in Chapter 2.1. As we have said, R_{eq} can be expressed as a function of T and d_{CO_2}. For this chapter, let us assume the pattern effect is negligible, so that we will mostly drop the eq subscript (e.g., $R = R_{eq}$) except where clarification might be helpful. In general, R acts to warm or cool the surface ($dT/dt \propto R$).

Suppose that the preindustrial Earth was in equilibrium (i.e., $R(T_{pi}, 0) = 0$, where $T_{pi} \approx 287K$ is the preindustrial values of T). If the CO_2 concentration is increased, R would increase by a radiative forcing F, causing the planet to gain energy and T to increase until R is 0 again, resulting in an equilibrium warming ΔT.

The slope of R at T_{pid} with respect to T is the preindustrial feedback, which we call λ_{pi} (i.e., $\lambda_{pi} \equiv \frac{\partial R}{\partial T}|_{T_{pi},0}$). We use "feedback" to describe the sum of *all* the ways T changes R, including the Planck effect. In our sign convention, a negative λ_{pi} implies a stable

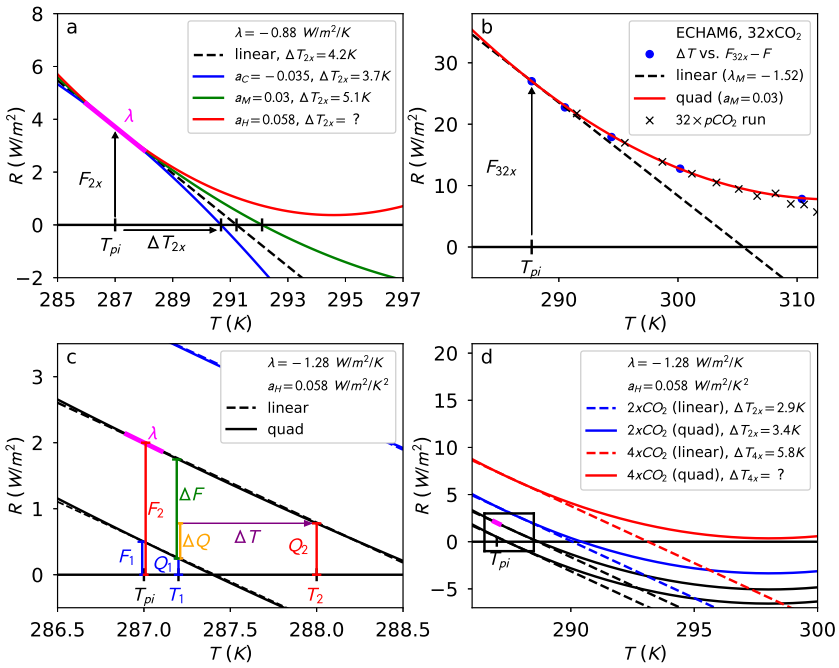

Figure 3.1: Global-annual-mean net top-of-atmosphere energy flux R (downwards positive) as a function of global-annual-mean surface temperature T, where each curve has a fixed CO_2 concentration. a) Doubling CO_2 increases R by a forcing $F_{2\times} \approx 3.71\ W/m^2$ (Collins et al., 2013). All four curves have the same preindustrial feedback, λ_{pi} (the slope of R at T_0), but different values for the feedback temperature dependence, a, resulting in different warming responses $\Delta T_{2\times}$. Colored lines have values of a where the subscripts correspond to GCMs in Figure 3.2. For a_H (red curve), the feedback becomes positive, causing runaway warming. b) ECHAM6 under five CO_2 doublings. We quadratically fit R to preindustrial conditions and the equilibrium response to four CO_2 forcings (blue circles, Meraner et al. (2013)), adding $F_{32\times}$ to estimate $R_{eq}(T, 5)$ (red curve). The time series of T vs. R_{inst} from the $32\times CO_2$ run (black $\times$s) follows this curve, until the model "blows up" (data from T. Mauritsen, 2015, personal communication). c) Illustrative scenario with fixed values λ_{pi} and a, showing how λ_{pi} can be estimated from observations of surface temperature (T), forcing (F), and heat uptake (Q) from times 1 and 2. Warming appears linear. d) Same scenario after one (blue) and two (red) CO_2 doublings, with the previous panel inset. The linear approximation works for one doubling, but under two doublings the quadratic model runs away.

preindustrial Earth, and a more negative λ_{pi} implies a less sensitive preindustrial Earth. If R is a linear function of T (Figure 3.1a, dashed black line), this feedback remains the same under warming, and we get the linear estimate $\Delta T = -F/\lambda_{pi}$, as stated above.

We can explore the accuracy of this linear estimate by adding a quadratic term representing the temperature dependence of the feedback (e.g., Roe and Baker, 2007; Zaliapin and Ghil, 2010). We call the coefficient of this term a (i.e., $a \equiv \frac{1}{2}\frac{\partial^2 R}{\partial T^2}|_{T_{pi},1}$). There is also a quadratic term representing the CO_2 dependence of the feedback. While this term can influence the exact value of ΔT, it does not have the same potentially extreme effect on ΔT as the feedback temperature dependence (see Appendix Section 7.1.3 and compare Figures 3.2 and 7.3). To clearly explain this latter effect, we follow these earlier studies in leaving out feedback CO_2 dependence.

To get the quadratic estimate of ΔT for a given forcing F, we solve the quadratic equation

$$-F = \lambda_{pi}\Delta T + a\Delta T^2. \tag{3.1}$$

Negative a (Figure 3.1a, blue line) implies that the feedback gets more negative under warming, giving less than linear warming. Positive a (Figure 3.1a, green and red lines) implies the feedback gets less negative under warming, and so giving more warming. For large enough a (Figure 3.1a, red line), the feedback becomes positive before equilibrium is reached, and the quadratic model warms indefinitely. If we include still higher-order nonlinear terms of R, these terms would ensure that indefinite warming would not occur (e.g., Figures 7.1 and 7.2).

To explore the applicability of the linear approximation, we compare linear and quadratic estimates of $\Delta T_{2\times}$, $\Delta T_{4\times}$, and $\Delta T_{8\times}$ for reasonable values of λ_{pi} and a. We estimate the range of reasonable λ_{pi} by dividing the minimum and maximum $\Delta T_{4\times}$ from the CMIP5 $abrupt4xCO_2$ experiment (4.16 and $9.34K$, Andrews et al. (2012)) by $-F_{4\times} \approx -7.42W/m^2$ (Collins et al., 2013), giving $-1.78 \leq \lambda_{pi} \leq -0.79W/m^2/K$.

We can estimate the range of likely a by estimating a for GCMs that were run to equilibrium under different CO_2 forcings using least squares linear regression (see Appendix Section 7.1.1; Appendix Section 7.1.3 also discusses the effect of CO_2 dependence on these estimates). Estimates of a for various GCMs are given by the letters in Figure 3.2, and suggest a range of $-0.04 \leq a \leq 0.06\ W/m^2/K^2$. This range is similar to an earlier survey, ($-0.06 \leq a \leq 0.06 W/m^2/K^2$, Roe and Armour (2011)). Our range is narrower because we only include studies that use CO_2 forcings, as opposed to solar forcings or conditions at the last glacial maximum.

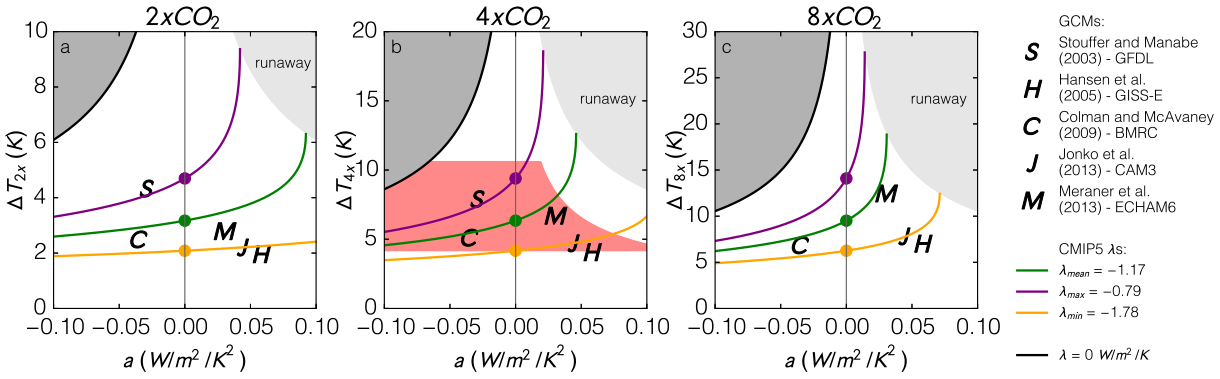

Figure 3.2: a (the feedback temperature dependence) vs. $\Delta T_{2\times}$, $\Delta T_{4\times}$, and $\Delta T_{8\times}$ (the equilibrium warming from one, two, and three doublings of CO_2 respectively). Letters represent GCMs and are centered on pairs of a (quadratic estimate, see Appendix Section 7.1.1) and $\Delta T_{n\times}$ (reported values). Colored lines show how quadratic estimates of $\Delta T_{n\times}$ (using Equation 3.1) vary with a for fixed values of λ_{pi}, assuming forcings of $F_{n\times} = log_2(n)3.71\ W/m^2$. (Note that all λs are λ_{pi}s.) Colored circles where these lines intersect the center vertical give linear estimates of $\Delta T_{n\times}$. Some lines intersect the shaded regions in the upper right corner of each panel. For values of a greater than these intersections, the quadratic model experiences runaway warming. If we assume that the preindustrial climate was stable, then $\lambda_{pi} \leq 0$ and the dark region in the upper left hand corner of each panel is inaccessible under the quadratic model. CMIP5 models must lie in the pink region in b) to avoid quadratic runaway under RCP8.5. Note that the y-axis increases proportionally in each panel.

As an example, Figure 3.1b shows a quadratic fit of R (red line) to equilibrium runs of ECHAM6 (blue dots, Meraner et al. (2013)), where the blue dots represent preindustrial conditions and the response to four successive doublings of CO_2. We have added a forcing $F_{32\times}$ to predict ECHAM6's response to five CO_2 doublings. Instantaneous values of T vs.

N for the $32 \times CO_2$ run closely tracks our curve (black crosses, T. Mauritsen, 2015, personal communication). The quadratic model successfully predicts that the five doubling run of ECHAM6 does not equilibrate, although the exact reasons for this blow up are unclear.

To understand and justify this range of feedback temperature dependence, we look at the various physical processes that contribute to it. The Planck feedback gets more negative under warming, contributing a term of $\frac{1}{2}(d^2[\sigma T^4]/dT^2|_{255K}) \approx -0.02\ W/m^2/K^2$, where $255K$ is Earth's equilibrium temperature. The water vapor feedback gets more positive, possibly offset by changes to the lapse rate feedback (Colman et al., 1997). Meraner et al. (2013) studied the changing sensitivity of a moist adiabat with fixed relative humidity and tropopause temperature, representing the combined water vapor, lapse rate, and Planck feedbacks. We estimate $a \approx 0.05\ W/m^2/K^2$ for their model. The surface albedo feedback weakens under warming as snow and ice disappear. An early study of this weakening with idealized geography (Manabe and Bryan, 1985) found $a \approx -0.1 W/m^2/K^2$ for the combined non-cloud feedbacks. The contributions to a from cloud feedbacks and cross terms between feedbacks are uncertain, given our uncertainty about the cloud feedback itself.

3.3 Results

The colored lines in Figure 3.2 show how quadratic estimates of $\Delta T_{2\times}$, $\Delta T_{4\times}$, and $\Delta T_{8\times}$ vary with a for fixed values of λ_{pi}. Linear estimates are given by the circles where these lines intersect the center vertical. When a is positive, warming makes the feedback become less negative. In turn, a less negative feedback causes additional warming. These two effects feed on each other, so that positive values of a can greatly increase warming from the linear estimate. This effect gets stronger for less negative λ_{pi} and for larger CO_2 forcing.

Some colored lines intersect the shaded region in the upper right corner of each subfigure. For values of a greater than these intersection points, these values of λ_{pi} make the quadratic model warm indefinitely, like the red curves in Figures 3.1a and d. In these cases, the stable equilibrium that the quadratic model is tracking disappears in a saddle-node bifurcation

(Strogatz (1994)). If Earth has values of λ_{pi} and a such that for a given forcing F, the quadratic model warms indefinitely, we say that Earth experiences a "quadratic runaway" for that forcing. If Earth has $\lambda_{pi,max} = -0.79 W/m^2/K$ and $a_J = 0.042 W/m^2/K^2$ (Jonko et al., 2012), it would experience a quadratic runaway from just one CO_2 doubling. Under two doublings, $\lambda_{pi,mean} = -1.17 W/m^2/K$ and $a_H = 0.058 W/m^2/K^2$ (Hansen et al., 2005) would lead to a quadratic runaway. This latter case appears linear under only one doubling, with a ΔT_{2x} of only $3.4K$ (Figure 3.1d).

If Earth experiences a quadratic runaway for a forcing F, the warming response to that forcing cannot be estimated from the quadratic model, since the quadratic model warms indefinitely, which is clearly unphysical. To know the warming caused by F, we would have to include higher-order terms, and the exact value of the resulting warming would depend sensitively on the value of these higher-order terms (see Figure 7.2). Since we only have detailed observations of Earth over a relatively small temperature range, our physical (GCM) modeling of these higher-order terms is poorly constrained. Further, as Earth warms, new feedbacks such as those from the melting of ice sheets, changes to the ocean circulation, and the release of soil carbon come into play (Lunt et al., 2010), further impeding our ability to estimate these higher-order terms.

As a result, if Earth quadratically runs away due to a forcing F, the resulting warming may be inherently uncertain. In fact, if these higher-order terms are negative enough, Earth might not run away at all, even though its quadratic approximation runs away (Figure 7.2, green curve). However, if the higher-order terms are not very negative, Earth will go through a period of runaway warming before ultimately stabilizing (e.g., Figure 7.1; Figure 7.2, red and blue curve). These jumps to warmer states can be hundreds of degrees, as in the case of the runaway greenhouse, or much fewer, depending on the higher-order terms discussed. Jumps to warmer states far short of the runaway greenhouse have been seen in lower-dimensional models (Abbot and Tziperman, 2008; Popp et al., 2014) and GCMs (Popp, 2014), and may explain Cenozoic hothouse climates (Pierrehumbert, 2013).

The quadratic model breaks down in the limit of high warming and large higher-order temperature dependence, just like the linear model. The quadratic runaway is simply a case where this breakdown is guaranteed. Care should be taken in using the quadratic model to estimate large warming. However, a main goal of this paper is to estimate the regions of parameter space for which linearity breaks down. Higher-order terms only grow these regions. Further, the quadratic approximation often works well (e.g., Figure 7.2a). Even if the quadratic model does run away, the forcing and temperature at which the quadratic estimate of R reaches its minimum can predict when Earth will begin a potential jump to a warmer state (e.g., Figure 3.1b and Figure 7.1e), and serves as a rough lower bound on warming for these quadratic runaway cases.

Jumps to warmer states and other large deviations from equilibrium linearity under only one or two doublings are uncommon in the published literature, outside of occasionally appearing in perturbed physics ensembles (e.g., Tomassini et al., 2015). When such deviations do appear, they are usually in response to much larger forcings (e.g., Boer et al., 2005). Of the five GCMs analyzed here, all but ECHAM6 occupy regions where the linear approximation works well, having either highly negative preindustrial feedbacks (λ_{pi}), negative or small feedback temperature dependence (a), or both. There are two main reasons why equilibrium nonlinearity may be uncommon in the published literature.

First, GCMs that experience quadratic runaway would behave in ways suggestive of a non-physical loss of stability, i.e. model "blow up." In a MIROC perturbed physics ensemble (e.g., Yokohata et al., 2010), 20% of parameter sets created GCMs that, after being subjected to an abrupt doubling of CO_2, did not regain stability after 70 years. As a result, these cases were disregarded. A similar selective exclusion may occur during the development of GCMs and tuning of parameter values for CMIP experiments (e.g., Mauritsen et al., 2012) or in the adaptation of new GCMs from existing models (e.g., Knutti et al., 2013). Given that there is evidence that CMIP GCMs may be biased towards a specific range of climate sensitivities (e.g., Huybers, 2010), it is still more plausible that there may be a bias against runaway

behavior.

As an example, under the RCP8.5 scenario from the CMIP5 experiments, the CO_2 concentration increases over 250 years to $1962ppm$, and is then held constant for 50 years. If we make the approximation that GCMs that would experience a quadratic runaway under a stabilized $1962ppm$ CO_2 concentration would also do so under RCP8.5, we can use the quadratic model to estimate what values of λ_{pi} and a CMIP5 GCMs are necessary to prevent this quadratic runaway. Specifically, for a given a and F, λ_{pi} will not cause a quadratic runaway if $\lambda_{pi} < -2\sqrt{Fa}$, and we assume $F_{1962} \approx \log_2(1962ppm/C_0) \times 3.71W/m^2$. We can combine this with the CMIP5 ΔT_{4x} range to draw the pink region in Figure 3.2b, where the curve on the right-hand side is caused by the quadratic runaway cut off. All CMIP5 GCMs presumably fall inside this region. Since three of the GCMs for which we estimate a fall not far from the curved edge, it would seem that this cutoff is artificial. Further, Meraner et al. (2013) find that one of the CMIP5 models (CSIRO-Mk3.6.0) comes close to runaway under RCP8.5.

In our discussion above, we argue that GCMs should not be expected to accurately model warming beyond a certain point, and so GCMs model λ_{pi} and a much better than higher order terms. If a GCM never stops warming after being subjected to a CO_2 increase, it may be accurately modeling a pair of λ_{pi} and a that cause quadratic runaway, and therefore may be just as physically accurate as a model that has these same values of λ_{pi} and a, but higher order terms that happen to cause the model to stabilize.

The second reason we may not often see extreme deviations from equilibrium linearity is that positive a combined with small-magnitude negative λ_{pi} may be unphysical. As an example, a highly positive low cloud feedback, which would result in a small-magnitude negative λ_{pi}, could saturate as clouds are lost, causing a negative a. Generally, it is not obvious that a should depend strongly on λ_{pi}, since the dominant term determining a can be different than the dominant term determining λ_{pi} (e.g., water vapor feedback vs. cloud feedback). More work needs to be done to understand how λ_{pi} and a covary.

3.4 Implications for observational estimates of sensitivity, and for the long tail

Studies often estimate equilibrium climate sensitivity from observations of the recent past (Gregory et al., 2002). Since the atmosphere equilibrates faster than the oceans, we assume the top-of-atmosphere energy imbalance N is balanced by the ocean heat uptake Q. If we have observations of Q, T, and forcing F for two reference periods, we can estimate λ_{pi} using

$$\lambda_{pi} \approx -\frac{\Delta F - \Delta Q}{\Delta T}, \tag{3.2}$$

as demonstrated by Figure 3.1c. $\Delta T_{2\times}$ can then be estimated using the linear model, i.e. $\Delta T_{2\times} = -F_{2\times}/\lambda_{pi}$. This estimation of λ_{pi} assumes that there is no equilibrium *or* transient nonlinearity.

Our knowledge of ΔF, ΔQ, ΔT, and $F_{2\times}$ is uncertain, causing uncertainty in our knowledge of λ_{pi} and $\Delta T_{2\times}$. Since many distributions of λ_{pi} have non-zero probability that λ_{pi} is arbitrarily close to 0, the linear model implies probability that $\Delta T_{2\times}$ is arbitrarily large, resulting in a long tail of high climate sensitivities (Roe and Baker, 2007). However, $\lambda_{pi} \to 0$ does *not* imply $\Delta T_{2\times} \to \infty$, but instead that the linear term is disappearing, so that nonlinear terms completely determine the nature of the warming. Accounting for nonlinearity is therefore essential for properly estimating the long tail (e.g., Zaliapin and Ghil, 2010).

For a given value or distribution of a, we must alter our analysis in two ways. First, our estimate of λ_{pi} must account for the feedback change from warming between the reference periods,

$$\lambda_{pi} \approx -\sqrt{\left(\frac{\Delta F - \Delta Q + a\Delta T^2}{\Delta T}\right)^2 + 4a(F_1 - Q_1)}, \tag{3.3}$$

where F_1 and Q_1 are values for the first reference period. As Figure 3.1c demonstrates, this effect is usually negligible, as there has not been enough warming for nonlinearity to act. This makes estimating a from observations of the recent past difficult. Second, we must account

for a in estimating long-term warming from λ_{pi} by using Equation 3.1. We demonstrated in Section 3.3 that this effect can be significant: Figure 3.1d shows that the scenario that appeared linear in Figure 3.1c experiences a quadratic runaway under two doublings.

In Figure 3.3, we use these two alterations to recalculate estimates of the distribution of $\Delta T_{4\times}$ from three studies: one with a distribution that roughly matches the distribution of CMIP models (Murphy et al., 2009), and two recent studies that find substantially lower climate sensitivities (Otto et al., 2013; Lewis and Curry, 2015). Assuming linearity (dashed black lines) recreates the studies' results (with $\Delta T_{4\times} = 2\Delta T_{2\times}$). Two of the nonlinear cases (colored lines) assume that a has either the minimum ($-0.035 W/m^2/K^2$, blue line) or maximum ($0.058 W/m^2/K^2$, red line) value seen in our GCM survey. The other two nonlinear cases assume a has a uniform distribution of possible values, with a range of either ± 0.06 (purple line) or 0 to $0.06 W/m^2/K^2$ (green line). The latter case is included because Meraner et al. (2013) argue that a is likely positive (e.g., the majority of CMIP5 GCMs appear to have positive a). Note that in calculating these new distributions of ΔT_{4x}, we are assuming λ_{pi} and a are independent, which need not be the case, as discussed above.

Distributions that allow for positive values of a (the purple, green, and red lines) have a nonzero probability that Earth will undergo a quadratic runaway for two CO_2 doublings (P_{qr}). Since the quadratic model has an infinite sensitivity in these cases, cumulative density functions of these distributions asymptote to values of $1 - P_{qr}$. As before, knowledge of higher-order terms would allow us to estimate what values of ΔT_{4x} this probability should be associated with. Since our uncertainty of these terms is large, the proper assignment of this probability is difficult. It is unclear, for example, how much of this probability should fall above or below $\Delta T_{4x} = 12K$. However, barring large negative higher-order terms, we would expect probabilities such as $P(\Delta T_{4\times} > 8K)$ to be accurate.

Adding a temperature-dependent feedback has a small effect on the less sensitive part of the distribution (i.e., the 5^{th}, 17^{th}, and 50^{th} percentiles), but a strong effect on the more sensitive part: $P(\Delta T_{4\times} > 8K)$ varies from 2% to 13% for Lewis and Curry (2015), from

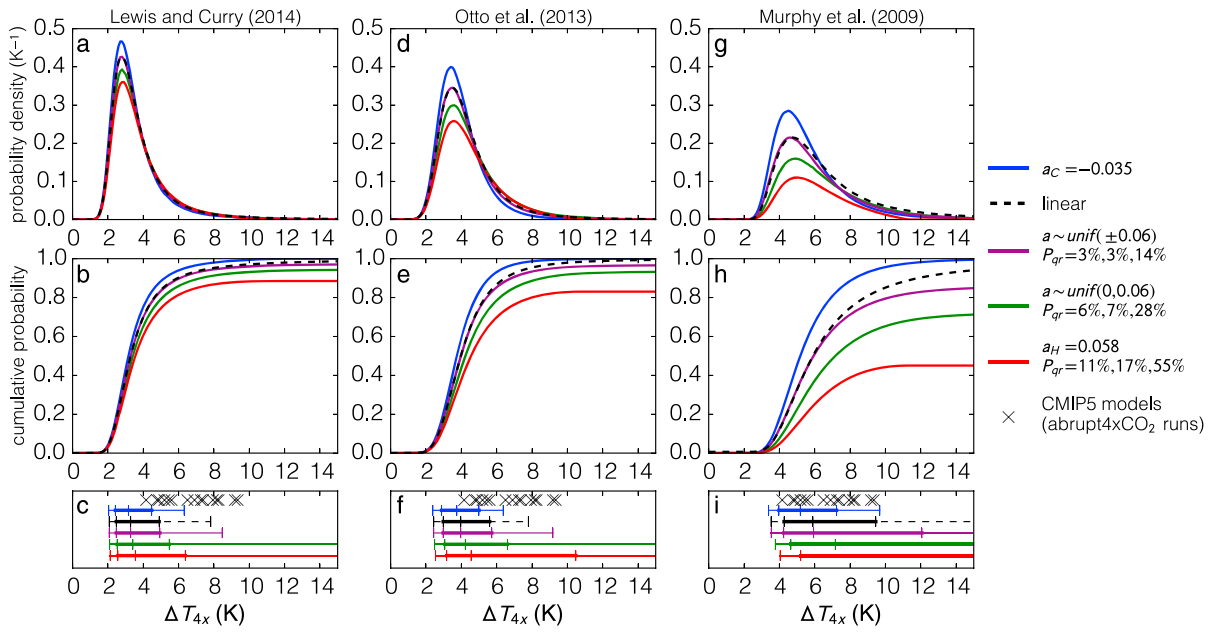

Figure 3.3: Distributions of $\Delta T_{4\times}$, recalculated from three observational studies to account for feedback temperature dependence. Linear distributions (dashed black lines) assume a constant feedback, recreating the results from the three studies, while the nonlinear distributions (colored lines) assume a has the minimum (blue lines) or maximum (red lines) value from our analyzed GCMs, or assume a follows a uniform distribution covering the range $\pm 0.06 W/m^2/K^2$ (purple lines) or 0 to $0.06 W/m^2/K^2$ (green lines). a,d,g) Probability density functions and b,e,h) cumulative density functions (CDFs) for these distributions. Distributions which allow for positive a (red, purple, and green lines) have a nonzero probability of quadratic runaway (P_{qr}), which causes CDFs to asymptote to values less than one. These probabilities are listed for each distribution in the legend, with the values for the three studies separated by commas. c,f,i) Uncertainty ranges, where ticks demarcate 5^{th}-17^{th}-50^{th}-83^{rd}-95^{th} percentiles, as compared with results from the CMIP5 $abrupt4xCO_2$ experiment (black crosses).

1% to 20% for Otto et al. (2013), and from 11% to 61% for Murphy et al. (2009) depending on the assumed value or distribution of a. For all three studies, adding a distribution of a that roughly matches our GCM range (purple line) increases the risk of high warming, despite this distribution having equal likelihood of positive and negative a. If a is likely positive (Meraner et al., 2013), the risk of high warming increases further, as does the risk of quadratic runaway. Accounting for transient nonlinearity would likely further increase the sensitivity of these distributions (Armour et al., 2012; Armour, 2017; Proistosescu and Huybers, 2017; Goodwin, 2018).

3.5 Conclusions

Estimates of the preindustrial climate feedback, λ_{pi}, and the feedback temperature dependence, a, suggest that equilibrium nonlinearity can strongly affect the warming caused by only a few CO_2 doublings, especially when a is positive. When λ_{pi} is small-magnitude negative and a is positive, the quadratic model can experience a runaway. In these cases, the amount of warming experienced by Earth depends sensitively on higher-order terms which are difficult to estimate. As a result, if Earth has values of λ_{pi} and a that cause the quadratic model to run away under a giving forcing, our estimate of the warming response to that forcing may have at best a rough lower bound.

Few GCMs appear to have both small-magnitude negative λ_{pi} and positive a, and therefore few GCMs exhibit the behavior of a extreme equilibrium nonlinearity, e.g. quadratic runaway. Models that experience a quadratic runaway would appear to behave in ways that suggest non-physical model "blow up," and may subsequently be selectively excluded from the published literature. Alternatively, combining small-magnitude negative λ_{pi} and positive a may be unphysical. Future work is need to understand how λ_{pi} and a covary.

Observational estimates of the risk of high climate sensitivity vary significantly based on the assumed feedback temperature dependence. Published studies that assume linearity likely underestimate the risk of high warming. The long tail contains precisely those quadratic runaway cases described above. For those hoping to constrain the risk of high climate sensitivity, understanding equilibrium nonlinearity is just as essential as understanding λ_{pi}.

CHAPTER 4

FEEDBACK TEMPERATURE DEPENDENCE IN A PERTURBED PHYSICS ENSEMBLE

4.1 Introduction

In Chapter 3, we discussed how accounting for feedback temperature dependence can change our probabilistic forecasts of equilibrium climate sensitivity. In this Chapter, we expand on this theme with two focused analyses. In the first, we use offline radiative calculations to determine the specific cause of positive feedback temperature dependence in the ECHAM6 model. We also use this technique to separate out the contributions of feedback temperature dependence, feedback CO_2 dependence, and changes in forcing per doubling of CO_2 to changes in the equilibrium climate sensitivity. In the second, we show results from a perturbed physics ensemble (Stainforth et al., 2005) in which we alter ECHAM6's convective parameters, showing that the presence of positive feedback temperature dependence leads to large differences in sensitivity at high temperatures in spite of only modest changes to the climate feedback itself.

4.2 Partial radiative perturbation: offline radiative feedback analysis

4.2.1 Methods

We ran the ECHAM6.1 atmospheric model with a slab ocean, no sea ice, and prescribed aerosols at T63 resolution. After a spinup of forty years, we ran the model for fifty years with either 1x, 2x, 4x, 8x, or 16x preindustrial CO_2 levels. We then ran the model for an additional three years, outputting instantaneous data every eighteen hours and fifteen minutes. We chose this frequency and number of years because it allowed globally averaged

values of TOA fluxes to match the output from monthly means to within 0.1 Wm^{-2} when we averaged over the entire period.

We take the means of these sets of 1442 time steps of instantaneous snapshots to be representations of $\Omega_{eq}(\Omega_{pi}, d_{CO_2})$ for one through five doublings. For each of these 1442 time steps, we performed offline radiative calculations where the CO_2 level was set at each of the five levels in turn. We then used the mean of each of these sets to estimate $R_{eq}(T, d_{CO_2})$, as given in Figure 4.1. By subtracting $R_{eq}(T, d_{CO_2} + 1)$ from $R_{eq}(T, d_{CO_2})$, we were able to estimate the (unadjusted, top-of-atmosphere) radiative forcing associated with doubling CO_2 as a function of T and d_{CO_2}, as given in Figure 4.2. Finally, by using a finite difference approximation to estimate the partial derivative of $R_{eq}(T, d_{CO_2})$ with respect to T, we were able to estimate λ_{eq} as a function of T and d_{CO_2}, as given by the first panel of Figure 4.3.

We assessed the contribution of the different component feedbacks to the overall feedback temperature dependence by using a version of the partial radiative perturbation (PRP) method (Wetherald and Manabe, 1988). Suppose that each climate Ω can be thought of us a collection of climatic variables, so that $\Omega = [T_{atm}, T, q, cl, ...]$. Suppose we have two climates $\Omega_1 = [T_{atm,1}, T_1, q_1, cl_1, ...]$ and $\Omega_2 = [T_{atm,2}, T_2, q_2, cl_2, ...]$ and we wish to know how radiatively important a change of one of the individual variables is; for the sake of example, suppose it is the humidity q. We can construct two new climates $\Omega_{1,q_2} = [T_{atm,1}, T_1, q_2, cl_1, ...]$ and $\Omega_{2,q_1} = [T_{atm,2}, T_2, q_1, cl_2, ...]$, perform offline radiative calculations to determine $R(\Omega_{1,q_2})$ and $R(\Omega_{2,q_1})$, and then take the average of the forwards perturbation $R(\Omega_{1,q_2}) - R(\Omega_1)$ and the backwards perturbation $R(\Omega_2) - R(\Omega_{2,q_1})$. We used both forwards and backwards calculations to account for correlations that may occur between different components (Colman and McAvaney, 1997).

We performed these perturbations for the pairs $\Omega_{eq}(\Omega_{pi}, d_{CO_2}), \Omega_{eq}(\Omega_{pi}, d_{CO_2} + 1)$ for all $d_{CO_2} \in [0, 1, 2, 3]$, dividing by the temperature difference $\Delta T = T(\Omega_{eq}(\Omega_{pi}, d_{CO_2} + 1)) - \Omega_{eq}(\Omega_{pi}, d_{CO_2})$ to create an estimate of the feedback associated with the perturbed variable. We performed these perturbations five times separately for each fixed CO_2 level in order

to understand how these feedbacks depend on CO_2 as well as temperature. We considered perturbations of water vapor, clouds (with LW and SW components considered separately), and changes in stratospheric and tropospheric temperature. Tropospheric temperature is split further between perturbations that are uniform with surface warming and perturbations associated with the residual to create estimates of the Planck and lapse rate feedbacks respectively.

Because of the computationally intensive nature of these calculations, we first tested to see what frequency of timestep was necessary to capture the pattern of feedbacks seen in the full 1442 time steps. We considered factors of 1442, and it was decided that taking a frequency of every seven time steps (206 in total) was sufficient, as it produced errors less than 0.1 $Wm^{-2}K^{-1}$.

Estimates of these component feedbacks using PRP at this lower frequency are given in Figure 4.3. The feedback strengths relevant to the trajectory seen in our experiments are given in Figure 4.4 (the values traced by the black line in Figure 4.3). The sum of the component feedbacks is within 0.2 $Wm^{-2}K^{-1}$ of the true value (top left panel of Figure 4.3), suggesting that cross terms are relatively small.

4.2.2 Results

Figure 4.1 shows an estimate of $R_{eq}(T, d_{CO_2})$. Let $T_{d_{CO_2}} \equiv T(\Omega_{eq}(\Omega_{pi}, d_{CO_2}))$. The black line traces the values $T_{d_{CO_2}}$ for the five forcing levels considered. The equilibrium climate sensitivities for each consecutive doubling (that is, $T_{d_{CO_2}+1} - T_{d_{CO_2}}$ for $d_{CO_2} \in [0, 1, 2, 3]$) are 2.7, 3.4, 5.6, and 9.3 K, suggesting a large increase in sensitivity. The line traces the zero contour of $R_{eq}(T, d_{CO_2})$, suggesting that these are in fact the equilibrium values of warming in response to their respective forcings, and that we are using sufficient years and time steps to characterize the GCM's equilibrium climates (Ω_{eq}s).

Figure 4.2 shows the TOA forcing $F(\Omega_{eq}(T), d_{CO_2} + 1) - F(\Omega_{eq}(T), d_{CO_2})$, that is the forcing associated with doubling CO_2 starting at a given pair of T and d_{CO_2}. The black

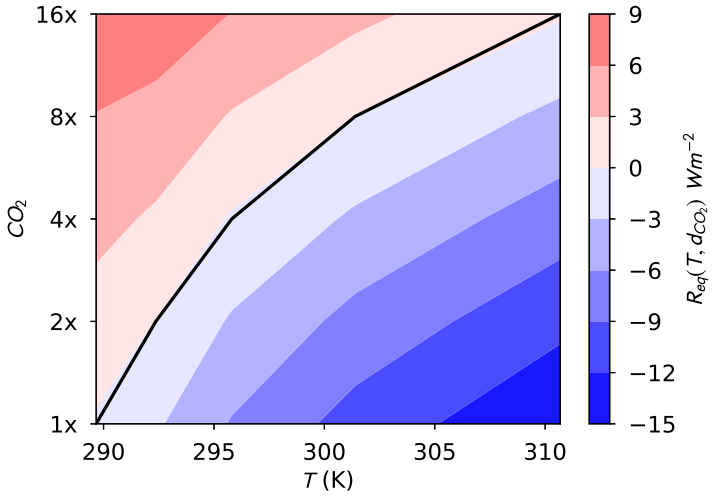

Figure 4.1: Estimates of $R_{eq}(T, d_{CO_2})$ for a slab ocean version of ECHAM6.1, made using offline calculations performed on roughly 18-hourly instantaneous snapshots across three years of data after the model has been brought to equilibrium under 1x, 2x, 4x, 8x, and 16x preindustrial CO_2 levels. The five equilibrium states are represented by the globally averaged surface temperatures along the x-axis. Offline calculations were performed for all CO_2 levels, showing the forcing that would occur if each equilibrium climate were changed to the y-axis CO_2 level. The black line shows the globally averaged surface temperature at the end of these runs. This traces the zero-contour of $R_{eq}(T, d_{CO_2})$, suggesting that R_{eq} does determine the equilibrium response. Equilibrium climate sensitivity increases over threefold between the first and last doubling.

line once more shows the trajectory of the equilibrium response. Forcing increases along this trajectory from about 2 Wm^{-2} to almost 4 Wm^{-2}, a twofold increase, but since climate sensitivity increases by more than a factor of three, the change in forcing alone cannot account for the sensitivity increase. It appears that much of this increase in forcing is balanced by the negative feedback temperature dependence of the stratospheric temperature feedback (middle left panel in Figure 4.3, blue line in Figure 4.4), so that the forcing found by taking the y-intercept of R_{inst} vs. T_{inst} (the "Gregory" estimate of forcing, Gregory et al., 2004) and which includes this stratospheric temperature change as a forcing adjustment is more proportional to CO_2 doubling.

It is worth noting that most of the increase in forcing does not appear to be to due to super-logarithmic potency of CO_2 on its own (i.e., the addition of a new optically thick

band at sufficiently high CO_2 concentrations), as this would create a stronger dependence of the top-of-atmosphere forcing on the CO_2 concentration. Instead, it appears that the climates associated with a warmer world have a larger forcing, due in part to the larger difference between surface temperature and the temperature at the effective atmospheric emission level, and in part to bands in the near-IR becoming more important. There does appear to be more of a dependence of top-of-atmosphere forcing on the CO_2 concentration at higher temperatures, perhaps related to these bands not yet being optically thick.

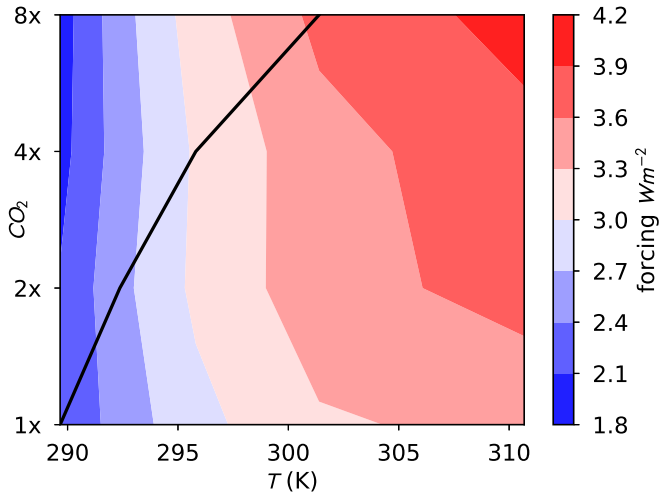

Figure 4.2: By taking $R_{eq}(T, d_{CO_2}+1) - R_{eq}(T, d_{CO_2})$, we can estimate how the forcing due to doubling CO_2 changes as a function of the equilibrium climate represented by T and the background CO_2 level. The forcing gets stronger in a warmer world, likely due to the greater difference between surface and tropopause temperatures and increased radiation in the near-IR, where there are additional CO_2 bands. There is little evidence for CO_2 dependence at lower temperatures. The black line shows the trajectory taken by the model's equilibrium response, showing that the forcing increases by a factor of two, explaining some but not all of the climate sensitivity increase.

The top left panel of Figure 4.3 shows the net feedback as a function of T and d_{CO_2}, estimated as follows:

$$\lambda((T_{d_{CO_2}+1} + T_{d_{CO_2}})/2, d'_{CO_2}) = \frac{R_{eq}(T_{d_{CO_2}+1}, d'_{CO_2}) - R_{eq}(T_{d_{CO_2}}, d'_{CO_2})}{T_{d_{CO_2}+1} - T_{d_{CO_2}}} \tag{4.1}$$

The figure suggests that there is an increase in feedback both with CO_2 level (vertical gradient) and, except for the few coldest degrees, with temperature. This same pattern is seen in the sum of the individual feedback components. The largest contribution to the CO_2 dependence appears to come from the Planck feedback, while the largest contribution to the temperature dependence appears to come from the water vapor feedback. Colder than 295 K, the temperature dependence appears to be more driven by a decreasing stratospheric temperature feedback, which could also be thought of as part of the CO_2 forcing. This contributes to the change in sign of the feedback temperature dependence, as does the change in sign of feedback temperature dependence of the SW cloud feedback at this temperature.

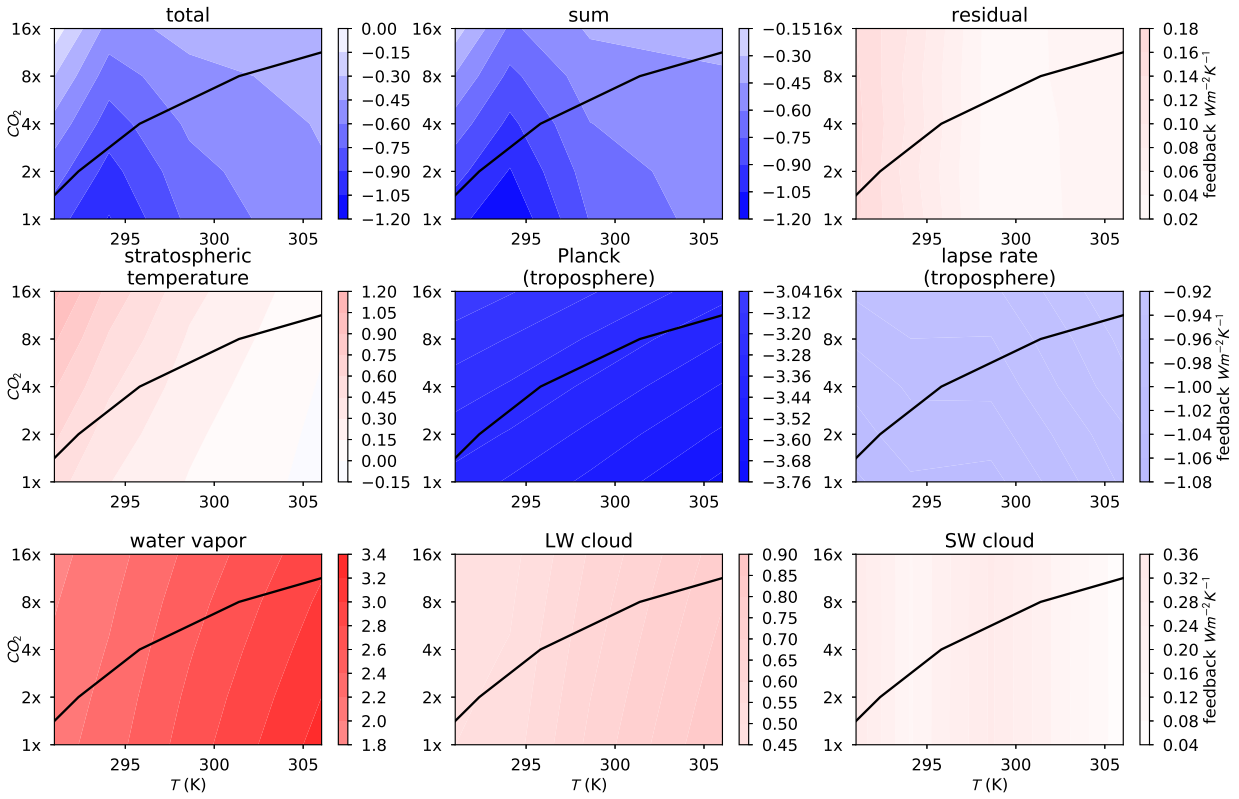

Figure 4.3: ECHAM6 feedbacks as a function of temperature and CO_2 estimated using offline radiation calculations. Upper left panel shows the total feedback; all panels in the lower two rows come from averaging the results of forwards and backwards partial radiative perturbations of the given climatic variable (see text for details). The upper middle panel shows the sum of the different individual feedbacks, and the upper right panel shows the residual. The black line shows the path taken by the equilibrium response of the model.

We can see which of these component feedbacks makes a larger contribution to the overall change in feedback strength by seeing what values are traced out by the line representing the equilibrium response (black lines in Figure 4.3). These component feedback values are given as lines in Figure 4.4. The total feedback initially gets more negative, but for most of the range gets more positive, primarily due to the water vapor feedback, in keeping with arguments from previous work (Meraner et al., 2013). The Planck feedback also contributes to this increase, with some offsetting decrease from the stratospheric temperature feedback. This positive change leads to larger and larger sensitivity increases due to the inverse relationship between these values. The SW cloud feedback appears to be responsible for the change in sign.

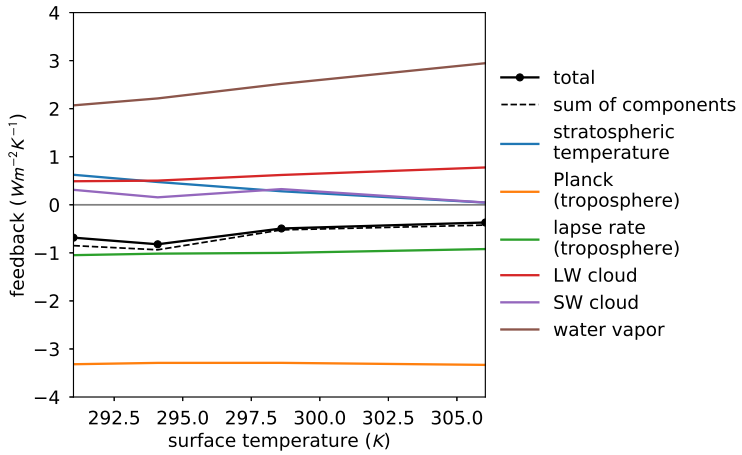

Figure 4.4: Feedback values along the black line shown in Figure 4.3. The total feedback (black line) increases with temperature over most of the range, primarily due to the water vapor feedback (brown line). The overall feedback appears to be influenced also by the change in sign of the SW cloud feedback temperature dependence (purple line). The negative feedback temperature dependence of the stratospheric temperature feedback (blue line) offsets the increase in the top-of-atmosphere forcing in Figure 4.2, so that the "adjusted" forcing is roughly constant with doubling.

We note that this analysis was performed using TOA fluxes. In future work, we will explore how accounting for the stratospheric feedback as part of the forcing changes our analysis, and also if part of the increase in the water vapor feedback comes from a spurious stratospheric increase due to prescribed ozone not tracking with a rising tropopause under

CO_2 forcing (Nowack et al., 2018).

We have demonstrated that ECHAM6.1 does have positive feedback temperature dependence, in addition to feedback CO_2 dependence and a change in forcing with warming. As discussed in the previous chapter, while the latter two can increase the climate sensitivity, the former can lead to much larger increases in sensitivity and even bifurcations. We now explore the consequences of this feedback temperature dependence in a perturbed physics ensemble.

4.3 Perturbed Physics Ensemble

4.3.1 Methods

We performed the same set of experiments (forty years of spinup followed by five fifty-year long abrupt CO_2 experiments), but with four parameters associated with atmospheric convection altered. The parameters, their perturbations, and their combinations are chosen to correspond in part to the perturbed physics ensemble performed in Tomassini et al. (2015), and are given in Table 4.3.1. *entrpen* has units of m^{-1} and represents the fractional organized entrainment/detrainment rate of deep convection. *entrscv* is the same, but for shallow convection. $cmfctop$ represents the fraction of convective mass flux that overshoots the top of the cloud, being deposited at the level above. *cprcon* represents the conversion rate of cloud water to rain. Most of the configurations represent simple increases and decreases of these values, while the last five correspond to scenarios from (Tomassini et al., 2015). $\Omega_{eq}(\Omega_{pi}, d_{CO_2})$ for each run is defined as the mean of the last five years.

4.3.2 Results

We calculate the equilibrium climate sensitivity at each forcing level as above by taking $T_{d_{CO_2}+1} - T_{d_{CO_2}}$. We plot the resulting sensitivities as a function of their initial temperature $T_{d_{CO_2}}$ in the left panel of Figure 4.5. Sensitivities tend to increase with initial

Table 4.1: Different parameter configurations used in a perturbed physics ensemble of ECHAM6.1. Definitions of four convective parameters *entrpen*, *entrscv*, *cmfctop*, and *cprcon* are given in the text. Ranges of reasonable values and the five bottom parameter configurations are based on the experiments performed in Tomassini et al. (2015) (our "strong conversion" is their "large autoconversion").

parameter configuration	entrpen (m^{-1})	entrscv (m^{-1})	cmfctop	cprcon
default	1×10^{-4}	3×10^{-4}	0.21	2×10^{-4}
entrpen low	2×10^{-5}	3×10^{-4}	0.21	2×10^{-4}
entrpen high	2×10^{-4}	3×10^{-4}	0.21	2×10^{-4}
entrscv low	1×10^{-4}	1.5×10^{-4}	0.21	2×10^{-4}
entrscv high	1×10^{-4}	3×10^{-3}	0.21	2×10^{-4}
cmfctop low	1×10^{-4}	3×10^{-4}	0.11	2×10^{-4}
cmfctop high	1×10^{-4}	3×10^{-4}	0.31	2×10^{-4}
cprcon low	1×10^{-4}	3×10^{-4}	0.21	2×10^{-5}
cprcon high	1×10^{-4}	3×10^{-4}	0.21	2×10^{-3}
weak deep mixing	1×10^{-3}	3×10^{-4}	0.21	2×10^{-3}
strong deep mixing	1×10^{-5}	3×10^{-4}	0.21	2×10^{-5}
weak cloud top mixing	1×10^{-4}	3×10^{-4}	0.11	2×10^{-3}
strong cloud top mixing	1×10^{-4}	3×10^{-4}	0.31	2×10^{-5}
strong conversion	1×10^{-3}	3×10^{-4}	0.31	2×10^{-5}

temperature, and the variance of sensitivity increases substantially with initial temperature. This is explained by considering the right panel of Figure 4.5, which shows estimates of the feedback strength of each of these parameter configurations as a function of temperature. These were estimated as above, except that instead of using offline calculations to estimate forcings (and in turn to estimate R_{eq}), the Gregory method defined in Section 4.2.2 was used (Gregory et al., 2004). As a result, these forcing estimates and their accompanying feedbacks each have a larger magnitude, as the positive stratospheric feedback (see middle right panel of Figure 4.3) is accounted as a forcing adjustment rather than a feedback, making the forcing more positive and the feedback more negative. This might also explain why the default feedback has positive feedback temperature dependence even at lower temperatures in Figure 4.5, as the stratospheric feedback has negative feedback temperature dependence.

With the exception of *entrscv high*, all parameter configurations have effectively the same $\lambda_{eq}(T)$, with a range of about 0.5 $Wm^{-2}K^{-1}$ around the black dashed line in right panel of

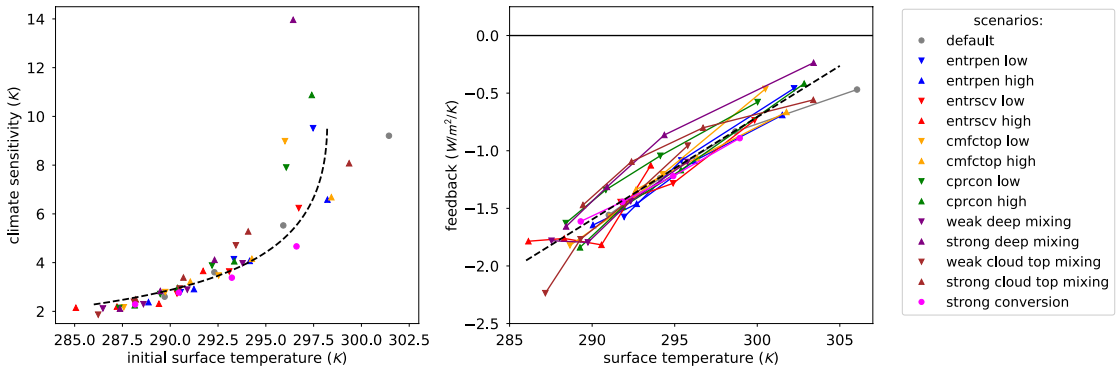

Figure 4.5: Results from an ECHAM6.1 perturbed physics ensemble in which each configuration is equilibrated at five different CO_2 levels, each a doubling of the previous level. The left panel shows all the equilibrium climate sensitivities (differences in surface temperature between doublings of CO_2) from all models. Sensitivity and variance in sensitivity both increase with temperature. The right panel gives estimates of $\lambda_{eq}(T)$ for each model, where the forcings needed to estimate $R_{eq}(T, d_{CO_2})$ for each model are calculated using the Gregory method. These feedbacks differ from those in Section 4.1 because stratospheric warming is considered an adjustment to forcing here. The dashed line in the right panel is made by taking the linear regression of all feedback values, and the dashed line in the left panel is the resulting estimates of sensitivity. As the range of probable feedbacks approaches zero, the range of probable sensitivities greatly increases due to the inverse relation between them.

Figure 4.5, which is the least squares linear regression of the feedbacks. Decomposing these feedbacks into LW/SW and clear sky/cloud components (Figure 4.6) suggests that with the exception of *strong conversion*, all the component feedbacks are quite similar among models (and track with the values seen in Figure 4.4). If we take the dashed-line average feedback from the right panel of Figure 4.5 and use it to estimate sensitivities (dashed line in the left panel of Figure 4.5), we see that it matches the dramatic increase in sensitivity at high temperatures. The increase in variance is then explained as the inverse of the same relatively small range in feedback moved closer and closer to 0.

4.4 Discussion

Figure 4.5 shows how even small variance in feedback values (or in the convective parameters that determine them), combined with water vapor-induced positive feedback temperature

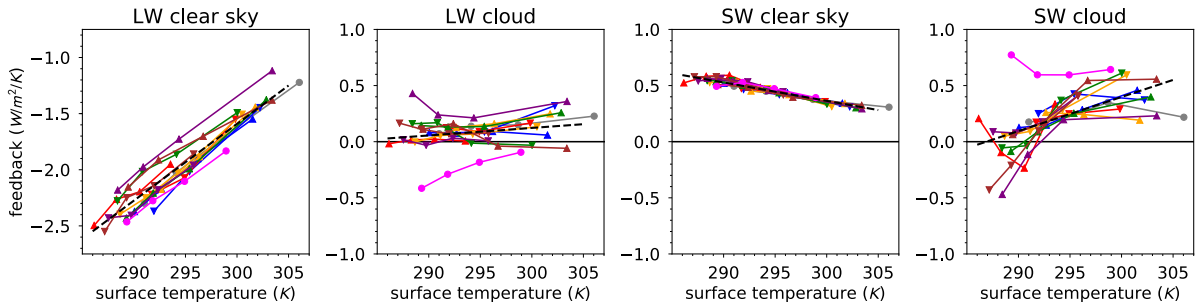

Figure 4.6: Individual flux feedbacks estimated as in the right panel of Figure 4.5. Feedbacks show mostly the same qualitative behavior across parameter configurations, except for *strong conversion*, and are generally in line with the feedbacks in Figure 4.4, with the strongest positive contribution coming from the LW clear sky feedback, presumably due to the water vapor feedback. The SW cloud feedback also changes sign, as in Figure 4.4.

dependence (Figure 4.4), can cause a climate sensitivity range of almost ten degrees at high forcings. This underscores the challenge positive feedback temperature dependence presents to forecasts of warming. Narrowing our physical understanding of convective processes or the feedbacks seen across the historical record may not eliminate the fat tail of high sensitivity probabilities. Future work should explore the nonlinearity of the water vapor feedback in detail considering, for example, whether models universally have positive water vapor feedback temperature dependence (as was found, for example, in Colman et al., 1997), and whether this effect is physical or the result of spurious stratospheric water vapor increases (Nowack et al., 2018). While other studies that find large increases in GCMs' sensitivity under warming attribute this increase to cloud feedbacks (Caballero and Huber, 2013; Popp, 2014), these feedbacks may act in concert with the water vapor feedback, so that cloud feedbacks create a spread around a typically positive feedback temperature dependence, increasing it in some models and cancelling it in others.

CHAPTER 5

SPATIAL FEEDBACKS FROM INTERANNUAL VARIABILITY USING MULTIPLE REGRESSION

5.1 Introduction

The preceding chapters show the important effect feedback temperature dependence can have on assessments of the risk of high climate sensitivity. We now turn to the pattern effect. As we mentioned in the preceding chapters, we define the *climate feedback* (λ) as the dependence of the globally averaged net top-of-atmosphere (TOA) radiative flux (R, the rate of planetary energy gain per unit area) on the globally averaged surface temperature (T); λ is defined so that a negative λ implies a stable climate. λ is the same as the climate feedback parameter defined in Gregory et al. (2004) with the exception that the sign convention is reversed; it is also the same as the thermal damping rate (Dessler, 2012). The climate feedback is a key parameter in determining the response of the Earth to a given scenario of radiative forcing, with less negative values of λ implying a larger climate sensitivity and warming response (Hansen et al., 1985).

If the climate feedback were a constant, then estimates of λ made using observations of the past would also hold in the future, allowing us to forecast warming. In particular, if we have observations (or reconstructions) of time series of R and T, and we remove changes in R and T due to radiative forcing, then the slope of the regression of the remaining year-to-year variation (the "interannual variability") in R against T should give an estimate of λ (Forster and Gregory, 2006; Murphy et al., 2009; Dessler, 2010), assuming that internal (non-T related) variations in R do not bias this regression.

In global climate models (GCMs), however, λ changes with time as the models respond to radiative forcing (e.g., Murphy, 1995; Watterson, 2000; Senior and Mitchell, 2000; Armour et al., 2012; Andrews et al., 2014). Further, these models exhibit different feedbacks when responding to forcing than would be inferred from interannual variability (Dessler, 2012;

Colman and Hanson, 2017). As a result, our observational estimates of future warming are significantly smaller when we assume a constant feedback (Otto et al., 2013; Lewis and Curry, 2015) than when we assume the climate feedback changes with time at a rate similar to that seen in GCMs (Armour, 2017; Proistosescu and Huybers, 2017; Goodwin, 2018). Ideally, we would estimate future warming by directly accounting for the physical mechanisms that cause the feedback to change with time, of which there are two main types, which we recap here:

- *Pattern effect.* The climate feedback can depend on the spatial pattern of warming (Armour et al., 2012; Andrews et al., 2014). The processes that determine the strength of the climate feedback, such as increases in blackbody radiation, the warming and moistening of the atmosphere, the change in position, magnitude, thickness, and composition of clouds, and the melting of snow and ice, are all spatially heterogeneous. As a result, surface warming in different regions elicits quite different radiative responses. Since only the proportional response matters for calculating the climate feedback, even modest forcings can cause changes in the climate feedback over time if the spatial patterns of warming change in response to forcing. This pattern effect appears to occur in most GCMs under anthropogenic forcing (Andrews et al., 2014).

- *Feedback temperature dependence.* The processes that determine the strength of the climate feedback may change when the surface temperature is higher (Colman and McAvaney, 2009; Meraner et al., 2013; Bloch-Johnson et al., 2015). For example, the water vapor feedback is expected to strengthen in a warmer world (Meraner et al., 2013) while the ice albedo feedback would decrease once most sea ice was melted (Colman and McAvaney, 2009). A significant feedback temperature dependence would cause the equilibrium warming response to be a nonlinear function of radiative forcing, while a climate with a pattern effect could still be linear. Such a nonlinearity is seen in some paleoclimate proxies (Shaffer et al., 2016; Köhler et al., 2017) and climate models (Colman and McAvaney, 2009; Jonko et al., 2012; Caballero and Huber, 2013; Meraner

et al., 2013). However, this effect does not appear to significantly affect the sensitivity of many models under common anthropogenic forcing scenarios (i.e., two doublings of CO_2), so that its primary influence is on the upper tail of forecast projections (Chapters 2 and 3).

The pattern effect affects the climate feedback generally, while feedback temperature dependence only comes into effect under high levels of warming. These two effects are not mutually exclusive, as the tendency of the pattern effect to increase the climate's sensitivity to forcing over time (Armour, 2017) makes the effects of feedback temperature dependence stronger, likely increasing the risk of extreme warming (Bloch-Johnson et al., 2015). However, since the pattern effect is far more common in simulations of anthropogenic global warming and is central to relating the climate feedback exhibited by interannual variability to that found in the forced response (Colman and Hanson, 2017), we focus on it in this study, and only discuss feedback temperature dependence briefly in Section 5.5.

While changes to the spatial pattern of warming may cause the climate feedback to vary with time, the spatial structure of these feedbacks may remain fixed, allowing us to estimate the values these spatial feedbacks had in the past to forecast future warming. The simplest case would be if feedbacks only acted locally, so that R_i (the net TOA radiative flux in region i) depended purely on T_i (the surface temperature in that region). In this case, one could simply regress each R_i against each T_i, generating a spatial collection of λ_i (Boer and Yu, 2003b; Crook et al., 2011; Armour et al., 2012; Feldl and Roe, 2013; Brown et al., 2015; Trenberth et al., 2015). However, recent work suggests that surface warming at a given location can have large nonlocal impacts on TOA fluxes (Rugenstein et al., 2016a; Zhou et al., 2017; Po-Chedley et al., 2018). This is illustrated in the top row of Figure 5.1 (taken from Zhou et al., 2017) which shows the change in the average spatial pattern of net TOA flux due to changes in cloudiness caused by warming the circled areas in a fixed-SST experiment. While warming in the extratropics or in the subsiding tropics causes primarily a local response, warming in the convecting tropics reverberates throughout the tropics and

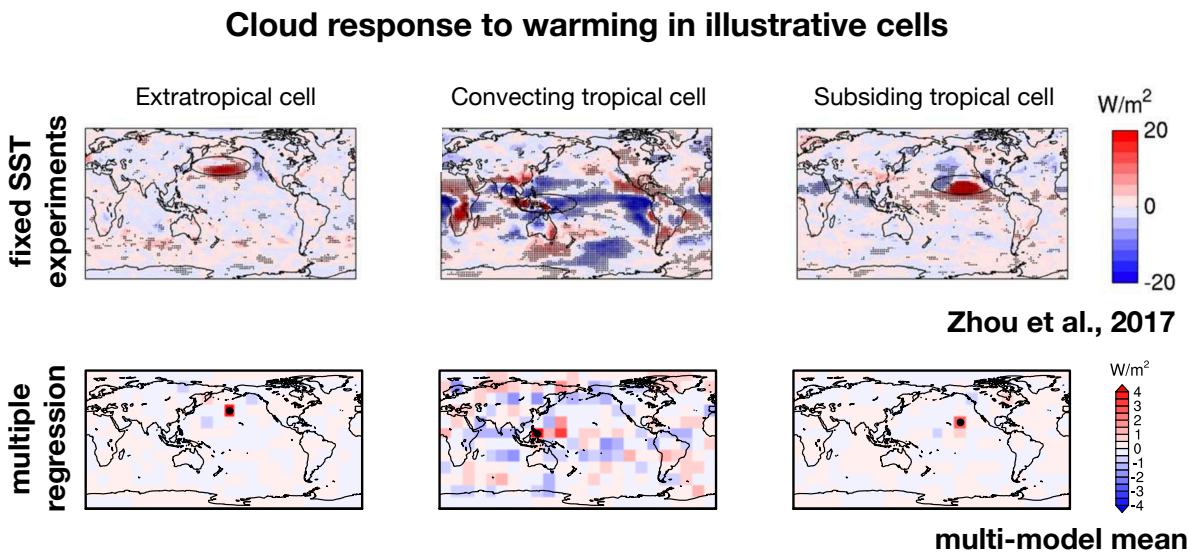

Figure 5.1: Cloud-induced changes in the net top-of-atmosphere (TOA) radiative flux from warming in illustrative regions indicated by ovals (top row) or dotted grid cells (bottom row). Cloud response in the top row, reproduced with permission from Zhou et al. (2017), is estimated by running CAM5.3 with a fixed-SST pattern consisting of an isolated warming perturbation in the indicated area. Cloud response in the bottom row is estimated using the multiple regression method proposed in Section 5.2 and is the mean response across six general circulation models. The two methods agree qualitatively: surface warming in regions which inhibit propagation of a warming perturbation (such as tropical subsiding regions with inversion-capped boundary layers and regions outside the tropics with high Coriolis force) tend to have predominately local responses, while surface warming in tropical convecting regions propagates throughout the free tropical and extratropical troposphere. Note that the bottom row scale bar covers a fraction of the range of the top row because surface warming in the bottom row occurs over a smaller region.

beyond, impacting TOA fluxes broadly.

However, Zhou et al. (2017) do find that the spatial feedbacks estimated using these fixed-SST experiments explain some of their model's response to forcing, in that some aspects of the TOA flux response can be approximated by a linear combination of these spatial feedbacks weighted by the pattern of surface warming from the forced response. This suggests that spatial feedbacks may be constant across different patterns of surface temperature change – including those patterns associated with interannual variability, which unlike the fixed-SST patterns from Zhou et al. (2017) occur naturally in observations of the real Earth.

In this paper, we propose a method for estimating spatial feedbacks from interannual vari-

ability by using multiple regression of TOA fluxes against the full spatial pattern of surface temperature. While Liu et al. (2008) used a similar technique to study surface feedbacks, we are not aware of previous studies applying this method to TOA fluxes. We demonstrate that using this multiple regression (MR) method on an ensemble of GCMs accurately captures many features of that ensemble's responses to CO_2 forcing. The success of the MR method suggests that it may be possible to forecast future warming from observations despite the time-varying nature of the climate feedback.

In Section 5.2, we explain the MR method and show results from applying it to an ensemble of GCMs. We demonstrate that this method, in conjunction with the spatial patterns of warming from a forced simulation, can recreate the time-varying climate feedback of these models seen in these forced simulations, as well as the spatial pattern of change in R. In Section 5.3, we explore the physical insight the MR method can provide into the climate feedback. In Section 5.4, we see how the MR method compares with existing methods for estimating spatial feedbacks. In Section 5.5, we discuss the challenges to applying the MR method to observations as opposed to GCM control simulations. Finally, in Section 5.6, we summarize our conclusions.

5.2 The multiple regression (MR) method

Figure 5.2 shows plots of anomalies of R (R') against anomalies of T (T') for preindustrial control simulations of six coupled atmosphere-ocean global climate models (AOGCMs). The runs are drawn from a new archive of thousand-plus year long AOGCM simulations called LongRunMIP (Rugenstein et al., forthcoming). If the climate feedback were constant under all conditions, we could regress these values of R' against T' to estimate λ. Instead, to account for the variation of feedbacks in time and space, we define $\vec{T}'(t)$ to be a vector time series where each entry $T'_i(t)$ in $\vec{T}'(t)$ is the surface temperature anomaly time series of the i^{th} region of the model (in other words, the columns of the maps in Figure 5.2 are stacked on top of each other to produce a one dimensional object). We then can define two versions

of our multiple regression (MR) method (for more details, see Appendix Section 7.2.1):

- *Global version.* To ascertain the contribution of warming in each location to the change in the globally averaged TOA radiative flux (R) and thus to the global climate feedback, take the multiple regression of the globally averaged value R against all values of $\vec{T}'(t)$ simultaneously using the statistical model

$$R'_{pi}(t) = \vec{\lambda} \cdot \vec{T}'_{pi}(t) + \epsilon(t) \tag{5.1}$$

  where the pi subscript indicates that these are time series of preindustrial runs, $\vec{\lambda}$ is a vector whose i^{th} entry (λ_i) is the dependence of R directly on the surface temperature in region i (T_i), and $\epsilon(t)$ is noise. Mathematically, the λ_i derived using this method are equivalent to first taking the multiple regression of R' and T'_i against all T'_j for which $j \neq i$, and then regressing the resulting residual of R' against the residual of T', with the resulting regression coefficient being λ_i. This ensures that, given sufficient information, λ_i is not contaminated by covariance between temperatures in region i and temperatures elsewhere. This effect is explored more fully in Section 5.4 (see Figure 5.11 in particular).

- *Full version.* To ascertain the spatial dependence of TOA fluxes everywhere on surface temperature everywhere, take the multiple regression of each region's R'_i against all values of $\vec{T}'$ simultaneously, which can be compactly written as:

$$\vec{R}'_{pi}(t) = \Lambda \vec{T}'_{pi}(t) + \vec{\epsilon}(t) \tag{5.2}$$

  where Λ is a square matrix with as many columns and rows as regions, and in which each entry $\lambda_{i,j}$ represents the dependence R_i on T_j. Once more, multiple regression removes the influence of covariance with values of T_j elsewhere.

Each of these regressions requires at least as many years of data as regions in order to

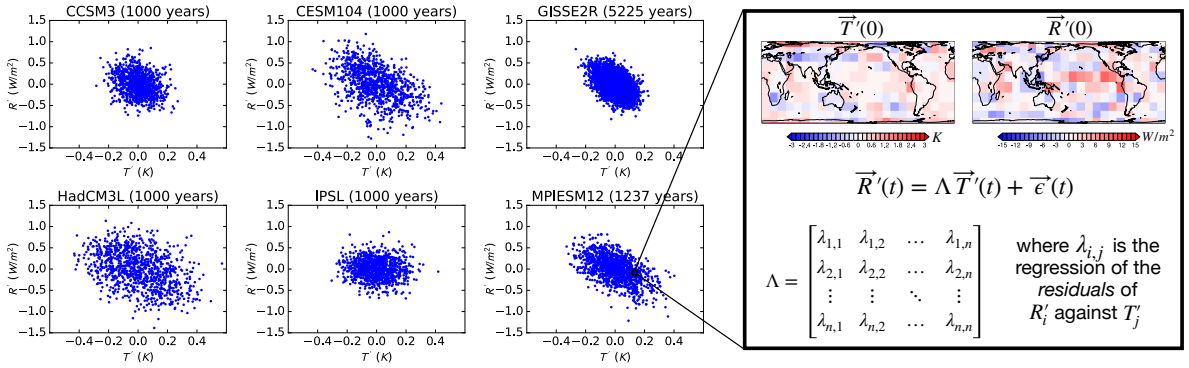

Figure 5.2: Schematic outlining of the multiple regression method for estimating spatial feedbacks from interannual variability. The six panels on the left show annual anomalies of the globally averaged net TOA radiative flux, R', against the globally averaged surface temperature, T', from millennial-length preindustrial simulations of six coupled atmosphere-ocean global climate models. The time series of the full spatial patterns of these variables are $\vec{T}'(t)$ and $\vec{R}'(t)$ respectively. For each grid cell i, we take the multiple regression of R'_i against the full $\vec{T}'$, producing a matrix of spatial feedbacks $\lambda_{i,j}$. Each $\lambda_{i,j}$ represents the dependence of R_i on T_j, with effects due to covariance with surface temperature elsewhere removed. This regression requires more years than grid cells, and so simulations are regridded to $15° \times 15°$ ($7.5° \times 7.5°$ for GISSE2R). This regression is performed separately for each month of the year, and also for each component flux (e.g., LW clear-sky).

not be underdetermined. Since all of our simulations have many more grid cells than years, we coarsened our grids to $15° \times 15°$ resolution, resulting in 288 regions. Since we have more years of data for the model GISSE2R, we regridded it instead to a $7.5° \times 7.5°$ resolution, which does not qualitatively effect our results but improves their accuracy. We discuss the choice of grid further in Section 5.5. We perform these regressions separately for each month of the year (first regressing all Januaries, then all Februaries, etc.), to capture the different local and nonlocal relationships that occur as circulations shift seasonally, resulting in twelve different sets of spatial feedbacks. Finally, we can perform this same analysis not just for the net TOA radiative flux, but for the individual components of this flux: the longwave (LW) and shortwave (SW) clear-sky fluxes and cloud radiative effects. Although this method will falsely apportion parts of the change in flux between the clear-sky and cloud feedbacks due to masking effects (Soden et al., 2004, 2008), we are constrained by the data available for

this analysis, and recent work suggests the use of cloud radiative effect may not strongly affect qualitative arguments about the spatial pattern of cloud feedbacks (Rugenstein et al., 2016a).

In theory, the full version of the MR method should contain the information of the global version, in that the area-weighted average of the i^{th} column of Λ should give the i^{th} entry in $\vec{\lambda}$. However, regridding can change global averages, causing slight discrepancies. We therefore use Equation 5.1 for global feedbacks and changes in R and Equation 5.2 for full spatial feedbacks. Both of these definitions differ from the "local contribution" and local feedback definitions used elsewhere in the literature, which are considered in detail in Section 5.4.

To get a sense of the MR method, we can compare the multi-model mean cloud response to warming in the illustrative regions of Figure 5.1 estimated by this method to the fixed-SST results from Zhou et al. (2017). The MR method predicts qualitatively similar cloud feedbacks to the fixed-SST method, namely mostly local positive cloud feedbacks in the extratropics and subsiding tropics, but significant (and often negative) nonlocal feedbacks in the convecting tropics. The fixed-SST approach includes some spurious cloud feedbacks due to land warming induced by the prescribed SST pattern (Zhou et al., 2017), while the MR method removes these cloud feedbacks by accounting for covariance of surface temperature with other regions. Note that the MR method allows us to estimate feedbacks caused by warming the land surface directly, which the fixed-SST approach does not.

We can more fully test the accuracy of this method by seeing how well these spatial feedbacks recreate the feedbacks of a forced simulation. The black dots in Figure 5.3 show the annually averaged values of R' vs. T' for *abrupt4x* simulations of each respective model (simulations in which initial conditions are drawn from a preindustrial control simulation, but in which the CO_2 concentration is set to be four times the preindustrial value, resulting in an abrupt initial forcing and resultant warming). The gold dots show estimates of R' for these runs ($\hat{R}'$) made by multiplying the spatial feedbacks $\vec{\lambda}$ estimated from the preindustrial control simulations by the spatial pattern of warming from the *abrupt4x* simulation (where

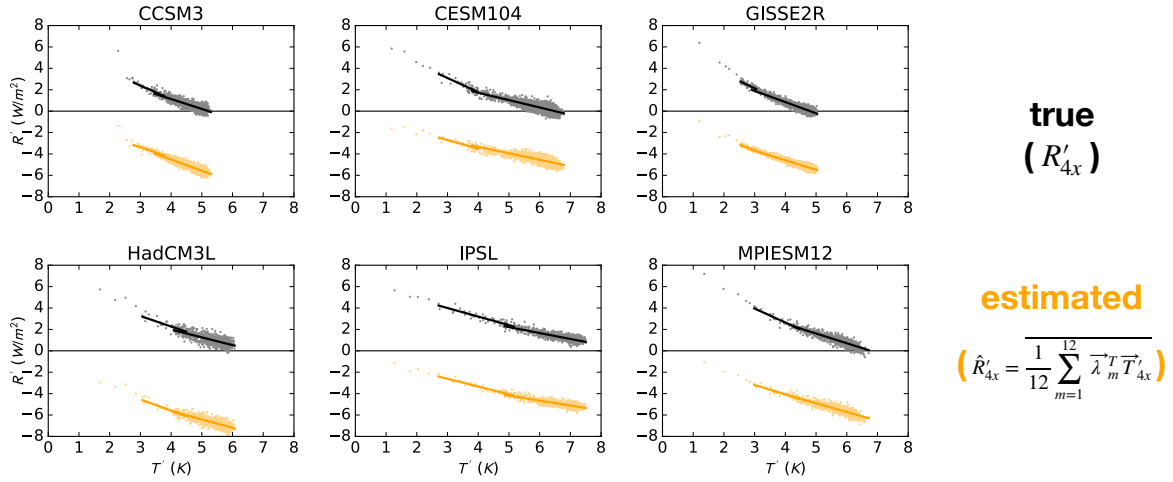

Figure 5.3: Plots of annual averages of R' vs. T' for *abrupt4x* simulations of the six models from Figure 5.2 (black dots), and plots of estimated values of R' ($\hat{R}'$) created by multiplying time series of the spatial pattern of surface temperature from these simulations with the models' respective spatial feedbacks (gold dots). Slopes of the regressions of R' against T' (black lines) and $\hat{R}'$ against T' (gold lines) give the true and estimated climate feedback respectively. Because our method does not estimate radiative forcing (or forcing adjustments), there is an offset, mostly constant after year 5, between the true and estimated values of R'. Because the climate feedback changes with time, we take two regressions, one for an early period (years 6 to 50, to leave out initial adjustments to forcing) and one for a late period (51 to end). True and estimated slopes (feedbacks) are similar across all models and periods.

this multiplication is done separately for each month of the year, and the resulting time series of $\hat{R}'$ are then annually averaged). For more details, see Appendix Section 7.2.2.

The resulting estimates $\hat{R}'$ will only include changes in R' caused by changes in the surface temperature, and therefore will not account for radiative forcing. This causes a mostly constant and positive offset between the true and estimated values. We can split the runs into an early period (years 6 to 50), leaving out the first few years which contain ongoing adjustments to radiative forcing (Sherwood et al., 2014; Chung and Soden, 2015; Kamae et al., 2015; Rugenstein et al., 2016b), and a late period (years 51 onwards). Slopes of R' vs. T' give the climate feedback λ, and so the similarity between each GCM's true and estimated slopes of R' vs. T' (black and gold lines, respectively) suggests that the MR

method is able to recreate this feedback.

Figure 5.4 assesses this similarity by showing the estimated global feedback (the slope of the gold lines in Figure 5.3) vs. the true values of these feedbacks (the slope of the black lines in Figure 5.3) for the early (left panel) and late periods (right panel) and for both the net feedbacks and the individual TOA flux component feedbacks for all six GCMs. 5%-95% confidence intervals for the true feedback values are calculated assuming independent and identically distributed normal noise, while the confidence intervals for the estimated feedback values are calculated using bootstrapping techniques that capture the uncertainty in the spatial feedbacks themselves; for more details, see Appendix Section 7.2.2. For most feedbacks, the estimates are close to their true values. The "MR" columns of Table 5.1 gives estimates for the root mean square error of these estimated feedbacks (see Appendix Section 7.2.3 for error definitions). For the late period, the MR method error for the net feedback is only 0.11 $Wm^{-2}K^{-1}$. Relatively low errors hold for the component feedbacks as well, with most estimates being within 0.3 $Wm^{-2}K^{-1}$ of the true answer. As seen in the third row of Table 5.1, our method also estimates the change in the global feedback between the early and late period well compared to the existing methods discussed in Section 5.4.

We can also test how well the MR method captures the change in the spatial pattern of TOA fluxes in response to forcing. The left panel of Figure 5.5 shows how the multi-model mean of the average value of $\vec{R}$ changes between the first and second halves of the early period of the *abrupt4x* simulation, normalized by the change in the globally averaged surface temperature T ($\Delta\vec{\tilde{R}}$; see Appendix Section 7.2.2 for more details). The temperature normalization allows for comparison between the early and late periods, and implies that the global average of $\Delta\vec{\tilde{R}}$ (shown in the title) gives an estimate of the global feedback for this period.

The right panel of Figure 5.5 shows the same as the left, except that $\Delta\vec{\tilde{R}}$ is estimated using the full spatial feedbacks for each model ($\hat{\vec{R}}'_{4x} = \Lambda\vec{T}'_{4x}$). The pattern is broadly similar, although there is some difference in the Southern Ocean, possibly related to processes that

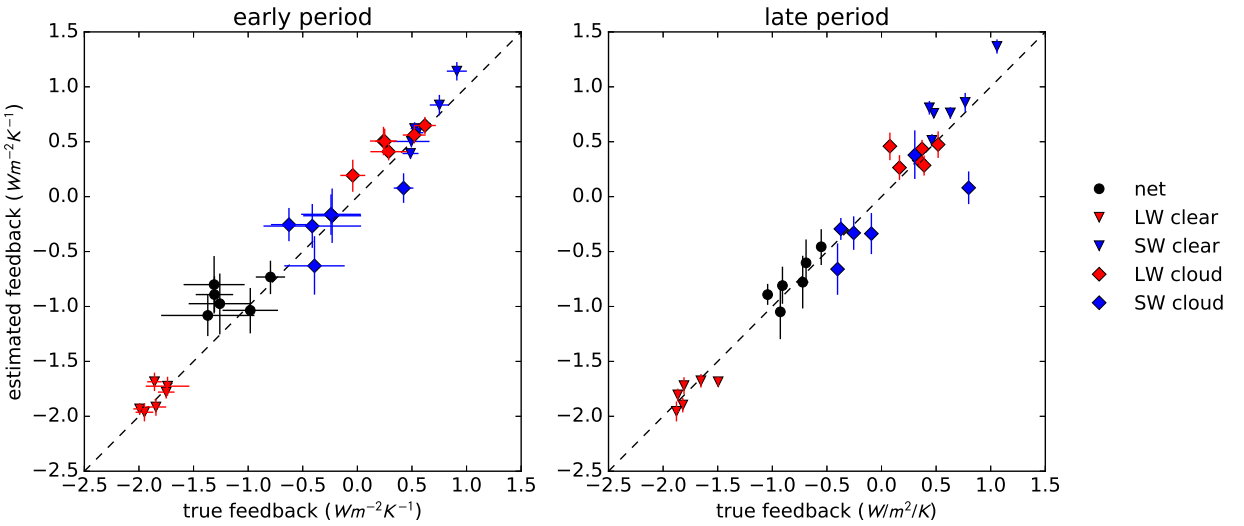

Figure 5.4: Estimates vs. true values of global feedbacks for the early and late periods of *abrupt4x* simulations of the six models considered in this study (plots of the slopes of the gold lines in Figure 5.3 against slopes of black lines, for all TOA fluxes). Error bars show 5%-95% confidence intervals (see Appendix Section 7.2.2 for details). Feedbacks estimated using the MR method generally show good agreement with their true values. In particular, estimates of the net late period feedback have an error of 0.11 $Wm^{-2}K^{-1}$ (Table 5.1). Note that error bars for the true feedback are smaller than their symbols in the late period panel.

can not be described as the linear combination of warming in individual locations (Senior and Mitchell, 2000; Graversen et al., 2014) or due to teleconnections that occur on scales longer than a month (Yuan et al., 2017). Similar results hold for individual fluxes and models. The area-weighted root mean square error of these spatial estimates averaged across all models is given in the bottom rows of Table 5.1 (see Appendix Section 7.2.3 for more details).

5.3 Gaining physical insight from the MR method

We can use the MR method to understand how these GCMs respond to warming in different regions. The left panel of Figure 5.6 shows the change in the globally averaged value of R caused by warming in a given region T_i (i.e., a map of $\vec{\lambda}$) for the multi-model and multi-month mean. This picture is very similar for five of the six models: warming in regions of tropical convection (Indonesia, the western Pacific, and the Caribbean) tends to make the

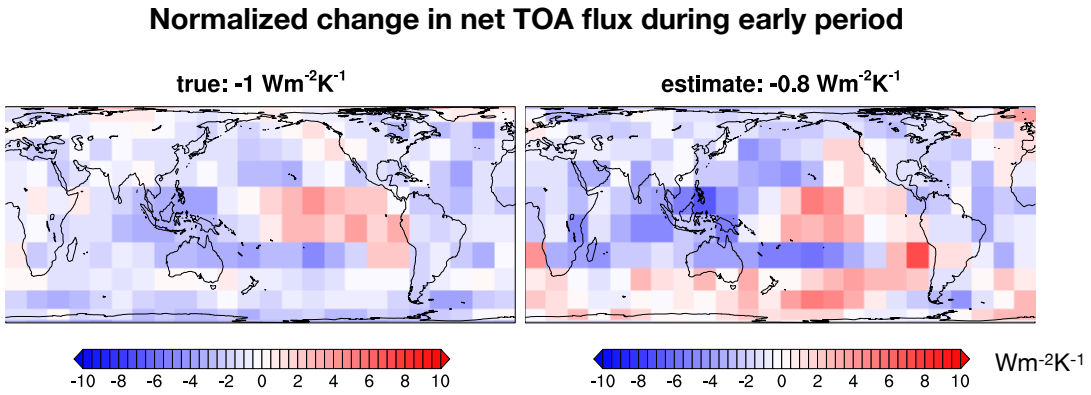

Figure 5.5: Multi-model mean of the change in the net TOA flux pattern between the first and second half of the early period divided by the change in globally averaged temperature. The left panel shows the true change from *abrupt4x* simulations, while the right panel shows the change estimated by the MR method. The area-weighted global average of this change (given in panel titles) gives an alternative estimate of the feedback in this period. The spatial error in Table 5.1 gives the multi-model mean of the average area-weighted root mean square difference for each individual model. Note that the spatial patterns broadly agree, with the exception of the Southern Ocean.

global feedback more negative, while warming in subsiding regions over the oceans tends to make the global feedback more positive. Warming poleward of 45° has relatively little global impact, although this is partly a methodological artifact, in that the grid cells in these regions have a smaller area (see the discussion at the end of Section 5.4). Note that the strong difference in feedback sign and strength across the tropics implies that anomalous tropical warming patterns have the greatest potential to change the global feedback strength, while homogeneous warming in the tropics may lead to a closer balance between negative and positive feedbacks. The spatial pattern of the net feedback appears to be primarily caused by SW cloud feedbacks (right panel of Figure 5.6).

We can use the full version of the MR method (Λ) to disaggregate the feedbacks in Figure 5.6 into local (Figure 5.7) and nonlocal (Figure 5.8) components. The local feedbacks are the direct influences of T_i on R_i (the diagonal elements of Λ). Note that these are not the same as the slopes of the individual regressions of T_i against R_i, as we discuss more fully

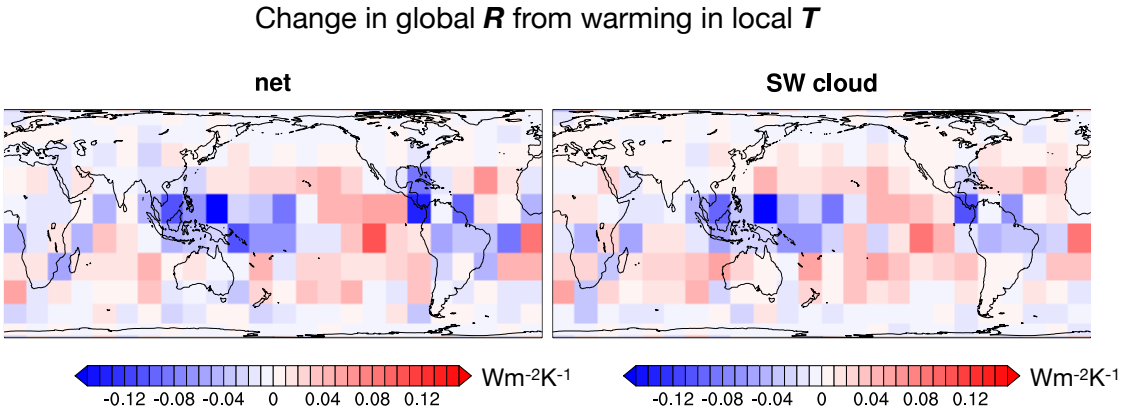

Figure 5.6: Multi-model and multi-month mean value of spatial feedbacks estimated using the MR method, where each grid cell shows the change in the global value of the given TOA flux per degree of warming in that grid cell. Feedbacks are strongly negative in regions of strong tropical convection (Indonesia, the western Pacific, and the Caribbean) and positive in tropical and subtropical subsiding regions over the oceans. The spatial pattern for the net feedback (left panel) is almost entirely determined by the spatial pattern of the SW cloud feedback (right panel).

in Section 5.4. The nonlocal feedbacks are the influences of T_i on all R_j where $j \neq i$ (the nondiagonal elements of Λ). In all cases, these spatial feedbacks are weighted by the area of their R grid cell to ensure that the sum of Figure 5.7 and Figure 5.8 produces Figure 5.6.

The left panel of Figure 5.7 shows the local component of the net spatial feedback, which is almost always positive. This holds for all GCMs except for GISSE2R. The MR method therefore implies that in the absence of nonlocal feedbacks, the five other GCMs would be unstable to radiative forcing (e.g., increasing CO_2 would cause runaway warming). These positive local feedbacks are primarily caused by tropical SW cloud feedbacks (middle panel). This is consistent with recent studies that suggest that surface warming in regions of low cloudiness weakens lower tropospheric stability, leading to entrainment of moist air out of the boundary layer and a subsequent loss of low clouds (Klein and Hartmann, 1993; Wood and Bretherton, 2006; Rose and Rayborn, 2016; Andrews and Webb, 2017; Ceppi and Gregory, 2017; Klein et al., 2017; Zhou et al., 2017). The relative lack of local LW cloud feedbacks in the subsiding regions of the tropics (right panel) would imply that the SW cloud feedback

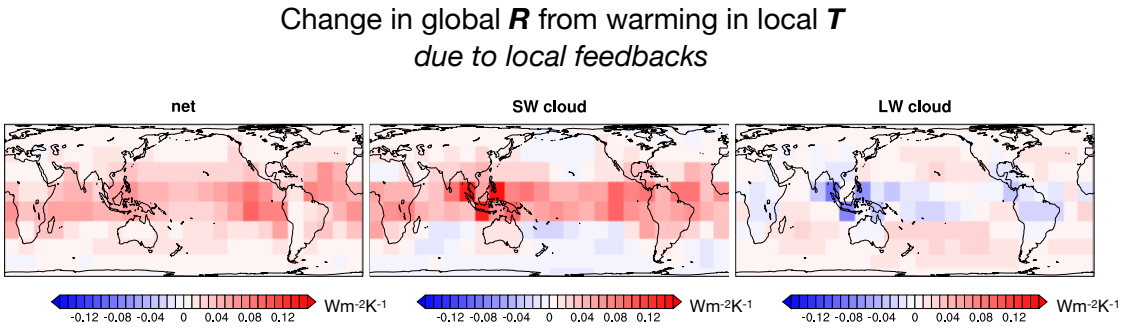

Figure 5.7: Same as Figure 5.6, but only considering changes in the global TOA flux from warming in each grid cell due to local feedbacks (changes in R_i due to T_i, i.e. the diagonal elements of Λ). For the net feedback (left panel), local feedbacks are positive almost everywhere, implying that in the absence of nonlocal feedbacks, the climate would be unstable to radiative forcing. In the tropics, local feedbacks are positive primarily because of the SW cloud feedback (middle panel), while in the extratropics this is primarily due to the LW cloud feedback (right panel). A similar picture holds for all models except for GISSE2R.

in these regions is primarily a low cloud effect, although in convecting areas like Indonesia there also appears to be strong and offsetting local LW cloud feedbacks likely related to high clouds. Outside of the tropics, the local SW cloud feedback tends to be slightly negative, possibly related to a mixed-phase cloud feedback (McCoy et al., 2018). The total local cloud feedback is positive in these regions, consistent with a positive LW cloud feedback likely due to increases in cloud top altitude and optical thickness (Zelinka et al., 2012).

These models are kept stable by the nonlocal component of their spatial feedbacks, which is primarily negative, especially in tropical convecting regions (left panel of Figure 5.8), in most part due to the SW cloud feedback pattern (middle panel). As shown in Figure 5.1, surface warming in the convecting tropics is able to propagate broadly, unlike warming in the subsiding tropics or the extratropics. In the subsiding tropics, warming is kept from propagating vertically by an inversion at the top of the boundary layer, while in the extratropics, local perturbations are confined horizontally by the Coriolis force. Since neither constraint holds in the convecting tropics, surface warming can warm the entire tropical free troposphere, increasing the lower tropospheric stability and inhibiting the loss of boundary layer moisture and low clouds (Klein and Hartmann, 1993; Wood and Bretherton, 2006; Rose and

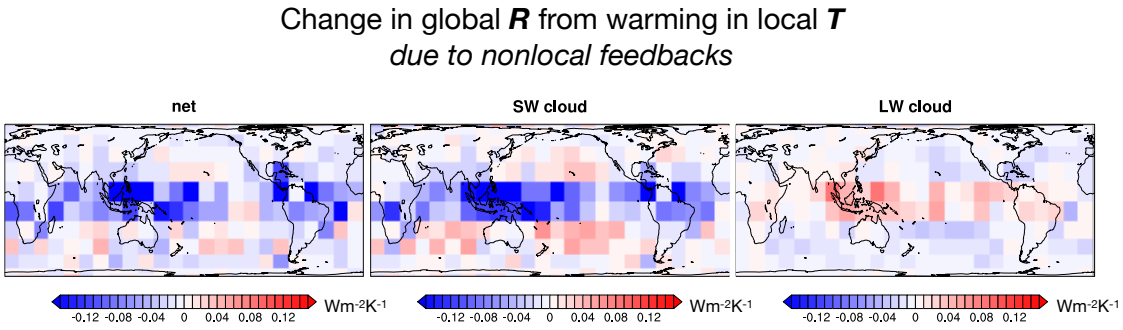

Figure 5.8: Same as Figure 5.7, but now only nonlocal feedbacks (nondiagonal elements of Λ) are considered. Once more, the net feedback pattern (left panel) is largely driven by the SW cloud feedback (middle panel), with some contribution from the LW cloud feedback (right panel). The strength of the negative nonlocal SW cloud feedback associated with regions of tropical convection ensures that the global feedback is negative. Once more, a similar picture holds for all models except for GISSE2R.

Rayborn, 2016; Andrews and Webb, 2017; Ceppi and Gregory, 2017; Klein et al., 2017; Zhou et al., 2017). Comparison with the nonlocal LW cloud feedbacks (right panel) confirms this is primarily, though not exclusively, a low cloud phenomenon. Outside of the tropics, the patterns once more reverse, perhaps because the nonlocal lapse rate effect also reverses sign, leading to a reduction in nonlocal mid and high-cloudiness. Note that the local and nonlocal LW cloud effects mostly cancel, so that the LW cloud feedback contributes negligibly to the net pattern in Figure 5.6.

This disaggregation between positive local and negative nonlocal feedbacks becomes particularly striking when we separate the MR method's estimates of the global climate feedback between local and nonlocal components (Figure 5.9). The average magnitude of the local and nonlocal feedbacks for the two periods is about 3 to 6 $Wm^{-2}K^{-1}$ with opposing signs. While it is possible that this cancellation is an artifact of our method, it is also possible that our method has identified two physically distinct cancelling cloud feedbacks (and there is observational evidence to suggest that it has, as we discuss in Section 5.4). If the latter is true, this suggests an important question both for assessing the risk of global warming and for understanding planetary habitability generally: is there a physical mechanism that ensures this cancellation, or is this cancellation one of many possibilities, which our present

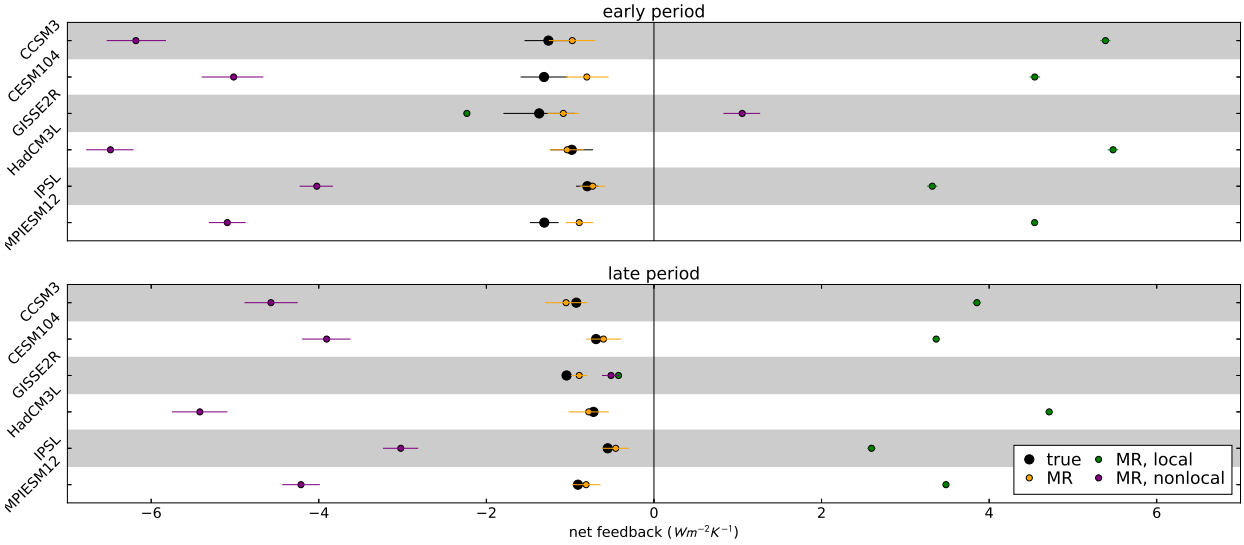

Figure 5.9: True (black dots) and estimated values of the global net feedback for the early and late periods of each model's *abrupt4x* simulation. The MR method estimates (gold dots) are the sum of strongly compensating local (green dots) and nonlocal (purple dots) feedbacks. Lines show 5-95% confidence intervals (see Appendix Section 7.2.2). The magnitude of these feedbacks is comparable to the Planck feedback. Strongly positive local feedbacks are consistent with observational results that find predominantly positive spatial feedbacks when local regressions are performed (Section 5.4). More work is needed to determine if the compensation between these components is a methodological artifact, a result of a physical mechanism, or simply a feature of the present climate that need not hold generally.

Earth (or our computer simulations of it) happens to have? There appears to be significant correlation between the magnitudes of the local and nonlocal feedbacks in different models, which could be a result of a methodological artifact, a physical compensation, or even a selection bias in the tuning of GCM parameters (Huybers, 2010; Mauritsen et al., 2012). If many combinations of local and nonlocal feedback strengths are possible, and if the strengths of these relative terms are determined by features as idiosyncratic as the current Earth's Hadley and Walker circulations, then we should expect that clouds should play a leading order role in determining planetary stability (Yang et al., 2013), and in fact some GCMs exhibit loss of stability due in part to cloud feedbacks under high CO_2 forcing (Caballero and Huber, 2013; Popp et al., 2016).

The relationship of local and nonlocal components is quite different for clear-sky feed-

backs. SW clear-sky feedbacks are almost entirely local and positive, although there is a small nonlocal contribution, presumably from local warming leading to a flux of moisture into surrounding regions. For LW clear-sky feedbacks, local and nonlocal feedbacks are both negative, and can be of comparable magnitude, primarily in the extratropics (Andrews and Webb, 2017; Po-Chedley et al., 2018). This suggests that studies that estimate planetary habitability using a single column model (e.g., Kopparapu et al., 2013) likely overestimate the LW clear-sky feedback and the inner edge of the habitable zone, as has been shown by using a 3D model (Leconte et al., 2013).

The MR method also allows us to investigate why the climate feedback changes with time. The left panel of Figure 5.10 shows the multi-model mean difference between the normalized warming in the early and late periods of the *abrupt*4*x* simulations. The MR method allows us to quantify the contribution of the change in regional warming to the change in the global feedback by multiplying a GCM's change in warming pattern by its estimated spatial feedbacks, resulting in the middle panel. As the amount of local warming per degree of global warming decreases in the tropics, the particularly negative feedback associated with warming in the tropics (left panels of Figure 5.6 and 5.8) contributes less to the global feedback, leading to a more positive global feedback, as shown by the red grid cells in the tropics in the middle panel. Comparison with the right panel suggests that SW clouds contribute predominantly to this effect (although tropical LW clear-sky feedbacks may also play a role).

While the MR method suggests that the tropics are the main cause of the change in global feedback between the early and late periods (e.g., the zonal plot on the left of the middle panel of Figure 5.10), the method does poorly in the Southern Ocean (Figure 5.5), the region some studies identify as the cause of feedback change (Senior and Mitchell, 2000; Armour et al., 2012). Specifically, Senior and Mitchell (2000) say cloud feedbacks in this region may get more positive over time due to differential warming between regions, which would be a nonlinear effect (and so not captured by the MR method). Further research is

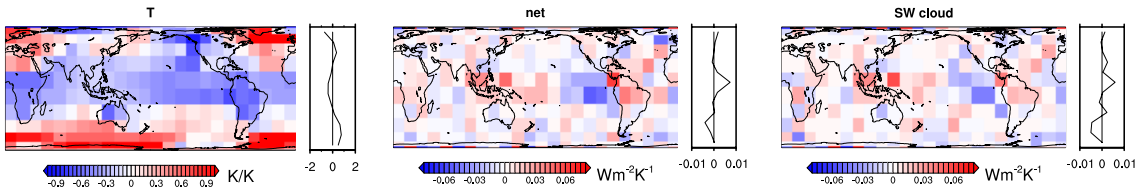

Figure 5.10: Left panel shows the change in the amount of local warming per degree of global warming between the early and late periods of each model. As the zonal mean on the right side of the left panel indicates, tropical warming makes up a smaller share of global warming over time. The middle panel shows the change in the global feedback between these two periods caused by the change in warming in each location. As shown in Figure 5.6, tropical feedbacks are particularly negative, and so a reduction in the share of tropical warming leads to an increase in the global feedback as seen in the zonal mean of the middle panel. This is primarily driven by SW cloud feedbacks, as suggested by the right panel. Note that the MR method has some degree of error over the Southern Ocean (Figure 5.5), so that changes in this region (possibly associated with nonlinear effects) may also contribute to the strengthening of the global feedback with time, rather than weakening it as the MR method suggests.

needed to understand feedbacks in this region.

The physical picture described in this section holds for all models except the GISSE2R model. The local component of the GISSE2R global net feedback is negative, and the model does not have the strongly positive local tropical cloud feedbacks exhibited by the others. This holds true whether we use the normal $15° \times 15°$ grid or a finer $7.5° \times 7.5°$ grid. This suggests that significantly different physical processes determine cloud feedbacks in this model, but it is a sign of the flexibility of the MR method that it is still able to predict the forced response from its interannual variability, suggesting it is able to capture its behavior. The difference in local feedbacks revealed by the MR method provides an opportunity to determine which of these different spatial cloud feedbacks is more consistent with observations, as we discuss briefly in Sections 5.4 and 5.5.

5.4 The MR method in the context of other spatial feedback methods

Previous studies have discussed other methods for deriving spatial radiative feedbacks from interannual variability. In this section, we show how our analysis complements this body of work.

As discussed above, the simplest way to estimate the climate feedback from interannual variability is to regress the globally averaged TOA flux anomaly R' against the globally averaged surface temperature anomaly T':

$$R'(t) = \lambda_{global} T'(t) + \epsilon(t) \tag{5.3}$$

(For more details on methods used in this section, see Appendix Section 7.2.4.) This can be modified to estimate spatial feedbacks by regressing the change in local net TOA flux anomaly R'_i against the globally averaged surface temperature (T):

$$\vec{R}(t) = \vec{\lambda}_{global} T(t) + \vec{\epsilon}(t) \tag{5.4}$$

Equation 5.4 is the "global" definition used in Feldl and Roe (2013), and is often referred to as the "local contribution" to the global feedback (Boer and Yu, 2003a,b; Crook et al., 2011; Zelinka et al., 2012; Andrews et al., 2014). Note that λ_{global} is simply the area-weighted average of $\vec{\lambda}_{global}$, and so the global feedback estimated using this method is always λ_{global}. This method has been found to do a relatively poor job of predicting the net and cloud feedbacks exhibited in response to forcing (Dessler, 2012; Colman and Hanson, 2017), and cannot predict changes in the feedback due to the change in pattern of surface warming, as it only uses globally averaged surface temperature.

The local contribution estimated using the "global" method largely reflects the ENSO patterns associated with globally cold vs. warm years. We can estimate global feedbacks

using the "global" method, which generally does not perform as well as the MR method (see the first three rows of Table 5.1). The "global" method does a poor job of replicating the spatial pattern of change in TOA flux as the method does not account for the spatial pattern of the underlying warming.

A second method for estimating spatial feedbacks is the "local" definition given in Feldl and Roe (2013), in which the local TOA flux anomaly R_i' is regressed directly against local surface temperature anomaly T_i':

$$R_i(t) = \lambda_{local,i} T_i(t) + \epsilon(t) \text{ for each region } i \tag{5.5}$$

producing a vector $\vec{\lambda}_{local}$. This approach has two major drawbacks:

- The "local" method does not explicitly account for nonlocal effects, which we found can be larger in magnitude than the local feedback itself. Since the nonlocal feedback is often negative, leaving it out can result in a large overestimate of the feedback associated with warming in a region. For example, two recent studies that have used this approach to estimate spatial feedbacks from observations have found that most regions of the Earth exhibit positive feedbacks, including in the global average (Brown et al. 2015, and upper right panel of Figure 10 in Trenberth et al. 2015), which would imply that for roughly uniform warming, the planet would be unstable to radiative forcing.

  Two other studies (Boer and Yu, 2003b; Crook et al., 2011) found that applying this method to a control simulation resulted in regressions whose error terms were too large to make the regressions useful, at least when a grid resolution similar to the one in the present study was used. Crook et al. (2011) instead used zonal averages, which accounts for nonlocal feedbacks by including more of them in the region being studied, but means that the feedbacks they derive may reflect the specific spatial pattern of surface temperature change occurring within the region when the regression was taken,

and therefore may not hold if those internal patterns differ under forcing (e.g., warming in a tropical zonal belt may produce a quite different feedback depending on where in the tropical belt the warming is occurring).

- Some of the covariance between R_i and T_i in Equation 5.5 could be due to the influence of surface temperatures outside of region i on both T_i and R_i. This can be understood by constructing a toy model with two regions in which the temperatures in both regions are given by a *sin* function with a small amount of noise added. The temperature in the two regions will therefore be strongly but not perfectly correlated. Suppose that the net TOA radiative fluxes in both regions are determined only by the surface temperature in the first region T_1, with a small positive local feedback (affecting Region 1) and a large negative nonlocal feedback (affecting Region 2):

$$\begin{aligned} T_1 &= \sin(t) + \epsilon(t); \quad T_2 = \sin(t) + \epsilon(t) \\ R_1 &= 0.5T_1 + \epsilon(t); \quad R_2 = -2T_1 + \epsilon(t) \end{aligned} \tag{5.6}$$

Figure 5.11 shows the result of applying the "local" and MR methods to more and more years of a realisation of this model. Since T_2 is correlated with T_1, and R_2 depends on T_1, there is a significant correlation between R_2 and T_2. As a result, the "local" method incorrectly predicts a negative feedback in Region 2 ($\hat{\lambda}_{2,local}$ in the left panel of Figure 5.11). Given sufficient years, the MR method correctly predicts that the local feedback in Region 2 is zero ($\hat{\lambda}_{2,2,MR}$ in the right panel of Figure 5.11).

In GCM simulations with strong positive radiative forcing (such as *abrupt*4x simulations), surface temperature covaries strongly across different regions, as they are all undergoing warming. As a result, studies that apply the "local" method to forced simulations (e.g., Armour et al., 2012; Colman and Hanson, 2017) are likely inadvertently conflating significant nonlocal effects with local feedbacks, so that the feedback associated with warming in a given region may reflect the consequences of warming

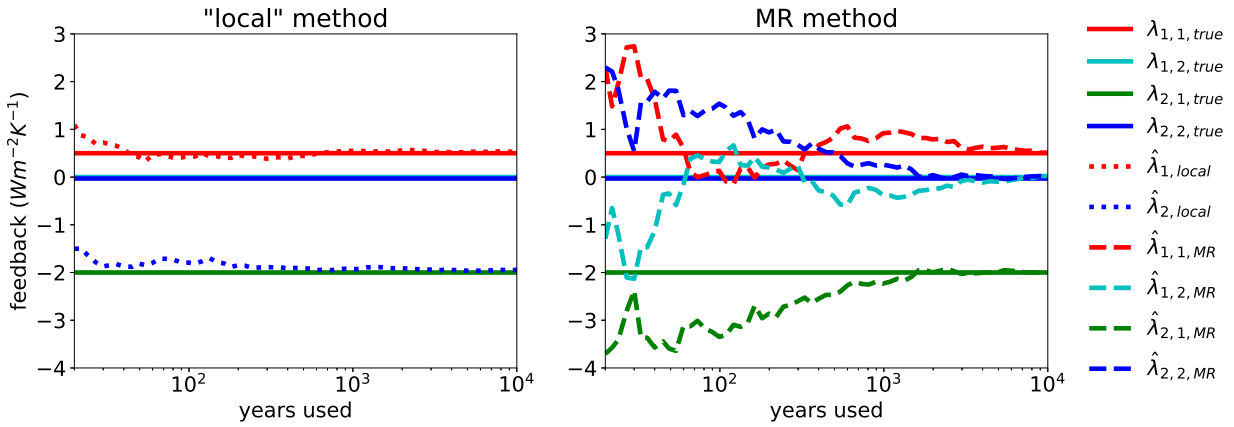

Figure 5.11: A toy model showing how covariance between surface temperatures in different regions can affect estimates of spatial feedbacks using the "local" method. In the model given in Equation 5.6, the temperature in Region 1 (T_1) controls the TOA fluxes in Regions 1 & 2 (R_1 and R_2), while the temperature in Region 2 (T_2) has no effect on TOA fluxes in either region. Because of the covariance between T_2 and T_1 and the dependence of R_2 on T_1, the Region 2 feedback estimated by the "local" method, which is the slope of the regression of R_2 against T_2 ($\hat{\lambda}_{2,local}$, dotted blue line in the left panel), converges to a spurious negative local feedback estimate for Region 2. Using multiple regression to estimate the local feedback ($\hat{\lambda}_{2,2,MR}$) instead converges to the correct value of 0.

far afield. This may explain why the feedbacks estimated in these studies tend to be less positive than those found in the studies discussed above, which apply the "local" method to observations dominated by interannual variability.

We now apply the "local" method to the simulations considered in this study. As might be expected from the results of Section 5.3, the "local" method net and SW cloud feedbacks of all models besides GISSE2R are mostly positive, similar to the observations from Brown et al. (2015) and Trenberth et al. (2015). As a result, the global *abrupt4x* feedbacks predicted by the "local" method are positive, except for GISSE2R, although not as positive as the local feedbacks estimated by the MR method, due to contamination from temperatures elsewhere (since nonlocal feedbacks are predominantly negative). If nonlocal feedbacks were also mostly negative for the real Earth, then studies like Brown et al. (2015) and Trenberth et al. (2015) would imply that the true local feedback would be even more strongly positive than their results. This would imply a local positive cloud feedback (since clear-sky feedbacks outside

of icy or snowy regions are not likely to be locally strongly positive), suggesting that the physical picture presented discussed in Section 5.3 for the non-GISSE2R models is more likely than that presented for the GISSE2R model. This also suggests that strongly compensating local and nonlocal cloud feedbacks may exist for the real Earth.

As a result, the "local" method does poorly at estimating the global feedback and its change compared to the other methods considered (top three rows of Table 5.1), except for the global SW clear-sky feedback, since this feedback is primarily due to changes in snow and ice albedo, which are local processes. However, the "local" method does a better job than the "global" method at capturing the spatial pattern of TOA flux change (last two rows of Table 5.1), perhaps because it is able to reflect changes in the spatial pattern of warming. The MR method has still smaller spatial errors, except for the SW clear-sky feedback.

As a methodological note, we use each model's native grid to estimate feedbacks for the "global" and "local" methods, and then regrid to the coarser grid to compare these methods' estimates of the spatial pattern of TOA flux change with the MR method. We also use the annual version of the "global" and "local" methods, rather than first finding monthly feedbacks as we do with the MR method. We do so because the annual versions of the "global" and "local" tend to produce smaller errors. Also note that while the spatial feedbacks estimated by the local and global methods do not depend on the area of the regions used, the spatial feedbacks given in Equation 5.1 ($\vec{\lambda}$) will depend on grid cell size. For a grid with equal divisions of latitudes and longitudes, more poleward grid cells will be smaller and (all else being equal) have less influence on globally averaged TOA fluxes than larger ones near the tropics. Similarly, for the full spatial feedbacks $\vec{\Lambda}$, given in Equation 5.2, the relative size of the regions i and j will be a factor in the size of the feedback $\lambda_{i,j}$.

5.5 Challenges to applying the MR method to observations

Our results suggest that for GCMs, it is possible to use interannual variability to estimate the spatial feedbacks at work in response to radiative forcing. This suggests it may be possible

to estimate time-invariant spatial feedbacks for the Earth from observations, and in turn to forecast future warming by forecasting climate feedbacks and changes in these feedbacks. There remain significant challenges to doing so:

- *The MR method requires many years of observations:* As mentioned above, our method requires at least as many years as grid cells used (which for most of our models is 288), and at least a thousand years to become accurate, far more than the few decades of satellite data available. Our method is very generalized, using far more information than is likely necessary. A better choice of grid might target specific areas, although Andrews and Webb (2017) demonstrates that box models that work well for one GCM might not work well for others, due to the variety in the mean state and sensitivity of tropical circulations among models. Choosing the right box model for the Earth may be difficult because of this uncertainty. A Bayesian approach that uses prior assumptions about the physical relationships underlying spatial feedbacks may allow a similar method to be employed with fewer years of data.

- *The MR method requires forecasts of the spatial pattern of warming:* In our analysis above, we knew the true spatial pattern of warming for the forced response ($\vec{T}_{4x}$). In reality, the future pattern is not known and would have to be predicted with other methods. Estimating the spatial feedbacks could help if they were incorporated into a broader spatial energy balance model that includes ocean heat uptake (e.g., by building on Marshall and Zanna, 2014) and horizontal energy transport (e.g., Roe et al., 2015), taking spatial trajectories of radiative forcing as an input.

- *The Earth may be more nonlinear than the GCMs included here:* Although our method appears to work for the six models studied (except for in the Southern Ocean), the real Earth may behave in a nonlinear fashion. A simple way this can occur is if feedbacks are temperature-dependent. Figure 5.12 shows true and estimated feedbacks for five different levels of abrupt CO_2 forcing of MPIESM12. The feedback strength gets

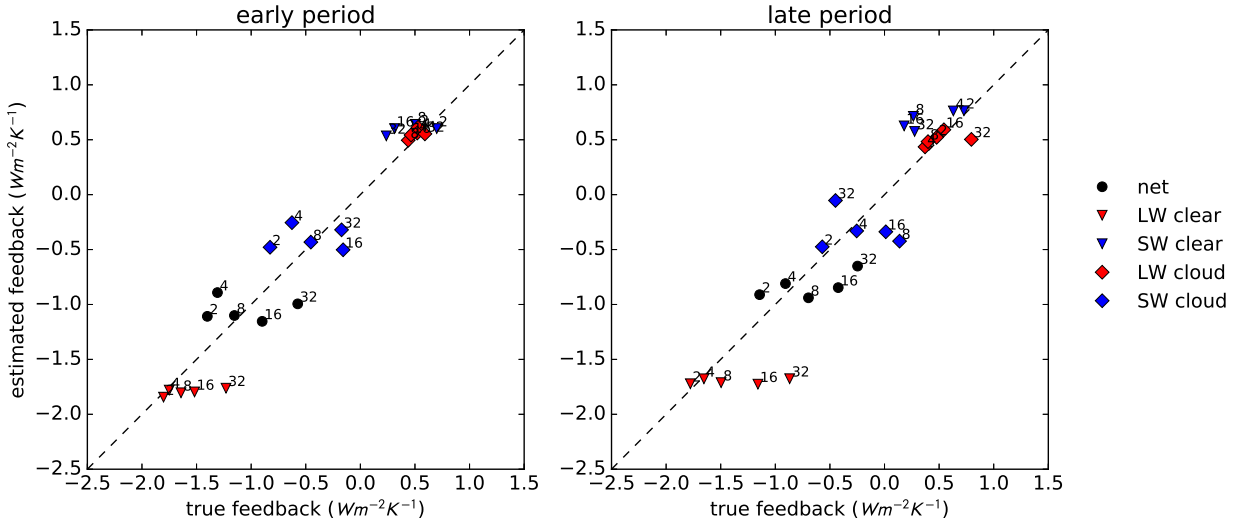

Figure 5.12: Same as Figure 5.4, except instead of different models, markers represent different simulations of the MPIESM12 model, each with a different level of abrupt CO_2 forcing indicated by numbers next to the markers. While the net and LW clear feedbacks become more positive at higher forcings, the MR method's estimate of these feedbacks remains roughly the same. Previous work (Meraner et al., 2013; Bloch-Johnson et al., 2015) suggests that these feedbacks are temperature-dependent, a nonlinearity that would not be accounted for by the linear MR method.

increasingly positive for higher forcings, primarily due to an increase in the LW clear-sky feedback (likely as a result of a stronger water vapor feedback in a warmer world, Meraner et al., 2013).

A more complex example would be if warming multiple regions simultaneously had a result that was different than the linear combination of the effects of warming each region alone. As discussed above, such an effect may be at work over the Southern Ocean in some models (Senior and Mitchell, 2000), or in general in polar regions (Graversen et al., 2014), leading to a discrepancy in the estimate of normalized TOA flux change (Figure 5.5). The real Earth might have similarly complex feedbacks over a larger region. For example, our method assumes that local and nonlocal feedbacks are constant functions of location and month of the year, but circulations change significantly within interannual variability, suggesting that the spatial feedbacks themselves may change if these oscillations and their frequency change under forcing (for example, spa-

tial feedbacks might differ strongly between El Nino and La Nina years; Pierrehumbert 1995).

It is possible to alter our method to account for some degree of nonlinearity by adding additional terms, although it is always possible that the warmer Earth will exhibit behaviors outside of the experience of present climate. To account for these effects, physical reasoning and paleoclimate proxies must be used, and the results of these studies can be incorporated into an observational MR model. Alternatively, the use of our linear method may provide a way of testing for the presence of nonlinearity by measuring the residual behavior between forecasts and outcomes in contexts where can be sure that other sources of error are small.

- *Forcing must be accounted for:* Unlike in preindustrial simulations, observational time series of TOA fluxes and surface temperature have some degree of forcing signal which must be removed. Understanding the spatial and seasonal pattern of this forcing and its response provides a particular challenge.

While these challenges are difficult, they are not insurmountable, and will serve as subjects of further research.

5.6 Conclusions

We have introduced a method for estimating a climate's spatial radiative feedbacks from records of its interannual variability. This method uses multiple regression of regional TOA fluxes against surface temperature in the same and all other regions to estimate the local and nonlocal effects of surface warming. Given a long enough record of interannual variability, this multiple regression (MR) method ensures that these spatial feedbacks are not contaminated by covariance with warming in other regions. We have shown that for six atmosphere-ocean general circulation models, the spatial radiative feedbacks estimated with our method provide improved forecasting of the global feedback, its change with time,

and the spatial pattern of change in net TOA flux compared to existing methods that use only global or local regressions. This improvement also holds for net feedbacks and for all component feedbacks that include significant nonlocal effects (all but SW clear-sky).

The MR method allows us to separate out local and nonlocal feedbacks, and suggests that for five of the six models studied, their local feedbacks are positive in almost all regions. In the absence of nonlocal feedbacks, these models would be unstable to radiative forcing. These positive local feedbacks appear to be due primarily to a tropical local low cloud feedback, which is consistent with studies suggesting that surface warming reduces lower tropospheric stability, reducing low cloudiness. These positive local feedbacks are countered by a negative nonlocal tropical cloud feedback, consistent with studies in which warming in tropical convecting regions warms the entire tropical free troposphere, increasing lower tropospheric stability and low cloudiness elsewhere. The large magnitude of these opposing feedbacks in the forced response (comparable to the underlying Planck feedback) invites the question of whether physical mechanisms keep these cloud feedbacks in balance. The answer is important for future climate change (for example, if local or nonlocal cloud feedbacks depend on temperature, warming can upset this balance, leading to a net positive feedback and subsequent transition to a warmer state) and planetary habitability (it may be less common than expected to have a liquid ocean, as planets with a positive local cloud feedback without an offsetting negative nonlocal one may be unstable in states between a snowball and a runaway greenhouse).

The MR method uses a more general definition of spatial feedback than is used in previous studies, which either defined spatial feedbacks as the local contribution to the global feedback (regression of local values of TOA flux against the globally averaged surface temperature) or as "local" feedbacks (regressions of local values of TOA flux against local surface temperature). The former method does not take the spatial pattern of warming into account, which is necessary to account for the pattern-induced changes in climate feedback over time. The latter does not account for the nonlocal impacts of local warming, and can lead to inac-

curate estimates of the local feedback due to covariance with warming in nonlocal cells. As a result, the "local" method can give very different answers depending on whether the time series it is being applied to has significant forcing, which would tend to cause large covariance between surface temperatures in different regions, and thus a lot of nonlocal contamination. Observational estimates of spatial feedbacks from interannual variability that use the "local" method tend to find broadly positive spatial feedbacks, consistent with the qualitative picture of compensating positive local and negative nonlocal feedbacks found by the MR method. While the "global" and "local" methods have relative advantages at estimating the global feedback and the spatial pattern of feedback, the MR method matches or improves on either in both categories.

Significant challenges remain to applying the MR method to observations. The method requires hundreds of years of observations. Customizing the grid and using prior assumptions about the physical nature of the Earth's climate may allow us to focus on important variability in the observational record, reducing the years needed. The MR method requires a forecast of future spatial patterns of warming, which could be created by complementing these spatial feedbacks with a full energy balance model, including spatial patterns of forcing, ocean heat uptake, and atmospheric energy transport. The method does not account for nonlinearity, either in the sense of temperature-dependent feedback processes or of cross-terms arising from simultaneous warming in different regions. Finally, since we only consider preindustrial simulations, the MR method does not address the need to remove radiative forcing and its responses from records of interannual variability.

In summary, our work suggests that interannual variability and forced response may set off the same spatial feedback processes, with a prime role for nonlocal feedbacks. We may not have to wait to observe how climate change will unfold – the nature of the Earth's response may be playing out all around us, waiting to be properly disentangled. Doing this work may allow us to forecast future warming from observations, in spite of the inconstancy of the climate feedback.

Table 5.1: Root mean square errors (in units of $Wm^{-2}K^{-1}$) of the methods for estimating spatial feedbacks from interannual variability considered in this paper, where error definitions are given in Appendix Section 7.2.3. "MR" is the multiple regression method proposed in Section 5.2, and "global" and "local" are existing methods described in Section 5.4. The feedback "early" and "late" rows show the errors in the estimate of the early and late period *abrupt*4*x* global feedbacks, while "diff" gives the error in the estimate of the change in feedback between the early and late period (note that the "global" estimate assumes a constant feedback, and therefore assumes no change with time). The spatial "early" and "late" rows give the area-weighted root mean square error of the change in spatial pattern of TOA flux normalized by globally averaged surface warming and regridded to match the grid used for the MR method. While the "global" method estimates the feedback better than the "local" method and the "local" method estimates the spatial pattern of TOA flux better than the "global" method, the "MR" method performs better at both, with the exception of SW clear for which nonlocal effects are negligible.

	feedback			*spatial*	
	early	late	diff	early	late
net					
MR	0.32	0.11	0.28	2.8	2.4
global	0.33	0.38	0.39	4.9	5.54
local	2.88	2.29	0.61	3.52	2.65
LW clear					
MR	0.08	0.10	0.09	0.84	0.75
global	0.13	0.21	0.13	2.57	3.13
local	0.71	0.47	0.27	1.17	0.82
SW clear					
MR	0.12	0.24	0.18	1.76	2.18
global	0.12	0.19	0.08	2.02	2.17
local	0.80	0.2	0.16	1.4	1.92
LW cloud					
MR	0.19	0.17	0.17	1.81	1.45
global	0.42	0.49	0.2	5.75	6.85
local	1.24	0.96	0.34	2.88	2.02
SW cloud					
MR	0.24	0.33	0.27	3.18	2.82
global	0.30	0.58	0.32	6.59	7.73
local	1.36	0.97	0.46	3.61	2.72

CHAPTER 6

CONCLUSIONS

In this thesis, we have set out to understand time-varying climate sensitivity. The equilibrium climate sensitivity, when defined as the warming response from doubling CO_2 relative to current CO_2 levels, changes over time because the two quantities it depends on, the radiative forcing and the equilibrium feedback, change with the surface temperature and the background CO_2 level (Figures 4.1 and 4.3), a phenomenon we refer to as "equilibrium nonlinearity." Of all these nonlinearities, the dependence of the equilibrium feedback on the surface temperature can have the largest effect, particularly when this dependence is positive (Figure 3.2). Our offline analysis of feedbacks in ECHAM6.1 suggests that the water vapor feedback has a positive feedback temperature dependence, in keeping with our understanding that at some point this feedback strengthens sufficiently to cause a runaway greenhouse. The possibility of positive feedback temperature dependence can greatly increase the risk of high warming in forecasts of future warming (Figure 3.3). Ruling out these scenarios can be difficult even when our knowledge of the current climate feedback is relatively certain (Figure 4.5).

The instantaneous climate sensitivity is the warming response one would get if one assumed that the instantaneous climate feedback held fixed as the Earth came to equilibrium. Unlike the equilibrium climate sensitivity, we can measure it from observations, as well as from CMIP coupled runs which are not run to equilibrium. While the instantaneous climate sensitivity can change over time because of *equilibrium* nonlinearity, we would not expect to observe this effect under the relatively small amount of warming seen so far in observations, nor even even under the forcing levels considered for most CMIP models (Figure 2.2). In other words, the equilibrium climate sensitivity has likely not changed much in the observational record, or during the first 150 years of a CMIP abrupt4x simulation. Therefore, while we should consider the inconstancy of the equilibrium climate sensitivity in projecting future warming, equilibrium nonlinearity is likely not impacting our estimates of current climate

sensitivity.

However, the instantaneous climate sensitivity can be different than the equilibrium climate sensitivity because the relative amounts of warming associated with different regions can be different for the same globally averaged warming, setting off different feedback processes in different regions, and therefore leading to different changes in R with T. In particular, when the spatial pattern has relatively more warming in tropical convecting regions, this sets off a strongly negative nonlocal low cloud feedback. Since these regions tend to warm more quickly than regions of strong ocean heat uptake, such as the Southern Ocean and North Atlantic, the current instantaneous climate sensitivity, as inferred from observations, likely underestimates the current equilibrium climate sensitivity. Our understanding of these spatial feedbacks come mostly from models, but the work in Chapter 5 suggests it may be possible to infer these spatial feedbacks directly from observations of interannual feedbacks.

Therefore, in order to forecast future warming, especially from observations, we must account for the present negative transient nonlinearity caused by anomalous warming in regions of the convecting tropics and the future positive equilibrium nonlinearity caused by the temperature dependence of the water vapor feedback – basically, the former causes our current estimates of λ_{pi} to be too negative, while the latter gives the capacity for a positive a. On their own, these two nonlinearities would cause only a mild underestimate of future warming. However, combined, they give the spark and the fuel source respectively for enormous increases in climate sensitivity, giving just enough of an increase in initial warming to make it more likely that feedback temperature dependence can kick in (i.e. the differences between the right side of the gold and green curves in the middle panel of Figure 3.2). Without this transient nonlinearity, we may not need to worry about equilibrium nonlinearity under anthropogenic forcings, but with it, understanding feedback temperature dependence becomes incredibly important. We hope this work is just the beginning of a more thorough study of these two different and distinct causes of time-varying climate sensitivity. While this work explored these ideas using general circulation models, we hope that future

work allows us to study these ideas directly in observations and using physical first principles.

CHAPTER 7

APPENDIX

7.1 Appendices to Chapter 3

7.1.1 Estimating a for a GCM

For each GCM, we have a collection of runs where atmospheric CO_2 was abruptly increased from preindustrial levels to d_{CO_2} (see Chapter 2.1) and then the model was run for long enough to estimate the equilibrium warming response, $\Delta T_{d_{CO_2}}$. The initial radiative imbalance is the forcing ($F_{d_{CO_2}} = R_{eq}(T_{pi}, d_{CO_2})$). We assume that there is no CO_2-dependent feedback (or equivalently, that $R_{eq}(T, d_{CO_2,1}) - R_{eq}(T, d_{CO_2,2})$ is independent of T). Sections 4.2.2 and 7.1.3 discuss the impact of this assumption. As a result of this assumption, for all ΔT,

$$\begin{aligned} R_{eq}(T_{pi} + \Delta T, d_{CO_2}) - R_{eq}(T_{pi} + \Delta T, 1) &= R_{eq}(T_{pi}, d_{CO_2}) - R_{eq}(T_{pi}, 0) && (7.1) \\ &= R_{eq}(T_{pi}, d_{CO_2}) - 0 && (7.2) \\ &= F_{d_{CO_2}}. && (7.3) \end{aligned}$$

So, $-F_{d_{CO_2}} = -(R_{eq}(T_{pi} + \Delta T_{d_{CO_2}}, d_{CO_2}) - R_{eq}(T_{pi} + \Delta T_{d_{CO_2}}, 0)) = -(0 - R_{eq}(T_{pi} + \Delta T_{d_{CO_2}}, 0)) = R_{eq}(T_{pi} + \Delta T_{d_{CO_2}}, 0)$.

If we have m different runs, this gives us $m + 1$ points on the curve $R_{eq}(T_{pi} + \Delta T, 0)$, including $R_{eq}(T_{pi}, 0) = 0$. We can then use least squares linear regression to fit a model of the form

$$R_{eq} = \lambda_{pi}\Delta T + a\Delta T^2 + \epsilon \tag{7.4}$$

to these points. (Note that this fit is constrained so that $R_{eq}(T_{pi}, 0) = 0$). The resulting values are plotted in Figure 3.2 and given in Table 7.1.1.

Table 7.1: Estimating feedback temperature dependence (a) with and without a CO_2-dependent feedback. b is the feedback CO_2 dependence.

GCM	a (w/o CO_2 dep.)	a (w/ CO_2 dep.)	b	# of model runs
GFDL	-0.030	-0.044	0.07	3
GISS-E	0.059	0.048	0.02	8
BMRC	-0.034	-0.016	-0.05	9
CAM3	0.042	0.186	-0.42	3
ECHAM6	0.031	0.019	0.10	4

7.1.2 Bifurcations and higher order terms

See Figures 7.1 and 7.2.

7.1.3 Adding a CO_2-dependent feedback

We start by adding a quadratic term representing the CO_2 dependence of the feedback to the quadratic model:

$$-F = \lambda \Delta T + a \Delta T^2 + b \Delta d_{CO_2} \Delta T \tag{7.5}$$

where b represents the CO_2 dependence of the feedback (i.e., $b \equiv \partial^2 R_{eq} / \partial T \partial d_{CO_2}$).

We first show the effects of adding this term on our GCM estimates of a. We perform the same regression as we did with the model 7.4, but with the added term $b \Delta d_{CO_2} \Delta T$. Results for a and b are shown in Table 7.1.1. Only one of the GCMs is affected by more than $0.02 W/m^2/K^2$ (CAM3). However, this GCM has only three runs, and has a value of b with a much larger magnitude than any of the other models, suggesting that the small sample size may limit the usefulness of this estimate of a and b.

Our regressions suggest a reasonable range of the feedback CO_2 dependence (b) of roughly ± 0.1 $W/m^2/K$ per doubling, disregarding the outlier associated with the under-sampled model. The effect of feedback CO_2 dependence b is qualitatively different than feedback temperature dependence a, because the nonlinearity associated with a positive a is self-amplifying (warming makes you more sensitive, and being more sensitive makes you warm more) in a way that can lead to jumps to warmer states, or extreme increases in sensitivity.

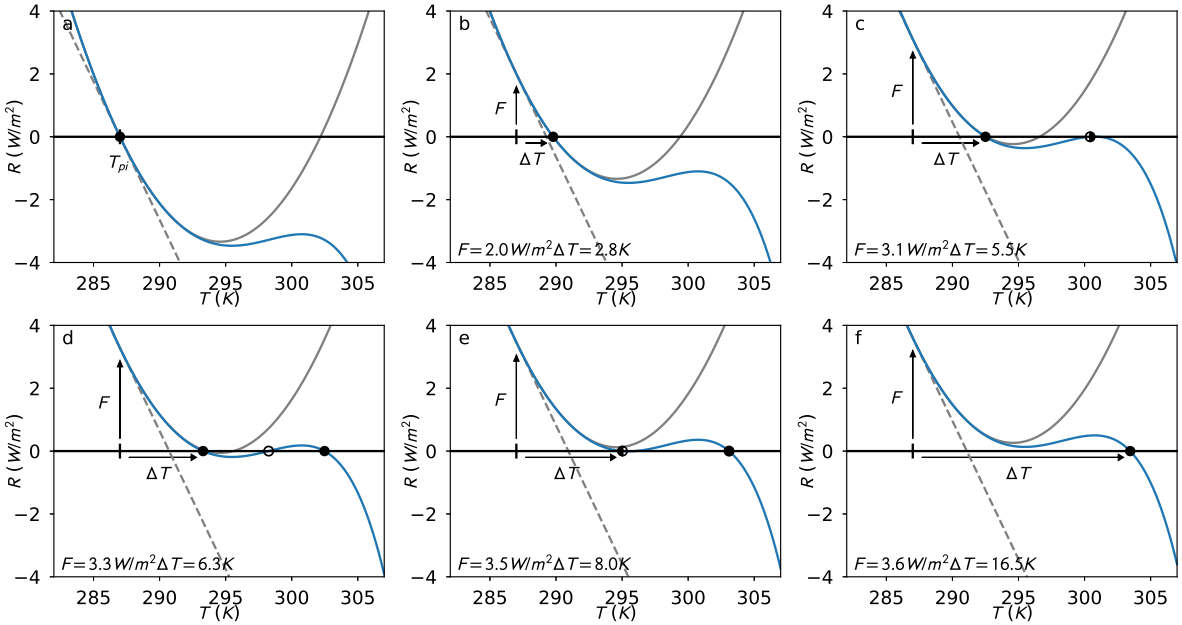

Figure 7.1: **Jumping to a warmer state.** Global annual-mean net top-of-atmosphere energy flux R_{eq} as a function of global annual-mean surface temperature T, where each panel has a successively higher CO_2 concentration. The blue curves have the same preindustrial feedback λ_{pi} ($-0.88 W/m^2/K$) and a ($0.058 W/m^2/K^2$) as the red curve in Figure 1a, but with a higher-order term ($-4 \times 10^{-6} \Delta T^5$) added. Gray dashed and solid lines show the linear and quadratic approximations of R_{eq} respectively. a) The planet is in preindustrial equilibrium. b) After $2 W/m^2$ of forcing, the two approximations estimate the warming well. c) After $3.1 W/m^2$ of forcing, another part of the blue curve intersects the x-axis, so that a pair of new equilibria, one unstable and the other stable, is created in a saddle-node bifurcation. d) As the forcing increases, this pair separates, until in e), the stable equilibrium that the Earth is tracking collides with the unstable equilibrium. f) Under further forcing, these two equilibria disappear in another saddle-node bifurcation, and the Earth warms until it reaches the new stable equilibrium.

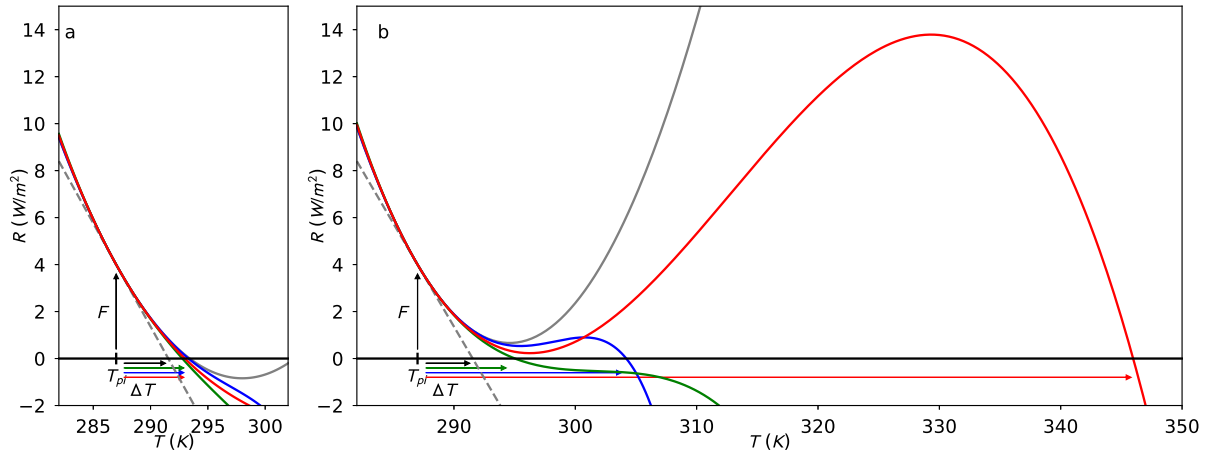

Figure 7.2: **The effect of higher-order terms.** Global annual-mean net top-of-atmosphere energy flux R_{eq} as a function of global annual-mean surface temperature T for a fixed CO_2 concentration, after CO_2 has been increased from the preindustrial value. a) The three colored curves have the same values of preindustrial feedback λ_{pi} ($-0.88 W/m^2/K$), feedback temperature dependence a ($0.02 W/m^2/K^2$), and CO_2 forcing F ($4\ W/m^2$), but different values of higher-order terms (red, $-7.5 \times 10^{-4} \Delta T^3$; green, $-1.3 \times 10^{-3} \Delta T^3$; blue, $-4 \times 10^{-6} \Delta T^5$, where each term is added to $\lambda_{pi} \Delta T + a \Delta T^2$ to estimate R_{eq}). Gray dashed and solid lines show the linear and quadratic approximations of R_{eq} respectively. For this collection of λ_{pi}, a, and F, the quadratic model does not run away, and the various higher-order terms do not significantly affect the total warming. b) The same as a), except that now a is ($0.058 W/m^2/K^2$), so that the quadratic model does run away. As a result, higher-order terms *must* come into play. The different higher-order terms cause very different warmings, and different qualitative behaviors (the green curve experiences no jump to a warmer state, while the red and blue curves do).

The positive CO_2 dependence has some capacity for self-amplifying if d_{CO_2} increases as a function of T (e.g. through the melting of methane hydrates, or the increased release of soil carbon). However, this effect is limited (e.g., there is only so much carbon to be released, and a certain threshold must be reached to release it), while the capacity for self-amplification due to positive temperature dependence does not have a threshold or limited capacity.

The qualitative difference in the two effects can be seen by comparing Figure 3.2 with Figure 7.3. While Figure 3.2 shows how equilibrium warming (ΔT) changes with feedback temperature dependence (a) for a fixed preindustrial feedback (λ_{pi}), Figure 7.3 shows how equilibrium warming changes with feedback CO_2 dependence (b) for a fixed preindustrial feedback. While CO_2 dependence clearly affects the exact value of warming associated with a given CO_2 increase, it does not cause the same extreme behavior, such as loss of stability, or greatly heightened warming, caused by positive feedback temperature dependence.

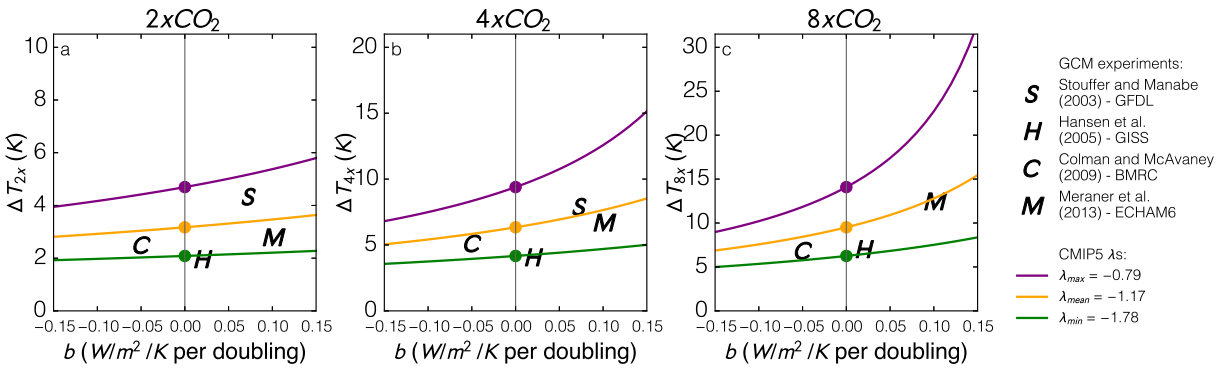

Figure 7.3: **The effect of feedback CO_2 dependence.** This figure is analogous to Figure 3.2, except instead of showing how equilibrium warming (ΔT) changes with feedback temperature dependence (a) for a fixed preindustrial feedback (λ_{pi}), this shows how warming changes with feedback CO_2 dependence (b) for a fixed preindustrial feedback (λ_{pi}). This effect is shown for one, two, and three doublings of CO_2 in panels a), b), and c) respectively. Letters are centered at values of b and equilibrium warmings for various GCMs. While warming changes with feedback CO_2 dependence (with less negative λ_{pi} and larger CO_2 increases causing a larger deviation), these changes do not exhibit the same extreme behavior as the positive values of feedback temperature dependence in Figure 3.2. As an example, there are no shaded areas where the model runs away.

7.2 Appendices to Chapter 5

For LongRunMIP data access, visit http://www.longrunmip.org/. The code used to perform the calculations in this paper is available at https://github.com/jsbj/spatial.

7.2.1 The multiple regression method

Suppose that we have a time series of surface temperatures and TOA radiative fluxes of the Earth, real or simulated, where the surface of the Earth is regridded into n_{grid} regions (288 for all models except for GISSE2R, for which it is 1152), and where we have n_{time} years of monthly observations. For each month of the year m (where m is an integer between 1 and 12), we can define an $n_{time} \times n_{grid}$ matrix $\mathbf{T}_m$, where the element in row i and column j, $T_{i,j,m}$, is the surface temperature in region j during month m of year i. We can also define a matrix of anomalies, $\mathbf{T}'_m$, where

$$\mathbf{T}'_m = \begin{bmatrix} T_{1,1,m} & T_{1,2,m} & \dots & T_{1,n_{grid},m} \\ T_{2,1,m} & T_{2,2,m} & \dots & T_{2,n_{grid},m} \\ \vdots & \vdots & \ddots & \vdots \\ T_{n_{time},1,m} & T_{n_{time},2,m} & \dots & T_{n_{time},n_{grid},m} \end{bmatrix} - \frac{1}{n_{time}} \begin{bmatrix} \sum_{i=1}^{n_{time}} T_{i,1,m} & \sum_{i=1}^{n_{time}} T_{i,2,m} & \dots & \sum_{i=1}^{n_{time}} T_{i,n_{grid},m} \\ \sum_{i=1}^{n_{time}} T_{i,1,m} & \sum_{i=1}^{n_{time}} T_{i,2,m} & \dots & \sum_{i=1}^{n_{time}} T_{i,n_{grid},m} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n_{time}} T_{i,1,m} & \sum_{i=1}^{n_{time}} T_{i,2,m} & \dots & \sum_{i=1}^{n_{time}} T_{i,n_{grid},m} \end{bmatrix} \tag{7.6}$$

We then define global and full versions of the multiple regression method.

Global version

To estimate the global spatial feedbacks associated with a TOA radiative flux of type f (where f is either *net*, *LW clear*, *SW clear*, *LW cloud*, or *SW cloud*) and month m, we

first define a vector $\vec{R}_{f,m}$ of length n_{time}, where the i^{th} element is the globally averaged value of the TOA flux f during month m of year i. If we subtract the average value of this vector from each element in the vector, we get the anomaly vector $\vec{R}'_{f,m}$. Assuming the statistical model defined in Equation 5.1, we can then perform a linear least squares fit to solve for the global spatial feedbacks ($\vec{\lambda}_f$):

$$\vec{\lambda}_f = \begin{bmatrix} \lambda_{f,1} \\ \lambda_{f,2} \\ \vdots \\ \lambda_{f,n_{grid}} \end{bmatrix} = (\mathbf{T}'^T_m \mathbf{T}'_m)^{-1} \mathbf{T}'^T_m \vec{R}'_{f,m} \tag{7.7}$$

Full version

To estimate the full spatial feedbacks associated with a TOA radiative flux of type f (where f is either *net*, *LW clear*, *SW clear*, *LW cloud*, or *SW cloud*) and month m, we first define an $n_{time} \times n_{grid}$ matrix of anomalies $\mathbf{R}'_{flux,m}$, which is the same as $\mathbf{T}'_m$ above, except instead of surface temperatures, TOA radiative fluxes of type f are considered. Assuming the statistical model defined in Equation 5.2, we can then perform a linear least squares fit to solve for the full spatial feedbacks (Λ_f):

$$\Lambda_f = \begin{bmatrix} \lambda_{f,1,1} & \lambda_{f,1,2} & \cdots & \lambda_{f,1,n_{grid}} \\ \lambda_{f,2,1} & \lambda_{f,2,2} & \cdots & \lambda_{f,2,n_{grid}} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{f,n_{grid},1} & \lambda_{f,n_{grid},2} & \cdots & \lambda_{f,n_{grid},n_{grid}} \end{bmatrix} = (\mathbf{T}'^T_m \mathbf{T}'_m)^{-1} \mathbf{T}'^T_m \mathbf{R}'_{f,m} \tag{7.8}$$

If $\vec{R}'_{f,m}$ is simply the spatial (area-weighted) average of $\mathbf{R}'_{f,m}$, then $\vec{\lambda}_f$ will be the spatial average of Λ_f.

7.2.2 Estimating the forced response

Forced feedbacks

Suppose that we have a $n_{time,forced}$-year long forced simulation of a GCM for which we have spatial feedbacks estimated from a control run. We then define an early period (years 6 to 50) and a late period (years 51 to $n_{time,forced}$). The true feedbacks $\lambda_{f,p}$ for this forced simulation during each period p (where p is either *early* or *late*) is defined as the slope of the least squares fit of the linear regression of the time series of globally and annually averaged TOA flux anomalies of type f from the forced simulation ($\vec{R}'_{f,forced,p}$), against the globally and annually averaged surface temperature anomalies from the forced simulation $\vec{T}'_{forced,p}$ (where $\vec{R}'_{f,forced,p}$ and $\vec{T}'_{forced,p}$ are vectors with as many entries as years in the time period p, and the $forced$ subscript distinguishes them from the control simulation time series above):

$$\lambda_{forced,f,p} = \frac{\vec{T}'_{forced,p} \cdot \vec{R}'_{f,forced,p}}{\|\vec{T}'_{forced,p}\|^2} \tag{7.9}$$

We can make estimates of these feedbacks using the MR method by first estimating the forced simulation's TOA radiative flux of type f for each month of the year m by multiplying the surface temperature time series of that forced simulation for that month, $\mathbf{T}'_{m,forced}$ (a $n_{time,forced} \times n_{grid}$ matrix) by either the global or full spatial feedbacks:

$$\hat{\vec{R}}'_{f,m,forced} = \mathbf{T}'_{m,forced}\vec{\lambda}_f \tag{7.10}$$

$$\hat{\mathbf{R}}'_{f,m,forced} = \mathbf{T}'_{m,forced}\Lambda_f \tag{7.11}$$

These monthly time series $\hat{\vec{R}}'_{f,m,forced}$ and $\hat{\mathbf{R}}'_{f,m,forced}$ can then be turned into annual averages $\hat{\vec{R}}'_{f,forced}$ and $\hat{\mathbf{R}}'_{f,forced}$, allowing us to estimate the feedbacks for period p by taking the subset of these time series that fall in that period, and then performing the same least squares fit:

$$\hat{\lambda}_{forced,f,p} = \frac{\vec{T}'_{forced,p} \cdot \hat{\vec{R}}'_{f,forced,p}}{\|\vec{T}'_{forced,p}\|^2} \tag{7.12}$$

Forced feedback uncertainty analysis

5-95% confidence intervals on true feedbacks in Figure 5.4 and 5.9 are calculated by assuming that noise is normally, identically, and independently distributed, in which case the intervals are:

$$\lambda_{forced,f,p} \pm \frac{t(n_{time,forced,p} - 2, 0.975)\|\vec{R}'_{f,forced,p} - \lambda_{forced,f,p}\vec{T}'_{forced,p}\|}{\|\vec{T}'_{forced,p}\|^2} \tag{7.13}$$

where t is Student's t distribution.

5-95% confidence intervals on estimated feedbacks in Figures 5.4 and 5.9 are calculated using bootstrapping. Specifically, spatial feedbacks $\vec{\lambda}$ are recalculated ten thousand times (except for GISSE2R for local/nonlocal feedbacks in 5.9, where the recalculation is performed a thousand times), with each recalculation drawing upon a set of n_{grid} surface temperature time series randomly chosen with replacement from all n_{grid} grid cells. (In other words, each draw will likely have multiple repeated grid cells, and leave some grid cells out.) Each resulting set of spatial feedbacks $\vec{\lambda}$ is applied as in Equations 7.10 and 7.12 to generate an estimated feedback $\hat{\lambda}_{forced,f,p}$. We then use a histogram of the resulting values of $\hat{\lambda}_{forced,f,p}$ to estimate the 5^{th} and 95^{th} percentiles.

Spatial patterns of TOA flux change

We can quantify the spatial pattern of TOA radiative flux change of flux type f across a period p by taking the mean value of $\vec{R}_{f,forced}$ during the first half of the period and subtracting it from the second half, and then dividing this by the average change in the

globally averaged surface temperature between these two periods:

$$\Delta\vec{\tilde{R}}_{f,forced,p} = \left(\sum_{i=t_{mid,p}}^{t_{end,p}} \begin{bmatrix} R_{f,forced,i,1} \\ R_{f,forced,i,2} \\ \vdots \\ R_{f,forced,n_{grid}} \end{bmatrix} - \sum_{i=t_{start,p}}^{t_{mid,p}} \begin{bmatrix} R_{f,forced,i,1} \\ R_{f,forced,i,2} \\ \vdots \\ R_{f,forced,n_{grid}} \end{bmatrix} \right) / \left(\sum_{i=t_{mid,p}}^{t_{end,p}} T_{forced,i} - \sum_{i=t_{start,p}}^{t_{mid,p}} T_{forced,i} \right) \quad (7.14)$$

where $t_{start,p}$ and $t_{end,p}$ are the first and last years in period p, respectively, where $t_{mid,p} - (t_{end,p} + t_{start,p} - 1)/2$ is the midpoint of this time period, where $R_{f,forced,i,j}$ is the element in the i^{th} row and j^{th} column of $\mathbf{R_{f,forced}}$, and where $T_{forced,i}$ is the i^{th} element in $\vec{T}_{forced}$. The multi-model mean change in the *net* TOA flux during the early period is given in the left panel of Figure 5.5.

We can estimate these spatial patterns using our spatial feedbacks by performing the same calculation as in Equation 7.14, except replacing $R_{f,forced,i,j}$ by $\hat{R}_{f,forced,i,j}$ as derived in Equation A6. An example of the estimate of the *net* TOA flux during the early period is given in the right panel of Figure 5.5.

7.2.3 Errors

We calculate three types of errors: feedback error (early and late), feedback difference error, and spatial error (early and late). We now add a subscript g to our feedbacks and spatial patterns of TOA flux change to signify that they belong to the GCM g, where g is one of CCSM3, CESM104, GISSE2R, HadCM3L, IPSL, and MPIESM12. The feedback error for flux f and period p is then given by the root mean square error:

$$\epsilon_{feedback,f,p} = \sqrt{\frac{1}{n_{GCMs}} \sum_{g \in GCMs} (\hat{\lambda}_{f,forced,p,g} - \lambda_{f,forced,p,g})^2} \quad (7.15)$$

where n_{GCMs} is the number of GCMs (in this case, 6). The feedback difference error measures the method's skill at estimating the change in feedback between the early and late period:

$$\epsilon_{difference,f} = \sqrt{\frac{1}{n_{GCMs}} \sum_{g \in GCMs} ((\hat{\lambda}_{f,forced,2,g} - \hat{\lambda}_{f,forced,1,g}) - (\lambda_{f,forced,2,g} - \lambda_{f,forced,1,g}))^2} \quad (7.16)$$

The spatial error for each flux f and period p is measured by taking the area-weighted root mean square error of the spatial estimate above for each GCM g, and then taking the average resulting error:

$$\epsilon_{spatial,f,p} = \frac{1}{n_{GCMs}} \sum_{g \in GCMs} \sqrt{\frac{\sum_{i=1}^{n_{grid,g}} (\Delta \hat{\vec{\tilde{R}}}_{f,forced,p,i,g} - \Delta \vec{\tilde{R}}_{f,forced,p,i,g})^2 a_{i,g}}{\sum_{i=1}^{n_{grid,g}} a_{i,g}}} \quad (7.17)$$

where $a_{i,g}$ is the area of GCM g's i^{th} grid cell.

7.2.4 Other methods

We consider two other methods for deriving spatial feedbacks, and subsequently for estimating forced feedbacks and spatial patterns of TOA flux change (and calculating associated errors using the equations of the previous section):

The "global" method

In order to estimate the forced feedback for TOA flux f using the "global" method, we perform a least squares fit on the regression in Equation 5.3:

$$\lambda_{global,f} = \frac{\vec{T}' \cdot \vec{R}'_f}{\|\vec{T}'\|^2} \quad (7.18)$$

and use the resulting $\lambda_{global,f}$ as our forced feedback estimate. The feedback difference estimate will always be 0. The normalized spatial pattern of TOA flux change in either period is simply given by the spatial feedbacks found by fitting Equation 5.4:

$$\vec{\lambda}_{global,f} = \frac{\mathbf{R}_f'^T \vec{T}'}{\|\vec{T}'\|^2} \tag{7.19}$$

The "local" method

We first use a least squares fit to estimate the spatial feedbacks given by the regression in Equation 5.5:

$$\vec{\lambda}_{local,f} = \begin{bmatrix} \lambda_{local,f,1} \\ \lambda_{local,f,2} \\ \vdots \\ \lambda_{local,f,n_{grid}} \end{bmatrix} = \begin{bmatrix} \frac{\vec{T}_1' \cdot \vec{R}_{f,1}'}{\|\vec{T}_1'\|^2} \\ \frac{\vec{T}_2' \cdot \vec{R}_{f,2}'}{\|\vec{T}_2'\|^2} \\ \vdots \\ \frac{\vec{T}_{n_{grid}}' \cdot \vec{R}_{f,n_{grid}}'}{\|\vec{T}_{n_{grid}}'\|^2} \end{bmatrix} \tag{7.20}$$

where $\vec{T}_i'$ and $\vec{R}_{f,i}'$ are the i^{th} rows of $\mathbf{T}'$ and $\mathbf{R}_f'$ respectively. We can then generate estimates of $\mathbf{R}_{f,forced}'$ by applying these spatial feedbacks to the forced surface temperature record

$$\hat{\mathbf{R}}_{f,forced,local}' = \begin{bmatrix} \lambda_{local,f,1} \vec{T}_{forced,1}' \\ \lambda_{local,f,2} \vec{T}_{forced,2}' \\ \vdots \\ \lambda_{local,f,n_{grid}} \vec{T}_{forced,n_{grid}}' \end{bmatrix} \tag{7.21}$$

and taking the global average to generate $\hat{\vec{R}}_{f,forced,local}'$. We can then apply these estimates to Equations 7.12 and 7.14 to estimate forced global feedbacks and spatial patterns of TOA flux change.

REFERENCES

Abbot, D. S. and Tziperman, E. (2008). A high-latitude convective cloud feedback and equable climates. *Quarterly Journal of the Royal Meteorological Society*, 134(630):165–185.

Andrews, T., Gregory, J. M., and Webb, M. J. (2014). The Dependence of Radiative Forcing and Feedback on Evolving Patterns of Surface Temperature Change in Climate Models. *Journal of Climate*, 28(4):1630–1648.

Andrews, T., Gregory, J. M., Webb, M. J., and Taylor, K. E. (2012). Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models. *Geophysical Research Letters*, 39(9).

Andrews, T. and Webb, M. J. (2017). The Dependence of Global Cloud and Lapse Rate Feedbacks on the Spatial Structure of Tropical Pacific Warming. *Journal of Climate*, 31(2):641–654.

Armour, K. C. (2017). Energy budget constraints on climate sensitivity in light of inconstant climate feedbacks. *Nature Climate Change*, 7(5):331.

Armour, K. C., Bitz, C. M., and Roe, G. H. (2012). Time-Varying Climate Sensitivity from Regional Feedbacks. *Journal of Climate*, 26(13):4518–4534.

Arrhenius, P. S. (1896). XXXI. 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):237–276.

Bloch-Johnson, J., Pierrehumbert, R. T., and Abbot, D. S. (2015). Feedback temperature dependence determines the risk of high warming. *Geophysical Research Letters*, 42(12):2015GL064240.

Boer, G. and Yu, B. (2003a). Climate sensitivity and response. *Climate Dynamics*, 20(4):415–429.

Boer, G. J., Hamilton, K., and Zhu, W. (2005). Climate sensitivity and climate change under strong forcing. *Climate Dynamics*, 24(7):685–700.

Boer, G. J. and Yu, B. (2003b). Climate sensitivity and climate state. *Climate Dynamics*, 21(2):167–176.

Brown, P. T., Li, W., Jiang, J. H., and Su, H. (2015). Unforced Surface Air Temperature Variability and Its Contrasting Relationship with the Anomalous TOA Energy Flux at Local and Global Spatial Scales. *Journal of Climate*, 29(3):925–940.

Budyko, M. I. (1969). The effect of solar radiation variations on the climate of the Earth. *Tellus*, 21(5):611–619.

Caballero, R. and Huber, M. (2013). State-dependent climate sensitivity in past warm climates and its implications for future climate projections. *Proceedings of the National Academy of Sciences*, 110(35):14162–14167.

Callendar, G. S. (1938). The artificial production of carbon dioxide and its influence on temperature. *Quarterly Journal of the Royal Meteorological Society*, 64(275):223–240.

Ceppi, P. and Gregory, J. M. (2017). Relationship of tropospheric stability to climate sensitivity and Earths observed radiation budget. *Proceedings of the National Academy of Sciences*, 114(50):13126–13131.

Charney, J. G., Arakawa, A., Baker, D. J., Bolin, B., Dickinson, R. E., Goody, R. M., Leith, C. E., Stommel, H. M., and Wunsch, C. I. (1979). *Carbon Dioxide and Climate: A Scientific Assessment*.

Chung, E.-S. and Soden, B. J. (2015). An Assessment of Direct Radiative Forcing, Radiative

Adjustments, and Radiative Feedbacks in Coupled OceanAtmosphere Models. *Journal of Climate*, 28(10):4152–4170.

Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., Gao, X., Gutowski, W. J., Johns, T., Krinner, G., Shongwe, M., Tebaldi, C., Weaver, A. J., and Wehner, M. (2013). Chapter 12 - Long-term climate change: Projections, commitments and irreversibility. In IPCC, editor, *Climate Change 2013: The Physical Science Basis. IPCC Working Group I Contribution to AR5*. Cambridge University Press, Cambridge.

Colman, R. and Hanson, L. (2017). On the relative strength of radiative feedbacks under climate variability and change. *Climate Dynamics*, 49(5-6):2115–2129.

Colman, R. and McAvaney, B. (2009). Climate feedbacks under a very broad range of forcing. *Geophysical Research Letters*, 36(1).

Colman, R. A. and McAvaney, B. J. (1997). A study of general circulation model climate feedbacks determined from perturbed sea surface temperature experiments. *Journal of Geophysical Research: Atmospheres*, 102(D16):19383–19402.

Colman, R. A., Power, S. B., and McAvaney, B. J. (1997). Non-linear climate feedback analysis in an atmospheric general circulation model. *Climate Dynamics*, 13(10):717–731.

Crook, J. A., Forster, P. M., and Stuber, N. (2011). Spatial Patterns of Modeled Climate Feedback and Contributions to Temperature Response and Polar Amplification. *Journal of Climate*, 24(14):3575–3592.

Dessler, A. E. (2010). A Determination of the Cloud Feedback from Climate Variations over the Past Decade. *Science*, 330(6010):1523–1527.

Dessler, A. E. (2012). Observations of Climate Feedbacks over 2000-10 and Comparisons to Climate Models. *Journal of Climate*, 26(1):333–342.

Feldl, N. and Roe, G. H. (2013). Four perspectives on climate feedbacks. *Geophysical Research Letters*, 40(15):4007–4011.

Forster, P. M. F. and Gregory, J. M. (2006). The Climate Sensitivity and Its Components Diagnosed from Earth Radiation Budget Data. *Journal of Climate*, 19(1):39–52.

Fourier, J.-B. J. (1827). Mémoire sur les températures du globe terrestre et des espaces planétaires. *Mémoires de l'Académie Royale des Sciences*, 7:569–604.

Goodwin, P. (2018). On the Time Evolution of Climate Sensitivity and Future Warming. *Earth's Future*, 6(9):1336–1348.

Graversen, R. G., Langen, P. L., and Mauritsen, T. (2014). Polar Amplification in CCSM4: Contributions from the Lapse Rate and Surface Albedo Feedbacks. *Journal of Climate*, 27(12):4433–4450.

Gregory, J. M., Ingram, W. J., Palmer, M. A., Jones, G. S., Stott, P. A., Thorpe, R. B., Lowe, J. A., Johns, T. C., and Williams, K. D. (2004). A new method for diagnosing radiative forcing and climate sensitivity. *Geophysical Research Letters*, 31(3):L03205.

Gregory, J. M., Stouffer, R. J., Raper, S. C. B., Stott, P. A., and Rayner, N. A. (2002). An Observationally Based Estimate of the Climate Sensitivity. *Journal of Climate*, 15(22):3117–3121.

Hansen, J., Lacis, A., Rind, D., Russell, G., Stone, P., Fung, I., Ruedy, R., and Lerner, J. (1984). Climate Sensitivity: Analysis of Feedback Mechanisms. In *Climate Processes and Climate Sensitivity*, pages 130–163. American Geophysical Union (AGU).

Hansen, J., Russell, G., Lacis, A., Fung, I., Rind, D., and Stone, P. (1985). Climate Response Times: Dependence on Climate Sensitivity and Ocean Mixing. *Science*, 229(4716):857–859.

Hansen, J., Sato, M., Ruedy, R., Nazarenko, L., Lacis, A., Schmidt, G. A., Russell, G., Aleinov, I., Bauer, M., Bauer, S., Bell, N., Cairns, B., Canuto, V., Chandler, M., Cheng, Y., Del Genio, A., Faluvegi, G., Fleming, E., Friend, A., Hall, T., Jackman, C., Kelley, M., Kiang, N., Koch, D., Lean, J., Lerner, J., Lo, K., Menon, S., Miller, R., Minnis, P., Novakov, T., Oinas, V., Perlwitz, J., Perlwitz, J., Rind, D., Romanou, A., Shindell, D., Stone, P., Sun, S., Tausnev, N., Thresher, D., Wielicki, B., Wong, T., Yao, M., and Zhang, S. (2005). Efficacy of climate forcings. *Journal of Geophysical Research: Atmospheres*, 110(D18).

Hegerl, G. C., Crowley, T. J., Allen, M., Hyde, W. T., Pollack, H. N., Smerdon, J., and Zorita, E. (2007). Detection of Human Influence on a New, Validated 1500-Year Temperature Reconstruction. *Journal of Climate*, 20(4):650–666.

Huybers, P. (2010). Compensation between Model Feedbacks and Curtailment of Climate Sensitivity. *Journal of Climate*, 23(11):3009–3018.

Jonko, A. K., Shell, K. M., Sanderson, B. M., and Danabasoglu, G. (2012). Climate Feedbacks in CCSM3 under Changing CO2 Forcing. Part II: Variation of Climate Feedbacks and Sensitivity with Forcing. *Journal of Climate*, 26(9):2784–2795.

Kamae, Y., Watanabe, M., Ogura, T., Yoshimori, M., and Shiogama, H. (2015). Rapid Adjustments of Cloud and Hydrological Cycle to Increasing CO2: a Review. *Current Climate Change Reports*, 1(2):103–113.

Klein, S. A., Hall, A., Norris, J. R., and Pincus, R. (2017). Low-Cloud Feedbacks from Cloud-Controlling Factors: A Review. *Surveys in Geophysics*, 38(6):1307–1329.

Klein, S. A. and Hartmann, D. L. (1993). The Seasonal Cycle of Low Stratiform Clouds. *Journal of Climate*, 6(8):1587–1606.

Knutti, R., Masson, D., and Gettelman, A. (2013). Climate model genealogy: Generation CMIP5 and how we got there. *Geophysical Research Letters*, 40(6):1194–1199.

Knutti, R., Rugenstein, M. A. A., and Hegerl, G. C. (2017). Beyond equilibrium climate sensitivity. *Nature Geoscience*, 10(10):727–736.

Köhler, P., Stap, L. B., von der Heydt, A. S., de Boer, B., van de Wal, R. S. W., and Bloch-Johnson, J. (2017). A State-Dependent Quantification of Climate Sensitivity Based On Paleodata of the Last 2.1 Million Years. *Paleoceanography*, 32:1102–1114.

Kopparapu, R. K., Ramirez, R., Kasting, J. F., Eymet, V., Robinson, T. D., Suvrath Mahadevan, Terrien, R. C., Domagal-Goldman, S., Meadows, V., and Deshpande, R. (2013). Habitable Zones around Main-sequence Stars: New Estimates. *The Astrophysical Journal*, 765(2):131.

Leconte, J., Forget, F., Charnay, B., Wordsworth, R., and Pottier, A. (2013). Increased insolation threshold for runaway greenhouse processes on Earth-like planets. *Nature*, 504(7479):268–271.

Lewis, N. and Curry, J. A. (2015). The implications for climate sensitivity of AR5 forcing and heat uptake estimates. *Climate Dynamics*, 45(3-4):1009–1023.

Liu, Z., Wen, N., and Liu, Y. (2008). On the Assessment of Nonlocal Climate Feedback. Part I: The Generalized Equilibrium Feedback Assessment. *Journal of Climate*, 21(1):134–148.

Lunt, D. J., Haywood, A. M., Schmidt, G. A., Salzmann, U., Valdes, P. J., and Dowsett, H. J. (2010). Earth system sensitivity inferred from Pliocene modelling and data. *Nature Geoscience*, 3(1):60–64.

Manabe, S. and Bryan, K. (1985). CO2-induced change in a coupled ocean-atmosphere model and its paleoclimatic implications. *Journal of Geophysical Research: Oceans*, 90(C6):11689–11707.

Manabe, S. and Wetherald, R. T. (1967). Thermal Equilibrium of the Atmosphere with a

Given Distribution of Relative Humidity. *Journal of the Atmospheric Sciences*, 24(3):241–259.

Marshall, D. P. and Zanna, L. (2014). A Conceptual Model of Ocean Heat Uptake under Climate Change. *Journal of Climate*, 27(22):8444–8465.

Mauritsen, T., Stevens, B., Roeckner, E., Crueger, T., Esch, M., Giorgetta, M., Haak, H., Jungclaus, J., Klocke, D., Matei, D., Mikolajewicz, U., Notz, D., Pincus, R., Schmidt, H., and Tomassini, L. (2012). Tuning the climate of a global model. *Journal of Advances in Modeling Earth Systems*, 4(3).

McCoy, D. T., Hartmann, D. L., and Zelinka, M. D. (2018). Mixed-Phase Cloud Feedbacks. Technical Report LLNL-BOOK-745709, Mixed-phase Clouds: Observations and Modeling, Mixed-Phase Cloud Feedbacks, Elsevier, Amsterdam, 2017, pp. 215-231.

Meraner, K., Mauritsen, T., and Voigt, A. (2013). Robust increase in equilibrium climate sensitivity under global warming. *Geophysical Research Letters*, 40(22):5944–5948.

Milankovitch, M. (1930). *Mathematische Klimalehre und astronomische Theorie der Klimaschwankungen.* Gebrder Borntraeger, Berlin. OCLC: 490063906.

Murphy, D. M., Solomon, S., Portmann, R. W., Rosenlof, K. H., Forster, P. M., and Wong, T. (2009). An observationally based energy balance for the Earth since 1950. *Journal of Geophysical Research: Atmospheres*, 114(D17).

Murphy, J. M. (1995). Transient Response of the Hadley Centre Coupled Ocean-Atmosphere Model to Increasing Carbon Dioxide. Part 1: Control Climate and Flux Adjustment. *Journal of Climate*, 8(1):36–56.

Nakajima, S., Hayashi, Y.-Y., and Abe, Y. (1992). A Study on the Runaway Greenhouse Effect with a One-Dimensional RadiativeConvective Equilibrium Model. *Journal of the Atmospheric Sciences*, 49(23):2256–2266.

Nowack, P. J., Abraham, N. L., Braesicke, P., and Pyle, J. A. (2018). The Impact of Stratospheric Ozone Feedbacks on Climate Sensitivity Estimates. *Journal of Geophysical Research: Atmospheres*, 123(9):4630–4641.

Otto, A., Otto, F. E. L., Boucher, O., Church, J., Hegerl, G., Forster, P. M., Gillett, N. P., Gregory, J., Johnson, G. C., Knutti, R., Lewis, N., Lohmann, U., Marotzke, J., Myhre, G., Shindell, D., Stevens, B., and Allen, M. R. (2013). Energy budget constraints on climate response. *Nature Geoscience*, 6(6):415.

Pierrehumbert, R. T. (1995). Thermostats, Radiator Fins, and the Local Runaway Greenhouse. *Journal of the Atmospheric Sciences*, 52(10):1784–1806.

Pierrehumbert, R. T. (2013). Hot climates, high sensitivity. *Proceedings of the National Academy of Sciences*, 110(35):14118–14119.

Po-Chedley, S., Armour, K. C., Bitz, C. M., Zelinka, M. D., Santer, B. D., and Fu, Q. (2018). Sources of intermodel spread in the lapse rate and water vapor feedbacks. *Journal of Climate*.

Popp, M. (2014). *Climate instabilities under strong solar forcing*. PhD thesis, Max-Planck-Institut fr Meteorologie.

Popp, M., Schmidt, H., and Marotzke, J. (2014). Initiation of a Runaway Greenhouse in a Cloudy Column. *Journal of the Atmospheric Sciences*, 72(1):452–471.

Popp, M., Schmidt, H., and Marotzke, J. (2016). Transition to a Moist Greenhouse with CO_2 and solar forcing. *Nature Communications*, 7:10627.

Proistosescu, C. and Huybers, P. J. (2017). Slow climate mode reconciles historical and model-based estimates of climate sensitivity. *Science Advances*, 3(7):e1602821.

Roe, G. H. and Armour, K. C. (2011). How sensitive is climate sensitivity? *Geophysical Research Letters*, 38(14).

Roe, G. H. and Baker, M. B. (2007). Why Is Climate Sensitivity So Unpredictable? *Science*, 318(5850):629–632.

Roe, G. H., Feldl, N., Armour, K. C., Hwang, Y.-T., and Frierson, D. M. W. (2015). The remote impacts of climate feedbacks on regional climate predictability. *Nature Geoscience*, 8(2):135–139.

Rose, B. E. J., Armour, K. C., Battisti, D. S., Feldl, N., and Koll, D. D. B. (2014). The dependence of transient climate sensitivity and radiative feedbacks on the spatial pattern of ocean heat uptake. *Geophysical Research Letters*, 41(3):1071–1078.

Rose, B. E. J. and Rayborn, L. (2016). The Effects of Ocean Heat Uptake on Transient Climate Sensitivity. *Current Climate Change Reports*, 2(4):190–201.

Rugenstein, M., Bloch-Johnson, J., Gregory, J., Andrews, T., Mauritsen, T., Li, C., Frölicher, T., Paynter, D., Danabasoglu, G., Yang, S., Dufresne, J.-L., Cao, L., Schmidt, G. A., Abe-Ouchi, A., and Knutti, R. (forthcoming). Approaching equilibrium in climate models under elevated forcing.

Rugenstein, M. A. A., Caldeira, K., and Knutti, R. (2016a). Dependence of global radiative feedbacks on evolving patterns of surface heat fluxes. *Geophysical Research Letters*, 43(18):9877–9885.

Rugenstein, M. A. A., Gregory, J. M., Schaller, N., Sedlek, J., and Knutti, R. (2016b). Multiannual OceanAtmosphere Adjustments to Radiative Forcing. *Journal of Climate*, 29(15):5643–5659.

Sagan, C. (1962). Structure of the lower atmosphere of Venus. *Icarus*, 1(1):151–169.

Sellers, W. D. (1969). A Global Climatic Model Based on the Energy Balance of the Earth-Atmosphere System. *Journal of Applied Meteorology*, 8(3):392–400.

Senior, C. A. and Mitchell, J. F. B. (2000). The time-dependence of climate sensitivity. *Geophysical Research Letters*, 27(17):2685–2688.

Shaffer, G., Huber, M., Rondanelli, R., and Pedersen, J. O. P. (2016). Deep time evidence for climate sensitivity increase with warming. *Geophysical Research Letters*, 43(12):6538–6545.

Sherwood, S. C., Bony, S., Boucher, O., Bretherton, C., Forster, P. M., Gregory, J. M., and Stevens, B. (2014). Adjustments in the Forcing-Feedback Framework for Understanding Climate Change. *Bulletin of the American Meteorological Society*, 96(2):217–228.

Soden, B. J., Broccoli, A. J., and Hemler, R. S. (2004). On the Use of Cloud Forcing to Estimate Cloud Feedback. *Journal of Climate*, 17(19):3661–3665.

Soden, B. J., Held, I. M., Colman, R., Shell, K. M., Kiehl, J. T., and Shields, C. A. (2008). Quantifying Climate Feedbacks Using Radiative Kernels. *Journal of Climate*, 21(14):3504–3520.

Stainforth, D. A., Aina, T., Christensen, C., Collins, M., Faull, N., Frame, D. J., Kettleborough, J. A., Knight, S., Martin, A., Murphy, J. M., Piani, C., Sexton, D., Smith, L. A., Spicer, R. A., Thorpe, A. J., and Allen, M. R. (2005). Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature*, 433(7024):403–406.

Strogatz, S. (1994). Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering. *Computers in Physics*, 8(5):532–532.

Tomassini, L., Voigt, A., and Stevens, B. (2015). On the connection between tropical circulation, convective mixing, and climate sensitivity. *Quarterly Journal of the Royal Meteorological Society*, 141(689):1404–1416.

Trenberth, K. E., Zhang, Y., and Fasullo, J. T. (2015). Relationships among top-of-

atmosphere radiation and atmospheric state variables in observations and CESM. *Journal of Geophysical Research: Atmospheres*, 120(19):2015JD023381.

Tyndall, J. (1861). XXIII. On the absorption and radiation of heat by gases and vapours, and on the physical connexion of radiation, absorption, and conduction.The bakerian lecture. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 22(146):169–194.

Watterson, I. G. (2000). Interpretation of Simulated Global Warming Using a Simple Model. *Journal of Climate*, 13(1):202–215.

Weitzman, M. L. (2011). Fat-Tailed Uncertainty in the Economics of Catastrophic Climate Change. *Review of Environmental Economics and Policy*, 5(2):275–292.

Wetherald, R. T. and Manabe, S. (1975). The Effects of Changing the Solar Constant on the Climate of a General Circulation Model. *Journal of the Atmospheric Sciences*, 32(11):2044–2059.

Wetherald, R. T. and Manabe, S. (1988). Cloud Feedback Processes in a General Circulation Model. *Journal of the Atmospheric Sciences*, 45(8):1397–1416.

Wood, R. and Bretherton, C. S. (2006). On the Relationship between Stratiform Low Cloud Cover and Lower-Tropospheric Stability. *Journal of Climate*, 19(24):6425–6432.

Yang, J., Cowan, N. B., and Abbot, D. S. (2013). Stabilizing Cloud Feedback Dramatically Expands the Habitable Zone of Tidally Locked Planets. *The Astrophysical Journal Letters*, 771(2):L45.

Yokohata, T., Webb, M. J., Collins, M., Williams, K. D., Yoshimori, M., Hargreaves, J. C., and Annan, J. D. (2010). Structural Similarities and Differences in Climate Responses to CO2 Increase between Two Perturbed Physics Ensembles. *Journal of Climate*, 23(6):1392–1410.

Yuan, X., Kaplan, M. R., and Cane, M. A. (2017). The Interconnected Global Climate SystemA Review of TropicalPolar Teleconnections. *Journal of Climate*, 31(15):5765–5792.

Zaliapin, I. and Ghil, M. (2010). Another Look at Climate Sensitivity. *Nonlinear Processes in Geophysics*, 17(2):113–122. arXiv: 1003.0253.

Zelinka, M. D., Klein, S. A., and Hartmann, D. L. (2012). Computing and Partitioning Cloud Feedbacks Using Cloud Property Histograms. Part I: Cloud Radiative Kernels. *Journal of Climate*, 25(11):3715–3735.

Zhou, C., Zelinka, M. D., and Klein, S. A. (2017). Analyzing the dependence of global cloud feedback on the spatial pattern of sea surface temperature change with a Green's function approach. *Journal of Advances in Modeling Earth Systems*, 9(5):2174–2189.